



Universidad Centroamericana José Simeón Cañas

Maestría en Estadística Aplicada a la Investigación

Exámen Parcial I – Diseños Experimentales

Ramírez Solís Luis Edmundo

de Carné 00048203

Caso I

El impacto de los costos de operación sobre el porcentaje de cierre de negocios (3.5/10.0)

Una compañía quiere validar la suposición de su CFO, quien ha indicado que *“El porcentaje de negocios cerrados es inversamente proporcional a los costos de operación de la sucursal”*.

Para validar este supuesto, el equipo de Analítica ejecutó un Diseño Experimental, en el cual se seleccionaron aleatoriamente 24 sucursales y las monitorearon durante un periodo de 6 meses, la operación de estas sucursales se ejecutó en condiciones similares para el periodo de estudio. Los datos obtenidos se muestran en la Tabla 1.

Tabla 1. Costos de Operación y Porcentaje de negocios cerrados en promedio por mes.

Ubicación sucursal	Costo de operación (dólares)	Costo discretizado (dólares)	Porcentaje de negocios cerrados (%)	Orden de toma de datos
Parla	8,864	8000-10500	26.00%	24
Fuenlabrada	9,391	8000-10500	26.80%	21
Leganés	9,877	8000-10500	22.50%	4
Móstoles	9,977	8000-10500	21.10%	3
Arganda	10,052	8000-10500	36.50%	7
Torrejón	10,369	8000-10500	31.00%	8
Getafe	10,555	10500-12000	19.80%	10
Coslada	10,736	10500-12000	17.00%	15
Pinto	10,846	10500-12000	26.60%	5
Alcorcón	10,888	10500-12000	25.80%	18
Alcalá de Henarés	10,942	10500-12000	27.80%	20
Collado	11,913	10500-12000	18.70%	13
Colmenar Viejo	12,587	12000-15000	22.50%	17

Arroyomolinos	12,740	12000-15000	24.70%	12
S. Sebastian de los Reyes	13,041	12000-15000	17.20%	11
S. Lorenzo del Escorial	13,189	12000-15000	20.60%	2
Rivas	13,903	12000-15000	11.90%	23
Aluvia	14,680	12000-15000	13.20%	9
Alcobendas	16,256	Más de 15000	19.00%	14
Tres Cantos	17,562	Más de 15000	11.20%	22
Torrelodones	18,812	Más de 15000	8.10%	1
Boadilla	19,368	Más de 15000	13.00%	6
Majadahonda	19,477	Más de 15000	16.00%	19
Pozuelo	22,050	Más de 15000	13.00%	16

1. ¿Cuál es su conclusión en relación con lo indicado por el CFO con los datos proporcionados? **(2.0/10.0)**

Creando una categoría de agrupación a través de una tabla pivote con los costos de operación discretizados.

Nivel	Etiquetas de fila	Cuenta de Porcentaje de negocios	Suma de Costo de operación
A	8000-10500	6	58530
B	10500-12000	6	65880
C	12000-15000	6	80140
D	Más de 15000	6	113525
	Total general	24	318075

Creándose 4 niveles luego de la discretización.

Se crea una nueva tabla con los niveles agrupados reflejados en la Tabla 1.1.

Ubicación sucursal	Costo de operación (dólares)	Costo discretizado (dólares)	Porcentaje de negocios cerrados (%)	Nivel	Orden de toma de datos
Parla	8,864	8000-10500	26.00%	A	24
Fuenlabrada	9,391	8000-10500	26.80%	A	21
Leganés	9,877	8000-10500	22.50%	A	4
Móstoles	9,977	8000-10500	21.10%	A	3
Arganda	10,052	8000-10500	36.50%	A	7

Torrejón	10,369	8000-10500	31.00%	A	8
Getafe	10,555	10500-12000	19.80%	B	10
Coslada	10,736	10500-12000	17.00%	B	15
Pinto	10,846	10500-12000	26.60%	B	5
Alcorcón	10,888	10500-12000	25.80%	B	18
Alcalá de Henarés	10,942	10500-12000	27.80%	B	20
Collado	11,913	10500-12000	18.70%	B	13
Colmenar Viejo	12,587	12000-15000	22.50%	C	17
Arroyomolinos	12,740	12000-15000	24.70%	C	12
S. Sebastian de los Reyes	13,041	12000-15000	17.20%	C	11
S. Lorenzo del Escorial	13,189	12000-15000	20.60%	C	2
Rivas	13,903	12000-15000	11.90%	C	23
Aluvia	14,680	12000-15000	13.20%	C	9
Alcobendas	16,256	Más de 15000	19.00%	D	14
Tres Cantos	17,562	Más de 15000	11.20%	D	22
Torrelorones	18,812	Más de 15000	8.10%	D	1
Boadilla	19,368	Más de 15000	13.00%	D	6
Majadahonda	19,477	Más de 15000	16.00%	D	19
Pozuelo	22,050	Más de 15000	13.00%	D	16

Tabla 1.1. Luego de realizar la agrupación.

