Introducción a los contrastes de hipótesis. Límites de confianza y pruebas estadísticas

[0011] DEFAD. Métodos de contraste de hipótesis y diseño de experimentos

00R Team

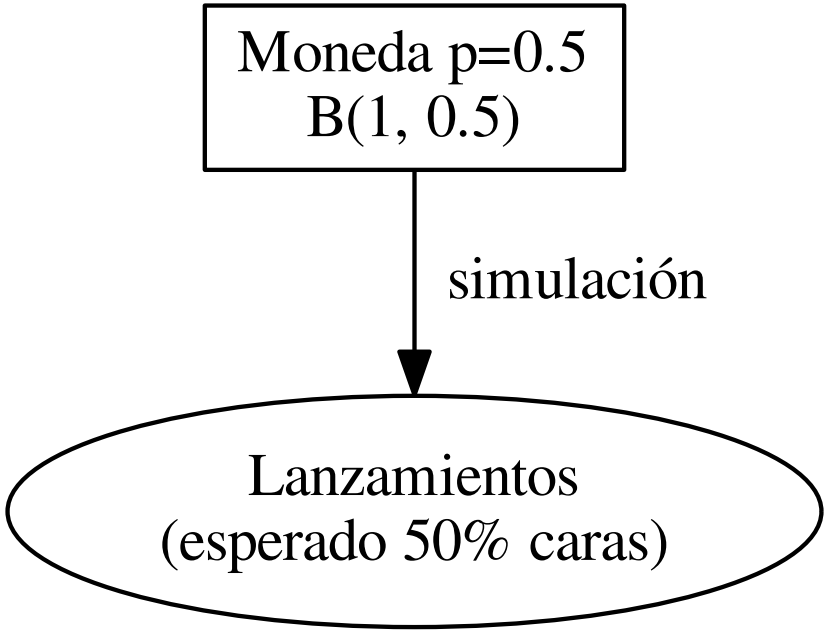
2014–15

1. Inferencia estadística
2. Contrastes de hipótesis
3. Prueba t de Student
4. Referencias y bibliografía

Inferencia estadística

Introducción. Simulación e inferencia

Lanzamiento de una moneda. Simulación

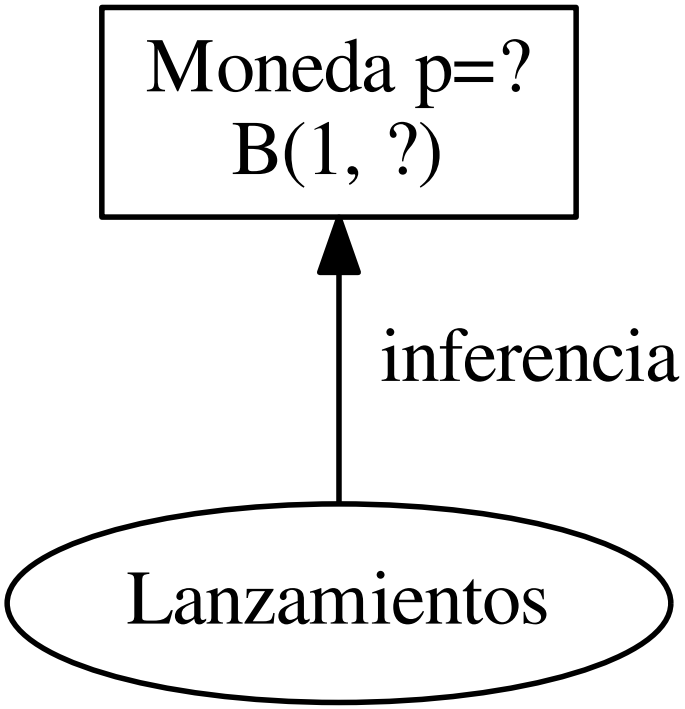


Lanzamiento de una moneda. Simulación

```r
# 100 lanzamientos con p = 0.5
table( rbinom( 100, 1, 0.5 ) )
```

```
##
##  0  1
## 59 41
```

Lanzamiento de una moneda. Inferencia



Lanzamiento de una moneda. Inferencia

```
moneda <- c( 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1,
             1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0 )
table( moneda ) # 8 / 25 = 0.32
```

```
## moneda
##  0  1
##  8 17
```

```
# binom.test( table(moneda), p = 0.5)
```

¿Es razonable pensar que la moneda no está trucada, es decir, que p=0.5? Contraste de hipótesis

Objetivo de la inferencia

La idea básica de las estadística es extrapolar, desde los datos recogidos, para llegar a conclusiones más generales sobre la población de la que se han recogido los datos.

