

## Pruebas de hipótesis paramétricas

### Casos para una población

### Casos para dos poblaciones

### Códigos R

# Taller1 Pruebas de hipótesis paramétricas

## Modulo 4- Unidad 4.2

dgonzalez

## Pruebas de hipótesis paramétricas

Procedimientos para realizar una prueba de hipótesis estadística en caso de realizar el procedimiento paso a paso.

- Determinar  $H_0$  y  $H_a$
- Establecer el  $EdeP$
- Calcular el Estadístico de Prueba  $EdeP$ , con los valores obtenidos en la(s) muestra(s)
- Construir la Región de Rechazo,  $RdeR$
- Concluir a partir de la Regla de decisión 1,  $RdeD1$

En caso de realizarlo con ayuda de un programa estadístico

- Determinar las hipótesis nula  $H_0$  y hipótesis alterna  $H_a$
- Determinar la prueba a realizar
- Procesar la información
- Concluir a partir de los resultados obtenidos teniendo en cuenta la  $RdeD2$  o  $RdeD3$

---

## Casos para una población

### Pruebas para una media

#### Problema 1

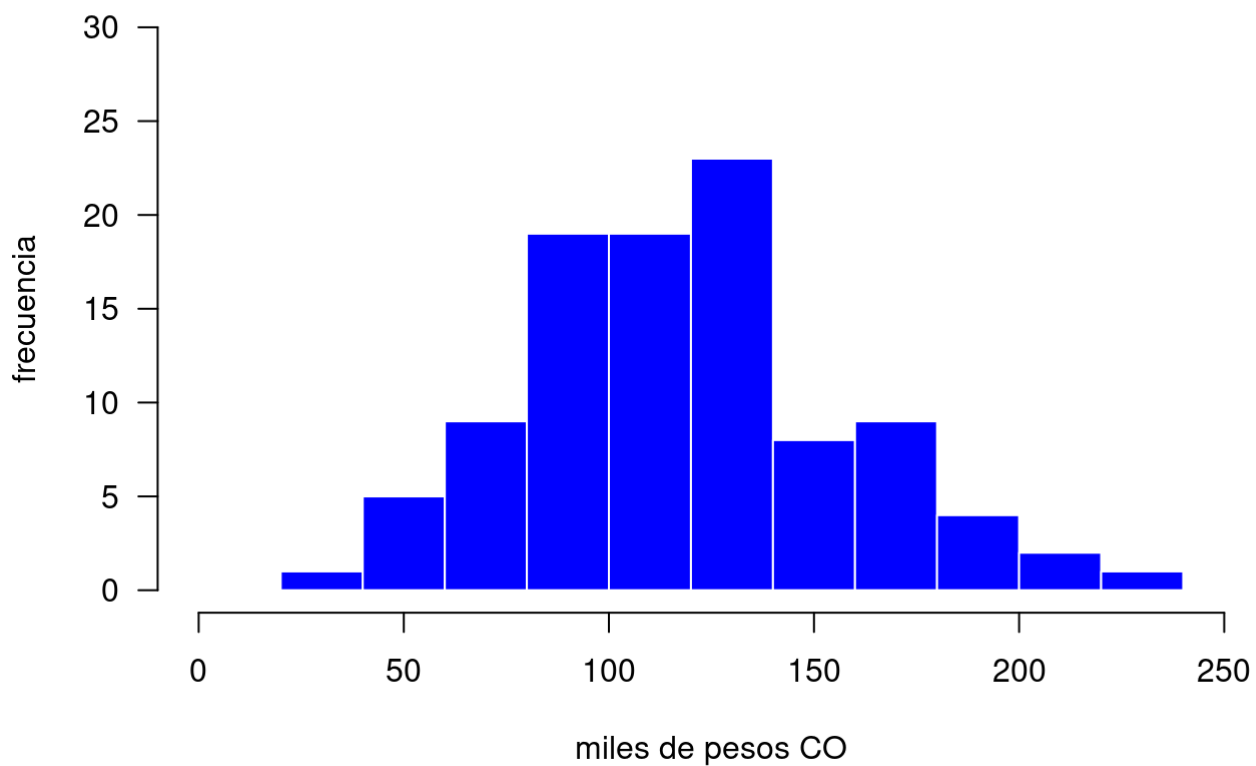
Un informe indica que el precio medio del pasaje de avión entre Cali y San Andrés es de \$ 120 mil pesos y su desviación estándar de \$ 40 mil pesos. Se toma una muestra que se registra a continuación.

