Ciencia Latina Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), marzo-abril 2024,
Volumen 8, Número 2.

https://doi.org/10.37811/cl_rcm.v8i2

EVALUACIÓN DE UN EXAMEN DE ADMISIÓN A PREGRADOS: METODOLÓGICA DE MEDIDAS REPETIDAS CON JASP

EVALUATION OF AN UNDERGRADUATE ADMISSION EXAM: METHODOLOGY OF REPEATED MEASURES WITH JASP

Héctor Francisco Ponce Renova
Universidad Nacional Autónoma de Ciudad Juárez, México

Ramses Jiménez Castañeda
Universidad Nacional Autónoma de Ciudad Juárez, México

Cely Celene Ronquillo Chávez
Universidad Nacional Autónoma de Ciudad Juárez, México

DOI: https://doi.org/10.37811/cl_rcm.v8i2.10763

Evaluación de un Examen de Admisión a Pregrados: Metodológica de Medidas Repetidas con JASP

Héctor Francisco Ponce Renova[1]
hector.ponce@uacj.mx
https://orcid.org/0000-0002-9302-3740
Universidad Nacional
Autónoma de Ciudad Juárez
México

Ramses Jiménez Castañeda
rjimenez@uacj.mx
https://orcid.org/0000-0003-0123-5576
Universidad Nacional
Autónoma de Ciudad Juárez
México

Cely Celene Ronquillo Chávez
cronquil@uacj.mx
https://orcid.org/0000-0002-7902-4544
Universidad Nacional
Autónoma de Ciudad Juárez
México

RESUMEN

La pregunta fue: ¿Existe una diferencia estadísticamente significativa entre los puntajes promedio de los aspirantes a través de las cinco ocasiones/generaciones? El objetivo fue dual: Analizar puntajes de aspirantes (n = 34,617) a 32 programas de pregrado de una universidad pública Mexicana con el Examen de Habilidades y Conocimientos Básicos con el uso del Análisis de Varianza de Medidas Repetidas; y enseñar como utilizar el Análisis de Varianza de Medidas Repetidas con el uso de JASP. El método consto de recolectar del examen de admisión de cinco generaciones de aspirantes (2016-2018), describir y analizar os datos con Análisis de Varianza de Medidas Repetidas. También, se detalló como se usa JASP para este análisis y como se interpreta la hoja de resultados. Los resultados mostraron que no existe diferencia de los puntajes entre las ocasiones/generaciones (p = .262) con una varianza explicada considerada pequeña (eta cuadrado parcial de .042). En conclusión, aparentemente no existe diferencia entre los puntajes de las cinco ocasiones/generaciones, y no aparecieron indicios de algún efecto. Asimismo, se encontraron correlaciones altas de .79 hasta .93 entre las ocasiones/generaciones. Estos resultados se pueden interpretar como una gran consistencia en aspirar a los mismos pregrados a través del tiempo.

Palabras clave: ANOVA de medidas repetidas, educación superior, EXHCOBA, pruebas de admisión, tamaño de efecto

[1] Autor principal
Correspondencia: hector.ponce@uacj.mx

Evaluation of an Undergraduate Admission Exam: Methodology of Repeated Measures with JASP

ABSTRACT

The question was: Is there a statistically significant difference between the average scores of the applicants across the five occasions/generations? The objective was dual: Analyze scores of applicants (n = 34,617) to 32 undergraduate programs from a Mexican public university with the Basic Skills and Knowledge Examination with the use of Repeated Measures Analysis of Variance; and teach how to use Repeated Measures Analysis of Variance with the use of JASP. The method consisted of collecting the admission exam of five generations of applicants (2016-2018), describing, and analyzing the data with Analysis of Variance of Repeated Measures. Also, it was detailed how JASP is used for this analysis and how the results sheet is interpreted. The results showed that there is no difference in scores between occasions/generations ($p = .262$) with an explained variance considered small (partial eta squared of .042). In conclusion, there appears to be no difference between the scores of the five occasions/generations, and no evidence of any effect appeared. Likewise, high correlations of .79 to .93 were found between occasions/generations. These results can be interpreted as great consistency in aspiring to the same undergraduate degrees over time.

***Keywords*:** admission test, ANOVA repeated measures, effect size, EXHCOBA, higher education

Artículo recibido 28 febrero 2024
Aceptado para publicación: 25 marzo 2024

INTRODUCCIÓN

Los objetivos fueron: *Analizar* puntajes de aspirantes a 32 programas de pregrado en un Examen de Habilidades y Conocimientos Básicos (EXHCOBA) con el uso del ANOVA de Medidas Repetidas (AMR); y *Enseñar* cómo hacer el AMR en JASP. Se analizaron puntajes de cinco generaciones de aspirantes (n = 34,617) a 32 programas de pregrado de una universidad pública en el norte de México. Cada generación de aspirantes produjo un set de puntajes durante cinco semestres consecutivos, llamados *ocasiones*: otoño del 2016 a la primavera del 2018. La pregunta de investigación fue: ¿Existe una diferencia estadísticamente significativa entre los puntajes promedio de los aspirantes a través de las generaciones/ocasiones? Se esperaba una correlación alta entre los puntajes promedio de los aspirantes de los pregrados.

Examen de Habilidades y Conocimientos Básicos (EXHCOBA)

La compañía Métrica Educativa A. C. (2018) explicó que, con el EXHCOBA, se evalúan habilidades, y conocimientos indispensables y básicos de aspirantes para cursar con éxito la licenciatura. Estas habilidades y conocimientos se desarrollan desde la primaria hasta la preparatoria. Esta compañía ha afirmado que existe una relación entre los puntajes del EXHCOBA y el graduarse de una universidad.

Revisión y Vacío en la Literatura.

Siguiendo con la literatura, aparentemente solo existen 10 publicaciones de revistas arbitradas acerca de los puntajes del EXHCOBA en los últimos 31 años, según la página de Métrica Educativa consultada el 20 de Junio del 2023.

- Backhoff y Tirado (1992) hablaron de la génesis y desarrollo del test.
- Backhoff et al. (1996) trataron el desarrollo de una interface para administrar el test: SICODEX.
- Tirado et al. (1997) presentaron la correlación entre los puntajes del test y las calificaciones del primer año de universidad.
- Backhoff et al. (1997) explicaron el proceso de validación para la administración del test.
- Tirado y Backhoff (1999) justificaron la opción de declarar No sé para evitar una posible penalización por contestar incorrectamente y para identificar las preguntas que los aspirantes no sabían.
- Backhoff et al. (2000) expusieron el nivel de dificultad y poder de discriminación del test.

- Antillón et al. (2006) enseñaron como igualar los puntajes de dos versiones diferentes del test.
- Antillón et al. (2008) expusieron cómo igualar los puntajes de tres diferentes versiones del test.
- Sánchez et al. (2008) resumieron el artículo de Backhoff y Tirado (1992).
- Backhoff et al. (2015) trataron sobre la Teoría de la Respuesta del Ítem aplicada a los puntajes de esta prueba.

Aunque dos artículos de 10 antes mencionados podrían ser considerados metodológicos (i.e., Antillón et al., 2006; Antillón et al., 2008), ninguno de éstos usó del AMR. Se realizó una búsqueda en Google Académico con las palabras: EXHCOBA y ANOVA de medidas repetidas. Aparecieron tres artículos: i.e. En uno, llevaron a cabo un AMR y aunque mencionaron al EXHCOBA vagamente, no usaron los puntajes de este examen; los otros dos artículos usaron el AMR con otros puntajes, pero solo mencionaron a este examen. En conclusión, se trata de llenar estos vacíos en el tiempo y en lo metodológico al usar el AMR para analizar los puntajes del EXHCOBA u otros datos similares.

Análisis de la Varianza de Medidas Repetidas (AMR)

El ANOVA y el AMR pertenecen al Modelo Lineal General (ver a Salkind, 2007; Tabachnick & Fidell, 2018). Un modelo lineal es una manera matemática y gráfica de representar la relación entre dos o más variables. Salkind definido al AMR (2007, pp. 544–545):

Una clase de procedimientos estadísticos interrelacionados enfocados en relaciones lineales entre variables o variables compuestas [Una variable compuesta es una combinación de distintas variables para hacer una sola]. El termino lineal se usa porque estas técnicas pueden ser representadas visualmente al poner una variable en contra de otra en una figura bidimensional y usando fórmulas matemáticas para determinar dónde dibujar una o más líneas que representan la relación visualmente entre las variables.