Población y muestra

Población y muestra

Población: Conjunto de referencia sobre el cual van a recaer las observaciones

Muestra: Subconjunto de elementos de la población. Se suelen tomar muestras cuando es difícil o costosa la observación de todos los elementos de la población estadística

Censo: Decimos que realizamos un censo cuando se observan todos los elementos de la población estadística

Parámetros y estadísticos

Parámetros y estadísticos

Parámetro: Medida o característica de una población

Estadístico: Medida sobre una muestra cuyo objetivo es estimar o inferir características de una población (parámetro)

Si un **estadístico** se usa para aproximar un **parámetro** también se le suele llamar **estimador**[1].

[1] Diferentes test estadísticos aquí, Wikipedia.

Teorema del límite central

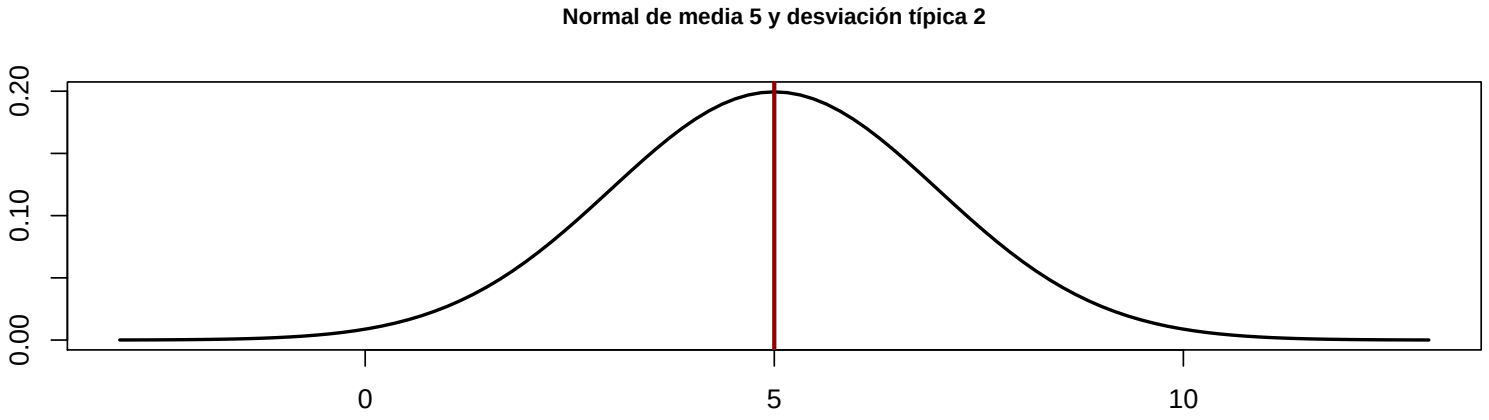
Teorema del límite central

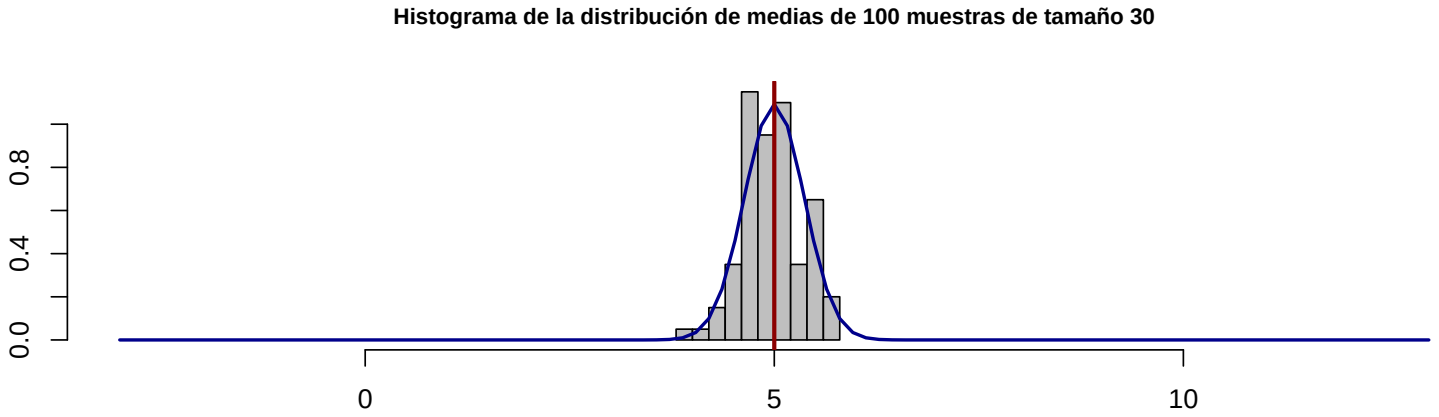
Si tenemos muestras independientes de una población, de tamaños suficientemente grandes, entonces las medias de estas muestras seguirán una distribución normal con la misma media que la de la población.

