

Volumen 1
Número 2
Diciembre 2019

SERENGUETI

Revista de Estadística

“Una nueva propuesta para el aprendizaje”



Foto: Ricardo Alvarado B.

Universidad de Costa Rica, Escuela de Estadística
Andrea Vargas, Ericka Valerio, Joshua Salazar, Susana García

COMITÉ EDITORIAL

El comité está integrado por estudiantes de bachillerato de la carrera de Estadística de la Universidad de Costa Rica. Este cumple la función de verificar que los artículos que se publican en esta revista cumplan con los lineamientos establecidos para su inclusión. Así mismo el comité tiene entre sus funciones difundir y promocionar la revista en plataformas o medios de interés para la escuela de estadística y sus estudiantes.

De parte del personal del comité se agradece la dedicación y el esfuerzo realizado por todos los participantes para que el volumen 2 de la revista Serengueti fuese posible. Así mismo, nuevamente se extiende la invitación a la población estudiantil para que formen parte del proyecto, tanto a través del proceso de publicación de los artículos como en el seguimiento activo de los trabajos que serán seleccionados semestralmente.

Ericka Valerio Salas, Universidad de Costa Rica, c.e. *erickavs189@gmail.com*

Susana García Calvo, Universidad de Costa Rica, c.e. *asgc3093@gmail.com*

Andrea Vargas Montero, Universidad de Costa Rica, c.e. *avargas2398@gmail.com*

Joshua Salazar Obando, Universidad de Costa Rica, c.e. *joshua.salazar1692@gmail.com*

DEDICATORIA

A nuestro profesor Ricardo Alvarado Barrantes, nuestro profesor guía y ayuda incondicional, también a Fernando Ramírez Hernández, director de la Escuela de Estadística, quienes han sido un gran apoyo y consejero en este proceso.

Gracias por su sincera disposición para colaborar con todo lo que esté a su alcance para que este proyecto siga marchando. Gracias por escuchar y apoyar nuestras ideas.

Reflexión

En el Simposio Académico de Estudiantes de Estadística¹, realizado en marzo del año 2019, fue presentado el Volumen I de la Revista digital Serengueti con un contenido realizado por y para los estudiantes de la carrera de Estadística. Fue una sorpresa muy agradable para todas las personas que conformamos el equipo de trabajo de la Escuela de Estadística: la dirección, el personal administrativo, el personal docente y el estudiantado. Ese primer volumen fue la culminación del trabajo de cuatro estudiantes guiados por el profesor Ricardo Alvarado Barrantes y se concentró en el tema de diseños experimentales. La revista tiene su acceso directo en la página Web de la Escuela.

Para el Volumen II de la Revista Serengueti, los estudiantes, que conforman su comité editorial, me invitaron a participar en la sección inicial de la Revista, con un mensaje motivador para que los estudiantes participen y compartan el conocimiento adquirido en los cursos de la carrera de Bachillerato en Estadística, así como sus aplicaciones. Todo dentro del contexto de mi función como director, profesor y especialmente como profesional con experiencia en el campo de la consultoría estadística, tanto en el sector privado como público.

Es muy importante este tipo de iniciativas dentro del proceso de mejora continua en que está inmersa actualmente la Escuela de Estadística, con la acreditación de la carrera del Bachillerato en Estadística ante el SINAES (2016), pues se enmarca dentro de nuestros compromisos, tanto con la acreditación como con el Plan Estratégico quinquenal. Esto responde a algunas metas propuestas en dichos compromisos: mejorar la metodología de la enseñanza, la estrategia de comunicación con nuestros estudiantes, ampliar la divulgación de la labor de nuestros estudiantes en la aplicación de la Estadística y la internacionalización de la Escuela, entre otros.

En varios cursos de la carrera los estudiantes realizan una serie de actividades y proyectos con aplicaciones que permiten solucionar problemas a entidades públicas y privadas. Esto ha generado una relación de la Escuela con la comunidad nacional que debemos divulgar como una actividad de docencia-acción social, que muchas veces resulta invisibilizada.

Quizás los estudiantes no aprecien con claridad la importancia de una serie de actividades que realizan durante su estancia en la universidad, las cuales no se dan a conocer fuera del ámbito de los cursos. Esto lo he observado con relativa frecuencia en los cursos de Práctica Profesional que imparto en la Escuela de Estadística. En estos cursos los estudiantes se ubican en una entidad pública o privada y desarrollan un proyecto específico durante dos ciclos lectivos. Los proyectos que desarrollan son de índole muy diverso. Específicamente, durante este año, se concentran en: aplicaciones complejas de análisis de datos, desarrollo de modelos de riesgo, creación de metodologías de análisis, propuestas de evaluación, estimación de indicadores, segmentación de mercados, enlace de series históricas, propuesta de índices de precios específicos, análisis de métodos de imputación, comparación de diseños muestrales, análisis de asociación cruzada entre series de tiempo, conformación de

¹ El Simposio Académico de Estudiantes, que se inició en el año 2014, es un evento que se realiza anualmente y es organizado por la Asociación de Estudiantes de Estadística apoyado por la Escuela de Estadística.

escenarios para medición de impacto, propuesta de un registro administrativo de pescadores artesanales, estimaciones de poblaciones por medio de registros administrativos, desarrollo de sistemas de información como apoyo para la toma de decisiones, entre otros. Situación similar se presenta en otros cursos de la carrera como Modelos de Regresión Aplicados, Aplicaciones de Diseños Experimentales y Diseño de Encuestas por Muestreo. Lo anterior evidencia el potencial de publicaciones para la Revista Serengueti que tienen los proyectos realizados por nuestros estudiantes.

Por lo tanto, nuestro grupo estudiantil debe ser proactivo con sus actividades y proyectos en los cursos y el personal docente debe apoyarlos e incentivarlos, de tal forma que los mejores proyectos se publiquen en la Revista Serengueti, o bien, participen en eventos académicos internacionales.



M.Sc. Fernando Ramírez Hernández
Director de la Escuela de Estadística
Universidad de Costa Rica

Contenido

INTRODUCCIÓN	7
I. DISEÑOS EXPERIMENTALES FACTORIALES	8
Análisis del efecto de la condición de escritura utilizando el adverbio especialmente como partícula focalizadora en el tiempo promedio de segunda lectura	9
Comparación de preservadores caseros para la conservación postcosecha de la rosa freedom	19
Análisis de la influencia a la exposición a la lengua inglesa en el mejoramiento de la pronunciación de consonantes para estudiantes del inglés como segunda lengua.....	27
II. SIMULACIONES	37
Impacto del desbalance en los tamaños de muestra por tratamiento sobre el desempeño de la prueba de comparaciones múltiples de Tukey	38
Afectación de la cobertura de los intervalos de confianza para diferencias de medias al incumplir el supuesto de homoscedasticidad, un enfoque de tres alternativas para analizar datos heteroscedásticos	46
III. TÉCNICAS DE AGRUPAMIENTO	54
Relación entre actitudes hacia el amor y signo zodiacal	55
Análisis sobre tipos de inteligencia y pasatiempos de los estudiantes activos de las Facultades de Ciencias Sociales, Educación, Ciencias Económicas e Ingeniería de la Universidad de Costa Rica.....	65
IV. MINERÍA DE DATOS	74
Predicción si lloverá mañana en Australia a través de algunos métodos de clasificación	75
Comparación de métodos de clasificación para la calidad del vino mediante árboles de decisión, bosques aleatorios y K-vecinos más cercanos.....	85
AGRADECIMIENTOS.....	93

INTRODUCCIÓN

Se destaca, de manera general, que la revista Serengueti se compone como una herramienta capaz de incentivar la participación estudiantil, en el cual principalmente se tiene como fin evidenciar y reconocer los esfuerzos realizados por parte de los estudiantes activos de la carrera de Estadística mediante artículos que son calificados y recomendados por los profesores del curso respectivo. Se promueve la revista como modelo para desarrollar en un futuro propuestas de enseñanza que permitan brindar diferentes formas y técnicas de aprendizaje.

El volumen 2 se toma como eje temático el análisis multivariado, cuya finalidad es mediante un conjunto de métodos estadísticos y matemáticos, describir e interpretar los datos que provienen de la observación de variables estadísticas y que han sido analizadas simultáneamente (Cuadras, 2007, p.9)².

A partir de ello, se proponen algunas técnicas multivariantes como lo son el análisis factorial, técnicas de agrupamiento y minería de datos para predicción. Además, se realiza un apartado de simulaciones en el que se asocia al análisis factorial por la cercanía de los temas propuestos, pero la técnica de análisis se elabora desde un enfoque distinto.

² Cuadras, C. M. (2007). Nuevos Métodos de Análisis Multivariante. Revista Española de Quimioterapia: Publicación Oficial de La Sociedad Española de Quimioterapia, 20(3), 249. Recuperado de <http://www.ncbi.nlm.nih.gov/pubmed/19406528>



I. DISEÑOS EXPERIMENTALES FACTORIALES

Un diseño experimental incluye todos los elementos que se van a considerar dentro de la ejecución y análisis de un experimento, de forma específica, un experimento factorial tiene un diseño que consta de dos o más factores, donde cada uno tiene distintos valores o niveles y se tienen tratamientos que se componen de la combinación entre los niveles de los distintos factores. Estos diseños fueron utilizados por primera vez en el siglo XIX por Henry J. Gilbert y John Bennet Lawes de la Estación Experimental de Rothamsted (Montes, 2018)³.

³ Montes, C. (2018). "Análisis de diseños factoriales". Tecnológico Nacional de México-Instituto Tecnológico de Tapachula. Recuperado de academia.edu/37969293/Punto_No._1.Qué_es_un_diseño?auto=download



Análisis del efecto de la condición de escritura utilizando el adverbio especialmente como partícula focalizadora en el tiempo promedio de segunda lectura

Dayana Gómez Quesada⁴, Joshua Salazar Obando⁴, Andrea Vargas Montero⁴

dayana.gomezquesada@gmail.com, joshua.salazar1692@gmail.com, vargas2398@gmail.com

RESUMEN

Escribir correctamente implica no solo tener buena ortografía, es un proceso complejo donde el uso adecuado de cada palabra incluida puede hacer la diferencia entre comunicar un mensaje de forma completa y hacerlo de forma parcial. Los adverbios son parte de este proceso de decisión y en este caso se analiza el adverbio *especialmente*. Se busca contrastar el uso de este adverbio bajo dos condiciones: cuando este hace referencia a una palabra incluida en un listado previo y cuando hace referencia a una palabra que no se encuentra dentro de este listado, con el objetivo de encontrar si existe diferencia en el tiempo promedio de la segunda lectura por palabra al utilizar una condición u otra, es decir, encontrar cuál condición facilita el proceso de lectura y comprensión. Aunque se tiene la hipótesis de que la primera condición logra un proceso de lectura más eficiente, se plantea un modelo factorial y se realiza un análisis de varianza con el cual no se encuentran diferencias entre los tiempos promedios de segunda lectura de ambas condiciones. Sin embargo, se realiza un análisis más a fondo de lo obtenido y se plantean recomendaciones para estudios futuros dentro de este marco de investigación.

Palabras clave: especialmente, adverbio, diferencia de medias, tiempo de segunda lectura, partícula focalizadora

ABSTRACT

Writing correctly means not only good spelling, it's a complex process where the correct use of each of the words included may make the difference between communicating your message completely or partially. Adverbs are a part of this decision process and in this case we analyze the adverb *especially*. This article seeks to contrast the use of this adverb under two conditions: when it references a word included in a previous list and when it references a word that is not included in this list, with the goal of finding if there is a difference in the time of the second reading per word when using one condition of the other, i.e., finding which condition achieves a more efficient reading and comprehension process. Even though the hypothesis is that the first condition achieves better results, a factorial model is made and an analysis of variance is made where no differences are found between the reading times of the second lecture of both conditions. However, a more in-depth analysis is made and recommendations for future studies in the same research framework are proposed.

Keywords: especially, adverb, mean difference, second reading time, focusing particle

INTRODUCCIÓN

Escribir un texto implica la intención de comunicar un mensaje al público que lee, sin embargo, no siempre este es el resultado que se obtiene. Aprender a escribir es más que simplemente conocer los conceptos básicos de ortografía y redacción. Dominar la escritura de textos también implica evitar confusiones y malentendidos para poder comunicar correctamente el mensaje deseado. Toda expresión lingüística implica un proceso de elección; en ocasiones, esta se produce dentro de un paradigma de

⁴ Estudiantes de Estadística de la Universidad de Costa Rica



posibilidades (Portolés, 2009), en el cual depende si se logra o no comunicar el mensaje que se pretende. Es claro que escribir correctamente es un proceso complejo, con muchos factores que afectan y que se deben tomar en cuenta a la hora de escribir, desde componentes más “complejos” como las relaciones causales y los complementos (directos o indirectos), hasta los más “simples” como los adjetivos y los adverbios (entendiéndose simples como “de una sola palabra y “complejos” como formados por distintas palabras). Sin embargo, el uso correcto de estos últimos puede ser determinante para la escritura de un texto comprensible y que logre su objetivo. Los adverbios, por ejemplo, son palabras invariables cuya función propia es la de complementar a un verbo, a un adjetivo o a otro adverbio, aunque también puede incidir sobre grupos nominales, preposicionales o sobre toda una oración (RAE, 2005). Por lo que, aunque parezcan de poca importancia, realmente en general si inciden en la conformación de un texto comprensible.

Adicionalmente, la importancia de estos también radica en que el buen uso de los componentes simples es parte esencial del correcto uso de los componentes complejos. Y, si bien su uso parezca intuitivo o fácil, no necesariamente todos los usan correctamente, se deben estudiar más a fondo para poder conocer cuál es la mejor forma de emplearlos dentro de un texto ya sea educativo, académico o de entretenimiento. Es pertinente señalar que existen distintos tipos de adverbios, estos son: de lugar, de tiempo, de cantidad, de modo, de afirmación, de negación, interrogativos, de duda y relativos (Lingolia, s.f.).

El presente artículo pretende analizar el cambio que se puede dar en el proceso de lectura cuando se muestran despliegues de opciones, con el fin de verificar si enfatizar o contrastar estas opciones acompañado del adverbio modal *especialmente*, facilita la fluidez con la cual se lee y se comprende el texto. A su vez, identificar la diferencia en tiempo de procesamiento que se genera en la fase de lectura, a partir de la estructura con la cual se presenta el texto, mediante el acompañamiento del adverbio modal *especialmente*. Y por último, lograr sugerir la forma con la que *especialmente* colabora de mejor manera en la comprensión de textos, a través del tiempo de lectura promedio por palabra.

METODOLOGÍA

Se utiliza un diseño experimental de un factor con dos niveles, se tienen dos tratamientos con bloques. El factor de diseño es el tipo de oración, esto se refiere a textos en los cuales se incluye la partícula focalizadora, en este caso el adverbio *especialmente*, bajo distintas condiciones. Se trabaja el diseño experimental con 2 condiciones y los tratamientos son siguientes tipos de oraciones:

- **A:** oración donde se presenta un listado de palabras, seguido, entre otros componentes, por el adverbio *especialmente*, y seguido a este, una de las palabras incluidas en el listado inicial.
- **B:** oración donde se presenta un listado de palabras, seguido, entre otros componentes, por el adverbio *especialmente*, y seguido a este, una palabra que no estaba incluida en el listado inicial.

Es importante mencionar que, en ambos casos, la oración de interés, aquella que incluye el adverbio estudiado (*especialmente*), es precedida por una introducción al tema tratado en dicha oración y a su vez, posterior a ella se encuentra un cierre. Se presentan de esta forma los textos por tres razones:

1. Por medio de estudios previos, se conoce que las personas al comenzar a leer un texto no ponen la misma atención que al avanzar su proceso de lectura dentro de este.
2. Además, se utilizan las oraciones de cierre para evitar que el proceso lectura se afecte por una finalización rápida e imprevista, por ende, con esas oraciones finales la persona continúa leyendo de manera normal, sin alterar los resultados.
3. Se requiere que los participantes no conozcan el objetivo exacto del experimento, es decir, no pueden conocer que la evaluación se centraba en la oración que incluía *especialmente*, es una

forma de cubrir el componente exacto dentro del texto al que se le está poniendo especial atención debido a que esto puede causar un sesgo dentro de los participantes.

En cuanto a la variable respuesta, esta es el tiempo promedio de segunda lectura por palabra, se mide en milisegundos (ms) utilizando el software SMI BeGaze. De manera más específica, se puede obtener una distinta variable respuesta ya sea para el proceso completo, como es el tiempo promedio total de lectura por palabra, o para las diferentes fases del proceso de lectura que se puede dividir en los promedios de la primera y segunda lectura por palabra, todos en milisegundos. Sin embargo, se utiliza el tiempo promedio de la segunda lectura como variable respuesta debido a que el experto considera que en este es donde se debe reflejar mayoritariamente la diferencia.

Esta diferencia relevante, llamada nivel de resolución se determina que un 20% de diferencia entre los promedios de cada estímulo (A y B) se considera importante, a forma de ilustración, se podría decir que, si un tratamiento obtiene un promedio de 100 ms y el otro obtiene un promedio de 120 ms, ya esta diferencia se considera relevante (realizando su respectivo intervalo de confianza). Como previamente se mencionó, se utilizan bloques, es decir, cada unidad se ve expuesta a ambos tratamientos, en este caso se cuenta con 43 participantes, los cuales son personas letradas entre 18 y 25 años, tanto hombres como mujeres, todas hispanohablantes.

Cada persona lee 2 oraciones de cada tratamiento, es decir 2 por cada tipo de oración, los tiempos de estas dos lecturas se suman y se dividen por la cantidad de palabras importantes a tomar en cuenta dentro de las oraciones. Se obtiene un promedio por tratamiento para cada persona, es decir, se obtienen 43 datos por tratamiento.

Para realizar la investigación se utiliza el siguiente modelo:

$$\mu_{ij} = \mu + \alpha_i + \gamma_j$$

Donde representa el efecto del factor de diseño sobre la variable respuesta, es decir el efecto de la condición del tipo de oración que se presenta sobre el tiempo promedio de lectura por palabra dentro de cada oración, por lo que i puede ser 1 o 2 pues se tiene dos tipos de oraciones, las cuales se especifican más adelante. Además representa el efecto del bloque sobre el tiempo promedio de lectura por palabra, en este caso j puede tomar valores de 1 a 43 ya que se trabaja con 43 bloques (personas).

Cabe mencionar que dentro del modelo que se utiliza no se considera desde un inicio la interacción entre los dos factores, esto debido a que cuando se trabaja con bloques se toma este supuesto, comúnmente se realiza esto ya que se supone que los bloques son unidades las cuales se seleccionan aleatoriamente de un conjunto mayor donde no interesa dar conclusiones específicas por bloque, sino de una manera global.

Los bloques se incluyen en el modelo con el objetivo de extraer la variabilidad que por naturaleza tienen estos, es decir, se reconoce que los bloques producen variabilidad porque tienen distintas características entre ellos que pueden influir en la variable respuesta. Para esta investigación, quiere decir que se incluyen los bloques dentro del modelo ya que se reconoce que las personas tienen ciertas características como sexo, edad, gusto por la lectura, velocidad a la que leen, actitud ante la realización de la prueba, entre otras cosas, que pueden llegar a introducir ruido en los resultados, lo cual puede causar dificultad para ver diferencias entre los promedios de la variable respuesta para cada tratamiento cuando sí las hay.

La investigación se realiza en conjunto con el Dr. Adrián Vergara Heidke, profesor de la Escuela de Filología, Lingüística y Literatura e investigador del Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica. Se utiliza un dispositivo llamado "eye tracker" (nombre en inglés) para la

recolección de datos, el cual es un monitor especial que lanza rayos infrarrojos a los ojos del que está mirando la imagen sujeta de análisis para extraer información de este analizando sus movimientos oculares (Canelles, 2012).

Por la naturaleza del estudio, se trabaja con personas voluntarias, las cuales se contactan de distintas formas. El día en que se aplica el experimento cada participante debe firmar una serie de documentos donde se da el consentimiento formalmente para ser parte del estudio. Seguidamente el participante ingresa a una cabina con aislante de ruido y se coloca en frente de una computadora. A su lado se encontraba el Dr. Adrián Vergara Heidke el cual le brindaba las instrucciones del caso, entre las que cabe mencionar:

- Sentarse con buena postura (erguido)
- No realizar un movimiento de ningún tipo
- Realizar el proceso de lectura solo con el movimiento ocular sin girar la cabeza.

Al terminar la etapa de instrucciones se procede con la etapa de calibración. Esta parte del experimento resulta de suma importancia, ya que el dispositivo de medición “*eye tracker*” necesita acoplarse a los ojos de cada persona para poder realizar su seguimiento de forma correcta. Cabe mencionar que no todos los participantes obtienen resultados adecuados en esta primera etapa, es decir, que el dispositivo no ubica correctamente el movimiento ocular, por lo cual dichos participantes deben ser excluidos de la muestra que se tiene inicialmente.

Una vez que se completa la fase de calibración, da inicio el experimento como tal, donde se le muestra a cada participante una serie de oraciones donde se incluyen las oraciones de interés. Simultáneo a esto, se vigila ciertas métricas importantes por parte del experto Dr. Vergara, por ejemplo, el “Tracking Ratio”, el cual es una medida de la precisión con la que se recolecta los datos de cada participante en las distintas partes del texto. Al finalizar el proceso de lectura, los resultados se guardan y se deja listo para el siguiente participante.

Se pretende observar la diferencia en el tiempo de lectura que genera los dos tratamientos a aplicar, con la partícula focalizadora *especialmente*, y se espera que la persona relacione que la opción enfatizada que se menciona dentro de las opciones, y por tanto el tiempo promedio de segunda lectura por palabra sea menor. En el caso alternativo en el cual se enfatiza, con el mismo adverbio, una opción relacionada la cual no se encuentra incluida en el texto y se espera que para la persona sea necesario una segunda lectura, esta puede ser más exhaustiva, lo cual incrementa la variable respuesta.

Se pone a prueba esta hipótesis, realizando un Análisis de Varianza (ANDEVA o ANOVA por sus siglas en inglés), el cual consiste en analizar si la variabilidad de la variable respuesta, en este caso el tiempo promedio de segunda lectura por palabra, está mayormente explicada por los tratamientos propuestos o por aquello que no se logró controlar en el experimento, es decir factores que no se tomaron en cuenta y que están afectando, esta se conoce como la variabilidad residual.

Para poder realizar este análisis de varianza, es necesario comprobar ciertos supuestos que se deben cumplir, estos son normalidad, homocedasticidad e independencia. Para el caso de la normalidad, se realiza la prueba Kolmogorov-Smirnov. Para probar la homocedasticidad se realiza la prueba Barlett. Y, por último, en el caso de la independencia, no se utiliza ninguna prueba formal, solamente se analiza si las observaciones obtenidas son independientes entre ellas, es decir, una respuesta no depende de la respuesta anterior ni de la posterior.

Se utiliza el software para realizar medición inicial fue SMI BeGaze. Mientras que el análisis fue realizado utilizando el lenguaje *R* (R Core Team, 2018), con la versión 3.6.0 con las librerías *ggplot2*

(Wickham, 2016), lattice (Sarkar, 2008), car (Fox & Weisberg, 2019) y readxl (Wickham & Bryan, 2019).fbeheh

RESULTADOS

Un modelo de esta naturaleza puede tener como variable respuesta el tiempo de lectura promedio lectura, segunda lectura o tiempo total (tiempo promedio de la primera y segunda lectura). Dado la revisión de estudios previos en conjunto con la decisión del experto se procede a realizar el análisis más en concreto a lo que sucede en el tiempo de relectura o de segunda lectura, puesto que en esta es donde usualmente se presentan mayores diferencias. Antes de entrar de lleno con la cuantificación de resultados extraídos del análisis experimental se precisa la verificación de los supuestos del modelo.

- 1) **Normalidad:** en la figura 1 (ver anexo) se aprecia como la mayoría de los puntos están distribuidos cerca de la línea teórica. También, se recurre a la prueba de Kolmogorov-Smirnov y se confirma el supuesto debido a que el valor p obtenido es de 1.
- 2) **Homocedasticidad:** se realiza la prueba Bartlett, de homogeneidad de varianzas, la cual indica que si se cumple una varianza constante para cada tratamiento, ya que se obtiene un valor p de 1.
- 3) **Independencia:** se asume dicho supuesto.

Ahora bien, se obtiene a través de la aplicación del experimento para el análisis del tiempo de lectura utilizando el uso del adverbio modal *especialmente* dentro de las oraciones de dos maneras distintas. En la tabla 1 (ver anexo) se tiene las medias de cada estímulo, en las cuales se evidencia un cierto distanciamiento, y de manera correspondiente para lo que se esperaba el estímulo B tiene una media mayor, esto porque las personas al no observar que la palabra posterior a *especialmente* no está dentro del listado, se prevé que van a tener que releer este listado para ratificar si lo que leyeron se está comprendiendo de la manera correcta. Caso contrario si se nombra el listado después de la partícula focalizadora se deduce que va a tener menor tiempo puesto que está utilizando la memoria de corto plazo, ese listado debe de estar presente durante la lectura.

Para verificar si existe diferencias entre estas dos condiciones se obtiene un valor p de 0.0962, el resultado concretamente haciendo uso del contraste de hipótesis posiciona a favor la hipótesis nula, diciendo que se debe asumir la igualdad de medias entre estos dos tratamientos, se asume que no hay un efecto del tipo de oración sobre la variable respuesta. Al obtener un resultado como este es necesario evaluar la potencia de la prueba realizada y se obtiene una potencia de 0.91, esto quiere decir que la prueba realizada tiene una probabilidad de detectar diferencias entre los dos tratamientos de 0.91, cuando realmente existen diferencias. Es decir, la prueba es sumamente capaz de encontrar las diferencias si es que las hay, lo cual resulta sorprendente, ya que como lo menciona el experto basándose en estudios previos, se espera que el tiempo promedio de la segunda lectura sea considerablemente menor cuando a las personas se les presenta un listado de palabras y luego se referencia mediante una palabra, como *especialmente*.

A partir de ello, se revisan diversos estudios en donde se muestra que experimentos con características similares han encontrado diferencias entre las condiciones que contemplan utilizando una palabra focalizadora distinta. Por tal motivo se ve la necesidad de buscar alternativas al respecto, de manera que estas respalden la decisión tomada anteriormente o se tenga que llevar a cabo una discusión más profunda que permita evidenciar diferencias.

Anteriormente se menciona que se prioriza el tiempo de la segunda lectura para buscar diferencias entre los tipos de oraciones, y se observa que pasa lo contrario. Bajo ese panorama se recurre hacer una revisión complementaria tomando un modelo con tiempo promedio de la primera lectura y otro con el tiempo total, con el mismo factor de diseño y las personas como bloques. A medida que se revisa cada uno

de ellos se aprecia que el tiempo de la primera lectura presenta valores extremos, los cuales inciden considerablemente en la estimación, estos hacen que el modelo no pudiese asumir la normalidad. Detalladamente se revisa estos valores y se da con que es un mismo participante que arroja estos resultados. Con el propósito de hacer una revisión preliminar se decide momentáneamente separar a este participante y revisar los dos modelos nuevos y el modelo prioritario que es tomando la segunda lectura. Al hacerlo en definitiva se aprecia que el supuesto de normalidad se cumple. Además, se puede asumir el supuesto de homocedasticidad en los tres modelos.

En cuanto al efecto de la condición se evidencia un aspecto muy particular y es que en este caso sí se encuentran diferencias por tipo de oración cuando se toma el modelo con el promedio de la primera lectura como variable respuesta. No así, si hacemos esto en el caso del promedio de la segunda lectura, donde al igual que en análisis original no se presentan diferencias. Además, particularmente se tiene dos posiciones distintas, al hacer el modelo con el tiempo total de lectura se concluye que la condición no tiene un efecto sobre la respuesta. Sin duda esto genera controversia, la primera lectura no posee un respaldo teórico de encontrar diferencias, no quiere decir que no las haya, simplemente se piensa que es más difícil encontrar las diferencias en la primera lectura que la segunda. Sin embargo, resulta importante mencionar que, en el caso de este análisis, la potencia de la prueba era de 0.99, por lo que en caso de haber diferencias era casi seguro que se iban a encontrar, y efectivamente así fue.

Se hace una cuantificación de la diferencia para la primera lectura y como se observa en la tabla 2 (ver anexo), se encuentran diferencias significativas entre los tratamientos para el promedio de primera lectura, pero estas no resultan relevantes para contexto del problema porque según el Dr. Adrián Vergara, para que una diferencia se considere importante, esta debe ser de 20%, esto quiere decir que por ejemplo, si se obtiene un promedio de 200 para una de las condiciones y en el otro un 240, ya se considera relevante. En este caso para que clasificara así la diferencia el intervalo de debía tener al menos como límite inferior 207.6, esto representa aproximadamente el 20% del tratamiento inicial, donde la palabra posterior a *especialmente* es referenciada anteriormente en un listado.

No obstante, no es así por lo que no se encuentra una diferencia relevante en términos del problema a estudiar. Este resultado si coincide más con lo que se espera en el caso de la segunda lectura, casualmente el experto considera que este análisis no es el más revelador puesto que la diferencia en los tiempos promedios de este caso no resulta tan importante o interesante. Probablemente sería útil, profundizar y analizar desde distintas perspectivas el porqué de la diferencia en significancia en la diferencia de las dos comparaciones (primera y segunda lectura). En el caso de la segunda lectura los resultados que se obtienen en este análisis complementario son similares a lo que tiene el análisis original. No se encuentran diferencias, pero de cierta manera la prueba si es capaz de encontrar esas diferencias.

Por otro lado, en la primera lectura hay diferencias, va en contra de lo que usualmente se presenta, pero muy posiblemente esto tiene un trasfondo más aplicable en relación con lo visto en el análisis de cada persona. Se comenta de esta manera, debido a que en este caso la comprensión como tal se queda en la primera lectura, no se ve la necesidad en muchos participantes de hacer una relectura. Esto a su vez hace que para la primera lectura se mantenga más la atención de lo expuesto en el experimento que en la propia segunda lectura, lo que resulta bastante llamativo y de una u otra forma difiere en cuanto a la importancia que se propone anteriormente por el experto referente a la primera lectura. Además, se obtiene que, con el tiempo total no se encuentran diferencias.

Ahora bien, observando la potencia de la prueba para este modelo, se aprecia en la tabla 2 (ver anexo) que es prácticamente 1, lo cual es concluyente con lo que se tiene previamente, en donde no existen diferencias, pero la prueba como tal si las está encontrando. En el caso de la segunda lectura se ratifica lo que propone el análisis original y ese enfoque no debe perderse de vista.

En la descripción de la prueba se puede hacer mención que cada persona leyó los dos tratamientos y por esto son partícipes en el modelo con la estructura de bloques. Sin embargo, este grupo de personas fue separado por especificación del experto, en donde se conformó para el experimento con tal de beneficiar el análisis posterior en el que se van a profundizar otras características y que no son valoradas para el análisis que el grupo de trabajo realiza. Se propiciaron dos listados en los cuales un grupo tuvo que leer más que otro, eso sí para ambos casos tenían las oraciones de interés y que son parte de los niveles del factor de diseño de este trabajo. Hablar del método de “parcelas divididas” es un tema por el cual puede prever cierta información direccionado propiamente con la segunda lectura, específicamente con el efecto que tiene el tipo de oración sobre la variable respuesta.

El porqué de la incursión de este concepto radica en que las personas estuvieron leyendo ambos tratamientos, pero no específicamente en las mismas condiciones, refiriéndose al proceso de lectura reflejado tanto en el tiempo, en la cantidad de frases, entre otros. De estos dos grupos el segundo fue el que leyó menos oraciones. Se elaboran dos modelos, tomando en este caso la misma variable respuesta, pero ahora va diferenciarse de acuerdo con los dos grupos de personas por separado. Seguidamente se procede a la revisión del efecto del estímulo y de nuevo resulta no ser significativo en ambos modelos. Tomando las condiciones por separado de la misma forma que en el análisis original no se encuentran diferencias. Cabe reiterar que el análisis para este método resulta más amplio y que para estudios posteriores se debe profundizar el análisis, ya que, si se obtienen resultados distintos para cada modelo, se podría sospechar que existe interacción entre la condición de escritura (A y B) y la lista de oraciones utilizada.

Por otra parte, en algunos casos durante la prueba, los participantes dejan de mostrar interés al momento de leer, es decir, solo escanean el texto, pero no se intenta tener una comprensión clara. Por contraparte, ciertos participantes se enfocan en retener el texto lo más que se pueda y para esto releen ambos tipos de oraciones en reiteradas veces, que igualmente afecta el tiempo que tardan en realizar la prueba, y esto está incidiendo en los resultados. En muchas ocasiones se debe a que los participantes no conocen el objetivo del experimento, por lo cual interpretan distintas cosas, como, por ejemplo, que se les aplica preguntas de comprensión al finalizar la prueba o al contrario, que deben leer lo más rápido posible. Estas distintas interpretaciones pueden causar que los resultados obtenidos no reflejan realmente lo que se pretende. Entre otros aspectos que afectan la prueba son la postura y movimiento de las personas, ya que como previamente se menciona, sentarse completamente erguido durante toda la prueba es de suma importancia para una buena medición, por consiguiente, puede que la medición en ciertas etapas no resulta con la mayor precisión.

Se realiza el análisis desde distintas perspectivas, pero el resultado del análisis inicial se mantiene, no existen diferencias en el tiempo de promedio de lectura cuando se utilizan las dos distintas condiciones de escritura con el adverbio *especialmente*. Esto no concuerda con estudios de la misma naturaleza, como el caso de los experimentos con otras partículas gramaticales efectuados por miembros del laboratorio de Lenguaje y Cognición de la Universidad de Heidelberg (HULC por sus siglas en inglés), Alemania, uno de estos, por ejemplo, con la conexión “*por tanto*” donde se encuentra un efecto de aceleración en el segundo segmento del tiempo total de lectura (Loureda, O., Nadal, L., & Recio Fernández, I., 2016). Sin embargo, aunque parezca extraño puede que, en el caso específico de *especialmente*, no afecte tanto la forma en cómo se trabaje con dicha palabra, es decir, para el lector resulta igual de entendible cualquiera de los métodos analizados en este estudio.

Lo anterior puede deberse a que este adverbio resulta fácil de entender y la persona, aunque lea un elemento que no se encuentra en el listado previo, pueda entender con facilidad que lo anterior tiene una justificación lógica y logre mantener el hilo conductor de la idea que expone la oración. Por ejemplo, al leer: “Marianela y Constanza compraron naranjas, mandarinas y manzanas. A Marianela le gustaron

especialmente los mangos y Constanza preparó un delicioso batido para ambas”, la persona puede entender que, aunque los mangos no están en el listado, tiene sentido que se vean aludidos de esta forma, puede ser que, aunque estos son los más gustados por Marianela no se compra porque, por ejemplo, tienen un precio muy elevado. Este resultado es valioso, debido a que se puede demostrar que *especialmente* produce resultados distintos a otras partículas focalizadoras y para futuras investigaciones, si se obtienen los mismos resultados, se puede analizar a fondo las razones por las que esto ocurre.

CONCLUSIONES

Utilizar *especialmente* para referenciar una palabra que no se encuentra en un listado previo no produce un cambio considerable en el tiempo promedio de la segunda lectura de las personas. Esto quiere decir que a las personas no les resulta más difícil comprender un texto cuando una palabra que no está previamente en el texto es referenciada con este adverbio. Lo anterior resulta sorprendente ya que se pensaba que sí estaba afectando y se había demostrado con estudios previos de otras partículas focalizadoras. *Especialmente* tiene un comportamiento distinto a otras palabras previamente estudiadas y esto puede deberse a que este adverbio produce que las personas puedan realizar un análisis de la situación que están leyendo y puedan conectar las ideas sin necesidad de que estas se presenten tan directas o explícitas como se espera.

Se recomienda, para futuros estudios similares que se analice desde una perspectiva más teórica fundamentado desde un punto de vista más de la Lengua Española para poder explicar a fondo por qué esta partícula focalizadora presenta un comportamiento tan distinto al esperado y encontrado en otros casos, por ejemplo en el estudio de la partícula *por tanto*, realizado en la Universidad de Heidelberg, Alemania, por Elisa Narváez García.

En el caso del tiempo promedio de la primera lectura si se encuentran diferencias significativas, pero no relevantes, esto coincide con lo considerado con el experto de que este análisis comúnmente no muestra resultados importantes para este tipo de investigaciones. Sin embargo, cabe mencionar que si este resultado persiste en investigaciones similares se debe analizar el por qué, sobre todo si en algún caso se encuentran diferencias significativas e importantes, se debe examinar si en realidad existe un efecto de la condición de escritura sobre el tiempo promedio de la primera lectura.

Se recomienda que se realicen investigaciones de este tipo, mejorando la planificación del diseño e implementación del experimento, de manera que con esto mejore la calidad de los datos. Entre las posibles recomendaciones se encuentra brindar una explicación previa a cada participante, en la cual se le comunique de manera adecuada la forma en la que se llevará a cabo el experimento, una idea general de lo que se pretende, sin revelar cuál será el objetivo principal de la aplicación, esto para evitar que la persona sesgue sus resultados prestando más atención a la partícula en gramatical en estudio, como es el caso del enfoque que se le da al adverbio modal *especialmente*. Esto producto de que muchos participantes no sabían de qué se trataba el experimento y esto trajo consigo mucha confusión.

BIBLIOGRAFÍA

- Bravo, H. (2007) “El arte de escribir bien para comunicar mejor”. [Mensaje en un blog]. Recuperado de: <http://www.maestrosdelweb.com/el-arte-de-escribir-bien-para-comunicar-mejor/>
- Canelles, E. (2012). “¿Qué es el “Eye Tracking” y para qué nos sirve?”. Recuperado de: <https://www.solucionesc2.com/que-es-el-eye-tracking-y-para-que-nos-sirve/>
- Fox, J.y Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

Lingolia, (s.f.) “Tipos de adverbios en español”. Recuperado de: <https://espanol.lingolia.com/es/gramatica/adverbios/tipos>

Loureda, Ó., Nadal, L. y Recio, I. (2016). Partículas discursivas y cognición: por tanto y la conexión argumentativa. In: Romanistisches Jahrbuch, Vol 67, 1, de Gruyter Mouton, 240-254.

