Comparación de medias

Dr. Marco Aurelio González Tagle

07 /08/ 2019

Índice

Ejercicio1

Un laboratorio de Estados Unidos preguntó a 24 personas con fobia a las arañas (Aracnofobia) si podrían participar en un experimento. El laboratorio dividio a los asistentes en 2 grupos. Solo a 12 personas se les pidio que jugaran con una tarantula por un momento y se midieron sus niveles de ansiedad. A las 12 personas restantes solo se les mostró una fotografía de una tarantula y sus valores de ansiedad fueron igualmente medidos.

Procedimiento general para la pruebas de t independientes

Ingresar los datos

```
Grupo <- gl(2, 12, labels = c("Fotografía", "Araña"))
Ansiedad <- c(30, 35, 45, 40, 50, 35, 55, 25, 30, 45, 40, 50, 40, 35, 50, 55,
              65, 55, 50, 35, 30, 50, 60, 39)
Datos <- data.frame(Grupo, Ansiedad)
head(Datos)
```

```
##        Grupo Ansiedad
## 1 Fotografía       30
## 2 Fotografía       35
## 3 Fotografía       45
## 4 Fotografía       40
## 5 Fotografía       50
## 6 Fotografía       35
```

Explorar los datos

Revise que sus datos posean las siguientes características: n =24, media grupo *Fotografía* = 40 y la media del grupo *Araña* = 47.

- Indique mediante una gráfica de **boxplot** las posibles diferencias que existen entre los grupos.

```
boxplot(Datos$Ansiedad ~ Datos$Grupo, col= "lightgreen", ylab="Nivel de ansiedad")
```

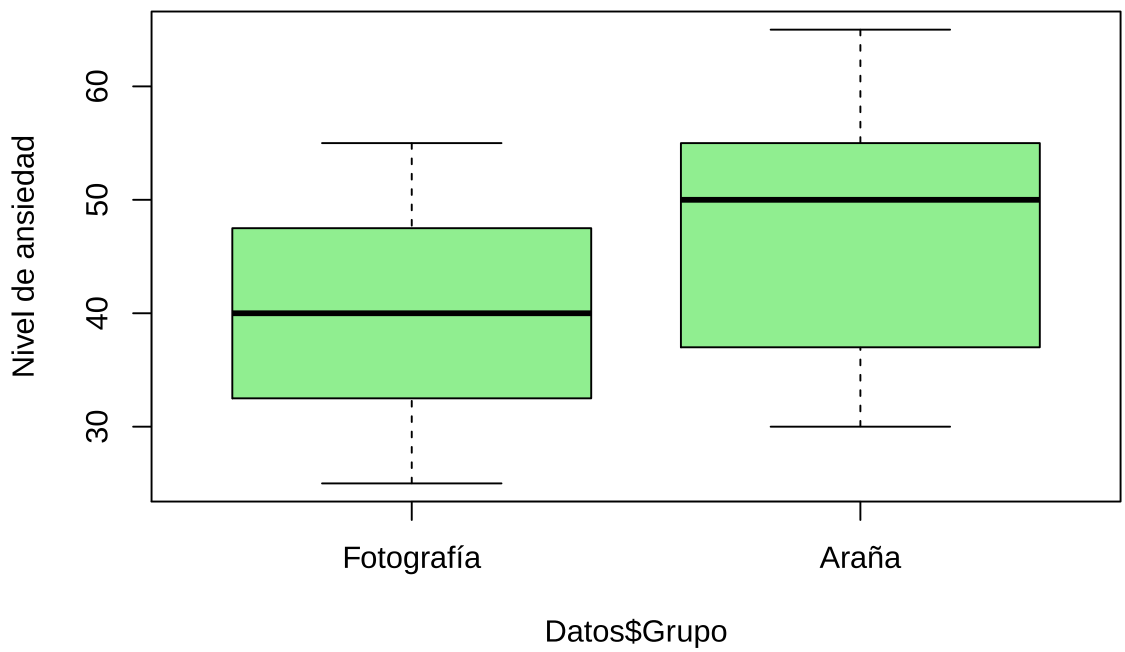

Figura 1: La gráfica muestra el nivel de ansiedad de los partcipantes en el experimento. El grupo de la izquierda solo se les mostró una fotografia (Fotografía) y al grupo de la derecha (Real) se les mostro una tarantula real.

- Describa la hipótesis nula y alternativa para este experimento

Ho = No existe una diferencia significativa entre los niveles de ansiedad de los 2 grupos participantes.

- ¿Cúales son las dos supuestos que deben cumplir los datos?
  - Normalidad aplicando la prueba de shapiro Wilkins
  - Homogenedidad de Varianzas

```
shapiro.test(Datos$Ansiedad)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Datos$Ansiedad
## W = 0.96282, p-value = 0.4977
```

```
bartlett.test(Datos$Ansiedad, Datos$Grupo)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Datos$Ansiedad and Datos$Grupo
## Bartlett's K-squared = 0.30702, df = 1, p-value = 0.5795
```