1. Dada una población con una distribución cualquiera
2. Aleatoriamente obtenemos varias muestras de esa población y calculamos sus medias
3. Construimos un histograma de la distribución de frecuencias de las **medias**
4. Esta distribución de medias sigue una distribución normal[2]

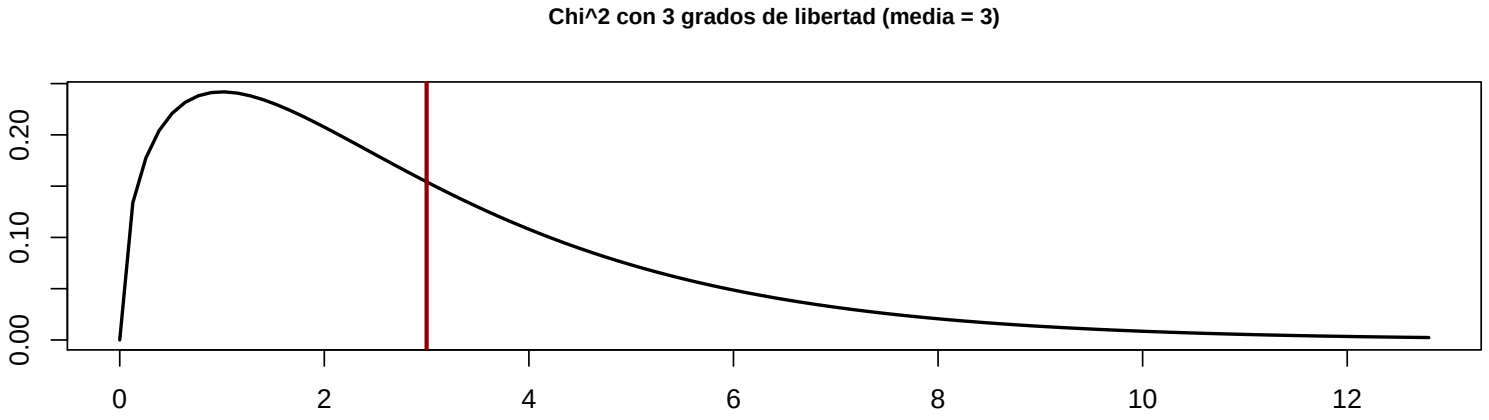
[2] Ver vídeo, López (2010)

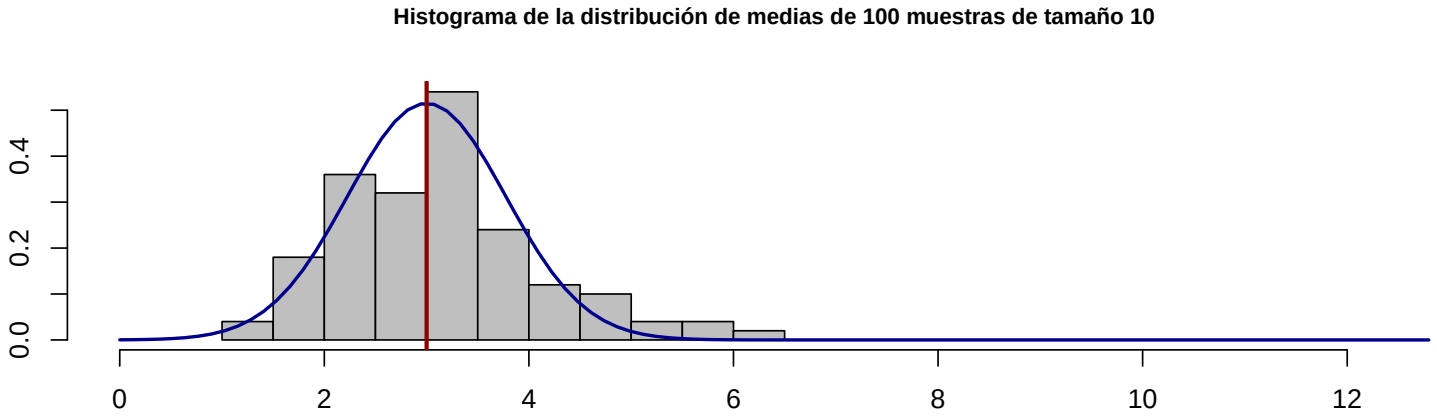
TCL sobre una población normal





TCL sobre una población χ^2





Contrastes de hipótesis

Contrastes de hipótesis

Contraste de hipótesis

H_0 : hipótesis nula

H_1 : hipótesis alternativa

1. Planteamos una hipótesis por defecto, que suele ser conservadora
2. Calculamos un valor a partir de los datos obtenidos (muestra)
3. Si el valor es razonable cuando la hipótesis nula es cierta, no hay razón para pensar que es falsa[3]

