

II Jornadas Internacionales de Estadística Aplicada 5 y 6 de diciembre de 2019

ANÁLISIS DEL RENDIMIENTO ACADEMICO DE LOS ESTUDIANTES APLICANDO DISEÑO FACTORIAL

López, David Alejandro, Marcial, Sofia Daiana

Facultad de Ingeniería, Universidad Nacional de Salta.

Salta-Capital. davidalelopez1998@gmail.com. Telf. 3874129200

RESUMEN:

A modo de experimentación para verificar los conocimientos aprendidos en la Cátedra se decidió realizar un estudio con el cual se pretendía conocer el efecto que tiene los diferentes factores sobre la materia de análisis matemático I con respecto a los promedios de los estudiantes y la significancia que tienen las diferentes carreras de ingeniería en estas notas; utilizando para ello las herramientas estadísticas que brinda el diseño experimental a través de sus modelos factoriales completos.

PALABRAS CLAVES

Aleatoriedad, Hipótesis, Independencia, Modelo Factorial, Normalidad, Significancia.

INTRODUCCION

Los diseños factoriales son ampliamente utilizados en experimentos en los que intervienen varios factores para estudiar el efecto conjunto de éstos sobre una respuesta. Existen varios casos especiales del diseño factorial general que resultan importantes porque se usan ampliamente en el trabajo de investigación y constituyen la base para otros diseños de gran valor práctico. El más importante de estos casos especiales ocurre cuando se tienen k factores, cada uno con dos niveles. Estos niveles pueden ser cuantitativos o cualitativos. En esta investigación se incluye tres factores de tipo cualitativo con los que se busca analizar los posibles efectos sobre el desempeño de los estudiantes de la Facultad de ingeniería en determinada materia en relación a su nota final. Esto permitirá establecer cierto tipo de caracterización del rendimiento académico del estudiante con base a un soporte estadístico robusto, tal como lo es el que se obtiene al aplicar un diseño experimental.

Por lo tanto, la finalidad de este trabajo es demostrar si el género, carrera y comisión asignada son los factores que influyen en el rendimiento académico de los estudiantes de primer año de la facultad.

METODOLOGIA

Con el desarrollo de esta investigación se pretende conocer la relación existente entre el rendimiento académico del estudiante universitario ingresante en la materia (AMI) con referencia al sexo del estudiante, la carrera en la que está matriculado y la comisión que le fue asignada. Para el estudio se eligió el periodo lectivo 2018. Los datos fueron cedidos por los responsables de la materia. Se trabajó con seis comisiones las cuales presentaban las mismas condiciones en los dos cuatrimestres.

Para determinar las comisiones se seleccionaron tres de ellas al azar por medio de un generador de números aleatorios. Teniendo el objeto de investigación ya identificado, se indagó en los listados existentes de calificaciones en los cuales están registrados: los códigos de los estudiantes, el programa académico, el promedio de la asignatura; teniendo las comisiones identificadas y conociendo los estudiantes matriculados por cada carrera se empezó a seleccionar los estudiantes agrupándolos en primera instancia por sus carreras.

Marco teórico

Las formulas desarrolladas en este trabajo para las diferentes pruebas se tomaron de la bibliografía básica, Roberto Kuel.

1) Prueba de normalidad de Shapiro – Wilks:

Este contraste nos permite probar que nuestra muestra tiene una distribución normal, mediante el siguiente procedimiento:

H_0 : tiene una distribución normal

H_a : no tiene distribución normal

$$W = \frac{A^2}{n \cdot s^2}$$

$$\text{siendo: } A = \sum (a_{j,n} * (x_{(n-j+1)} - x_j))^2$$

$$n \cdot s^2 = \sum (x_i - \bar{x})^2$$

Con $h = n/2$ si es par

$h = (n-1)/2$ si es impar

Si $w_{obs} < w_c$, rechazo H_0 .