Se procede a realizar el análisis en R. (Código en ANEXO1)

ANALISIS DE ANOVA:

Planteamiento de Hipótesis:

H_0 : No existe diferencias significativas entre los grupos de medidas

$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$

H_1 : Al menos un par de medias son estadísticamente diferentes

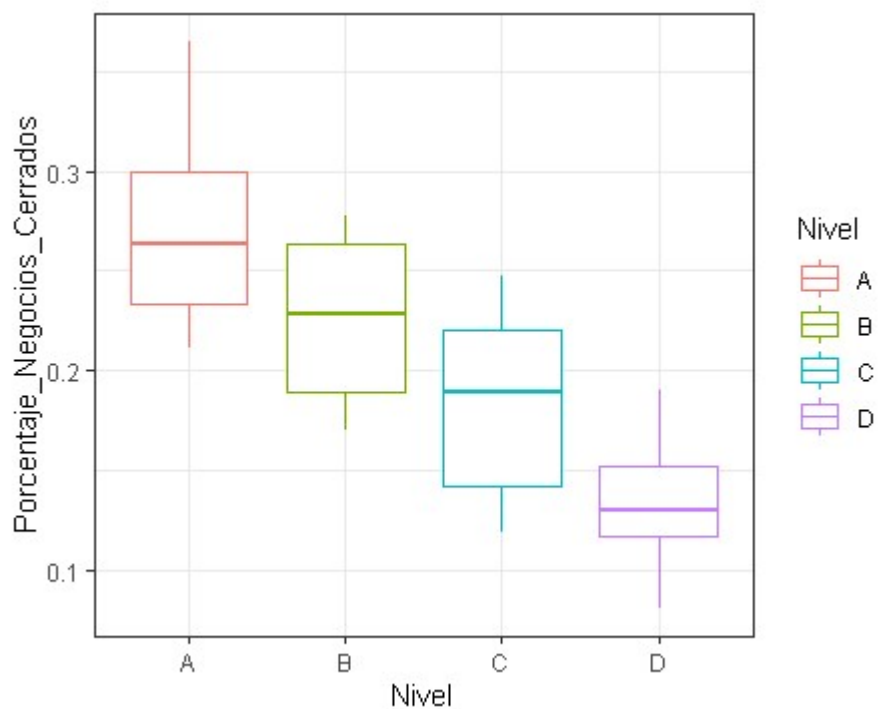
$\mu_i \neq \mu_j$

```
> model_parameters(anova)
```

Parameter	Sum_Squares	df	Mean_Square	F	p
Nivel	0.06	3	0.02	8.98	< .001
Residuals	0.05	20	2.36e-03		

Para un nivel de significancia $\alpha = 0.05$ el p-valor es menor, por lo que se rechaza H_0 y se acepta que al menos un par de medias son estadísticamente diferentes.

Diagrama de cajas y Bigotes resultante en el análisis de R.



La variable Nivel explica en un 57% la variación en los negocios cerrados. La variación se encuentra en un intervalo de confianza al 95% y va desde [27%-100%]. Como la variación nos da arriba del 50% se considera que posee influye de manera significativa en el modelo.

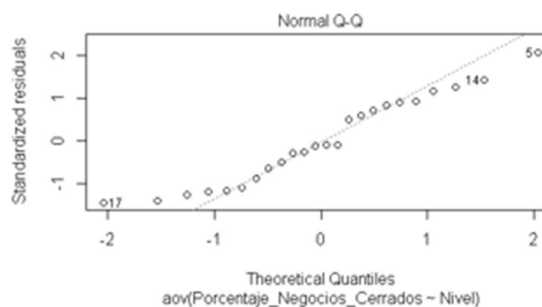
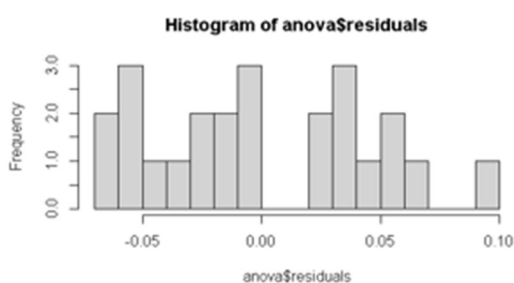
```
> #EXPLICA LA VARIACIÓN Y DA UN INTERVALO DE CONFIANZA
> eta_squared(anova, partial = FALSE)
# Effect Size for ANOVA (Type I)
```

Parameter	Eta2	95% CI
Nivel	0.57	[0.27, 1.00]

COMPROBACIÓN DE LOS SUPUESTOS DEL MODELO:

a) NORMALIDAD

La distribución a través de los gráficos de normalidad “aparenta” una existencia de normalidad tanto en el histograma como en el Gráfico Q-Q.