Otra manera de usar estos modelos es al comparar grupos: ej., ¿Existe una diferencia estadísticamente significativa entre los aprendizajes de los hombres y las mujeres? En resumen, estos análisis (ej., ANOVA y regresión) sirven para modelar datos. Si el modelo cabe perfectamente en los datos, éste explicará el 100% de la variación entre participantes, observaciones u objetos. Si el modelo no cabe en los datos, la variación explicada será cercana al 0%. Kotz (2006) señaló que la variancia no explicada es independiente de la variación en x (variable independiente): i.e., es igual a 1 menos la varianza explicada. Cuando se usa un ANOVA o un AMR que tenga uno o más factores, la variable independiente es la manera para formar grupos: ej., sexo (hombres vs. mujeres). Esta manera de agrupamientos es una variable nominal: no implica una jerarquía. La variable dependiente puede ser una escala Tipo Likert o preferentemente en forma continua (ej., puntajes en un test). Girden (1992, p. 1) definió al ANOVA como "…un procedimiento en general para aislar las fuentes de variabilidad de un set de mediciones." Esta última autora agregó que el objetivo de este análisis es determinar la magnitud de un efecto principal de una variable independiente en la dependiente.

Análisis de la Varianza de un factor abreviada como ANOVA es un procedimiento para comparar si las diferencias entre promedios de dos grupos o más es probablemente al azar (cf., Hinkle et al., 2003). El análisis de la varianza puede ser multifactorial: algunas variables independientes y una sola dependiente (ANOVA con varios Factores: ej., sexo, nivel educativo, grupo étnico, etc.). En ANOVA de un factor, la hipótesis nula plantea (H_0): no hay diferencia entre los promedios de cierto número de grupos. La contraparte es la hipótesis alternativa: por lo menos, uno de los promedios es diferente. La prueba F es usada para el ANOVA y el AMR. Esta prueba se basa en una distribución de una razón de varianzas: i.e., variación entre los grupos (considerada la varianza de interés) dividida por la variación dentro de los grupos (varianza del error). De esta división, se obtiene un valor Fcalculado (razón de varianzas) para ser comparado con un Fcrítico. Las reglas para rechazar la hipótesis nula son las siguientes:

$F_{calculado} > F_{critico}$ o probabilidad calculada (p) < alfa (rechazo la H_0);

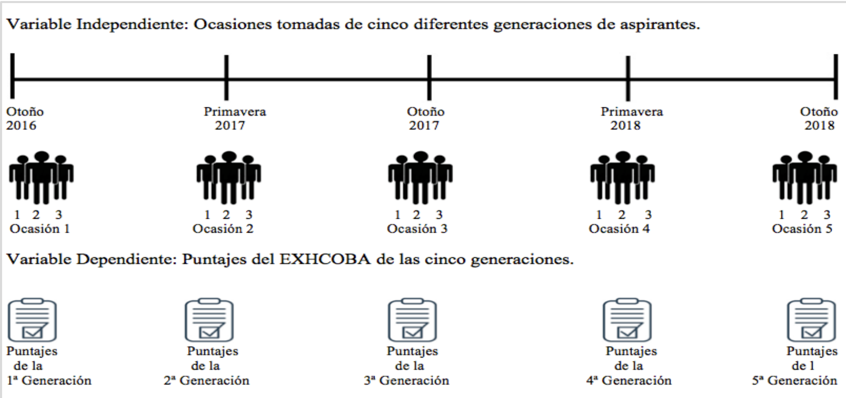
$F_{calculado} \leq F_{critico}$ o $p \geq$ alfa (no rechazo la H_0).

El AMR fue definido por Hinkle et al. (2003, p. 738) como:

Un ANOVA en la cual los participantes son medidos en dos o más ocasiones y el total de la varianza es partida en tres componentes: (1) variación entre los individuos; (2) variación entre las ocasiones o niveles; y (3) variación residual.

La variación entre los individuos de un grupo (ocasión) es que no todos obtuvieron el mismo puntaje en una toma de datos. La variación entre las ocasiones representa esas posibles diferencias entre cada una de ellas (ej., diferencias de puntajes entre un examen y otros con el mismo grupo de estudiantes). La variación residual no se debe ni a los individuos ni a las ocasiones. La variación residual no tiene explicación (cf. Girden, 1992). Hinkle et al. (2003) dijeron que, en AMR, el test de ocasiones es el efecto de interés primario: variable independiente. No hay término de error apropiado para poner a prueba el efecto de las diferencias entre los individuos y no se pone a prueba. El F calculado se obtiene de una razón: la varianza entre las ocasiones dividida por la varianza residual. Un AMR señala cuando hay una diferencia estadística entre dos o más ocasiones, porque es un test ómnibus, pero se necesita un test post hoc para señalar cuáles precisamente son los promedios de las ocasiones que difieren: el test post hoc de Bonferroni (ver Laerd Statistics, 2018). La Figura 1 muestra la variable dependiente e independiente del presente estudio con el AMR y otros detalles.

Figura 1. ANOVA de Medidas Repetidas de un Factor del Presente Estudio



Nota. La fuente fue elaboración propia.

Hinkle et al. (2003) definió la variación: qué tanto se dispersan los puntajes en alguna distribución de datos (ej., distribución normal). Estos puntajes se dispersan por una serie de factores que tienen un efecto en éstos.

Para complementar los análisis de significancia estadística de un ANOVA o AMR, se usan el coeficiente Eta Cuadrado (η^2) y el eta cuadrado parcial (η^2_p) que son análogos (similares) al coeficiente de determinación (r^2) para explicar la variación de una variable independiente en una dependiente. Otra forma de llamar a la varianza explicada es el efecto práctico o tamaño del efecto. La Asociación Americana de Psicología ha defendido el principio de incluir en toda investigación un efecto práctico para estimar la magnitud de un tratamiento o bien cuanta varianza explica una variable independiente en la dependiente. Las siguientes son las definiciones matemáticas de los coeficientes de

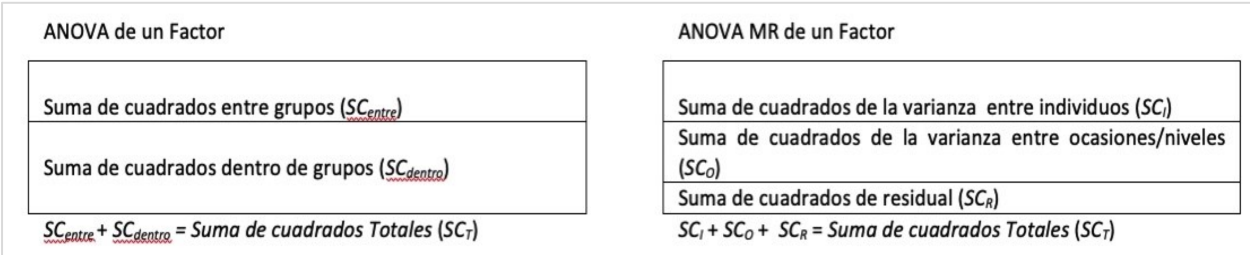
η^2 y η^2_p: $\eta^2 =$ (Suma de cuadrados entre grupos) / (Suma de cuadrados Totales)

η^2_p = (Suma de Cuadrados Varianza entre Ocasiones) / (Suma de Cuadrados Residual + Suma de Cuadrados Varianza entre Ocasiones)

Además, Cohen (1988) explicó que depende del campo de conocimiento para designar el tamaño del efecto en tres grandes categorías: pequeño, mediano y grande. Sin embargo, cuando no se tenga una base de investigaciones previas, Cohen (1988) sugirió para el η^2 los siguientes tamaños: pequeño = .0099; mediano = .0588; y grande = .1379. Richardson (2011) dijo que el η^2 mide la proporción de la varianza total en una variable dependiente (ej., puntajes en el EXHCOBA) explicada por una variable independiente: i.e., las ocasiones en AMR.

