Pruebas de hipótesis - Ejercicios

Ejercicio 1.

Descripción.

Para verificar si el proceso de llenado de bolsas de café con 500 gramos está operando correctamente se toma una muestra aleatoria de diez bolsas cada cuatro horas. Una de estas muestras está compuesta por las siguientes observaciones:

$$502, 501, 497, 491, 496, 501, 502, 500, 489, 490$$

Suponiendo que los datos son normales, ¿considera que el proceso de llenado es correcto? Use un nivel de significancia del 5%.

Solución.

1. **Planteamiento de las hipótesis:**

Según la información proporcionada, el parámetro a analizar es μ, es decir el promedio de contenido de café en las bolsas de muestra. Por lo tanto, las hipótesis que se pueden plantear son:

$$H_0 : \mu = 500 \text{ (El proceso de llenado es correcto)}$$
$$H_1 : \mu \neq 500 \text{ (El proceso de llenado no es correcto)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como los datos son normales, se desconoce la varianza de la población y la cantidad de datos no es muy grande, entonces el estadístico de prueba es:

$$t_{n-1} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Calculamos los datos que nos hacen falta a partir de la muestra proporcionada para luego sustituirlos en la ecuación anterior y asi obtener el valor de $t_{n-1} = t_9$

```r
data <- c(502, 501, 497, 491, 496, 501, 502, 500, 489, 490)
mu <- 500

x <- mean(data)
s <- sd(data)
n <- length(data)

t <- (x - mu) / (s / sqrt(n))
round(t, 2)
```

```
## [1] -1.89
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y tenemos una prueba bilateral, tenemos que calcular el valor de:

$t_{\frac{\alpha}{2},n-1} = t_{0.025,9}$, el cual es:

```r
round(qt(0.025, 9), 2)
```

```
## [1] -2.26
```

y de $t_{1-\frac{\alpha}{2},n-1} = t_{0.975,9}$, cuyo valor es:

```r
round(qt(0.975, 9), 2)
```

```
## [1] 2.26
```

Por lo tanto, la regla de decisión es rechazar H_0 si $t_9 < -2.26$ o si $t_9 > 2.26$

5. **Toma de decisión:**

Dado que el estadístico de prueba $t_9 = -1.89 \not< t_{0.025,9} = -2.26$ asi como $t_9 = -1.89 \not> t_{0.975,9} = 2.26$. Por lo tanto con un valor de significancia de 5% no se puede rechazar H_0 y se descarta H_1. Es decir, tomamos la decisión de que $H_0 : \mu = 500$, concluyendo que el proceso de llenado es correcto.

Podemos corroborar la decisión, usando la prueba p-valor de R. Donde podemos observar que el $p - value = 0.09$ es mayor que el nivel de significancia con el que teniamos que evaluar la prueba $\alpha = 0.05$, por lo cual no podemos rechazar H_0.

```r
t.test(x = data, y = NULL, mu = 500, alternative = "two.sided", conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  data
## t = -1.8948, df = 9, p-value = 0.09064
## alternative hypothesis: true mean is not equal to 500
## 95 percent confidence interval:
##  493.199 500.601
## sample estimates:
## mean of x
##     496.9
```

Ejercicio 2.

Descripción.

Se afirma que los automóviles recorren en promedio más de 20,000 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. ¿Estaría usted de acuerdo con la afirmación si la muestra aleatoria indicara un promedio de 19,500 kilómetros y una desviación estándar de 3,900 kilómetros? Utilice el *p-valor* en su conclusión y use una significancia del 3%.

Solución.

1. **Planteamiento de las hipótesis:**

Según la información proporcionada, el parámetro a analizar es μ, es decir el promedio de kilómetros que recorren por año los automóviles, por lo tanto las hipótesis a plantear son:

$$H_0 : \mu = 20,000 \text{ (Los automóviles recorren en promedio 20,000 kilómetros)}$$
$$H_1 : \mu > 20,000 \text{ (Los automóviles recorren en promedio más de 20,000 kilómetros)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.03$

3. **Cálculo del estadístico de prueba:**

Como de desconoce la varianza de la población pero la muestra es grande, entonces nuestro estadístico de prueba es:

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Sustituyendo los datos tenemos que el valor de Z (redondeado a 2 decimales) es:

```r
mu <- 20000
n <- 100
x <- 19500
s <- 3900

Z <- (x - mu) / (s / sqrt(n))
round(Z, 2)
```

```
## [1] -1.28
```

4. **Formulación de la regla de decisión:**

Como la prueba es unilateral, y $\alpha = 0.03$, tenemos que calcular $Z_{1-\alpha} = Z_{0.97}$, dando como resultado:

```r
round(qnorm(0.97), 2)
```

```
## [1] 1.88
```

Entonces, la regla de decisión es rechazar H_0 si $Z > 1.88$

5. **Toma de decisión:**

Dado que el estadístico de prueba $Z = -1.28 \not> Z_{0.97} = 1.88$, por lo tanto no podemos rechazar H_0.

Si calculamos el p-valor para corroborar la decisión, es decir $P(Z > -1.28) = 1 - P(Z < -1.28)$, redondeando a 2 decimales obtenemos:

```r
p_value <- 1 - pnorm(Z)
round(p_value, 2)
```

```
## [1] 0.9
```

Donde podemos ver que no se puede rechazar H_0 porque el valor de $\alpha = 0.03$ es menor que el $p-valor = 0.9$.

Por lo tanto con $\alpha = 0.03$, no se rechaza H_0 y se concluye que no tenemos evidencia significativa para afirmar que los automóviles recorren en promedio más de 20,000 kilómetros por año.

Ejercicio 3.

Descripción.

Un fabricante de un quitamanchas afirma que su producto quita el 90% de todas las manchas. Para poner a prueba esta afirmación se toman 200 camisetas manchadas de las cuales a solo 174 les desapareció la mancha. Prueba la afirmación del fabricante a un nivel $\alpha = 0.05$

Solución.

1. **Planteamiento de las hipótesis:**

Según la información proporcionada, el parámetro a analizar es **p**, es decir la proporción de manchas que puede quitar el producto del fabricante. Por lo tanto, las hipótesis a plantear son:

$$H_0 : p = 0.9 \text{ (La afirmación del fabricante es correcta)}$$
$$H_1 : p < 0.9 \text{ (La afirmación del fabricante no es correcta)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como tenemos una muestra grande (200 camisetas) y además el parámetro que queremos analizar es **p**, el estadístico de prueba es:

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

Al calcular el valor de Z con los datos proporcionados, obtenemos (redondeado a 2 decimales):

```
p_hat <- 174 / 200
p <- 0.90
n <- 200
Z <- (p_hat - p) / sqrt((p * (1-p)) / n)

round(Z, 2)
```

```
## [1] -1.41
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y que se tiene una prueba unilateral, tenemos que calcular el valor de $Z_\alpha = Z_{0.05}$, cuyo valor redondeado a 2 decimales es:

```
round(qnorm(0.05), 2)
```

```
## [1] -1.64
```

Entonces la regla de decisión es rechazar H_0 si $Z < -1.64$.