Narváez, E. (2019). Causality and its processing paths: an experimental study of the Spanish por tanto. Universidad de Heidelberg. [A publicar]

Parodi, G., Julio, C. y Recio, I. (2018). When words and graphs move the eyes: The processing of multimodal causal relations. Journal of Eye Movement Research,11(1). <https://doi.org/10.16910/jemr.11.1.5>

Portolés, J. (2009). Censura y pragmática lingüística. Circulo de lingüística aplicada a la comunicación, ISSN 1576-4737, No. 38, 2009.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.

RStudio Team. (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL: <http://www.rstudio.com/>.

Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham, H. y Bryan, J. (2019). readxl: Read Excel Files. R package version 1.1.0. <https://CRAN.R-project.org/package=readxl>

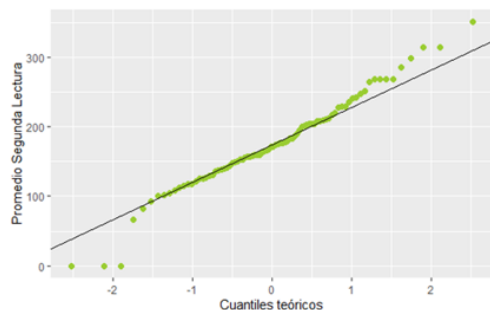
ANEXOS

Tabla 1. Resumen del modelo con el análisis original, (segunda lectura).

Condiciones de cada estímulo	Tiempo promedio de la segunda lectura	Intervalo de confianza para el tiempo promedio de la segunda lectura
A: contiene la opción a enfatizar con especialmente	166.5 ms	146.5 ms - 186.3 ms
B: no contiene la opción a enfatizar con especialmente	184.5 ms	164.9 ms - 203.9 ms
Intervalo de confianza para la diferencia de medias		
-3.338291 - 39.27507		
Potencia de la prueba		
0.909239		

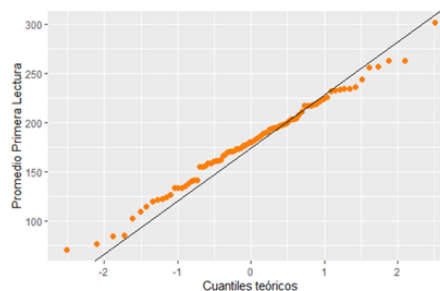
Fuente: Elaboración propia 2019.

Gráfico 1. Cuantiles teoricos de una distribución Normal contra cuantiles obtenidos con el Promedio Segunda Lectura



Fuente: Elaboración propia. 2019

Gráfico 2. Cuantiles teoricos de una dist. normal contra cuantiles obtenidos con el Promedio Primera Lectura eliminando los valores extremos



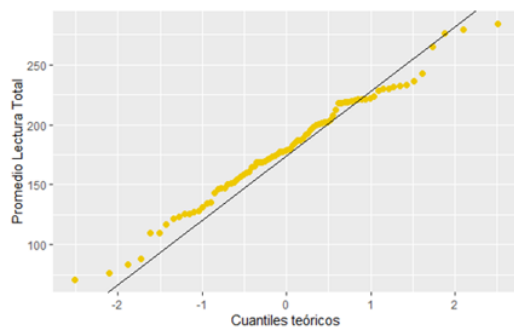
Fuente: Elaboración propia. 2019

Tabla 2. Resumen de los modelos una vez quitado los valores extremos para el análisis complementario

Condiciones de cada estímulo	Tiempo promedio total de lectura	Intervalo de confianza para el tiempo promedio total de lectura
A: contiene la opción a enfatizar con especialmente	338.5 ms	309.5 ms – 367.7 ms
B: no contiene la opción a enfatizar con especialmente	388.2 ms	339.4 ms – 437.1 ms
Intervalos de confianza para la diferencia de medias en el total de lectura		
-1.407585 - 22.32925		
Potencia de la prueba		
0.9999787		
Condiciones de cada estímulo	Tiempo promedio de la primera lectura	Intervalo de confianza para el tiempo promedio de la
A: contiene la opción a enfatizar con especialmente	171.9 ms	157.66 ms – 186.26 ms
B: no contiene la opción a enfatizar con especialmente	203.7 ms	164.8 ms – 242.65 ms
Intervalos de confianza para la diferencia de medias en la primera lectura		
0.8865 - 22.0471		
Potencia de la prueba *		
0.9999993		
*Se había rechazado la hipótesis de que las medias de los tratamientos son iguales		
Condiciones de cada estímulo	Tiempo promedio de la primera lectura	Intervalo de confianza para el tiempo promedio de la
A: contiene la opción a enfatizar con especialmente	166.4	146.1- 186.6
B: no contiene la opción a enfatizar con especialmente	183.1	163.4- 202.9
Intervalos de confianza para la diferencia de medias en la segunda lectura		
-4.9392 - 38.4415		
Potencia de la prueba		
0.8963		

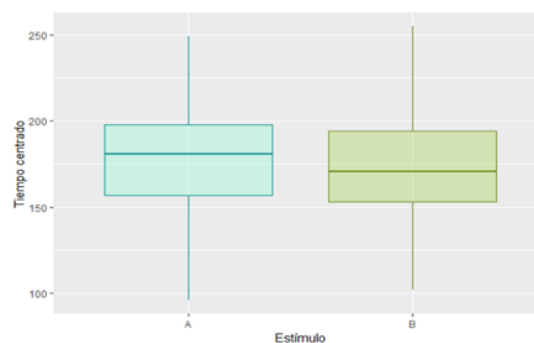
Fuente: Elaboración propia 2019.

Gráfico 3. Cuantiles teóricos de una distribución Normal contra cuantiles obtenidos con el Promedio Lectura Total



Fuente: Elaboración propia. 2019

Gráfico 4. Distribución del Tiempo Promedio de la Segunda Lectura de cada Condición, eliminando la variabilidad causada por las personas (bloques)



Fuente: Elaboración propia. 2019

Comparación de preservadores caseros para la conservación postcosecha de la rosa freedom

Josué González⁵, Jesús Reyes⁵, Sergio Varela⁵

josuegonzalezc10@gmail.com, chuz212009@gmail.com, sergiovarelasoto@gmail.com

RESUMEN

El motivo de este estudio es determinar cuál es el mejor método de conservación de rosas *Freedom*, conocida como rosa de floristería, entre aquellos que sean de fácil acceso para la población en general. Se utilizaron cuatro formas de conservación en agua: 1) agua, 2) agua con cloro, 3) agua con aspirina y 4) cambiar el agua todos los días; también se utilizó el uso o no uso del spray para el cabello como método reciente y menos conocido. De la combinación de estas dos formas de conservación se obtienen ocho tratamientos que se analizaron en busca de aquel que preservara la flor por más tiempo. Se encontró interacción entre los factores. Los resultados arrojaron que el agua con aspirina, cambiar el agua todos los días y el agua con cloro fueron significativamente mejores cuando no son mezclados con el spray que el resto de tratamientos. No se encontraron diferencias consideradas relevantes, ya que se buscaba que las rosas lograran conservarse por al menos dos días entre un método y otro para que fuese importante. El spray no alargó la vida de las rosas, pero mejoró su aspecto los primeros días. El experimento fue llevado a cabo bajo un modelo con bloques; sin embargo, estos no redujeron la variabilidad e hicieron más tediosa la fase de preparación del experimento.

Palabras clave: Rosa Freedom, análisis factorial, modelo con interacción, modelo con bloques, métodos de conservación.

ABSTRACT

The reason for this study is to determine the best method of conservation of *Freedom* roses, known as florist rose, among those that are easily accessible to the general population. Four forms of water conservation were used: 1) water, 2) water with chlorine, 3) water with aspirin and 4) change the water every day; The use or non-use of hair spray was also used as a recent and less known method. From the combination of these two forms of conservation eight treatments are obtained that were analyzed in search of the one that preserved the flower for the longest time. Interaction between the factors was found. The results showed that the water with aspirin, changing the water every day and chlorinated water were significantly better when they are not mixed with the spray than the rest of treatments. No differences considered relevant were found, since it was sought that the roses be preserved for at least two days between one method and another for it to be important. The spray did not extend the life of the roses, but improved its appearance the first few days. The experiment was carried out under a block model; however, these did not reduce the variability and made the preparation phase of the experiment more tedious.

Keywords: Freedom roses, factor analysis, model with interaction, model with blocks, conservation methods.

INTRODUCCIÓN

Los cuidados luego de adquirir flores son primordiales para su conservación. La flor más vendida, la rosa Freedom, no es la excepción. Alargar su vida y buena apariencia no es tan sencillo; en la gran mayoría de los casos, se hace con métodos caseros entre los que se destacan, el uso de hipoclorito de sodio (NaClO),

⁵ Estudiantes de Estadística de la Universidad de Costa Rica



que se obtiene comercialmente con el nombre de cloro, y el ácido acetilsalicílico (AAS), conocido como aspirina, se mezclan con el agua donde se encuentra la flor. Otro método conocido es cambiar el agua cada día desde que se obtiene la rosa. Constantemente se buscan nuevas formas para prolongar el buen aspecto de la rosa, un ejemplo de esto es el uso de spray para el cabello, por lo que este experimento tomará en cuenta este método de conservación.

Como se sabe el cloro se utiliza con mucha frecuencia comercial para la desinfección, es conocido por favorecer la eliminación de microorganismos que puedan estar en agua logra reduciendo el pH; la aspirina tiene una función similar respecto al pH del agua y sirve para proveer carbohidratos. La obtención de ambos productos es fácil, ya que cambiar el agua donde se encuentra la flor es un método que evita la formación de bacterias en el agua, puesto que esta no tiene movimiento la ausencia de oxigenación permite a las bacterias su proliferación y dañan la rosa. Si se maneja adecuadamente, la mayoría de las rosas comerciales durarán fácilmente en el florero durante 10 días (Reid, 2004).

El problema con los remedios caseros es que es difícil obtener las proporciones idóneas: si se administra demasiada lejía, se podría matar a las flores (Hockenberry, 2018). Se recomienda que tanto para el cloro como la aspirina se utilicen dosis controladas.

El spray para el cabello se utiliza como un nuevo método para conservar la apariencia de las flores por más tiempo, este protege a la flor debido a que los tallos cortados son incapaces de producir la humedad que requieren. Al cubrir las rosas Freedom con este spray, la humedad que hay en los pétalos se conserva de mejor manera, además, proporciona un retardo en la apertura del pétalo conservando la humedad y la apariencia.

Normalmente, las disoluciones conservantes contienen un agente microbicida, compuestos tampón para bajar el pH y una fuente nutritiva, que suele ser sacarosa (Serrano, 2015), por eso la escogencia de los compuestos anteriormente seleccionados. El uso de métodos caseros proporciona facilidades gracias a que la utilización de los componentes es sencilla y es de fácil obtención, pero conlleva a dificultades en cuanto a las dosis, por lo tanto, se define la cantidad que se proporciona de cada sustancia en el tratamiento respectivo. En el caso del cloro, la concentración de 3 ml de cloro por cada 250 ml agua es la recomendada, si se excede el uso del cloro podría provocar deshidratación prematura de la rosa. Para la aspirina, se disolvió de 200 mg por cada 250 ml de agua y, en el caso del spray para el cabello, se rocía ligeramente la parte del pétalo. Además, se utiliza como tratamiento el uso de agua solamente, basados en la recomendación de la experta Carolina Chacón

Esta investigación tiene se propone: 1) encontrar un efecto en el uso de spray debido a que su uso es un nuevo método y aún es poco utilizado con este fin y 2) verificar si hay una mayor preservación de las rosas en los métodos donde se utiliza componentes bactericidas, debido a que el uso de agua como único componente no tiene forma de reducir el pH.

METODOLOGÍA

El estudio experimental se llevó a cabo contemplando dos factores de diseño: 1) los medios de conservación de la rosa, (solo agua, cambiar agua todos los días, agua con aspirina y agua con cloro) y 2) el uso de spray. Además, se implementaron bloques para tratar de reducir la variabilidad del error experimental. Los bloques fueron diversas zonas, descritas al final de esta sección, donde se realizó el experimento, en las cuales factores como la temperatura, ventilación y exposición al sol fueron homogéneas.

Se utilizó, en primera instancia, un modelo saturado que contempló los factores, la interacción posible entre los factores y, también, el bloque seleccionado. El modelo utilizado fue:

$$\mu_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$$

Donde:

- μ = Media general de la cantidad de días que se conservó viva la flor.
- α_i = Efecto de los distintos tipos de agua sobre la media general.
- β_j = Efecto del spray sobre la media general.
- γ_k = Efecto de los bloques sobre la media general.
- $(\alpha\beta)_{ij}$ = Efecto de interacción entre el tipo de agua y el spray sobre la media general.

Para determinar el tamaño de muestra necesario con el que se obtendría una potencia de 0.95, la cual es la probabilidad de encontrar diferencias en el caso de que las haya, se llevó a cabo el análisis del tratamiento de agua con cloro sin spray. Se encontró que era necesario contar con un total de 5 repeticiones por tratamiento, es decir, 5 rosas para cada tratamiento, por lo que se necesitaron 40 rosas para todo el experimento para obtener resultados similares a la potencia deseada.

Se tomaron en cuenta las 8 posibles combinaciones entre los factores, denominadas tratamientos, y a estos se les asignó el mismo tamaño de muestra a cada tratamiento para que el experimento fuese balanceado. El tamaño de muestra por tratamiento fue de 5 rosas y los tratamientos fueron:

- *Agua con spray*: Se les proporcionó una cantidad de 250 ml de agua a cada una, y se le aplicó spray en la flor a cada una (extrayendo la rosa del agua para evitar contaminarla y añadir ruido) una vez cada dos días, siempre a la misma hora durante los 10 días.
- *Agua todos los días con spray*: Se les proporcionó la medida de 250 ml de agua a cada una y esta cantidad se les fue cambiando todos los días, todas a la misma hora. Además, se les aplicó spray en la flor una vez cada dos días.
- *Agua con aspirina y spray*: A la medida de 250 ml de agua se le añadió 200 mg de Aspirina, y también se roció la flor con el spray 1 vez cada 2 días durante los 10 días.
- *Agua con cloro y spray*: Se hizo la mezcla entre 250 ml de agua y 3 ml de cloro. Así mismo, la flor fue rociada con spray 1 vez cada 2 días durante los 10 días.
- *Agua sin spray*: En este tratamiento únicamente se tuvo la rosa en 250 ml de agua.
- *Agua todos los días sin spray*: Se les cambió la medida de 250 ml de agua a las flores de este tratamiento todos los días y a la misma hora durante los 10 días.
- *Agua con aspirina sin spray*: Se diluyó en los 250 ml de agua 200 mg de Aspirina.
- *Agua con cloro sin spray*: A los 250 ml de agua se le agregaron 3 ml de cloro durante los 10 días.

El diseño con bloques se realizó debido a que se quería encontrar si existían diferencias en la conservación de la rosa, dada la separación en diversas zonas, por cambios en la temperatura, ventilación, iluminación o por estar acompañadas de otras plantas. Los bloques fueron realizados en una misma casa, esto para que aspectos como la temperatura y la ventilación a la cual estaban expuestas las rosas fuera la misma y que esto no produjera confusión en las mediciones. En total fueron 5 bloques distribuidos en espacios de la casa donde estuvieran expuestos a distintos niveles de iluminación. También se trató de que algunos bloques estuvieran acompañados de otro tipo de plantas. La distinción de los bloques es la siguiente:

- *Zona 1*: En esta zona las rosas estaban expuestas a luz solar y poca sombra. No estaban acompañadas de otro tipo de plantas.
- *Zona 2*: En la zona 2 las rosas tenían poca exposición al sol y tenían sombra. Estaban acompañadas de otras plantas.

- *Zona 3:* Para esta zona se trató de mantener las rosas sin exposición al sol y con mucha sombra. No había presencia de otras plantas.
- *Zona 4:* La particularidad de esta zona fue que estaba dentro de la casa, con nula exposición a la luz solar, pero con presencia de luz artificial, y también mucha sombra. No estaban acompañadas de otras plantas.
- *Zona 5:* En la zona 5 las plantas estaban expuestas al sol y con poca sombra. Estaban acompañadas de otras plantas.

Bajo instrucción de la experta, se decidió que la rosa estaba muerta cuando su flor estaba inclinada por completo y el color de los pétalos se había tornado oscuro y con apariencia de deshidratación. Para dar la flor por muerta, Carolina indicó que debía de cumplir con las dos condiciones.

Las rosas fueron cultivadas en condiciones homogéneas y fueron cortadas en el mismo momento, esto bajo pedido, para no tener diferencias dada la manipulación antes del experimento. Se asignó aleatoriamente un bloque a cada rosa, en el cual se estableció el orden (de 1 a 40) de las rosas, y a cada uno se le asignó de forma aleatoria un número entre 1 y 5 para asignar los bloques, de la misma forma un número entre 1 y 8, para aleatorizar el tratamiento bajo el cual iban a estar siendo analizadas las rosas.

El procedimiento analítico se llevó a cabo mediante el software R versión 3.6.0 (R Core Team, 2019). En primera instancia, se verificó si se estaban cumpliendo los supuestos de normalidad y homoscedasticidad. Las pruebas utilizadas para verificar ambos supuestos fueron la prueba de normalidad de Shapiro-Wilks y la prueba de normalidad de Kolmogorov Smirnov. Para la verificación de homoscedasticidad se llevó a cabo con la prueba de Bartlett.

Luego de llevar a cabo la verificación de los supuestos, se procedió a probar la interacción. Así mismo, se procedió a analizar si el bloque provocaba algún efecto en las diferencias observadas. Para analizar el efecto del bloque, se procedió a calcular el efecto del diseño, que se calcula con la siguiente fórmula:

$$ED = \frac{CME_{SB}}{CME_{CB}} * \frac{(gl.res_{CB} + 1)(gl.res_{SB} + 3)}{(gl.res_{CB} + 3)(gl.res_{SB} + 1)}$$

Donde:

CME_{SB} = Cuadrado Medio del Error del modelo sin bloques
 CME_{CB} = Cuadrado Medio del Error del modelo con bloques
 $gl.res_{CB}$ = Grados de libertad residuales del modelo con bloques
 $gl.res_{SB}$ = Grados de libertad residuales del modelo sin bloques

Para el modelo resultante, se procedió a realizar los contrastes respecto a cada tratamiento. Primero, se realizaron las comparaciones entre los tratamientos que presentaron el mismo factor de conservación (agua sin spray contra agua con spray, agua todos los días sin spray contra agua todos los días con spray, agua con aspirina sin spray contra agua sin aspirina con spray y agua con cloro sin spray contra agua con cloro con spray) mediante sus contrastes, luego, se procedió a crear el estadístico studentizado para compararlo contra el umbral del valor p corregido por la cantidad de contrastes realizados. En los casos en los que no se encontró diferencias, se procedió a calcular los intervalos de confianza, esto para establecer el rango en el cual se podían encontrar las diferencias si es que las había, comparándolo con la medida de la diferencia importante que estableció la experta ($\delta = 2$).

Para finalizar el análisis, se procedió a calcular la potencia de la prueba, esto lo que nos indica es la probabilidad de encontrar diferencias de 2 o más en caso de que las haya.

RESULTADOS

Primeramente, se evaluó el cumplimiento de los supuestos de normalidad y homoscedasticidad, para esto se realizó un gráfico de los residuales que permitiera dar una idea visual de si se cumplía la normalidad (ver Gráfico 1 en Anexos), así como también se realizaron las pruebas estadísticas de Shapiro Wilk y Kolmogorov Smirnov. Por su parte, para el supuesto de homoscedasticidad se utilizó la prueba Bartlett, para la visualización de la variabilidad y si esta es similar en los tratamientos se hizo un gráfico de los datos centrados para todos los tratamientos (ver Gráfico 2 en Anexos). Se obtuvo que ninguno de los supuestos se rechazó (ver Tabla 4 en Anexos).

Se analizó, la existencia de interacción entre los tipos de agua y el spray de manera gráfica para visualizar si esta podía darse (ver Gráfico 3 en Anexos) y luego mediante un análisis de variancia (ANOVA), se determinó que hay suficiente evidencia estadística para rechazar la hipótesis nula de que el efecto de interacción entre los tipos de agua y el spray es igual a cero, dado que la probabilidad asociada obtenida fue de 0.049, tal y como se aprecia en la Tabla 1.

Tabla 1. Resultados del análisis de variancia para tipo de agua y spray e interacción de ambos

Prueba	Probabilidad Asociada
Agua	0.063
Spray	p<00.1
Agua y Spray	0.049

Fuente: Elaboración Propia

Al obtener una interacción significativa se procedió a realizar contrastes y así determinar cuál de los tratamientos logró una mayor preservación de la rosa. Los contrastes se realizaron comparando el promedio más alto en spray contra el promedio más bajo en spray para cada tipo de agua. Los contrastes que se crearon son los presentados en la Tabla 2:

Tabla 2. Contrastes de Spray dentro de cada Tipo de Agua.

Contraste	Probabilidad Asociada
Aspirina sin spray vs aspirina con spray	0.004
Agua sin spray vs agua con spray	0.134
Agua todos los días sin spray vs agua todos los días con spray	p<0.001
Cloro sin spray vs cloro con spray	p<0.001

Fuente: Elaboración Propia

Luego de realizado el análisis correspondiente se aplicó la corrección de Bonferroni, dando como resultado un p de 0,0125. Se rechaza todos los contrastes menos el segundo, correspondiente a Agua sin Spray contra Agua con Spray. Para aquellos contrastes que resultaron significativos se les calculó el límite inferior y así poder decidir si las diferencias ya estadísticamente significativas son además relevantes en términos del experimento y satisfacen los objetivos correspondientes.

De acuerdo a lo observado en la Tabla 3, se obtuvo que ninguno de los tratamientos supera los dos días establecidos como diferencia relevante para afirmar que un tratamiento es mejor que otro, por

consiguiente, se hicieron intervalos de dos colas y ver si es posible que en algún caso se llegue a dar esa diferencia.

Tabla 3. Intervalos de confianza para contrastes significativos.

Contraste	Límite Inferior	Límite Superior
Aspirina sin spray vs aspirina con spray	0.16	2.71
Agua sin spray vs agua con spray	1.04	2.31
Cloro sin spray vs cloro con spray	0.92	3.48

Fuente: Elaboración Propia

Por último, se quiso ver si el diseño con bloques logró hacer diferencia respecto a si el experimento hubiese sido llevado a cabo con un modelo sin bloques, en este caso se hicieron gráficos para dar una idea (ver Figura 1 en Anexos) y se usó la eficiencia relativa, que consiste en una razón, donde en el numerador se encuentra la eficiencia del modelo con bloques (MCB) y en el denominador la eficiencia del modelo sin bloques (MSB). Se obtuvo una eficiencia relativa de 1, lo que indica que los bloques no tuvieron impacto en el experimento.

Luego de realizado el experimento se obtiene que, con una significancia de 5%, hay interacción entre el tipo de agua y spray, es decir que el efecto del spray no es el mismo en cada tipo de agua, en la Tabla 1; el uso de spray en las rosas no ayudó en la preservación de estas por más días, todo lo contrario, en todos los casos donde se usó spray, a excepción de agua, la diferencia del tiempo promedio de vida de las rosas sin spray y el tiempo promedio de vida de las rosas con spray fue significativamente menor, esto con una significancia de 5%, más allá de que estas diferencias no resultaron relevantes en términos de la cantidad de días que se consideró para elegir un método sobre otro, puesto que con una confianza de 95% se puede esperar que la media de días vida de las rosas en agua con aspirina sin spray sea al menos 0.31 días mayor que para las rosas en agua con aspirina con spray, que la media de días vida de las rosas a las que se les cambió el agua todos los días sin spray sea al menos 1.11 días mayor que para las rosas en este tipo de agua sin spray y que la media de días vida de las rosas en agua con cloro sin spray sea al menos 1.08 días mayor que para las rosas en agua con cloro con spray.

Por otro lado, existe la posibilidad de que se llegue a ver una diferencia relevante entre tratamientos, ya que los 3 intervalos de confianza si incluyen al dos, lo cual implica que puede llegar a darse el caso en que un par de promedios disten en la diferencia señalada. Debido a este resultado se calculó la potencia del experimento, dando como resultado una potencia mayor a 93%, al ser tan alta indica que se contaba con una cantidad adecuada de observaciones para ver diferencias si las había.

En los casos donde no se utilizó spray, es decir, los métodos caseros con agua, que son además los más utilizados por las personas cuando de mantener flores se trata, el experimento muestra que la media de agua y aspirina es muy similar, de 10 días de vida para esas rosas aproximadamente, las rosas que estuvieron el agua con cloro lograron vivir en promedio 11 días y a las rosas que se les cambió el agua todo los días superaron, en promedio, los 11 días, alcanzando casi los 11 días y medio (ver Tabla 5 en Anexos). Si se ve desde el punto de vista del comprador o aquel que mantiene las rosas, al superar los 11 días se pueden dejar hasta el final del día y ser desechadas.

CONCLUSIONES

Basándose en la información brindada por la persona experta se tenía una idea de los resultados que podría arrojar la investigación, sin embargo, la implementación del spray podía afectar dichos resultados, ya que normalmente sólo se utiliza un método casero para la conservación de las rosas y el uso

de dos métodos químicos (aspirina con spray y cloro con spray) podía llegar a superar el que fue establecido como mejor método de conservación por la experta, que fue agua todos los días, pero no fue así.

Pese a esto la coloración, brillo, firmeza y olor de las rosas con spray fue mejor durante los primeros días, en este experimento se aplicó el spray cada 48 horas, ya que al estar expuesta a condiciones de viento y sol el efecto del spray en la rosa podía perderse y hubiese afectado uno de los fines del experimento, podría intentarse utilizar varios niveles de spray y ver si los resultados son distintos. Por último, el diseño experimental utiliza bloques, los cuales no aportaron nada e hicieron más tedioso el proceso previo al análisis, por lo que se sugiere prescindir de ellos en caso de querer llevar a cabo un experimento similar a este.

BIBLIOGRAFÍA

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

Reid, M. (2004). *Rosas, Rosa De Amor, Rosa De Amor. Recomendaciones para Mantener la Calidad Postcosecha*. Postharvest Center. Universidad de California. Disponible en: http://postharvest.ucdavis.edu/Commodity_Resources/Fact_Sheets/Datastores/Ornamentals_English/?uid=29&ds=801

Serrano, M. (2015). *Cómo prolongar la vida de la flor cortada*. (Departamento de Biología Aplicada, Escuela Politécnica Superior – Universidad Miguel Hernández. Recuperado de: <http://www.blogdelflorista.es/index.php/2015/08/16/como-prolongar-la-vida-de-la-flor-cortada>

ANEXOS

Gráfico 1. Residuales contra cuantiles teóricos de la distribución normal

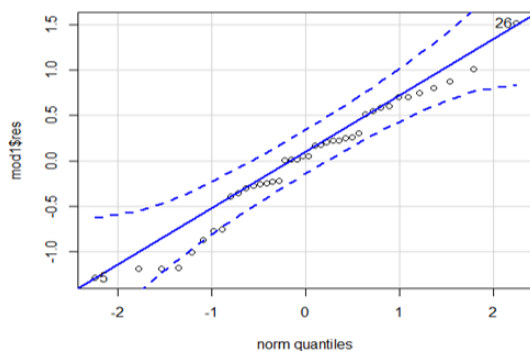


Gráfico 2. Datos centrados, para cada tratamiento

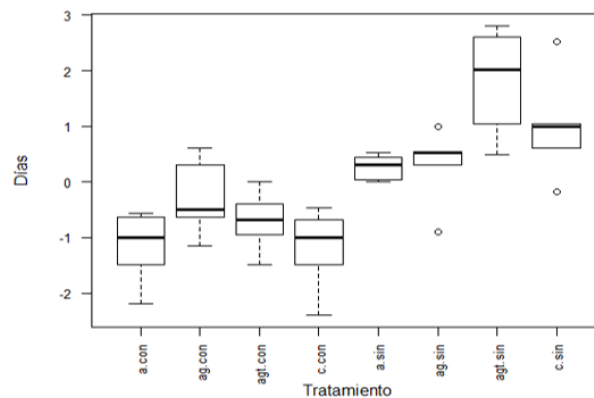


Tabla 4. Pruebas para comprobación de supuestos.

Prueba	Probabilidad Asociada
Shapiro-Wilk	0.35
Kolmogorov-Smirnov	0.76
Bartlett	0.43

Fuente: Elaboración Propia

Tabla 5. Comparación de medias estimadas para cada contraste

Agua / Spray	Con Spray	Sin Spray
Aspirina	8.42	9.85
Solamente agua	9.32	9.88
Agua todos los días	8.88	11.38
Cloro	8.38	10.58

Fuente: Elaboración Propia

Gráfico 3. Días según tipo de agua y spray.

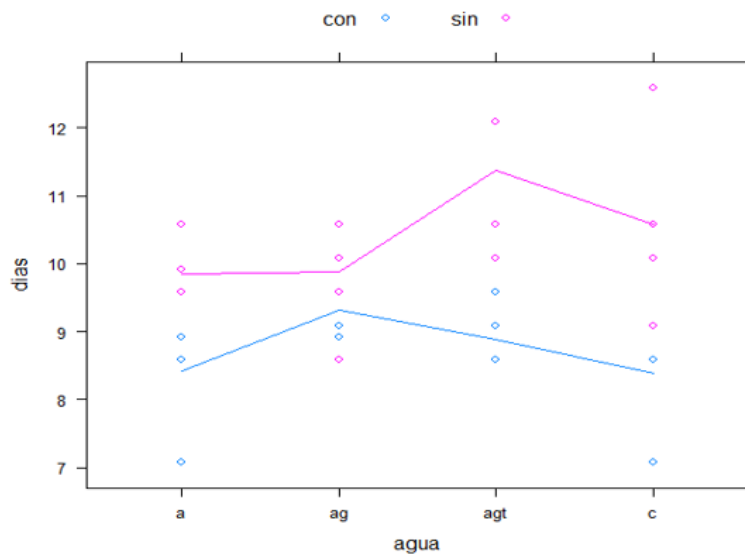
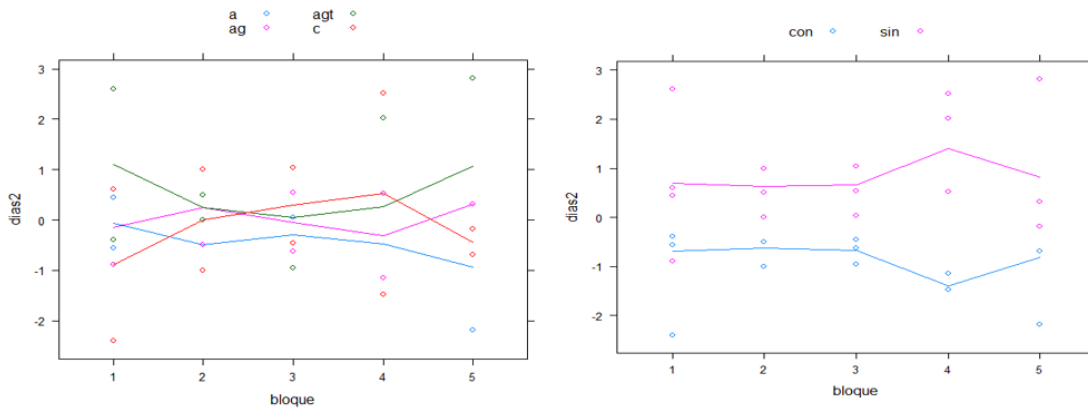


Figura 1. Datos centrados para eliminar el efecto del bloque, separados por bloque según tipo de agua y separados por bloque según spray



Análisis de la influencia a la exposición a la lengua inglesa en el mejoramiento de la pronunciación de consonantes para estudiantes del inglés como segunda lengua

Alejandro Fallas Godínez⁶, Moisés Vargas Herrera⁶

ale.fallas@gmail.com, moisesvh123456@gmail.com

RESUMEN

El aprendizaje de una segunda lengua conlleva un proceso exhaustivo en el cual influyen diversos factores. La lengua inglesa al ser una lengua fonológica conlleva a un mayor esfuerzo de parte de los aprendices para mejorar o adquirir una pronunciación más adecuada. En el caso específico de nativo hablantes del español, estos encuentran difícil la producción fonética específicamente de consonantes que no son homólogas en su lengua nativa. Por ende, el presente estudio analiza el “input” o grado de exposición al idioma como factor determinante en la mejora continua de la pronunciación de consonantes para nativo hablantes del español, así como otras variables como la motivación y el tipo de educación secundaria previa. Para lo cual, este estudio utiliza estudiantes de las carreras de inglés y enseñanza del inglés como objetos de estudio debido a su exposición a la lengua; se obtiene como resultado final que el tiempo de exposición a la lengua inglés incide en la mejora de la pronunciación de consonantes, esto se ve reflejado en el impacto de factores como el tiempo de exposición del idioma en la carrera y la educación secundaria previa.

Palabras clave: input, pronunciación, lengua inglesa, segunda lengua, consonantes.

ABSTRACT

Second language acquisition requires an exhausting process influenced by different factors. Due to being a phonological language, the English language implies to second language learners a higher effort for developing a better pronunciation. In the case of Spanish speaker, they find challenging the production of specific consonants' phonemes, which are non-homologous in their language. Subsequently, the current study analyzes the “input” or language exposure as a decisive factor in the continuous improvement of consonants' production in Spanish native speakers. As well as this, other variables such as motivation and the type of previous education were considered. For this, the study uses English and English Teaching students as study's subjects due to their language exposure. Then, it was found as a final result that the time of English language exposure boosts the improvement of consonant's production. This is reflected by the impact of factors such as the time of language exposure in the major and the previous secondary education's experience.

Keywords: input, pronunciation, English language, second language, consonants.

INTRODUCCIÓN

Las exigencias personales y profesionales que impone este mundo globalizado dan vital importancia a la formación de personas bilingües para obtener mejor capacidad de comunicación. Según España (2010) el manejo de un segundo idioma, como lo es el inglés, pasó de ser un valor agregado a una competencia lingüística que requiere desarrollar el individuo para integrarse con efectividad y eficacia a la nueva concepción del mundo globalizado (p.64). De igual manera, Gutiérrez & Landeros (2010), mencionan que el dominio de una segunda lengua implica mayores oportunidades de insertarse en el mundo global en términos laborales y culturales (p.98). Bajo las condiciones previamente mencionadas es coherente pensar

⁶ Estudiantes de Estadística de la Universidad de Costa Rica



que aprendices del Inglés como Segunda Lengua (ESL de ahora en adelante por sus siglas en inglés) deseen desarrollar o potencializar sus competencias lingüísticas en áreas tales como la pronunciación, en una Segunda Lengua (L2 de ahora en adelante).

Los procesos de aprendizaje de una segunda lengua varían según sea el caso; por consiguiente, las competencias en una segunda lengua generan resultados distintos según el método de aprendizaje en el cual se vea expuesto un aprendiz. Sin embargo, no solo los procesos de aprendizaje influyen en la adquisición de una L2, sino además factores externos como el grado de exposición a la lengua. Yule (2010) describe el término "input" como la exposición a la lengua a la que se puede enfrentar un aprendiz; para el cual un adecuado "input" generaría una mejor adquisición de la segunda lengua (p.192-193). Bajo esta misma línea, Gass y Mackey (2014) mencionan cómo el input supone en el contexto comunicativo, un elemento clave en el cual los aprendices se exponen, refiérase estos a lecturas o producciones orales (p.181), las cuales forman parte de la exposición a la lengua, por ende influyen en el contexto del aprendizaje de una L2. De este modo, resulta de suma relevancia analizar el impacto que genera el "input" en el desarrollo de habilidades lingüísticas, y en específico para este estudio, su alcance en el mejoramiento de la pronunciación para estudiantes de una segunda lengua.

El desarrollo de competencias lingüísticas se ve marcada por métodos de aprendizaje, los cuales se pueden enfocar en distintas macro áreas de la lengua (escucha, habla, escritura y lectura); cada una con la finalidad de desarrollar habilidades en una L2. Así por ejemplo, Tornel (2015) en su proyecto experimental documenta el alcance del uso de canciones como una herramienta pedagógica con gran relevancia para el mejoramiento de la pronunciación (p.69). Por otra parte, Lai, Tsai, Yu y Hsu (2017) analizan el uso de un sistema Multimedia de Aprendizaje del inglés (MEL por sus siglas en inglés) para el desarrollo de conciencia fonológica en aprendices taiwaneses del inglés como L2.

Gass y Mackey (2014) describen al input como un componente esencial para el aprendizaje de una lengua, ya que este proporciona evidencia crucial, a partir de la cual, los alumnos pueden formular hipótesis lingüísticas (p.181). Con ello se muestra la importancia que posee este componente en el aprendizaje de una L2. Este estudio se basa en la medición del mejoramiento que desarrollan en la lengua inglesa los estudiantes de la Escuela de Lenguas Modernas de la Universidad de Costa Rica, a partir del primer y hasta el cuarto año de exposición al idioma. Con la visión de cuantificar el efecto de la exposición a la lengua, este estudio pretende evidenciar un progreso relevante en la pronunciación de consonantes del inglés para nativo hablantes del español, al variar el grado de exposición a la lengua extranjera en estudiantes de las carreras Bachillerato en Inglés y Enseñanza del Inglés, pertenecientes a la Escuela de Lenguas Modernas Universidad de Costa Rica durante el I ciclo lectivo del 2019.