En el AMR, se detalla la varianza que explican las cinco ocasiones (variable independiente) en los puntajes promedio de los diferentes pregrados (variable dependiente). En el AMR, la varianza se divide en tres: variación entre los individuos (Suma de Cuadrados de la Varianza de Individuos), variación entre las diferentes ocasiones (Suma de Cuadrados de la Varianza entre Ocasiones / Niveles), y la varianza del error (Suma de Cuadrados Residual). La suma de estas tres variaciones da la Suma de Cuadrados Totales. La Figura 2 muestra todas estas sumas de varianzas.

Figura 2. División de la Varianza en estos análisis



Nota especifica 1. Cada rectángulo representa la varianza total y como se divide en otras varianzas.
Nota especifica 2. Para ambos análisis, la suma de estos cuadrados da como resultado el 100% de la varianza. La Suma de Cuadrados dentro de los Grupos (SC dentro) para el ANOVA de un Factor se parte en dos para el AMR de un Factor: i.e., Suma de Cuadrados de la Varianza de Individuos (SCI) y en Suma de Cuadrados Residual (SCR) para el AMR de un Factor.
Nota especifica 3. La fuente fue elaboración propia.

JASP

El software estadístico y psicométrico llamado JASP (creado en 2010) indica las iniciales de *Jeffreys' Amazing Statistics Program*, en reconocimiento de Sir Harold Jeffreys (1891-1989; Goss-Sampson, 2020). Fue un matemático, estadista, geofísico, y astrónomo británico cuyo libro, *Theory of Probability* (1939) jugó un importante rol en revivir el objetivo de la *Estadística Bayesiana*. Goss-Sampson (2020) declaró que JASP es un paquete estadístico de multiplataforma gratuita de código abierto (*open-source*), desarrollado y continuamente actualizado por un grupo de investigación de la Universidad de Ámsterdam. Otra ventaja es que produce hojas de resultados con el estilo de APA. Este software promueve el uso de la ciencia abierta al haberse integrado con *Center of Open Science Framework* (Centro del Marco de Ciencia Abierta). Es una organización tecnológica sin fines de lucro con sede en Charlottesville, Virginia, con la misión de aumentar la apertura, la integridad y la reproducibilidad de la investigación científica. La interfaz de JASP es similar al programa comercial de SPSS, pero utiliza el código de *R*. En JASP (s.f.) se encuentran las instrucciones básicas para bajarlo y usarlo.

METODOLOGÍA

La pregunta de investigación fue: ¿Existe una diferencia estadísticamente significativa entre los puntajes promedio de los aspirantes a través de las cinco ocasiones? Los puntajes fueron obtenidos en el examen de admisión por los aspirantes a una universidad. Los resultados esperados fueron una expectativa de una alta correlación entre las ocasiones. El método constó de la recolección y análisis de los puntajes del EXHCOBA por medio del AMR para comparar e interpretar los resultados. El método del AMR fue obtenido de los siguientes autores como sustento teórico y práctico: Girden (1992), Hinkle et al. (2003), Kots (2006), y Salkind (2007). Se utilizó el material por parte de la organización llamada Laerd Statistics (2018). También, se muestra cómo se llevó a cabo el AMR en JASP.

Durante cinco semestres consecutivos (2016–2018), los datos fueron obtenidos de la página de una universidad pública del norte de México. Los aspirantes (n = 34617) tomaron el EXHCOBA con un límite de tiempo de 180 minutos para responder 190 reactivos de opción múltiple con cuatro posibles respuestas y una quinta opción para declarar: *no sé*. Se oprime el botón (*no sé*) porque se penalizaba con un cuarto de punto una respuesta incorrecta. El EXHCOBA se tomó por computadora en las instalaciones de la universidad como uno de los criterios de admisión al pregrado de su elección, un semestre antes de ser admitidos(as). La Tabla 1 muestra el semestre de inicio el pregrado, así como las carreras, promedios de cada generación de aspirantes, desviación estándar (*SD*), Promedio por nivel/ocasión/semestre, Intervalo del Confianza del 95%, *n* por semestre de aspirantes y Promedio Global/General.

Tabla 1. Promedio y Desviación estándar por cada ocasión de Aspirantes Pregrado

Pregrado	**Otoño 2016 (*SD*)**	**Primavera 2017(*SD*)**	**Otoño 2017 (*SD*)**	**Primavera 2018(*SD*)**	**Otoño 2018(*SD*)**
Medicina	95.81(33.63)	105.7(29.85)	96.1(31.71)	99.81(30.36)	93.53(31.38)
Química	91.95(3.6)	91.95(25.87)	92.35(32.69)	77.13(30.37)	93.37(2.9)
Biología	91.95 (32.1)	90.98(22.96)	86.95(32.89)	77.23(26.68)	85.6(30.25)
Químico Fármaco Biólogo	87.21(33.38)	89.12(31.9)	87.66(32.36)	85.63(29.21)	90.63(19.27)
Ing. Eléctrica	76.29(31.63)	88.01(30.65)	81.41(33.04)	78.32(33.19)	79.54(32.22)
Sistemas Digitales	88.33(32.29)	87.99(32.82)	80.57(29.04)	91.66(36.07)	86.11(31.91)
Finanzas	86.92(29.74)	87.79(30.03)	92.22(30.17)	86.4(27.07)	90.46(28.51)
Aeronáutica	87.74(32.65)	86.61(27.92)	96.64(29.36)	88.64(34.29)	96.28(34.47)
Manufactura	76.92(31.12)	85.42(25.92)	74.92(30.15)	69.77(31.1)	79.14(34.93)
Diseño Digital de Medios	83.74(30.49)	85.24(28.82)	92.75(29.43)	89.98(28.64)	87.41(30.64)
Arquitectura	76.78(31.26)	83.04(32.57)	72.97(30.5)	77.28(30.14)	77.57(27.97)
Contabilidad	73.77(30.22)	80.99(30.39)	79.23(30.06)	79.84(29.89)	80.31(28.25)
Odontología	73.57(33.05)	79.65(29.46)	73.24(29.48)	76.12(27.94)	75.37(29.88)
Sistemas Computacionales	76.06(30.99)	79.57(31.99)	75.39(32.85)	77.53(30.99)	80.25(32.99)
Ing. Mecatrónica	87.43(34.61)	78.82(32.39)	80.91(33.44)	91.99(29.01)	84.29(34.15)
Diseño Industrial	74.6(29.39)	77.11(29.74)	78.61(31.59)	80.1(31.02)	76.6(25.29)
Derecho	75.54(28.11)	76.89(26.02)	76.66(29.27)	74.34(27.94)	79.51(29.05)
Ing. Civil	82.68(30.31)	74.77(27.02)	80.76(28.27)	78.58(32.68)	81.03(29.26)
Educación	65.7(25.82)	74.05(26.6)	70.01(28.25)	67.99(25.76)	69.61(14.66)
Enfermería	64.82(31.56)	72.58(30.22)	64.48(26.98)	64.15(25.83)	63.79(27.61)
Ing. Industrial de Sistemas	72.41(28.07)	71.55(29.35)	69.77(29.41)	69.27(26.87)	74.52(29.07)
Ing. Mecánica	77.47(32.76)	71.33(29.64)	72.36(30.6)	76.18(31.54)	77.6(30.23)
Médico Veterinario Zootecnista	69.87(29.54)	70.71(25.23)	65.43(29.78)	56.79(26.78)	60.76(28.3)
Administración de Empresas	66.61(28.16)	69.9(26.52)	67.57(27.16)	63.44(26.74)	66.93(25.92)
Seguridad y Políticas Publicas	62.82(28.53)	69.15(24.97)	66.99(24.01)	69.21(22.66)	69.1(22.97)
Ing. Ambiental	82.49(32.24)	68.76(26.53)	83.34(30.98)	72.20(24.65)	79.32(28.81)
Nutrición	66.63(30.9)	67.62(28.85)	67.36(34.74)	64.12(33.15)	68.27(31.56)
Turismo	65.34(28.73)	62.33(26.26)	70.21(30.58)	63.37(25.52)	67.98(25.4)
Ing. Automotriz	70.21(29.95)	62.27(25.89)	70.98(27.84)	71.18(26.12)	73.86(29.6)
Diseño de Interiores	63.83(29.1)	60.86(22.47)	62.25(28.15)	70.9(28.87)	67.67(29.21)
Trabajo Social	57.81(22.82)	59.83(21.64)	65.29(23.88)	62.37(26.17)	58.91(18.69)
Entrenamiento Deportivo	57.58(26.65)	55.82(24.43)	49.44(23.58)	53.92(26.28)	50.48(24.3)
Promedio por nivel/ocasión/semestre	75.97(1.83)*	77.08(1.98)*	76.35(1.92)*	75.17(1.89)*	77.05(1.91)*
Intervalo del Confianza del 95%	[72.23, 79.70]	[73.04, 81.12]	[72.44, 80.27]	[71.31, 79.03]	[73.16, 80.94]
n por semestre de aspirantes	8698	4448	8113	4307	9051
Promedio Global/General	76.323(1.80)*	Limite Bajo	72.646	Limite Alto	80

Nota especifica 1. (…) = SD; (…)* = error estándar. El Limite Bajo y Alto señalaron un IC $_{95\%}$.
Nota especifica 2. La fuente fue elaboración propia.

