1 Test para la media de una población normal

Sean $X_1, \ldots, X_n \sim N\left(\mu, \sigma^2\right)$ una muestra aleatoria (v.a.i.i.d. variables aleatorias independientes idénticamente distribuidas). Se quiere testear

$$H_0 : \mu = \mu_0 \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \mu \neq \mu_0 \qquad H_1 : \mu > \mu_0 \qquad H_1 : \mu < \mu_0$$

1.1 Caso varianza conocida

Estadístico del test

$$\frac{\overline{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} N(0,1)$$

Las instrucciones del R para hacer este test.

```
library(BSDA)
z.test(x = datos, alternative = c("two.sided", "less", "greater"), mu = mucero,
sigma.x = desvio.conocido, conf.level = 0.95)
```

La instrucción `z.test` hace el test recién descripto. En estos tests y los restantes hay que completar lo siguiente: en `datos` el vector con los valores observados en la muestra, `mucero` $= \mu_0$ y `desvio.conocido` $= \sigma$. En `alternative`, poner la hipótesis alternativa de interés: `"two.sided"` para $\mu \neq \mu_0$, `"less"` para $\mu < \mu_0$, o `"greater"` para $\mu > \mu_0$. En `conf.level`, completar el nivel de confianza deseado para el intervalo de confianza.

Para calcular la potencia del test: `power.t.test`.

1.2 Caso varianza desconocida

Estadístico del test

$$\frac{\overline{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} t_{n-1}$$

donde

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{1}$$

es la varianza muestral de las observaciones.

Instrucciones para el R:

```
t.test(x = datos, alternative = c("two.sided", "less", "greater"), mu = mucero, conf.level = 0.95)
```

Asumimos que en el vector `datos` se guardan las observaciones. Donde dice `alternative`, hay que poner entre comillas cuál opción de hipótesis alternativa del test es la que se desea: `"two.sided"`: $H_1 : \mu \neq \mu_0$, o sea test a dos colas, `"less"` si uno quiere $H_1 : \mu < \mu_0$ o `"greater"` si uno quiere $H_1 : \mu > \mu_0$. En `mucero` se anota el valor de μ_0 propuesto. Por default, el nivel de confianza con el que se construyen los intervalos de confianza es 0.95, pero se puede cambiar.

2 Test para la varianza de una población normal

Sean $X_1, \ldots, X_n \sim N\left(\mu, \sigma^2\right)$ una muestra aleatoria (v.a.i.i.d. variables aleatorias independientes idénticamente distribuidas). Se quiere testear

$$H_0 : \sigma = \sigma_0 \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \sigma \neq \sigma_0 \qquad H_1 : \sigma > \sigma_0 \qquad H_1 : \sigma < \sigma_0$$

y la media poblacional μ es desconocida. Estadístico del test

$$(n-1)\frac{S^2}{\sigma_0^2} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} \chi^2_{n-1}$$

donde S^2 es la varianza muestral definida en (1).

Instrucciones de R:

```
library(TeachingDemos)
sigma.test(x = datos.x, sigma = sigma_0, alternative = c("two.sided", "less", "greater"), conf.level =
0.95)
```

Completar con el vector de datos en `datos.x` y el valor `sigma_0` $= \sigma_0$ de la hipótesis nula.

3 Test asintótico (o aproximado) para la media de una población

Sean $X_1, \ldots, X_n \sim F$ una muestra aleatoria donde F es una distribución no necesariamente normal y n es un número suficientemente grande. Sea $\mu = E(X_1)$. Se quiere testear

$$H_0 : \mu = \mu_0 \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \mu \neq \mu_0 \qquad H_1 : \mu > \mu_0 \qquad H_1 : \mu < \mu_0$$

Estadístico del test

$$\frac{\overline{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\overset{\text{aprox}}{\sim}} N(0,1)$$

donde S^2 es la varianza muestral definida en (1). Instrucciones para el R: `t.test`, se completa igual que en el Test 1.2.