5. **Toma de decisión:**

Dado que el estadístico de prueba $Z = -1.41 \nless Z_{0.05} = -1.64$, por lo tanto no se rechaza H_0 y se concluye que con un valor $\alpha = 0.05$ no existe evidencia significativa para afirmar que el producto quitamanchas del fabricante quite menos del 90% de todas las manchas.

Ejercicio 4.

Descripción.

Un asadero de pollos asegura que el 90% de sus órdenes se entregan en menos de 10 minutos. En una muestra de 200 órdenes, 170 se entregaron dentro de ese lapso. ¿Puede concluirse en el nivel de significancia $\alpha = 0.05$, que menos del 90% de las órdenes se entregan en menos de 10 minutos?

Solución.

1. **Planteamiento de las hipótesis:**

Según la información proporcionada, el parámetro a analizar es **p**, es decir la proporción de ódenes de pollos que se pueden entregar en menos de 10 minutos. Por lo tanto, las hipótesis a plantear son:

$$H_0 : p = 0.9 \text{ (El 90\% de las órdenes se entregan en menos de 10 minutos)}$$
$$H_1 : p < 0.9 \text{ (Menos del 90\% de las órdenes se entregan en menos de 10 minutos)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como tenemos una muestra grande (200 órdenes) y ademas el parámetro que queremos analizar es **p**, el estadístico de prueba es:

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

Al calcular el valor de Z con los datos proporcionados, obtenemos (redondeado a 2 decimales):

```
p_hat <- 170 / 200
p <- 0.90
n <- 200
Z <- (p_hat - p) / sqrt((p * (1-p)) / n)

round(Z, 2)
```

```
## [1] -2.36
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y que se tiene una prueba unilateral, tenemos que calcular el valor de $Z_\alpha = Z_{0.05}$, cuyo valor redondeado a 2 decimales es:

```
round(qnorm(0.05), 2)
```

```
## [1] -1.64
```

Entonces la regla de decisión es rechazar H_0 si $Z < -1.64$.

5. **Toma de decisión:**

Dado que el estadístico de prueba $Z = -2.36 < Z_{0.05} = -1.64$, debemos rechazar H_0 y quedarnos con H_1.

Por lo tanto se concluye que con un valor $\alpha = 0.05$ tenemos evidencia suficiente para rechazar H_0 lo que significa que H_1 se cumple, es decir menos del 90% de las órdenes de pollos se entregan en menos de 10 minutos

Ejercicio 5.

Descripción.

Para verificar si el proceso de llenado de bolsas de café está operando con la variabilidad permitida se toma una muestra de diez bolsas cada cuatro horas. Una de estas muestra está compuesta por las siguientes observaciones:

$$502, 501, 497, 491, 496, 501, 502, 500, 489, 490$$

El proceso de llenado está bajo control si presenta una varianza de 40 o menos. ¿Está el proceso de llenado de bolsas bajo control? Use un nivel de significancia del 5%.

Solución.

1. **Planteamiento de las hipótesis:**

El parámetro de interés para este problema es la varianza. Por lo tanto las hipótesis que se pueden plantear son:

$$H_0 : \sigma^2 = 40 \text{ (El proceso de llenado de bolsas está bajo control)}$$
$$H_1 : \sigma^2 > 40 \text{ (El proceso de llenado de bolsas no está bajo control)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como el parámetro de interés del problema es la varianza, entonces el estadístico de de prueba es:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Calculamos los datos necesarios de la muestra, para despues calcular el valor de χ^2 con la formula anterior, obteniendo como resultado (redondeado a 2 decimales):

```
data <- c(502, 501, 497, 491, 496, 501, 502, 500, 489, 490)
S <- sd(data)
n <- length(data)
var <- 40

chisq <- ((n - 1) * S**2) / var
round(chisq, 2)
```

```
## [1] 6.02
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y que tenemos una prueba unilateral, tenemos que calcular el valor de $\chi^2_{1-\alpha} = \chi^2_{0.95}$, cuyo valor es (redondeado a 2 decimales):

```
round(qchisq(0.95, n-1), 2)
```

```
## [1] 16.92
```

Entonces la regla de decisión es rechazar H_0 si $\chi^2 > 16.92$

5. **Toma de decisión:**

Dado que el estadístico de prueba $\chi^2 = 6.02 \ngtr \chi^2_{0.95} = 16.92$ no se rechaza H_0

Podemos corroborar la decisión tomada, apoyandonos del cálculo del $p - valor$

$$P(\chi^2 > 6.02) = 1 - P(\chi^2 < 6.02)$$

```r
p_value <- 1 - pchisq(chisq, n-1)
round(p_value, 4)
```

```
## [1] 0.7377
```

Como el $p - valor = 0.7377$, lo cual es muy superior al nivel $\alpha = 0.05$, no tenemos evidencia significativa para rechazar H_0.

Por lo tanto podemos concluir que con un valor $\alpha = 0.05$ no tenemos evidencia significativa para afirmar que el proceso de llenado de bolsas no está bajo control.

Ejercicio 6.

Descripción.

El arsénico en agua potable es un posible riesgo para la salud. Un artículo reciente reportó concentraciones de arsénico en agua potable en partes por billón (ppb) para diez comunidades urbanas y diez comunidades rurales. Los datos son los siguientes;

$$Urbana : 3, 7, 25, 10, 15, 6, 12, 25, 15, 7$$

$$Rural : 48, 44, 40, 38, 33, 21, 20, 12, 1, 18$$

¿Considera que las varianzas de las concentraciones son iguales o diferentes? Usa $\alpha = 0.02$

Solución.

- **Planteamiento de las hipótesis:**

Considerando el subindice **u** como los datos de las comunidades urbanas y el subíndice **r** como los datos de las comunidades rurales. Entonces las hipótesis a plantear para este problema son:

$$H_0 : \frac{\sigma_u^2}{\sigma_r^2} = 1 \text{ (Las varianzas de las concentraciones son iguales para ambas zonas)}$$

$$H_1 : \frac{\sigma_u^2}{\sigma_r^2} \neq 1 \text{ (Las varianzas de las concentraciones no son iguales para ambas zonas)}$$

- **Método p-valor**

Con el apoyo de la función *var.test()* de R podemos calcular el $p - valor$ del cociente de las varianzas, obteniendo como resultado lo siguiente:

```r
data_u <- c(3, 7, 25, 10, 15, 6, 12, 25, 15, 7)
data_r <- c(48, 44, 40, 38, 33, 21, 20, 12, 1, 18)

var.test(x = data_u, y = data_r, ratio = 1, alternative = "two.sided", conf.level = 0.98)
```

```
##
##  F test to compare two variances
##
## data:  data_u and data_r
## F = 0.24735, num df = 9, denom df = 9, p-value = 0.04936
## alternative hypothesis: true ratio of variances is not equal to 1
## 98 percent confidence interval:
##  0.04622339 1.32358740
## sample estimates:
## ratio of variances
##          0.2473473
```

- **Toma de decisión**

Del resultado anterior podemos ver que el *p-valor* es 0.04936 lo cual es ligeramente superior al nivel $\alpha = 0.02$, por lo que no tenemos evidencia significativa para rechazar H_0.