Para el análisis, se promediaron los puntajes del EXHCOBA de los aspirantes (variable dependiente) por cada uno de los 32 pregrados ofrecidos respectivamente para cinco semestres que también se les llamó ocasiones o niveles (variable independiente con cinco niveles; Tabla 1). Cada pregrado sería el equivalente a un individuo que tomó un mismo test en cinco ocasiones distintas para ser congruente con el AMR. A través de las cinco ocasiones, los puntajes promedio más bajos fueron aproximadamente en los cincuenta (ej., trabajo social y entrenamiento deportivo) y los más altos aproximadamente llegaron a los 106 (i.e., medicina). La desviación estándar fue similar entre pregrados de las cinco ocasiones: aproximadamente 30 puntos. Otras características de los datos fueron: La curtosis y simetría estuvieron entre -1 y 1, lo cual indica una posible distribución normal a nivel de cada ocasión. Los cinco promedios por ocasión fueron de 75.17 a 77.08 con sus respectivos errores estándar e intervalos de confianza (*IC*) del 95% (Tabla 1). Se calculó el promedio global con $IC_{95\%}$ de las cinco ocasiones: 76.323 [72.66, 80.00] para indicar dónde se encontraría probablemente el verdadero promedio de la población (Tabla 1). Aunque se ofertaron más de 32 pregrados durante estos cinco semestres, solo se consideraron estos últimos porque coincidían en las cinco convocatorias consecutivamente para el análisis. Aparte, el AMR al igual que el resto de otros análisis estadísticos es más preciso para estimar los parámetros de las poblaciones cuando se tienen sets de datos completos. Se optó usar pregrados con todos los datos: aproximadamente 15 pregrados quedaron fueron del análisis porque no se ofertaron consecutivamente. Los datos usados no tenían información demográfica de los aspirantes y solo mostraron un número de identificación. Se implementó el test de normalidad con el Chi cuadrada de Kolmogorov-Smirnov que arrojó una $p = .20 > \text{alfa} = .05$. Bajo la hipótesis nula de que una distribución normal fue igual a la distribución observada, no se rechazó esta hipótesis y se asume que ambas son iguales: los datos tuvieron una distribución normal.

Los datos fueron analizados para observar si había valores atípicos que rebasaban el límite del valor absoluto de 3. Para tal análisis, se calculó el gran promedio (76.51: promedios de las cinco ocasiones que fueron 160 valores) con una $SD = 10.85$ para convertir cada uno de los valores usando la fórmula: $z_i = (x_i - \bar{x}) / SD_x$, donde x_i = es cada valor individual de un set de datos llamado x; $\bar{x}$ = promedio del set de datos llamado x; y SD_x = Desviación estándar del set de datos llamado x. El análisis de los valores atípicos mostró que los valores *z* estuvieron entre -2.49 y 2.69. Por lo tanto, no se detectaron valores atípicos.

El poder estadístico es la habilidad para detectar una diferencia que en realidad está allí: i.e., Es una probabilidad de correctamente rechazar una hipótesis nula falsa y así detectar un efecto genuino (Cohen, 1988; Autor, 2019). Por ejemplo, si existe una diferencia entre los promedios de dos poblaciones en alguna variable dependiente de interés y esta diferencia se detecta en las correspondientes muestras que se están comparando con un test de significancia estadística y con un poder de 80% o más, se rechazará acertadamente la hipótesis nula: i.e., H_0: el promedio de la población uno es igual al promedio de la población dos. Como consecuencia, los resultados del test estadístico apoyarán acertadamente la hipótesis alternativa: el promedio de la población uno no es igual al promedio de la población dos. Para los datos, se encontró un poder estadístico del 99.99% en el programa *Gpower* (Faul et al., 2007). Las hipótesis de interés fueron:

H_0: $\mu_{\text{ocasión 1}} = \mu_{\text{ocasión 2}} = \mu_{\text{ocasión 3}} = \mu_{\text{ocasión 4}} = \mu_{\text{ocasión 5}}$

H_A: Por lo menos, el promedio de una ocasión de una población es diferente a los demás.

El AMR tiene una serie de supuestos:

- La muestra fue elegida al azar de la población para posiblemente ser representativa: En el caso de la presente investigación, no se seleccionaron al azar los aspirantes con sus pregrados porque los pregrados debían ser ofertados en cinco ocasiones consecutivas. Sin embargo, la muestra contó con aproximadamente el 68% del total de pregrados (32 / 47).
- La variable dependiente está normalmente distribuida en la población: Se cumplió.
- Esfericidad: La varianza de la diferencia es homogénea. Este no, pero se llevó a cabo una corrección en la parte de los resultados.
- Los coeficientes de correlación entre los pares de variables son iguales. Los coeficientes de la *r* de Pearson fueron todos grandes (Tabla 2): .79 a .95. No existe diferencia estadísticamente significativa entre estos coeficientes de acuerdo con la Calculadora de Psychometrica (s.f.). Cohen (1988) sugirió los siguientes tamaños cuando no hay algo más en la literatura que lo indique de otra manera: efecto pequeño $r = .10$; mediano $r = .30$; y grande $r = .50$.

Tabla 2. Correlaciones entre Cinco Ocasiones

Ocasión	Otoño 2016	Primavera 2017	Otoño 2017	Primavera 2018	Otoño 2018
Otoño 2016	1				
Primavera 2017	0.86	1			
Otoño 2017	0.91	0.83	1		
Primavera 2018	0.84	0.79	0.84	1	
Otoño 2018	0.93	0.85	0.95	0.89	1

Nota especifica 1. La calculadora de Psychometrica (s.f.). fue usada para poner a prueba las diferencias entre los coeficientes de Pearson mediante una prueba z (alfa = .05) de Comparaciones de Correlaciones de Muestras Dependientes.
Nota especifica 2. La fuente fue elaboración propia.

Esfericidad es una condición en la cual las varianzas de las diferencias entre todas las combinaciones de grupos relacionados (ocasiones/niveles) son iguales (Girden, 1993). El alfa es comparado con una *p*, y los criterios son: $\alpha > p$ (rechazar la hipótesis nula: no se cumple con el supuesto de esfericidad), y $\alpha \leq p$ (no rechazar la hipótesis nula: se cumple con el supuesto de esfericidad). Cuando no se cumple el supuesto de esfericidad (varianzas de las diferencias que no son iguales), se le considera una violación seria que implica algún tipo de remedio para la situación o abandonar el test. Si se prosigue con el test, esta violación puede hacer el test muy liberal, lo cual quiere decir que se incrementa la posibilidad de incurrir en el Error Tipo I (probabilidad de rechazar una hipótesis nula que es cierta: parámetros de las poblaciones son iguales; ver a Cohen, 1988). Existen correcciones para producir un valor F_{critico} más valido (reducción de la probabilidad del Error Tipo I). La corrección se aplica a los *df* de la distribución *F* (Girden, 1993). En otras palabras, las correcciones a la violación de esfericidad incrementan el F_{critico} lo que causa que sea menos probable obtener un resultado estadísticamente significativo. El test de esfericidad es de Mauchly (test no-paramétrico; ver a Girden, 1993) que aparece en JASP al efectuar un AMR. Para los datos del presente estudio, las hipótesis son:

