

### Elementos de un contraste de hipótesis

#### Hipótesis nula $H_0$

hipótesis que estamos dispuestos a aceptar si no encontramos evidencia suficiente de la hipótesis alternativa. Suele plantearse en términos de "no hay diferencia".

#### Hipótesis alternativa $H_1$

aquella de la que buscamos evidencia en nuestro estudio

#### Contraste unilateral

contraste en el que la hipótesis alternativa viene definida por un  $>$  o un  $<$ : que un parámetro sea mayor que un valor dado, que un parámetro sobre una población sea menor que el mismo parámetro sobre otra población, ...

#### Contraste bilateral

contraste en el que la hipótesis alternativa viene definida por un  $\neq$ : que un parámetro sea diferente de un valor dado, que un parámetro sobre una población sea diferente del mismo parámetro sobre otra población, ...

#### Error de tipo I

Falso positivo: concluir que la hipótesis alternativa es verdadera cuando en realidad es falsa.

#### Nivel de significación alpha

la probabilidad de cometer un error de tipo I, es decir, de la  $H_1$  sea falsa. Suele fijarse en 5%

#### Nivel de confianza (1-alpha)

la probabilidad de no cometer un error de tipo I

#### Error de tipo II

Falso negativo: concluir que la hipótesis alternativa es falsa cuando en realidad es verdadera.

### Elementos de un contraste de hipótesis (cont)

#### Potencia (1-b)

la probabilidad de no cometer un error de tipo II.

#### Estadístico de contraste

el valor que se calcula a partir de la muestra obtenida en el estudio y que se usará para tomar la decisión en el contraste planteado.

#### p-valor

valor de probabilidad, por lo que oscila entre 0 y 1. El p-valor nos muestra la probabilidad de haber obtenido un resultado satisfactorio suponiendo que la hipótesis nula  $H_0$  es cierta. Es decir, lo utilizamos para saber si rechazamos o no  $H_0$ .

#### Intervalo de confianza

intervalo en el que, en el  $(1-\alpha) \times 100\%$  de las ocasiones, contiene el valor del estadístico del contraste

#### Regla de decisión

rechazamos la hipótesis nula en favor de la alternativa con un nivel de significación  $\alpha$  cuando el p-valor es menor o igual que el nivel de significación  $\alpha$ .

### Contraste de hipótesis de bondad de ajuste

#### 1. Fijar la familia de distribuciones teóricas

Si la familia es la Bernoulli, el parámetro es  $p$ : la probabilidad poblacional de éxito.

Si la familia es la Poisson, el parámetro es  $\lambda$ : la esperanza.

Si la familia es la binomial, los parámetros son  $n$  y  $p$ : el tamaño de las muestras y la probabilidad de éxito, respectivamente.

Si la familia es la normal, los parámetros son  $\mu$  y  $\sigma$ : la esperanza y la desviación típica, respectivamente

### Contraste de hipótesis de bondad de ajuste (cont)

2. Si el diseño del experimento no fija sus valores, tendremos que **estimar** a partir de la muestra los valores de los parámetros que mejor se ajusten a nuestros datos. Ya hemos tratado este asunto en el Notebook 1 con la función `fitdist`, pero también existe una librería llamada `library(MASS)` que tiene la función `fitdistr` que nos estima directamente los parámetros de todas las distribuciones vistas anteriormente según la familia. Tenemos que saber, a priori, que distribución se ajusta mejor a nuestro conjunto de datos.

```
library(MASS)
fitdistr(x, densfun="poisson")
```

3. Realizar el contraste y comprobar si rechazamos (o no) la **hipótesis nula** planteada. Se usa el **test de Kolmogorov-Smirnov (K-S)**. El test K-S se puede utilizar para saber si 1 muestra se ajusta a una distribución determinada o para comparar si dos muestras se ajustan a la misma distribución.

Su sintaxis básica es `ks.test(x, y, parámetros)` donde:

$x$  es la muestra de una variable continua.

$y$  puede ser un segundo vector, y entonces se contrasta si ambas muestras han sido generados por la misma distribución continua, o el nombre de la función de distribución (empezando con  $p$ ) que queremos contrastar, entre comillas; por ejemplo " `pno rmpa` para la distribución normal.

Los parámetros de la función de distribución si se ha especificado una; por ejemplo `mean=0, sd=1` para una distribución normal estándar. Si el p-valor es mayor que 0.05 NO podemos rechazar nuestra hipótesis nula  $H_0$ . Es decir, Sí provienen de la misma distribución de probabilidad (con mismos valores de los parámetros).