[3] pág. 108, Grima (2010)

Contraste sobre una media: ejemplo

Quiero saber si la media de un conjunto de valores normales x es diferente a 0.

$$H_0 : \mu = 0$$
$$H_1 : \mu \neq 0$$

```
norm <- c( 3.2005, 0.2608, 1.5324, 1.92, 1.4173, 0.0164,
           -0.9709, 1.8213 )
med <- mean(norm); sd <- sd(norm)
c( med, sd )
```

```
## [1] 1.149725 1.311472
```

Contraste sobre una media: ejemplo

Teorema: Si $X \sim N(\mu, \sigma^2)$ y $\bar{X}$ y S^2 son la media y la varianza en una muestra X de tamaño n entonces se cumple:

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

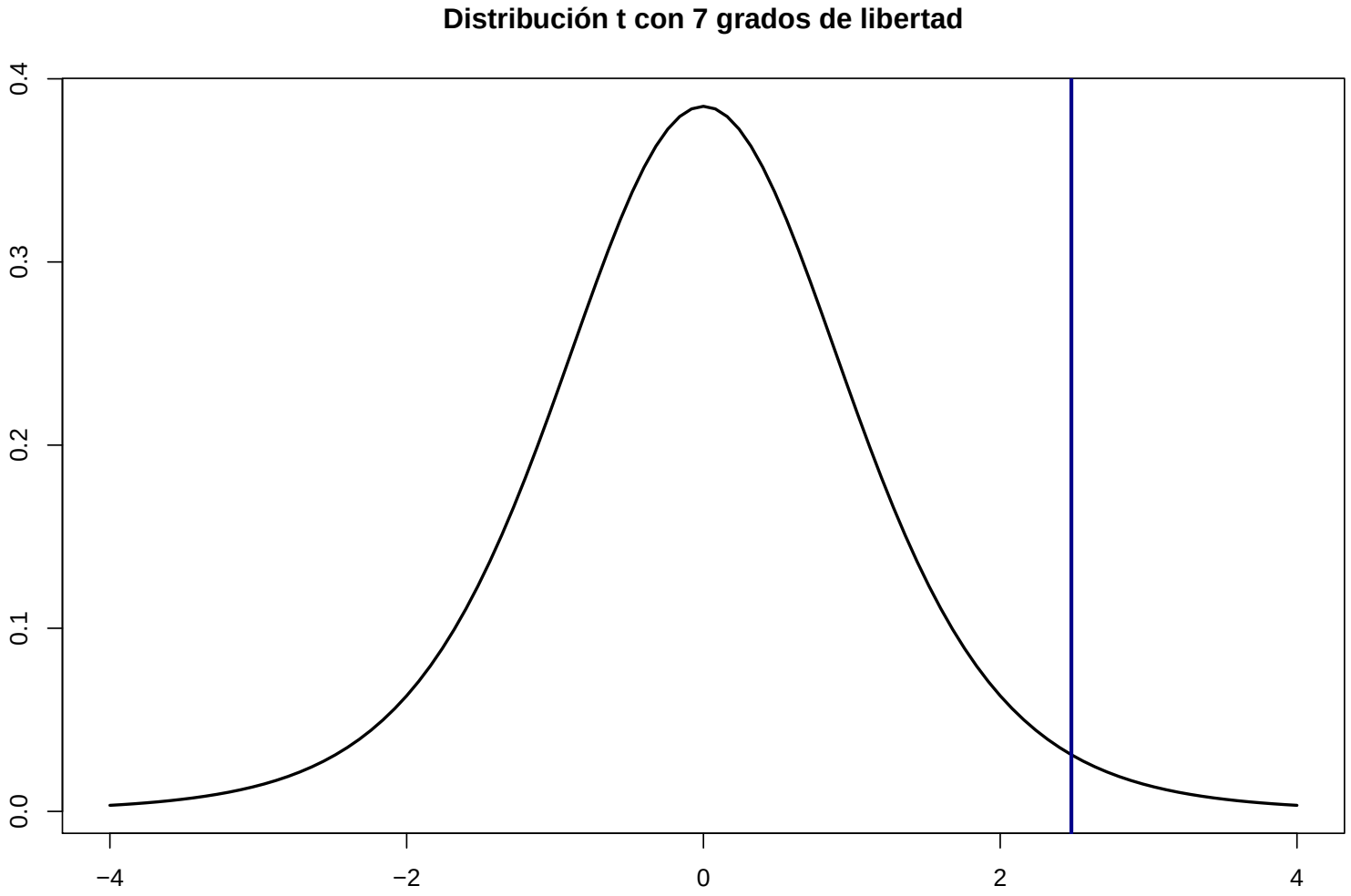
Contraste sobre una media: ejemplo

```
tstat <- (med - 0) / (sd/sqrt(8)) # estadístico t
gl    <- length(norm) - 1         # grados de libertad
tstat; gl
```

```
## [1] 2.47959
```

```
## [1] 7
```