2) Diseño factorial aleatorizado

Modelo lineal:

$$y_{ijkl} = \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{ijk} + \varepsilon_{ijkl}$$

$i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$; $k = 1, 2, \dots, c$; y $l = 1, 2, \dots, n$,

donde:

μ : media general

a_i, b_j y c_k : efectos principales (genero, carrera y docente)

$(ab)_{ij}, (ac)_{ik}$ y $(bc)_{jk}$: efectos de interacción de 2 factores (carrera docente, carrera género y docente genero)

$(abc)_{ijk}$: efecto de interacción de 3 factores (carrera, docente, genero)

ε_{ijkl} : error experimental

Supuestos:

$$a_i \sim N(0, \sigma^2) \text{ e independiente}$$

$b_j \sim N(0, \sigma^2)$ e independiente

$c_k \sim N(0, \sigma^2)$ e independiente

$(ab)_{ij} \sim N(0, \sigma^2)$ e independiente

$(ac)_{ik} \sim N(0, \sigma^2)$ e independiente

$(bc)_{jk} \sim N(0, \sigma^2)$ e independiente

$(abc)_{ijk} \sim N(0, \sigma^2)$ e independiente

$\varepsilon_{ijkl} \sim N(0, \sigma^2)$ e independiente

Cuadro ANOVA:

A partir de las siguientes formulas calcularemos las sumas de cuadrados de los factores, la interacción, el error y el total

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl}^2 - \frac{y_{\dots}^2}{abcn}$$

Ecuación (1): Suma total de cuadrados.

$$SS_A = \sum_{i=1}^a \frac{y_{i\dots}^2}{bcn} - \frac{y_{\dots}^2}{abcn}$$

$$SS_B = \sum_{j=1}^b \frac{y_{\cdot j \cdot \cdot}^2}{acn} - \frac{y_{\dots}^2}{abcn}$$

$$SS_C = \sum_{k=1}^c \frac{y_{\cdot \cdot k \cdot}^2}{abn} - \frac{y_{\dots}^2}{abcn}$$

Ecuación (2): Suma de cuadrados para los efectos principales.

$$\begin{aligned} SS_{AB} &= \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij \cdot \cdot}^2}{cn} - \frac{y_{\dots}^2}{abcn} - SS_A - SS_B \\ &= SS_{\text{subtotales (AB)}} - SS_A - SS_B \end{aligned}$$

$$\begin{aligned} SS_{AC} &= \sum_{i=1}^a \sum_{k=1}^c \frac{y_{i \cdot k \cdot}^2}{bn} - \frac{y_{\dots}^2}{abcn} - SS_A - SS_C \\ &= SS_{\text{subtotales (AC)}} - SS_A - SS_C \end{aligned}$$

$$\begin{aligned} SS_{BC} &= \sum_{j=1}^b \sum_{k=1}^c \frac{y_{\cdot jk \cdot}^2}{an} - \frac{y_{\dots}^2}{abcn} - SS_B - SS_C \\ &= SS_{\text{subtotales (BC)}} - SS_B - SS_C \end{aligned}$$

Ecuación (3): Suma de cuadrados de interacción de dos factores.

$$SS_{ABC} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \frac{y_{ijk}^2}{n} - \frac{y_{...}^2}{abcn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$= SS_{\text{subtotales (ABC)}} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

Ecuación (4): Suma de cuadrados de interacción de tres factores.

$$SS_E = SS_T - SS_{\text{subtotales (ABC)}}$$

Ecuación (5): Suma de cuadrados del error.

ANOVA:

F. de variabilidad	GL	SC	CM	F
A	(a-1)	SC_A	CM_A	CM_A/CM_{ERROR}
B	(b-1)	SC_B	CM_B	CM_B/CM_{ERROR}
C	(c-1)	SC_C	CM_C	CM_C/CM_{ERROR}
AB	(a-1)(b-1)	SC_{AB}	CM_{AB}	CM_{AB}/CM_{ERROR}
AC	(a-1)(c-1)	SC_{AC}	CM_{AC}	CM_{AC}/CM_{ERROR}
BC	(b-1)(c-1)	SC_{BC}	CM_{BC}	CM_{BC}/CM_{ERROR}
ABC	(a-1)(b-1)(c-1)	SC_{ABC}	CM_{ABC}	CM_{ABC}/CM_{ERROR}
ERROR	abc(r-1)	SC_{ERROR}	CM_{ERROR}	
TOTAL	abcr-1	SC_{TOTAL}	CM_{TOTAL}	

Tabla (1): Tabla de análisis de varianza para el modelo de efectos aleatorios de tres factores.