METODOLOGÍA

El presente estudio conformó un diseño cuasi-experimental, en el cual se analizaron los efectos que posee la exposición al inglés como L2 en el mejoramiento de la pronunciación. Se seleccionó una muestra de 48 estudiantes pertenecientes a la Escuela de Lenguas Modernas de la Universidad de Costa Rica, matriculados en las carreras de Bachillerato en Inglés y Bachillerato en Enseñanza del Inglés durante el primer ciclo lectivo del año 2019. Los estudiantes que conformaron la muestra eran estudiantes regulares de los grupos correspondientes a los siguientes cursos: LM-1001 Inglés integrado I, LM-1230 Comunicación oral I, LM-1351 Técnicas de comunicación y pronunciación I, y LM-1471 Técnicas de comunicación y pronunciación III. Los cuales son cursos compartidos en ambas carreras, y que en su mayoría se enfocan en el desarrollo de competencias orales.

La distribución de la muestra consistió en 16 elementos para dos cursos correspondientes a primer y segundo nivel, ya que por la cantidad de estudiantes matriculados, se esperaba que la variabilidad dentro

de los grupos fuera mayor. Por otro lado, debido a la menor cantidad de estudiantes y por asunción de mayor homogeneidad dentro de los grupos, en los últimos dos grupos se seleccionaron muestras de 8 estudiantes por curso. La selección dentro de los grupos se realizó aleatoriamente. La muestra se definió con base en una prueba piloto que tomó en cuenta estudiantes de todos los años, la cual medía la variabilidad del puntaje de pronunciación tomando una lista de 10 palabras. Los estudiantes leían la lista de palabras, y con base en el número de aciertos, se calculó la variabilidad en los números de aciertos de pronunciación de los estudiantes.

La recolección de los datos necesarios para el análisis se realizó bajo el uso dos instrumentos. El primero correspondió a un cuestionario (ver apéndice A), el cual recolectó información básica sobre el input y la motivación de los aprendices de ESL. El instrumento consta de cuatro preguntas cerradas relacionadas a la exposición a la L2; además de una pregunta sobre la cantidad de horas de exposición extracurricular de los estudiantes. Por último, mediante escalas de Likert (ver anexo A) con base en la tesis realizada por Eckstein (2017), se buscó medir la percepción del input recibido, así como el elemento motivacional del aprendizaje de los estudiantes. El segundo instrumento consta de una lista de 30 oraciones (ver anexo B), las cuales debían ser leídas por cada uno de los participantes, con el fin de cuantificar el número de palabras pronunciadas incorrectamente; la lectura fue grabada para su posterior revisión y verificación del conteo. Las oraciones contenían palabras cuyos sonidos son mayormente problemáticos para hispanohablantes (Amaro 2010; Helman 2004) como por ejemplo: labiodentales /f/,/v/; dentales /θ/,/ð/; post-alveolares /ʃ/,/ʒ/; alveolares en posición inicial /s/; plurales /s/,/z/,/iz/; terminaciones en /t/,/d/,/id/; velar /ŋ/; bilabiales /b/,/p/; y consonantes silenciosas (Ver anexo C).

El cuasi-experimento utilizó como unidad experimental y de observación, al estudiante matriculado en los cursos de interés, especificados en la descripción de la muestra, del cual se obtuvieron los datos mediante los dos instrumentos descritos en el párrafo anterior, y que tuvieron como objetivo medir las siguientes variables:

- *Variable dependiente:* Correspondió al conteo de palabras pronunciadas incorrectamente por cada uno de los participantes en el experimento, en el cual, el rango de palabras se encontraba entre 0 y 400 palabras aproximadamente distribuidas en 30 oraciones (ver anexo B).
- *Factores:*
 - *Año de carrera:* Para efectos de este experimento, se empleó como factor de diseño el nivel de carrera en la cual se encontraba el estudiante, ya que este representaba un “input” directo en los estudiantes. Este factor se midió en cuatro niveles (primer, segundo, tercero o cuarto año) que correspondió al grado de exposición a través de los años en la carrera de los estudiantes. La selección de los niveles (año) se asignó con base al curso en el cual el estudiante se encontraba matriculado.
 - *Experiencia laboral:* Bajo la misma línea, se analizó si el participante había desempeñado algún puesto de trabajo relacionado con la lengua inglesa por un período de tiempo de al menos un mes, ya que esta condición favorece considerablemente la exposición a la L2.
 - *Educación secundaria:* Se tomó como factor la modalidad de educación secundaria (privada, pública, semiprivada) que cursaron los estudiantes antes de ingresar a la universidad, esto debido a que en la educación costarricense la brecha entre la educación privada y pública es marcada. Así, por ejemplo, Fernández y Del Valle (2013) describen como en las pruebas PISA, la diferencia entre ambas modalidades es clara, esto “obedece sobre todo a las diferencias que existen en los factores familiares y personales que afectan al estudiante, en especial el nivel en que se encuentra en el momento de realizar la prueba” (p.1). Por ende, se consideró relevante analizar esta variable dentro del estudio.
- *Covariables*

- *Input*: Corresponde al grado de exposición que se expone un aprendiz de L2. Académicos en la teoría de interacción para la adquisición de L2, han evidenciado efectos positivos del input en el desarrollo de competencias lingüísticas para aprendices de ESL (Gass y Mickey, 2014, p.182). Covariable resultante del promedio obtenido en la escala adaptada de la tesis de Eckstein (2007, p. 100).
- *Horas extracurriculares*: Se consideró el tiempo dedicado a prácticas extracurriculares; las cuales implicaban una exposición a la L2, tales como entablar conversaciones, observar televisión o series, o como lo documenta Villalobos (2012) escuchar música en una L2. Bajo el supuesto de que estas prácticas podrían mejorar las competencias lingüísticas en pronunciación de los estudiantes.
- *Motivación*: El desarrollo de procesos de aprendizaje en una L2, se ven influenciados por un elemento motivacional. Yule (2010) explica como aquellos aprendices que experimentan una motivación al adquirir una lengua tienen una mayor posibilidad de éxito en una L2, que aquellos que poseen una motivación negativa (p.192). Debido a esto, se midió el nivel de motivación como covariable en el diseño experimental. Se utilizó un promedio obtenido de una escala de Likert con una puntuación de 1 a 5; escala adaptada de la tesis de Eckstein (2007, p. 57), que utiliza ítems para medir la motivación.

El presente estudio se basó en la siguiente hipótesis: las destrezas en pronunciación tienden a mejorar conforme aumenta el input en los estudiantes. Esta hipótesis se planteó según el análisis gráfico previo a las estimaciones de los modelos de regresión, obtenidos en la Figura C3 (Ver Anexo C), para lo cual se utilizó un modelo de regresión lineal que permitió evaluar los distintos niveles del factor de diseño, es decir, que el nivel de carrera tiene un efecto relevante con respecto al mejoramiento en la pronunciación. El modelo saturado inicial se muestra a continuación, el cual contenía todos los factores, covariables e interacciones necesarios para el estudio.

$$\mu_{ijkX_1X_2X_3} = \alpha_i + \gamma_j + \varphi_k + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + (\alpha * \gamma)_{ij} + (\gamma * \varphi)_{jt} + (\alpha * \varphi)_{it} + \alpha_i^{(1)} X_1 + \alpha_i^{(2)} X_2 + \alpha_i^{(3)} X_3 + \gamma_j^{(1)} X_1 + \gamma_j^{(2)} X_2 + \gamma_j^{(3)} X_3$$

Donde:

- α_i = Tipo de educación secundaria, con α_1 = Privada, α_2 = Pública
- γ_j = Año de carrera, con γ_1 = Primer año, γ_2 = Segundo año, γ_3 = Tercer año, γ_4 = Cuarto año
- φ_t = Labora, con φ_1 = Si, φ_2 = No
- X_1 = Puntuación en Input
- X_2 = Horas extracurriculares (semanales) de exposición
- X_3 = Puntuación en motivación
- $(\alpha * \gamma)_{ij}$ = Interacción entre Tipo de educación secundaria y Año de carrera
- $(\gamma * \varphi)_{jt}$ = Interacción entre Año de carrera y Labora
- $(\alpha * \varphi)_{it}$ = Interacción entre Tipo de educación secundaria y Labora
- $\alpha_i^{(1)}$ = Interacción entre el factor y covariable: Tipo de educación y Puntuación en Input
- $\alpha_i^{(2)}$ = Interacción entre Tipo de educación y Horas extracurriculares
- $\alpha_i^{(3)}$ = Interacción entre Tipo de educación y Puntuación en motivación
- $\gamma_j^{(1)}$ = Interacción entre Año de carrera y Puntuación en Input
- $\gamma_j^{(2)}$ = Interacción entre Año de carrera y Horas extracurriculares
- $\gamma_j^{(3)}$ = Interacción entre Año de carrera y Puntuación en motivación

Bajo el modelo anterior, se eliminaron dos valores extremos encontrados en dos estudiantes del primer nivel en las carreras de inglés y enseñanza del inglés, con previa aprobación de los expertos. Se

calcularon 12 modelos diferentes en los cuales se realizó un proceso de contraste de hipótesis para analizar si cada una de las covariables, factores y sus interacciones, estadísticamente aportaba al modelo. De este modo, se excluyeron aquellos factores, covariables e interacciones que no aportaban al modelo, y así obtener un modelo más parsimonioso.

El análisis de los datos se realizó con ayuda del paquete estadístico R (R Core Team, 2018), versión 3.6.1 y con ayuda de los paquetes Car (Fox, Weisberg y Price, 2019) versión 3.0-5, lattice (Sarkar, 2018) versión 0.20-38 y lmttest (Hothorn, Achim, Farebrother y Cummins, 2019) versión 0.9-37.

RESULTADOS

Para analizar los datos, se estableció un nivel de significancia de 0.05, con lo cual se inició con el análisis de supuesto de normalidad. La prueba de Shapiro-Wilk indicó una probabilidad de 0.1233 confirmando el supuesto de normalidad, y se comprueba el análisis exploratorio visual en la Gráfico 1 (ver Anexo). Se analizó de la misma manera el supuesto de homoscedasticidad en la Gráfico 2 (ver Anexo) que en conjunto con la prueba de Brush-Pagan, la cual dio como resultado una probabilidad asociada de 0.1706, da como resultado la corroboración del supuesto de homoscedasticidad en los datos.

Tabla 1. Contrastes de hipótesis para la significancia estadísticas de factores, covariables e interacciones para

	Nº1	Nº2	Nº3	Nº4	Nº5	Nº6	Nº7	Nº8	Nº9	Nº10	Nº11	Nº12
Variables												
α_i	**	**	*	*	**	**	**	**	**	**	***	**
γ_j	**	***	***	***	***	***	***	***	***	***	***	***
φ_t	0.65	0.64	0.62	0.63	0.62	0.62	0.61	0.62	0.61	--	--	--
Covariables												
X_1	0.50	0.49	0.47	0.47	0.46	0.46	0.45	0.46	0.46	0.38	0.38	--
X_2	0.80	0.79	0.78	0.79	0.78	0.78	0.78	0.78	0.78	0.89	--	--
X_3	0.06	0.05	*	*	*	*	*	*	*	*	*	**
Interacciones												
$(\alpha * \gamma)_{ij}$	*	**	**	**	**	**	**	**	**	**	**	**
$(\gamma * \varphi)_{jt}$	0.50	0.61	0.58	0.59	--	--	--	--	--	--	--	--
$(\alpha * \varphi)_{it}$	0.51	0.49	0.47	0.49	0.47	0.46	0.46	0.47	--	--	--	--
$\alpha_i^{(1)}$	0.21	0.19	0.17	0.18	0.17	0.16	0.16	--	--	--	--	--
$\alpha_i^{(2)}$	0.82	0.81	0.80	0.81	0.80	0.80	--	--	--	--	--	--
$\alpha_i^{(3)}$	0.54	0.52	0.80	0.51	0.50	--	--	--	--	--	--	--
$\gamma_j^{(1)}$	0.64	0.35	0.31	--	--	--	--	--	--	--	--	--
$\gamma_j^{(2)}$	0.82	0.80	--	--	--	--	--	--	--	--	--	--
$\gamma_j^{(3)}$	0.63	--	--	--	--	--	--	--	--	--	--	--

Nota: *p<0.05 **p<0.01 ***p<0.001

Con los resultados anteriores, se validó el uso de un modelo de regresión lineal para efectos del análisis en el estudio. La Tabla 1 presenta aquellos factores, covariables e interacciones que estadísticamente aportaron para el análisis después de ser sometidas a un contrastes de hipótesis.

Tras deducir aquellos factores, covariables e interacciones estadísticamente significativas, el modelo reducido obtenido para el análisis se describe a continuación:

$$\mu_{ijx_3} = \alpha_i + \gamma_j + (\alpha * \gamma)_{ij} + \beta_3 X_3$$

Donde, α_i = Tipo de educación secundaria, con α_1 = Privada, α_2 = Pública

γ_j = Año de carrera, con γ_1 = Primer año, γ_2 = Segundo año, γ_3 = Tercer año, γ_4 = Cuarto año

$(\alpha * \gamma)_{ij}$ = Interacción entre el los factores: Tipo de educación y Año de carrera

X_3 = Puntuación en motivación

Se analizó si existía efecto relevante entre los factores de diseño en la variable respuesta. Una vez observadas las diferencias, se procedió a realizar un estudio detallado sobre los efectos en los distintos niveles de los factores del modelo donde se reportaron diferencias, con el objetivo de determinar si estas diferencias eran relevantes para el presente estudio.

Una vez establecido el modelo, se analizó la presencia de diferencias significativas entre cada uno de los niveles del factor de estudio, año de carrera. El análisis se hace separado para cada tipo de educación secundaria (pública y privada), debido a que el modelo final, se encontró interacción entre el tipo de educación secundaria y el año de carrera, dando los siguientes resultados contenidos en la Tabla 2.

Tabla 2. Comparaciones entre promedios por nivel según tipo de educación secundaria

Comparación	Probabilidad asociada	*Efecto significativo
Educación Secundaria Privada		
1st-2nd	0**	SI
1st-3rd	0.0002	SI
1st-4th	0**	SI
2nd-3rd	0.3348	NO
2nd-4th	0.0527	NO
3rd-4th	0.1496	NO
Educación Secundaria Pública		
1st-2nd	0**	SI
1st-3rd	0**	SI
1st-4th	0**	SI
2nd-3rd	0.0081	SI
2nd-4th	0.0196	NO
4th-3rd	0.3771	NO

Nota: *Comparaciones por corrección de 0.05/6, **p<0.0001

Una vez encontradas aquellas comparaciones con promedios relevantes entre sí. Dada la complejidad regular de las palabras, y tal como lo propone el Mag. William Charpentier Jiménez (comunicación personal 24 de mayo de 2019), coordinador e instructor de cursos orales de inglés, Escuela de Lenguas Modernas, es relevante encontrar una disminución de al menos 25 palabras en la media de palabras pronunciadas incorrectamente, por los estudiantes avanzados en la carrera (tercer y cuarto año) en comparación con aquellos establecidos en primer año.

A su vez, el Mag. Randolph Zúñiga Coudin (comunicación personal 24 de mayo de 2019), instructor de la Escuela de Lenguas Modernas, confirma esta diferencia esperada de una reducción de 25 palabras en la media de estudiantes avanzados contra estudiantes de primer año, pero que además, se perciban diferencias menos marcadas en el conteo incorrecto de palabras al comparar las medias de estudiantes de tercer año contra aquellos en segundo año; y esta diferencia aún menor entre estudiantes avanzados (tercer y cuarto año), donde se espera encontrar una diferencia que revele el mejoramiento, pero que a su vez, no sea tan marcada como en los casos anteriores. Para lo cual, el presente estudio tomó como relevante únicamente una diferencia de al menos 25 palabras mal pronunciadas; desarrollando el análisis en aquellas comparaciones verdaderamente relevantes para efectos del estudio, con ello se obtuvieron como resultado los siguientes intervalos de confianza.

De este modo como se observa en la Tabla 3, se obtuvo finalmente que, en estudiantes con educación privada previa, sólo existe un mejoramiento relevante entre estudiantes de primer año en comparación con aquellos en cuarto nivel. Contrario al caso anterior, para estudiantes provenientes de

instituciones públicas, se encontró una mejora relevante entre los años avanzados de la carrera en comparación con los estudiantes de primer nivel.

Tabla 3. Límite inferior de Intervalos de confianza para la cantidad palabras pronunciadas incorrectamente

Comparación	Límite Inferior*
Educación Secundaria Privada	
1st-2nd	18
1st-3rd	20
1st-4th	26
Educación Secundaria Pública	
1st-2nd	38
1st-3rd	51
1st-4th	49
2nd-3rd	10

Nota: *Valores redondeados hacia arriba

Finalmente, se cuantificó la potencia de la prueba para el experimento con un tamaño de muestra mínimo de 8 unidades de estudio y para una diferencia mínima de 25 palabras entre grupos, el cual dio una potencia de la prueba de 99.65%, esto indica que, si se repitiera el estudio 100 veces, el 99.65 de las ocasiones se detectarían diferencias de al menos 25 palabras de diferencia entre los grupos si en realidad existieran con un tamaño de muestra de 8 estudiantes por grupo.

CONCLUSIONES

Se logró mostrar que existe una relación entre el grado de exposición a la lengua con la mejora de en la pronunciación, para lo cual se evidenció que conforme los estudiantes avanzan en la carrera, la pronunciación de consonantes mejoraba considerablemente. Lo cual es de esperarse, ya que los estudiantes reciben formación especializada, conforme avanzan en la carrera, obtienen un notorio mejoramiento en la producción de los fonemas.

Es importante destacar que la mejora continua, anteriormente descrita, difiere para estudiantes provenientes de una secundaria pública con aquellos que provienen de una privada. Esto se relaciona con lo mencionado por Mag. Charpentier (comunicado personal 24 de junio de 2019), el cual explica que los estudiantes que estudian en secundarias privadas tienen un aprendizaje más profundo en la lengua, lo que implica un conocimiento previo y una mejora menos marcada, debido a que en algunos fonemas ya se han adquirido previamente.

Por otro lado, se evidenció la motivación como única covariable importante en el mejoramiento de la pronunciación, lo cual es consistente con la literatura como componente externo, que influencia el desarrollo de competencias lingüísticas durante el proceso de aprendizaje. Debido a los resultados obtenidos en la mejora a través de los años, se puede concluir que el “input” o grado de exposición a una lengua, en este caso a través de los años, tiene un efecto directo en la mejora de la pronunciación en los estudiantes. Tanto los estudiantes más avanzados en la carrera, como aquellos que previamente presentaron un contacto directo con el idioma, cometieron menos errores de pronunciación. Lo cual, evidencia que a mayor exposición, aumenta positivamente los resultados en la mejora de la pronunciación.

Lo anterior sugiere que, a mayor tiempo de exposición, como los experimentan estudiantes de las carreras de Bachillerato en Inglés y Enseñanza del Inglés, mejora positivamente la pronunciación en la lengua inglesa. Pero además, la motivación personal durante el proceso de aprendizaje puede contribuir positivamente en la mejora de la pronunciación, ya que cuando los estudiantes se sienten motivados, el proceso de adquisición de una L2 se ve beneficiado.

En relación al tipo de educación secundaria, se puede mostrar que en Costa Rica la educación privada, semiprivada o bilingüe brinda a los estudiantes mayores herramientas para el desarrollo de habilidades lingüísticas en una segunda lengua. Esto indica que aun cuando los estudiantes de instituciones públicas reciben educación en la L2, inglés, los alcances de estos programas son limitados frente a los otros sistemas educativos en el país. Es por lo que aquellos estudiantes del sistema educativo público presentan una desventaja sobre sus similares que provienen de otros sistemas.

El presente estudio no analizó el impacto de prácticas extracurriculares individuales; por lo cual, se recomienda en futuros estudios considerar el alcance de estas prácticas, aunque estos datos fueron recolectados, no se incluyeron en el análisis. De igual manera, resultaría de gran apoyo el incluir más variables relacionadas con el “input”, así como elevar la dificultad del instrumento incluyendo palabras que represente un resto mayor, en cuanto a la dificultad de pronunciación para aprendices hispanohablantes.

BIBLIOGRAFÍA

- Amaro, A. B. (2010). Las consonantes plosivas en inglés: dificultades para los estudiantes españoles y posibles orientaciones pedagógicas. Recuperado de https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero_27/ANA_BUENO_1.pdf
- Eckstein, G. T. (2007). A correlation of pronunciation learning strategies with spontaneous English pronunciation of adult ESL learners.
- Fernández, A., y Del Valle, A. (2013). Desigualdad educativa en Costa Rica: La brecha entre estudiantes de colegios públicos y privados. Análisis con los resultados de la evaluación internacional PISA. *Revista cepal*.
- Fox J., Weisberg, S. y Price, B. (2019). Companion to Applied Regression. Recuperado de <https://cran.r-project.org/web/packages/car/car.pdf>
- Helman, L. A. (2004). Building on the sound system of Spanish: Insights from the alphabetic spellings of English-language learners. *The Reading Teacher*, 57(5), 452-460.
- Hothorn, T., Achim, Z., Farebrother, R. W. y Cummins C. (2019). Testing Linear Regression Models. Recuperado de <https://cran.r-project.org/web/packages/lmtest/lmtest.pdf>
- España, C. (2010). El idioma inglés en el currículo universitario: importancia, retos y alcances. *Revista electrónica educare*, (14), 63-65. Recuperado de <https://www.redalyc.org/html/1941/194115606005/>
- Gass, S., y Mickey, A. (Eds.). (2014). *Theories in second language acquisition: An introduction*. Routledge.
- Gutiérrez, M., Landeros, A. I. (2010). Importancia del lenguaje en el contexto de la aldea global. *Revista Horizontes Educativos*, (15), 98-99. Recuperado de <https://www.redalyc.org/pdf/979/97916218008.pdf>
- Lai, Y. S., Tsai, H. H., Yu, P. T., y Hsu, J. M. (2007). A multimedia learning system using hmms to improve phonemic awareness for english pronunciation. In *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, 1064-1069. IEEE.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical <https://www.r-project.org/>

Parker, M. (2000). Pronunciation and grammar: Using video and audio activities. In *English Teaching Forum* 38(1), 24-31.

Samuel, C. (2010). Pronunciation pegs. *TESL Canada Journal*, 103-103.

Saran, M., Seferoglu, G., y Cagiltay, K. (2009). Mobile assisted language learning: English pronunciation at learners' fingertips. *Eurasian Journal of Educational Research (EJER)*, (34).

Sarkar, D. (2018). Trellis Graphics for R. Recuperado de <https://cran.r-project.org/web/packages/lattice/lattice.pdf>

Villalobos, N. (2012). Using songs to improve EFL students' pronunciation.

Tornel, P. (2015). El uso de canciones en inglés como recurso didáctico en el aula de idiomas para la enseñanza de vocabulario y pronunciación a aprendices españoles adultos.

Yule, G. (2010). *The Study of Language*. Cambridge, UK: Cambridge University Press. ISBN: 9780511686320

ANEXOS

Anexo A. Cuestionario

XS-3150 Principios de Diseños Experimentales

Escuela de Estadística

Universidad de Costa Rica

ID: _____

General description: This instrument attempts to collect data about the influence of input practices for the development of English pronunciation proficiency. Data collected will be only used for research purposes, and your answers will be completely anonymous.

Part A) Instruction: Answer carefully each of the following questions.

a) What type of high school did you attend to?

1- Private 2- Public 3- Other: _____

b) Which course from the following list are you currently taking?

1- Integrated English I 3- Pronunciation & Comm. Techniques I
2- Oral Communication I 4- Pronunciation & Comm. Techniques III

c) Have you ever worked before (for at least 1 month or more) for an English field?

1- Yes 2- No

d) By filling a circle in the following scale where 1 is the lowest and 5 is the highest score, rate your personal perception about your pronunciation.

I	1	2	3	4	5
When I am trying to learn new English sounds, I use English media such as television, movies, and the radio.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am listening to someone speaking English, I identify sounds that are difficult for me to produce.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am listening to someone speaking English, I notice when they make mistakes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am listening to someone speaking English, I listen for new sounds.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am listening to someone speaking English, I repeat their words silently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

e) Besides classroom time, do you practice any activity that exposes you to the English language (Watching series, listening to music or talking with friends/relatives in English)?

1- Yes

2- No

If your answer is No, go to question h)

f) From the following list, mark the activities that you usually used to do.

Watching series in English	Listening to music in English	Talking with a friend/relative in English	Other English practice
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If your answer is other English practice, specify this activity: _____.

g) How many hours in total per week do you spend doing the previous activities?

h) By filling a circle in the following scale where 1 is the lowest and 5 is the highest score, rate the following list of statements based on your learning experiences in the English language.

M	1	2	3	4	5
When I feel stressed about learning English pronunciation, I know how to cut this stress down.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I feel stressed about my English pronunciation learning, I solve this problem immediately.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I study English pronunciation, I look for a good learning environment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When learning English pronunciation, I keep working until I reach the goals that I make for myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Number of Words

Part B) Instruction: Read carefully the following list of sentences.

- 1- Athletes have stronger muscles than those young college students.
- 2- Mrs. Smith still grieves for the absence of her husband.
- 3- Deep on the chamber, Louis could listen to a beautiful choir song.
- 4- Lana found a new scratch on her ankle after she attended to her doctor appointment on Wednesday.
- 5- The knight shield possesses a hexagon with two swords under it.
- 6- A skull was found in the catacombs of the castle.
- 7- I played for many years on the island before it was damaged by the devastating hurricane.
- 8- My new watch is so small that it couldn't use it on my wrist.
- 9- The new boy behaves in a very old chauvinistic way.
- 10- In the jungle, animals refused to refuge in the new shelters.
- 11- Mom replaced the light bulb after paying enough for the debt on the store.
- 12- I looked through the window because I heard a knock on the front door. Then, I asked "Who's there?"
- 13- #WeAreAllEqual became another sign of love against xenophobia.
- 14- Diverse obstacles challenge the hero in his quest.
- 15- Lauren succeeds as a boss for the new division opened in the south headquarters. He feels comfortable with his position.
- 16- Priest Hawthorne tells his monotheist followers about charity. Next week, they will bring clothes for homeless people.
- 17- Statisticians investigated the outcomes of the current vaccines.
- 18- Marry hangs out the phone and places her head under the table. Her financial world brings her a mixture of feelings and thoughts.
- 19- John is forced to take the knife; he aims to release the pressure of his heart. Although he is a man of good will, a catastrophe becomes inevitable.
- 20- New clients have a query. They expect us to be strict with the standard unit for pressure measurement, Pascal.
- 21- Sales experienced a significant growth since the king left his throne. Now, he checks every aspect for the commercial trades.

- 22- Scholars have inquired about a new variation for the asthma virus. For example, the new homologous chain of the virus would threaten the population's health.
- 23- The young boy was bewitched and transformed into a squirrel by the powerful warlock.
- 24- X-rays were enough to diagnose my acquaintance of her broken leg.
- 25- The average of person likes juice and shrimp.
- 26- In the garage, I found exotic vegetation, but nobody ever answered me where it came from.
- 27- Even though some methods of teaching are tough, they might be effective.
- 28- Lina rehearses a smooth character for her next debut.
- 29- Sally sells shells in the seashore.
- 30- Can you feel my love? It bombs away.

THANKS FOR YOUR PARTICIATION!!!

Gráfico 1. Gráfica sobre distribución normal de los residuos con bandas de confianza

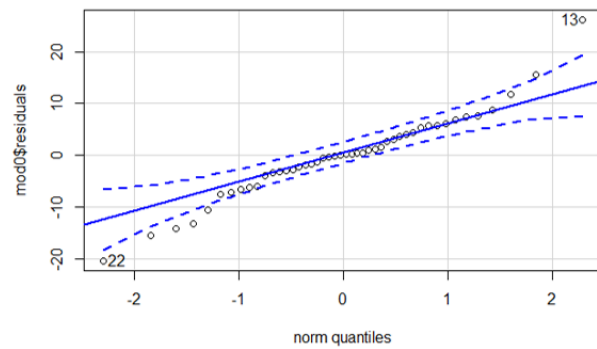


Gráfico 2. Gráfica de residuales contra predichos

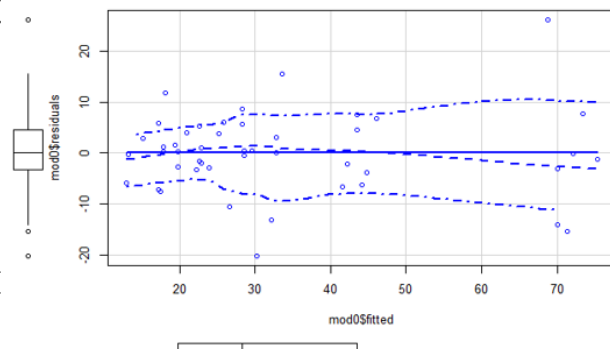
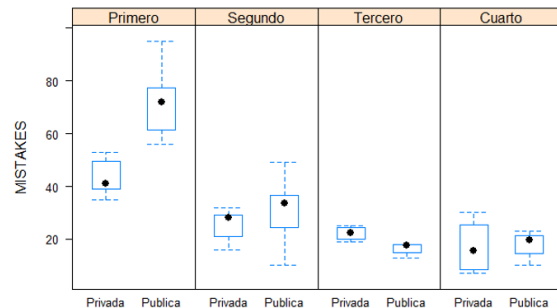


Figura 1. Distribución del conteo de errores por año según tipo de educación secundaria



Anexo B. Consonantes enfatizadas en la evaluación

Consonante	Silent	/ð/	/kw/	/s/ initial Pos.	/ks/ vrs /z/
Palabras	Knight Wednesday	Although Another	Quest Query	Student Smooth	Example Hexagon
	Sign	Then	Acquaintance	Statisticians	Mixture
	Knock	They	Bewitch	Squirrel	Xenophobia
	Debt		Headquarter	Skull	X-ray
	Island		Window	Sword	Exotic
	Castle		Warlock	Store	
	asthma			Scholars	
	Answer				
	Wrist				
	Debut				
	Bought				
Consonante	/v/	/f/	/t/	Velar n /ŋ/	/dʒ/
Palabras	Love	Shrimp	Effective	Song	Vegetation
	Grieves	Shelter	Enough	Stronger	Jungle
	Of	Financial	Tough	Young	Juice
	Variation	Pressure	Catastrophe	Hang out	Average
	Have	Seashore	Knife	Ankle	Refuge
	Vaccine	Shells	phone		
	Behave				
Consonante	/ʃ/ /k/	s/z/tz	/θ/	t/d/id	/ʒ/
Palabras	Chauvinistic	Clients	Monotheist	Damaged	Garage
	Scratch	Places	Athletes	Played	Division
	Watch	Works	Method	Looked	Measure
	Choir	Brings	Growth	Investigated	
	Challenge	Tells	Throne	Refused	
	Character	Is	Through	Replaced	
	Chain	Aims	South	Forced	
	Charity Chambe	Succeeds	Thought	Watched	
		Checks	Threat	Experienced	
		Possesses	Health		
Consonante	/h/	/b/	/p/		
Palabras	Homologous	Bombs	People		
	Husband				
	Hurricane				
	Hexagon				

Nota: Todas las palabras eran consideradas para el conteo de errores

II. SIMULACIONES

En ocasiones interesa resolver una problemática o interrogante y no es posible lograrlo con pruebas estadísticas tradicionales. En casos como estos, se puede recurrir a los procesos conocidos como simulaciones para obtener respuesta de forma empírica. Robert Shannon define una simulación como “el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias” (Shannon y Johannes, 1976)⁷.

⁷ Shannon, Robert; Johannes, James D. (1976). «Systems simulation: the art and science». IEEE Transactions on Systems, Man and Cybernetics. 6(10). pp. 723-724.



Impacto del desbalance en los tamaños de muestra por tratamiento sobre el desempeño de la prueba de comparaciones múltiples de Tukey

Dayana Gómez Quesada⁸, Joshua Salazar Obando⁸, Andrea Vargas Montero⁸
dayana.gomezquesada@gmail.com, joshua.salazar1692@gmail.com, avargas2398@gmail.com

RESUMEN

La prueba de Tukey HSD es de las pruebas de comparaciones múltiples más utilizadas; sin embargo, comúnmente se utiliza cuando en el estudio experimental se tiene una muestra balanceada. En este trabajo se evalúa el desempeño de una prueba de Tukey de forma manual y automática, bajo 6 escenarios distintos. Estos escenarios consisten en utilizar muestras balanceadas y desbalanceadas, variando el tamaño de muestra y la magnitud del desbalance de la muestra, con el objetivo de evaluar el desempeño de la prueba Tukey por las diversas condiciones planteadas, utilizando intervalos de confianza como la herramienta de evaluación. Se encuentra que utilizar una muestra desbalanceada afecta negativamente el desempeño de la prueba Tukey pero también que la magnitud de este desbalance y el tamaño de muestra tienen un papel importante en esta afectación.

Palabras clave: Prueba Tukey HSD, comparación de medias, muestra balanceada, muestra desbalanceada, TukeyHSD

ABSTRACT

The Tukey HSD test is one of the most popular post-hoc multiple comparison tests, however, it's mostly used when the experiment is balanced. This article evaluates the performance of the Tukey test made manually and automatically, under 6 different cases. The cases consist in using balanced and unbalanced sample sizes, while also changing the sample size and the magnitude of the unbalance, with the objective of evaluating how the Tukey test's performance is affected by the different conditions, using confidence intervals as the evaluation tool. It is found that using an unbalanced sample negatively affects the performance of the Tukey test, but also that the magnitude of this imbalance and the sample size have a very important role on this affectation.

Key words: Tukey HSD test, comparison of means, balanced sample, unbalanced sample, TukeyHSD

INTRODUCCIÓN

Las comparaciones múltiples conforman un grupo de análisis de suma importancia y popularidad en los análisis experimentales, ya que, en muchos casos, un estudio experimental desea contrastar los distintos tratamientos que fueron tomados en cuenta en el diseño y, en muchos casos, hacerlo de forma simultánea, es decir, se desea utilizar un método de análisis que logre realizar estas distintas comparaciones al mismo tiempo, controlando que el error tipo 1 no incremente.

Dentro del marco de este tipo de análisis se encuentra la prueba o método de Tukey, creada por John W. Tukey. La cual consiste en la comparación de los promedios para todos los pares de tratamientos posibles. Por ejemplo, si se tienen 3 tratamientos, se hacen 3 comparaciones en total, si se tienen 6 tratamientos 15 comparaciones, y así sucesivamente. Esta es la prueba más aplicada y preferida por los estadísticos, pues controla de mejor manera los dos errores ampliamente conocidos en la estadística (α y β) (Montgomery 2003, citado por Wong, 2010).

⁸ Estudiantes de Estadística de la Universidad de Costa Rica



En la mayoría de casos, esta prueba se trabaja con muestras balanceadas, esto quiere decir que todos los tratamientos poseen la misma cantidad de repeticiones u observaciones. Lo anterior sucede así porque dentro de los cálculos realizados en esta prueba se asume esta condición, por lo que algunos autores consideran que la prueba de Tukey solo se utiliza bajo un panorama de diferencias grandes en los tamaños de muestra de los tratamientos (Fallas, J. 2012). Sin embargo, en la práctica, con frecuencia sucede que no se puede tener el mismo número de repeticiones para cada tratamiento, por lo cual, se debe optar por una muestra desbalanceada. En ciertas ocasiones estas muestras desbalanceadas no son planeadas en el diseño, sino que por distintos motivos como fallas en la implementación de tratamientos, o incluso causas desconocidas que se acontecen durante el proceso de aplicación se pierden observaciones, con lo cual se obtienen distinto número de datos para cada tratamiento y por consiguiente complica el análisis (Reyes, 2014). Esto no significa que no se pueda utilizar la prueba mencionada, existen diversas formas de trabajar un panorama como el descrito.

Según Glen (2016) la prueba de Tukey de honesta diferencia significativa (Tukey HSD por sus siglas en inglés) tiene supuestos que se deben seguir para su utilización, los cuales son:

- Las observaciones son independientes dentro y entre los grupos.
- Los grupos para cada media en la prueba provienen de una distribución normal.
- Igualdad de variancia dentro de los grupos asociados a cada media en la prueba (homoscedasticidad).

Ahora bien, a partir de lo mostrado se propone los objetivos de la investigación:

- Ilustrar el desempeño que presenta la prueba Tukey en dos formatos distintos como la prueba manual y la prueba automática HSD, esto al aplicarlo a diferentes muestras desbalanceadas con distintas magnitudes de desbalance.
- Ilustrar el cambio en el desempeño de cada formato de aplicación de la prueba de Tukey al utilizar distintos escenarios como muestras balanceadas o desbalanceadas en dos niveles (10% y 30%).

METODOLOGÍA

El modelo utilizado se conforma de un solo factor, con 3 niveles, es decir, se trabajan 3 tratamientos. Por lo que el modelo se puede escribir de la siguiente forma:

$$\mu_{ij} = \mu + \alpha_i$$

Donde α_i se refiere al efecto del tratamiento i sobre la media de la variable respuesta y μ es la media general. Se utiliza el modelo de tratamiento de referencia, por lo que $\alpha_1 = 0$.

En esta prueba el primer paso es encontrar las diferencias entre los distintos tratamientos que se estén trabajando, donde se restan los promedios observados de cada par de tratamientos. Comúnmente se hace esta resta de tal forma que sea el promedio mayor menos el menor, con el fin de que todos los cálculos posteriores sean mayores a cero. Seguidamente, se debe calcular el Cuadrado Medio Residual del modelo, el cual tiene la forma:

$$CME = \frac{\sum(n_j - 1)s_j^2}{n - k} = \frac{SCE}{n - k}$$

Donde n corresponde al tamaño de muestra total, mientras que n_j corresponde al tamaño de muestra del j -ésimo, también conocido como r de "repeticiones" por tratamiento. Además, s_j^2 se refiere a

la varianza del j -ésimo tratamiento y la suma de estas por su respectivo $(n_j - 1)$ corresponde a la *SCE* o *SCRes*, es decir, a la Suma de Cuadrados Residual. El Cuadrado Medio Residual o del Error, se obtiene dividiendo la *SCRes* entre los grados de libertad residuales, es decir $n - k$, donde k corresponde al número de tratamientos. Una vez obtenido el *CME* se procede a calcular el error estándar de estimación de la diferencia, esto se obtiene con la siguiente fórmula:

$$ee = \sqrt{\frac{2CME}{r}} \text{ o } \sqrt{CME \left(\frac{1}{r_j} + \frac{1}{r_j} \right)}$$

Donde el primer caso es con muestras balanceadas y el segundo con muestras desbalanceadas.