Contraste sobre una media: ejemplo



P-valor

P-valor

p-valor: Probabilidad de obtener un resultado como el obtenido (o peor, en el sentido de la hipótesis alternativa), suponiendo cierta la hipótesis nula

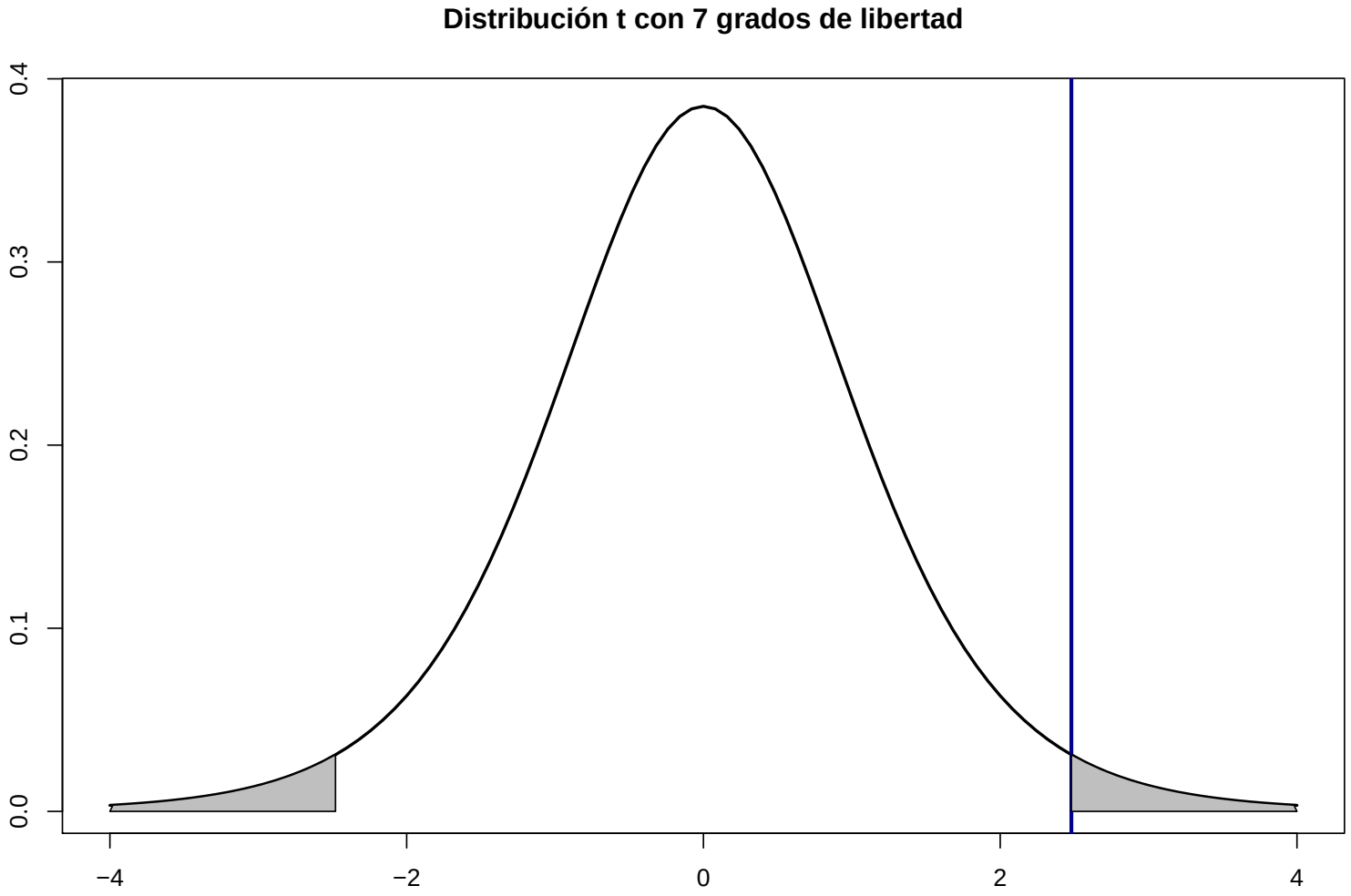
Si es menor a 0.05 se suele considerar un resultado *raro* bajo la hipótesis nula, así que, se rechaza esta hipótesis.