Para la comprobación estadística de la normalidad se ocupará las pruebas de hipótesis no paramétricas.

H_0 : Los datos provienen de una distribución normal con media cero y varianza desconocida

H_1 : Los datos no provienen de una distribución normal con media cero y varianza desconocida

$$H_0: \varepsilon_{ij} \approx N(0, \sigma)$$

$$H_1: \varepsilon_{ij} \not\approx N(0, \sigma)$$

b) Prueba K-S.

Por medio de “R” se realiza la prueba de Kolmogorov-Smirnov a dos colas.

```
> #PRUEBA K-S ("greater") COLA SUPERIOR
> #PRUEBA K-S
> ks.test(anova$residuals, pnorm, mean(anova$residuals), sd(anova$residuals), alternative = c("greater"))

One-sample Kolmogorov-Smirnov test

data:  anova$residuals
D+ = 0.11702, p-value = 0.5182
alternative hypothesis: the CDF of x lies above the null hypothesis
```

Para un nivel de significancia $\alpha = 0.05$ el p-valor es mayor, por lo que se rechaza H_0 y aceptamos la hipótesis alternativa. Para confirmar realizamos la prueba de Anderson – Darling.

c) Prueba ANDERSON-DARLING.

H_0 : Los datos provienen de una distribución normal con media cero y varianza desconocida

H_1 : Los datos no provienen de una distribución normal con media cero y varianza desconocida

```
> #PRUEBA ANDERSON-DARLING
> #p-valor 0.05 > 0.2908
> ad.test(anova$residuals)

Anderson-Darling normality test

data:  anova$residuals
A = 0.42559, p-value = 0.2908
```

Para un nivel de significancia $\alpha = 0.05$ el p-valor es mayor, por lo que se rechaza H_0 y aceptamos la hipótesis alternativa. Confirmando que los datos no provienen de una distribución normal con media cero y varianza desconocida.

2. Con los datos brindados no discretizados, elabore un modelo de regresión lineal simple: **(1.5/10.0)**

A través del programa Minitab Versión 20 se estimó la ecuación de Regresión.

Ecuación de regresión

Porcentaje de negocios cerrados = $0,3859 - 0,000014$ Costo de operación (dólares)

A través de “R” Se realizó el Modelo de Regresión Calculando el Intercepto y la Variable Dependiente.

```
> anova2$coefficients
      (Intercept) Costo_Operación 
3.858847e-01    -1.371133e-05
```

a) ¿Cuál es su conclusión con relación a lo indicado por el CFO?

De acuerdo a los datos obtenidos en el modelo y el haber comprobado los supuestos, se posee la suficiente evidencia estadística para indicar que la afirmación del CFO sobre el impacto de los costos de operación posee la suficiente evidencia estadística para indicar que el porcentaje de cierre de los negocios si influye de manera inversamente proporcional a los costos de operación de la sucursal que se monitorearon durante el periodo de 6 meses en el estudio realizado.

b) Para una sucursal con costos de operación promedio mensuales de 15,000 dólares ¿Cuál es el valor pronosticado de porcentaje promedio de negocios cerrados con el modelo de regresión y con el diseño experimental realizado en el punto a.? **(Recordatorio: El valor esperado de una observación en el Diseño Experimental está dado por el promedio de las observaciones de ese grupo)**

Ecuación de Regresión a través del programa Minitab 20.

Ecuación de regresión

Porcentaje de negocios cerrados = $0,3859 - 0,000014$ Costo de operación (dólares)

Se configuró el promedio de predicción mensual de Costo de Operación a \$15,000

Configuración

Variable	Valor de configuración
Costo de operación (dólares)	15000

El valor de Predicción resultante es del 18.02% para un Costo de operación de \$15,000 con un Nivel de confianza al 95% e intervalo [15.77%-20.26%]

Predicción

Ajuste	EE de ajuste	IC de 95%	IP de 95%
0,180215	0,0108257	(0,157764; 0,202666)	(0,0784257; 0,282004)

c) Si debiese seleccionar uno de los dos valores esperados, ¿Cuál seleccionaría? Justifique.