4 Test asintótico para una proporción poblacional

Sean $X_1, \ldots, X_n \sim Bi(1,p)$ una muestra aleatoria con n suficientemente grande. Se quiere testear

$$H_0 : p = p_0 \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : p \neq p_0 \qquad H_1 : p > p_0 \qquad H_1 : p < p_0$$

Estadístico del test

$$\frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\overset{\text{aprox}}{\sim}} N(0,1)$$

donde $\widehat{p} = \overline{X}$ es la proporción de éxitos en la muestra. Instrucciones en el R

```
prop.test(x = cantidad.de.exitos, n = cantidad.de.intentos, p = p.cero,
 alternative = c("two.sided", "less", "greater"), conf.level = 0.95, correct = TRUE)
```

Completar con los valores `cantidad.de.exitos, cantidad.de.intentos, p.cero, alternative, conf.level` de interés. La opción `correct = TRUE` indica que se utilice corrección por continuidad, `FALSE` indica que no se utilice.

5 Test para comparar medias de dos poblaciones normales independientes

Sean $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$, e $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ dos muestras aleatorias independientes entre sí. Se quiere testear

$$H_0 : \mu_1 - \mu_2 = \delta \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \mu_1 - \mu_2 \neq \delta \qquad H_1 : \mu_1 - \mu_2 > \delta \qquad H_1 : \mu_1 - \mu_2 < \delta$$

5.1 Caso varianzas conocidas

Estadístico del test

$$\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \underset{\text{bajo } H_0}{\underset{\uparrow}{\sim}} N(0,1)$$

Instrucciones para el R:
`z.test(x = datos.x, y = datos.y, alternative = "two.sided", mu = delta, sigma.x = sigma1, sigma.y = sigma2, conf.level = 0.95)`

Completar como antes, pero en `datos.x`, y `datos.y` completar con los dos vectores de datos, `delta` $=\delta$, `sigma1` $=\sigma_1$, `sigma2` $=\sigma_2$.

5.2 Caso varianzas desconocidas pero iguales

Estadístico del test

$$\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \underset{\text{bajo } H_0}{\underset{\uparrow}{\sim}} t_{n_1+n_2-2}$$

donde

$$\begin{aligned} S_p^2 &= \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} \\ &= \frac{1}{n_1+n_2-2}\left(\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2\right), \end{aligned}$$

S_1^2 y S_2^2 son las varianzas muestrales de las *X´s* y las *Y´s* respectivamente, es decir

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2 \tag{2}$$

$$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2 \tag{3}$$

Instrucciones para el R:
`t.test(x = datos.x, y = datos.y, alternative = c("two.sided", "less", "greater"), mu = delta, paired = FALSE, var.equal = TRUE, conf.level = 0.95)`

Completar como antes, además: `var.equal = TRUE` para que asuma que las varianzas son iguales, `paired = FALSE` porque se trata de dos muestras independientes.

5.3 Caso varianzas desconocidas y no necesariamente iguales

En este caso tenemos el test de Welch que tiene nivel aproximado. Estadístico del test

$$\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \underset{\text{bajo } H_0}{\underset{\uparrow}{\overset{\text{aprox}}{\sim}}} t_K$$

donde

$$K = \left[\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}\right]$$

y S_1^2 y S_2^2 están definidos en (2) y (3) respectivamente.

Las instrucciones del R son las mismas que para el test anterior, pero en la opción `var.equal` completar con `var.equal = FALSE`.

6 Test para comparar varianzas de dos poblaciones normales independientes

Sean $X_1, \ldots, X_{n_1} \sim N\left(\mu_1, \sigma_1^2\right)$, e $Y_1, \ldots, Y_{n_2} \sim N\left(\mu_2, \sigma_2^2\right)$ dos muestras aleatorias independientes entre sí. Se quiere testear

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Estadístico del test

$$\frac{S_1^2}{S_2^2} \underset{\text{bajo } H_0}{\overset{}{\underset{\uparrow}{\sim}}} F_{n_1-1, n_2-1}$$

donde S_1^2 y S_2^2 están definidos en (2) y (3) respectivamente.