Seguidamente, se encuentran los cuantiles de una distribución de Tukey correspondientes

$$a: q = \frac{d}{ee}$$

Donde d corresponde a cada diferencia entre promedios. Y con estos cuantiles se calcula el valor p para cada diferencia, esto con el fin de determinar si una diferencia se considera importante. Para encontrar estos valores p se utiliza la distribución de Tukey. Igual que con cualquier otro contraste de hipótesis, se define un nivel de significancia para rechazar o no la hipótesis de igualdad de medias. En el caso de que se rechace, se procede a obtener el intervalo de confianza para la diferencia. Para esto se obtiene un cuantil de la distribución t , pero al estar haciendo comparaciones múltiples es necesario hacer un ajuste, en lugar de calcular estos cuantiles con probabilidad $1 - \alpha$, donde α es el nivel de significancia, se calculan con $1 - \frac{\alpha}{b}$, donde b corresponde a la cantidad de intervalos que se están calculando.

Con estos cuantiles se obtienen los intervalos de confianza:

$$\pm t * ee$$

Y con estos se decide si la diferencia que resulta relevante en términos del problema se encuentra en intervalo o no.

El análisis se llevó a cabo mediante el software de RStudio, dentro del paquete *stats* se encuentra la función *TukeyHSD*, la cual realiza el procedimiento anterior automáticamente. Solo es necesario indicar el modelo que se está utilizando y la función devuelve la estimación puntual de la diferencia, el límite inferior, el límite superior y el valor p ajustado. Es importante mencionar que el modelo creado debe ser utilizando la función *aov* para que el procedimiento funcione. Todo el procedimiento se realizó utilizando el lenguaje *R* (*R Core Team, 2018*), con la versión 3.6.0 el programa de integración del mismo *RStudio* (1.2.1335).

Con la finalidad de responder a los objetivos de la investigación, primeramente, se plantean dos escenarios generales, el primero, donde se realiza la prueba Tukey manualmente y el segundo donde se utiliza la función *TukeyHSD* para realizar la prueba. En ambos casos se realizan las comparaciones de medias necesarias, las cuales se establecieron en $\mu_1 = 10$, $\mu_2 = 12.5$ y $\mu_3 = 15$, para fines didácticos, donde particularmente lo que se pretende ilustrar es que partir de escenarios que se comentarán más adelante se evidencie diferencia entre las medias. Las comparaciones que se realizan son tres casos, en el cual se

compara cada par de medias de la siguiente manera: $\mu_2 - \mu_1$ en el caso 1, para el caso 2 es $\mu_3 - \mu_1$ y por último $\mu_3 - \mu_2$, como caso 3. Ambos procedimientos (manual y automático) se realizan creando una función.

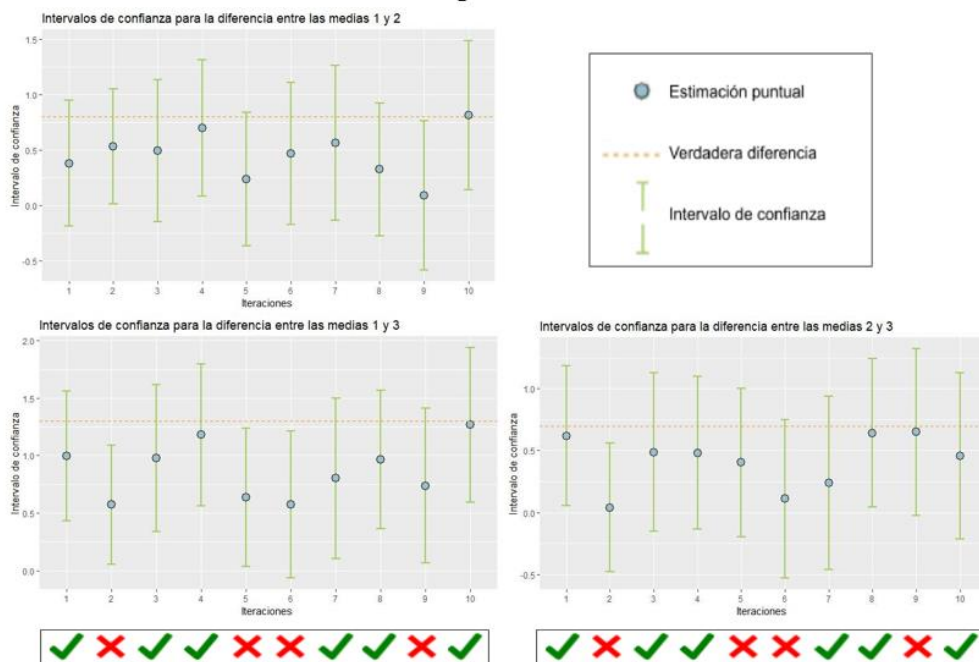
Sobre la simulación en sí, antes de entrar de lleno en la función inicialmente se debe asignar la prueba Tukey correspondiente, así como el tipo de muestra, la cual depende del balance o no y del tamaño. Ambas selecciones como se observa en el anexo esta parte del código son fabricados a partir de condicionales o estructuras de control cuyo propósito es facilitar el acceso al tipo de escenario que se pretende evaluar.

Posteriormente dentro de la función para ambas pruebas de Tukey se generan datos aleatorios condicionadas por lo que el usuario asigna, las repeticiones en cada tratamiento (r), la lista de medias teóricas(μ) y la variancia, que por motivos prácticos se recomienda utilizarla con el valor de 1 hasta comprender realmente el procedimiento. Asimismo cabe recalcar que las observaciones son independientes dentro y entre los grupos.

Estos datos aleatorios son provenientes de una distribución normal, cuyo propósito es hacer referencia a medias condicionales de la variable respuesta dado los tres tratamientos provenientes de la misma distribución. Con la generación de datos respectiva se asegura el cumplimiento de los tres supuestos para la creación de un modelo lineal (normalidad, homocedasticidad e independencia). Seguidamente se crea un modelo utilizando esos datos y un factor "x" que también es generado aleatoriamente. En el caso de la prueba automática solo se aplica la función *TukeyHSD* al modelo generado.

Para la prueba manual se deben seguir el procedimiento explicado en la metodología (ver función en anexo). Se realiza el procedimiento 1000 veces (esto puede ser modificado fácilmente si se desea) y para evaluar el desempeño de la prueba en cada caso se revisa en cuántas iteraciones se logra encontrar la diferencia real entre las medias respectivas. Se considera acierto o éxito cuando, para una iteración, los tres intervalos contienen a su respectivo parámetro, en caso de que al menos uno de los intervalos no lo incluya se considera un fallo. Para ejemplificar el proceso de categorización "fallo o éxito".

Figura 1



Fuente: Elaboración propia, 2019

La figura 1 muestra 3 gráficos, correspondientes a las 3 comparaciones posibles entre medias: 1 contra 2, 1 contra 3 y 2 contra 3, respectivamente. Debajo de los 3 gráficos se muestra una simbología que indica si en esa iteración se obtuvo un éxito o un fallo, como previamente fue mencionado para que una iteración se considere un éxito, los 3 intervalos deben contener a la diferencia verdadera que se pretende estimar. Si se evalúan estos intervalos se obtienen 6 éxitos de 10 iteraciones, es decir se obtendría un 60 % de aciertos.

Este procedimiento se realiza para cada uno de los 6 escenarios específicos que interesa analizar:

1. Muestra pequeña balanceada ($r=10$).
2. Muestra grande balanceada ($r=30$).
3. Muestra pequeña con desbalance de 10% ($r_i = 9-10-11$).
4. Muestra pequeña con desbalance de 30% ($r_i = 7-10-13$).
5. Muestra grande con desbalance de 10% ($r_i = 27-30-33$).
6. Muestra grande con desbalance de 30% ($r_i = 21-30-39$).

Y se aplica cada uno de estos escenarios específicos a los dos escenarios generales mencionados anteriormente (prueba manual y prueba automática), por lo que se tienen al final 12 casos.

RESULTADOS

Luego de realizar el procedimiento previamente mencionado se obtuvieron los distintos porcentajes de éxito, de las 1000 iteraciones realizadas, cuál fue el porcentaje de estas que, para cada caso, obtuvo dentro de los 3 intervalos el parámetro respectivo. Los resultados obtenidos se resumen en la Tabla 2.

Tabla 2. Porcentajes de éxito por 1000 iteraciones realizadas en diferentes escenarios

Escenarios	Tukey Manual	Tukey HSD
1. Muestra pequeña balanceada ($r=10$)	91.9%	94.9%
2. Muestra grande balanceada ($r=30$)	92.4%	94.9%
3. Muestra pequeña con desbalance de 10% ($r_i=9-10-11$)	92.1%	94.5%
4. Muestra pequeña con desbalance de 30% ($r_i=7-10-13$)	90.2%	93.1%
5. Muestra grande con desbalance de 10% ($r_i=27-30-33$)	91.8%	95.1%
6. Muestra grande con desbalance de 30% ($r_i=21-30-39$)	91.6%	94.7%

Fuente: Elaboración propia, 2019

La Tabla 2 permite realizar comparaciones entre los porcentajes de éxito que se extraen de las pruebas aplicadas y además entre los diferentes escenarios para cada una de las pruebas. Como se puede observar, en general la prueba automática de *TukeyHSD* tiene porcentajes de éxitos más altos que los resultados que muestra la prueba manual. Las pruebas donde se cuenta con tamaño de muestra balanceada tienen mayor facilidad para su realización, ya que ciertos cálculos pueden ser simplificados; en cuanto al porcentaje de éxitos obtenidos no se ve mucha variación en aumentar estos tamaños, aun así, hay una serie de detalles que se deben analizar para la comprensión de la temática antepuesta. Para ello se pretende describir lo que sucede en cada escenario de acuerdo con su prueba respectiva, así como una breve comparación entre ellas. El análisis no va dirigido a ratificar cuál prueba es mejor que la otra, sino precisamente el enfoque primordial será su comparación dado los escenarios.

La postura inicial, como bien se comenta al inicio del escrito, el método de Tukey usualmente se asume una muestra con la misma cantidad de repeticiones en cada tratamiento y por tal motivo computacionalmente resulta muy común utilizar esta condición para sus procedimientos. Si bien es cierto

en los escenarios 1 y 2 mostrados en la Tabla 2 no tienen ningún inconveniente, los datos respaldan a que la prueba *TukeyHSD* presenta un mayor porcentaje de éxitos.

La prueba Tukey manual puede verse perjudicada por una cierta particularidad y es que la conformación de los intervalos de confianza es dada por un proceso en el cual se hace la revisión de cuáles diferencias de medias son estadísticamente significativas y de acuerdo con eso se procede a la cuantificación de la diferencia por medio de los intervalos de confianza. Debido a esto se recurre a una corrección, que es parte de los intervalos de confianza necesarios para dicha prueba, los cuales son los intervalos de Bonferroni. Dicha corrección se conforma dependiendo de cantidad de diferencias significativas, dado un tamaño de contraste fijado. Una vez conocido esta cantidad se divide el nivel de significancia entre esta cantidad. Se puede mencionar que esta prueba es muy conservadora, por lo que es posible no contemplar diferencias en ciertos casos. (Dagninos, J. 2014).

Dentro del proceso de simulación esta corrección permanece constante, es decir, dado que se pretende visualizar las diferencias entre las medias, se toma como si todas las diferencias en su momento fuesen significativas, para luego posteriormente ver la posibilidad si con las iteraciones correspondientes puede modificarse esta situación. Indirectamente el proceso que se lleva a cabo se limita al enfoque de estudio y de alguna manera debido a que esta corrección es bastante conservadora, puede estar incidiendo en que las estimaciones por intervalo correspondientes no logren captar correctamente las diferencias como debería. En el caso de la prueba *TukeyHSD* si está encontrando mayores diferencias. No obstante, la prueba *TukeyHSD* presenta cierta flexibilidad, como bien se comentaba en la descripción de la prueba, lo cual esta poca rigurosidad está haciendo que las estimaciones tengan mayor acceso a encontrar diferencias, que si bien es cierto en general si se compara con los escenarios de la prueba manual obtienen porcentajes más altos.

Por otra parte, se debe hacer una aclaración con respecto a los escenarios con una muestra desbalanceada y es que según la descripción de esta prueba automática que brinda el software de R, solo se puede utilizar cuando existe una muestra balanceada, o en casos muy específicos donde una muestra la diferencia entre las repeticiones de los grupos o tratamientos no llegue a desestabilizar la prueba, prácticamente a niveles de desbalance sumamente pequeños. Esto deja mucho en qué pensar puesto que a nivel práctico es muy posible encontrar casos donde la muestra tenga desbalances más pronunciados y efectuar una prueba como esta deja mucha incertidumbre. Aun así, se procede a verificar estos resultados para observar qué tan lejanos pudiesen estar en comparación con la prueba manual y con lo mostrado en la figura 2 no hay diferencias llamativas, pero sí hay detalles que se pueden comentar.

En las situaciones donde se cuenta con muestras pequeñas desbalanceadas (escenarios 3 - 4), los porcentajes están siendo más afectados por el cambio en 30%, es decir, a mayor porcentaje de desbalance en la muestra las pruebas arrojan un porcentaje menor de éxitos en cuantos a los intervalos, por tanto indica que menor cantidad de estos incluyen a sus parámetros respectivos.

Lo anterior tiene sentido ya que, como previamente fue mencionado, la prueba Tukey es comúnmente utilizada para muestras balanceadas y, al presentarle un escenario donde el desbalance es considerablemente grande, la prueba muestra problemas en ambos, no solo a partir de *TukeyHSD*. Además una vez realizado el proceso de simulación se nota como en este caso se incrementa la cantidad de intervalos que contienen al cero, dichos intervalos no están encontrando diferencias entre las medias asignadas y por tal motivo no se pueden tomar en cuenta.

De todos los escenarios posibles las muestras pequeñas con desbalances son las que obtuvieron la mayor cantidad de estos intervalos, haciendo pensar que posiblemente aunque se tenga una diferencia de medias considerable como parámetros (diferencia de 2.5, 5 y 2.5 respectivamente) y una varianza ajustada,

la cual no llega a incidir tanta en la precisión de los otros escenarios, no hay duda que generan sospechas de que a menor muestra y a mayor desbalance se obtenga el peor de los escenarios. Aunque no sea tan alarmante en la Tabla 2 se aprecia como el escenario 4 para ambas pruebas son los de menor porcentaje de éxitos.

Ahora bien, esa situación también se analiza si se incrementa el tamaño de muestra; sin embargo, resulta importante analizar que tanto la prueba manual como la prueba automática utilizando la función *TukeyHSD*, el cambio en el porcentaje de éxito no es tan superior en comparación con los escenarios anteriores.

En los escenarios 5 - 6 las diferencias entre los porcentajes de éxito en los intervalos entre los distintos desbalances son pequeñas, con lo cual se muestra que para este caso el impacto del nivel de desbalance se ve minimizado para tamaños de muestra relativamente grandes para cada tratamiento que se aplique.

Como dato a destacar se aprecia como en el escenario 5 la prueba *TukeyHSD* obtiene el mayor porcentaje de éxitos, el tamaño de muestra con unas repeticiones considerablemente grandes y el poco desbalance indudablemente está beneficiando a la prueba, pero la limitación de dicha prueba con una muestra no balanceada no permite aclarar si este resultado pudiese ser el indicado para esta situación, ya que el caso donde debería estar sin ningún problema es el escenario 1, donde la muestra es balanceada, cumple con todas las condiciones requeridas y tener un resultado mejor que ese, pero que no cumple en su totalidad con las condiciones no resulta tan confiable.

CONCLUSIONES

La prueba Tukey tanto manual como automática presenta una baja en su desempeño cuando se presenta un desbalance grande en la muestra. Esto expone por qué muchos prefieren utilizar la prueba bajo un panorama de balance; sin embargo, esto no significa que la prueba es completamente imposible de utilizar bajo un panorama de muestra desbalanceada, solo se debe tener en cuenta al utilizarla en estas condiciones, que si existe una afectación en el desempeño de la prueba, pero igual puede producir resultados relevantes para la investigación en cierta manera, esto mediante desbalances pequeños o desbalances en que no se refleje una afectación importante en la prueba. La creación de escenarios con diferentes porcentajes de desbalances, así como la utilización de una prueba manual y otra automática puede resultar efectivo para la medición que se quiere implementar

Si se utiliza la prueba de Tukey con una muestra desbalanceada, se debe considerar la magnitud del desbalance que se está utilizando, ya que un desbalance mayor tendrá un desempeño considerablemente menor al de la misma prueba con un tamaño de muestra similar, pero con un desbalance menor.

La forma en que el desbalance afecta la prueba de Tukey también se diferencia con respecto a los tamaños de muestra. Un panorama donde se realiza la prueba con una muestra pequeña y sumado a esto, un desbalance considerable entre las muestras de cada tratamiento, obtendrá resultados en donde no se podrá encontrar con más claridad las diferencias entre medias a como si se realizara con una muestra balanceada o una muestra grande y desbalanceada. El desbalance debe evaluarse con mayor profundidad, aunque en los casos que se planteó se refleja como un tamaño de muestra grande opaca que la distribución de la muestra en los diferentes tratamientos pudiese ser una dificultad en la estimación por intervalo.

Los resultados mostrados son bajo una situación particular, es sobre un modelo con un solo factor, una varianza que permanece constante en todos los casos, una diferencia de medias ajustada para captar en cierta medida esas diferencias en los intervalos de confianza. Es un caso muy específico y por tanto no

debe ser una generalización, no debe ser un enfoque que se permita contextualizar para diferentes modelos o estudios más complejos que requieran del método de Tukey. Si bien es cierto los porcentajes de éxito en todos fueron altos, esto es precisamente porque se quería ver que sucedía en situaciones con condiciones muy parecidas cómo afectaba o no el desbalance de muestra en ambas pruebas. No obstante, el estudio presenta aportes que pueden ser tomados en cuenta, debido a que se hace un análisis del método de Tukey no desde un punto de vista de probabilidades sino se profundiza más allá de una decisión, intentando por medio de los intervalos de confianza contemplar una cuantificación de la diferencia, que bajo estas condiciones es la búsqueda de este parámetro de interés dentro ellos.

BIBLIOGRAFÍA

- Dagninos, J. (2014). Comparaciones múltiples. Recuperado de <https://revistachilenadeanestesia.cl/comparaciones-multiples/>
- Fallas, J. (2012). Análisis de Varianza: comparando tres o más medias [archivo PDF]. Recuperado de http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/analisis_de_varianza_2012.pdf
- Glen, S. (2016). What is the Tukey Test / Honest Significant Difference? Recuperado de <https://www.statisticshowto.datasciencecentral.com/tukey-test-honest-significant-difference/>
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Reyes, L. (2014). PRUEBA DE TUKEY PARA EXPERIMENTOS DESBALANCEADOS. Recuperado de <http://reyesestadistica.blogspot.com/2014/05/prueba-de-tukey-para-experimentos.html>
- RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL: <http://www.rstudio.com/>.
- Wong, E. (2010). ¿Después de un análisis de varianza...qué? Ejemplos en Ciencia de Alimentos [archivo PDF]. Recuperado de http://www.mag.go.cr/rev_meso/v21n02_349.pdf

Afectación de la cobertura de los intervalos de confianza para diferencias de medias al incumplir el supuesto de homoscedasticidad, un enfoque de tres alternativas para analizar datos heteroscedásticos

Jose Pablo Aguilar Umaña⁹, Fernando Alvarado Prado⁹, Daniely Hernández Orama⁹
jose.aguilarumana@ucr.ac.cr, fernando.alvaradoprado@ucr.ac.cr, daniely.hernandez@ucr.ac.cr

RESUMEN

Los modelos estadísticos utilizados para el análisis de varianza dentro del diseño experimental para determinar la diferencia entre dos o más promedios, pueden o no estar basados en diferentes supuestos. Si se decide modelar bajo supuestos es necesario verificar si estos se cumplen o no, para no caer en un análisis erróneo. La presente investigación se centra en el incumplimiento del supuesto de homoscedasticidad, el cual se estudia en cuatro niveles con magnitud distinta y para tamaños de muestra diferentes. Se propusieron tres métodos para analizar datos que presentan heteroscedasticidad: uno en donde se toma en cuenta el incumplimiento del supuesto de homoscedasticidad mediante el método de mínimos cuadrados ponderados (MCP), otro en donde simplemente se ignora el incumplimiento del supuesto ya mencionado, y por último, uno donde se toma en cuenta la heteroscedasticidad al sumar las varianzas de las medias de cada tratamiento. La comparación del funcionamiento de los métodos se da al comparar un resumen de la proporción de veces que se logró captar la diferencia de los promedios para cada modelo en cada nivel de incumplimiento del supuesto, a esta medida resumen se le denomina cobertura. La cobertura es medida para los intervalos de confianza (IC) bilaterales para la diferencia de medias con una confianza del 95%. Para esto se obtuvieron datos simulados provenientes de distribuciones normales con diferentes promedios y varianzas, para luego utilizarlos como base para plantear modelos utilizando los tres métodos mencionados y proceder a calcular los IC al 95% para corroborar si se obtuvo o no la confianza esperada en cada caso. Se encontró que al tomar en cuenta la heteroscedasticidad mediante cualquiera de los dos métodos el nivel de cobertura de los IC es generalmente mayor que si simplemente se ignora la heteroscedasticidad, en especial cuando se tiene un tamaño de muestra considerable. Lo anterior se cumple para los cuatro niveles de incumplimiento planteados. De igual forma la cobertura de los métodos que toman en cuenta la heteroscedasticidad se comporta de manera igual en todos los escenarios.

Palabras clave: homoscedasticidad, heteroscedasticidad, Mínimos Cuadrados Ponderados, sumatoria de varianzas, intervalo de confianza, cobertura.

ABSTRACT

Statistical models used for analysis of variance within the realm of experimental design, which have the objective of finding differences between two or more means, may or may not satisfy its different assumptions. If one decides to model using said assumptions, it's necessary to verify if these are met or not in order to not carry out any erroneous analysis. The current study is focused on the breach of homoscedasticity, which is studied in four different levels and for different sample sizes. Three methods were proposed to face this heteroscedasticity: one where the heteroscedasticity is taken into account using Weighted Least Squares (WLS), another where the heteroscedasticity is simply ignored and another where the heteroscedasticity is taken into account by adding the variances of the mean of the response variable for each treatment. This comparison is done by observing how many times the theoretical difference of means is correctly estimated for each model and for each level of heteroscedasticity, which we call

⁹ Estudiantes de Estadística de la Universidad de Costa Rica



coverage. We define coverage as what proportion of bilateral confidence intervals for the difference of means contains the real difference, considering a confidence level of 95%. To that end, we created data through a simulation that came from a normal distribution with different means and variances. This was done in order to use them to model the three options mentioned and to calculate the 95% confidence intervals for the purpose of seeing if the expected confidence level was met for each method or not. We found that when heteroscedasticity was considered in the model, be it through WLS or through the sum of variances, the coverage was generally bigger than if heteroscedasticity was ignored, especially when there is a big enough sample size. This phenomenon happens for all four levels of the breach of the assumption. Also, the coverage for the methods which take into account the heteroscedasticity has the same behavior for all of the proposed scenarios.

Keywords: homoscedasticity, heteroscedasticity, Weighted Least Squares, sum of variances, confidence interval, coverage.

INTRODUCCIÓN

Al utilizar modelos para describir la relación del promedio de una variable respuesta o dependiente con ciertos factores independientes, muchas veces se debe trabajar bajo diferentes supuestos teóricos. Cochran (1974) menciona que, si se cumplen con los supuestos requeridos para un modelo de análisis de varianzas, se pueden obtener del modelo medidas de precisión como errores estándar estimados o intervalos de confianza, razonablemente, libres de sesgo. De lo mencionado por el autor se puede inferir que si se incumple algún supuesto la calidad de las medidas de precisión se puede ver afectada.

Algunos modelos de regresión deben cumplir supuestos, entre ellos el de normalidad, linealidad, no multicolinealidad y homoscedasticidad, esto para que su construcción sea posible y logren generar estimaciones aceptables, es decir, con una alta precisión. El incumplir supuestos teóricos que los modelos deben satisfacer y trabajar como si se estuvieran cumpliendo, puede acarrear errores o un aumento en la imprecisión en las estimaciones obtenidas del modelo, esto resulta peligroso ya que decisiones tomadas de acuerdo a resultados obtenidos pueden resultar erróneas.

Justamente señalan Barker y Shaw (2015) que una de las consecuencias de incumplir con los supuestos ante los modelos de análisis de varianza es que los intervalos de confianza (IC) estimados no contengan los valores reales en una frecuencia diferente de lo esperado. En otras palabras, la cobertura empírica de los modelos resultará deficiente pues no coincidirá con lo que se espera a nivel teórico. Lógicamente, esto significa también que las probabilidades de error tipo I y de error tipo II a nivel empírico no son igual a las esperadas.

En este trabajo el enfoque estuvo en el supuesto de homoscedasticidad (igualdad de varianzas) que debe existir entre cada distribución que se tiene para la variable respuesta dado cada tratamiento y cómo afecta el incumplir este supuesto a los IC bilaterales construidos para una confianza de 95%. Dentro de la simulación se incumplió solamente este supuesto con el fin de aplicar tres métodos diferentes para lidiar con la falta de esta premisa. Además, interesó ver el desempeño de cada método bajo diferentes casos determinados por la cantidad de réplicas por tratamiento y la magnitud del incumplimiento de la homoscedasticidad, por lo que se incluyeron diferentes tamaños de muestra y diferentes niveles de incumplimiento de la homoscedasticidad en la simulación.

La meta de este trabajo es comparar la cobertura de los intervalos de confianza para la diferencia de medias en dos niveles de un factor, al enfrentarse al incumplimiento del supuesto de homoscedasticidad (en diferente intensidad y para cantidades diferentes de réplicas) con tres métodos diferentes:

1. No tomar en cuenta la heteroscedasticidad

2. Mínimos Cuadrados Ponderados
3. Tomar en cuenta la heteroscedasticidad mediante la suma de las varianzas

METODOLOGÍA

Se definió un factor de diseño, el cual tiene tres niveles que se llamaron tratamientos ya que solo se cuenta con este factor. Se propuso el siguiente modelo de suma nula:

$$\mu_i = \mu + \alpha_i$$

Donde:

μ_i : Media del tratamiento i

μ : Media general

α_i : Efecto del tratamiento i -ésimo

y se tiene como restricción:

$$\sum_{i=1}^n \alpha_i = 0 \Rightarrow \alpha_1 = - \sum_{i=2}^n \alpha_i$$

Específicamente con 3 tratamientos el modelo en forma no resumida es:

$$\mu_{trat} = \mu + \alpha_1 X_1 + \alpha_2 X_2$$

Si interesa el promedio del tratamiento 1 se puede extraer del modelo si X_i toma los siguientes valores: $X_1 = 1$ y $X_2 = 0$, de igual manera para la media del tratamiento 2: $X_1 = 0$ y $X_2 = 1$ y para tratamiento 3: $X_1 = -1$ y $X_2 = -1$.

Se utilizaron tres formas de estimar los parámetros en el modelo.

- *Método A:*

Se ignoró el incumplimiento de la homoscedasticidad y se construyen los intervalos de confianza como si se estuviera modelando bajo un modelo que satisface dicho supuesto.

- *Método B:*

En este método se aplicó el método de Mínimos Cuadrados Ponderados (MCP) para tomar en cuenta la heteroscedasticidad.

Se definió MCP como la ponderación de cada observación por el inverso de la variabilidad dado el tratamiento, con lo cual se tomó en cuenta la heteroscedasticidad en el modelo utilizado.

Según Alvarado (2018), el procedimiento para realizar este método es:

1. Ajustar el modelo de regresión sin ponderar y analizar los residuales.
2. Estimar la función de varianza o desviación estándar.
3. Usar los valores ajustados para obtener las ponderaciones:

$$w_i = \frac{1}{(\hat{s}_i)^2} \quad W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_n \end{bmatrix}$$

donde \hat{s}_i^2 es la varianza de cada tratamiento.

4. Estimar los coeficientes de regresión usando estos pesos:

$$\hat{\beta}_w = (X'WX)^{-1} X'WY$$

Después de realizar este procedimiento se dice que el modelo está tomando en cuenta la heteroscedasticidad y se pudo seguir realizando los cálculos con ese mismo modelo sin temor de alterar características de un modelo que sí cumple los supuestos.

- *Método C:*

Para el último método se trabajó calculando las varianzas de las diferencias entre medias tomando en consideración la heteroscedasticidad mediante la suma de las varianzas de las medias de cada tratamiento, sabiendo que las medias entre los tratamientos son independientes.

Nótese como la parte aleatoria y la sistemática resultan ser la misma para los tres métodos al escribir la ecuación respectiva, sin embargo, la diferencia entre A, B y C está dada por el método aplicado a cada uno para enfrentar la heteroscedasticidad. El modelo general era saturado.

Se utilizó el proceso de Simulación de Monte Carlo para generar tres poblaciones de variables aleatorias con distribución Normal y con varianzas heterogéneas, una de ellas con media 10, otra con media 20 y la última con media 15. De estas poblaciones se obtuvieron 12 diferentes conjuntos de datos, que se derivaron de la combinación de nivel de gravedad con el que se incumplió el supuesto de homoscedasticidad y de diferentes cantidades de réplica por tratamiento. Se tuvieron cuatro niveles incumplimiento en donde la diferencia de las varianzas era, en el nivel “moderado”, de 1.25 veces de una varianza con respecto a la otra, en el nivel “considerable” de 1.50 veces, en el nivel “fuerte”, de 2 veces y por último en el nivel “alarmante” de 4 veces. Las cantidades de las réplicas (r) por tratamiento se categorizaron en “pocas” si el número de observaciones por tratamiento era menor a 10, “intermedias” si las observaciones por tratamiento estaban entre 10 y 30 y por último si la cantidad de réplicas por tratamiento era mayor a 30 se dijo que la cantidad de réplicas era “grande”.

En la Tabla 1 se presentan los conjuntos de datos generados, denominados de ahora en adelante set de datos, según el nivel de incumplimiento y el número de réplicas. Cada set de datos cuenta con 10.000 datos.

Tabla 1. Creación de los casos para la simulación.

Categoría según cantidad réplicas por tratamiento	Nivel de incumplimiento de la homoscedasticidad			
	Moderado (25% mayor)	Considerable (50% mayor)	Fuerte (100% mayor)	Alarmante (300% mayor)
Pocas (r<10)	Set 1.1	Set 1.2	Set 1.3	Set 1.4
Intermedias (10≤r≤30)	Set 2.1	Set 2.2	Set 2.3	Set 2.4
Muchas (r>30)	Set 3.1	Set 3.2	Set 3.3	Set 3.4

Dado que la cantidad de réplicas en cada set de datos era el mismo para cada nivel del factor X, se dijo que se trabaja con un modelo balanceado en cada ocasión. Note que esto permitió que si se desea conocer el tamaño de muestra total (n), se puede multiplicar el número de réplicas en cada caso por tres (que corresponde al número de tratamientos).

Cada método propuesto para afrontar la heteroscedasticidad (A, B y C) fue utilizado 12 veces, una por cada set de datos presentes en Tabla 1, con el fin de ver su cobertura ante datos provenientes de diferentes escenarios en cuanto al nivel de incumplimiento de la homoscedasticidad y el número de réplicas. Dentro de cada uno de estos 12 escenarios se calcularon los intervalos de confianza bilaterales para la diferencia de medias y se observó la proporción de cuántos de los IC calculados lograron contener la diferencia de las medias poblacionales. Estos intervalos de confianza se realizaron con una distribución t con n-k grados de libertad, donde k es el número de tratamientos. Buscando que los intervalos sean al 95% de confianza. Este proceso fue el mismo para los tres métodos utilizados. El resto del intervalo de confianza se calculó de manera convencional:

$$IC = (\bar{x} - \bar{y}) \pm t_{1-\alpha/2} \cdot ee$$

En cada método se calculó el error estándar de una manera diferente.

Para el método A el cálculo del error estándar estuvo dado por:

$$ee = \sqrt{2 \cdot \frac{CMRes}{r}}$$

dónde:
$$\text{CMRes} = \frac{\text{SCRes}}{(n - k)} \quad \text{y} \quad \text{SCRes} = \sum_{j=1}^k \sum_{i=1}^{r_j} (y_{ij} - \bar{y}_j)^2$$

Para calcular el error estándar en el método B se debió usar el método de matriz de covarianzas del modelo, para poder tomar en cuenta los pesos que provienen del MCP. El procedimiento fue:

1. Se calculó el Cuadrado Medio Residual del modelo ponderado:

$$\text{CMRes}^{(w)} = \frac{\sum w_i (y_i - \hat{y}_i)^2}{n - p}$$

2. Se calculó la matriz de covarianzas y se le sacó la raíz para sacar el error estándar:

$$\hat{V} \left(\hat{\beta}^{(w)} \right) = \text{CMRes}^{(w)} (X^T W X)^{-1}$$

por lo tanto:

$$ee = \sqrt{c^T \cdot \hat{V} \left(\hat{\beta}^{(w)} \right) \cdot c}$$

Donde c es el vector de contrastes, dado por la diferencia de los contrastes respectivos para la media del tratamiento 1 (0,1,0) y la media del tratamiento 2 (0,01), o sea, $c = (0,1,-1)$. No se consideró la media del tercer tratamiento dado que no es diferente por lo que no interesa.

Por último, en el método C el error estándar se calculó como una ponderación de varianzas, de la siguiente manera:

$$ee = \sqrt{\frac{S_1^2 + S_2^2}{r}}$$

Este proceso se pudo realizar porque el modelo es balanceado.

Luego se graficaron las coberturas para cada intervalo de confianza realizado y de esta manera se compararon el desempeño de A, B y C al romper con la homoscedasticidad en diferentes niveles (Moderado, Considerable, Fuerte y Alarmante) y para cada categoría del tamaño de la cantidad de réplicas por tratamiento.

Para fines de este estudio, se creó una función utilizando el software R (2018) mediante RStudio (2016) en donde se simuló la creación de un modelo ANOVA sin pesos y otro con los pesos obtenidos del método Mínimo Cuadrado Ponderado. Esto se realizó para poder calcular los errores estándar de la diferencia de media entre un tratamiento y otro, en donde su diferencia es igual a un número definido anteriormente que denominamos delta, el cual para nuestro estudio fue de 10 unidades.

El cálculo de los errores estándar para los tres diferentes métodos definidos sirvió para crear los intervalos de confianza de la diferencia de medias. Una vez creados estos, se verificó si el valor delta se encontraba dentro del intervalo.

Para lograr todo lo anterior, se simularon medias condicionales provenientes de una distribución normal y con varianzas desiguales dependiendo de los tres niveles del factor diseño. Estas medias se utilizaron para poder crear los modelos ANOVA. Este proceso se realizó con 10.000 repeticiones para poder calcular la cobertura empírica con suficiente precisión para poder llegar a conclusiones fiables. Además, se simuló para los cuatro niveles de incumplimiento de la homoscedasticidad y para los tamaños de réplicas por tratamiento desde 3 hasta 50 para poder obtener la tendencia de la cobertura dependiendo de diferentes tamaños de muestra.

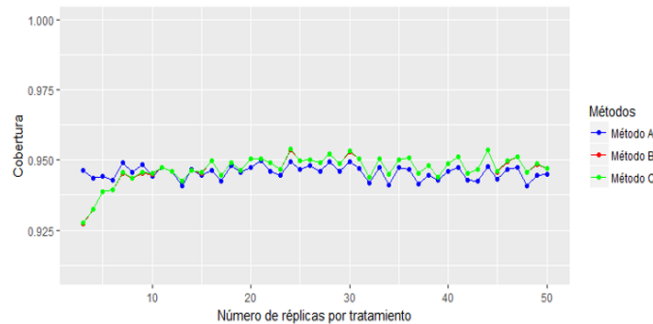
Para visualizar los resultados de manera gráfica, se utilizó el paquete ggplot2 (Wickham, 2009). El código se puede ver en el Anexo 1.

RESULTADOS

A continuación, se presentan los resultados de las simulaciones: Observe que entre más cerca de 1 estén los valores del eje y, mejor es la cobertura de los intervalos de confianza. Se utiliza un nivel de confianza de 95% para simular los datos, este es la cobertura ideal que se busca cumplir.

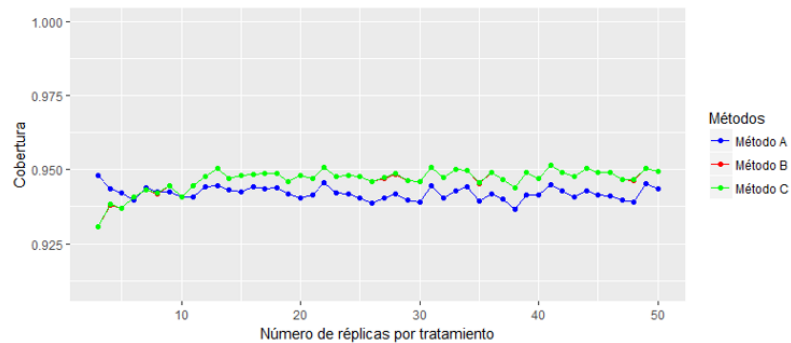
El Gráfico 1 muestra que al estar ante un incumplimiento moderado y en la categoría de cantidad de réplicas “Intermedias”, específicamente a partir de $r \geq 15$, los métodos B y C tuvieron el mismo desempeño, estando ambos cerca del 95% de confianza teórico. Mientras que el desempeño del método A fue menor que el de los otros dos métodos, excepto al encontrarse en la categoría de “Pocas” réplicas.

