## Estadística I
## Tema 8: Contraste de hipótesis.

### POBLACIÓN NORMAL: CONTRASTE DE HIPÓTESIS PARA LA MEDIA

**1)** Para la realización de este ejercicio se necesitan los siguientes paquetes: `RcmdrPlugin.TeachingDemos` y `tkrplot`. El primero de ellos se carga a través del menú de `Rcmdr`:
`Herramientas ->Cargar plugin(s) de Rcmdr`
En la ventana de diálogo, elige `RcmdrPlugin.TeachingDemos`. `R` pedirá permiso para reiniciar `Rcmdr` (Aceptar), y añadirá la opción `Demostraciones` al menú. El segundo paquete, `tkrplot`, se cargará de forma automática.
Para ver cómo varían la distribución del estadístico de contraste bajo las hipótesis nula y alternativa, así como las regiones de aceptación y rechazo, en el contraste de hipótesis

$$H_0: \ \mu \leq \overbrace{\mu_0}^{\text{aquí}=0} \qquad \text{contra} \quad H_1: \ \mu > \mu_0$$

en función del nivel de significación $\alpha$, el tamaño de muestra $n$ y la varianza poblacional <u>conocida</u> $\sigma^2$, vete a:
`Demostraciones ->Potencia del test`
En la parte superior de la ventana gráfica apaprecerán las siguientes cantidades:

`se` = (error estándar) desviación típica de $\bar{X}$, $\frac{\sigma}{\sqrt{n}}$
$z^*$ = valor crítico del contraste, $z_\alpha \frac{\sigma}{\sqrt{n}}$
`power` = potencia del contraste cuando $\mu_1 =$ `diff`, $P(\text{rechazar } H_0 | \mu = \mu_1)$
`n` = tamaño de muestra
`sd` = desviación típica de la población, $\sigma$
`diff` (= `True Difference`)= la diferencia $\mu_1 - \mu_0$; aquí $\mu_1 - \mu_0 = \mu_1$ porque $\mu_0 = 0$
`alpha` = nivel de significación del contraste, $\alpha$

a) Aumenta el tamaño de muestra, manteniendo fijos los valores de las otras variables. ¿Qué pasa con el error estándar, el valor crítico y la potencia? ¿Aumentan o disminuyen?

b) Aumenta el valor de la desviación típica poblacional $\sigma$, manteniendo fijos los valores de las otras variables.
¿Qué pasa con el error estándar, el valor crítico y la potencia?
Disminuye el valor de la desviación típica poblacional $\sigma$, manteniendo fijos los valores de las otras variables.
¿Qué pasa con el error estándar, el valor crítico y la potencia?

c) Aumenta la diferencia entre $\mu_0$ y $\mu_1$, manteniendo fijos los valores de las otras variables.
¿Qué pasa con la potencia?

Disminuye la diferencia entre $\mu_0$ y $\mu_1$, manteniendo fijos los valores de las otras variables.
¿Qué pasa con la potencia?

d) Aumenta el nivel de significación, manteniendo fijos los valores de las otras variables.
¿Qué pasa con el valor crítico y la potencia?
Disminuye el nivel de significación, manteniendo fijos los valores de las otras variables.
¿Qué pasa con el valor crítico y la potencia?

**2)** Carga el conjunto de datos `trees` del paquete `datasets` ejecutando la siguiente instrucción:

```
data(trees, package = "datasets")
```

Además, si escribes en la `Ventana de instrucciones`:

```
attach(trees)
```

vas a poder referirte a las variables de este conjunto de datos por sus nombres sin tener que escribir delante del nombre `trees$`.
La variable de interés es `Height`.

a) Comprueba si la hipótesis de normalidad de la población de `Heights` es aceptable examinando el histograma de la muestra:

```
hist(Height)
```

b) Para un nivel de significación $\alpha = 0{,}1$ contrasta las hipótesis

$$H_0: \ \mu \leq 78 \quad \text{contra} \quad H_1: \ \mu > 78$$

```
t.test(x=Height,alternative="greater",mu=78,conf.level=0.90)
```

Expresa las hipótesis nula y alternativa con palabras (explicando lo que establece cada una).
Indica el número de grados de libertad de la distribución $t$ del estadístico del contraste, el valor del estadístico en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

c) Indica qué hipótesis son necesarias para justificar los cálculos realizados en el apartado b).

d) Repite el apartado b) para $H_1: \ \mu < 78$ (escribe `less` en vez de `greater`).

e) Repite el apartado b) para $H_1: \ \mu \neq 78$ (escribe `two.sided` en vez de `greater`).

