Trabajo 2

Jhonatan Smith Garcia Muñoz 1039705595 Natalia Valencia Nuñez 1006502467

Enero 2021

Preparacion de los datos

Se utiliza la cedula de jhonatan smith garcia.

```
base <- read.table(file.choose(), header=T)
```

```
genera <- function(cedula){
set.seed(cedula)
data <- base[sample(1:2000,120),]
data
}
df = genera(1039705595)
```

```
df[,4] <- as.factor(df[,4])
df[,5] <- as.factor(df[,5])
```

1.

Se tiene la creencia de que la ESTATURA media de los estudiantes es inferior a 170 cms. ¿Es esto cierto? Justifique su respuesta.

Para identificar que tan cierto es el problema, se empieza por analizar la media de los datos y ver siquiera si tiene sentido dicha hipotesis.

```
attach(df)
mean(ESTATURA)
```

```
## [1] 169.7375
```

AL analizar la media de los datos, se tiene el resultado anterior.

¿Los datos se distribuyen de manera normal?

```
shapiro.test(ESTATURA)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ESTATURA
## W = 0.98896, p-value = 0.4461
```

H_0 : Datos normales vs H_1 : Datos NO normales.

Dado el valor p, donde es mayor que un nivel de significancia dado, por tanto NO se rechaza la hipotesis nula y se concluye que los datos distribuyen normal.

Para realiazr dicho procedimiento, se plantea la prueba de hipotesis donde $H_0 : \mu = 170$ vs $H_1 : \mu \leq 170$

```
t.test(x = ESTATURA, alternative = "less", mu = 170, conf.level = 0.95 )
```

```
##
##  One Sample t-test
##
## data:  ESTATURA
## t = -0.32781, df = 119, p-value = 0.3718
```

```
## alternative hypothesis: true mean is less than 170
## 95 percent confidence interval:
##     -Inf 171.065
## sample estimates:
## mean of x
##  169.7375
```

Con esto, se tiene que el valor P segun la prueba dada es de 0.3718 y por tanto NO se rechaza la hipotesis nula y se concluye que en efecto la estatura de los estudiantes es de 170 y no menor. Con un nivel de confianza del 95%.

2 ¿Puede afirmarse que la MASA promedio de los Hombres es superior a la de las Mujeres? Justifique su respuesta

Para verificar la afirmacion, primero observemos la proporcion de hombres.

Siguiendo con el analisis anterior se debe comprobar normalidad en las masas para hombres y mujeres.

```r
mean(df$MASA[df$SEXO=="Hombre"]) # Media de los hombres
```

```
## [1] 64.07213
```

```r
mean(df$MASA[df$SEXO=="Mujer"]) # Media masa mujeres
```

```
## [1] 62.61695
```

En un principio, parece haber cierta diferencia de medias. Verificando normalidad.

```r
shapiro.test(df$MASA[df$SEXO=="Hombre"]) # Para hombres
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$MASA[df$SEXO == "Hombre"]
## W = 0.97245, p-value = 0.1841
```

Se concluye datos normales a un nivel de significancia del 0.95

```r
shapiro.test(df$MASA[df$SEXO=="Mujer"]) # Para mujeres
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$MASA[df$SEXO == "Mujer"]
## W = 0.96461, p-value = 0.08378
```

Se concluyen datos normales a un nivel de significancia del 0.95

Ahora, respecto a la pregunta problema:

```r
t.test(x = df$MASA[df$SEXO=="Hombre"], y = df$MASA[df$SEXO=="Mujer"], alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  df$MASA[df$SEXO == "Hombre"] and df$MASA[df$SEXO == "Mujer"]
## t = 0.69901, df = 117.99, p-value = 0.243
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -1.996148       Inf
## sample estimates:
## mean of x mean of y
##  64.07213  62.61695
```

Planteado el juego de hipotesis siguiente; donde mu1 es la media de los hombres y mu2 la media de las mujeres:

$H_1 : \mu_2 = \mu_1$

$H_0 : \mu_2 \leq \mu_1$

Y teniendo en cuenta el resultado del valor-p, no se rechaza la hipotesis nula y se concluye a un nivel de confianza del 95% que la masa promedio de los hombres es mayor a la de las mujeres.

```r
summary(SEXO)
```

```
## Hombre  Mujer
##     61     59
```

```r
# Por tanto se tienen segun estimacion puntual

p1<-61/120 # Porcentaje hombres
p1
```

```
## [1] 0.5083333
```

```r
p2<-59/120 # Porcentaje de mujeres
p2
```

```
## [1] 0.4916667
```

En ambos casos, la muestra es grande. Se realiza la siguiente prueba de hipotesis dado teorema del limite central.

$H_0 : \hat{p}_1 = 0.52$ vs $H_1 : \hat{p}_1 \geq 0.52$

Esto, segun la prueba de Wald:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

```r
z <- (p1 - 0.52) / sqrt(0.52 * (1 - 0.90) / 120)
z  # Para obtener el valor del estadístico
```

```
## [1] -0.5604485
```

Para obtener el valor-P de la prueba debemos tener en cuenta el sentido en la hipótesis alternativa, que es H1<0.52 por esa razón el valor-P será P(Z<z):

```r
pnorm(q=z, lower.tail=TRUE)
```

```
## [1] 0.2875868
```

Se tiene que el valor p es mayor que un nivel de significancia del 0.95 por tanto no se rechaza la hipotesis nula y por tanto no se puede concluir que la proporcion de hombres sea mayor al 52% con una confianza del 95%.

4 Un investigador afirma que la proporción de estudiantes que proviene de cada una de las 6 ciudades es la misma. ¿Es cierta esta afirmación?

```r
summary(CIUDAD)
```

```
##      ARMENIA BARRANQUILLA       BOGOTA         CALI    CARTAGENA     MEDELLIN
##           20           26           18           17           21           18
```

Al analizar las proporciones, se tiene lo siguiente:

```r
# Proporciones por orden de ciudad dado el summary anterior
Ar = 20/120
br = 26/120
bt = 18/120
cl = 17/120
ct <- 21/120
md <- 18/120
```

H0: Ar=br=..=md vs H1: Alguna de ellas diferente

```r
prop.test(x = c(20,26,18,17,21,18), n = c(120,120,120,120,120,120),  )
```

```
##
##  6-sample test for equality of proportions without continuity
##  correction
##
## data:  c(20, 26, 18, 17, 21, 18) out of c(120, 120, 120, 120, 120, 120)
```

```
## X-squared = 3.24, df = 5, p-value = 0.663
## alternative hypothesis: two.sided
## sample estimates:
##    prop 1    prop 2    prop 3    prop 4    prop 5    prop 6
## 0.1666667 0.2166667 0.1500000 0.1416667 0.1750000 0.1500000
```

Dado el p-valor No se rechaza la hipotesis nula y se concluye que las proporciones son iguales.