Gráfico 1. Cobertura de los IC bilaterales calculados utilizando los métodos A, B y C según la cantidad de réplicas por tratamiento, cuando el incumplimiento del supuesto de homoscedasticidad es moderada



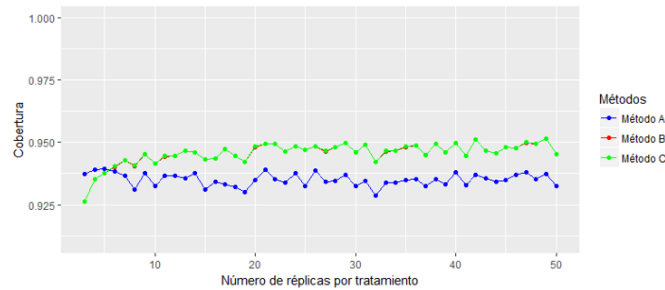
El gráfico 2 también muestra que los métodos B y C tuvieron el mismo desempeño, estando ambos cerca del 95% de confianza teórico al entrar en la categoría de cantidad de réplicas Intermedia, específicamente a partir de $r \geq 15$. Mientras que el desempeño del método A fue menor que el de los otros dos métodos, excepto al encontrarse en la categoría de “Pocas réplicas”, de hecho, al encontrarse en esta categoría y con un incumplimiento considerable hay ocasiones en que el desempeño de B y C, en cuanto a cobertura, fue menor que el de A.

Gráfico 2. Cobertura de los IC bilaterales calculados utilizando los métodos A, B y C según la cantidad de réplicas por tratamiento, cuando el incumplimiento del supuesto de homoscedasticidad es considerable



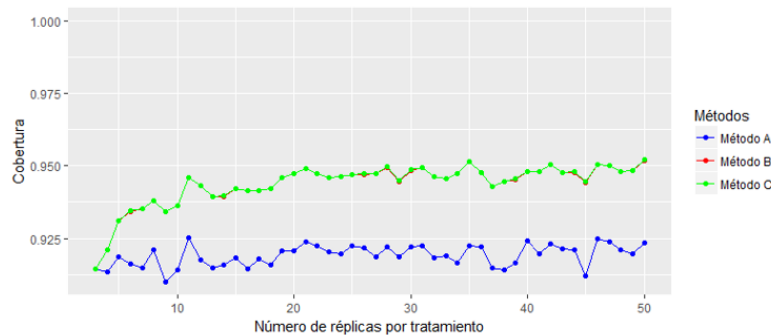
En el gráfico 3 se vuelve a tener que el método B y el método C tuvieron el mismo desempeño a pesar de estar en un nivel de incumplimiento mayor que en los otros dos gráficos. En el nivel de incumplimiento Fuerte el método A siguió presentando un mucho peor desempeño que los otros métodos, pero de nuevo se tiene que en ocasiones con pocas réplicas A tuvo un mejor desempeño que B y C.

Gráfico 3. Cobertura de los IC bilaterales calculados utilizando los métodos A, B y C según la cantidad de réplicas por tratamiento, cuando el incumplimiento del supuesto de homoscedasticidad es fuerte



Por último, al analizar el gráfico 4, se evidencia que si el incumplimiento es “Alarmante” el método A siempre tuvo un desempeño peor que los métodos B y C, quienes nuevamente tuvieron el mismo desempeño. En este nivel de incumplimiento la diferencia en desempeño entre A y los otros métodos llega a ser de 4% después de 20 réplicas por tratamiento.

Gráfico 4. Cobertura de los IC bilaterales calculados utilizando los métodos A, B y C según la cantidad de réplicas por tratamiento, cuando el incumplimiento del supuesto de homoscedasticidad es alarmante



CONCLUSIONES

Para el método A, el cual consiste en estimar los parámetros del modelo utilizando máxima verosimilitud e ignorando el incumplimiento de la homoscedasticidad, para posteriormente construir los intervalos de confianza como si se estuviera modelando bajo dicho supuesto, el desempeño se encontró por debajo de la cobertura teórica del 95% prometida. Al cumplir con el supuesto de homoscedasticidad y también por debajo del desempeño de los métodos B y C prácticamente en todas las categorías del número de réplicas y en todos los niveles del incumplimiento del supuesto, esto resulta de interés debido a que ambos métodos toman en cuenta el incumplimiento del supuesto de homoscedasticidad pero de formas diferentes. El método B consiste en aplicar el método MCP para tomar en cuenta la heteroscedasticidad, mientras que el método C consiste en calcular las varianzas de las diferencias de medias para tomar en consideración la heteroscedasticidad pero mediante la suma de las varianzas de las medias de cada tratamiento, esto sabiendo que las medias entre los tratamientos son independientes. Este menor desempeño del método A con respecto al desempeño del método B y C evidencia de manera empírica la pérdida de precisión al ignorar el incumplimiento del supuesto de homoscedasticidad.

Se obtuvo que el método A tuvo un menor desempeño que los otros dos métodos prácticamente en todas las categorías para número de réplicas, debido a que con cantidades de réplicas que se encuentran en las categorías “Intermedias” y “Muchas” lo anterior se cumple, pero, en ocasiones el método A tuvo un mejor desempeño que B y C al estar en la categoría de “Pocas” réplicas. Por lo tanto, las categorías de

Intermedias y muchas réplicas no tuvieron un efecto considerable sobre la tendencia que sigue la cobertura de los IC de cada modelo para ninguno de los niveles de incumplimiento del supuesto, pero la categoría de Pocas réplicas si tuvo efecto.

Sobre los diferentes niveles del incumplimiento de homoscedasticidad, se tuvo que el desempeño del método A, en cuanto a la cobertura se refiere, difirió del desempeño de B y C. Cuando el incumplimiento fue moderado el desempeño fue parecido para todos los métodos. Mientras que al aumentar la intensidad del incumplimiento el método A se desempeñó peor que los otros dos métodos. Qué tanto peor se desempeña el método A fue proporcional al nivel de incumplimiento de homoscedasticidad. Esto indicó que es importante tomar en consideración en qué nivel se está incumpliendo la homoscedasticidad.

Por otro lado, tanto el método de MCP como el de tomar en cuenta la heteroscedasticidad mediante la suma de las varianzas cumplieron de manera aceptable con el fin de mantener la precisión prometida en los IC bilaterales al 95% de confianza, en todos los niveles de incumplimiento de la homoscedasticidad y para las categorías de la cantidad de réplicas Intermedias y Muchas. De hecho, ambos métodos tuvieron el mismo desempeño.

Como recomendación para futuras investigaciones, puede resultar interesante llevar a cabo un análisis similar al realizado en este estudio, pero enfocado en observar el comportamiento de la cobertura utilizando modelos que no son ni saturados ni balanceados, dado que podrían generar diferentes resultados para los métodos aplicados.

Por otro lado, vale la pena estudiar más a profundidad el por qué el método B y del método C dieron resultados idénticos y si existen condiciones donde estos dos no son semejantes. Sería provechoso probar diferentes condiciones tales como la inclusión de otros factores, de interacciones, de bloques y de covariables para ver si los dos métodos siguen teniendo los mismos resultados.

Como último punto, sería interesante aplicar la misma simulación, pero para diferentes valores de varianza para poder observar si los incumplimientos porcentuales de la homoscedasticidad tienen algún efecto diferente, dependiendo de la magnitud en valores absolutos de las varianzas. Por ejemplo, en lugar de empezar con una varianza igual a 4 y aumentar en 25%, empezar con un valor igual a 100 e ir aumentando de igual manera, de esta forma a pesar de que el aumento porcentual es el mismo, la diferencia en los valores absolutos puede afectar de manera diferente lo métodos planteados.

BIBLIOGRAFÍA

- Alvarado, R. (2018) Modelos de Regresión Aplicados: IV. Construcción del modelo. Universidad de Costa Rica. Presentaciones del curso.
- Barker, L. & Shaw, K. (2015). Best (but of t-forgotten) practices: checking assumptions concerning regression residuals. *The American Journal of Clinical Nutrition*. 102(3), 533–539. doi:10.3945/ajcn.115.113498.
- Cochran, W. (1974) Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, vol. 3, no. 1, p. 22-25.
- H. Wickham (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

III. TÉCNICAS DE AGRUPAMIENTO

Las técnicas de agrupamiento o clustering son una colección de métodos estadísticos que buscan grupos de instancias con características similares mediante el análisis de los atributos que interesan estudiar. Son muy utilizadas en la minería de datos como una técnica de aprendizaje no supervisado, pero se utilizan en muchas áreas como la psicología, robótica y biología. Este tipo de técnicas fueron originadas en la antropología por Driver y Kroeber en 1932 (Driver y Kroeber, 1932)¹⁰.

¹⁰ Driver and Kroeber (1932). "Quantitative Expression of Cultural Relationships". University of California Publications in American Archaeology and Ethnology. Quantitative Expression of Cultural Relationships: 211–256 – via <http://dpg.lib.berkeley.edu>.



Relación entre actitudes hacia el amor y signo zodiacal

Daniela Chavarría Guevara¹¹, Noelia Jiménez Mena, Gabriel Zarate Artavia
danichava15@gmail.com , noeliaj11@gmail.com , gabozarate1998@gmail.com

RESUMEN

Dado el interés de las personas en los estudios e investigaciones de las estrellas y los astros que se han generado a través de los años para predecir el futuro o caracterizar a la persona, se decide analizar si existe una relación entre el signo zodiacal y las actitudes hacia el amor de una persona. Se aplica un análisis de agrupamiento, con el método de K-medias para generar cuatro conglomerados en el que se consideran 6 tipos de amor, los cuales corresponden a: Eros, Ludus, Storge, Manía, Pragma y Ágape. Con base en los resultados, se obtiene que los individuos quedan distribuidos en los diferentes grupos independientemente de su signo zodiacal y del elemento al que pertenece su signo, por lo que se concluye que no hay relación entre los signos zodiacales y los puntajes en los diferentes tipos de amor.

Palabras clave: Amor, Astrología, Conglomerados, Horóscopo, Zodíaco, Relaciones sentimentales.

ABSTRACT

Given the timeless interest in studying the stars to predict the future or try to characterize people attitudes is decided to analyze if it exists a relationship between the zodiac sign and one's love attitudes. A K-means cluster analysis is performed to separate the sample in four groups, considering 6 kinds of love: Eros, Ludus, Storge, Manía, Pragma y Ágape, using the Love Attitude Scale developed by Hendrick & Hendrick (1986). Based on the results, it is shown that people is distributed in the groups regardless of the zodiac sign or its element; therefore it is concluded that in the sample there is no relationship between the zodiac sign and the score obtained in the different kinds of love.

Keywords: Love, Astrology, Clusters, Horoscope, Zodiac, Love relationships

INTRODUCCIÓN

A través del tiempo, los seres humanos se han interesado en el estudio de las estrellas y los astros, esto ha generado distintas corrientes de investigación; por ejemplo, la astronomía busca conocer las leyes que rigen a los astros, mientras que la astrología es el “estudio de la posición y del movimiento de los astros como medio para predecir hechos futuros y conocer el carácter de las personas” (RAE, 2019).

La astrología trata de explicar cómo diferentes aspectos en el momento de nacimiento de una persona influyen en su personalidad a lo largo de su vida. El signo zodiacal solar es uno de esos aspectos, el cual es de los más conocidos y sencillos de entender, generalmente se ve en los horóscopos que se publican a diario en periódicos, revistas o en línea. Son doce signos, que se establecen dependiendo de la fecha de nacimiento. Según la astróloga Linda Goodman, el sol es el más poderoso de todos los cuerpos estelares y es por mucho, la más importante entre las consideraciones aisladas (Goodman, SF).

Por otro lado, es conocido que los signos se agrupan según cuatro elementos, de acuerdo la página oficial del gobierno de México, la clasificación es la siguiente:

-Fuego: Conformado por Aries, Leo y Sagitario. Son perspicaces y poseen agudeza de ingenio. Son capaces de percibir de manera inmediata la realidad.

¹¹ Estudiantes de Estadística de la Universidad de Costa Rica



-Tierra: Conformado por Tauro, Virgo y Capricornio. Se caracterizan por ser sensatos, organizados, perseverantes y realistas.

-Aire: Conformado por Géminis, Libra y Acuario. Los signos de este elemento saben moverse, viajar e interactuar.

-Agua: Conformado por Cáncer, Escorpio y Piscis. Los signos de agua se adaptan con facilidad, se amoldan.

Los astrólogos afirman que, dependiendo del día, hora, incluso la latitud y longitud en que nacen las personas, se pueden deducir características de su personalidad, comportamiento y preferencias en torno a distintos temas. Entre esos temas, uno de los que genera más interés es las preferencias en cuanto a las relaciones sentimentales, ya que, como afirman Lascurain, Lavandera y Manzanares (2017) citando a Graham y Christiansen (2009) el ser humano siempre se ha interesado por este tema, y todo lo relacionado a la búsqueda de una pareja y, en especial, el amor, que es considerado como un aspecto importante de la experiencia humana. Generalmente, al ser un vínculo tan estrecho que se da entre individuos, se ha tratado de establecer algún tipo de criterio para saber en qué condiciones alguien resulta ser “compatible” con otra persona, y para muchos, el zodiaco es la guía utilizada para ese propósito.

Aunque la astrología es una disciplina muy antigua y tiene bastantes seguidores, lo cierto es que también genera bastante escepticismo puesto que verificar las teorías que se generan a partir de ella es muy difícil, aunque algunos astrólogos dicen usar modelos matemáticos para sus predicciones, en realidad no se ha logrado dar un respaldo científico a las predicciones que establecen. Por lo anterior, esta investigación busca explorar si existe una relación entre el signo zodiacal y la actitud hacia el amor que presentan las personas.

METODOLOGÍA

Para conocer la actitud amorosa, se han desarrollado escalas psicométricas, que se vienen usando desde hace décadas. Se miden seis tipos de amor propuestos por Lee en 1973, según García et.al (2012) este autor diferenció tres estilos amorosos básicos, basado en la teoría platónica en la obra “El Banquete”:

- Eros: amor pasional o romántico, intenso, atracción física y sexual.
- Ludus: amor lúdico, poca implicación emocional, sin expectativas futuras.
- Storge: amor amistoso, basado en la intimidad, la amistad, el cariño y el compromiso.

García et.al (2012) mencionan que esto dio lugar para que Lee propusiera tres estilos secundarios:

- Manía: amor obsesivo formado por Eros y Ludus, determinado por los celos intensos, la intimidad, la desconfianza y la posesividad.
- Pragma: amor pragmático, establecido por Ludus y Storge, búsqueda de la pareja compatible ideal.
- Ágape: amor altruista constituido por Eros y Storge, entrega total y desinteresada por la pareja.

Lascurain, Lavandera y Manzanares (2017) explican que, Hendrick y Hendrick (1986) continuaron con la teoría de Lee, mediante la confección de un instrumento de medición de los estilos de amor denominado Escala de Actitudes sobre el Amor, conformado por 42 ítems. Desde entonces se han hecho adaptaciones de esta escala.

Para efectos de la presente investigación, se utiliza el cuestionario validado por Rodríguez, Fernández, Carrera y Vallejo en el 2013, en su investigación *Validación de la Escala de Actitudes hacia el Amor en una muestra de adolescentes*, este instrumento consta de 18 ítems (tres por cada tipo de amor), el cual es una escala de Likert; posee un formato de respuesta que va desde 1 hasta 5 (que corresponde desde totalmente en desacuerdo a totalmente de acuerdo). Esta escala permite diferenciar los seis tipos de amor antes descritos: amor Eros (Apasionado), amor Ludus (Lúdico), amor Storge (Amistoso), amor

Pragma (Lógico), amor Mania (Posesivo) y amor Agape (Altruista). La escala consiste en 18 preguntas, en la que por cada tipo de amor se tienen 3 preguntas; por lo que, para obtener el puntaje del individuo en cada tipo, se calcula el promedio de respuesta en las preguntas asociadas al mismo.

El cuestionario incluye también preguntas generales sobre la persona, como características sociodemográficas, opiniones respecto a diversos temas y otras características de la persona (orientación sexual, religión, cantidad de parejas que ha tenido), para observar si existe una asociación entre estas y las preferencias amorosas antes descritas. La unidad de estudio corresponde a una persona entre los 18 y los 30 años, que se encuentra en una relación sentimental; entendiéndose como relación sentimental si entre las partes existe un vínculo, algún tipo de atracción, y un acuerdo mutuo sobre los términos de la relación.

A cada unidad de estudio se le aplica un cuestionario, realizado en Formularios de Google, con preguntas relacionadas a relaciones sentimentales. Cabe mencionar que la recolección de los datos no es aleatoria; ya que, el formulario se aplica a personas con disponibilidad de contestar. Se obtienen 151 respuestas, pero de estas solo 133 personas indican estar en una relación sentimental (acorde a la definición antes mencionada).

Posteriormente, al obtener las respuestas de los entrevistados se aplica un análisis de agrupamiento o clusters para realizar una partición de los sujetos en grupos basados en la similitud de los puntajes obtenidos en los tipos de amor de los mismos, para de esta manera desarrollar una clasificación objetiva de las personas según las actitudes hacia el amor.

En el análisis, inicialmente se determina cuántos grupos son los indicados, para esto, se calcula la suma de cuadrados dentro de grupos con el método de K-medias; la cual es una medida de la variabilidad dentro de los grupos que se van a formar, por lo que se busca que este valor sea lo más pequeño posible, en la **Figura 1 (ver anexos)** se puede apreciar que conforme aumenta la cantidad de grupos, la suma de cuadrados disminuye considerablemente; sin embargo, cerca de los cuatro grupos tiende a estabilizarse. Cabe destacar que para este análisis se toma en cuenta la distancia euclídea, que es la menor distancia existente entre dos puntos.

Adicionalmente, se hace un análisis a partir de un dendograma, que se muestra en la **Figura 2 (ver anexos)**, para realizar el agrupamiento de los sujetos, en el que se utiliza el método de Ward (criterio aplicado en el análisis de cluster jerárquico que realiza una serie de particiones a partir del tamaño de la muestra donde cada uno de ellos contiene un solo individuo). Se decide realizar 4 clusters en el análisis, pues dan un ajuste adecuado.

Una vez asignados los individuos a alguno de los cuatro grupos se procede a observar por medio de gráficos de cajas, un panorama de las respuestas promedio por tipo de amor. A partir de esto, se determina que tres de los seis tipos de amor (Eros, Manía y Ágape) no aportan mayor información para la caracterización, puesto que los resultados para los cuatro grupos son similares. En el caso de Eros, el grupo uno muestra una mediana más baja, pero al tener tanta variabilidad no se reconocen diferencias importantes respecto a los otros grupos, tal y como se muestra en la **Figura 3 (ver anexos)**. Por esta razón, se decide retirar estos tres tipos de amor, y se procede a crear nuevos conglomerados utilizando solo los tipos de amor Ludus, Storge y Pragma.

Posteriormente, se agrupan los individuos mediante tres diferentes técnicas de agrupamiento: Agrupamiento Jerárquico, donde se realizan particiones según los conglomerados que se fijan y el proceso no es reversible, esto implica que una vez asignado a un grupo, un elemento no podrá pasar a otro grupo; K-medias, en el que se calcula el centroide de cada conglomerado y se ubican los puntos cercanos a ese centroide por medio del cálculo de la distancia (euclídea para efectos de esta investigación); y finalmente K-medoids que es un método similar a K-medias, pero que toma un punto inicial arbitrario. En las **Tablas 5**

y 6 (ver anexos) se observa que los resultados de los distintos agrupamientos son consistentes entre sí. Se tomó la decisión de trabajar con k-medias; ya que, además de ser de fácil implementación, utilizar el centroide como punto inicial resulta adecuado para iniciar el proceso de agrupación.

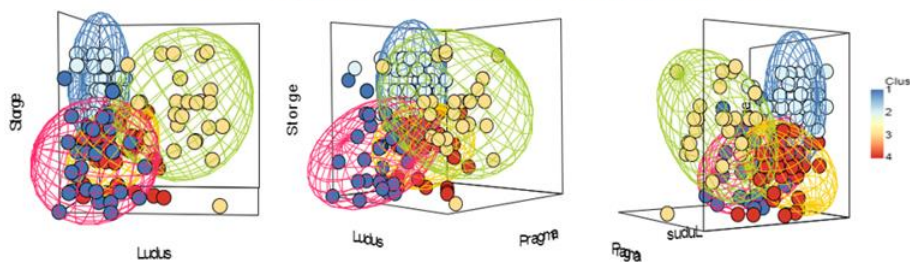
Una vez que cada individuo ha sido asignado a uno de los cuatro grupos, se realiza un análisis de resultados para observar si los grupos según el tipo de amor poseen relación con el signo del zodiaco los cuales están asociados con el mes y el día de nacimiento de las personas. Además, de las relaciones que pueden tener estas actitudes con otras variables de interés.

El software que se utilizó para el análisis de los datos fue R, versión 3.5.0., y diversos paquetes: lubridate, dplyr, cluster, rgl, canvasXpress, pvclust, ggplot2 y GGally (Bates et al., 2015; Deepayan, 2008; Sturtz, Set al., 2005).

RESULTADOS

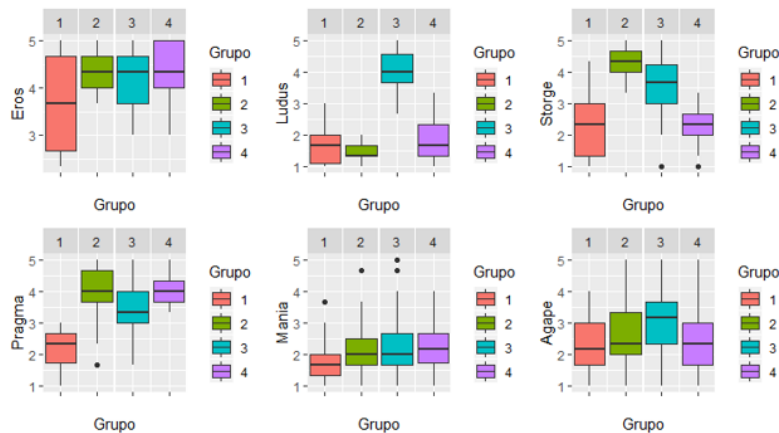
Tras asignar cada individuo a uno de los grupos, se puede observar (Figura 4), la formación de cuatro conglomerados bien definidos, que se proceden a caracterizar para posteriormente evaluar si existe una relación entre las actitudes hacia el amor y el signo zodiacal.

Figura 4. Diagrama de dispersión de los 4 conglomerados formados



Se puede observar en la Figura 5, que los tipos de amor Eros, Manía y Ágape siguen sin ser relevantes y por lo tanto no aportan mucha información para la caracterización. En el caso de Eros, se observa un comportamiento similar al descrito anteriormente, donde el único grupo que presenta una mediana diferente presenta mucha variabilidad. Respecto a los tipos de amor Ludus, Storge y Pragma se observa que poseen diferencias entre los grupos, por lo cual estos se consideran importantes para caracterizar los grupos.

Figura 5. Variabilidad de puntajes en los grupos según tipos de amor



En la **Tabla 1**, se muestra que los grupos 1, 2 y 4 poseen puntaje promedio bajo en el amor Ludus, mientras que el grupo 3 presenta puntaje promedio alto. Respecto al amor Pragma el grupo 1 presenta un puntaje promedio más bajo que el resto de los grupos. Finalmente, respecto al amor Storge, los grupos 1 y 4 presentan puntaje promedio más bajo que los grupos 2 y 3.

Es importante mencionar que, aunque el tipo de amor Manía como se mencionó anteriormente no posee diferencias importantes entre los clusters, en la **Tabla 2**, se observa que los puntajes son muy bajos (indistintamente del grupo), esto podría deberse a que para las personas es más difícil reconocer si tienen una actitud un tanto más posesiva hacia sus parejas.

Tabla 1. Puntaje promedio para cada tipo de amor según grupo

Tipo de Amor	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Eros	3.64	4.43	4.21	4.28
Ludus	1.17	1.42	3.96	1.81
Storge	2.27	4.25	3.54	2.29
Pragma	2.17	3.98	3.47	4.15
Mania	1.82	2.09	2.26	2.25
Agape	2.25	2.64	3.00	2.36

Respecto al género, se observa una mayor proporción de mujeres en los grupos 2 y 4, como se muestra en la **Figura 6 en anexos**. Por el contrario, en el grupo 3, se observa que este posee una mayor proporción de hombres. En la **Figura 7 en anexos**, se puede observar que con respecto a la orientación sexual todos los grupos están formados por una importante proporción de personas heterosexuales.

Con relación a la práctica de una religión, en la **Figura 8 en anexos**, se observa que no se encuentran grandes diferencias entre los grupos. El que no se detecten diferencias se puede deber a la cantidad de personas de la muestra, lo cual podría estar afectando los resultados del estudio. Respecto a la edad, en la **Figura 9 en anexos**, se aprecia que la distribución de edades es muy similar en los cuatro grupos.

Finalmente, respecto a la opinión de las personas sobre las relaciones abiertas, la **Figura 10 en anexos**. En general, las características de cada grupo se encuentran resumidas en la **Tabla 2** que se presenta a continuación.

Tabla 2. Resumen de caracterización según número de grupo

Grupo	Característica	Tendencia*	Promedio	Porcentaje
1	Ludus	Bajo**	1.17	
	Storge	Bajo**	2.27	
	Pragma	Bajo**	2.17	
	Considera estar en relación abierta	Alto		80%
2	Ludus	Bajo	1.42	
	Storge	Alto**	4.25	
	Sexo	Más mujeres		72%
	Considera estar en relación abierta	Alto		89.7%
3	Ludus	Alto**	3.96	
	Sexo	Más hombres		73%
4	Storge	Bajo	2.29	
	Pragma	Alto**	4.15	
	Sexo	Más mujeres		79%

* Respecto a los otros grupos

**Más alto o bajo de todos

Posteriormente, se realiza una comparación respecto al objetivo de la investigación, que es contrastar el signo zodiacal con los 4 grupos formados. Como se aprecia en el **Figura 11**, todos los grupos realizados mediante el análisis de agrupamiento tienen proporciones similares de participación en cada signo del zodiaco. El signo Leo es una excepción pues este signo zodiacal posee una proporción mayor de personas en el grupo 2, además, no cuenta con personas en el grupo 1. Por otro lado, nadie con los signos Géminis y Acuario forma parte del grupo 3. Esto puede deberse al tamaño de muestra y no se logra establecer un contraste importante entre el signo zodiacal y los grupos, por lo cual se procede a mostrar los resultados de los grupos con respecto a los elementos en los que son agrupados los signos zodiacales.

Figura 11. Distribución porcentual de cada grupo según signo zodiacal

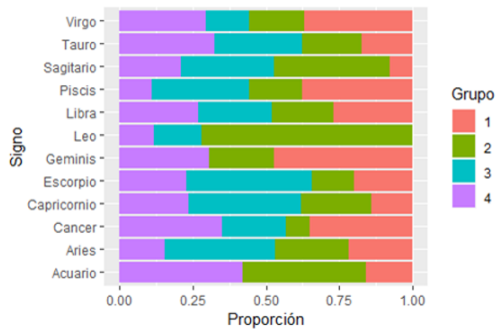
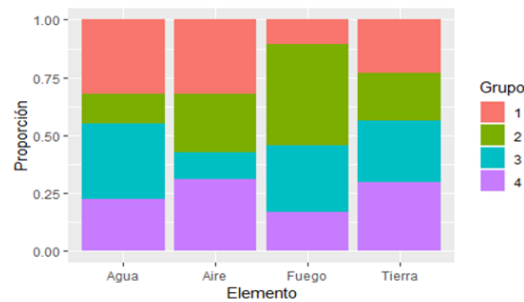


Figura 12. Distribución porcentual de cada grupo según elemento



En la **Figura 12**, se refleja que los elementos poseen individuos de todos los conglomerados, por lo cual no se presenta una relación importante entre el tipo de elemento y los grupos formados en el análisis, pero se muestra que los individuos del elemento Fuego se encuentran principalmente en el grupo 2, esto se debe a que el signo del zodiaco denominado Leo tiene como elemento el fuego. Adicionalmente, se analizan el sexo y los signos zodiacales tomando en cuenta los seis tipos de amor iniciales, esto para conocer el comportamiento de estas variables desde otra perspectiva.

En la **Tabla 3** se presenta el puntaje promedio de respuesta de actitud hacia cada tipo de amor según sexo y se observa que, para ambos, el tipo de amor Eros es en el que obtienen un mayor puntaje, mientras que Ludus y Manía tienen los puntajes más bajos, esto podría deberse a que para las personas es más difícil reconocer si tienen una actitud un tanto más desinteresada y/o posesiva hacia sus parejas. Por otro lado, resalta el hecho de que para amor Pragma las mujeres respondieron en promedio 0,46 puntos más alto que los hombres. Siendo la diferencia más alta entre ambos sexos.

Tabla 3. Puntaje de amor promedio según sexo

Tipo de amor	Hombres	Mujeres
Eros	4.08	4.22
Ludus	2.56	1.81
Storge	3.19	3.05
Pragma	3.24	3.70
Mania	2.26	2.02
Agape	3.11	2.20

Finalmente se compara cada signo con respecto a los puntajes en cada tipo de amor, tal como se muestra en la **Tabla 4**, donde se puede apreciar que en general, nuevamente es Eros el tipo de amor con promedios más altos.

Tabla 4. Puntaje promedio por tipo de amor según signo

Tipo de amor	Aries	Tauro	Geminis	Cáncer	Leo	Virgo	Libra	Escorpio	Sagitario	Capricornio	Acuario	Piscis
Eros	3.8	4.25	3.86	3.94	4.63	4.18	3.94	4.33	4.67	4.18	4.10	4.24
Ludus	2.17	2.21	1.92	2.17	1.78	1.78	2.29	2.43	2.18	2.43	1.24	2.10
Storge	3.40	3.10	2.50	2.61	3.29	2.85	2.81	3.21	3.67	3.29	3.00	3.21
Pragma	3.40	3.38	3.11	3.08	4.26	3.19	3.85	3.42	3.74	3.81	4.14	3.24
Mania	2.40	1.97	2.03	2.50	2.11	2.15	2.00	1.94	1.90	2.00	2.14	2.24
Agape	2.73	2.95	2.31	2.97	2.44	2.38	2.46	2.64	2.15	2.43	2.62	2.48

CONCLUSIONES

En el grupo 1, se encuentra individuos con bajos puntajes en Ludus, Storge y Pragma, lo que indica que son personas que tienen cierta implicación emocional en las relaciones, no consideran que la intimidad y el cariño sean fundamentales en su pareja y tampoco dan mucha importancia a aspectos lógicos a la hora de seleccionar su pareja. Adicionalmente, la mayoría de las personas que pertenecen a este grupo no han considerado estar en una relación abierta. El grupo 2 está formado por personas con bajo puntaje en amor Ludus y alto puntaje en amor Storge, esto implica que son individuos que tienen una implicación emocional en sus relaciones y consideran el amor, la amistad y la intimidad como fundamentales en sus relaciones. Este grupo está formado en su mayoría por mujeres y un 89.7% de las personas en este grupo nunca han considerado estar en una relación abierta. El grupo 3, se caracteriza por tener altos puntajes en el tipo de amor Ludus, su promedio es bastante alto respecto a los otros grupos, lo cual indica una poca implicación emocional y pocas expectativas futuras. Es importante observar que el grupo 3 se encuentra formado por una mayor proporción de hombres.

Finalmente, el grupo 4 está formado por personas con bajos puntajes en el tipo de amor Storge y alto amor Pragma. Esto se traduce en que no dan tanta importancia al hecho de que sus vínculos con la pareja procedan de una amistad o fuerte cariño, pero si a aspectos como si su pareja los puede apoyar en su carrera profesional o son potencialmente buenos padres o madres. Está formado principalmente por mujeres.

Con base en los resultados obtenidos, se observa que los individuos quedan repartidos en los diferentes grupos indiferentemente de su signo zodiacal, aquellos que comparten signo no quedan agrupados en un mismo conglomerado. Lo mismo sucede con el elemento al que pertenece el símbolo de la persona. Esto indica que la teoría astrológica, no está acertando, pues de hacerlo se notaría una tendencia a agrupar los mismos signos o elementos en un solo grupo. La única excepción es con el signo zodiacal Leo, que se caracteriza de acuerdo a la teoría por ser bondadoso, generoso, fiel, cariñoso y comprensivo con los demás, cualidades relacionadas al grupo 2, en el que se encuentra la mayoría de las personas de este signo.

Sin embargo, la relación entre los dos factores de estudio solo se encuentra en uno de los doce signos, se concluye que el signo del zodiaco no presenta una relación con las actitudes amorosas de las personas presentes en la muestra.

Por otro lado, cabe mencionar que existe un hallazgo interesante, aunque no está relacionado con el objetivo principal del estudio. Este se aprecia en la **Tabla 4** y es el hecho de que los puntajes en Ludus y Manía son los más bajos en la muestra, y esto no es común en estudios internacionales consultados como referencia. Esto puede deberse a la vergüenza o incapacidad de admitir actitudes desinteresadas y con poca implicación emocional o posesivas y controladoras. Por su parte, la puntuación promedio en el amor Agape presenta la diferencia más grande entre hombres y mujeres, esto es interesante puesto que los ítems para identificar esta actitud hacen referencia a cuestiones que implican sacrificar los propios intereses por los de la pareja, y el hecho de que muchas menos mujeres respondieron alto en esta sección, puede deberse a los actuales discursos de empoderamiento femenino, sobre todo tomando en cuenta el rango de edad y que la muestra en su mayoría pertenece a personas que cursan la universidad.

Siguiendo esa línea también resalta el hecho de que, las mujeres en su mayoría tienen actitudes que implican un razonamiento lógico a la hora de elegir pareja.

Como recomendaciones para posteriores estudios en relación del tema, se sugiere obtener una mayor muestra de personas, para así contar con más individuos en cada signo zodiacal y mejorar la calidad de los resultados. Además, sería importante dar una selección aleatoria de los sujetos de estudio.

BIBLIOGRAFÍA

- García, M. E., Garnica, M. C., González, N., Márquez, M., Martín, M., Pérez, M. M., & Vico, M. N. (2012). Las mujeres viven la relación romántica diferente al hombre. *Revista Electrónica de Investigación Docencia Creativa*, 1, 95-100. Recuperado de <http://digibug.ugr.es/handle/10481/21967#>. WUIEuGNpg.
- Goodman, L. (S.F). Los signos del zodiaco y su carácter. Recuperado de <https://www.astroworld.es/recursos/lindagoodman.pdf>
- Lascurain, P.C., Lavandera, M.C., Manzanares, E. (2017). Propiedades psicométricas de la escala de actitudes sobre el amor en universitarios peruanos. *Acta colombiana de Psicología*, 20(2), 270-281. doi: <http://www.dx.doi.org/10.14718/ACP.2017.20.2.13>
- Lotería Nacional para la Asistencia Pública. (2017). Este es tu elemento según tu signo del zodiaco. México: *Gobierno de México*. Recuperado de <https://www.gob.mx/lotenal/articulos/este-es-tu-elemento-segun-tu-signo-del-zodiaco>
- RAE. (2019). Diccionario de la lengua española. España: *Real academia española*. Recuperado de <https://dle.rae.es/?id=46Pv1Wk>
- RAE. (2019). Diccionario de la lengua española. España: *Real academia española*. Recuperado de <https://dle.rae.es/?id=46ggSe5>
- RAE. (2019). Diccionario de la lengua española. España: *Real academia española*. Recuperado de <https://dle.rae.es/?id=46ggSe5>
- Rodríguez, Y., Fernández, Carrera, M., Vallejo, P. (2013). Validación de la Escala de Actitudes hacia el Amor en una muestra de adolescentes. *Estudios de Psicología*, 2013, 34 (2), 209-219. DOI: 10.1174/021093913806751429

ANEXOS

Figura 1. Gráfico lineal de Suma de Cuadrados dentro de grupos según número de grupos.

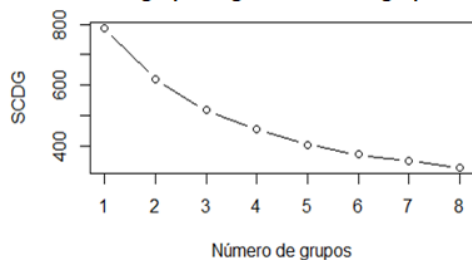


Figura 2. Dendrograma obtenido a partir de los datos recolectados.

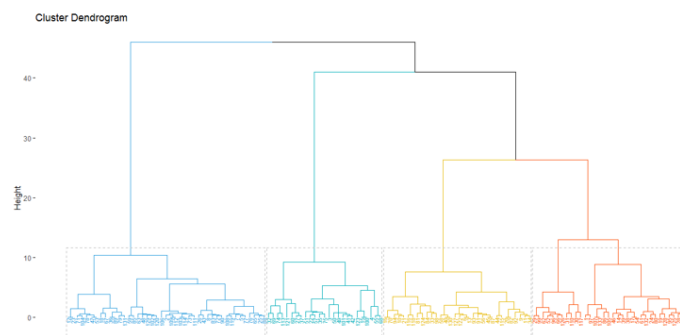


Figura 3. Puntaje de amor según grupo.

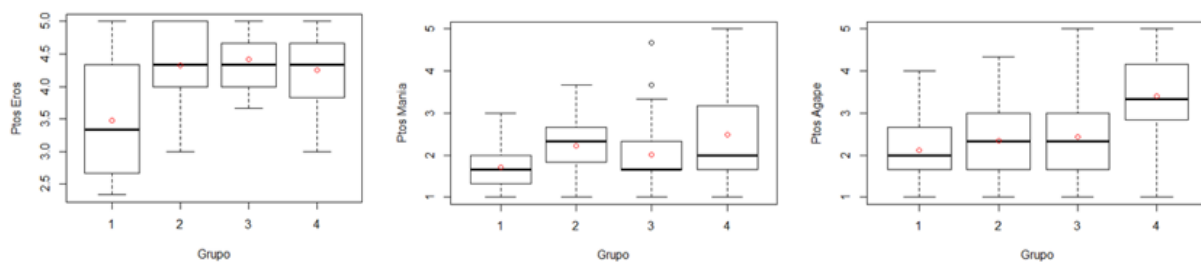


Tabla 5. Comparación de agrupamientos según K-Means y K-Medoids.