Analizando los valores en los 4 grupos que se realizaron anteriormente y evaluando una predicción en el modelo de regresión quedaría:

Nivel	Etiquetas de fila	Valor	Ec. Reg
A	8000-10500	9000	0.262486
B	10500-12000	10501	0.241905
C	12000-15000	12500	0.214497
D	Más de 15000	15001	0.180206

PREDICCIÓN PARA GRUPO A

Para el grupo “A” el valor de Predicción resultante es del 26.24% para un Costo de operación de \$9000 con un Nivel de confianza al 95%

Predicción

Ajuste	EE de ajuste	IC de 95%	IP de 95%
0,262483	0,0149715	(0,231434; 0,293532)	(0,158459; 0,366507)

PREDICCIÓN PARA GRUPO B

Para el grupo “B” el valor de Predicción resultante es del 24.19% para un Costo de operación de \$10,501 con un Nivel de confianza al 95%

Predicción

Ajuste	EE de ajuste	IC de 95%	IP de 95%
0,241902	0,0122214	(0,216556; 0,267248)	(0,139436; 0,344368)

PREDICCIÓN PARA GRUPO C

Para el grupo “B” el valor de Predicción resultante es del 24.19% para un Costo de operación de \$12,500 con un Nivel de confianza al 95%

Predicción

Ajuste	EE de ajuste	IC de 95%	IP de 95%
0,214493	0,0099763	(0,193803; 0,235183)	(0,113078; 0,315908)

PREDICCIÓN PARA GRUPO D

CONCLUSIÓN

Con un mejor valor de predicción a seleccionar en promedio para el grupo "D", cercano al valor en costos promedio mensuales resultantes de \$15,000, con un 18.02% en sus costos de operación.

Predicción

Ajuste	EE de ajuste	IC de 95%	IP de 95%
0,180201	0,0108268	(0,157747; 0,202654)	(0,0784114; 0,281990)

Caso II

El nivel educativo y el nivel socioeconómico (6.5/10.0)

Un informe elaborado por la Fundación BBVA indica que:

1. **Los niveles educativos de los jefes de familia impactan significativamente en los ingresos familiares.** Textualmente, el informe indica *"el jefe de familia con estudios universitarios tiene una familia cuyos ingresos aumentan de manera significativa contra los ingresos percibidos por una familia cuyo jefe tiene estudios de bachillerato; esta diferencia se incrementa a medida se especializa en la carrera seleccionada."*
2. **"En general, el nivel de pobreza de una familia condiciona el número de miembros presentes en esta"**

Con los datos brindados:

1. Para todo el conjunto de datos del departamento asignado (San Salvador), categorice a las familias de acuerdo con el índice de pobreza monetaria, el cual utiliza como parámetro el valor de la Canasta Básica Alimentaria (CBA). Este indicador categoriza a las familias en los siguientes niveles de pobreza.
(1.0/10.0)

Nivel de pobreza	Condición
Extrema	Ingreso per cápita por miembro por debajo del CBA
Relativa	Ingreso per cápita por miembro por debajo del CBA ampliado (dos veces el CBA)
No pobreza	Ingreso per cápita por miembro por arriba del CBA ampliado (dos veces el CBA)

Los datos del CBA por año para El Salvador se muestran a continuación:

Año	Area	CBA
2009	Rural	\$27.86
2009	Urbana	\$44.33
2010	Rural	\$27.80
2010	Urbana	\$45.12
2011	Rural	\$33.93
2011	Urbana	\$49.08
2012	Rural	\$31.28
2012	Urbana	\$46.83
2013	Rural	\$29.36
2013	Urbana	\$46.77
2014	Rural	\$30.73
2014	Urbana	\$49.53
2015	Rural	\$34.23
2015	Urbana	\$53.85
2016	Rural	\$33.45
2016	Urbana	\$53.63
2017	Rural	\$32.73
2017	Urbana	\$53.08