## POBLACIÓN NORMAL: CONTRASTE DE HIPÓTESIS PARA LA VARIANZA

**3)** Utiliza el mismo conjunto de datos que en el ejercicio 2).

a) Para un nivel de significación $\alpha = 0{,}05$ contrasta las hipótesis

$$H_0: \ \sigma^2 \leq 20 \quad \text{contra} \quad H_1: \ \sigma^2 > 20$$

```
n=length(Height)
pop.var.underH0=20
sample.var=var(Height)
test.statistic=(n-1)*sample.var/pop.var.underH0
quantile=qchisq(p=0.05,df=n-1,lower.tail=FALSE)
p.value=pchisq(q=test.statistic,df=n-1,lower.tail=FALSE)
```

Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

b) Para un nivel de significación $\alpha = 0{,}05$ contrasta las hipótesis

$$H_0: \ \sigma \geq 5 \quad \text{contra} \quad H_1: \ \sigma < 5$$

```
n=length(Height)
pop.var.underH0=25
sample.var=var(Height)
test.statistic=(n-1)*sample.var/pop.var.underH0
quantile=qchisq(p=0.95,df=n-1,lower.tail=FALSE)
p.value=pchisq(q=test.statistic,df=n-1,lower.tail=TRUE)
```

Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

## DOS POBLACIONES NORMALES: CONTRASTE DE HIPÓTESIS PARA LA DIFERENCIA DE MEDIAS Y PARA EL COCIENTE DE VARIANZAS

**4)** Genera dos muestras aleatorias simples independientes a partir de $X_1 \sim N(\mu_1 = 5, \sigma_1^2 = 4^2)$, con tamaño $n_1 = 15$, y de $X_2 \sim N(\mu_2 = 5, \sigma_2^2 = 5^2)$, con tamaño $n_2 = 20$, ejecutando:

```
x1=rnorm(n=15,mean=5,sd=4)
x2=rnorm(n=20,mean=5,sd=5)
```

a) Para un nivel de significación $\alpha = 0{,}1$ contrasta las hipótesis

$$H_0: \ \mu_1 - \mu_2 \leq 1 \quad \text{contra} \quad H_1: \ \mu_1 - \mu_2 > 1$$

```
t.test(x=x1,y=x2,alternative="greater",mu=1,conf.level=0.90,var.equal=FALSE)
```

Obtén el valor del estadístico del contraste en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

b) Repite el apartado a) para $H_1 : \mu_1 - \mu_2 < 1$.

c) Repite el apartado a) para $H_1 : \mu_1 - \mu_2 \neq 1$.

d) Para un nivel de significación $\alpha = 0{,}05$ contrasta las hipótesis

$$H_0 : \ \sigma_1^2/\sigma_2^2 \leq 2 \quad \text{contra} \quad H_1 : \ \sigma_1^2/\sigma_2^2 > 2$$

```
var.test(x=x1,y=x2,ratio=2,alternative="greater",conf.level=0.95)
```

Obtén el valor del estadístico del contraste en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

e) Repite el apartado d) para $H_1 : \sigma_1^2/\sigma_2^2 < 2$.

f) Repite el apartado d) para $H_1 : \sigma_1^2/\sigma_2^2 \neq 2$.