Por lo tanto se concluye que con un nivel $\alpha = 0.02$ las varianzas de las concentraciones de arsénico en el agua potable para las comunidades urbanas y rurales son similares.

Ejercicio 7.

Descripción.

El arsénico en agua potable es un posible riesgo para la salud. Un artículo reciente reportó concentraciones de arsénico en agua potable en partes por billón (ppb) para diez comunidades urbanas y diez comunidades rurales. Los datos son los siguientes;

$$Urbana : 3, 7, 25, 10, 15, 6, 12, 25, 15, 7$$

$$Rural : 48, 44, 40, 38, 33, 21, 20, 12, 1, 18$$

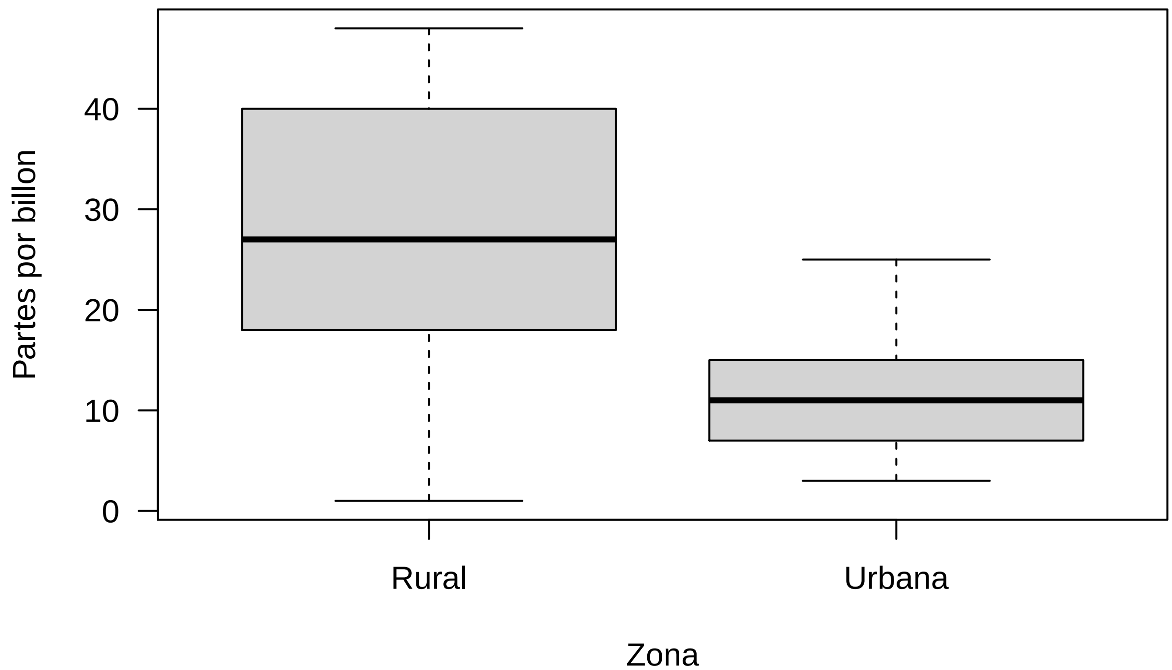
¿Existen diferencias entre las concentraciones de arsénico de la zona urbana y rural? Usar un nivel de significancia del 6%.

Solución.

Podemos primero construir un *boxplot* con los datos proporcionados para comparar las concentraciones de arsenico en PPB según la zona. A continuación podemos ver el comparativo.

```r
data_u <- c(3, 7, 25, 10, 15, 6, 12, 25, 15, 7)
data_r <- c(48, 44, 40, 38, 33, 21, 20, 12, 1, 18)

data <- data.frame(ppb = c(data_u, data_r), zone = rep(c("Urbana", "Rural"), each = 10))
boxplot(ppb ~ zone, data = data, las = 1, xlab = "Zona", ylab = "Partes por billon")
```



En la figura anterior podemos observar que si bien el mínimo y máximo de la zona Rural y la zona Urbana respectivamente translapan con su contraparte, los datos centrales de las cajas no se translapan.

- **Planteamiento de las hipótesis:**

Ademas en la figura anterior como se puede observar la parte central de la caja para la zona rural esta encima de la parte central de la caja para la zona urbana.

Representando como μ_1 y μ_2 el promedio de las concentraciones de arsenico en las comunidades rurales y urbanas respectivamente. Entonces las hipótesis a plantear son las siguientes:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (No existen diferencias entre las concentraciones de arsénico de la zona urbana y rural)}$$

$$H_1 : \mu_1 - \mu_2 \neq 0 \text{ (Existen diferencias entre las concentraciones de arsénico de la zona urbana y rural)}$$

- **Estadístico de prueba:**

Para este tipo de prueba, como las varianzas poblacionales son desconocidas y ademas la muestra no es tan grande, el estadístico de prueba a utilizar es **t**. Solo que no conocemos si las varianzas poblaciones son iguales o distintas pero como las muestras son de zonas diferentes podemos asumir que son distintas, por lo tanto nuestro estadístico de prueba sera:

$$t_k = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$k = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$$

- **Cálculo del p-valor con R:**

Para obtener el **p-valor** de una manera más facil nos podemos apoyar de la función *t.test()* de R, con la cual al pasarle los parámetros que ya conocemos se obtiene:

```
t.test(x = data_r, y = data_u, alternative = "two.sided",
       mu = 0, var.equal = FALSE, conf.level = 0.94)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_r and data_u
## t = 2.7669, df = 13.196, p-value = 0.01583
## alternative hypothesis: true difference in means is not equal to 0
## 94 percent confidence interval:
##   3.847898 26.152102
## sample estimates:
## mean of x mean of y
##      27.5      12.5
```

- **Toma de decisión:**

Como podemos observar de la salida anterior, $p-value = 0.01583$ lo cual es menor al valor $\alpha = 0.06$ por lo tanto tenemos evidencia suficiente para rechazar H_0 y quedarnos con H_1.

Concluyendo que con un $\alpha = 0.06$ si existen diferencias entre las concentraciones de arsénico de la zona urbana y rural.

Ejercicio 8.

Descripción.