Contraste sobre una media: ejemplo

```
pval <- 2 * pt( -abs(tstat), gl ) # p-valor
pval
```

```
## [1] 0.04223546
```

Contraste sobre una media: ejemplo



Intervalos de confianza

Intervalos de confianza

Un intervalo de confianza es un rango de valores (calculado en una muestra) en el cual se encuentra el verdadero valor del parámetro con una probabilidad determinada.

- Nivel de confianza $1 - \alpha$: probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo
- Nivel de significación α: probabilidad de equivocarnos

Normalmente $1 - \alpha = 0.95$ ($\alpha = 0.05$)

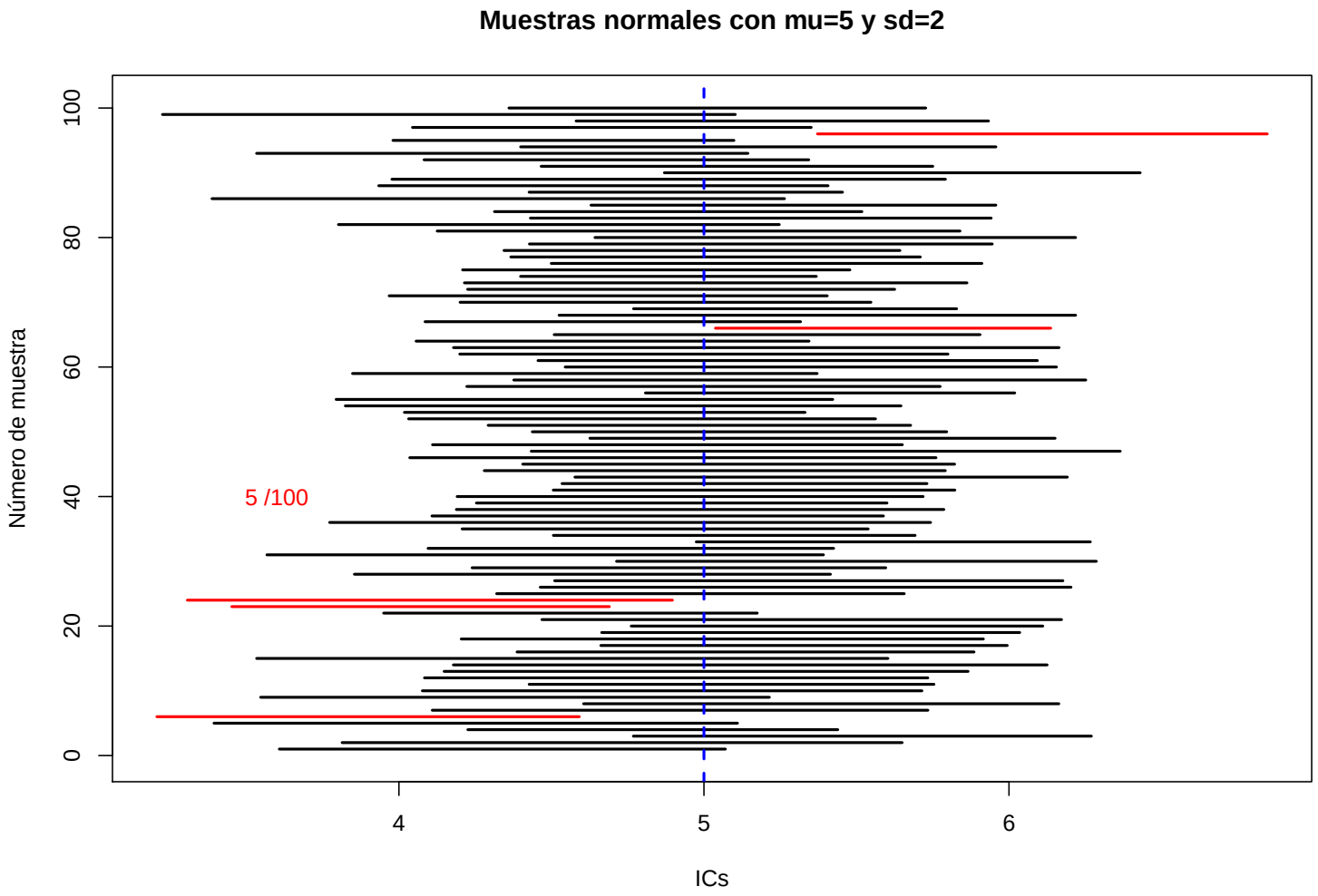
Intervalos de confianza

$$P\left(-t_{\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

Se dice que $\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$ es un intervalo de confianza al nivel $1-\alpha$ del parámetro μ.