Método de Tukey:

Lo usamos para comparar las comisiones. El procedimiento es:

$$H_0: Y_i = Y_j \text{ para todo } i \neq j$$

$$H_0: Y_i \neq Y_j \text{ para algun } (i, j)$$

$$T = \frac{t*(t-1)}{2}, \text{ numero de comparaciones}$$

El estadístico de prueba es:

DHS: Diferencia honestamente significativa

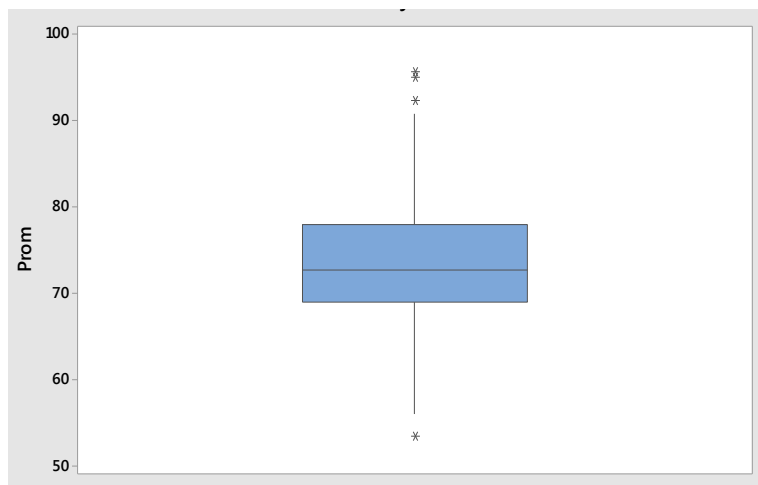
$$DHS = q_{[\alpha_e; t; (N-t)]} * \sqrt{\frac{CM_{error}}{r}} \text{ Para cuando las réplicas son iguales } r_j = r$$

El intervalo de confianza:

$$|\bar{y}_i - \bar{y}_j| \pm DHS(k, \alpha_e)$$

DESARROLLO

En primer lugar, verificamos que no existan valores atípicos en la muestra para ello realizamos un diagrama de caja y eliminamos los datos atípicos.



Gráfica (1): Diagrama de caja.

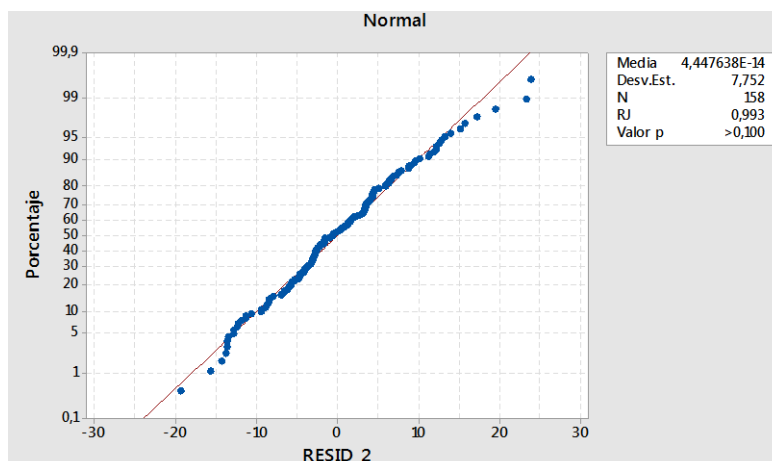
Para que un experimento planteado por cualquier persona sea aceptado (fiable) debe pasar dos condiciones estadísticas importantes las cuales están dadas en la aleatoriedad de los errores los cuales siguen una distribución normal con media cero y varianza constante, siendo independientes entre sí.

Análisis de Varianza

Fuente	GL	SC	CM	Valor F	Valor p
Carrera	3	16,03	5,345	0,09	0,968
Comisión	2	223,36	111,679	1,79	0,171
Genero	1	36,95	36,951	0,59	0,443
Error	151	9435,56	62,487		
Total	157	9738,66			

Tabla (2): Anova para la variable promedio.

Podemos concluir con los valores p obtenidos que la varianza es uniforme, procedemos a realizar el análisis de los residuos para corroborar que estos se distribuyan normal.



Gráfica (2): Prueba de Normalidad.

Se observa en esta prueba de normalidad que el Valor P es de 0.100 con lo que se puede concluir que se cumple con el supuesto de normalidad con lo cual podemos proseguir con el análisis del promedio final con las tres comisiones seleccionadas que cumplan con las condiciones solicitadas.