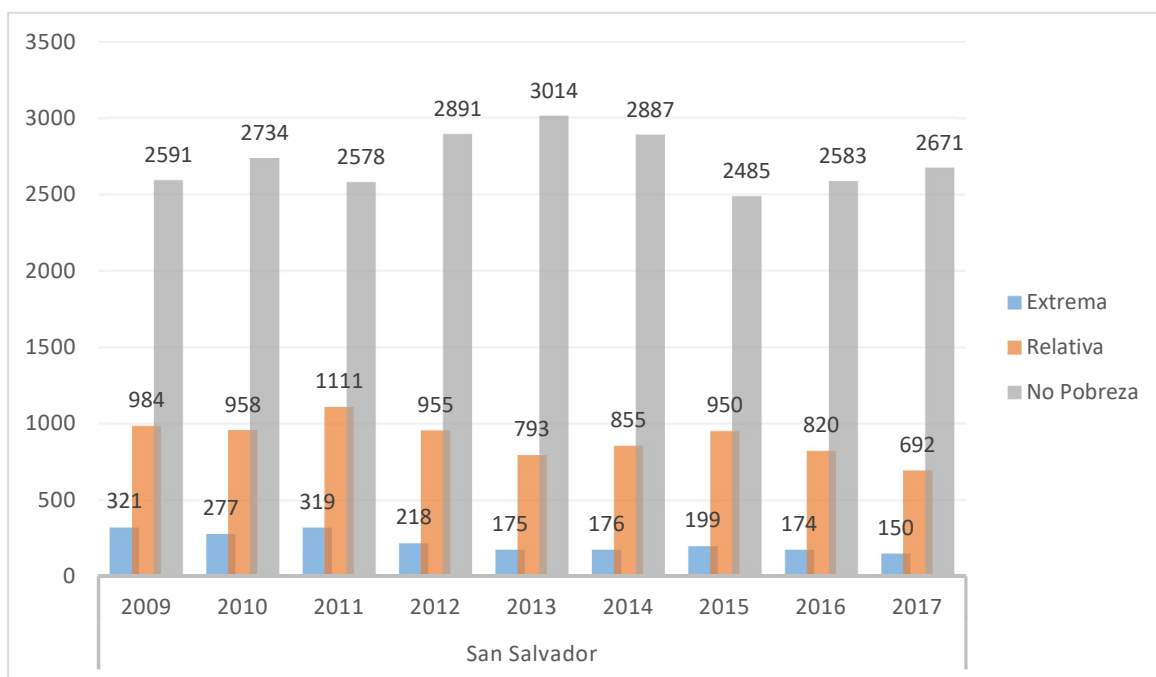
Por ejemplo, una familia del área urbana en 2016 con 4 miembros y un ingreso per cápita total de \$375.00 (\$93.75 por cada miembro) estaría en la categoría de “Pobreza Relativa”, ya que su ingreso está entre un CBA (\$53.63) y debajo del CBA ampliado (2 veces CBA \$107.26)

2. Construya una gráfica (la que considere conveniente) que indique cómo se distribuyen los niveles de pobreza por año. **(1.0/10.0)**

Tabla pivote construida en Excel con los niveles de pobreza con los datos para la Gráfica.

Cuenta de NivelPobreza		Etiquetas de columna			
Etiquetas de fila		Extrema	Relativa	No Pobreza	Total general
San Salvador		2009	8118	24434	34561
2009		321	984	2591	3896
2010		277	958	2734	3969
2011		319	1111	2578	4008
2012		218	955	2891	4064
2013		175	793	3014	3982
2014		176	855	2887	3918
2015		199	950	2485	3634
2016		174	820	2583	3577
2017		150	692	2671	3513
Total general		2009	8118	24434	34561

Gráfica 1. Niveles de Pobreza (Extrema, Relativa y No Pobreza para San Salvador Periodo 2009-2017.
Fuente de datos EHPM).



3. Con los resultados de la EHPM del año 2016, proponga dos diseños experimentales, cada uno con el objetivo de validar las afirmaciones del informe. **(Nota: Para la segunda afirmación del informe, utilice como factor la variable que construyó con la CBA en el punto 1 del caso) (3.0/10.0)**

✓ DISEÑO EXPERIMENTAL 1

“Los niveles educativos de los jefes de familia impactan significativamente en los ingresos familiares”.

1	Unidad Experimental	<i>Las Familias del Departamento de San Salvador en el año 2016</i>
2	Factores y Niveles (Independientes y de fondo)	
	<i>Independientes</i>	
	<i>a) Nivel Académico del jefe del hogar</i>	
		<i>0: No asistió a centro educacional</i>
		<i>1-3 Grados para Kindergarder</i>
		<i>4-12 Educación Básica</i>
		<i>13-15 Educación Bachillerato</i>
		<i>16-20 Educación Universitaria</i>
		<i>22-25 Educación Postgrado</i>
	<i>b) Área</i>	
		<i>Urbana</i>
		<i>Rural</i>
	<i>De fondo</i>	
		<i>Vulnerabilidad al crimen organizado</i>
3	Variable de respuesta	
		<i>¿Cómo se medirá el ingreso por remesas?(mensual, anual, promedio de un periodo)</i>
		<i>¿Es recomendable incluir todos los tipos de remesas recibidos?</i>

DISEÑO COMPLETO DEL EXPERIMENTO-1

Tratamiento	Área	Nivel Académico del Jefe del hogar
1	Rural	No asistió a centro educacional
2	Rural	Grados para Kindergarder
3	Rural	Educación Básica
4	Rural	Educación Bachillerato
5	Rural	Educación Universitaria
6	Rural	Educación Postgrado
7	Urbano	No asistió a centro educacional
8	Urbano	Grados para Kindergarder
9	Urbano	Educación Básica
10	Urbano	Educación Bachillerato
11	Urbano	Educación Universitaria
12	Urbano	Educación Postgrado

Un total de 12 tratamientos

✓ DISEÑO EXPERIMENTAL 2

"En general, el nivel de pobreza de una familia condiciona el número de miembros presentes en esta"