¿Se puede aceptar, con un nivel de significación igual a 0.1, las afirmaciones acerca de la media y de la desviación?

```
# data
x1=c(165.1,91.8,128.3,160.3,103.6,153.2,50.7,150.3,160.8,38.1,98.8,88.7,123.1
',
    106.5,79.5,84.1,216.1,139.8,98.4,127.1,103.0,83.7,158.8,126.1,113.4,61.3
',
    79.8,121.8,89.9,97.4,134.0,141.2,154.3,122.4,107.6,137.7,121.0,187.7,56.
8,
    79.5,183.8,114.7,80.2,97.3,113.8,92.4,123.5,138.4,114.7,48.8,178.1,106.0
',
    103.1,100.8,180.7,113.1,121.4,163.0,51.4,135.5,104.5,202.1,171.0,100.7,
    56.4,99.7,93.0,118.3,110.4,136.9,112.3,69.8,138.3,87.2,121.8,118.6,174.5
',
    77.7,158.5,85.3,71.4,63.0,140.9,126.9,194.2,136.7,98.8,222.5,131.3,175.0
',
    127.7,88.5,104.5,144.7,73.0,97.2,131.0,120.5,166.6,96.0)
```

```
hist(x1,
     main= "Precio del pasaje aéreo Cali- San Andres Islas", # titulo
     xlab = "miles de pesos C0", # etiqueta x
     ylab= "frecuencia",        # etiqueta y
     las=1,                     # etiquetas horizontales
     ylim=c(0,30),             # rango de y
     xlim=c(0,250),            # rango de x
     col="blue",                # color del histograma
     border="white"             # color de linea separadora
     )
```

## Precio del pasaje aéreo Cali- San Andres Islas



```
shapiro.test(x1) # prueba de normalidad
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  x1  
## W = 0.9853, p-value = 0.3337
```

Como el valor-p correspondiente a la prueba de normalidad es grande, no existe suficiente evidencia en la muestra que permita rechazar la  $H_0$ . Asumimos que la variable tiene distribución normal.

```
# prueba t para una media  
# Ho:  $\mu = 120$   
# Ha:  $\mu \neq 120$   
t.test(x1,                                # variable  
       alternative = "two.sided",          # desigualdad de Ha  
       mu = 120,                          # valor a comparar  
       conf.level = 0.95)                 # nivel de confianza / nivel de significancia
```

```
##
##  One Sample t-test
##
## data:  x1
## t = -0.39343, df = 99, p-value = 0.6949
## alternative hypothesis: true mean is not equal to 120
## 95 percent confidence interval:
##  110.9228 126.0732
## sample estimates:
## mean of x
##  118.498
```

Como el valor-p es grande no existe evidencia que permita rechazar la hipótesis  $H_0$ , asímimos que la media es igual a 120. Por tal razón la información suministrada concuerda con los resultados obtenidos

## Problema 2

La duración de las bombillas de 100 W que fabrica una empresa la cual garantiza que funciona al menos 800 horas. Se escoge al azar una muestra de los tiempos de 25 bombillas de un lote y se les mide su tiempo de duración y esta información se reporta a continuación