K-means	K-medoids			
	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Grupo 1	0	8	18	4
Grupo 2	1	36	0	2
Grupo 3	19	0	1	6
Grupo 4	0	6	0	32

Tabla 6. Comparación de agrupamiento según Ward y K-Means

K-means	Ward (Cutree)			
	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Grupo 1	26	0	1	3
Grupo 2	4	27	7	1
Grupo 3	2	0	24	0
Grupo 4	0	0	7	31

Figura 6. Proporción de género por grupo

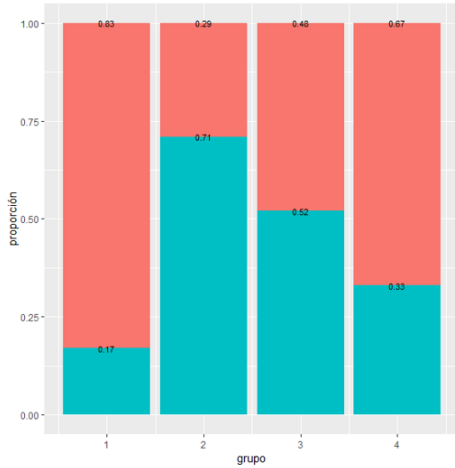


Figura 7. Proporción de orientación sexual por grupo

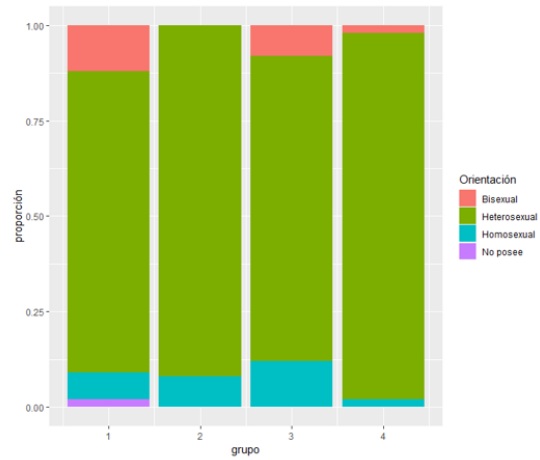


Figura 8. Proporción de practicantes de una religión según grupo

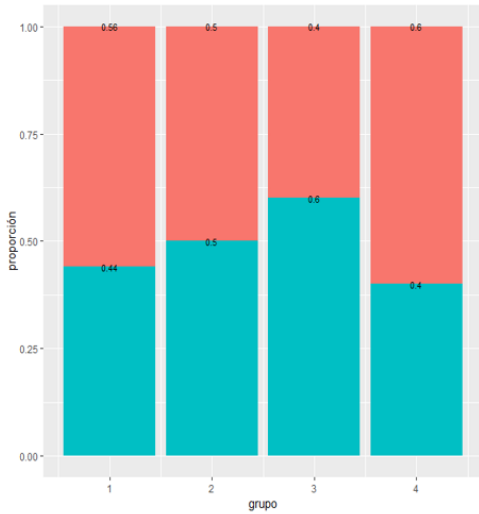


Figura 9. Proporción de edad según grupos

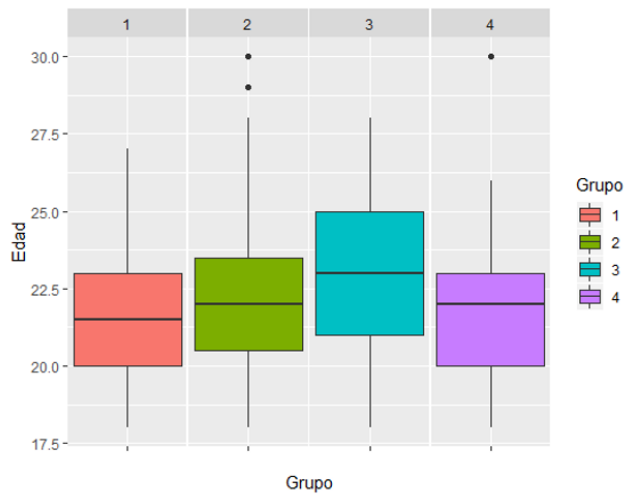
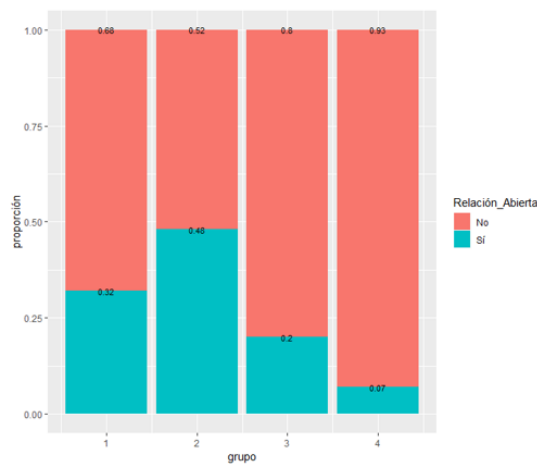


Figura 10. Proporción de relación abierta según grupos



Instrumento

Preferencias en relaciones sentimentales

El presente cuestionario es parte del curso Introducción al Análisis Multivariado, del Bachillerato de Estadística de la Universidad de Costa Rica. Le solicitamos cordialmente su colaboración para responderlo. El objetivo del proyecto es conocer las preferencias de las personas en cuanto a las relaciones sentimentales y determinar si existe alguna relación entre estas preferencias y la fecha y hora de su nacimiento.

Para propósitos del estudio se entenderá como relación si entre las partes existe un vínculo, algún tipo de atracción, y un acuerdo mutuo sobre los términos de la relación. Además, la edad debe estar entre los 18 y 30 años cumplidos.

La información proporcionada es totalmente confidencial. Por la naturaleza de la investigación, por favor solo responda el cuestionario si en este momento se encuentra en una relación acorde a la definición antes mencionada y si se encuentra dentro del rango de edad deseado.

*Obligatorio

1. Edad *

2. Fecha de Nacimiento *

Ejemplo: 15 de diciembre de 2012

3. Hora de Nacimiento

FORMATO DE 24 HORAS

Ejemplo: 8:30 a.m.

4. País de Residencia Actual *

5. Género *

Marca solo un óvalo.

Femenino Masculino Otro

6. Orientación Sexual

Marca solo un óvalo.

Heterosexual Homosexual Bisexual

Otro:

7. ¿Practica alguna religión?

Marca solo un óvalo.

Si No

8. En caso de responder sí, ¿Cuál religión practica?

9. Respecto a su último grado académico alcanzado

Marca solo un óvalo.

Primaria Incompleta Primaria Completa Secundaria Incompleta Secundaria Completa Cursando

Universidad Universitaria Completa

10. De acuerdo con la definición dada en la introducción; ¿Se encuentra actualmente en una relación sentimental? *

Definición: Entre las partes existe un vínculo, hay algún tipo de atracción, y un acuerdo mutuo sobre los términos de la relación.

Marca solo un óvalo.

Si

No Pasa a la pregunta 18.

Características

11. Por favor indique la cantidad de parejas sentimentales que ha tenido a lo largo de su vida.

12. ¿Alguna vez ha considerado estar en una relación abierta o practicar la poligamia?

Marca solo un óvalo.

Si No

13. ¿Considera que el amor es indispensable para mantener una relación sentimental con alguien?

Marca solo un óvalo.

Si No

14. ¿Cree en el amor a primera vista?

Marca solo un óvalo.

Si No

15. Respecto a su pareja actual normalmente, ¿Con cuánta frecuencia se ven físicamente?

Marca solo un óvalo.

Menos de una vez al mes Entre 1 y 3 veces al mes.

Al menos una vez por semana

Varias veces a la semana Todos los días

Viven juntos

16. Respecto a su pareja actual, indique brevemente ¿Cuál es la principal causa de conflictos en su relación?

Por ejemplo: Problemas de celos, diferencia de intereses, falta de atención, problemas relacionados a la familia, problemas económicos.

Escala de Actitudes Amorosas

17. A continuación se presentan una serie de afirmaciones, por favor indique que tan de acuerdo o desacuerdo está con cada una de ellas, donde el 1 es muy en desacuerdo y el 5 es muy de acuerdo.

•
Marca solo un óvalo por fila.

1 2 3 4 5

Mi pareja y yo tenemos "buena química".

Siento que mi pareja y yo estamos hechos el uno /a para el otro /a.

Mi pareja responde a mi ideal de belleza física.

Hay cosas sobre mí que prefiero que mi pareja no sepa

A veces he tenido que ocultar cosas sobre mis parejas anteriores.

Pienso que mi pareja se enfadaría si supiera algunas cosas que he hecho con otras personas.

El amor más profundo nació de una larga amistad.

Entre mi pareja y yo nuestra amistad inicial se transformó gradualmente en amor.

Mis relaciones amorosas más satisfactorias se han desarrollado a partir de relaciones de amistad.

Una de las cosas esenciales a la hora de elegir a la pareja es la aceptación de nuestra familia hacia él/ella.

Un criterio importante en la elección de pareja es saber si él/ella podrá ser un buen padre/madre.

Un criterio importante a la hora de elegir pareja es saber si él /ella podrá apoyarnos en nuestra carrera profesional.

Cuando mi pareja no me hace caso, me siento fatal.

Cuando estoy enamorado /a me cuesta concentrarme sobre cualquier tema que no sea mi pareja.

Si mi pareja me ignora durante cierto tiempo, hago cosas sin sentido para tratar de atraer de nuevo su atención.

Prefiero sufrir yo antes de que sufra mi pareja.

Mi felicidad depende de la felicidad de mi pareja.

Estoy generalmente dispuesto/a a sacrificar mis propios deseos para que mi pareja pueda realizar los suyos.

Muchas Gracias

18. En caso de que desee recibir los resultados de esta investigación, indique el correo electrónico y se le hará llegar una vez concluida.

Análisis sobre tipos de inteligencia y pasatiempos de los estudiantes activos de las Facultades de Ciencias Sociales, Educación, Ciencias Económicas e Ingeniería de la Universidad de Costa Rica

Shirley Herrera Delgado¹², Noelia Rojas Ramírez¹², María José Solís Quirós¹²

auxirleyucr31@gmail.com, mnrr.10@gmail.com, mari.solisquiros@gmail.com

RESUMEN

La inteligencia es una característica que se ha estudiado durante muchos años con el objetivo de descifrar y poder entender cuáles son los factores que influyen en ella y su impacto a nivel educativo, profesional y personal. Sin embargo, hay teorías que proponen que no hay una única inteligencia y que de la misma forma en que cada individuo cuenta con experiencias, conocimientos, habilidades y limitaciones diferentes, también posee diversas formas de razonar y percibir el mundo. Entre estas teorías destaca la de Howard Gardner, quien propone las inteligencias lingüísticas, lógico-matemática, visual-espacial, corporal y kinestésica, musical, intrapersonal, interpersonal y naturalista. Por lo que, este análisis se enfocó en explorar las similitudes de estudiantes activos de 17 años en adelante de la Universidad de Costa Rica pertenecientes a las Facultades de Ciencias Sociales, Educación, Ciencias Económicas e Ingeniería con el objetivo de distinguir grupos dependiendo de los tipos de inteligencia y caracterizarlos según sus hábitos y preferencias relacionadas al tiempo libre. Se realizó un análisis estadístico de conglomerados tomando como variables de agrupamiento los tipos de inteligencia y como variables de caracterización los hábitos y preferencias del tiempo libre. Se concluye que los estudiantes del grupo 1 tienen puntajes más altos en las inteligencias lógico-matemática y espacial mientras que los estudiantes del grupo 2 tienden a puntuar más alto en la inteligencia de tipo interpersonal.

PALABRAS CLAVE: tipos de inteligencia, pasatiempos, agrupamiento, habilidades cognitivas.

ABSTRACT

Intelligence is a characteristic than have been studied for a long time with the goal of discover and understand which are the factors involved in it and the impact of education, professional and personal area. However, there are theories that propose that there is no a unique intelligence and in the same way people have their experience, knowledge, abilities and limitations, also have different ways to think and perceive the world. One of this is Howard Gardner's theory, who proposed the following intelligence: linguistic, logical-matematical, visual-spatial, kinesthetic, musical, intrapersonal, interpersonal and naturalist. Thus, this analysis was focused to explore the similarities of students of University of Costa Rica from 17 years belonging to the faculty of social sciences, economics, education and engineering with the aim to distinguish groups depend of the type of intelligence and describe by the habits and preference in the free time. It has been made a clustering analysis taking as grouping variables the types of intelligence and as characterization variables the hobbies. It concludes that students of group 1 they have higher scores in logical-matematical and spatial while the students of group 2 have higher scores in interpersonal intelligence.

KEYWORDS: intelligence types, clustering, hobbies, cognitive abilities.

¹² Estudiantes de Estadística de la Universidad de Costa Rica



INTRODUCCIÓN

La sociedad durante muchos años se ha caracterizado por valorar la inteligencia, considerándola como un factor determinante para alcanzar las metas educativas, un trabajo exitoso y el desarrollo socioeconómico de los pueblos. Por lo que, durante siglos los psicólogos, filósofos, educadores y el público en general, se han interesado en su estudio analizando el concepto, medición, los tipos de inteligencia, las implicaciones sociales y políticas, sus orígenes evolutivos, entre otros (Ardila, 2011).

Es importante aclarar que se entiende por inteligencia la capacidad de resolver problemas genuinos, dificultades, además de un producto efectivo valioso en uno o más ambientes culturales. Con esto se pueden establecer bases para adquirir un nuevo conocimiento, según Gardner (2001).

Las personas poseen múltiples características en común entre unos y otros, pero también se pueden hallar aspectos únicos que identifican a cada quién. Aun cuando han contado con la misma estructura genética y comportamientos como consecuencia de igual aprendizaje temprano, aptitudes, intereses, habilidades y limitaciones; cada persona es única por cómo se combinan esos elementos y factores (Ardila, 2011).

Además, cuando se considera la forma de conocer el mundo, esta también se realiza de forma distinta, por lo que se han propuesto teorías. Entre ellas, la de Howard Gardner, que propone las siguientes ocho inteligencias, lingüística, lógico-matemática, visual-espacial, corporal y kinestésica, musical, intrapersonal, interpersonal y naturalista. Cada tipo de inteligencia y sus posibles combinaciones en un individuo se deben considerar en el ámbito de la educación, ya que, mantiene elementos propios para atender la diversidad de los estudiantes y satisfacer sus tendencias de enseñanza-aprendizaje (Sánchez, 2015).

Por otro lado, los hábitos de tiempo libre o pasatiempos se consideran determinantes en la construcción de identidades, en especial en el caso de los jóvenes, como los universitarios, profesionales, entre otros (Ureta, 2015). Además, Sánchez (2015) menciona en su trabajo de investigación que algunas teorías se desarrollan con la intención de explicar la inteligencia no sólo como un factor genético, sino que depende de la sociedad, el ambiente, la cultura y otros factores, resaltando que los otros factores ambientales tienen una cierta relación con factores externos, pudiendo considerar como esos factores externos los pasatiempos que las personas prefieren, es decir, que esto puede ser una característica que puede dar indicios del tipo de inteligencia que tiene el individuo.

Según Pérez (2014), los avances que se han realizado en el campo de la neurociencia han ofrecido información relevante para el campo educativo, por el hecho de que al conocer cómo funciona el cerebro de las personas se permite diseñar y programar actividades para desarrollar determinadas áreas cerebrales en función de sus propósitos. Estos estudios junto con los que se han realizado en la rama de neuropsicología llevan a la importancia del fomento de ciertas capacidades humanas como lo es la creatividad y las inteligencias múltiples.

Es por estas razones que con este análisis se pretende describir las características de los distintos grupos dependiendo de los tipos de inteligencia de los estudiantes activos de la UCR que pertenecen a la Facultad de Ciencias Sociales, Educación, Ciencias Económicas e Ingeniería, de acuerdo con características sociodemográficas y referentes a sus hábitos y preferencias en su tiempo libre.

METODOLOGÍA

La toma de datos se realizó con un cuestionario autoadministrado a través de Google forms, aplicado a estudiantes activos de las facultades de Ciencias Sociales, Educación, Ciencias Económicas e

Ingeniería de la Universidad de Costa Rica, sede Rodrigo Facio Brenes, durante el 01 al 15 de mayo 2019. Se contó con la participación total de 325 estudiantes distribuidos en 208 mujeres y 117 hombres, de los cuales 69 estudiantes pertenecen a la Facultad de Ciencias Económicas, 46 estudiantes a la de Ciencias Sociales, 58 estudiantes a la de Educación y 152 estudiantes de Ingeniería. Debido a la forma en que se recolectaron los datos y al carácter voluntario del cuestionario, los resultados que se generen podrán ser únicamente descriptivos y no se pretende generalizar al resto de la población.

El formulario se dividió en 3 secciones: información personal, pasatiempos y tipos de inteligencia; además de una pequeña introducción donde se detalló el objetivo del estudio y se aclararon los fines académicos del mismo (ver Anexo 1). Se tomó en cuenta las siguientes variables:

Variables de caracterización:

- Edad de 17 años en adelante.
- Sexo hombre o mujer.
- Facultad a la que pertenece su carrera.
- Pasatiempos o hobbies: 11 preguntas relacionadas a qué tanto le gusta una serie de pasatiempos con opciones de respuesta de “me gusta mucho”, “me gusta algo” y “no me gusta” denotando la intensidad con la que a cada estudiante le gusta hacer determinada actividad. Además, otras preguntas que contemplan cuál es el objetivo de realizar hobbies, cuánto tiempo en horas les dedica a la semana, si los hobbies requieren de inversión económica o no y en caso de que sí, cuánto debe invertir aproximadamente por semana.

Variables de agrupamiento:

- Tipos de inteligencia: Se incluyó una batería de 35 preguntas que determinan características para 7 tipos de inteligencia, de acuerdo con las preferencias para aprender y percibir el mundo. En cada pregunta se asigna un 1 en caso de ser una respuesta afirmativa y un 0 en caso de ser falsa, esto para posteriormente obtener los puntajes para cada tipo de inteligencia según la escala estructurada obtenida de una serie de observaciones antropológicas realizadas a diversas culturas por el psicólogo, investigador y profesor de la Universidad de Harvard, Howard Gardner (Gamandé, 2014). Por lo tanto, el puntaje para cada tipo de inteligencia varía en un rango de 0 a 5 evaluando los siguientes tipos de inteligencia: lingüística, lógico-matemática, visual-espacial, corporal y kinestésica, musical, intrapersonal e interpersonal.

Para explorar los datos antes del análisis de agrupamiento, se identificaron los valores extremos por medio de los leverages y se eliminaron 6 observaciones para evitar que afectara el análisis. Además, se obtuvo la correlación entre los tipos de inteligencia para evaluar el requerimiento de hacer o no un análisis de componentes principales concluyendo que no era necesario. Las variables de los índices no se estandarizaron debido a que las 7 tenían el rango de 0 a 5, por lo que no resultó indispensable.

Posteriormente, para obtener grupos de estudiantes de acuerdo a similitudes en los tipos de inteligencia, se realizó el análisis de conglomerados de forma exploratoria con las variables de agrupamiento. Para esto, se tomó en cuenta las técnicas de agrupamiento jerárquico y k-medias, así como de las diferentes distancias entre individuos como la euclídea, de mahalanobis, de manhattan y chebyshev; y se contempló el agrupamiento por medio del método del vecino más cercano, vecino más lejano y salto promedio.

Este análisis exploratorio de los posibles grupos obtenidos de las combinaciones entre las diferentes distancias entre variables y las distancias entre los grupos según las variables incluidas en el agrupamiento; se hizo con el objetivo de conseguir el mayor porcentaje de variabilidad explicada. Gráficamente, se utilizaron dendrogramas para explorar la cantidad de grupos y gráficos de cajas para indagar el comportamiento de los tipos de inteligencia dentro de cada uno de los posibles grupos.

Adicionalmente, se exploró combinaciones entre los tipos de inteligencia hasta lograr un porcentaje considerable de esa variabilidad que no estuviera siendo afectado por variables que incluían ruido en el análisis de agrupamiento.

El agrupamiento que se eligió para la caracterización fue con la técnica de k medias y diferenciando a los estudiantes en dos grupos: de donde se obtiene la suma de cuadrados dentro de grupos, representado de la siguiente forma:

$$SCDG = \sum_{h=1}^k \sum_{i \in G_h} \sum_{j=1}^q (x_{ij} - x_j^{(h)})^2.$$

Donde:

x_{ij} : es el individuo ij

$x_j^{(h)}$: es el promedio en cada variable dentro de cada grupo

Para caracterizar los grupos se comparó la conformación de cada uno de acuerdo a las mismas variables de agrupamiento y se contrastó con los datos de los pasatiempos y variables como el sexo y la facultad de los estudiantes.

Para el análisis se utilizó el lenguaje de programación R, mediante el software libre *Rstudio* (R Core Team, 2019) en su versión 3.6.0 y haciendo uso de las librerías *corrgram* (Wright, 2018), *ggplot2* (Wickham, 2016), *tidyverse* (Wickham, 2017) y *biotools* (Da Silva et al., 2017).

RESULTADOS

Al analizar las correlaciones entre las variables resultaron ser bajas en su mayoría, ya que las más altas son entre los índices de inteligencia kinestésica e interpersonal, entre interpersonal e intrapersonal y entre la inteligencia lógica y espacial con 0.24, 0.36 y 0.37, respectivamente (Ver Anexo 2, Figura 1).

Con base a la comparación del método k medias y el agrupamiento jerárquico usando vecino más cercano, según tres de las distancias entre individuos, Euclídea, Mahalanobis y Manhattan (Ver Anexo 3, Cuadro 1), además de los dendrogramas generados (Ver Anexo 2, Figura 2), se decidió usar dos grupos de estudiantes con el método K medias.

De la indagación previa con los distintos índices de tipos de inteligencia, se obtuvo como opción el agrupamiento con las variables lógico-matemática, visual-espacial e interpersonal, formando dos grupos con el método de k medias con aproximadamente un 80% de variabilidad explicada (Anexo 2, Figura 3). La composición de cada grupo se detalla a continuación:

- Grupo 1: consta de 196 estudiantes activos con predominancia de puntuaciones altas para las inteligencias lógica-matemática y espacial.
- Grupo 2: se conforma con 123 estudiantes activos con predominancia de puntuaciones altas para la inteligencia interpersonal.

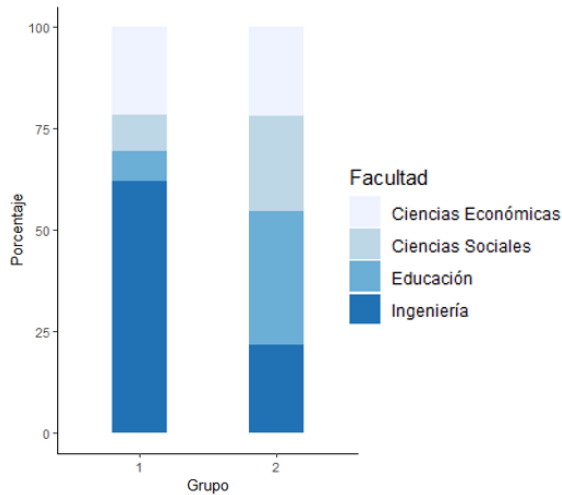
De acuerdo con la Facultad a la que pertenecen los estudiantes se observa que la distribución de estos se diferencia entre los grupos, dado que, en el grupo 1 la mayoría son de carreras que pertenecen a las Facultades de Ingeniería y Ciencias Económicas mientras que en el grupo 2 hay mayor porcentaje de personas de las carreras que pertenecen a las Facultades de Ciencias Sociales y Educación (Ver Figura 4).

Al contrastar las características se obtiene que la mayoría de los estudiantes de estas facultades se diferencian acorde con los tipos de inteligencia. Ya que, es de esperarse que las personas que tengan puntajes más altos en el tipo de inteligencia lógica-matemática, sean de facultades que demanden ciertas habilidades acorde con las carreras, mientras que las personas que estudien carreras relacionadas al área

social sean personas que tengan más habilidades para comprender y relacionarse con los demás (Ver Anexo 2, Figura 5).

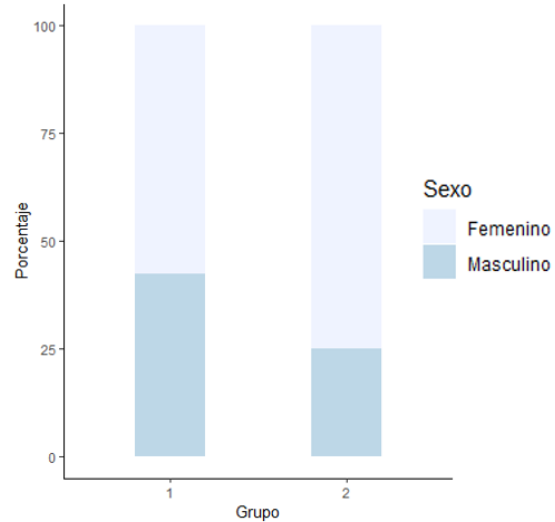
Con respecto al sexo se obtiene que el grupo 1 hay mayor porcentaje de estudiantes de sexo masculino en comparación con el porcentaje de hombres que hay en el grupo 2, tal y como se muestra en la Figura 6.

Figura 4. Facultad a la que pertenecen los estudiantes



Fuente: Elaboración propia.

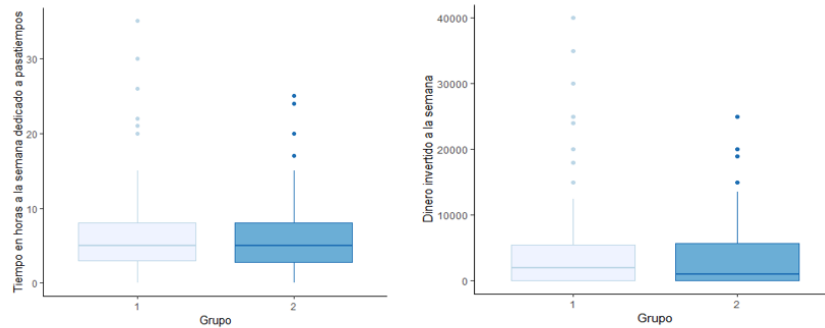
Figura 6. Sexo de los estudiantes por grupo



Fuente: elaboración propia.

Por otra parte, se encontró que los estudiantes dedican aproximadamente cinco horas y dos mil colones a la semana para practicar sus pasatiempos. La inversión en tiempo y dinero en los pasatiempos, es similar entre los grupos (Ver Figura 7).

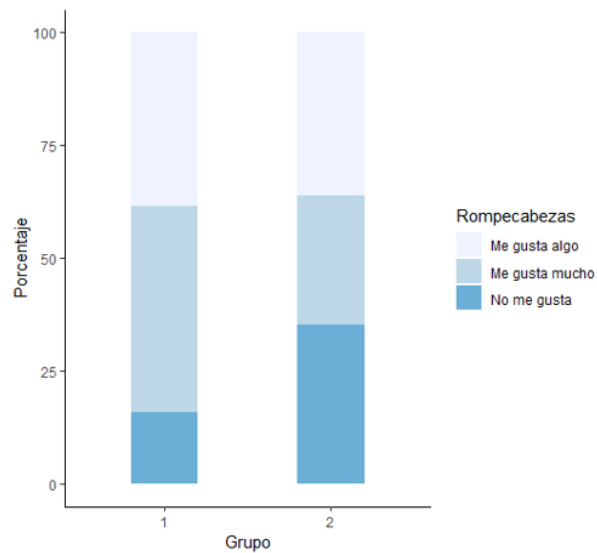
Figura 7. Tiempo invertido y dinero (colones) por semana en los pasatiempos según el grupo



Fuente: elaboración propia.

Al contrastar los grupos obtenidos con la intensidad con que les gustan a los estudiantes determinados pasatiempos o no, se revela que cerca de un 84% de los estudiantes del grupo 1 manifiesta que les gusta algo o mucho armar rompecabezas o variantes como crucigramas o sudokus, mientras que en el grupo 2, el porcentaje llega a cerca del 65% para las mismas categorías (Ver Figura 8).

Figura 8: Preferencia del pasatiempo rompecabezas por grupo



Fuente: elaboración propia.

CONCLUSIONES

El análisis mostró la presencia de un perfil de estudiantes en el que prevalecen y se caracterizan principalmente según tres tipos de inteligencias: lógico-matemática, espacial e interpersonal. Con estos tipos de inteligencia se logran diferenciar dos grupos de estudiantes donde se da una combinación de las dos primeras inteligencias observadas mayoritariamente en estudiantes de las Facultades de Ingeniería y Ciencias Económicas mientras que para el tipo de inteligencia interpersonal resulta ser un rasgo más característico de los estudiantes que cursan carreras de las Facultades de Ciencias Sociales y Educación.

Esto resulta importante porque al considerar los tipos de inteligencia como un conjunto de fortalezas para desempeñarse en distintos ámbitos de la vida, es de esperarse que los estudiantes de cada una de esas facultades compartan similitudes relacionadas con el tipo de inteligencia, dando señales de que eventualmente también comparten habilidades debido al desempeño actual en sus carreras y a futuro a nivel profesional.

Para los estudiantes de las carreras de la Facultad de Ingeniería y Ciencias Económicas, la combinación de estos dos tipos de inteligencia que resaltan puede darse por las similitudes individuales que hay en las preferencias y los comportamientos personales de cada estudiante. Es decir, entre todos estos estudiantes comparten características comenzando por haber escogido estudiar dichas carreras afines sin dejar de lado las diferencias individuales de cada uno, de la misma forma sucede para los estudiantes de las carreras de la Facultad de Ciencias Sociales y Educación.

Se puede considerar que esta conformación de grupos toma en cuenta las exigencias de las carreras en cada facultad analizada, ya que podrían requerir dentro de su perfil de ingreso las inteligencias que más predominaron en cada grupo como destrezas o habilidades esperadas. Por ejemplo, las carreras de la Facultad de Ingeniería y Ciencias Económicas podrían requerir que sus estudiantes tengan destrezas matemáticas mientras que para las carreras de la Facultad de Ciencias Sociales y Educación es necesario que sean estudiantes con empatía y capacidad de comprender lo que sucede con otras personas.

No obstante, es necesario resaltar que independientemente de la carrera o Facultad en que estén empadronados los estudiantes que aportaron a este estudio, todos tienen un poco de cada tipo de inteligencia, pero con desarrollos diferentes, por lo que cada uno tiene un perfil de inteligencia específico en las que todas las demás inteligencias hacen pequeñas contribuciones para dar ese perfil al estudiante.

La importancia de conocer los perfiles de inteligencias múltiples obtenidos en la investigación pueden ser un insumo para reestudiar las prácticas en la educación universitaria llegando a ser un recurso docente para idear técnicas y estrategias de aprendizaje que más se adecúen a sus estudiantes, con el propósito de lograr una mejor transferencia de conocimientos por parte del educador, independientemente del perfil de inteligencia que tenga y tomando en cuenta las similitudes en los tipos de inteligencia y habilidades de cada grupo de estudiantes.

REFERENCIAS

- Ardila, R. (2011). Inteligencia ¿Qué sabemos y qué nos falta por investigar? [en línea]. Recuperado de <http://www.scielo.org.co/pdf/racefn/v35n134/v35n134a09.pdf> [consultado el 10 de mayo 2019].
- Da Silva, A.R.; Malafaia, G.; Menezes, I.P.P. (2017). biotools: An R function to predict spatial gene diversity via an individual-based approach. *Genetics and Molecular Research*, 16: gmr16029655.
- Gardner, H. (2001). Estructuras de la Mente, la Teoría de Las Inteligencias Múltiples [en línea]. Recuperado de: https://utecno.files.wordpress.com/2014/07/howard_gardner_-_estructuras_de_la_mente.pdf [consultado el 10 de mayo 2019].
- Gamandé, N. (2014). Las inteligencias múltiples de Howard Gardner: unidad piloto para propuesta de cambio metodológico. [en línea]. Recuperado de <https://reunir.unir.net/bitstream/handle/123456789/2595/gamande%20villanueva.pdf?sequence=1&isAllowed=y> [consultado el 10 de mayo 2019].
- Pérez, M. (2014). Proyecto creativo a través de las inteligencias múltiples en primaria [en línea]. Recuperado de: https://reunir.unir.net/bitstream/handle/123456789/3024/Maria_Perez_Espinosa.pdf?sequence=1&isAllowed=y [consultado el 10 de mayo 2019].
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. url: <https://www.R-project.org/>.
- Sánchez, I. (2015). La teoría de las inteligencias múltiples en la educación [en línea]. Recuperado de: http://unimex.edu.mx/Investigacion/DocInvestigacion/La_teor%C3%ADa_de_las_inteligencias_m%C3%9Altiples_en_la_educacion.pdf [consultado el 10 de mayo 2019].
- Ureta, X. (2015). Estilos de vida y prácticas de ocio en estudiantes de la Licenciatura en Educación Física, Recreación y Deporte del IUACJ. *Revista Universitaria de la Educación Física y el Deporte*. Año 8, Nº 8, Noviembre de 2015. Montevideo (Uruguay). p. 49 - 58. ISSN 1688-4949.
- Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. url: <https://CRAN.R-project.org/package=tidyverse>
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. url: <https://ggplot2.tidyverse.org>
- Wright, K. (2018). corrgram: Plot a Correlogram. R package version 1.13. url: <https://CRAN.R-project.org/package=corrgram>

ANEXOS

1

Anexo 1. Cuestionario Tipos de Inteligencia

El siguiente instrumento tiene como finalidad recolectar información para el curso XS-4110 Introducción al Análisis Multivariado de cuarto año de la carrera de Estadística de la Universidad de Costa Rica. Se pretende conocer si los estudiantes activos de la UCR pertenecientes a las facultades de Ciencias Sociales, Ciencias Económicas, Educación e Ingeniería tienen habilidades en común que caracterizan alguno de los tipos de inteligencia y poder relacionarlo con los hobbies de los estudiantes. La información proporcionada será confidencial y utilizada únicamente para fines académicos.

Información personal

Edad

Sexo

Femenino Masculino

¿Su colegio de procedencia es?

Público Privado Otro

¿Cuál era el enfoque (modalidad) de la institución en la cual cursó el ciclo de Educación Diversificada (últimos 2 o 3 años de colegio)?

Científico Académico Artístico Otro: _____

Facultad en la que tiene la mayoría de los créditos matriculados durante este semestre:
Ciencias Sociales Ingeniería Educación

Carrera en la que está empadronado (primera opción de carrera)

Primero queremos saber sobre sus hobbies o pasatiempos.

¿Indique en la siguiente lista de hobbies si le gusta mucho, le gusta algo o no le gusta cada uno de ellos?

	Me gusta mucho	Me gusta algo	No me gusta
Practicar deportes			
Tocar un instrumento musical			
Arte plástico: pintura, escultura, diseño gráfico			
Coleccionismo: colecciona cualquier tipo de objeto que le apasione.			
Lectura			
Escritura creativa: escribe literatura de diversos géneros			
Cocinar			
Confeccionar manualidades			
Senderismo: es el hobby de salir a caminar a lugares			
Danza: baile popular, danza moderna, flamenco, ballet			
Rompecabezas: crucigramas y sus variantes como el sudoku y las sopas de letras			

¿Qué le gustaría lograr a través de su hobby?

Mejorar la condición física Mejorar la memoria Encontrar nuevos amigos Desenvolverse en el arte Otro:

¿Con cuánto tiempo (en horas) dispone por semana para dedicarle a su hobby?

¿Su hobby requiere de una inversión económica?

Sí No

¿Cuánto dinero debe invertir por semana? (en caso de que su hobby no requiere de inversión económica, continúe con la siguiente sección)

Tipos de inteligencia

A continuación, se presentan una serie de afirmaciones. Si expresan características fuertes en su persona y le parece que la afirmación es veraz entonces indique que es verdadera y si no lo es, entonces indique que es falsa.

	VERDADERO	FALSO
Prefiero hacer un mapa que explicarle a alguien como tiene que llegar.		
Si estoy enojado(a) o contento (a) generalmente sé exactamente por qué.		
Sé tocar (o antes sabía tocar) un instrumento musical.		
Asocio la música con mis estados de ánimo.		
Puedo sumar o multiplicar mentalmente con mucha rapidez.		
Puedo ayudar a un amigo a manejar sus sentimientos porque yo lo pude hacer antes en relación a sentimientos parecidos.		
Me gusta trabajar con calculadoras y computadores.		
Aprendo rápido a bailar un ritmo nuevo.		
No me es difícil decir lo que pienso en el curso de una discusión o debate.		
Disfruto de una buena charla, discurso o sermón.		
Siempre distingo el norte del sur, esté donde esté.		
Me gusta reunir grupos de personas en una fiesta o en un evento especial.		
La vida me parece vacía sin música.		
Siempre entiendo los gráficos que vienen en las instrucciones de equipos o instrumentos.		
Me gusta hacer rompecabezas y entretenerme con juegos electrónicos.		
Me fue fácil aprender a andar en bicicleta, (o patines).		
Me enoja cuando oigo una discusión o una afirmación que parece ilógica.		
Soy capaz de convencer a otros que sigan mis planes.		
Tengo buen sentido de equilibrio y coordinación.		
Con frecuencia veo configuraciones y relaciones entre números con más rapidez y facilidad que otros.		
Me gusta construir modelos (o hacer esculturas).		
Tengo agudeza para encontrar el significado de las palabras.		
Puedo mirar un objeto de una manera y con la misma facilidad verlo cuando le doy vuelta o lo pongo revés.		
Con frecuencia hago la conexión entre una pieza de música y algún evento de mi vida.		
Me gusta trabajar con números y figuras.		
Me gusta sentarme silenciosamente y reflexionar sobre mis sentimientos íntimos.		
Con sólo mirar la forma de construcciones y estructuras me siento a gusto.		
Me gusta tararear, silbar y cantar en la ducha o cuando estoy solo (a).		
Soy bueno(a) para el atletismo.		
Me gusta escribir cartas detalladas a mis amigos.		
Generalmente me doy cuenta de la expresión que tengo en la cara.		
Me doy cuenta de las expresiones en la cara de otras personas.		
Me mantengo "en contacto" con mis estados de ánimo. No me cuesta identificarlos.		
Me doy cuenta de los estados de ánimo de otros.		
Me doy cuenta bastante bien de lo que otros piensan de mí.		