Para ello tomamos el promedio de cada cuatrimestre separado por categorías como se indica a continuación

DOCENTE	CARRERA							
	ING. INDUSTRIAL		ING. ELECTROMECANICA		ING.QUIMICA		ING.CIVIL	
	F	M	F	M	F	M	F	M
1	73,06	71,79	69,91	71,84	70,45	71,59	71,54	71,13
	73,06	72,74	62,6	73,1	75,66	72,52	72,14	71,94
2	71,56	73,69	79,09	72,19	73,71	73,45	73,8	73,14
	71,08	70,68	73,72	70,94	71,96	70,19	58,78	71,69
3	83,54	74,56	82,91	74,93	73,91	74,69	74,33	74,62
	61,46	74,95	85,73	74,87	74,82	75,42	75,29	75,5

Tabla (3): Promedio de los estudiantes por género.

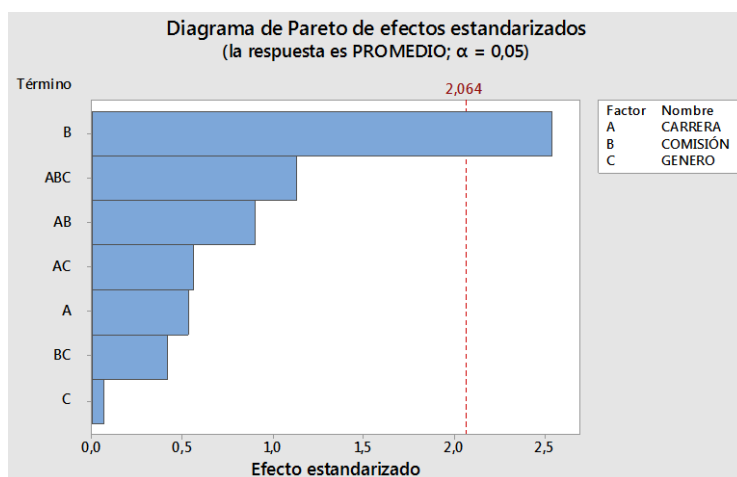
Procedemos a realizar el análisis factorial de un modelo lineal aleatorio utilizando un nivel de significancia de 0.05

Análisis de Varianza

Fuente de variabilidad	GL	SC	CM	Valor F	Valor p
CARRERA	3	34,666	11,5555	0,64	0,596
COMISIÓN	2	172,182	86,0912	4,78	0,018
GENERO	1	0,079	0,0792	0,00	0,948
CARRERA*COMISIÓN	6	122,257	20,3761	1,13	0,375
CARRERA*GENERO	3	36,164	12,0545	0,67	0,579

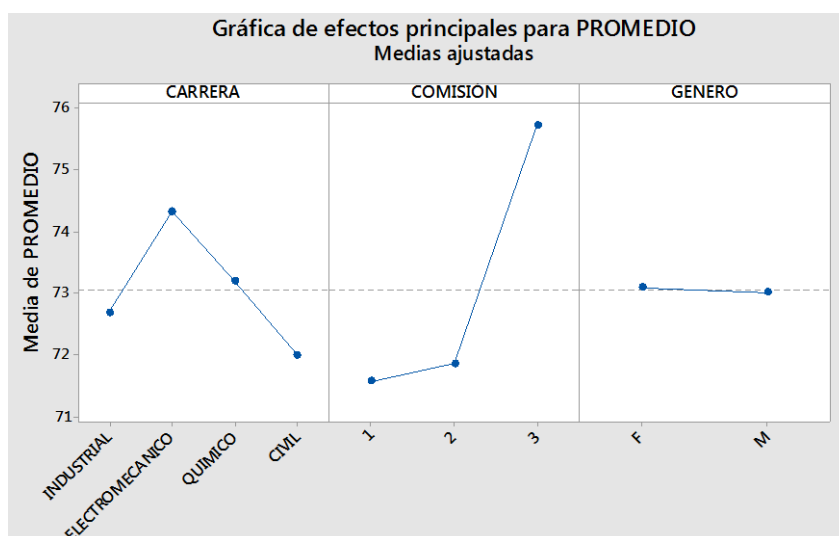
COMISIÓN*GENERO	2	14,164	7,0819	0,39	0,679
CARRERA*COMISIÓN*GENERO	6	147,024	24,5041	1,36	0,270
Error	24	432,363	18,0151		
Total	47	958,899			

Tabla (4): Anova para la variable promedio.