Se quiere determinar si un cambio en el método de fabricación de una pieza ha sido efectivo o no. Para esta comparación se tomaron 2 muestras, una antes y otra después del cambio en el proceso y los resultados obtenidos se muestran en la siguiente tabla:

Números de piezas	Antes	Después
Defectuosas	75	80
Analizadas	1500	2000

Supón un nivel de significancia del 10%.

Solución.

1. **Planteamiento de las hipótesis:**

Dado que nos dan la cantidad de piezas defectuosas y analizadas tanto antes y después del nuevo proceso, podemos obtener la media de piezas defectuosas del antes y después. Por otra parte como queremos saber si el método que se uso despues es mas efectivo, entonces la media de piezas defectuosas del método que se usaba antes tendria que ser mayor al que se usa después.

Representando como μ_1 y μ_2 el promedio de piezas defectuosas que se usaban antes y después respectivamente. Entonces las hipótesis a plantear son las siguientes:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (El método que se uso después no es más efectivo del que se usaba antes)}$$
$$H_1 : \mu_1 - \mu_2 > 0 \text{ (El método que se uso después es más efectivo del que se usaba antes)}$$

2. **Cálculo del estadístico de prueba:**

Como desconocemos las varianzas poblacionales y las muestras son grandes, el estadístico de prueba es:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Al sustituir los datos que tenemos en la ecuación anterior obtenemos que **Z** es igual a (redondeado a 2 decimales):

```
data_before <- c( rep(1, 75), rep(0, 1500-75) )
data_after <- c( rep(1, 80), rep(0, 2000-80) )

n1 <- 1500
n2 <- 2000
x1 <- mean(data_before)
x2 <- mean(data_after)
s1 <- sd(data_before)
s2 <- sd(data_after)

Z <- ((x1 - x2) - (0)) / sqrt(((s1**2)/n1) + ((s2**2)/n2))
round(Z, 2)
```

```
## [1] 1.4
```

3. **Cálculo del p-valor**

Como la prueba es unilateral, tenemos que calcular la probabilidad de que **Z** supere el valor del estadístico de prueba, es decir:

$$P(Z > 1.4) = 1 - P(Z < 1.4)$$

Si calculamos la operación anterior, con el valor completo de **Z**, obtenemos que el **p-valor** es: (redondeado a 4 decimales):

```
p_value <- 1 - pnorm(Z)
round(p_value, 4)
```

```
## [1] 0.0805
```

4. **Toma de decisión:**

Como podemos observar del resultado anterior, el $p - valor = 0.0805$ lo cual es menor al valor $\alpha = 0.10$ por lo tanto tenemos evidencia suficiente para rechazar H_0 y quedarnos con H_1.

Por lo tanto podemos concluir que con un nivel $\alpha = 0.10$ el método de fabricacion de una pieza que se uso después es más efectivo del que se usaba antes.

Ejercicio 9.

Descripción.

Diez individuos participaron de programa para perder peso corporal por medio de una dieta. Los voluntarios fueron pesados antes y después de haber participado del programa y los datos en libras se pueden observar en la tabla de abajo. ¿Hay evidencia que soporte la afirmación que la dieta disminuye el peso promedio de los participantes? Usar nivel de significancia del 5%.

Sujeto	1	2	3	4	5	6	7	8	9	10
Antes	195	213	247	201	187	210	215	246	294	310
Después	187	195	221	190	175	197	199	221	278	285

Solución.

- **Planteamiento de las hipótesis:**

Como queremos probar si la dieta hizo que el peso promedio de los participantes disminuyera, entonces el peso promedio del peso que tenian antes deberia ser mayor que el peso promedio después de la dieta.

Representando como μ_1 y μ_2 el peso promedio de los participante antes y después del programa respectivamente. Entonces las hipótesis a plantear son las siguientes:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (El peso promedio de los participantes no disminuyo)}$$
$$H_1 : \mu_1 - \mu_2 > 0 \text{ (El peso promedio de los participantes disminuyo)}$$

- **Estadístico de prueba:**

Para este tipo de prueba, como las varianzas poblacionales son desconocidas y ademas la muestra no es tan grande, el estadístico de prueba a utilizar es **t**. Solo que no conocemos si las varianzas poblaciones son iguales o distintas pero como las muestras pertenecen a los mismos participantes podemos asumir que las varianzas son iguales, por lo tanto nuestro estadístico de prueba sera:

$$t_{n_1+n_2-2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{k}}$$
$$k = n_1 + n_2 - 2$$

- **Cálculo del p-valor con R:**

Para obtener el *p-valor* de una manera más facil nos podemos apoyar de la función *t.test()* de R, con la cual al pasarle los parámetros que ya conocemos se obtiene:

```
weight_before <- c(195, 213, 247, 201, 187, 210, 215, 246, 294, 310)
weight_after <- c(187, 195, 221, 190, 175, 197, 199, 221, 278, 285)

t.test(x = weight_before, y = weight_after, alternative = "greater",
       mu = 0, var.equal = TRUE, conf.level = 0.95)
```

```
##
##  Two Sample t-test
##
## data:  weight_before and weight_after
## t = 0.95087, df = 18, p-value = 0.1771
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -14.0022      Inf
## sample estimates:
## mean of x mean of y
##     231.8     214.8
```

- **Toma de decisión:**

Como se puede observar del resultado anterior, $p-value = 0.1771$ lo cual es superior al nivel de significancia $\alpha = 0.05$ por lo tanto si bien el promedio del antes y despues si disminuyo (lo podemos ver casi al final del resultado anterior), con un valor de significancia igual al 5% no tenemos evidencia suficiente para rechazar H_0.

Por lo que se concluye que con un nivel $\alpha = 0.05$ no se tiene evidencia significativa para afirmar que el peso promedio de los participantes disminuyo.

Ejercicio 10.

Descripción.

Una muestra aleatoria de 64 bolsas de palomitas con queso pesa, en promedio, 5.23 onzas, con una desviación estándar de 0.24 onzas. El dueño asegura que cada bolsa tiene 5.5 onzas. ¿Considera que dice la verdad según la muestra obtenida? Usa un nivel de significancia de 5%.

Solución.