H_0: $\sigma^2_{\text{diferencia 1 y 2}} = \sigma^2_{\text{diferencia 1 y 3}} = \sigma^2_{\text{diferencia 1 y 4}} = \sigma^2_{\text{diferencia 1 y 5}} = \sigma^2_{\text{diferencia 2 y 3}} = \sigma^2_{\text{diferencia 2 y 4}} = \sigma^2_{\text{diferencia 2 y 5}} = \sigma^2_{\text{diferencia 3 y 4}} = \sigma^2_{\text{diferencia 3 y 5}} = \sigma^2_{\text{diferencia 4 y 5}}$

H_A (Hipótesis alternativa): Por lo menos una de las varianzas es diferente al resto

Respecto a la esfericidad, Girden (1992, p. 53) explico: “Tal homogeneidad entre las varianzas de las diferencias es un evento raro en estudios que involucran más de dos medidas repetidas de un comportamiento.” Para solucionar la falta de homogeneidad, Box (1954) propuso la estadística épsilon (ε). Esta propuesta fue perfeccionada por Geisser y Greenhouse (1958). Cuando épsilon es cercano a 1, las varianzas de las diferencias son más homogéneas y, por lo tanto, la extensión de la esfericidad es más grande (Girden, 1993). Una ε de un valor de 1 indica que la condición de esfericidad se cumple exactamente. Por otro lado, cuanto ε disminuya por debajo de 1, mayor será la violación de la esfericidad (ver a Girden,1993). Se puede pensar en ε como una estadística que describe el grado en que se ha violado el supuesto de esfericidad para llevar a cabo un AMR.

Después de haber corrido el modelo en JASP, la hoja de resultados mostró tres correcciones: (a) el valor más bajo que puede tomar ε se denomina estimación de límite inferior (lower-bound); (b) procedimientos de Greenhouse-Geisser con su ε (Greenhouse y Geisser, 1959); y (c) Huynd-Feldt con su ε. Estas correcciones intentan estimar ε, aunque de diferentes formas. Recordando, las estadísticas son estimaciones que provienen de muestras, y los parámetros vienen de las poblaciones completas. Como se están utilizando tres procedimientos diferentes, las estimaciones de ε tienden a ser diferentes también. Al estimar ε, todos estos procedimientos corrigen los *df* de la distribución *F* para incrementar los valores críticos del *F* y hacer la obtención de significancia estadística menos probable (ver a Girden, 1993). Por ejemplo, la disminución en los *df* causó que el $F_{critico}$ aumentara (Tabla 3 con los datos del presente estudio).

Tabla 3. Cambios en $F_{critico}$ causados por cambios en df

Estadística	$df_{numerador}$	$df_{denominador}$	$F_{critico}$	$F_{calculada}$
Asumiendo Esfericidad	4	124	2.44	1.344
Greenhouse-Geisser	3.056 (3)	94.728 (95)	2.7	1.344
Huynh-Feldt	3.444 (3)	106.764 (107)	2.69	1.344
Lower-bound	1	31	4.159	1.344

Nota especifica 1. Los valores críticos de la distribución F fueron calculados mediante la calculadora en línea de Finance Train. (s.f.). Se tuvieron que redondear algunos de los grados de libertad a los valores que aparecen en (…) porque la calculadora no admite decimales. Nota especifica 2. La fuente fue elaboración propia.

Para contrarrestar la violación del supuesto de esfericidad, hay un aumento en los errores de Tipo I, debido a que los valores críticos en una tabla *F* son demasiado pequeños cuando se usan los *df* originales (Girden, 1993). El valor real de la estadística $F_{calculada}$ no cambia como resultado de la aplicación de las correcciones. Para recordar los errores de las inferencias estadísticas, se presenta lo siguiente:

- Error Tipo I: Probabilidad de rechazar una hipótesis nula verdadera, y pasa cuando se rechaza una hipótesis verdadera (i.e., los promedios de dos poblaciones son iguales) y se concluye que existe algún efecto cuando en realidad no existe: falso positivo.
- Error Tipo II: Probabilidad de no rechazar una hipótesis falsa (i.e., los promedios de dos poblaciones no son iguales), y pasa cuando no se rechaza una hipótesis falsa y se concluye que no existe algún efecto cuando en realidad existe: falso negativo.

La recomendación es usar Greenhouse-Geisser cuando $\hat{\varepsilon}$ (épsilon estimada) < .75 porque esta corrección sub-estima a $\hat{\varepsilon}$ cuando este valor se acerca a 1: i.e., es una corrección conservadora. Por otro lado, Huynd-Feldt sobreestima $\hat{\varepsilon}$, así que se recomienda que se use cuando $\hat{\varepsilon}$ > .75 (cf. Huynh,1978). Dado que se violó el supuesto de esfericidad para los datos y el $\hat{\varepsilon}$ > .75, se usó la corrección de Huynh-

Feldt que se muestra en seguida: Numerador: *df* ocasiones = $\hat{\varepsilon}$ (k - 1); Para Huynh-Feldt: .861 (5 - 1) = 3.444; $\hat{\varepsilon}$ = épsilon estimada

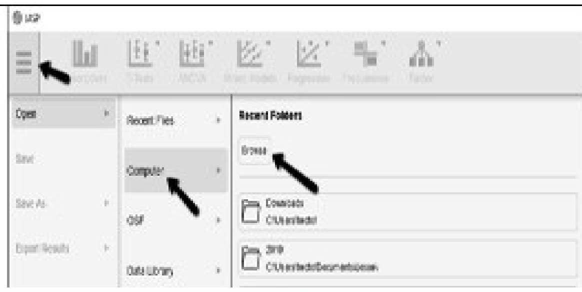
Denominador: *df* error = $\hat{\varepsilon}$ (k - 1) (*n* - 1); Para Huynh-Feldt: .861 (5 - 1) (32 - 1) = 106.764

Este test se llevó a cabo en JASP versión 0.16.3 para Windows. Se muestran la serie de pasos en las Figuras

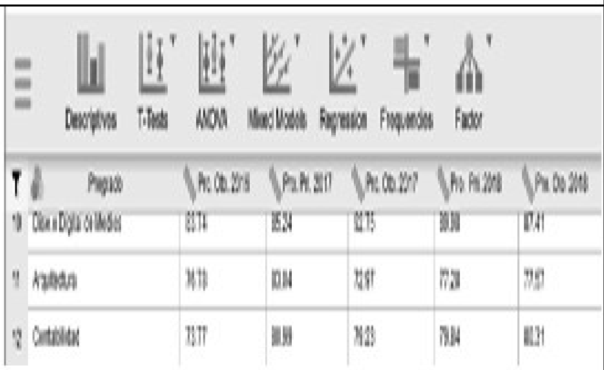
Figura 3. Primera Parte. Pasos para efectuar el AMR en JASP 0.16.3

1º Solo los promedios de La Tabla 1 se capturarán en un documento de Excel para salvarlo como .csv. Como resultado quedarían 32 pregrados en una columna y cinco columnas más para las ocasiones.

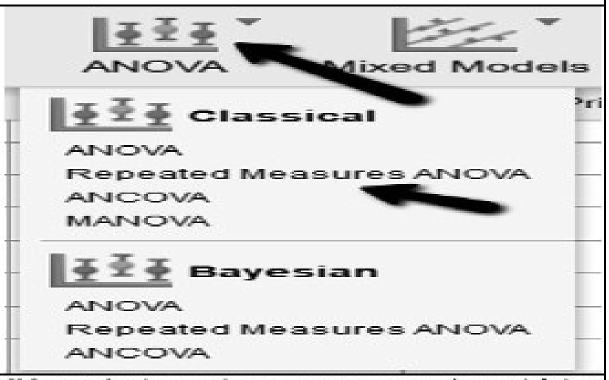
2º Se presiona la gráfica con las tres rayas < Computer < Browse para subir el documento de .csv a JASP.



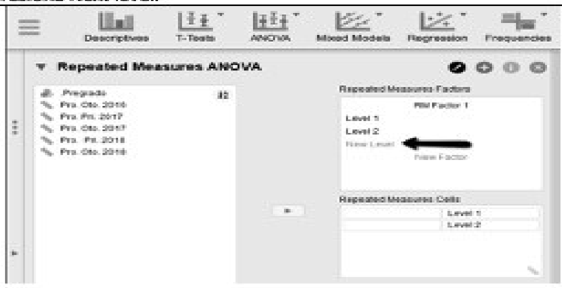
3º Aparecerá el set de datos. Solo habría que cerciorarse que aparezcan la columna con el nombre de las carreas medidas en una escala nominal (icono de diagrama de Venn) y las cinco variables con los datos medidos en una escala numérica (una regla).