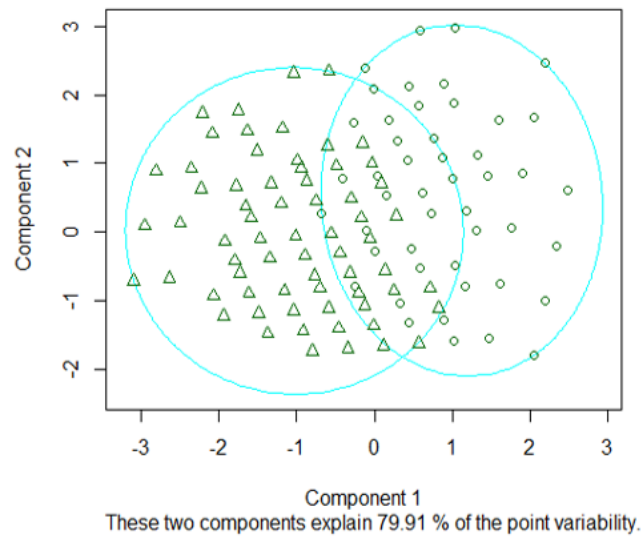
2

Figura 1. Correlaciones entre las distintas inteligencias

kinest	0.12 (0.01, 0.22)	0.24 (0.14, 0.34)	0.10 (-0.00, 0.21)	0.02 (-0.09, 0.12)	0.05 (-0.06, 0.16)	0.17 (0.06, 0.27)
musical	0.15 (0.04, 0.26)	-0.01 (-0.12, 0.10)	0.06 (-0.05, 0.17)	-0.05 (-0.16, 0.05)	-0.00 (-0.11, 0.11)	
inter	0.35 (0.25, 0.44)	0.21 (0.11, 0.31)	-0.05 (-0.16, 0.05)	0.03 (-0.08, 0.13)		
intra	0.14 (0.03, 0.24)	-0.12 (-0.23, -0.01)	-0.01 (-0.12, 0.10)			
verbal	-0.15 (-0.26, -0.04)	-0.04 (-0.15, 0.07)				
logica	0.35 (0.25, 0.44)					
espacial						

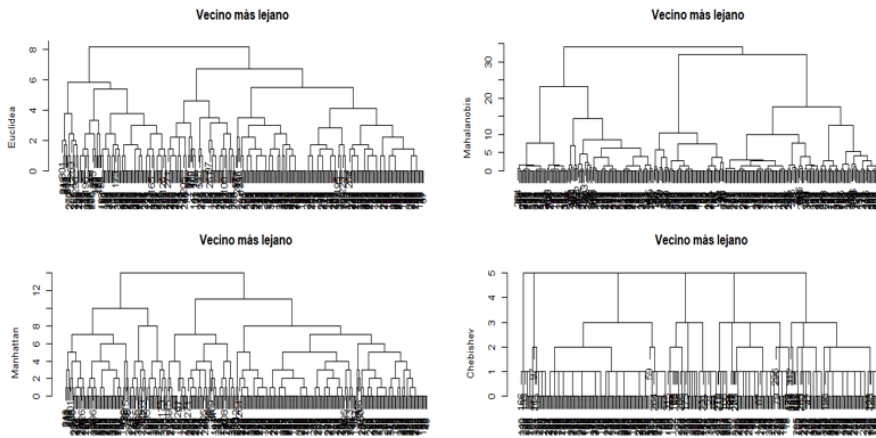
Fuente: elaboración propia.

Figura 3. Agrupamiento con dos clúster usando el método de k-medoids



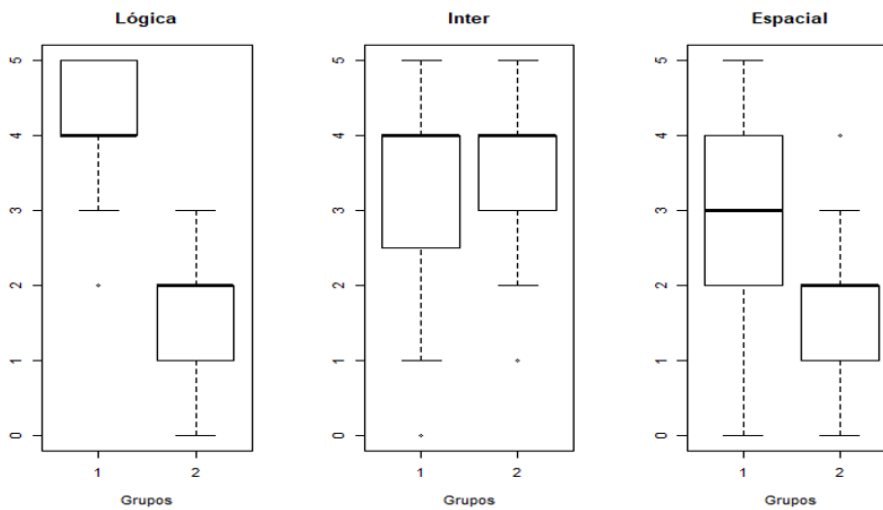
Fuente: elaboración propia.

Figura 2: Dendrogramas de agrupamiento jerárquico según cuatro de las distancias entre grupo y con el método del vecino más lejano



Fuente: elaboración propia.

Figura 5. Comportamiento de los tipos de inteligencia para cada grupo de estudiantes



Fuente: elaboración propia.

3

Cuadro 1. Comparación de los grupos obtenidos con agrupamiento jerárquico y k medias

Tipo de distancia entre grupos	Núm. de grupo	K-medias	
		Grupo 1	Grupo 2
Euclidea	1	195	31
	2	1	92
Mahalanobis	1	131	68
	2	65	55
Manhattan	1	196	33
	2	0	90

IV. MINERÍA DE DATOS

La minería de datos es el conjunto de técnicas y tecnologías que posibilitan manejar grandes bases de datos, de manera automática o semiautomática, con la intención de encontrar ciertos patrones repetitivos, tendencia o medidas que expliquen el comportamiento de un determinado contexto (Gibert, K., Ruiz, R., & Riquelme, J., 2019)¹³. Los patrones desconocidos se clasifican en grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). Cabe destacar que la minería de datos es muy utilizada en el análisis adicional y análisis predictivo. Dichas técnicas fueron apareciendo a principios de los años ochenta, por medio de Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, que empezaron a consolidar los términos de minería de datos (Virsedá, F., & Román, J. 2010)¹⁴.

¹³ Gibert, K., Ruiz, R., & Riquelme, J. (2019). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, ISSN 1137-3601, No. 29, 2006 (Ejemplar Dedicado a: Minería de Datos), Pags. 11-18. Recuperado de https://www.researchgate.net/publication/28140440_Presentacion_Mineria_de_Datos.

¹⁴ Virsedá, F., & Román, J. (2010). Minería de datos y aplicaciones. *Revista Latinoamericana de Ingeniería de Software*, 8. Recuperado de <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>

Predicción si lloverá mañana en Australia a través de algunos métodos de clasificación

Luis Diego Quirós Gómez¹⁵, Noelia Rojas Ramírez¹⁵, Shirley Herrera Delgado¹⁵
luis25quiros@gmail.com, mnrr.10@gmail.com, auxirleyucr31@gmail.com

RESUMEN

En los últimos años las personas a través de sus actividades para desarrollarse han generado un cambio en el clima, uno de los países que se ha visto afectado es Australia, sufriendo de sequías y oleadas de calor. La lluvia es un recurso natural escaso en esta zona del mundo, por lo que es de mucha utilidad contar con técnicas que predigan si lloverá. En este estudio se utilizaron los siguientes tres métodos de clasificación, regresión logística, árboles de decisión y bosques aleatorios. Se tuvo como objetivo establecer la mejor técnica para predecir si va a llover o no al siguiente día en Australia, esto se comprobó a través de distintos indicadores de desempeño, error de clasificación, curva de característica operativa del receptor (ROC), área bajo la curva (AUC) y Kolmogorov- Smirnov (KS). Se obtuvo que las tres técnicas se desempeñan bien para clasificar si lloverá o no mañana en Australia, sin embargo, el bosque aleatorio obtuvo mejores puntajes. Por lo tanto, este es el método que se recomienda usar para el contexto de predicción de lluvias en Australia.

PALABRAS CLAVE: Australia, lluvia, predicción, modelos de clasificación, indicadores desempeño, regresión logística, árbol de decisión y bosque aleatorio.

ABSTRACT

In recent years, human activity has generated a climate change, one of the countries that has been affected is Australia, which suffers droughts and heat waves. Rain is a scarce natural resource in this area, and it is very useful to have techniques that helps to predict if it will rain. In this study the following methods of classification are used, logistic regression, decision trees and random forests are used. The objective is to determine which is the better technique to predict whether or not it will rain the next day in Australia. To measure the performance of the models some indicators are used, classification error, receiver operating characteristic curve (ROC), area under the curve (AUC) and Kolmogorov-Smirnov (KS). It was obtained that the three techniques performed well, however, the random forest obtained better scores. Therefore, that is the method that is recommended to be used for the prediction within the context of rainfall in Australia.

KEYWORDS: Australia, rainfall, prediction, classification models, performance indicators, logistic regression, decision tree and random forest.

INTRODUCCIÓN

Debido al cambio climático generado por las actividades humanas en los últimos años, uno de los países que se ha visto afectado es Australia, colocando en riesgo su riqueza natural, incluyendo sus especies endémicas. Según la UNESCO (1988), este país se caracteriza por tener un territorio vasto y una población de sólo 16 millones de habitantes, dado que se encuentra situado entre 15° y 35° de latitud, es decir en la zona donde se hallan la mayoría de los desiertos del mundo, donde aproximadamente un 80% de su territorio se caracteriza por ser de clima árido o semiárido, contando con precipitaciones (lluvias) y recursos hídricos escasos.

¹⁵ Estudiantes de Estadística de la Universidad de Costa Rica



Dada la progresiva escasez de agua que se ha presentado, el gobierno australiano se ha dedicado a desarrollar desde el año 2006, proyectos para controlar y administrar mejor sus recursos hídricos, siendo los agricultores y ganaderos los más afectados. Esta iniciativa incluye tanto a científicos de centros estatales como privados, que trabajan en colaboración para combatir esta problemática (Fayanás, 2011).

Debido a lo anterior, se ha impulsado la utilización de experimentos de modelización en los que se ha determinado que la probabilidad de que ocurriese un verano caluroso sin precedentes en Australia es de cinco veces mayor actualmente. El año 2013 fue el más cálido para Australia desde que se iniciaran los registros en 1910, superando la temperatura en 1,20 °C a la media. Además, para ese año, las temperaturas del aire en superficie sobre las zonas terrestres en el hemisferio sur fueron muy cálidas, lo que produjo olas de calor generalizadas. Las condiciones más secas a la media a nivel global prevalecen en el este de Australia y entre otros países como oeste de Estados Unidos, gran parte de Canadá y el norte de Siberia (Organización Meteorológica Mundial, 2004).

Por estas razones, resulta imprescindible tomar en cuenta las limitaciones que impone el entorno, un medio para ello es la meteorología en la que una de sus principales aplicaciones es la predicción del tiempo futuro, siendo de enorme utilidad para el aprovechamiento de los recursos naturales. Cabe destacar que la meteorología ha sido estudiada desde la antigüedad, para explicar y predecir la conducta futura de los fenómenos naturales (Palomares,2008), esto debido a que una de las principales condiciones que limitan las actividades diarias es el tiempo atmosférico, que se entiende como la descripción de las condiciones atmosféricas en un momento y lugar concreto (Fundación Española de Ciencia y Tecnología,2004).

Con este fin Rodríguez (2004) indica que se han desarrollado métodos para realizar pronósticos eficientes, como lo son la clasificación a través de modelos, siendo de gran utilidad para el desarrollo de las actividades humanas y reduciendo la vulnerabilidad frente al cambio climático, sin embargo, los pronósticos dependen de la cantidad de información disponible y de la dificultad que se presenta en la recolección de los datos y la confianza deseada.

El objetivo de este estudio es establecer el mejor método de clasificación para predecir si va a llover o no al siguiente día en Australia, a partir de distintos indicadores de desempeño.

METODOLOGÍA

El conjunto de datos de lluvia en Australia se obtiene de la plataforma Kaggle (link adjunto en las referencias), este contiene 56.240 observaciones meteorológicas diarias recopiladas por la Oficina de meteorología de Australia durante el periodo comprendido entre el 31 de octubre del 2007 al 24 de junio del 2017. Para esta investigación, se utilizan tres métodos para predecir la clasificación de las observaciones, las técnicas usadas son la regresión logística, árbol de decisión y bosque aleatorio, a las cuales se calculan a cada una cuatro indicadores de desempeño obtenidos a través de validación cruzada para realizar una comparación y llegar a la escogencia de la mejor técnica de clasificación para estos datos.

La variable respuesta tiene dos categorías y explica si al día siguiente llueve o no, además se tienen otras 21 variables de diversos niveles de medición y escalas. Las variables fecha y lugar de la observación son excluidas del análisis debido a que se busca una predicción general que sea de utilidad en cualquier localidad y día del año. Por lo que al final se realiza el estudio con 19 variables que son consideradas como predictoras (ver anexos, cuadro 1).

La regresión logística estima la probabilidad de que un evento ocurra en función de las variables explicativas, es importante mencionar que esta técnica se basa en el método de la máxima función de verosimilitud, la cual es un método que estima los valores del parámetro de acuerdo con una muestra finita

de datos (Molinero, L. 2002). La regresión se ajusta incluyendo todas las variables de la base. Después de que se verifican los supuestos necesarios se realiza la selección de predictores hacia atrás utilizando la prueba de razón de verosimilitudes, para solamente incluir las variables que tienen un aporte significativo. El modelo es:

$$\log\left(\frac{\pi_{llueva}}{1 - \pi_{llueva}}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_{58} x_{58}$$

Donde:

π_{llueva} : probabilidad que el día de mañana llueva

β_0 : intercepto con el eje y

x_i : i-ésima variable con $1 \leq i \leq 58$.

β_i : i-ésimo coeficiente correspondiente a la i-ésima variable con $1 \leq i \leq 58$.

Para decidir cuál umbral se utiliza como corte se toma en cuenta el error y la precisión. Los intervalos de confianza y las pruebas realizadas en este estudio se utiliza una significancia de 0.05.

El árbol de decisión se caracteriza por ser un método sencillo de utilizar, puede ser usado en distintas áreas como por ejemplo el diagnóstico médico, control de calidad y predicciones meteorológicas, siendo el modelo de aprendizaje inductivo supervisado no paramétrico más utilizado (Solarte.G,2009). Además, según Solarte (2009), a partir de una base de datos esta técnica se encarga de hacer diagramas de construcción lógica, semejante a los sistemas de predicción basados en reglas, ya que, los datos son registros en ocurren de manera iterativa para la predicción de un fenómeno natural.

El método utilizado para generar el árbol de decisión es *Classification and Regression Trees* (CART). El algoritmo realiza particiones de los datos mediante reglas y se detiene hasta que ya no sea posible obtener una mejor separación que es cuando se llega a un nodo terminal. Asimismo, algunas reglas que permiten detener el algoritmo son el mínimo número de observaciones que debe de haber para poder dividir un nodo que se establece en un 5% (2.821) de los datos, el mínimo número de observaciones que debe de tener un nodo terminal que se considera que 2% (1.128) es suficiente y además se establece que la máxima profundidad es 10. Es importante recalcar que estos valores son establecidos para evitar una gran extensión del árbol que pueda llevar a un sobreajuste del conjunto de datos de entrenamiento. Además, se requiere establecer un parámetro de complejidad para el árbol.

Por otro lado, el método de clasificación de bosques aleatorios crea una multitud de árboles de decisión aleatoriamente, es decir es un conjunto de árboles de decisión que se ha creado a partir del mismo conjunto de datos, para producir un mejor modelo que un árbol individual, el resultado final es tomado a partir de la cantidad mayor de predicciones dadas por cada uno de los árboles que conforman el bosque.

Los parámetros que se establecen para la generación del bosque son el mínimo tamaño que debe de tener un nodo terminal que se establece en 12, aunque este número parezca pequeño si se toma en cuenta que esto es para cada árbol al final la decisión se tomaría con $12 * a$, donde a es el número de árboles. Además, no se establece límite para el número máximo de nodos que un árbol puede tener. El número de árboles y la cantidad de variables aleatorias para seleccionar se ajusta tomando en cuenta el error y la precisión.

Para lograr determinar el rendimiento del modelo se van a utilizar cuatro indicadores de desempeño, para todos hace el cálculo utilizando el modelo completo y validación cruzada con 10 pliegues. A continuación, se presentan los diferentes indicadores utilizados en este estudio:

- Error de clasificación: Es la proporción de observaciones que son mal clasificadas por el modelo.

- Curva de característica operativa del receptor (ROC): Este indicador muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación. Para lograr esto se compara los falsos positivos contra la precisión positiva al cambiar la probabilidad que se usa como corte para clasificar.
- Área bajo la curva (AUC): Es el área bajo la curva ROC y esta indica la probabilidad de que el modelo clasifique una observación positiva aleatorio más alto que una negativa aleatoria. Puede tomar valores de 0 a 1, en caso de que las predicciones sean todas incorrectas el indicador toma el valor de 0 y en caso contrario un valor de 1.
- Kolmogorov- Smirnov (KS): Este método es una prueba no paramétrica de bondad de ajuste que comprueba si dos muestras tienen la misma distribución, es decir, compara la distribución de que sí va a llover y no va a llover. El indicador calcula la diferencia máxima absoluta entre ambas distribuciones acumulativas relativa. Asimismo, para este indicador si es menor a 20 es considerado como malo, si es mayor a 40 es muy bueno y en caso de que sea mayor a 70 es excelente, sin embargo, se puede sospechar que se está sobre ajustando.

Debido al contexto en que nos encontramos y a que la cantidad de observaciones clasificadas como que al siguiente día va a llover, es necesario no solamente enfocarse en medidas globales. Sino en medidas que indiquen específicamente el desempeño en ambas clases. Por lo tanto, además de los indicadores previamente mencionados también se desea encontrar un equilibrio entre la precisión a la hora de predecir ambas clases, ya que no solo queremos precisión para predecir los días que no llueven, sino que es aún más importante predecir los días que sí va a llover.

Para el análisis se utiliza el software R (R Core Team, 2009) junto con otros varios paquetes. Para la manipulación de los datos se utilizan dplyr (Wickham et al, 2019) y forcats (Wickham, 2019), para generar las figuras ggplot2 (Wickham, 2016), para el ajuste de los modelos rpart (Therneau, 2019) y randomForest (Liaw, 2002) y para el cálculo de los indicadores de desempeño se utilizan caret (Kuhn, 2019) y ROCR (Sing et al, 2005).

RESULTADOS

Como resultados de un análisis exploratorio previo se encuentra que en el conjunto de datos aproximadamente el 78% de la variable respuesta se clasifica como no, y solamente el 22% corresponden a sí. Las variables predictoras poseen distintas escalas y categorías (ver anexo, Cuadro 2).

Se ajusta la regresión logística con las 16 variables seleccionadas (ver anexo, Cuadro 3). A partir del modelo seleccionado se obtienen los coeficientes estandarizados para determinar cuáles variables tienen mayor aporte, se destacan las siguientes dos variables, la humedad a las 3 p.m. ya que cuando esta aumenta en 20% la propensión de que al día siguiente llueva será entre 3 y 3.3 veces la que se tenía antes del aumento. La segunda es la presión a las 3 p.m. ya que cuando esta aumenta en 7 hPa, la propensión de que al día siguiente llueva se disminuye entre 69% y 76% respecto a la que se tenía antes de realizar el aumento. En ambos casos se debe de mantener constante todas las demás variables.

Para realizar la predicción de la clasificación se necesita establecer un valor que funcione como punto de corte para que las observaciones con probabilidad menor a este valor se clasifican como que no va a llover mañana y las que son iguales o mayores se clasifican como que sí lloverá. Tomando como se utiliza tradicionalmente una probabilidad de 0,50 se obtiene que el error de clasificación es de 14,6%, este parece estar muy bien, aunque al tomar en cuenta la precisión del sí es muy baja ya que alcanza apenas un 53% mientras que la del no es de 94%. Con el fin de tratar de mejorar estos indicadores se busca un nuevo valor de la probabilidad para usar como punto de corte.

Para encontrar el umbral de clasificación se decide observar cómo se comporta el error y la precisión para ambas clases, ambos obtenidos a través de validación cruzada, al cambiar la probabilidad

que se utiliza como límite (ver anexo, Figura 1). Al buscar un balance entre estos indicadores se toma 0.21 como punto de corte. En este valor no se está obteniendo el menor error posible, pero se encuentra muy cerca del mínimo y lo más importante es que se obtiene gran ganancia en la precisión del sí.

Con este nuevo valor de la matriz de confusión se obtiene el error de clasificación de 14%. Al observar cada una de las clases se tiene que no llover mañana cuenta con una precisión del 94%, pero esta decrece hasta un 54% en el caso de la precisión para predecir si el día siguiente lloverá. Lo cual confirma las sospechas de que al haber menor cantidad de observaciones positivas el modelo es muy preciso para la categoría contraria. Estos resultados son muy similares a los obtenidos a través de validación cruzada (ver anexo, Cuadro 4).

A pesar de los problemas en los indicadores previos se observa que la curva ROC (ver anexo, Figura 2) tiene un comportamiento que se aleja de la línea no discriminante, indicando que el modelo no es sensible a los umbrales de clasificación, y esto se confirma al obtener un AUC de 0,88 que se encuentra muy cerca del mejor valor que es 1. Por otro lado al observar el KS (ver anexo, Figura 3) se observa una muy buena separación entre ambas distribuciones acumuladas y la mayor distancia es de 60,2 que se considera muy buena. Por tanto, se evidencia que el modelo no tiene problemas y desempeña la clasificación.

Para el método de árbol de decisión, antes de ajustar el árbol final se requiere encontrar el valor del parámetro de complejidad que genere los mejores indicadores de desempeño. De igual modo que con la regresión logística se cambia este valor y se grafican el error de clasificación y las precisiones obtenidas al utilizar validación cruzada (ver anexo, Figura 4). Al buscar un balance entre los tres indicadores se escoge un parámetro de 0.023, ya que, aunque se tenga más error de clasificación hay una ganancia en la precisión de la clase sí.

El árbol que se obtiene con el parámetro de complejidad escogido es de 4 niveles, (ver anexo, Figura 5). Se puede observar que la variable que mejor discrimina es la Humedad a las 3 p.m. ya que, esta se encuentra presente en los primeros 2 niveles. Además, en la primera división es capaz de clasificar gran cantidad de observaciones pertenecientes a la clase no. Se cuenta con 4 nodos finales, los cuales la mitad tienen mayoría de observaciones para el día de mañana que no llueve y los otros para sí.

Además, es relevante que de las 17 variables predictoras con las que se cuenta, esta técnica de clasificación solamente toma en cuenta dos de ellas. En los nodos finales se observa que a pesar de los pocos niveles del árbol se da una buena clasificación, ya que el nodo terminal más malo posee 60% la clase sí lloverá y 40% de la otra. Mientras que el mejor tiene 86% y 14% respectivamente. Esto indica que la probabilidad de que una observación se clasifique bien es alta, para comprobar esto se calculan los indicadores de desempeño.

El error de clasificación del modelo es de un 16%, pero de igual modo al observar por clases se tiene que la precisión de la clase no es de 95%, siendo mucho más alta que la de la clase sí con apenas un 36%, presentado el mismo problema que la regresión logística. En la curva ROC (ver anexo, Figura 6) se logra observar que el modelo no se ve afectado por los umbrales de clasificación y se obtiene una medida del AUC de 0,70 la cual es buena. De igual modo el KS (ver anexo, Figura 7) que se obtiene se encuentra por debajo del límite para ser considerado como bueno. Al contrastar estos indicadores contra los obtenidos a través de validación cruzada se observan que todos son similares menos la precisión de la clase sí ya que esta aumenta aproximadamente 4 puntos porcentuales, mostrando que el modelo realmente no es capaz de hacer una predicción de los días que van a llover (ver anexo, Cuadro 5).

Para el método de bosques aleatorios, los parámetros de este modelo se escogen de manera análoga a como se reporta anteriormente. Se toma 100 como número de árboles ya que al tomar más de esta cantidad no hay ninguna ganancia observable (ver anexo, Figura 8) y el tiempo de computación se

torna bastante extenso. Para el número de variables aleatorias se escoge 6 porque de igual modo no se observa ninguna ganancia al aumentar este número y a partir de ese punto se nota estabilidad en los indicadores (ver anexo, Figura 9).

Al ajustar el bosque con los parámetros establecidos se observa que los indicadores de desempeño del modelo (ver anexo, Cuadro 5) son casi perfectos, ya que el error es muy bajo y las precisiones para ambas categorías son muy buenas. De igual modo el AUC (ver anexo, Figura 10) está muy cerca del valor de 1 y el KS (ver anexo, Figura 11) que alcanza 94,5 reflejando la mayor separación entre las distribuciones acumuladas obtenidas hasta el momento.

Estos indicadores en lugar de generar confianza en el modelo producen lo contrario ya que, pueden indicar un sobreajuste del modelo, lo que lo haría menos potente a datos distintos a los que se encuentran en el conjunto de entrenamiento. Para comprobar si esta situación se está dando se observan los mismos indicadores obtenidos a través de validación cruzada (ver anexo, Cuadro 6). Donde se obtuvo que efectivamente sí hay un sobreajuste, lo que se refleja a través de la baja en los indicadores, pero aun así en estos se observa un buen desempeño, incluida la precisión de la clase sí lloverá al día siguiente.

Como el modelo es más complejo que los usados anteriormente, para ver cuáles variables son las que tienen más peso se debe calcular cuáles de ellas son las que más ayudan a la disminución de la impureza cuando están presentes (ver anexo, Figura 12). Se obtiene que la humedad a las 3 p.m. tiene la mayor importancia con casi el doble de la mayoría de las demás y en segundo lugar se encuentra la cantidad de brillo solar que su aporte también es mucho mayor.

Al utilizar el método de regresión logística se obtiene un modelo aceptable con indicadores de desempeño que se encuentran dentro del rango que se considera bueno. Este además posee una ventaja muy clara que es que se pueden interpretar sus coeficientes y se puede cuantificar como cada una de las variables afectan la probabilidad de que llueva al día siguiente.

El árbol también posee gran interpretabilidad y posee un menor error global que la logística, pero a la hora de comparar el desempeño por clase la precisión positiva se encuentra por debajo del modelo clásico. Además, el AUC y el KS también se encuentran por debajo de la logística. Dada esta situación al comparar los dos modelos interpretables se encuentra que la regresión es superior en la mayoría de los indicadores.

Al introducir el bosque aleatorio a las comparaciones este se destaca debido a su superioridad no solamente en medidas globales como el error, KS y AUC, sino que también su desempeño en ambas clases es mucho mejor que cualquiera de las otras dos. Esto debido a su naturaleza de ensamblaje de modelos, que logra potenciar un modelo que resultó ser el más débil, el árbol. A pesar de que se pierde interpretabilidad la ganancia que se obtiene es mucho mayor.

Cuadro 7. Indicadores de desempeño con validación cruzada para las tres técnicas de clasificación

Indicador de desempeño	Logística	Árbol de decisión	Bosque aleatorio
Error de clasificación	0.146	0.168	0.139
Precisión de la clase no	0.944	0.952	0.946
Precisión de la clase sí	0.538	0.405	0.562
AUC	0.886	0.703	0.896
KS	60.632	39.352	62.189

Fuente: Elaboración propia.

CONCLUSIONES

Al utilizar una medida global de desempeño como lo es el error de clasificación todos los modelos muestran un desempeño similar. Por eso es importante utilizar otros indicadores de desempeño que puedan señalar problemas que el modelo presente en el momento de predecir una de las categorías, especialmente cuando estas se encuentran desbalanceadas en cantidad, por esta razón se utilizó la precisión positiva ya que, en este trabajo la categoría lloverá mañana representa aproximadamente la quinta parte de las observaciones. Al usar este indicador se contribuye a discriminar entre los modelos y logra señalar deficiencias que al utilizar otros indicadores de desempeño no se muestran.

A pesar del distinto desempeño de los modelos cabe destacar que los tres coinciden en que la variable más importante es la Humedad a las 3 p.m., por lo que esta se puede considerar como el predictor más importante que se tiene para predecir si el día siguiente llueve o no. Tomando en cuenta que Australia es un país que se encuentra entre las latitudes donde se hallan la mayor parte de desiertos del mundo, la variable humedad es una variable que aporta gran información en el campo de la predicción en la meteorología.

Cuando se requiera una técnica para predecir la clasificación de si mañana va a llover, en que se cuenta con contextos como los de Australia, que es un país con condiciones climáticas extremas, dado que la mayor parte de su territorio es de clima árido o semiárido, se recomienda la utilización de bosque aleatorio. Ya que, al observar conjuntamente todos los indicadores de desempeño y contrastarlos entre las diferentes técnicas, el bosque aleatorio es el modelo que brinda la mejor clasificación y que se desempeña mejor.

Utilizar este tipo de técnicas para clasificar el tiempo atmosférico como lo es si lloverá al siguiente día, son de mucha utilidad para países o territorios en que el recurso del agua es tan limitado. Brindan información confiable, que puede ser utilizada para la toma de decisiones e implementación de acciones. De ahí la importancia y la aplicación estos métodos.

REFERENCIAS

- Fundación Española de Ciencia y Tecnología. (2004). Meteorología y climatología. Recuperado de <https://cab.inta-csic.es/uploads/culturacientifica/adjuntos/20130121115236.pdf>.
- Fayanás, E. (2011). Australia, el extremismo hídrico. Recuperado de <https://www.nuevatribuna.es/articulo/medio-ambiente/australia-el-extremismo-hidrico/20110421133253053631.html>
- Kaggle. Lluvia en Australia, predecir si lloverá mañana. Recuperado de <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>
- Kuhn, M. (2019). caret: Classification and Regression Training. R package version 6.0-84. <https://CRAN.R-project.org/package=caret>
- Liaw, A. y Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), pp. 18-22.
- Molinero, L. (2002). Métodos estadísticos de clasificación. Recuperado de <https://www.alceingenieria.net/bioestadistica/clasifica.pdf>
- Organización Meteorológica Mundial. (2014). Declaración de la OMM sobre el estado del clima mundial en 2013. Recuperado de https://library.wmo.int/pmb_ged/wmo_1130_es.pdf

- Palomares, M. (2008). La meteorología y la predicción del tiempo ¿Cuál es actualmente su nivel de fiabilidad? ¿Por qué se equivocan los meteorólogos? Recuperado de <http://www.divulgameteo.es/uploads/Fiabilidad-predicci%C3%B3n.pdf>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Dirección electrónica <https://www.R-project.org/>.
- Rodríguez, R. (2004). Meteorología y climatología. Recuperado de <https://cab.inta-csic.es/uploads/culturacientifica/adjuntos/20130121115236.pdf>
- Sing, T., Sander, O., Beerenwinkel, N. y Lengauer, T. (2005). ROCr: visualizing classifier performance in R. *Bioinformatics*, 21(20), pp. 78-81.
- Solarte Martínez, G., & Soto Mejía, J. (2011). Árboles de decisiones en el diagnóstico de enfermedades cardiovasculares. *Scientia Et Technica*, XVI (49), 104-109.
- Therneau, T. y Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. <https://CRAN.R-project.org/package=rpart>
- UNESCO. (1988). Australia, un continente por descubrir. Recuperado de https://unesdoc.unesco.org/ark:/48223/pf0000081732_spa
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, New York, Estados Unidos, Springer-Verlag.
- Wickham, H. (2019). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.4.0. <https://CRAN.R-project.org/package=forcats>
- Wickham, H., François, R., Henry, L. y Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.1. <https://CRAN.R-project.org/package=dplyr>

ANEXOS

Cuadro 1. Descripción de las variables disponibles en el archivo de datos

Nombre	Descripción	Nivel de medición	Unidades o categorías
Llueve mañana	Indica si al día siguiente llueve o no. Es la respuesta que se busca predecir.	Cualitativa ordinal	Sí No
Temperatura mínima	Temperatura más baja	Cuantitativa continua	Grados Celsius
Temperatura máxima	Temperatura más alta		
Temperatura a las 9 a.m.	Temperatura a las 9 a.m.		
Temperatura a las 3 p.m.	Temperatura a las 3 p.m.	Cuantitativa continua	Milímetros
Cantidad de lluvia	Cantidad de lluvia		
Cantidad de evaporación	Cantidad de agua evaporada de un tanque diseñado especialmente para este tipo de mediciones	Cuantitativa continua	Milímetros
Brillo solar	Cantidad de horas de luz de sol	Cuantitativa continua	Horas
Dirección de la ráfaga	Dirección de la ráfaga más fuerte de viento	Cualitativa ordinal	16 posibles direcciones*
Dirección del viento a las 9 a.m.	Dirección del viento a las 9 a.m.		
Dirección del viento a las 3 p.m.	Dirección del viento a las 3 p.m.		
Velocidad de la ráfaga	Velocidad de la ráfaga más fuerte de viento	Cuantitativa continua	Kilómetros por hora
Velocidad del viento a las 9 a.m.	Velocidad del viento a las 9 a.m.		
Velocidad del viento a las 3 p.m.	Velocidad del viento a las 3 p.m.		
Humedad a las 9 a.m.	Porcentaje de humedad a las 9 a.m.	Cuantitativa continua	Porcentaje
Humedad a las 3 p.m.	Porcentaje de humedad a las 3 p.m.		
Presión a las 9 a.m.	Presión atmosférica al nivel del mar a las 9 a.m.	Cuantitativa continua	Hectopascales (hPa)
Presión a las 3 p.m.	Presión atmosférica al nivel del mar a las 3 p.m.		
Nubes a las 9 a.m.	Cantidad de octavos del cielo oscurecidos por nubes a las 9 a.m.	Cuantitativa discreta	octavas
Nubes a las 3 p.m.	Cantidad de octavos del cielo oscurecidos por nubes a las 3 a.m.		

Fuente: Oficina de meteorología, Australia. (<http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml>)

Cuadro 4. Indicadores de desempeño para la regresión logística

Indicador de desempeño	Modelo	Validación cruzada
Error	0,199	0,146
Precisión de la clase no	0,801	0,944
Precisión de la clase sí	0,801	0,538
AUC	0,886	0,886
KS	60,298	60,632

Fuente: elaboración propia.