Una forma alternativa de correr los análisis anteriores es instalando y usando la libreria *pastecs*.

```
library(pastecs)
by(Datos$Ansiedad, Datos$Grupo, stat.desc, basic= FALSE, norm =TRUE)
```

```
## Datos$Grupo: Fotografía
##       median        mean     SE.mean CI.mean.0.95          var
##   40.0000000  40.0000000   2.6827168    5.9046200   86.3636364
##      std.dev    coef.var    skewness     skew.2SE     kurtosis
##    9.2932038   0.2323301   0.0000000    0.0000000   -1.3939289
##     kurt.2SE  normtest.W  normtest.p
##   -0.5656047   0.9650165   0.8522870
## --------------------------------------------------------
## Datos$Grupo: Araña
##        median          mean       SE.mean  CI.mean.0.95           var
##  50.000000000  47.000000000   3.183765638   7.007420922 121.636363636
##       std.dev      coef.var      skewness      skew.2SE      kurtosis
##  11.028887688   0.234657185  -0.005590699  -0.004386224  -1.459758279
##      kurt.2SE    normtest.W    normtest.p
##  -0.592315868   0.948872904   0.620569431
```

Aplique la prueba de t

```
gr.t <- t.test(Datos$Ansiedad  ~ Datos$Grupo, var.equal = TRUE)
gr.t
```

```
##
##  Two Sample t-test
##
## data:  Datos$Ansiedad by Datos$Grupo
## t = -1.6813, df = 22, p-value = 0.1068
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -15.634222   1.634222
## sample estimates:
## mean in group Fotografía      mean in group Araña
##                       40                       47
```

Determinar el efecto del tamaño de muestra

$$r = \sqrt{\frac{t^2}{t^2 + df}} \tag{1}$$

```
## [1] 0.34
```

Reportar los datos

En promedio, los participantes experimentaron niveles de ansiedades mayores en el grupo que obtuvieron una tarantula en sus manos (M = 47, SE = 3.18), que aquellos que solo se les mostro una fotografía (M = 40, SE = 2.68). Esta diferencia no fue significativa t (22) = -1.6813456, p = 0.05. Sin emabrgo el efecto de tamaño de muestra es mediano (r= 0.34).

Ejercicio 2

Una empresa de alimento para venados esta interesado en determinar si el peso neto medio del contenido de sus costales es de 80 kg como se anuncia en su producto. Digamos que un consumidor al paso del tiempo ha comprado y pesado de forma precisa el contenido neto de 44 costales de 80 kg seleccionados al azar. El consumidor reclama a la empresa que el contenido de sus costales ha sido menor porque sus datos no provienen de una distribución con media de 80 kg. Para investigar esta demanda, la empresa plantea una hipótesis usando el grado de confiabilidad de $\alpha = 0.05$.

La hipótesis nula será que no existen diferencias entre la media es igual a 80 kg, y la hipótesis alternativa es que la media observada es menor a 80 kg.

$$H_0 : \mu = 80 \tag{2}$$

$$H_0 : \mu < 80 \tag{3}$$

Procedimiento general para la pruebas de t una muestra

Ingresar los datos

```
costal <- c(87.7, 80.01, 77.28, 78.76, 81.52, 74.2, 80.71, 79.5, 77.87, 81.94, 80.7,
            82.32, 75.78, 80.19, 83.91, 79.4, 77.52, 77.62, 81.4, 74.89, 82.95,
            73.59, 77.92, 77.18, 79.83, 81.23, 79.28, 78.44, 79.01, 80.47, 76.23,
            78.89, 77.14, 69.94, 78.54, 79.7, 82.45, 77.29, 75.52, 77.21, 75.99,
            81.94, 80.41, 77.7)
```

Explorar los datos

La media y la desviación estándar deben ser estimados de la muestra

```
# Determinar el número de observaciones
n <- length(costal)
n
```

```
## [1] 44
```

```
# Determinar la media
costa.media <- mean(costal)
costa.media
```

```
## [1] 78.91068
```

```
# Desviación estándar
costa.sd <- sd(costal)
costa.sd
```

```
## [1] 3.056023
```

Necesitamos responder la siguiente pregunta: Dada la desviación estándar (3.056), ¿Cuál es la probabilidad de observar una media de la muestra (cuando n = 44) de 78.91 Kg o de menor cantidad si la media verdadera es de 80 kg.

La prueba de T de una muestra puede ser resulta con la siguiente formula:

$$T = \frac{\bar{x} - \mu_0}{(s/\sqrt{n})} \tag{4}$$

El denominador se le conoce como error estándar de la media (se).

```
costa.se <- costa.sd/ sqrt(n)
costa.se
```

```
## [1] 0.4607128
```

Entonces podemos calcular el valor de T

```
costa.T <- (costa.media - 80)/ costa.se
costa.T
```

```
## [1] -2.364419
```

Finalmente, el valor de p puede ser calculado.

```
pt(costa.T, df = n-1)
```

```
## [1] 0.01132175
```

Podemos concluir que si la H_0 fuera aceptada, existiría tan solo un poco mas del 1 % de probabilidad que se observe que los costales obtenidos por el comprador fueran de una media de $\bar{x}$ = 78.91 kg o menores. Puesto que el valor de *p-value* (0.0113217) es menor que el α =0.05, la empresa concluye que existe suficiente evidencia para rechazar la H_0 y aceptar la H_1, suguiriendo que el valor de μ es de hecho menor que 80 kg.

Procedimiento en R

El resultado de la prueba de t de una muestra puede ser verificado con la fórmula ya establecida en R

```
t.test(costal, mu= 80, alternative = "less")
```

```
##
##  One Sample t-test
##
```

```
## data:  costal
## t = -2.3644, df = 43, p-value = 0.01132
## alternative hypothesis: true mean is less than 80
## 95 percent confidence interval:
##      -Inf 79.68517
## sample estimates:
## mean of x
##  78.91068
```