4º Se presiona el logo de ANOVA < Repeated Measures ANOVA.



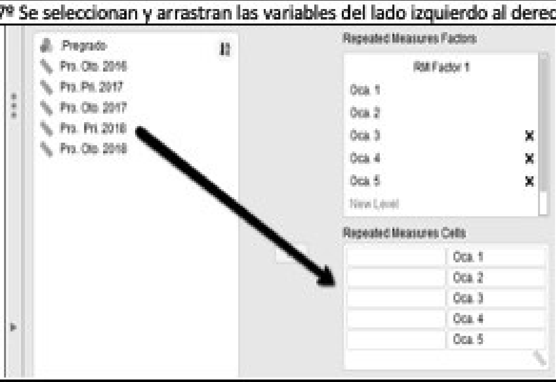
5º Aparecen las variables del lado izquierdo. La flecha indica que hay que crear las ocasiones que corresponden (5) así que se presiona Next level.

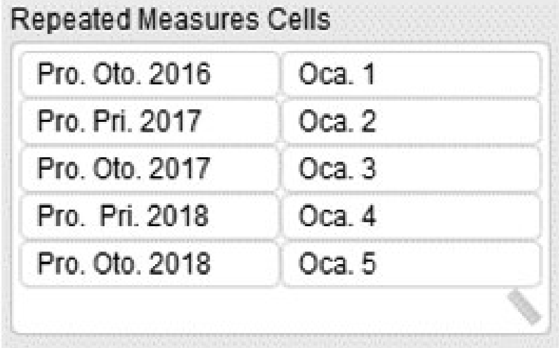


6º Se crean las cinco ocasiones como se muestra en la parte inferior derecha.

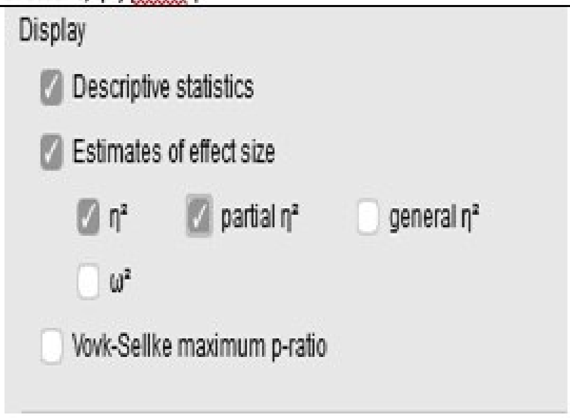
Figura 4. Segunda Parte. Pasos para efectuar el AMR en JASP 0.16.3

7º Se seleccionan y arrastran las variables del lado izquierdo al derecho. Deben de aparecer de esta manera.

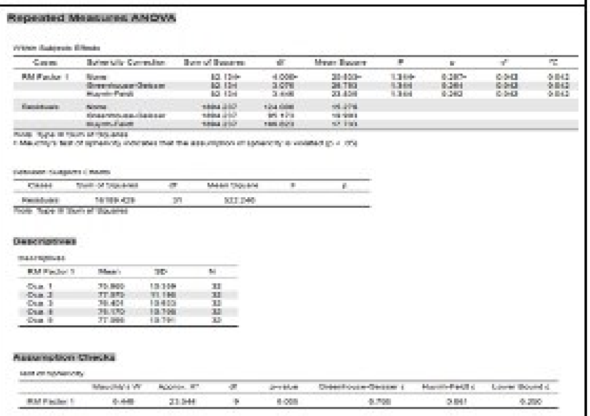




8º De lado izquierdo se presiona la pestaña para que aparezcan los efectos. Luego, se selecciona: Descriptive statistics, Estimates of effect size, η^2 y partial η^2.



9º Los resultados de los análisis aparecen automáticamente.



10º Para la Prueba de esfericidad se presiona la pestaña de Contrast. Se selecciona Polynomial, Assume Equal variances, y Confidence Intervals.

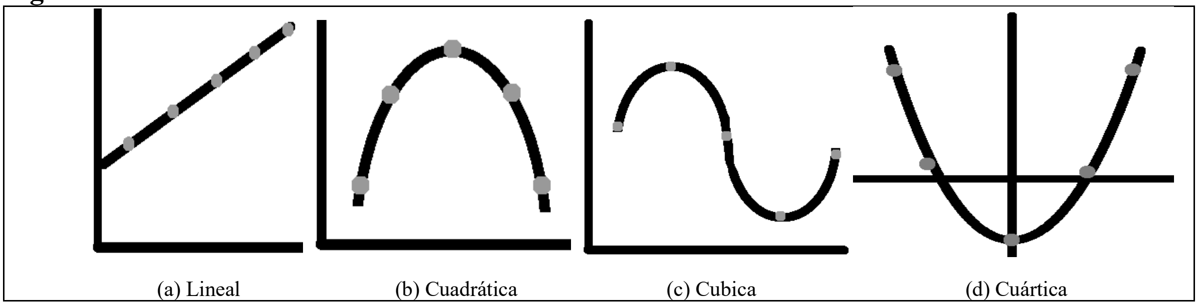
Polynomial Contrast - RM Factor 1

Comparison	Estimate	95% CI for Mean Difference Lower	Upper	SE	df	t	p
linear	0.088	-1.280	1.455	0.691	124	0.127	0.899
quadratic	0.266	-1.101	1.634	0.691	124	0.385	0.701
cubic	1.550	0.183	2.918	0.691	124	2.244	0.027
quartic	0.292	-1.075	1.660	0.691	124	0.423	0.673

Nota. La instrucciones son propias y la fuente es de las gráficas fue JASP.

Con el AMR, se puede probar la tendencia de los promedios de tres o más ocasiones con un test de significancia estadística. Detallando, Girden (1992) explicó que, si existe significancia estadística en alguno de los tipos de tendencia, esto quiere decir que los datos siguen ese patrón: La suma de cuadrados de cierto patrón explica la variabilidad de las ocasiones. Una tendencia es un movimiento de largo plazo que resulta de fenómenos tales como cambios en el ámbito demográfico, tecnológico, productivo, entre otros. JASP arroja cuatro posibles tendencias: lineal, cuadrática, cúbica y cuártica (Figura 5). Por ejemplo, esto puede ayudar a ver si algún fenómeno como puntajes del EXHCOBA van incrementando a través de las ocasiones si su probabilidad calculada es menor a algún alfa seleccionado (ej., $\alpha = .05$).

Figura 5. Tendencias de los Promedios



Nota. Fuente: Elaboración propia.

RESULTADOS Y DISCUSIÓN

Comenzando con los supuestos de AMR (Figura 6) fue el test de esfericidad de Mauchly ($w = .499$) resultó en $\chi^2 = 23.544$, $df = 9$, $p = .005 < \alpha =.05$, así que se rechazó la hipótesis nula: estadísticamente significativo porque se violó el supuesto de esfericidad. Esto implicó que el valor $F_{\text{calculado}}$ estuvo positivamente sesgado, y, por ello, sería inválido porque aumentaría el riesgo de un Error de Tipo I. Esto ameritó un remedio que consistió en aplicar correcciones a los df para obtener un F_{critico} válido.

Figura 6. El Test de Mauchly y el Épsilon de Greenhouse-Geisser; de Huynh-Feldt; y de Lower-bound.

Assumption Checks
Test of Sphericity

	Mauchly's W	Approx. Χ²	df	p-value	Greenhouse-Geisser ε	Huynh-Feldt ε	Lower Bound ε
RM Factor 1	0.449	23.544	9	0.005	0.768	0.861	0.250

Nota especifica 1. Lower-bound se pudo estimar manualmente con 1 / (k - 1): i.e., 1/ (5 -1) = .25. Este representa el peor escenario posible de violación a la esfericidad, así que La recomendación de varios autores ha sido: No usar esta corrección.
Nota especifica 2. La Fuente fue el Resultado del JASP con formato de APA.

En segundo lugar, JASP mostró en dos tablas las varianzas entre ocasiones, individuos y residual (Figura 7). Las dos tablas que arrojó JASP se tomaron los valores de la corrección de Huynh-Feldt porque eran las que correspondían por ser $\varepsilon > .75$ (Figura 7 muestra estos valores sombreados) y el $\eta^2_p = .041$.