### Contraste de hipótesis de bondad de ajuste (cont)

Uso de gráficos para comparar una distribución con otra

```
boxplot(data1, data2, main="Boxplot
data1 vs data2")
#o
par(mfrow=c(1,2))
hist(data1,freq=FALSE,col="lightsalmon",main="Histograma",sub="Datos 1")
hist(data2,freq=FALSE,col="lightsalmon",main="Histograma",sub="Datos 2")
par(mfrow=c(1,1))
```

3.2. Si estamos trabajando con una distribución normal también podemos utilizar el test de Kolmogorov-Smirnov-Lilliefors (K-S-L). Al contrario que el test K-S, el test K-S-L nos dice si los datos provienen de una distribución normal **sin aportar los parámetros**.

```
library(nortest)
lillie.test(data)
```

### Contrastes de hipótesis medias paramétricos

#### Requisitos

Que las muestras procedan de poblaciones en las que las variables se distribuyan **normalmente**. Test de **Kolmogorov y Smirnov** o **KSL**

Que las **varianzas** en ambas poblaciones no difieran significativamente. La  $H_0$  en este caso es que las varianzas son iguales. **Test de Barlett**

```
bartlett.test(list(data1,data2))
```

Ahora se realiza el contraste de hipótesis de medias con el test t de Student.

### Contrastes de hipótesis medias paramétricos (cont)

```
t.test(x, y, mu=..., alternative=..., conf.level=..., paired=..., var.equal=...)
```

**mu** para especificar el valor de la media que queremos contrastar en un test de una media.

**paired** para indicar si en un contraste de dos medias usamos muestras independientes o emparejadas. El parámetro **paired** solo lo tenemos que especificar si llevamos a cabo un contraste de dos medias. En este caso, con **paired = TRUE** indicamos que las muestras son emparejadas, y con **paired = FALSE** (que es su valor por defecto) que son independientes

**var.equal** para indicar en un contraste de dos medias usando muestras independientes si las varianzas poblacionales son iguales (igualándolo a **TRUE**) o diferentes (igualándolo a **FALSE**, que es su valor por defecto).

**conf.level** es el nivel de confianza  $1-\alpha$ . En esta función, y en todas las que explicamos a lo largo de este capítulo, su valor por defecto, que no es necesario especificar, es 0.95, que corresponde a un nivel de confianza del 95%, es decir, a un nivel de significación  $\alpha=0.05$ .

```
alternative="two.sided" "lower" "greater"
```

```
t.test(x, mu=2)$p.value
```

```
t.test(x, mu=2)$conf.int
```

```
t.test(x, mu=2)$conf.int[1]
```

```
t.test(x, mu=2)$conf.int[2]
```

```
t.test(S, V, alternative="less", var.equal=TRUE)
```

Ejemplo con lontananza de sépalos de iris setosa y virginica

```
t.test(Hiosciamina, Hioscina, alternative="less", paired=TRUE)
```

Ejemplo test somníferos

### Contrastes de hipótesis medias paramétricos (cont)

```
t.test(bwt~smoke, data=birthwt, alternative="greater", paired=FALSE, var.equal=TRUE)
```

Ejemplo con birthweight, donde smoke puede ser 0 o 1

### Con no p mediana

#### Test de signos

permite contrastar si la mediana de una variable aleatoria cualquiera es un valor dado  $k$  estudiando la distribución de los signos de las diferencias entre este valor y los de una muestra (si la mediana fuera  $k$ , los números de diferencias positivas y negativas en muestras aleatorias seguirían distribuciones binomiales con  $p=0.5$ ).

contrasta si la mediana de una sola variable difiere de un valor elegido. Es decir,  $H_0: \mu = k$

$H_0: \mu = k$

NO se puede aplicar a dos muestras independientes

se puede aplicar a dos mue. En este caso, la hipótesis nula del contraste que realiza es que  $H_0$ : la mediana de las diferencias de las dos variables es 0.

```
library(BSDA)
```

```
SIGN.test(colesterol, md=220, alternative="two.sided", conf.level=0.95)
```

#### Test de Wilcoxon

Asume que las variables son simétricas, es decir, media=mediana

contrasta si la mediana de una sola variable difiere de un valor elegido. Es decir,  $H_0: \mu = k$

$H_0: \mu = k$

NO se puede aplicar a dos muestras independientes.

se puede aplicar a dos muestras emparejadas.

```
wilcox.test(colesterol, mu=220, alternative="two.sided", conf.level=0.95)
```

### Con no p mediana (cont)

```
wilcox.test(Hiosciamina, Hioscina, alternative="less", paired=TRUE)
```

En este caso R nos avisa de nuevo de que el p-valor no es exacto, pero esto no afecta a la conclusión dado que el p-valor es muy pequeño: rechazamos la hipótesis nula en favor de la alternativa y también concluimos con este test no paramétrico que la hioscina tiene un mayor efecto somnífero que la hiosciamina.

### Test de Mann-Whitney

No utiliza la varianza

NO se puede aplicar para una muestra.

NO se puede aplicar a dos muestras emparejadas.

se puede aplicar a dos muestras independientes.

```
wilcox.test(hijos.20, hijos.30, alternative="-two.sided",paired=FALSE)
```

C

By **julenx**  
[cheatography.com/julenx/](https://cheatography.com/julenx/)

Published 29th November, 2022.  
Last updated 24th November, 2022.  
Page 3 of 3.

Sponsored by **ApolloPad.com**  
Everyone has a novel in them. Finish Yours!  
<https://apollopad.com>