Instrucciones del R para este test:

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

7 Test asintótico (o aproximado) para comparar las medias de dos poblaciones

Sean $X_1, \ldots, X_{n_1}$ v.a.i.i.d. y sean $\mu_1 = E(X_1)$ y $\sigma_1^2 = Var(X_1)$, e $Y_1, \ldots, Y_{n_2}$ v.a.i.i.d. independientes de las anteriores y sean $\mu_2 = E(Y_1)$ y $\sigma_2^2 = Var(Y_1)$, donde n_1 y n_2 son números suficientemente grandes. Se quiere testear

$$H_0 : \mu_1 - \mu_2 = \delta \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \mu_1 - \mu_2 \neq \delta \qquad H_1 : \mu_1 - \mu_2 > \delta \qquad H_1 : \mu_1 - \mu_2 < \delta$$

Estadístico del test

$$\frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \underset{\text{bajo } H_0}{\underset{\uparrow}{\overset{\text{aprox}}{\sim}}} N(0,1)$$

donde S_1^2 y S_2^2 son las varianzas muestrales de las $X's$ y las $Y's$ respectivamente y están dadas en (2) y (3). Instrucciones para el R: `t.test`, se completa igual que en el Test 1.2.

8 Test asintótico (o aproximado) para comparar dos proporciones poblacionales

Sean $X_1, \ldots, X_{n_1} \sim Bi(1, p_1)$ v.a.i.i.d. e $Y_1, \ldots, Y_{n_2} \sim Bi(1, p_2)$ v.a.i.i.d. independiente de las anteriores, con n_1 y n_2 suficientemente grandes. muestra aleatoria con n suficientemente grande. Se quiere testear

$$H_0 : p_1 - p_2 = \delta \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : p_1 - p_2 \neq \delta \qquad H_1 : p_1 - p_2 > \delta \qquad H_1 : p_1 - p_2 < \delta$$

Estadístico del test

$$\frac{\widehat{p}_1 - \widehat{p}_2 - \delta}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}} \underset{\text{bajo } H_0}{\underset{\uparrow}{\overset{\text{aprox}}{\sim}}} N(0,1)$$

donde $\widehat{p}_1 = \overline{X}$ y $\widehat{p}_2 = \overline{Y}$ son, respectivamente, las proporciones de éxitos en la primer y segunda muestra. Instrucciones en el R

```
prop.test(x = c(cantidad.de.exitos1,cantidad.de.exitos2), n = c(n1,n2),
 alternative = c("two.sided", "less", "greater"), conf.level = 0.95, correct = TRUE)
```

Completar con los valores `cantidad.de.exitos1` $= \widehat{p}_1 n_1$, `cantidad.de.exitos2` $= \widehat{p}_2 n_2$, `n1` $= n_1$,`n2` $= n_2$, `alternative`, `conf.level` de interés. La opción `correct = TRUE` indica que se utilice corrección por continuidad, `FALSE` indica que no se utilice.

9 Test para comparar medias de dos muestras apareadas

Sean $(X_1, Y_1), \ldots, (X_n, Y_n)$ una muestra aleatoria de los resultados de dos mediciones realizadas sobre la misma unidad experimental (o que puede considerarse la misma unidad experimental). Supongamos que las diferencias $D_i = X_i - Y_i \sim N(\mu_D, \sigma_D^2)$ independientes entre sí. Se quiere testear

$$H_0 : \mu_D = \delta \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \mu_D \neq \delta \qquad H_1 : \mu_D > \delta \qquad H_1 : \mu_D < \delta$$

Estadístico del test

$$\frac{\overline{D} - \delta}{\sqrt{\frac{S_D^2}{n}}} = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{S_D^2}{n}}} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} t_{n-1}$$

donde S_D es el desvío estándar de las D_i, es decir

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \overline{D})^2.$$

Instrucciones para el R: iguales que las del Test 5.2, pero con la opción `paired = TRUE`. Otra posibilidad es construir un nuevo vector con las diferencias D_i y usar el `t.test` para una muestra.

10 Tests no paramétricos para una y dos muestras

10.1 Test del signo para la mediana de una población

Sea $X_1, \ldots, X_n \sim F$ una muestra aleatoria donde F es una distribución continua. Sea $\widetilde{\mu}$ la mediana de F. Se quiere testear

$$H_0 : \widetilde{\mu} = m_0 \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \widetilde{\mu} \neq m_0 \qquad H_1 : \widetilde{\mu} > m_0 \qquad H_1 : \widetilde{\mu} < m_0$$

Sean $D_i = X_i - m_0$. El estadístico del test es $U =$ la cantidad de D_i positivos.

$$U \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} Bi\left(n - k, \frac{1}{2}\right).$$

donde k es la cantidad de D_i iguales a cero que hay en la muestra. Instrucciones en R:

```
library(BSDA)
sign.test(x = datos, md = m_0, alternative = c("two.sided", "less", "greater"), conf.level = 0.95)
```

Completar el vector de `datos`, `m_0` $= m_0$ y la hipótesis alternativa de interés.

