UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

POSTGRADO DE INGENIERIA

ESPECIALIZACION EN INVESTIGACION CIENTIFICA

TAREA No.5

Santiago Enrique Jacinto Cumátz 999003402

Oliver Eliu Valdez Ortega 999003506

Marvin Estuardo Vásquez Pérez 999003487

15 octubre del 2020

Dirección github:

https://github.com/MEvp22/EIC999003487/blob/master/Script%20tarea%205%20grupal\_EIC.R

# Ejercicio 1:

Se afirma que los automóviles recorren en promedio más de **20000** kilómetros por año, pero usted cree que el promedio es en realidad **menor**. Para probar tal afirmación se pide a una muestra de **100** propietarios de automóviles seleccionada de manera aleatoria que lleven un registro de los kilómetros que recorren. Los resultados de la muestra aleatoria indican un promedio de **19500** kilómetros y una desviación estándar de **3900** kilómetros. Utilice prueba de hipótesis y use una significancia **del 3%.**

1. Definir la hipótesis nula

* h0: Mu => 20000

1. Definir la hipótesis alternativa

* h1: Mu < 20000

1. Definir el nivel de significancia y alfa

* alfa <-0.03

1. Definir valores (n, media, sd, var, etc)

* n<- 100
* media<- 19500
* sd<- 3900
* mu<- 20000

1. Obtener valor del estadístico de prueba

* -1.282051

1. Obtener valor de la distribución

* -1.880794

1. Validar la hipótesis

* FALSE
* se acepta la hipótesis nula, donde se define la media como mayor a 20000
* no hay suficiente evidencia estadística para inferir que la hipótesis nula es falsa

# Ejercicio 2:

Para verificar si el proceso de llenado de bolsas de café está operando con la variabilidad permitida se toman aleatoriamente muestras de **10** cada hora. El proceso de llenado está bajo control si presenta una varianza de **40 o menos**. Al tomar una muestra al azar se obtiene una variabilidad de **27**. Utilice un nivel de significancia del **5%.**

1. Definir la hipótesis nula

* h0: sigma=>40

1. Definir la hipótesis alternativa

* h1: sigma ≤ 40

1. Definir el nivel de significancia y alfa

* alfa <-0.05

1. Definir valores (n, media, sd, var, etc)

* n= 10
* s2 <- 27
* sigma<- 40
* alfa <- 0.05

1. Obtener valor del estadístico de prueba

* 6.075

1. Obtener valor de la distribución

* 3.325113

1. Validar la hipótesis

* FALSE
* se acepta la h0 que indica que el valor de sigma es mayor o igual a 40
* hay suficiente evidencia estadística para inferir que la hipótesis nula no es falsa

# Ejercicio 3:

Para una muestra de 30 elementos con una media de 35 y una desviación estándar de 3.5, calcule el intervalo de confianza con un nivel de significancia del 5%

\*En este ejercicio por ser una muestra de 30 elementos se considera utilizar la fórmula de t, pero los resultados comparados con qnorm salen invertidos por lo que se dejan los dos, siendo el primer valor de qnorm y el segundo y en azul de t.

1. Definir variables

* media <- 35
* n <-30
* desv <- 3.5
* alfa <- 0.05/2

1. Defina normal ajustada

* 1.959964 , -1.699127

1. Defina Error

* 0.6390097, 0.6390097

1. Defina limites superior e inferior del intervalo de confianza

* Límite superior: 36.25244, 33.91424
* Límite inferior: 33.74756, 36.08576

# Ejercicio 4:

Para una muestra de 5000 elementos con una media de 500 y una desviación estándar de 100, calcule el intervalo de confianza con un nivel de significancia del 10%

1. Definir variables

* media <- 500
* n <-5000
* desv <- 100
* alfa <- 0.1/2

1. Defina normal ajustada

* 1.644854

1. Defina Error

* 1.414214

1. Defina limites superior e inferior del intervalo de confianza

* Límite superior: 502.3262
* Límite inferior: 497.6738

# Ejercicio 5:

Utilizando en conjunto de datos “Trees” realice una regresión lineal entre sus variables y defina los siguientes enunciados

1. Definir regresión entre Girth (y) y Height(x)

lm.trees <- lm(Girth~Height, data=trees)

summary (lm.trees)

plot (trees$Girth ~trees$Height)

abline (lm.trees)

1. Indique la formula generada

Call:

lm(formula = Girth ~ Height, data = trees)

1. Indique su intercepto, coeficiente de correlación R2 y R ajustado

* Intercepto: -6.18839
* Coeficiente de correlación R2: 0.2697
* R ajustado: 0.2445

1. Definir regresión entre Girth y volumen (x)

lm.trees <- lm(Girth~Volume, data=trees)

summary (lm.trees)

plot (trees$Girth ~trees$Volume)

abline (lm.trees)

1. Indique la formula generada

Call:

lm(formula = Girth ~ Volume, data = trees)

1. Indique su intercepto, coeficiente de correlación R2 y R ajustado

* Intercepto: 7.677857
* Coeficiente de correlación R2: 0.9353
* R ajustado: 0.9331

1. ¿Cuál de las dos regresiones se ajusta mejor al modelo?

La regresión que mejor se ajusta es la de Girth y volumen ya que presenta un coeficiente de correlación R2 de 0.9353, lo cual indica que los datos están menos dispersos. Comparado con el coeficiente de correlación R2 de Grith y Height que es de 0.2697.