1	Unidad Experimental	<i>Familias del departamento de San Salvador en el año 2016</i>
2	Factores y Niveles (Independientes y de fondo)	
	<i>Independientes</i>	
	<i>a) Nivel de Pobreza</i>	
		<i>Extrema</i>
		<i>Relativa</i>
		<i>No Pobreza</i>
	<i>b) Área</i>	
		<i>Urbana</i>
		<i>Rural</i>
	<i>De fondo</i>	<i>La violencia, El trabajo formal e informal</i>
3	Variable de respuesta	
		<i>Número de miembros presentes en la familia que conforman el hogar.</i>

DISEÑO COMPLETO DEL EXPERIMENTO-2

Tratamiento	Área	Nivel de Pobreza
1	Rural	Extrema
2	Rural	Relativa
3	Rural	No Pobreza
4	Urbana	Extrema
5	Urbana	Relativa
6	Urbana	No Pobreza

Un total de 6 tratamientos

4. Con los resultados de la EHPM del año 2016, simule con los datos brindados la ejecución de **uno de los dos experimentos propuestos** en el punto 3 del caso. Utilice al menos 15 individuos por nivel del factor. **(1.5/10.0)**

ANEXO1. Código- Ejercicio1

```
#####  
# Luis Edmundo Ramírez #  
#####  
  
#####  
# Carga de Librerías####  
#####  
library(openxlsx) # Para abrir documentos de Excel  
library(tidyverse) # ggplot2, para gráficos  
library(moments) # Estadísticos varios  
library(nortest) # Test de normalidad no paramétricos  
library(lmtest) # Test de normalidad no paramétricos  
library(parameters) # Análisis de parámetros ANOVA  
library(effectsize)  
library(lsr) # Efectos  
library(agricolae) # Test Multimedias  
library(DescTools) # Test Multimedias  
library(pwr2) #Potencia  
library(car)  
library(gridExtra)  
  
#####  
# Carga de datos #####  
#####  
data.df <- read.xlsx(xlsxFile ="Data/Caso 1P.xlsx",sheet = "Datos")
```

```

data.df

#####Visualización de los datos
View(data.df)
head(data.df)

#####Cambio de nombre las columnas
colnames(data.df)
colnames(data.df) <-
c("Ubicación", "Costo_Operación", "Costo_Discretizado", "Porcentaje_Negocios_Ce
rrados", "Nivel", "Orden")

#?str: Verificación del dataframe 24obs. y 6 variables.
str(data.df)

#Ordeno los datos
data.df <- data.df[order(data.df$Orden),]

#####
# Exploración de datos #
#####
#Creación de tabla cruzada de la combinación de niveles de la variable
"Ambiente" del dataframe.
table(data.df$Ubicación)
#MEDIA
aggregate(Porcentaje_Negocios_Cerrados ~ Nivel, data = data.df, FUN = mean)
#DESVIACIÓN ESTANDAR
aggregate(Porcentaje_Negocios_Cerrados ~ Nivel, data = data.df, FUN = sd)

ggplot(data = data.df, aes(x = Nivel, y = Porcentaje_Negocios_Cerrados,
color = Nivel)) +
  geom_boxplot() + theme_bw()

#ggplot: Función para graficar
#theme_bw(): Controla visualización del gráfico
#geom_boxplot(): Propiedad de Diagrama de cajas y bigotes
?geom_boxplot()

#####
##### ANOVA #####
#####
#DECLARACIÓN DE VARIABLE
anova <- aov(Porcentaje_Negocios_Cerrados ~ Nivel, data = data.df)
anova

```

```

anova2 <- aov(Porcentaje_Negocios_Cerrados ~ Costo_Operación, data =
data.df)
anova2

#TABLA ANOVA
summary(anova)

#MODELOS DEL PARÁMETRO DE LA ANOVA
model_parameters(anova)

#EXPLICA LA VARIACIÓN Y DA UN INTERVALO DE CONFIANZA
eta_squared(anova, partial = FALSE)

#VARIANZA PARCIAL EXPLICADA POR EL FACTOR SIN INTERVALO DE CONFIANZA
etaSquared(anova)

#####
##### SUPUESTOS #####
#####

#####
### 1.NORMALIDAD#####
#####

# MÉTODO GRÁFICO
#HISTOGRAMA DE ANOVA DE RESIDUALES
hist(anova$residuals)

#SE PUEDE INDICAR CUÁNTOS GRUPOS SE DESEA CON EL COMANDO BREAK
hist(anova$residuals, breaks = 20)

#GRÁFICA DE LOS PUNTOS (GRÁFICA Q-Q)
plot(anova, which = 2)

#GRÁFICA DE LOS RESIDUALES SIN LA LÍNEA
qqnorm(anova$residuals)

#GRÁFICA DE LOS RESIDUALES CON LA LÍNEA
qqline(anova$residuals)

#####
##### ESTADÍSTICOS #####