10.2 Test de rangos signados de Wilcoxon para la mediana de una población

Sea $X_1, \ldots, X_n \sim F$ una muestra aleatoria y F una distribución continua y simétrica. Sea $\widetilde{\mu}$ la mediana de F. Se quiere testear

$$H_0 : \widetilde{\mu} = m_0 \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \widetilde{\mu} \neq m_0 \qquad H_1 : \widetilde{\mu} > m_0 \qquad H_1 : \widetilde{\mu} < m_0$$

Sean $D_i = X_i - m_0$. Luego se ordenan en orden creciente las observaciones $|D_i|$ y se las ranquea. El estadístico del test es

$$W = \text{ suma de los rangos de las } D_i \text{ positivas}$$

Puede calcularse (con la tabla apropiada) el p-valor exacto de este test. Por default, el R calcula el p-valor exacto cuando el tamaño de muestra es menor a 50 y no hay observaciones repetidas. En caso contrario, calcula el p-valor aproximado. La instrucción de R:

```
wilcox.test(x = datos, alternative = c("two.sided", "less", "greater"), mu = m_0, paired = FALSE, exact
= NULL, correct = TRUE, conf.int = FALSE, conf.level = 0.95)
```

Como siempre, en `datos` va el vector con las observaciones, `m_0` $= m_0$, y hay que completar con el tipo de hipótesis alternativa de interés, `exact = TRUE` indica que compute el p-valor exacto, `correct = TRUE` indica que use una corrección por continuidad en el cálculo del p-valor aproximado.

Cuando se tiene una muestra apareada $(X_1, Y_1), \ldots, (X_n, Y_n)$ para testear si las medianas de ambas poblaciones coinciden se puede aplicar el test de rangos signados de Wilcoxon o el test del signo sobre las $D_i = X_i - Y_i$.

10.3 Test de Mann-Whitney-Wilcoxon para dos muestras independientes

Sean $X_1, \ldots, X_{n_1} \sim F_1$, e $Y_1, \ldots, Y_{n_2} \sim F_2$ dos muestras aleatorias independientes entre sí, donde F_1 y F_2 son dos distribuciones continuas no necesariamente normales. Se quiere testear

$$H_0 : F_1 = F_2$$
$$H_1 : F_1 \neq F_2.$$

Para construir el estadístico se juntan ambas muestras y se ranquean las observaciones, a esos números se los llaman rangos de las observaciones. El estadístico del test es la suma de los rangos del grupo correspondiente al grupo con menor cantidad de observaciones (digamos m observaciones, que será n_1 o n_2, el menor de ambos) menos la cantidad $\frac{m(m+1)}{2}$. Puede calcularse (con la tabla apropiada) el p-valor exacto de este test, por default el R lo hace si $n_1 + n_2$ es menor a 50 y no hay observaciones empatadas. Si no calcula el p-valor aproximado. Instrucciones de R:

```
wilcox.test(x = datos.x, y = datos.  y, alternative = c("two.sided", "less", "greater"), paired = FALSE,
exact = NULL, correct = TRUE, conf.int = FALSE, conf.level = 0.95)
```

11 Test de normalidad para una muestra

Sean $X_1, \ldots, X_n \sim F$ v.a.i.i.d. Se quiere testear

$$H_0 : F = N\left(\mu, \sigma^2\right) \text{ para algún valor } \mu \in \mathbb{R} \text{ y } \sigma > 0.$$
$$H_1 : \text{no existe ningún valor } \mu \in \mathbb{R} \text{ y } \sigma > 0 \text{ para los que valga } F = N\left(\mu, \sigma^2\right)$$

Un test posible para decidir entre estas hipótesis es el Test de Shapiro Wilk. La instrucción de R: `shapiro.test`