1. **Planteamiento de las hipótesis:**

Según la información proporcionada, el parámetro a analizar es μ, es decir, el promedio de onzas que tienen las palomitas con queso. Por lo tanto, las hipótesis que se pueden plantear son:

$$H_0 : \mu = 5.5 \text{ (La afirmación del dueño es correcta)}$$

$$H_1 : \mu \neq 5.5 \text{ (La afirmación del dueño no es correcta)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como tenemos una muestra relativamente grande (64 bolsas) y desconocemos la varianza poblacional, entonces el estadístico de prueba es:

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Al calcular el valor de **Z** con los datos proporcionados, obtenemos (redondeado a 2 decimales):

```
mu <- 5.5
n <- 64
x <- 5.23
S <- 0.24

Z <- (x - mu) / (S / sqrt(n))
round(Z, 2)
```

```
## [1] -9
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y que tenemos una prueba bilateral, tenemos que calcular el valor de:

$Z_{\alpha} = Z_{0.05}$, lo cual redondeado a 2 decimales es:

```
round(qnorm(0.05), 2)
```

```
## [1] -1.64
```

y $Z_{1-\alpha} = Z_{0.95}$, lo cual redondeado a 2 decimales es:

```
round(qnorm(0.95), 2)
```

```
## [1] 1.64
```

Por lo tanto, la regla de decisión es rechazar H_0 si $Z < -1.64$ o $Z > 1.64$.

5. **Toma de decisión:**

Dado que el estadístico de prueba $Z = -9 < Z_{0.05} = -1.64$, debemos rechazar H_0 y quedarnos con H_1.

Por lo que se concluye que con un valor $\alpha = 0.05$ la afirmación del dueño no es correcta, es decir el promedio del contenido de las palomitas con queso no es de 5.5 onzas.

Ejercicio 11.

Descripción.

La estatura promedio de mujeres en el grupo de primer año de cierta universidad ha sido, históricamente, de 162.5 centímetros, con una desviación estándar de 6.9 centímetros. ¿Existe alguna razón para creer que ha habido un cambio en la estatura promedio, si una muestra aleatoria de 50 mujeres del grupo actual de primer año tiene una estatura promedio de 165.2 centímetros? Utilice el *p-valor* para su conclusión. Suponga que la desviación estándar permanece constante.

Solución.

1. **Planteamiento de las hipótesis:**

Como el parámetro de interés para este problema es μ, es decir la estatura promedio de mujeres en el grupo de primer año de cierta universidad. Entonces las hipótesis que se pueden plantear son:

$$H_0 : \mu = 162.5 \text{ (No ha habido un cambio en la estatura promedio de mujeres)}$$
$$H_1 : \mu \neq 162.5 \text{ (Ha habido un cambio en la estatura promedio de mujeres)}$$

2. **Cálculo del estadístico de prueba:**

Dado que conocemos la desviación estándar poblacional, el estadístico de prueba es:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Al calcular el valor de Z con los datos proporcionados, obtenemos (redondeado a 2 decimales):

```r
mu <- 162.5
sigma <- 6.9
n <- 50
x <- 165.2

Z <- (x - mu) / (sigma / sqrt(n))
round(Z, 2)
```

```
## [1] 2.77
```

3. **Cálculo del p-valor**

Como la prueba es bilateral y el valor del estadístico de prueba de la muestra es positivo $Z = 2.77$, entonces tenemos que calcular la siguiente probabilidad:

$$2P(Z > 2.77) = 2(1 - P(Z < 2.77))$$

Lo cual da como resultado (redondeado a 4 decimales):

```r
p_value <- 2 * (1 - pnorm(Z))
round(p_value, 4)
```

```
## [1] 0.0057
```

4. **Toma de decisión**

Como se puede apreciar del resultado anterior el *p-valor = 0.0057* es muy pequeño. Lo podemos interpretar como la probabilidad de que H_0 se cumpla es aproximadamente 0.0057. Es decir, es muy improbable suponer cierta H_0.

Por lo tanto se concluye que H_1 se cumple, es decir que ha habido un cambio en la estatura promedio de mujeres en el grupo de primer año de cierta universidad.

Ejercicio 12.

Descripción.

Un estudio de la Universidad de Colorado revela que correr aumenta el porcentaje de la tasa matabólica basal (TMB) en mujeres ancianas. La TMB promedio de 30 ancianas corredoras fue 34% más alta que la TMB promedio de 30 ancianas sedentarias, en tanto que las desviaciones estándar reportadas fueron de 10.5% y 10.2% respectivamente. ¿Existe un aumento significativo en la TMB de las corredoras respecto a las sedentarias? Suponga que las poblaciones se distribuyen de forma aproxidamante normal con varianzas iguales. Utilice el *p-valor* en sus conclusiones.

Solución.

1. **Planteamiento de las hipótesis:**

Representando con μ_1 y μ_2 la TMB en mujeres ancianas corredoras y la TMB en mujeres ancianas sedentarias respectivamente. Entonces, las hipótesis que se pueden plantear son:

$H_0 : \mu_1 - \mu_2 = 0$ (No existe un aumento significativo en la TMB de las corredoras respecto a las sedentarias)

$H_1 : \mu_1 - \mu_2 > 0$ (Existe un aumento significativo en la TMB de las corredoras respecto a las sedentarias)

2. **Cálculo del estadístico de prueba:**

Para esta prueba como las poblaciones se distribuyen de forma aproximadamente normal y con varianzas iguales, entonces el estadístico de prueba es:

$$t_{n_1+n_2-2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{k}}$$

$$k = n_1 + n_2 - 2$$

Al sustituir los datos proporcionados en la formula del estadístico de prueba, tenemos que el valor de t es (redondeado a 2 decimales):

```
n1 <- 30
s1 <- 10.5
n2 <- 30
s2 <- 10.2

k <- n1 + n2 - 2
sp <- sqrt(( (n1-1)*s1**2 + (n2-1)*s2**2 ) / k)
t <- ((34) - (0)) / (sp * sqrt(1/n1 + 1/n2))

round(t, 2)
```

```
## [1] 12.72
```

3. **Cálculo del p-valor**

Como la prueba es unilateral tenemos que calcular la probabilidad de que t_{58} supere el valor obtenido en el estadistico de prueba, es decir:

$$P(t_{58} > 12.72) = 1 - P(t_{58} < 12.72)$$

```
p_value <- 1 - pt(t, n1 + n2 - 2)
p_value
```

```
## [1] 0
```

4. **Toma de decisión**

Como el $p - valor = 0$ es decir que es practicamente improbable se cumpla H_0, por lo tanto nos quedamos con H_1. Por lo que se puede concluir que si existe un aumento significativo en la TMB de las corredoras respecto a las sedentarias

Ejercicio 13.

Descripción.

De acuerdo con Chemical Engineering, una propiedad importante de la fibra es su absorbencia de agua. Se encontró que el porcentaje promedio de absorción de 25 pedazos de fibra de algodón seleccionados al azar es 20, con una desviación estándar de 1.5. Una muestra aleatoria de 25 pedazos de acetato reveló un porcentaje promedio de 12 con una desviación estándar de 1.25. ¿Existe evidencia sólida de que el porcentaje promedio de absorción de la población es significativamente mayor para la fibra de algodón que para el acetato? Suponga que el porcentaje de absorbencia se distribuye de forma normal y que las varianzas de la población en el porcentaje de absorbencia para las dos fibras son iguales. Utilice un nivel de significancia de 0.03.

Solución.

1. **Planteamiento de las hipótesis:**

Si representamos como μ_1 y μ_2 las medias poblacionales del porcentaje de absorción de la fibra de algodón y la fibra de acetato respectivamente. Entonces las hipótesis que se pueden plantear son:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (El porcentaje de absorbencia para las dos fibras son iguales)}$$
$$H_1 : \mu_1 - \mu_2 > 0 \text{ (El porcentaje de absorbencia de la fibra de algodón es mayor que la del acetato)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.03$

3. **Cálculo del estadístico de prueba:**

Para esta prueba como el porcentaje de absorbencia se distribuye de forma normal y las varianzas son iguales, entonces el estadístico de prueba es:

$$t_{n_1+n_2-2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{k}}$$
$$k = n_1 + n_2 - 2$$

Al sustituir los datos proporcionados en la formula del estadístico de prueba, tenemos que el resultado de t es (redondeado a 2 decimales):