Figura 7. Test de Entre sujetos y Resultados de las Varianzas

Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
RM Factor 1	None	82.134 [a]	4.000 [a]	20.533 [a]	1.344 [a]	0.257 [a]	0.042
	Greenhouse-Geisser	82.134	3.070	26.753	1.344	0.264	0.042
	Huynh-Feldt	82.134	3.446	23.835	1.344	0.262	0.042
Residuals	None	1894.237	124.000	15.276			
	Greenhouse-Geisser	1894.237	95.173	19.903			
	Huynh-Feldt	1894.237	106.823	17.733			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
Residuals	16189.429	31	522.240		

Note. Type III Sum of Squares

Nota especifica 1. Las partes sombreadas sirvieron para armar la Tabla 4 con los resultados que corresponde al AMR dada la violación del supuesto de esfericidad.
Nota especifica 2. La Fuente fue el Resultado del JASP con formato de APA.

La Tabla 4 muestra la forma de organizar un AMR con un factor. JASP no dio el resultado de la suma de cuadrados de la variación total (en negrillas), pero ésta se estimó al sumar las otras tres sumas de cuadrados: 1,6189.429 + 82.134 + 1,894.237 = **1,8165.80**. También, se muestra aquí como calcular el η^2_p de manera manual también como su $IC_{95\%}$ [0, .1053].

Tabla 4. Resultados Organizados de una Tabla de AMR

Fuente	Suma de Cuadrados	*df*	Promedio de la Suma de Cuadrados	$F_{calculada}$	*p*	η^2_p
Varianza de Individuos	Ecuación 1 SS_i = 16,189.429	Ecuación 5 31	Ecuación 9 MS_i = 522.24			
Varianza de Ocasiones/ Niveles	Ecuación 2 SS_o = 82.134	Ecuación 6 $df_{numerador}$ = 3.45	Ecuación 10 MS_o = 23.84	Ecuación 12 1.344	.262	.042
Varianza Residual	Ecuación 3 $SS_{Residual}$ = 1,894.237	Ecuación 7 $df_{denominador}$ = 106.82	Ecuación 11 MS_{Res} = 17.733			
Variación Total	Ecuación 4 SS_{Total} = 18,165.80	Ecuación 8 159				

La Tabla 5 complementa a la Tabla 4 ya que se exponen las fórmulas que subyacen en JASP para poder calcular los coeficientes.

Tabla 5. Fórmulas y Explicaciones del AMR

Suma de Cuadrados	*df*	Promedio de la Suma	$F_{calculada}$	*p*	η^2_p

de Cuadrados				
Ecuación 1 $SSi = \sum_{i=1}^{n}\left(\frac{T_i^{2}}{k}\right) - \frac{T^2}{n}$	Ecuación 5 $(n - 1)$	Ecuación 9 $MS_i = SS_i / (n - 1)$		
Ecuación 2 $SSo = \sum_{k=1}^{k}\left(\frac{T_k^{2}}{n}\right) - \frac{T^2}{n}$	Ecuación 6 *df numerador* $= (k - 1)$*	Ecuación 10 $MS_o = SS_o / (k - 1)$	Ecuación 12 *F cal.* $= MS_o / MS_{Res}$	Ecuación 13 $\eta^2_p = SS_o / (SS_o + SS_{Residual})$
Ecuación 3 $SS_{Residual}$ = Suma de Cuadrados de la Variación Total (–) Suma de Cuadrados de la Varianza de Individuos (–) Suma de Cuadrados de la Varianza entre Ocasiones/ Niveles	Ecuación 7 *df denomiador* $= (n - k)\ (k - 1)$*	Ecuación 11 $MS_{Res} = SS_{Residual} / (n - k)\ (k - 1)$		
Ecuación 4 $SSTotal = \sum_{k=1}^{k}\sum_{i=1}^{n} x_{ik}^{2} - \frac{T^2}{N}$	Ecuación 8 $N - 1$			
Explicación de términos	Variación Total $\sum$ = suma de un set de valores k = número de niveles o de ocasiones n = número de individuos T = Suma de todos los individuos en todos los niveles u ocasiones: $y_{ki} + \ldots + y_{kn}$ N = nk = número total de puntajes Varianza de Individuos T_i = Suma de los puntajes por cada uno de los individuos (ith): i.e., del presente estudio; $y_{1i} + y_{2i} + y_{3i} + y_{4i} + y_{5i}$ Varianza de Ocasiones/ Niveles T_k = Suma de los puntajes por cada uno de los niveles u de las ocasiones (kth): i.e., $y_{1i} + \ldots + y_{1n}$			

Nota especifica 1. *Estos df fueron multiplicados por $\hat{\varepsilon}$ debido a la violación del supuesto de esfericidad.
Nota especifica 2. La formulas provienen de Girden (1992) y existen otras fórmulas para obtener estos mismos coeficientes (ver a Laerd Statistics, 2018). Nota especifica 3. La Elaboración con datos de JASP fue propia.

El test de significancia de una tendencia (Figura 8). En resumen, la parte sombreada mostró que hubo una probabilidad calculada menor al alfa de .05 (i.e., la tendencia cubica). Cuando se implementan múltiples comparaciones o análisis exploratorios, es recomendable usar un nivel de significancia más conservador al tradicional de .05 como un alfa de .01 (cf. Bonferroni, 1936. En los análisis de tendencias, se implementó un alfa = .01 porque eran exploratorios aunque con esto se aumentó el riesgo del Error Tipo II.

Figura 8. Resultados de Tendencias

Polynomial Contrast - RM Factor 1

		95% CI for Mean Difference					
Comparison	**Estimate**	**Lower**	**Upper**	**SE**	**df**	**t**	**p**
linear	0.088	-1.280	1.455	0.691	124	0.127	0.899
quadratic	0.266	-1.101	1.634	0.691	124	0.385	0.701
cubic	1.550	0.183	2.918	0.691	124	2.244	0.027
quartic	0.292	-1.075	1.660	0.691	124	0.423	0.673

Nota. La Fuente fue el Resultado del JASP con formato de APA.

Se estimó un *IC* del 95% para el η^2_p usando el SPSS versión 23 de acuerdo con las indicaciones dadas por Lakens (2018): $IC_{95\%}$ = [0, .1053]. Tanto Richardson (2011) como Cummings (2013) habían recomendado el cálculo de un *IC* para estimar el parámetro de la población a la cual se desea hacer la inferencia.

La respuesta a la pregunta de investigación fue: *No*. La razón para esta respuesta es que el AMR con la corrección de Huynh-Feldt mostró que los promedios de las muestras no son estadísticamente significativos F (3.446, 106.823) = 1.344 y p = .262. El resultado de lo anterior fue que *no* se pudo rechazar la H_0 porque no hubo evidencia que la contradijera. El tamaño del efecto estuvo entre pequeño y mediano η_p^2 = .042, y el $IC_{95\%}$ = [0, .1053]. No hubo tendencia de los promedios porque todas las probabilidades calculadas fueron mayores al alfa de .01.

CONCLUSIONES

No existe una diferencia estadísticamente significativa entre los puntajes promedio de los aspirantes en sus carreras de elección a través de las cinco ocasiones, fue la respuesta a la pregunta de investigación: i.e., la evidencia apoyó la hipótesis nula de que no hay diferencia entre los puntajes de las ocasiones. Aunque la muestra de 32 pregrados no incluyó a todos los pregrados porque quedaron fuera 15 por no tener datos completos, se podrían generalizar estos resultados hasta cierto punto porque se tuvo una muestra de aproximadamente el 68%. Las ocasiones o generaciones están altamente correlacionadas entre sí: ej., los resultados de las muestras mostraron que el patrón de los aspirantes a carreras como medicina con los más altos puntajes, intermedios como los de ingeniería y de los bajos puntajes de entrenamiento deportivo prácticamente se repitieron a través de las cinco generaciones. Debido a que no hubo una diferencia estadísticamente significativa y se contó con un poder estadístico del 99.99%, las diferencias entre las ocasiones se debieron probablemente a una variación natural de los promedios de los puntajes sin que se haya detectado algún efecto grande de η^2_p (i.e., Cohen, 1988, con un efecto grande: $\eta^2_p = .14$). Con lo anterior se cumplió también uno de los objetivos que era el análisis de los datos, así como el otro objetivo al explicar cómo llevar a cabo el AMR de un factor en JASP. Al alcanzar estos dos objetivos se contribuyó al conocimiento.