Cuadro 2. Variables cuantitativas disponibles en el archivo de datos

Nombre	Mínimo	Máximo	Promedio
Temperatura mínima	-6.70	31.40	13.46
Temperatura máxima	4.10	48.10	24.22
Temperatura a las 9 a.m.	-0.70	39.40	18.20
Temperatura a las 3 p.m.	3.70	46.10	22.71
Cantidad de lluvia	0.00	206.20	2.13
Cantidad de evaporación	0.00	81.20	5.50
Brillo solar	0.00	14.50	7.74
Velocidad de la ráfaga	9.00	124.00	40.88
Velocidad del viento a las 9 a.m.	2.00	67.00	15.67
Velocidad del viento a 3 p.m.	2.00	76.00	19.79
Humedad a las 9 a.m.	0.00	100	65.87
Humedad a las 3 p.m.	0.00	100	49.60
Presión a las 9 a.m.	980.50	1040.40	1017.20
Presión a las 3 p.m.	977.10	1038.90	1014.80
Nubes a las 9 a.m.	0.00	8.00	4.24
Nubes a las 3 p.m.	0.00	9.00	4.33

Fuente: Elaboración propia a partir de Oficina de meteorología, Australia. (<http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml>)

Figura 1. Error y precisión del modelo logístico según cambia el umbral de clasificación

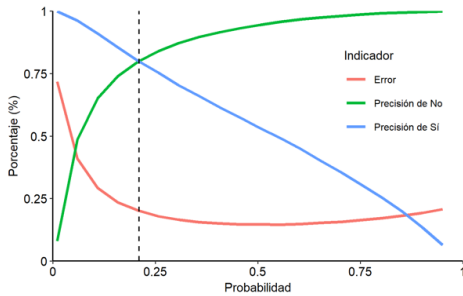


Figura 2. Curva ROC y AUC obtenida del modelo logístico

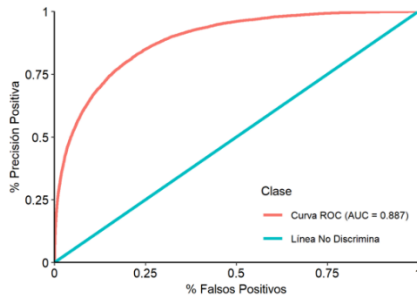
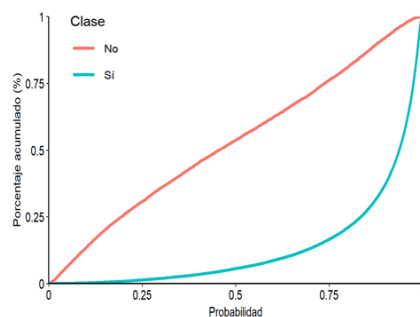


Figura 3. KS obtenido del modelo logístico



Cuadro 3. Coeficientes estimados de la regresión logística

Variable	Coefficiente estimado	Variable	Coefficiente estimado
Intercepto	51.199	Dirección del viento a las 9 a.m. SSW	-0.106
Temperatura mínima	-0.033	Dirección del viento a las 9 a.m. SW	0.181
Cantidad de lluvia	0.025	Dirección del viento a las 9 a.m. W	0.192
Brillo solar	-0.144	Dirección del viento a las 9 a.m. WNW	0.064
Dirección de la ráfaga ENE	-0.131	Dirección del viento a las 9 a.m. WSW	0.215
Dirección de la ráfaga ESE	0.057	Dirección del viento a las 3 p.m. ENE	-0.063
Dirección de la ráfaga N	-0.254	Dirección del viento a las 3 p.m. ESE	0.017
Dirección de la ráfaga NE	-0.1	Dirección del viento a las 3 p.m. N	0.099
Dirección de la ráfaga NNE	-0.118	Dirección del viento a las 3 p.m. NE	-0.06
Dirección de la ráfaga NNW	-0.012	Dirección del viento a las 3 p.m. NNE	0.096
Dirección de la ráfaga NW	0.03	Dirección del viento a las 3 p.m. NNW	0.217
Dirección de la ráfaga S	0.148	Dirección del viento a las 3 p.m. NW	0.227
Dirección de la ráfaga SE	0.173	Dirección del viento a las 3 p.m. S	-0.157
Dirección de la ráfaga SSE	0.225	Dirección del viento a las 3 p.m. SE	0.123
Dirección de la ráfaga SSW	0.172	Dirección del viento a las 3 p.m. SSE	-0.096
Dirección de la ráfaga SW	0.23	Dirección del viento a las 3 p.m. SSW	-0.139
Dirección de la ráfaga W	0.091	Dirección del viento a las 3 p.m. SW	-0.265
Dirección de la ráfaga WNW	0.123	Dirección del viento a las 3 p.m. W	0.048
Dirección de la ráfaga WSW	0.203	Dirección del viento a las 3 p.m. WNW	0.234
Velocidad de la ráfaga	0.059	Dirección del viento a las 3 p.m. WSW	-0.115
Dirección del viento a las 9 a.m. ENE	0.288	Velocidad del viento a las 9 a.m.	-0.011
Dirección del viento a las 9 a.m. ESE	-0.165	Velocidad del viento a 3 p.m.	-0.027
Dirección del viento a las 9 a.m. N	0.207	Humedad a las 9 a.m.	0.005
Dirección del viento a las 9 a.m. NE	0.373	Humedad a las 3 p.m.	0.057
Dirección del viento a las 9 a.m. NNE	0.525	Presión a las 9 a.m.	0.126
Dirección del viento a las 9 a.m. NNW	-0.043	Presión a las 3 p.m.	-0.183
Dirección del viento a las 9 a.m. NW	0.018	Nubes a las 9 a.m.	-0.016
Dirección del viento a las 9 a.m. S	-0.209	Nubes a las 3 p.m.	0.111
Dirección del viento a las 9 a.m. SE	-0.224	Temperatura a las 9 a.m.	0.045
Dirección del viento a las 9 a.m. SSE	-0.165		

Fuente: elaboración propia.

Figura 4. Cambio de indicadores de desempeño según cambia el parámetro de complejidad en el árbol de decisión

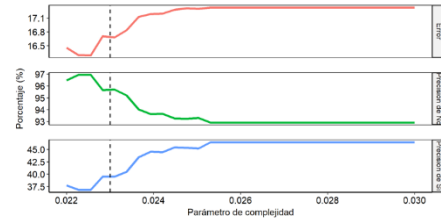


Figura 5. Árbol de decisión

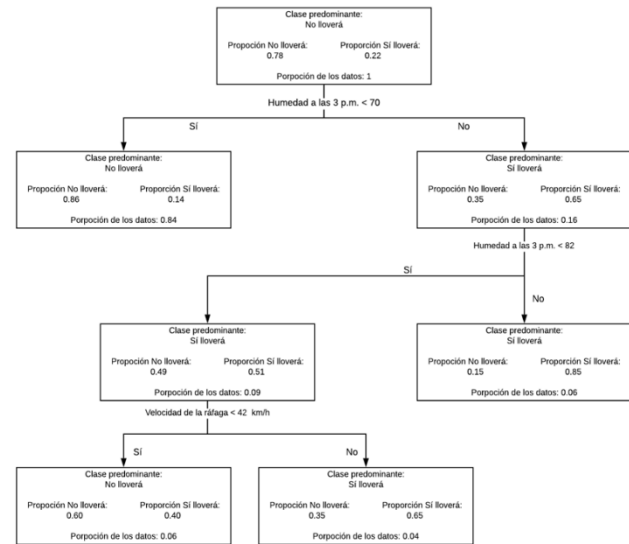
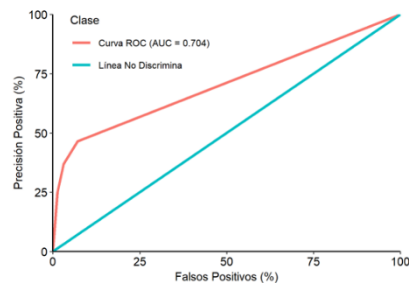
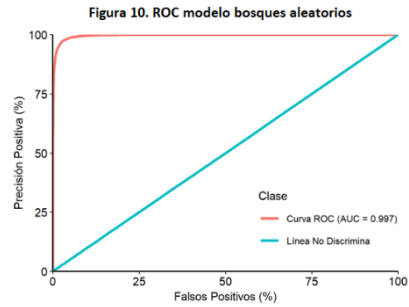
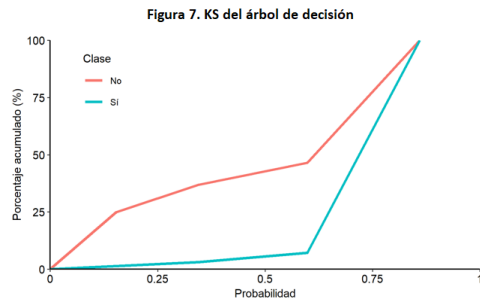


Figura 6. Curva ROC y AUC del árbol de decisión





Cuadro 5. Indicadores de desempeño para el árbol de decisión

Indicador de desempeño	Modelo	Validación cruzada
Error	0.163	0.168
Precisión de la clase no	0.969	0.952
Precisión de la clase sí	0.368	0.405
AUC	0.704	0.703
KS	39.352	39.352

Fuente: elaboración propia.

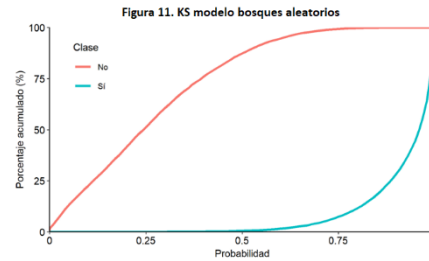


Figura 8. Número de árboles modelo bosques aleatorios

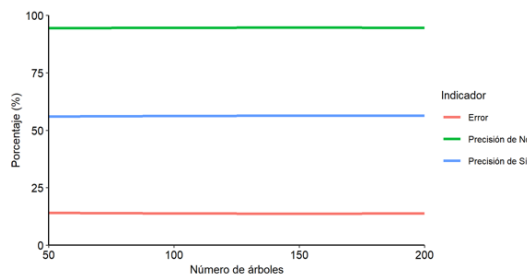
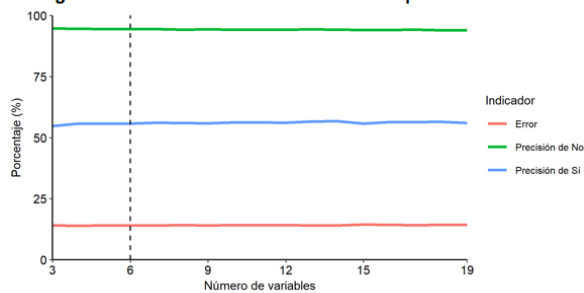


Figura 9. Número de variables modelo bosques aleatorios

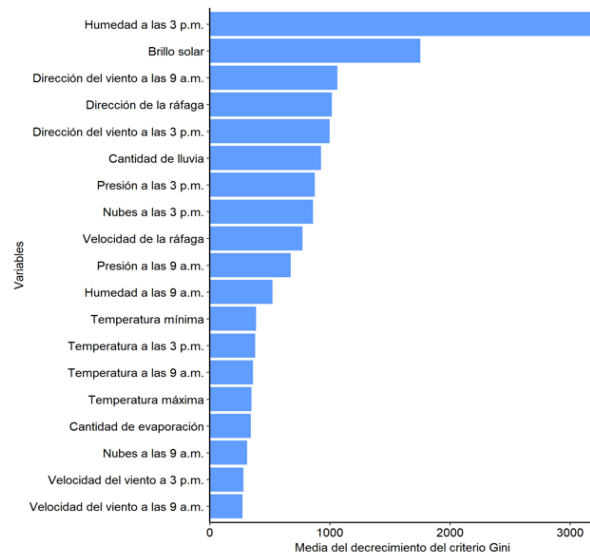


Cuadro 6. Indicadores de desempeño para Bosques Aleatorios

Indicador de desempeño	Modelo	Validación cruzada
Error	0.032	0.139
Precisión de la clase no	0.996	0.946
Precisión de la clase sí	0.867	0.562
AUC	0.997	0.896
KS	94.532	62.189

Fuente: Elaboración propia.

Figura 12. Importancia de variables para modelo bosques aleatorios



Comparación de métodos de clasificación para la calidad del vino mediante árboles de decisión, bosques aleatorios y K-vecinos más cercanos

Ericka Valerio Salas¹⁶, Juan José Montero Solórzalo¹⁶, Silvia Solera Vázquez¹⁶
erickavs189@gmail.com, jmontero639@gmail.com, solera271@gmail.com

RESUMEN

El siguiente estudio compara y determina el método de clasificación más adecuado para el vino según su calidad, esto mediante los métodos de árboles de decisión, bosques aleatorios y K-vecinos más cercanos. La evaluación de los métodos se realiza mediante los indicadores de desempeño: 1) el error de la matriz de confusión, 2) la curva de característica operativa del receptor (ROC) y 3) la máxima diferencia entre las distribuciones acumuladas relativas (KS); la validación de estos métodos se hizo mediante validación cruzada. Los datos utilizados fueron escogidos de la plataforma Kaggle con una base que cuenta con suficientes datos para hacer el análisis confiable. Se obtuvo que el método más efectivo es el de bosques aleatorios, con mejores indicadores de desempeño y relativamente consistentes, aunque este fue el que presentó mayores problemas de ajuste del modelo. El método de árboles de decisión obtuvo indicadores de desempeño menos favorables, pero fue el modelo que presentó un mejor ajuste y resultados más consistentes comparando el modelo con validación cruzada, aunque sólo utiliza dos variables predictoras de todas las disponibles para la base.

PALABRAS CLAVE: Árboles de decisión, bosques aleatorios, clasificación, k vecinos más cercanos, validación cruzada, calidad del vino

ABSTRACT

The following study compares and establish the most adequate classification method for wine quality, by using methods like decision trees, random forests and K-nearest neighbors. The method evaluation was made by performance measurement analysis: 1) the confusion matrix error, 2) the receiver operating characteristic curve (ROC) and 3) the maximum difference between the accumulative relative distributions (KS); the method validation was made through the cross-validation technique. The used data was chosen from Kaggle looking for data sets with enough records to make analysis reliable. The most effective method was found to be random forests with better and relatively consistent performance indicators, also, this technique was the one with the worst model adjustment problems though. The decision tree method obtained less favorable performance indicators, but it was the model that presented a better fit and more consistent results comparing the model with cross validation, although we only used two predictive variables of all those available for the base.

INTRODUCCIÓN

El vino se ha convertido en una parte importante desde el punto de vista cultural, alimentario, social y económico, principalmente en los países occidentales, aunque su consumo se presenta en la mayoría de los países a nivel mundial. Como cualquier otro alimento, este presenta una serie de características que son determinadas por diferentes elementos, como la variedad de uva, el tamaño y grado de maduración de esta, las características edáficas del suelo en que se ha cultivado, el clima, el tiempo de maceración, el método de elaboración y los posibles aditivos añadidos, la crianza y el envejecimiento, principalmente. (Contreras, 2017)

¹⁶ Estudiantes de Estadística de la Universidad de Costa Rica

Se ha demostrado que el vino presenta una serie de beneficios si se presenta un consumo moderado, como lo son: la prevención de enfermedades cardiovasculares y otras asociadas al envejecimiento, (alzhéimer), reducción en el desarrollo de diversos tipos de cáncer, de cálculos renales o mejoras en el metabolismo. Estos beneficios se deben principalmente a que cuentan con la presencia de polifenoles, entre ellos el resveratrol, en el vino (Zaldívar, 2012).

Debido a que la magnitud de estos beneficios mantiene una relación positiva con la calidad del vino consumido, se desea conocer los factores que influyen en la calidad y con esto, poder clasificar la calidad del vino que se produce, ya que esto ha tomado una mayor importancia por la alta asociación que se tiene con dietas saludables, como la mediterránea por ejemplo (Martínez-Carrasco, Brugarolas, del Campo y Martínez-Poveda, 2004).

Los métodos de clasificación nos permiten agrupar los datos que presenten características más homogéneas entre cada grupo y a su vez estas presentan heterogeneidad entre grupos, esto con el fin de poder realizar un mejor análisis de los datos. La clasificación tiene como objetivo principal encontrar un modelo que permita, de la manera más precisa posible, asignar la clase a la que pertenezca cada una de las observaciones.

Se cuenta con diferentes modelos que permiten realizar este tipo de análisis. Para artículo se trabajó con tres de estos: 1. árboles de decisiones, 2. bosques aleatorios y 3. K vecinos más cercanos, KNN por sus siglas en inglés. El estudio busca determinar cuál de los modelos, determinados a partir de los métodos anteriores, cuenta con la forma más eficiente para asignar las observaciones a la clase que pertenezcan. Se espera que el modelo que presente una mayor eficiencia sea el de bosques aleatorios, ya que es el modelo más robusto de los tres, es decir, que es el que se ve menos afectado cuando se producen anomalías en las muestras, si estas cuentan con valores atípicos.

METODOLOGÍA

Los datos se obtuvieron mediante la página Kaggle, de donde fueron descargados. La base seleccionada fue la base sobre variables que podrían definir la calidad del viñedo portugués “Vinho Verde”, para vinos blancos y tintos. Inicialmente se procedió a revisar la base y a realizar modificaciones, primero se encontró que la base contenía no respuesta en algunas variables, esto sucedía únicamente en treinta y ocho observaciones de 6497 por lo que se decidió eliminarlas. Seguidamente se procedió a codificar la variable respuesta, dicha variable era un puntaje de calidad del vino del uno al diez por lo que se codificó de la siguiente manera: buena si la calificación era mayor a cinco y mala si esta era menor o igual a cinco. De igual manera la base cuenta con las siguientes variables: tipo de vino (tinto o blanco), acidez fija, acidez volátil, ácido cítrico, azúcar residual, cloruros, dióxido de azufre libre, dióxido de azufre total, densidad, pH, sulfatos y nivel de alcohol (Parmar, 2018).

En la descripción disponible se explica que por cuestiones de privacidad existen variables relevantes como el tipo de uva que no son brindadas en la base, sólo se brindaron variables fisicoquímicas y sensoriales que pueden derivarse de estas variables no compartidas y que se desconoce la relevancia de todas las variables que sí están presentes en la base para explicar la calidad del vino. Además, se hace referencia al trabajo de Paulo Cortéz y otros, quienes son los autores que realizaron un estudio a esta base de datos, pero no fue posible tener acceso al artículo ya que este se encontraba solo en plataformas de pago.

Una vez corregida la base, se aplicaron tres métodos de clasificación, los cuales fueron: árbol de decisión, bosques aleatorios y k vecino más cercano. Para cada método se generó un modelo y posteriormente, se calcularon tres indicadores de desempeño, el error de matriz de confusión, la curva ROC y la máxima diferencia entre las distribuciones acumuladas relativas Kolmogorov-Smirnov (KS). Para cada método se calcularon los indicadores de precisión para el modelo completo y con el método de validación

cruzada. Finalmente, se realizó una comparación de los indicadores obtenidos con cada método para así determinar cuál de los tres métodos utilizados es el mejor.

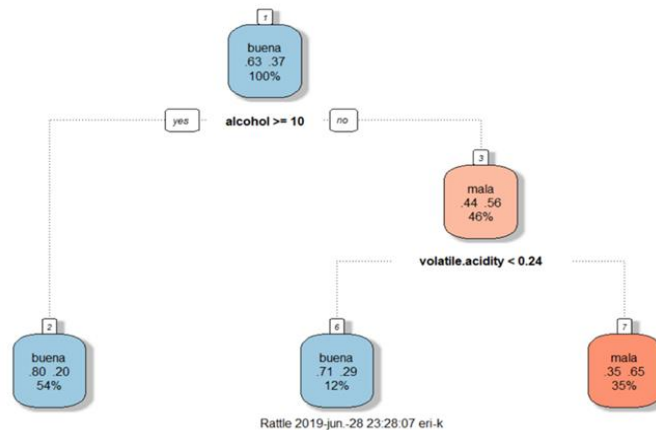
El primer método, el de árboles de decisión, realiza divisiones binarias para todas las variables, sean numéricas o categóricas, y estas se toman como una regla para realizar la clasificación. Para este método se utilizaron los parámetros del paquete de R rpart utilizados de manera predeterminada, minsplit=20, minbucket= minsplit/3, maxdepth= 30 y cp=0.01. El método de bosques aleatorios utiliza muestras bootstrap para combinar distintos árboles de decisión donde estos son generados con variables aleatorias. Por último, el método de k vecino más cercano clasifica datos nuevos según la cercanía con vecinos de un grupo.

El paquete estadístico utilizado para realizar el análisis fue R versión 3.5.2 (R Core Team, 2018). Las librerías usadas fueron: dplyr (Wickham, François, Henry y Müller, 2019), caret (Kuhn, 2018), rpart (Therneau y Atkinson, 2019), randomForest (Liaw y Wjener, 2002), ROCR (Sing et al., 2005), plotly (Sievert, 2018), kknn (Schliep y Hechenbichler, 2011) y rattle (Williams, 2011).

RESULTADOS

Para el método de árboles de decisión, se puede apreciar en la Figura 1 que para el caso, al tener tan pocos nodos se logra un modelo menos robusto que se está basando solamente en las dos variables con más correlación con la variable continua quality, alcohol y volatil.acidity. Con lo cual se estaría desaprovechando el resto de información disponible para clasificar el vino según la calidad.

Figura 1. Representación gráfica del árbol de decisiones del modelo

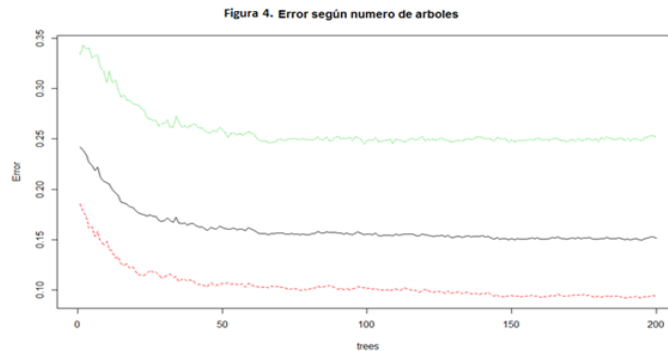


Al comparar los indicadores de desempeño del modelo completo con el método de validación cruzada se obtiene, para la matriz de confusión, una precisión de 0.0735 para el modelo completo y 0.0728 con validación cruzada; el error resultó 0.265 y 0.271 respectivamente con lo cual se observa que no existe una gran diferencia entre el modelo completo y la validación cruzada, esta última toma en cuenta los problemas de ajuste del modelo.

Algo similar se obtiene con la curva de característica operativa del receptor (ROC) donde, para ambos casos, el área bajo la curva (AUC) es de 0.721, ver Anexo 1. Esto también resulta consistente al analizar el ajuste del modelo completo. Este ajuste, al ser mayor que 0.5, es suficientemente alto para asumir que el modelo sí sirve para discriminar, aunque anteriormente se mencionó que sólo toma en cuenta el alcohol y el ácido volátil presente en el vino y no toda la información con la que se cuenta.

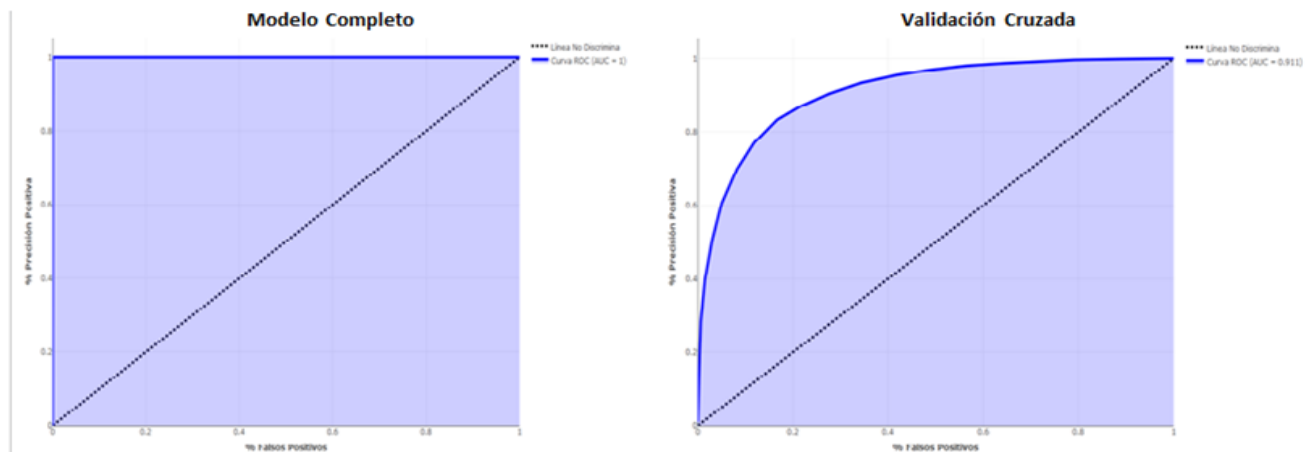
Con el indicador de desempeño relacionado a la máxima diferencia entre las distribuciones acumuladas de las clases, analizadas con el método Kolmogorov-Smirnov, se obtiene un KS de 41.8% con el modelo completo y 39.7% con validación cruzada, es decir presentan una pequeña diferencia de 2.1% lo cual es consistente con los resultados obtenidos para la matriz de confusión y el ROC obtenidos. Esto se evidencia con la representación gráfica, pues los gráficos para el modelo completo y con validación cruzada resultan muy similares, ver Anexo 2.

Para el método de bosques aleatorios, se puede observar en el Gráfico 4, que los errores se mantienen constantes luego de generar 50 árboles, por lo que empleó esta cantidad, ya que un mayor número no disminuiría el error por lo que no generaba una mayor ganancia.



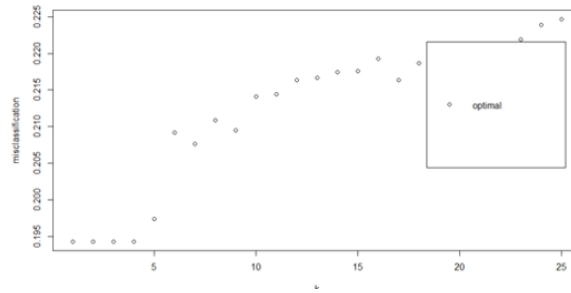
Para el modelo completo, se obtuvo un error de matriz de confusión de 1, mientras que con validación cruzada se obtuvo uno de 0.16. Al realizar la curva ROC con el modelo completo, se obtuvo una curva perfecta, un AUC de 1, en contraste, con validación cruzada, se obtuvo una AUC de 0.91; en la Figura 5 se puede observar las curvas mencionadas anteriormente para ambos modelos. Del mismo modo, al obtener el indicador KS, se consiguió un valor de 100, es decir, el máximo y para validación cruzada el KS es de 67.16, esto se evidencia en el Anexo 3, en el análisis gráfico si se observaron diferencias evidentes entre ambos modelos.

Figura 5. Curva de característica operativa (ROC) para modelo completo y validación cruzada



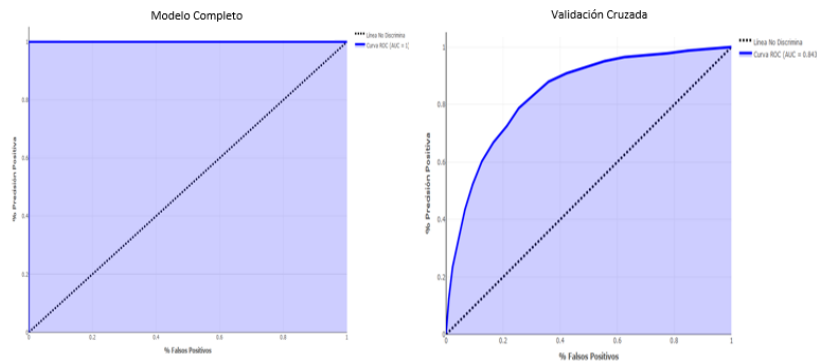
El método de K vecinos más cercanos se basa en la medición de la distancia que existe entre un elemento nuevo y otros elementos ya existentes, ordenando las distancias de menor a mayor, para poder determinar a cuál grupo pertenece el nuevo elemento. En Gráfico 7 se aprecia como el punto óptimo se encuentra en un k= a 19 aproximadamente. Por lo tanto, se utilizó este valor como parámetro.

Figura 7. Proporción de errores de clasificación por K-cantidad de vecinos



Con el modelo completo se obtuvo un error de matriz de confusión de 0, al realizar la curva ROC se obtuvo una curva perfecta y un AUC de 1, como se puede observar en la figura 8. Cuando se realiza para la validación cruzada, se obtiene un error de 0,2067, y un AUC de 0,843. Al usar la máxima diferencia Kolmogorov-Smirnov para el modelo completo se obtuvo un ks= 100, mientras que para validación cruzada se obtuvo un ks= 55, como se puede observar en el Anexo 4, donde se observan grandes diferencias entre los gráficos para ambos modelos.

Figura 8. Curva de característica operativa del receptor (ROC) para modelo completo y validación cruzada



Una vez obtenidos los indicadores de desempeño para los tres métodos aplicados, con el fin de concluir cuál de éstos es el método de clasificación más adecuado para determinar la calidad del vino, se procedió a comparar estos indicadores.

Cuadro 1. Resumen comparativo de indicadores de desempeño para modelo completo y validación cruzada por tipo de método de clasificación utilizado

Método utilizado	Matriz de confusión (e_T)		ROC		KS	
	Modelo completo	Validación Cruzada	Modelo completo	Validación Cruzada	Modelo completo	Validación Cruzada
Árboles de decisión	0.265	0.269	0.721	0.721	41.833	39.463
K-vecinos más cercanos	0.000	0.207	1.000	0.780	100	55.798
Bosques aleatorios	0.000	0.170	1.000	0.910	100	67.160

Fuente: Elaboración propia.

En el cuadro anterior se presenta que el error para el caso de árboles de decisión presenta una menor diferencia entre el modelo completo y utilizando validación cruzada; seguido de bosques aleatorios, pero para este último, el error es menor, aunque hay mayores diferencias entre modelos. El método K-NN presenta un error menor, que el de árboles de decisión, pero con resultados menos consistentes al comparar el modelo completo con validación cruzada. Para el ROC tenemos mayores valores con bosques aleatorios y con K-NN, pero al igual que en la matriz de confusión, se sugiere que hay problemas de ajuste al realizar validación cruzada. Los árboles de decisión presentan un ROC menor, pero con validación cruzada se corrobora que no existen problemas de sobreajuste. Respecto al KS del método de árboles de decisión, este es el método que presenta más consistencia al realizar validación cruzada, seguido de bosques aleatorios y por último K-NN.

CONCLUSIONES

Al comparar los indicadores de desempeño con los tres métodos, se puede notar que con el modelo completo los métodos de bosque aleatorio y K-vecinos más cercanos se encuentran sobre ajustados mientras que con el método de árboles de decisión esto no sucede. Al comparar los indicadores realizados con validación cruzada puede verse que en los tres indicadores el método de bosques aleatorios es el mejor ya que posee el error relativo de la matriz de confusión más pequeño, el AUC de la curva ROC y el KS más altos, en contraste, el método de árboles de decisión obtuvo los peores indicadores de los tres.

Al momento de comparar los indicadores realizados con el modelo completo y con validación cruzada se observa que validación es mejor debido a que elimina el sobreajuste que se tenía con el modelo completo. Basándonos en los indicadores realizados con validación cruzada y el cuadro 1 podemos concluir que el método de bosques aleatorios es el mejor de los tres métodos aplicados y que árboles de decisión el menos potente, como se esperaba al inicio del experimento.

Es importante mencionar que no se tiene conocimiento sobre la relevancia de todas las variables predictoras establecidas en la base, además no se realizó una calibración al modelo, ya que esto se escapaba de los objetivos del estudio. Por esto, no se hizo el análisis respectivo y se trabajó con toda la base y con parámetros definidos por defecto en el software utilizado, sin embargo, eso podría mejorar los resultados de los indicadores de precisión de los métodos, así como el ajuste del modelo. Por esto, para realizar un estudio más confiable se recomienda realizar el análisis de significancia de las variables incluidas en el modelo, así como la calibración de este.

BIBLIOGRAFÍA

Liaw, A. y Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), 8--22.

Carson Sievert (2018) plotly for R. <https://plotly-r.com>

Contreras, M. J. (2017). Estudio de la evolución de parámetros de color en vinos de Ribera del Duero, utilización de herramientas estadísticas y aplicaciones de interés para la industria vinícola. Universidad de Burgos. Obtenido de riubu.ubu.es/bitstream/10259/4508/1/Contreras_Fernández.pdf

Wickham, H., François, R., Henry, L. y Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.1. <https://CRAN.R-project.org/package=dplyr>

Schliep, K. y Hechenbichler, K. (2016). kkn: Weighted k-Nearest Neighbors. R package version 1.3.1. <https://CRAN.R-project.org/package=kkn>

Martínez-Carrasco, L., Brugarolas, M., del Campo, F. y Martínez-Poveda, A. (2004). El efecto de la preocupación por la salud, el interés por la gastronomía y por la actividad social sobre la intensidad de consumo de vino de calidad. *Economía Agraria y Recursos Naturales*. 04. 10.7201/earn.2004.07.02.

Kuhn, M. (2019). *caret: Classification and Regression Training*. R package version 6.0-84. <https://CRAN.R-project.org/package=caret>.

Parmar, R.A.J. (2018). *Wine Quality*. Recuperado 12 junio, 2019, de <https://www.kaggle.com/rajyellow46/wine-quality>

Zaldívar, E., Palacios, A., López, E., Crespo, I. y del Pueyo, I. (2012). Premios Eco Vino Awards 2019. Obtenido de http://www.premiosecovino.com/prensa/Ecovino_y_salud_SEVI21abr2012.pdf

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.

Sing, T., Sander, O., Beerenwinkel, N. y Lengauer, T. (2005). "ROCR: visualizing classifier performance in R." *_Bioinformatics_*, 21*(20), 7881. <URL: <http://rocr.bioinf.mpi-sb.mpg.de>>.

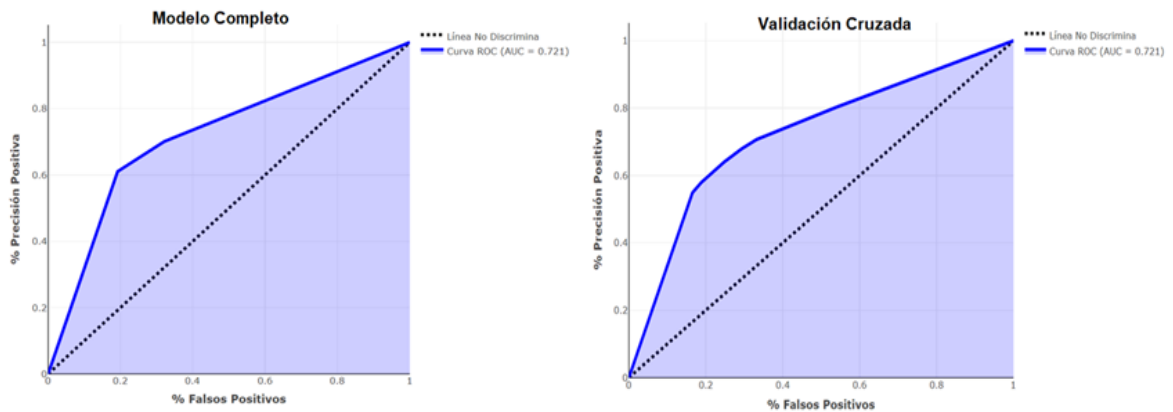
Therneau, T. y Atkinson, B. (2019). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-15. <https://CRAN.R-project.org/package=rpart>

Williams, G. J. (2011), *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Use R*, Springer.

ANEXOS

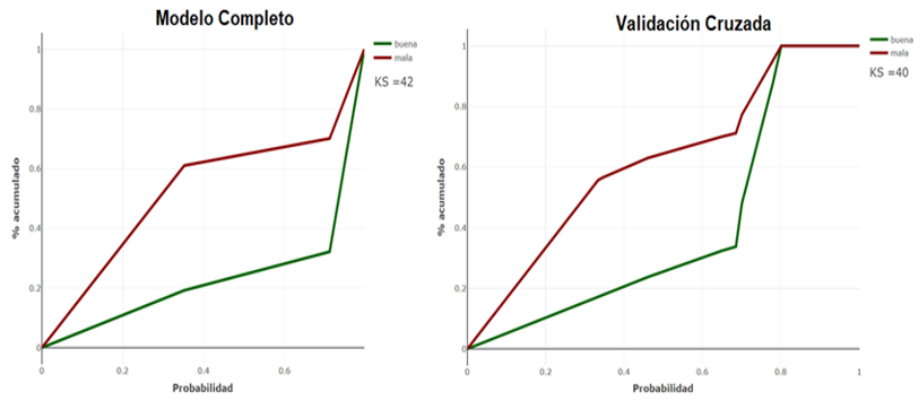
1

Figura 2. Curva de característica operativa del receptor (ROC) para modelo completo y validación cruzada



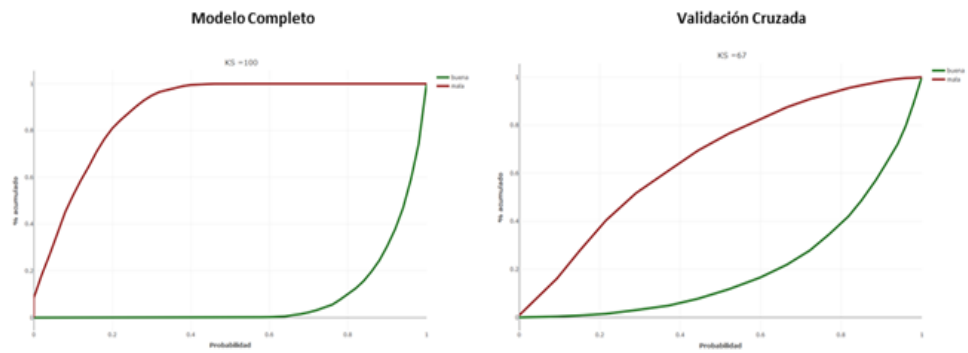
2

Figura 3. Máxima diferencia Kolmogorov-Smirnov para el modelo completo de validación cruzada



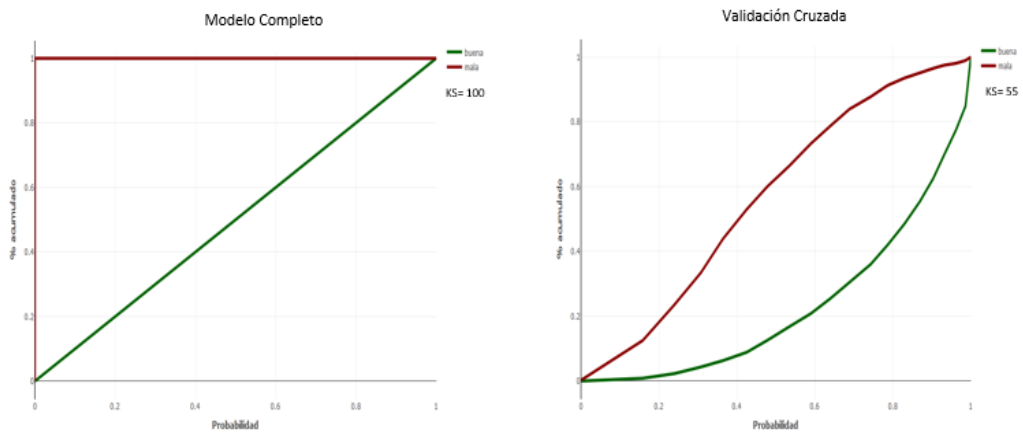
3

Figura 6. Máxima diferencia Kolmogorov-Smirnov para el modelo completo y validación cruzada



4

Figura 9. Máxima diferencia Kolmogorov-Smirnov para el modelo completo y validación cruzada



AGRADECIMIENTOS

Los integrantes del proyecto Revista Serengeti quieren agradecer de manera muy especial a todas las personas que de una u otra manera colaboración para llevar a cabo esta revista, tanto en la parte de asesoría como en el diseño.

Asesorías: Ricardo Alvarado B. (Profesor de Estadística), Samuel Muñoz T. (Profesor de Estadística), Adrián Vergara H. (Profesor e investigador de la Escuela de Filología, Lingüística y Literatura).

Diseño: José Aguilar U. (estudiante de Estadística y Comunicación Colectiva).

Sin el apoyo de todos ellos no hubiera sido posible terminar con éxito este proyecto tan importante.