- Con un nivel de significación de 0.05, ¿habría que rechazar el lote por no cumplir la garantía? Responder usando intervalos de confianza, región de rechazo y valor p para la prueba de hipótesis planteada.
- ¿Es la variabilidad de los tiempos inferior a 10 horas? Responder usando intervalos de confianza, región de rechazo y valor p para la prueba de hipótesis planteada.
- Si la media y la varianza de los tiempos es realmente 700 horas y 121 horas<sup>2</sup> respectivamente ¿En qué porcentaje de las muestras de 25 tiempos de bombillas la suma de los tiempos de duración es superior a 17500 horas?
- Graficar la información para respaldar las conclusiones de los literales anteriores del presente ejercicio.

```
x2=c(751,749,749,745,740,743,737,749,734,755,742,758,754,750,772,741, 750,76
9,
     732,753,761,750,774,751)
```

## Problema 3

Se está calibrando una balanza al pesar una patrón de prueba de 1000 g, 60 veces, arroja los siguientes datos. Realice un contraste de hipótesis para determinar si la balanza se encuentra calibrada. Utilice un  $\alpha = 0.05$ .

```
x3=c(1000.62, 1001.80, 999.10, 1003.21, 997.56, 1001.77, 1001.11, 1000.69,
999.93, 1001.52, 998.72, 1002.08, 1001.56, 999.52, 997.76, 1002.36,
998.70, 998.16, 1002.29, 1002.10, 1001.33, 1002.86, 1006.61, 997.82,
999.64, 1000.90, 999.66, 1000.37, 1000.12, 996.68, 1005.97, 1003.08,
1000.79, 1001.82, 1002.54, 999.01, 998.95, 999.08, 1001.48, 998.71,
998.10, 1001.10, 999.96, 1000.27, 1004.46, 1003.77, 1000.04, 1000.54,
999.17, 996.17, 1001.85, 998.46, 999.91, 1001.07, 1004.19, 1002.49,
998.99, 1000.48, 1000.42, 999.32)
```

#### Problema 4

En una sucursal bancaria se desarrolla un proceso para mejorar el servicio de sus clientes durante el horario del medio día (12:00 a 14:00) El tiempo de espera en la fila durante este tiempo se registra durante un día seleccionado de manera aleatoria a 15 clientes con los siguientes resultados:

Existe evidencia en los datos que permitan afirmar que un cliente espera en la fila para ser atendido un tiempo promedio inferior a 5 minutos?

```
x4=c(4.21, 5.55, 3.02, 5.13, 4.77, 2.34, 5.42, 4.50, 6.10, 3.80, 5.12, 6.46, 6.19, 3.79, 3.54)
```

## Pruebas para una proporción

#### Problema 5

Una empresa al seleccionar su personal lo somete a un curso de entrenamiento. Por su experiencia se conoce que el 76% de los aspirantes aprueban el curso. Los encargados del proceso efectúan cambios al programa de entrenamiento, al cual se inscriben 40 aspirantes de los cuales lo aprueban 24. ¿Podría afirmarse que los cambios realizados en el programa de entrenamiento reducen la selección?

```
n=40
x=24
```

## Pruebas para una varianza

#### Problema 6

Un fabricante de baterías para celulares afirma que el tiempo de duración de sus baterías se distribuyen aproximadamente normal con una desviación estándar de 9 horas. Un comprador antes de realizar una compra de este producto le solicita una muestra de 10 baterías para verificar la información que le está dando. Con este propósito son enviadas al laboratorio y se obtienen los siguientes datos

Es cierta la información suministrada por el fabricante?

```
x6=c(11.1, 15.6, 11.1, 7.5, 7.9, 14.7, 6.3, 8.5, 8.0, 7.6)
```

## Casos para dos poblaciones

### Pruebas para diferencia de medias - grupos pareados

#### Problema 7

El director de un gimnasio requiere determinar si contrata o no a un instructor para su campaña de reducción de peso. Para tomar la decisión decide tomar un grupo de 16 personas que acuden normalmente para darle una oportunidad a un aspirante al cargo. Los siguientes datos corresponden a los resultados obtenidos antes y después de un mes de estar en el programa.