Una de las limitaciones fue que no se tuvo acceso a la información como: ej., edad, sexo, escuela de origen, nivel socioeconómico, calificaciones previas etc. Con esta información se pudieron haber llevado a cabo análisis más complejos para poder explicar y predecir los puntajes del EXHCOBA: ej., regresión múltiple (ver a Tabachnick y Fidell, 2018). La descripción, explicación y predicción de los puntajes del EXHCOBA, entre otros, podrían ser muy útiles para la toma de decisiones de admisión de universidades (ver a Autor et al., 2016).

Dada la importancia de la admisión, habría que estudiar más los datos del EXHCOBA para poder estimar qué variables se asocian a sus puntajes. Por ejemplo, si el nivel socio-económico se asocia altamente con los puntajes donde los aspirantes considerados afluentes obtienen los puntajes más altos. Una universidad podría remediar las desventajas de los aspirantes no afluentes al poner cuotas por nivel socio-económico. Otra sería el continuar analizando los datos del EXHCOBA con diversos análisis estadísticos de comparaciones entre grupos y relaciones entre variables. Esto sería para tratar de llenar los grandes huecos que existen en el tiempo y en los temas de análisis de este test. Como señala Backhoff et al. (2011) los exámenes de ingreso a las instituciones de educación superior pueden proporcionar información importante a nivel grupal, regional o nacional, sobre el tipo de habilidades y conocimientos que dominan los estudiantes, así como identificar las áreas que presentan dificultades para su dominio. Esto permite que el sistema educativo pueda implementar políticas educativas dirigidas a mejorar la educación en función de las debilidades que se presentan en una institución, estado o país.

Los resultados de los exámenes de admisión se utilizan, probablemente las más de las veces, para seleccionar a los estudiantes que desean estudiar un pregrado. Dichos resultados podrían ser de utilidad para diseñar cursos propedéuticos para tratar de llenar los vacíos de conocimientos y habilidades necesarios para cursar una universidad. También, servirían para investigar los efectos que ejercen sobre el aprendizaje las características de las escuelas, las actividades extracurriculares, el logro escolar y los niveles socioeconómicos de los estudiantes. Resultaría interesante generar índices educativos por estados, lo que permitiría evaluar el sistema educativo a nivel bachillerato y así, poder realizar estudios comparativos de logro educativo entre los estados.

Otro aspecto importante que se pudiera considerar en trabajos posteriores a raíz de esta investigación es el seguimiento en las trayectorias escolares, así como su participación en el mercado laboral una vez concluidos los estudios. Ya que de esto dependerá el desempeño de los individuos y su impacto en el desarrollo económico de sus regiones. Estudios recientes demuestran la correlación existente entre el proyecto académico de los aspirantes y su contribución al desenvolvimiento del capital humano como componente del desarrollo.

Volviendo a cierta información antes mencionada, el AMR se puede utilizar en diseños experimentales donde exista un pretest, un tratamiento y un postest para poder inferir causa y efecto. El AMR se puede aplicar a diseños no experimentales como el llevado a cabo en el presente estudio. Existen AMR que además de hacer comparaciones dentro de los grupos o personas también pueden hacer como comparaciones entre grupos (ver a Maxwell et al., 2018). Por ejemplo, si se hubiera tenido el sexo de los aspirantes del presente estudio, se hubieran podido comparar el grupo de mujeres contra el grupo de hombres a través de las cinco ocasiones.

REFERENCIAS BIBLIOGRAFICAS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Autor. (2016).

Autor. (2019).

Antillón, L., Larrazolo, N., & Backhoff, E. (2006). Igualación equipercentil del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *RELIEVE, 12*(2), 205–217. http://doi.org/10.7203/relieve.12.2.4227

Antillón, L., Larrazolo, N., & Backhoff, E. (2008). Igualación lineal de tres versiones del examen de habilidades y conocimientos básicos (EXHCOBA). *Revista Iberoamericana de Evaluación*

Educativa, 1(2), 192–203.

Backhoff, E., Ibarra, M., & Rosas, M. (1996). Desarrollo y validación del sistema computarizado de exámenes (SICODEX). *Revista de la Educación Superior, 1*(97), 41–54.

Backhoff, E., Ibarra, M., & Rosas, M. (1997). Evaluación por computadora: una nueva tecnología para la aplicación de exámenes de admisión. *Revista Psicología Contemporánea, 4*(2), 4–11.

Backhoff, E., Larrazolo, N., & Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa, 2*(1), 11–27.

Backhoff, E., Larrazolo, N., & Tirado, F. (2011). Habilidades verbales y conocimientos del español de estudiantes egresados del bachillerato en México. *Revista de la Educación Superior, 4*(160), 9–27.

Backhoff, E., & Tirado, F. (1992). Desarrollo del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista de la Educación Superior, 3*(83), 95–118.

Backhoff, E., Larrazolo, N., Pérez, J.C., & Rojas, G. (2015). Análisis de la estructura cognitiva del área de habilidades cuantitativas del EXHCOBA mediante el modelo LLTM de Fisher. *Revista Internacional de Educación y Aprendizaje, 3*(1), 25–38.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.

Box, G. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, 25, 484–498.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Psychology Press.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* Routledge.

Finance Train. (s.f.). *F Value Calculator Online*. https://financetrain.com/calculator/f-value

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://www.gpower.hhu.de/

Girden, E. R. (1992). *ANOVA: Repeated measures.* Sage University Paper.

Geisser, S., & Greenhouse, S. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics,* 29, 885–891.

Greenhouse, S., & Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika, 24*(2), 95–112.

Goss-Sampson, M.A. (2020). *Statistical analysis in JASP: A guide for students* (4ª ed., v0.14). JASP.

Hinkle, D. W., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for Behavioral Sciences* (5ª ed.). Houghton Mifflin.

Huynh, H. (1978). Some approximate tests for repeated measurement design. *Psychometrika, 43*(2), 161–175.

JASP (s.f.). *JASP*. The JASP Team. Recuperado de https://jasp-stats.org/

Kotz, S. (2006). Encyclopedia of Statistical Sciences: 16 volúmenes (2nd ed.). Nueva York, NY: Wiley-Interscience.

Laerd Statistics. (2018). *Repeated Measures ANOVA - Understanding a Repeated Measures ANOVA.* Recuperado de Laerd Statistics website: https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide.php

Lakens, D. (2018). *The 20% Statistician: A blog on statistics, methods, and open science. Understanding 20% of statistics will improve 80% of your inferences.* Recuperado de http://daniellakens.blogspot.com/2014/06/calculating-confidence-intervals-for.html

Métrica Educativa, A.C. (2018). Página Principal. © Metrica Educativa, A.C. Recuperado de Metrica website: http://metrica.edu.mx/examenes/exhcoba/

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing Experiments and Analyzing Data* (3ª ed.). Routledge.

Psychometrica. (s.f.). Correlation Coefficients: An Overview. June 27, 2023, Recuperado de https://www.psychometrica.de/correlation.html#dependent

Richardson, J. (2011). Eta squared and partial eta squared as a measure of effect size in educational research. *Educational Research Review, 6*(2), 135–147. http://doi.org/10.1016/j.edurev.2010.12.001

Salkind, N. (2007). *Encyclopedia of Measurement and Statistics* (Vol. 1-3). Sage.

Sánchez, C., Larrazolo, N., & Rosas, M. (2008). El Examen de Habilidades y Conocimientos Básicos (EXHCOBA): Desarrollo, resultados y perspectivas. *Revista Mexicana de Psicología*, Número Especial, 18–52.

Tabachnick, B., & Fidell, L. (2018). *Using Multivariate Statistics* (7ª ed.). Pearson.

Tirado, F., & Backhoff, E. (1999). La compleja elaboración de exámenes, 16 razones para utilizar la opción no sé. *Revista Mexicana de Investigación Educativa, 4*(7), 13–26.

Tirado, F., Backhoff, E., Larrazolo, N., & Rosas, M. (1997). Validez predictiva del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Mexicana de Investigación Educativa, 2*(3), 67–84.