```



```
#####

#COEFICIENTE DE ASIMETRÍA DE LOS RESIDUALES DE LA ANOVA
#(debe estar entre -1 Y 1)
skewness(anova$residuals)

#CURTOSIS
moments::kurtosis(anova$residuals)

#EXCESO DE CURTOSIS
moments::kurtosis(anova$residuals) - 3

#####
# PRUEBA DE HIPÓTESIS NO PARAMÉTRICAS####
#####

#PRUEBA K-S ("greater") COLA SUPERIOR
#PRUEBA K-S
ks.test(anova$residuals, pnorm, mean(anova$residuals), sd(anova$residuals),
alternative = c("greater"))

#PRUEBA SHAPIRO-WILK
#p-valor 0.05 < 0.2497
shapiro.test(anova$residuals)

#PRUEBA ANDERSON-DARLING
#p-valor 0.05 < 0.2908
ad.test(anova$residuals)

#####
### 2.INDEPENDENCIA#####
#####
#GRÁFICO DE RESIDUALES DE ANOVA
plot(anova$residuals)

#TEST DURBIN WATSON (ANOVA)
#DW = 2.0075, p-value = 0.5436
dwtest(anova)

#TEST DURBIN WATSON (Autocorrelación)
durbinWatsonTest(anova)

#?bgtest: BREUCH-GODFREY TEST
```

```

#bgtest(anova,order = 2)
#LM test = 0.08751, df = 2, p-value = 0.9572
#bgtest(anova,order = 1)
#LM test = 0.031485, df = 1, p-value = 0.8592

#Auto-and-Cross-Covariance and Correlation Function Estimation
acf(anova$residuals, ylim=c(-1,1))

#####
### 5.COMPARACIÓN DE TRATAMIENTOS#####
#####

#par(mfrow=c(1,1))

#####
## Fisher###
## -----###
#(anova, "ambiente", desplegarlo en consola=true)
LSD.test(anova, "Nivel",console=T)
#Gráfico
plot(LSD.test(anova, "Nivel",console=T))

#####
## TUKEY####
## -----###
TukeyHSD(anova)
#Gráfico
plot(TukeyHSD(anova))

#Tukey HSD Test Con grupos
HSD.test(anova, "Nivel",console=T)
#Gráfico
plot(HSD.test(anova, "Nivel",console=T))

#####
## DUNCAN###
## -----###
duncan.test(anova, "Nivel",console=T)
#Gráfico
plot(duncan.test(anova, "Nivel",console=T))

#####
## NEWMAN###

```

```

## -----###
SNK.test(anova, "Nivel", console = T)
#Gráfico
plot(SNK.test(anova, "Nivel", console = T))

#####
## DUNNET### (un grupo de control)
## -----###
DunnettTest(x=data.df$Porcentaje_Negocios_Cerrados, g=factor(data.df$Nivel))
DunnettTest(x=data.df$Porcentaje_Negocios_Cerrados, g=factor(data.df$Nivel),
control = "B")

#####
# Modelo de Regresión #
#####

# Parametros del modelo
# -----
anova2$coefficients
modelo1=lm(Porcentaje_Negocios_Cerrados ~ Costo_Operación, data = data.df,
na.action=na.exclude)
summary(modelo1)

# Valores estimados
# -----
anova2$fitted.values
par(mfrow=c(2,2))
data.df['Estimado'] <- anova2$fitted.values

view(data.df)

plot1 <- ggplot(data = data.df, aes(x = Costo_Operación, y =
Porcentaje_Negocios_Cerrados, color = Costo_Operación)) +
  geom_point() +
  geom_text(label=data.df$Porcentaje_Negocios_Cerrados,nudge_x= 0.5, nudge_y=
0.5)

plot2 <- ggplot(data = data.df, aes(x = Costo_Operación, y = Estimado, color
= Costo_Operación)) +
  geom_point() + geom_text(label=data.df$Estimado,nudge_x= 0.5, nudge_y=
0.5)

grid.arrange(plot1, plot2, ncol=2)