Utilice un nivel de significancia  $\alpha = 0.10$ , para determinar si contrata o no al aspirante.

```
pant=c(104.5, 89, 84.5, 106, 90, 96, 79, 90, 85, 76.5, 91.5, 82.5, 100.5, 89.5, 121.5, 72)
pdes=c(98, 85.5, 85, 103.5, 88.5, 95, 79.5, 90, 82, 76, 89.5, 81, 99.5, 86.5, 115.5, 70)
```

### Pruebas para diferencia de medias - grupos independiente

### Pruebas para igualdad de varianzas

#### Problema 8

Una empresa desarrolla un curso de entrenamiento para sus empleados, formando dos grupos y aplicando dos metodologías diferentes de entrenamiento con el fin de poder evaluar que método produce los mejores resultados. El primer grupo (g1) esta conformado por 36 empleados y el segundo grupo (g2) por 40 empleados. Se puede afirmar que el método aplicado al segundo grupo produce mejores resultados que el aplicado al primer grupo? ¿Que supuestos debe tener en cuenta?

```
g1=c(6.8, 6.1, 5.8, 5.9, 5.8, 6.4, 5.7, 6.0, 5.9, 6.4, 6.0, 5.7, 6.5, 6.5, 6.0,
      5.9, 5.7, 5.8, 5.9, 5.8, 6.0, 6.0, 5.8, 5.7, 6.1, 5.9, 5.2, 6.3, 5.4, 6.5,
      5.5, 5.9, 7.0, 6.4, 5.1, 6.3)
g2=c(8.8, 8.5, 8.4, 8.5, 7.6, 8.7, 8.0, 7.9, 8.2, 8.0, 7.8, 8.6, 8.5, 7.9, 8.5,
      8.3, 8.4, 8.2, 8.3, 7.9, 8.2, 7.7, 7.8, 7.7, 8.1, 8.0, 8.3, 8.2, 8.1, 8.3,
      8.1, 8.8, 7.7, 9.1, 7.6, 8.4, 8.2, 8.3, 8.1, 8.7)
```

### Pruebas para diferencia de proporciones

## Problema 9

Una encuesta realizada a 100 usuarios de una tarjeta de crédito seleccionados aleatoriamente, 57 dijeron que sabían que empleando su tarjeta podían ganar millas de viajero. Después de una campaña publicitaria para difundir este beneficio, se realizó una encuesta independiente entre 200 usuarios de la tarjeta de crédito y 135 informaron que conocían el beneficio. ¿Se puede concluir que el conocimiento de este beneficio aumento después de la campaña publicitaria?

```
n1=100
x1=57

n2=200
x2=135
```

## Problema 10

Un estudio realizado por MasterCard revelo que 131 de las 468 mujeres que efectuaron compras en almacén lo hicieron utilizando la tarjeta de crédito propia del almacén, mientras que 57 de 237 hombres utilizaron la misma tarjeta para sus compras en el almacén. ¿ Existe evidencia suficiente en los datos que permita concluir que la proporción de mujeres es mayor que la proporción de hombres que utilizan la tarjeta de crédito propia del almacén para realizar sus compras?

```
x1=131
n1=468

x2=57
n2=237
```

# Codigos R

```
# Codigos R

# t.test(x, y = NULL,
#       alternative = c("two.sided", "less", "greater"),
#       mu = 0, paired = FALSE, var.equal = FALSE,
#       conf.level = 0.95, ...)

# prop.test(x, n, p = NULL,
#          alternative = c("two.sided", "less", "greater"),
#          conf.level = 0.95, correct = TRUE)

# var.test(x, y, ratio = 1,
#          alternative = c("two.sided", "less", "greater"),
#          conf.level = 0.95, ...)

# rmarkdown::render("Puebas_de_hipotesis1.Rmd", encoding = "UTF-8")
```