**5)** Carga los siguientes datos de tasas de inflación en países desarrollados ($X_1$) y en vías de desarrollo ($X_2$) en la `Ventana de Instrucciones` ejecutando:

```
developed=c(3.99, 4.07, 3.70, 1.79, 5.30, 3.47, 2.39, 3.33, 4.14, 3.11)
developing=c(4.73, 5.01, 5.07, 4.66, 4.49, 4.00, 4.33, 5.14, 3.15, 3.46)
```

a) Comprueba si la hipótesis de normalidad en ambas poblaciones de es aceptable examinando el histograma de las muestras:

```
par(mfrow=c(1,2))
hist(developed,breaks=2)
hist(developing,breaks=2)
```

(El comando `par(mfrow=c(1,2))` permite presentar dos figuras en la misma ventana gráfica).

b) Para un nivel de significación $\alpha = 0{,}05$ contrasta la igualdad de varianzas

$$H_0 : \ \sigma_1^2/\sigma_2^2 = 1 \quad \text{contra} \quad H_1 : \ \sigma_1^2/\sigma_2^2 \neq 1$$

```
var.test(x=x1,y=x2,ratio=1,alternative="two.sided",conf.level=0.95)
```

Obtén el valor del estadístico del contraste en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

c) A partir del resultado obtenido en el apartado anterior (fija `var.equal` igual a `TRUE` si las varianzas poblacionales se pueden considerar iguales e igual a `FALSE` si no), contrasta las siguientes hipótesis para $\alpha = 0{,}01$

$$H_0 : \ \mu_1 - \mu_2 = 0 \quad \text{contra} \quad H_1 : \ \mu_1 - \mu_2 \neq 0$$

```
t.test(x=x1,y=x2,alternative="two.sided",mu=0,conf.level=0.99,var.equal=...)
```

Obtén el valor del estadístico del contraste en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula. ¿La media de la tasa de inflación en los países desarrollados es la misma que la de los países en vías de desarrollo?

d) Indica todas las hipótesis necesarias para justificar los cálculos realizados en los apartados b) y c).

## MUESTRAS GRANDES: UNA POBLACIÓN (CONTRASTE DE HIPÓTESIS PARA LA PROPORCIÓN), DOS POBLACIONES (CONTRASTE DE HIPÓTESIS PARA LA DIFERENCIA DE PROPORCIONES)

6) Genera dos muestras aleatorias simples independentes a partir de $X_1 \sim Bernoulli(p_1 = 0{,}3)$, con tamaño $n_1 = 40$, y de $X_2 \sim Bernoulli(p_2 = 0{,}4)$, con tamaño $n_2 = 50$, ejecutando:

```
x1=sample(x=c(0,1),size=40,replace=TRUE,prob=c(0.7,0.3))
x2=sample(x=c(0,1),size=50,replace=TRUE,prob=c(0.6,0.4))
```

a) Para un nivel de significación $\alpha = 0{,}05$ contrasta las hipótesis (sin corrección por continuidad, `correct = FALSE`),

$$H_0 : \ p_1 \leq 0{,}4 \quad \text{contra} \quad H_1 : \ p_1 > 0{,}4$$

```
prop.test(sum(x1),n=length(x1), p = 0.4, alternative = "greater",
conf.level = 0.95, correct = FALSE)
```

O alternativamente,

```
z.test(x=x1,stdev=sqrt(0.4*(1-0.4)),alternative="greater",
mu=0.4, conf.level=0.95)
```

Obtén el valor del estadístico del contraste en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

b) Para un nivel de significación $\alpha = 0{,}05$ contrasta las hipótesis

$$H_0 : \ p_1 - p_2 = 0 \quad \text{contra} \quad H_1 : \ p_1 - p_2 \neq 0$$

```
prop.test(c(sum(x1),sum(x2)),n=c(length(x1),length(x2)), p = NULL,
alternative = "two.sided",conf.level = 0.95, correct = FALSE)
```

Obtén el valor del estadístico del contraste en la muestra y el p-valor.
Justifica si se rechaza o no la hipótesis nula y establece la conclusión final con palabras.

c) ¿Qué teorema permite llevar a cabo los contrastes de los apartados a) y b)? ¿Bajo qué hipótesis sobre $n_1$ and $n_2$?