```

ANEXO2. Código- Ejercicio2

```
#####  
# Luis Edmundo Ramírez #  
#####  
  
#####  
# Carga de Librerías####  
#####  
library(openxlsx) # Para abrir documentos de Excel  
library(tidyverse) # ggplot2, para gráficos  
library(moments) # Estadísticos varios  
library(nortest) # Test de normalidad no paramétricos  
library(lmtest) # Test de normalidad no paramétricos  
library(parameters) # Análisis de parámetros ANOVA  
library(effectsize)  
library(lsr) # Efectos  
library(agricolae) # Test Multimedias  
library(DescTools) # Test Multimedias  
library(pwr2) #Potencia  
library(car)  
library(gridExtra)  
#####  
# Carga de datos #####  
#####  
data.df <- read.xlsx(xlsxFile ="Data/Caso2SS.xlsx",sheet = "Datos")  
data.df  
  
#####Visualización de los datos  
#data.df['Aleatorio'] <-sample(1:10000,191789,replace=T)  
#View(data.df)  
#data.df <- data.df[order(data.df$Aleatorio),]  
#head(data.df)  
#?str: Verificación del dataframe 24obs. y 6 variables.  
str(data.df)  
  
# 2. Muestreo  
  
#Creación de números Aleatorios  
set.seed(12345)  
#install.packages('sampling')  
#library(sampling)  
  
estratos <- strata(data.df, stratanames = c('NivelPobreza'), size =  
c(15,15,15), method = "srswor")  
estratos
```

```

data.df.muestra <- getdata(data.df, estratos)
View(data.df.muestra)
table(data.df.muestra$NivelPobreza)

#Ordenamos los datos
rows <- sample(nrow(data.df.muestra))
rows
data.df.muestra <- data.df.muestra[rows, ]
View(data.df.muestra)

#####
# Exploración de datos #
#####

table(data.df.muestra$NivelPobreza)
aggregate(Miembros ~ NivelPobreza, data = data.df.muestra, FUN = mean)
aggregate(Miembros ~ NivelPobreza, data = data.df.muestra, FUN = sd)

ggplot(data = data.df.muestra, aes(x = NivelPobreza, y = Miembros, color =
NivelPobreza)) +
  geom_boxplot() + theme_bw()

# Convirtiendo a factor (los levels se pueden modificar dependiendo de los
resultados obtenidos)
data.df.muestra$NivelPobreza <- parse_factor(data.df.muestra$NivelPobreza,
      levels = c('Extrema',
                  'Relativa',
                  'No Pobreza'
                ))

#####
##### ANOVA #####
#####

anova <- aov(Miembros ~ NivelPobreza, data = data.df.muestra)
summary(anova)
model_parameters(anova)

anova2 <- aov(Miembros ~ NivelPobreza, data = data.df.muestra)
anova2

eta_squared(anova, partial = FALSE)
etaSquared(anova)

```

```
#####
##### SUPUESTOS #####
#####
#####
### 1.NORMALIDAD#####
#####

#
par(mfrow=c(1,1))

# MÉTODO GRÁFICO
#HISTOGRAMA DE ANOVA DE RESIDUALES
hist(anova$residuals)

#SE PUEDE INDICAR CUÁNTOS GRUPOS SE DESEA CON EL COMANDO BREAK
hist(anova$residuals, breaks = 20)

#GRÁFICA DE LOS PUNTOS (GRÁFICA Q-Q)
plot(anova, which = 2)

#GRÁFICA DE LOS RESIDUALES SIN LA LÍNEA
qqnorm(anova$residuals)

#GRÁFICA DE LOS RESIDUALES CON LA LÍNEA
qqline(anova$residuals)

#####
# PRUEBA DE HIPÓTESIS NO PARAMÉTRICAS####
#####

#PRUEBA K-S ("greater") COLA SUPERIOR
#PRUEBA K-S
ks.test(anova$residuals, pnorm, mean(anova$residuals), sd(anova$residuals),
alternative = c("greater"))

#PRUEBA SHAPIRO-WILK
#p-valor 0.05 < 0.2497
shapiro.test(anova$residuals)

#PRUEBA ANDERSON-DARLING
#p-valor 0.05 > 0.2908
ad.test(anova$residuals)
```

```
#####
### COMPARACIÓN DE TRATAMIENTOS#####
#####

#####
## Fisher###
## -----###
#(anova, "ambiente", desplegarlo en consola=true)
LSD.test(anova, "NivelPobreza",console=T)
#Gráfico
plot(LSD.test(anova, "NivelPobreza",console=T))

#####
## TUKEY###
## -----###
TukeyHSD(anova)
#Gráfico
plot(TukeyHSD(anova))

#Tukey HSD Test Con grupos
HSD.test(anova, "Nivel",console=T)
#Gráfico
plot(HSD.test(anova, "Nivel",console=T))

#####
## DUNCAN###
## -----###
duncan.test(anova, "Nivel",console=T)
#Gráfico
plot(duncan.test(anova, "Nivel",console=T))

#####
## NEWMAN###
## -----###
SNK.test(anova, "Nivel", console = T)
#Gráfico
plot(SNK.test(anova, "Nivel", console = T))

#####
# Modelo de Regresión #
#####
```

```
# Parametros del modelo
# -----
anova2$coefficients
modelo1=lm(Miembros ~ NivelPobreza, data = data.df.muestra,
na.action=na.exclude)
summary(modelo1)
```

[ENLACE GITHUB.](#)

https://github.com/00048203/EXAMEN_PARCIAL_DISENOS_EXPERIMENTALES