Intervalos de confianza



Contraste sobre una media: ejemplo

```
norm <- c( 3.2005, 0.2608, 1.5324, 1.92, 1.4173, 0.0164,
           -0.9709, 1.8213 )
med <- mean(norm); sd <- sd(norm)
c( med, sd )
```

```
## [1] 1.149725 1.311472
```

```
tt   <- qt( 1 - 0.05/2, gl )
cint <- med + c(-tt, tt) * sd/sqrt(8)
cint
```

```
## [1] 0.05330683 2.24614317
```

Contraste sobre una media: ejemplo

```r
t.test( norm )
```

```
##
##  One Sample t-test
##
## data:  norm
## t = 2.4796, df = 7, p-value = 0.04224
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.05330683 2.24614317
## sample estimates:
## mean of x
##  1.149725
```

Potencia y errores

Potencia y errores

Si $p > 0.05$ se concluye diciendo que no se ha encontrado una *'diferencia significativa'* pero esto no implica que no haya diferencia; simplemente no la hemos encontrado.

Error de tipo II o β: Cuando 'afirmamos' que no hay diferencias ($p > 0.05$) y, en realidad, sí las hay. Falso negativo.

Error de tipo I o α: Cuando 'afirmamos' que sí hay diferencias ($p < 0.05$) y, en realidad, no las hay. Falso positivo.

La **potencia estadística** de un test es la probabilidad de encontrar diferencias cuando realmente existen. Es el complementario del error tipo II, $1 - \beta$.

Tabla errores tipo I y tipo II

	H_0 Verdadera	H_1 Verdadera
Aceptar H_0	**Verdadero positivo** $(1-\alpha)$	**Error de tipo II** (β o falso negativo)
Rechazar H_0	**Error de tipo I** (α o falso positivo)	**Verdadero negativo** $(1-\beta)$[4]

[4] Más información aquí, McDonald (2014)

Prueba t de Student

Introducción

Prueba t de Student

Contraste paramétrico para comparar las medias entre dos grupos. Tres tipos:

- Prueba t para una muestra
- Prueba t para dos muestras dependientes
- Prueba t para dos muestras independientes (con corrección de Welch si las varianzas son diferentes)

Contrastes paramétricos y no paramétricos

- Contrastes paramétricos
  Necesita (o asume) cierta información sobre la distribución de probabilidad de la población. Se decide sobre los parámetros.

- Contrastes no paramétricos
  No necesita información sobre la distribución de probabilidad de la población (libres de distribución).

Supuesto de normalidad

La normalidad se puede visualizar con los gráficos Q-Q. Para contrastarla podemos utilizar:

- El test de Shapiro-Wilk con `shapiro.test()`. Funciona bien con muestras pequeñas (menores a 50)
- El test de Kolmogorov-Smirnov con `ks.test()`. Contrasta distribuciones (no solo la normal)
- El test de Jarque-Bera con `jarque.bera.test()` del paquete `tseries`

La hipótesis nula es la hipótesis de normalidad, esto es, *no hay diferencias entre nuestra distribución y una distribución normal con esa media y esa desviación típica*.

Supuesto de homocedasticidad (HOV)

También llamada homogeneidad de varianzas. La hipótesis nula es: *la varianza es constante (no varía) en los diferentes grupos*. Para contrastarla podemos utilizar:

- El test F de Snedecor con `var.test()`. Solo dos grupos
- El test de Levene con `leveneTest()` del paquete `car`
- El test de Bartlett con `bartlett.test()`. Es mejor que Levene si los datos son normales (más robusto)
- El test de Fligner-Killen con `fligner.test()`. De los más robustos a la falta de normalidad

Reglilla: Balanceadas y no HOV: ok si $\frac{\widehat{S}^2_{Max}}{\widehat{S}^2_{Min}} < 3$ (si no hay balanceo, cambiar por un dos)

Prueba t para una muestra

Prueba t para una muestra: hipótesis

Utilizado cuando tenemos una variable de medida y un valor esperado para la media. Se supone normalidad de los datos (o muestra grande)[5].

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

```
t.test( x, mu = 0, alternative = "two.sided" )
```

[5] Más información aquí, McDonald (2014).