```
n1 <- 25
x1 <- 20
s1 <- 1.5
n2 <- 25
x2 <- 12
s2 <- 1.25

k <- n1 + n2 - 2
sp <- sqrt(( (n1-1)*s1**2 + (n2-1)*s2**2 ) / k)
t <- ((x1 - x2) - (0)) / (sp * sqrt(1/n1 + 1/n2))

round(t, 2)
```

```
## [1] 20.49
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.03$ y que tenemos una prueba unilateral, se tiene que calcular el valor de $t_{1-\alpha,n_1+n_2-2} = t_{0.97,48}$, cuyo valor es:

```r
round(qt(0.97, 48), 2)
```

```
## [1] 1.93
```

Entonces la regla de decisión es rechazar H_0 si $t > 1.93$

5. **Toma de decisión**

Dado que $t = 20.49 > t_{0.97,48} = 1.93$, por lo tanto tenemos evidencia suficiente para rechazar H_0. Por otra parte como se cumple H_1 y la diferencia es relativamente grande entre el valor de t y $t_{0.97}$ esto indica que el porcentaje de absorción de la fibra de algodón es relativamente mayor respecto al de la fibra de acetato.

Podemos comprobar la decisión calculando el $p-valor$, el cual es:

$$P(t_{48} > 20.49) = 1 - P(t_{48} < 20.49)$$

```r
p_value <- 1 - pt(t, n1+n2-2)
round(p_value, 2)
```

```
## [1] 0
```

Como se puede observar de la salida anterior el $p-valor = 0$, lo que significa que es practicamente improbable que H_0 se cumpla.

Por lo tanto, con un nivel de significancia $\alpha = 0.03$ se concluye tras los 2 métodos aplicados que existe evidencia sólida de que el porcentaje promedio de absorción es significativamente mayor para la fibra de algodón que para el acetato.

Ejercicio 14.

Descripción.

La experiencia indica que el tiempo que requieren los estudiantes de último año de preparatoria para contestar una prueba estandarizada es una variable aleatoria normal con una media de 35 minutos. Si a una muestra aleatoria de 20 estudiantes de último año de preparatoria le toma un promedio de 33.1 minutos contestar esa prueba con una desviación estándar de 4.3 minutos, ¿se puede asegurar que el tiempo para tomar la prueba es menor a los 35 minutos?. Usa un nivel de significancia de 0.04.

Solución.

1. **Planteamiento de las hipótesis:**

Según la información proporcionada, el parámetro a analizar es μ, es decir el tiempo promedio que los estudiantes de último año de preparatoria tardan en contestar una prueba estandarizada. Por lo tanto, las hipótesis que se pueden plantear son:

$$H_0 : \mu = 35 \text{ (El tiempo para tomar la prueba es igual a 35 minutos)}$$
$$H_1 : \mu < 35 \text{ (El tiempo para tomar la prueba es menor a los 35 minutos)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.04$

3. **Cálculo del estadístico de prueba:**

Como se tiene una variable aleatoria normal con una muestra pequeña (20 estudiantes) y se desconoce la varianza poblacional, por lo tanto el estadístico de prueba es:

$$t_{n-1} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Al sustituir los datos proporcionados en la formula anterior, tenemos que el valor de t es (redondeado a 2 decimales):

```
mu <- 35
n <- 20
x <- 33.1
s <- 4.3

t <- (x - mu) / (s / sqrt(n))
round(t, 2)
```

```
## [1] -1.98
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.04$ y que tenemos una prueba unilateral, tenemos que calcular el valor de $t_{\alpha,n-1} = t_{0.04,19}$ lo cual es:

```
round(qt(0.04, 19), 2)
```

```
## [1] -1.85
```

Entonces la regla de decisión es rechazar H_0 si $t < -1.85$

5. **Toma de decisión:**

Dado que el estadístico de prueba $t = -1.98 < t_{0.04,19} = -1.85$ entonces tenemos evidencia suficiente para rechazar H_0 y quedarnos con H_1.

Por lo tanto con un nivel de significancia $\alpha = 0.04$ se concluye que se puede asegurar que el tiempo para tomar la prueba es menor a los 35 minutos.

Ejercicio 15.

Descripción.

Una muestra aleatoria de tamaño $n_1 = 25$, tomada de una población normal con una desviación estándar $\sigma_1 = 5.2$, tiene una media $\overline{x}_1 = 81$. Una segunda muestra aleatoria de tamaño $n_2 = 36$, que se toma de una población normal diferente con una desviación estándar $\sigma_2 = 3.4$, tiene una media $\overline{x}_2 = 76$. ¿Considera que las medias de las dos poblaciones son iguales? Utilice el *p-valor* para su conclusión.

Solución.

1. **Planteamiento de las hipótesis:**

Las hipótesis a plantear son:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (Las medias de las dos poblaciones son iguales)}$$
$$H_1 : \mu_1 - \mu_2 \neq 0 \text{ (Las medias de las dos poblaciones son diferentes)}$$

2. **Cálculo del estadístico de prueba:**

Como se conocen las desviaciones estándar poblacional, entonces el estadístico de prueba es:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Al sustituir los datos proporcionados en la formula anterior, tenemos que Z es (redondeado a 2 decimales):

```r
n1 <- 25
sigma1 <- 5.2
x1 <- 81
n2 <- 36
sigma2 <- 3.4
x2 <- 76

Z <- ((x1 - x2) - (0)) / sqrt(sigma1**2/n1 + sigma2**2/n2)

round(Z, 2)
```

```
## [1] 4.22
```

3. **Cálculo del p-valor**

Como la prueba es bilateral y el valor del estadístico de prueba de la muestra es positivo $Z = 4.22$, entonces tenemos que calcular la siguiente probabilidad:

$$2P(Z > 4.22) = 2(1 - P(Z < 4.22))$$