Gráfica (3): Diagrama de Pareto efectos estandarizados.

Como se observa en la tabla(4) y en la grafica(3), después de aplicar un diseño factorial completo, la única variable que tiene significancia para el experimento es “Comision” (esto se ve en la última columna en donde su valor P no es significativo ya que es menor al alfa que ha sido establecido de 0,05), con lo cual se puede afirmar que la carrera y el genero, no influyen en la nota promedio de la asignatura.



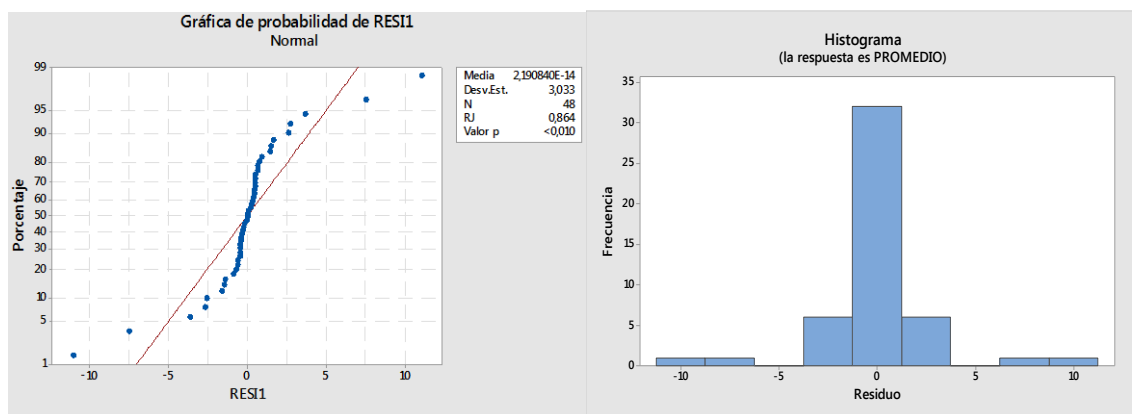
Gráfica (4): Efectos estandarizados para el promedio.

En la gráfica (4) se observa en la variable comisión en especial 1 y 2 presentan un efecto similar en la nota definitiva del estudiante y parece que hay alguna diferencia con la comisión 3, para estar más seguros con esta apreciación, se recurre a la prueba

de intervalos de confianza simultáneos de TUKEY, con lo que se podrá comprobar si existe una diferencia significativa entre las distancias observadas en dicha gráfica.

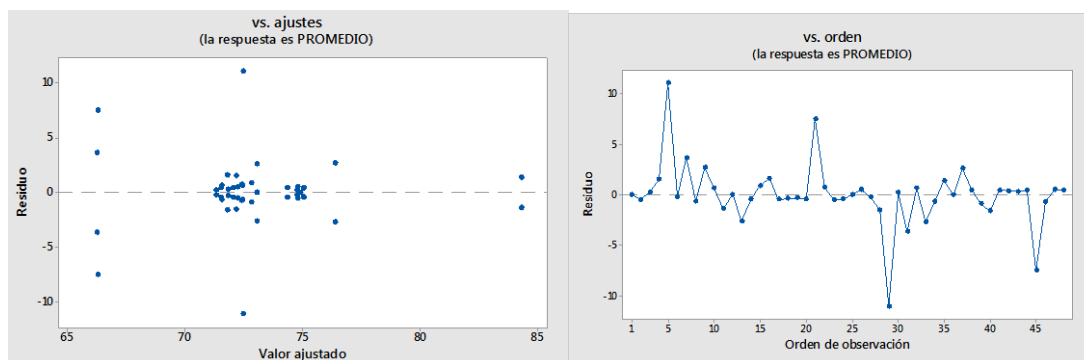
Análisis de Residuos:

De la gráfica (5) obtenemos que los residuos no se distribuyen normal ya que el valor p es de 0,010 pero en este caso podemos decir que se distribuyen aproximadamente normal debido a que en la gráfica 6 se observa una campana aguda con colas alargadas.