Prueba t dependiente

Prueba t dependiente: hipótesis

Utilizado cuando tenemos dos variables dependientes (p.e. sobre los mismos individuos). Es equivalente al de una muestra si tomamos la variable diferencia. Se supone normalidad de las diferencias (o muestra grande)[6]

$$H_0 : \mu_1 = \mu_2 \qquad H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 \neq \mu_2 \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

```
t.test( x, y, paired = TRUE )
```

[6] Más información aquí, McDonald (2014)

Ejemplo freshman: conjunto de datos para trabajar

¿Se modifica el BMI de los estudiantes de primer año entre septiembre y abril?

```
freshman <- read.table("http://people.sc.fsu.edu/~jburkardt/datasets/triola/freshman_15.csv",
               header = TRUE, sep = ",",
               col.names = c("Sex", "SeptWeight", "AprWeight", "SeptBMI", "AprBMI"))
head( freshman )
```

```
##   Sex SeptWeight AprWeight SeptBMI AprBMI
## 1   M         72        59   22.02  18.14
## 2   M         97        86   19.70  17.44
## 3   M         74        69   24.09  22.43
## 4   M         93        88   26.97  25.57
## 5   F         68        64   21.51  20.10
## 6   M         59        55   18.69  17.40
```

Ejemplo freshman: normalidad

```
qqnorm( freshman$SeptBMI - freshman$AprBMI )
qqline( freshman$SeptBMI - freshman$AprBMI )

shapiro.test( freshman$SeptBMI - freshman$AprBMI )
```

Ejemplo freshman: test t dependiente

```r
# t.test(freshman$SeptBMI - freshman$AprBMI)
t.test( freshman$SeptBMI, freshman$AprBMI, paired = TRUE )
```

```
##
##  Paired t-test
##
## data:  freshman$SeptBMI and freshman$AprBMI
## t = -2.9516, df = 66, p-value = 0.004374
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -0.7614044 -0.1469539
## sample estimates:
## mean of the differences
##               -0.4541791
```

Prueba t independiente

Prueba t independiente: hipótesis

Utilizado cuando tenemos dos variables independientes. Es equivalente a un ANOVA de una vía con dos categorías. Se supone normalidad (o muestra grande) y homocedasticidad u homogeneidad de varianzas (en caso contrario, corrección de Welch)[7]

$$H_0 : \mu_1 = \mu_2 \qquad H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 \neq \mu_2 \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

```
t.test( x, y, paired = FALSE, var.equal = TRUE )
```

[7] Más información aquí, McDonald (2014).

Ejemplo 2 freshman: conjunto de datos para trabajar

¿Difiere el BMI de abril de los estudiantes de primer año entre hombres y mujeres?

```
head( freshman, 6 )
```

```
##   Sex SeptWeight AprWeight SeptBMI AprBMI
## 1   M         72        59   22.02  18.14
## 2   M         97        86   19.70  17.44
## 3   M         74        69   24.09  22.43
## 4   M         93        88   26.97  25.57
## 5   F         68        64   21.51  20.10
## 6   M         59        55   18.69  17.40
```

```
# boxplot( freshman$AprBMI ~ freshman$Sex )
```

Ejemplo 2 freshman: normalidad

```
table( freshman$Sex ) # tamaños muestrales > 30 => Ok
shapiro.test( *vector BMI de abril de las mujeres* ) # p =
shapiro.test( *vector BMI de abril de las hombres* ) # p =
```

Ejemplo 2 freshman: homocedasticidad

```
fligner.test( freshman$AprBMI, freshman$Sex )
# leveneTest( freshman$AprBMI ~ freshman$Sex )
```

Ejemplo 2 freshman: prueba t independiente

```
t.test( freshman$AprBMI ~ freshman$Sex, var.equal = T )
```

```
##
##  Two Sample t-test
##
## data:  freshman$AprBMI by freshman$Sex
## t = -1.2802, df = 65, p-value = 0.205
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -2.873888  0.628638
## sample estimates:
## mean in group F mean in group M
##        21.94800        23.07062
```

Para reflexionar

Para reflexionar

```r
# Caso 1
x <- c( 150, 50, 31, 45 )
t.test( x, mu = 0 )

# Caso 2
x <- c( 0.03, -0.021, 0.038, 0.035, 0.034, 0.042 )
t.test( x, mu = 0 )
```

Referencias y bibliografía

Referencias y bibliografía

Referencias y bibliografía

Grima, P. (2010). *La certeza absoluta y otras ficciones: Los secretos de la estadística*. RBA.

López, F. J. B. (2010). *Bioestadística. Teorema del límite central - YouTube*. Retrieved from `https://www.youtube.com/watch?v=FcDcJnw00hk`

McDonald, J. (2014). *Handbook of Biological Statistics* (3rd ed.). Sparky House Publishing, Baltimore, Maryland. Retrieved from `http://www.biostathandbook.com/`