```r
p_value <- 2 * (1 - pnorm(Z))
p_value
```

```
## [1] 2.424822e-05
```

4. **Toma de decisión**

Dado que el *p-valor* obtenido es muy pequeño, lo podemos interpretar como la probabilidad de que H_0 se cumpla es aproximada a 2.424822e-05 es decir es muy improbable que se cumpla H_0, por lo tanto decidimos quedarnos con H_1.

Por lo tanto tenemos evidencia suficiente con la cual podemos concluir que hay diferencias entre las medias de las dos poblaciones.

Ejercicio 16.

Descripción.

Un fabricante afirma que la resistencia promedio a la tensión del hilo A excede a la resistencia a la tensión promedio del hilo B en al menos 12 kilogramos. Para probar esta afirmación se pusieron a prueba 50 pedazos de cada tipo de hilo en condiciones similares. El hilo tipo A tuvo una resistencia promedio a la tensión de 86.7 kilogramos con una desviación estándar de 6.28 kilogramos; mientras que el hilo tipo B tuvo una resistencia promedio a la tensión de 77.8 kilogramos con una desviación estándar de 5.61 kilogramos. Pruebe la afirmación del fabricante usando un nivel de significancia de 0.05

Solución.

1. **Planteamiento de las hipótesis:**

Representando como μ_1 y μ_2 el promedio a la tensión del hilo A y del hilo B respectivamente. Entonces las hipótesis a plantear son las siguientes:

$$H_0 : \mu_1 - \mu_2 = 12 \text{ (La resistencia del hilo A excede al menos 12 kg a la del B)}$$
$$H_1 : \mu_1 - \mu_2 < 12 \text{ (La resistencia del hilo A excede menos de 12 kg a la del B)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como las muestras son relativamente grandes (50 pedazos), entonces el estadístico de prueba es:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Al sustituir los datos proporcionados en la formula anterior, tenemos que el valor de Z es (redondeado a 2 decimales):

```r
n1 <- 50
x1 <- 86.7
s1 <- 6.28
n2 <- 50
x2 <- 77.8
s2 <- 5.61

Z <- ((x1 - x2) - (12)) / sqrt(s1**2/n1 + s2**2/n2)
round(Z, 2)
```

```
## [1] -2.6
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y que tenemos una prueba unilateral, tenemos que calcular el valor de $Z_\alpha = Z_{0.05}$ cuyo valor es (redondeado a 2 decimales):

```r
round(qnorm(0.05), 2)
```

```
## [1] -1.64
```

Entonces, la regla de decisión es rechazar H_0 si $Z < -1.64$

5. **Toma de decisión:**

Dado que el estadístico de prueba $Z = -2.6 < Z_{0.05} = -1.64$ tenemos evidencia suficiente para rechazar H_0 y quedarnos con H_1.

Por lo tanto con un nivel de significancia $\alpha = 0.05$ se concluye que la afirmación del fabricante es incorrecta, es decir que la resistencia promedio a la tensión del hilo A excede menos de 12 kilogramos a la resistencia a la tensión promedio del hilo B.

Ejercicio 17.

Descripción.

Se llevó a cabo un estudio para saber si el aumento en la concentración de sustrato tiene un efecto apreciable sobre la velocidad de una reacción química. Con una concentración de sustrato de 1.5 moles por litro, la reacción se realizó 15 veces, con una velocidad promedio de 7.5 micromoles por 30 minutos y una desviación estándar de 1.5. Con una concentración de sustrato de 2.0 moles por litro, se realizaron 12 reacciones que produjeron una velocidad promedio de 8.8 micromoles por 30 minutos y una desviación estándar muestral de 1.2. ¿Hay alguna razón para creer que este incremento en la concentración de sustrato ocasiona un aumento en la velocidad media de la reacción de más de 0.5 micromoles por 30 minutos? Utilice un nivel de significancia de 0.01 y suponga que las poblaciones se distribuyen de forma aproximadamente normal con varianzas iguales.

Solución.

1. **Planteamiento de las hipótesis:**

Si representamos como μ_1 y μ_2 las velocidades medias con el incremento y sin el incremento de sustrato respectivamente. Entonces las hipótesis que se pueden plantear son:

$$H_0 : \mu_1 - \mu_2 = 0.5 \text{ (El incremento de sustrato no tiene un efecto apreciable sobre la velocidad de reacción)}$$

$$H_1 : \mu_1 - \mu_2 > 0.5 \text{ (El incremento de sustrato tiene un efecto apreciable sobre la velocidad de reacción)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor de $\alpha = 0.01$

3. **Cálculo del estadístico de prueba:**

Como las muestras no son tan grandes y sabemos que que las poblaciones se distribuyen de forma aproximadamente normal con varianzas iguales, entonces el estadístico de prueba es:

$$t_{n_1+n_2-2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{k}}$$

$$k = n_1 + n_2 - 2$$

Al sustituir los datos proporcionados en la formula del estadístico de prueba, tenemos que el valor de t es (redondeado a 2 decimales):

```
# Con el incremento
n1 <- 12
x1 <- 8.8
s1 <- 1.2
# Sin el incremento
n2 <- 15
x2 <- 7.5
s2 <- 1.5

k <- n1 + n2 - 2
sp <- sqrt(( (n1-1)*s1**2 + (n2-1)*s2**2 ) / k)
t <- ((x1 - x2) - (0.5)) / (sp * sqrt(1/n1 + 1/n2))

round(t, 2)
```

```
## [1] 1.5
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.01$ y que tenemos una prueba unilateral, tenemos que calcular el valor de $t_{1-\alpha,n_1+n_2-2} = t_{0.99,25}$ cuyo valor es (redondeado a 2 decimales):

```
round( qt(0.99, n1+n2-2), 2)
```

```
## [1] 2.49
```

Entonces, la regla de decisión es rechazar H_0 si $t > 2.49$

5. **Toma de decisión:**

Dado que el estadístico de prueba $t = 1.5 \not> t_{0.99,25} = 2.49$, entonces no tenemos suficientes evidencias para rechazar H_0.

Por lo tanto, tras el aumento en las concentaciones de sustrato de 1.5 a 2.0 moles por litro y con un nivel de significancia $\alpha = 0.01$ no se rechaza H_0 y se concluye que no hay evidencia suficiente para afirmar que este incremento en la concentración de sustrato ocasiona un aumento en la velocidad media de la reacción de más de 0.5 micromoles por 30 minutos.

Ejercicio 18.

Descripción.

En un invierno con epidemia de influenza los investigadores de una conocida empresa farmacéutica encuestaron a los padres de 2000 bebés para determinar si el nuevo medicamento de la empresa era eficaz después de dos días. De 120 bebés que tenían influenza y que recibieron el medicamento, 29 se curaron en dos días o menos. De 280 bebés que tenían influenza pero no recibieron el fármaco, 56 se curaron en dos días o menos. ¿Hay alguna indicación significativa que apoye la afirmación de la empresa sobre la eficacia del medicamento?

Solución.

1. **Planteamiento de las hipótesis:**

Representando como μ_1 al promedio de bebés que se curaron y que recibieron el medicamento y como μ_2 el promedio de bebés que se curaron sin recibir el medicamento, donde en ambos casos los bebés se curaron en 2 días o menos. Entonces las hipótesis a plantear son las siguientes:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (El medicamento no es eficaz)}$$
$$H_0 : \mu_1 - \mu_2 > 0 \text{ (El medicamento es eficaz)}$$

2. **Cálculo del estadístico de prueba:**

Como desconocemos las varianzas poblaciones y las muestras son grandes, entonces el estadístico de prueba es:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Al sustituir los datos proporcionados en la formula anterior obtenemos que Z es igual a (redondeado a 2 decimales):