Gráfica (5): Prueba normalidad residuos.

Gráfica (6): Histograma distribución residuos.



Gráfica (7): Residuos vs ajustes.

Gráfica (8): Residuos vs orden.

Se puede apreciar que la varianza es constante y hay independencia de los datos.

Comparaciones:

Nos interesa saber qué comisión genera un efecto distinto a los demás para ello se emplea comparaciones múltiples por el método de tukey por ende se obtienen los valores promedios para poder realizar las comparaciones

Medias

COMISIÓN	N	Media
1	16	71,567
2	16	71,85

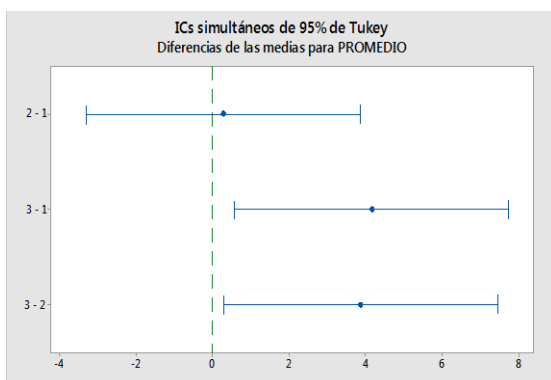
3	16	75,72
---	----	-------

Tabla (5): Valores de las medias.

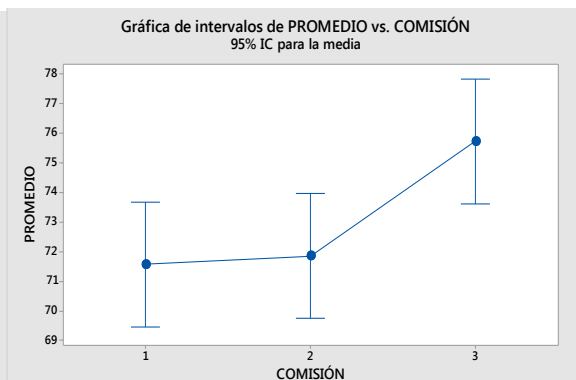
Pruebas simultáneas de Tukey para diferencias de las medias

Diferencia de comisiones	Diferencia de las medias	EE de diferencia	IC de 95%	Valor T	Valor p ajustado
2 - 1	0,29	1,48	(-3,30; 3,87)	0,19	0,979
3 - 1	4,15	1,48	(0,57; 7,74)	2,81	0,020
3 - 2	3,87	1,48	(0,28; 7,45)	2,62	0,032

Tabla (6): Cuadro de comparaciones.



Gráfica (9): Intervalos de confianza.



Gráfica (10): Promedio vs comisión.

Con la prueba de TUKEY se observa como la variable comisión genera distintos efectos en la nota definitiva, esto debido a que, al realizar las restas correspondientes entre sus diferentes niveles, no se encuentra el cero en dos de ellas, lo que sugiere que poblacionalmente no tienen el mismo promedio de nota.

CONCLUSION

- El diseño y análisis de experimentos, es una herramienta estadística robusta que puede utilizarse bajo cualquier escenario donde se quiera observar el efecto que puedan tener ciertas variables explicativas sobre una variable respuesta.
- Por medio de la presente investigación se demostró estadísticamente utilizando el análisis de diseño factorial que las diferentes comisiones presentes en la materia de Análisis Matemático I, generan variabilidad en los promedios de los estudiantes ya que la comisión tres se destaca, esto podría deberse a diversos factores que no se tuvieron en cuenta a la hora de estudiar el factor comisión.
- El género de la persona matriculada en cada una de las comisiones no es una variable que genere significancia para la nota final de la materia.

- Para las carreras estudiadas en esta investigación, se encontró que éstas no generan variabilidad significativa en los promedios de los estudiantes.

BIBLIOGRAFÍA

Robert O. Kuehl. "Principios estadísticos para el diseño y análisis de investigaciones". Capítulo 6 "Diseño Factorial"

Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye. "Probabilidad y estadística para ingeniería y ciencias". Capítulo 14 "Experimentos factoriales (dos o más factores)."

Douglas C. Montgomery." Probabilidad y estadísticas para ingeniería y administración". Capítulo 13 "Diseño de experimentos con varios factores."