```
n1 <- 120
n2 <- 280
x1 <- 29 / n1
x2 <- 56 / n2
with_med <- c( rep(1, 29), rep(0, n1-29) )
without_med <- c( rep(1, 56), rep(0, n2-56) )
s1 <- sd(with_med)
s2 <- sd(without_med)

Z <- ((x1 - x2) - (0)) / sqrt(((s1**2)/n1) + ((s2**2)/n2))
round(Z, 2)
```

```
## [1] 0.91
```

3. **Cálculo del p-valor**

Como la prueba es unilateral tenemos que calcular la probabilidad de que Z supere el valor obtenido en el estadistico de prueba, es decir:

$$P(Z > 0.91) = 1 - P(Z < 0.91)$$

```
p_value <- 1 - pnorm(Z)
round( p_value, 2 )
```

```
## [1] 0.18
```

4. **Toma de decisión**

Respecto al $p - valor \approx 0.18$, la conclusión estaria condicionada al nivel de significancia que se desee:

- Rechazamos H_0, con un valor minimo de $\alpha = 0.18$, es decir con un 82% máximo de confianza podemos rechazar H_0 y concluimos el medicamento es eficaz.
- No rechazamos H_0, con un valor máximo de $\alpha = 0.17$, es decir, con un 83% mínimo de confianza no podemos rechazar H_0 y concluimos que el medicamento no es eficaz.

Ejercicio 19.

Descripción.

Se asegura que el contenido de los envases de un lubricante específico se distribuye normalmente con una varianza de 0.03 litros. Se toma una muestra aleatoria de 10 envases y se obtiene:

$$10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, 9.8$$

¿Se puede concluir valido lo que se asegura?. Usa el *p-valor* para tus conclusiones.

Solución.

1. **Planteamiento de las hipótesis:**

Como el parámetro de interes para este problema es la varianza, entonces las hipotesis que se pueden plantear son:

$$H_0 : \sigma^2 = 0.03 \text{ (Es valido lo que se asegura)}$$
$$H_1 : \sigma^2 \neq 0.03 \text{ (No es valido lo que se asegura)}$$

2. **Cálculo del estadístico de prueba:**

Tenemos una muestra de 20 envases que se distribuyen normalmente, por lo tanto el estadístico de prueba es:

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma^2}$$

Al sustituir los datos proporcionados en la formula anterior, tenemos que el valor de χ^2 es (redondeado a 2 decimales):

```r
data <- c(10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, 9.8)
n <- 10
s_square <- var(data)
sigma_square <- 0.03

chisq <- ((n - 1) * s_square) / sigma_square
round(chisq, 2)
```

```
## [1] 18.13
```

3. **Cálculo del p-valor**

Como la prueba es bilateral la probabilidad que tenemos que calcular es:

$$P(\chi_9^2 > 18.13 \text{ o } \chi_9^2 < -18.13) = [1 - P(\chi_9^2 < 18.13)] + P(\chi_9^2 < -18.13)$$

```r
p_value <- (1 - pchisq(chisq, n-1)) + pchisq(-chisq, n-1)
round(p_value, 2)
```

```
## [1] 0.03
```

4. **Toma de decisión**

Como el $p - valor = 0.03$, la probabilidad de que se cumpla H_0 es minima, por lo que se rechaza H_0 y nos quedamos con H_1, por lo tanto concluimos que no es valido lo que se asegura, es decir que la varianza del contenido de los envases de un lubricante específico no es de 0.03.

Ejercicio 20

Descripción.

Se dice que una máquina despachadora de bebida gaseosa está fuera de control si la varianza de los contenidos excede a 1.15 decilitros. Se toma una muestra aleatoria de 25 bebidas de esta máquina y se obtiene una varianza de 2.03 decilitros, ¿esto indica, a un nivel de significancia de 0.05, que la máquina esta fuera de control? Suponga que los contenidos se distribuyen de forma aproximadamente normal.

Solución.

1. **Planteamiento de las hipótesis:**

El parámetro de interes para este problema es la varianza por lo que las hipótesis que se pueden plantear son:

$$H_0 : \sigma^2 = 1.15 \text{ (La máquina esta bajo control)}$$
$$H_1 : \sigma^2 > 1.15 \text{ (La máquina esta fuera de control)}$$

2. **Selección del nivel de significancia:**

El nivel de significancia ya es dado con un valor $\alpha = 0.05$

3. **Cálculo del estadístico de prueba:**

Como se tiene una muestra de 25 bebidas cuyo contenidos siguen una distribución normal, entonces el estadístico de prueba es:

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma^2}$$

Al sustituir los datos proporcionados en la formula anterior, tenemos que el valor de χ^2 es (redondeado a 2 decimales):

```
sigma_square <- 1.15
n <- 25
s_square <- 2.03

chisq <- ((n - 1) * s_square) / sigma_square
round(chisq, 2)
```

```
## [1] 42.37
```

4. **Formulación de la regla de decisión:**

Dado que $\alpha = 0.05$ y que tenemos una prueba unilateral, se tiene que calcular el valor de $\chi_{1-\alpha,n-1} = \chi_{0.95,24}$, dicho valor es (redondeado a 2 decimales):

```
round( qchisq(0.95, n-1), 2 )
```

```
## [1] 36.42
```

Entonces, la regla de decisión es rechazar H_0 si $\chi^2 > 36.42$

5. **Toma de decisión:**

Dado que el estadístico de prueba $\chi^2 = 42.37 > \chi_{0.95,24} = 36.42$, se rechaza H_0 y nos quedamos con H_1.

Podemos comprabar la decisión tomada calculando el *p-valor* el cual es:

$$P(\chi^2 > 42.37) = 1 - P(\chi^2 < 42.37)$$

```
p_value <- 1 - pchisq(chisq, n-1)
round(p_value, 2)
```

```
## [1] 0.01
```

Del resultado anterior podemos ver que el *p-valor = 0.01* es menor que el nivel de significancia $\alpha = 0.05$ con lo cual corroboramos que H_0 es rechazada.

Por lo tanto con un nivel de significancia $\alpha = 0.05$ tenemos evidencia suficiente para rechazar H_0 y se concluye que la máquina esta fuera de control.