Sesión 8: Inferencia Estadística

EUSEBIO ANGULO SÁNCHEZ-HERRERA

2022/2023

ESCUELA SUPERIOR
DE INFORMÁTICA
CIUDAD REAL

Contenidos Principales

- Intervalos de confianza (CI) y Constraste de Hipótesis (HT)
  - Contraste de Hipótesis para una muestra
  - Contraste de Hipótesis para dos muestras
  - Reglas de significación
  - p-valor
- Intervalos de confianza (CI) y Contraste de Hipótesis (HT)
  - Media de población (μ)
  - Proporción de población (π)

Intervalos de confianza y Contraste de Hipótesis

- **Población:** muestra → medida

  muestra → medida

  …

  - <u>Ejemplo</u>: Longitudes…: 10.32mm, 10.03mm…
  - Errores de medición, selección aleatoria de la muestra, hacen que la variable actúe como un variable aleatoria.

Ejemplo Grosor de Aislamiento

>insulation=c(10.3,10.1,9.8,9.9,10.2,10.1,9.7,9.9,9.7,10.2)

t student no conocemos varianza

Intervalo de confianza para μ

- Suposiciones: $X_i \sim$ Distribución Normal, desconocido σ^2

TWO-SIDED CI: (L_L, L_U) $being$ $\begin{cases} L_L = \bar{X} - \epsilon \\ L_U = \bar{X} + \epsilon \end{cases}$; $\epsilon = t_{n-1, \frac{1+\gamma}{2}} \cdot \frac{S_c}{\sqrt{n}}$

Distribución T de Student's

- Longitud proporcional, del intervalo de confianza (directo) y exactitud de la estimación (indirecta)

t distribución

- > t.test(insulation)$conf.int

  [1] 9.833818 10.146182 la media entre estos dos valores

- attr(,"conf.level") por defecto es 0,85

  [1] 0.95

- > t.test(insulation, conf.level = 0.99)$conf.int se puede cambiar el nivel de confianza

  ... el intervalo es más grande crece para albergar más valores y fallar menos

- > t.test(insulation, conf.level = 0.90)$conf.int

  ...

- Estamos al % seguros que la media está entre… (Intervalo Confianza)

t.test()


```
> t.test(insulation, conf.level=0.95)

        One Sample t-test

data:  insulation
t = 144.7, df = 9, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  9.833818 10.146182
sample estimates:
mean of x
     9.99
```

significación = 1 - grado de confianza
significación = 0,05

estadístico - paso 3

grado libertad

menor que la significación rechazo H1 y acepto H0

unilateral - una region critica

cuando no ponemos nada entiende que es bilateral

estimación puntual

t.test() para CI

- Estadístico t es para ver si se encuentra en la región de aceptación de la H0 o de rechazo.
- Se obtiene el p-valor, si es menor que la significación (0,05) se rechaza la H0 y se acepta H1 (hipótesis alternativa)
- df son los grados de libertad, n-1
- El intervalo de confianza te dice que la media está entre 9.83 y 10.14 con un 95% de probabilidad.
- La media de x es 9.99

Test de hipótesis para μ

- ¿Podríamos asegurar la protección del aislamiento es 10 mm?

  $H_0 : \mu = 10$

  $H_1 : \mu \neq 10$ (two-sided)

bilateral

```
>t.test(insulation,mu=10,alt="two.sided")
```

One Sample t-test

data: insulation

t = -0.14484, df = 9, p-value = 0.888 > 0,05 significación -> no se puede rechazar H0

alternative hypothesis: true mean is not equal to 10 (two-side, por defecto)

95 percent confidence interval:

9.833818 10.146182

sample estimates: Decisión???

mean of x

9.99

Reglas basadas en α (por defecto 0.05)

- Rechazar H_0 cuando:
  - Estadístico está en la región de rechazo (α) o critica
  - P-valor ≤ α
  - (1- α)% El intervalo de confianza CI no contiene el valor de referencia para el parámetro

  $H_1 : \mu \neq 10$ (two-sided)

- Example:
  - t = -0.14484 in Rejection region?
  - p-value = 0.888 ≤ α?
  - 95% CI: 9.833818 10.146182 contains 10(=mu, H0)?

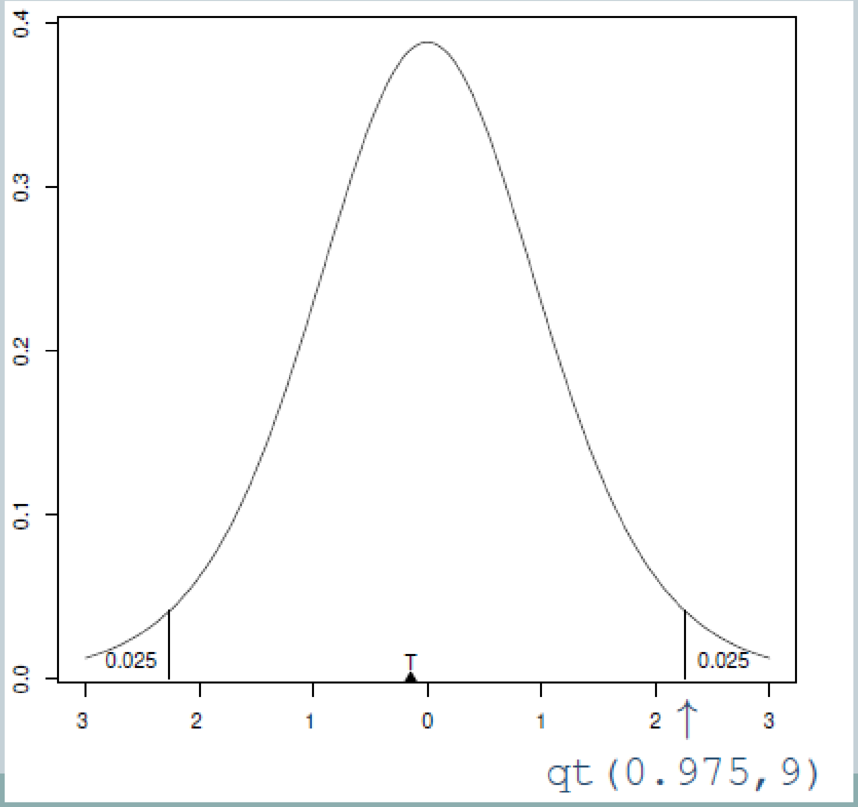
Test de Hipótesis bilateral

valor media   tipo contraste

>t.test(..., mu=..., alt=..., conf.level=...)

HT   CI

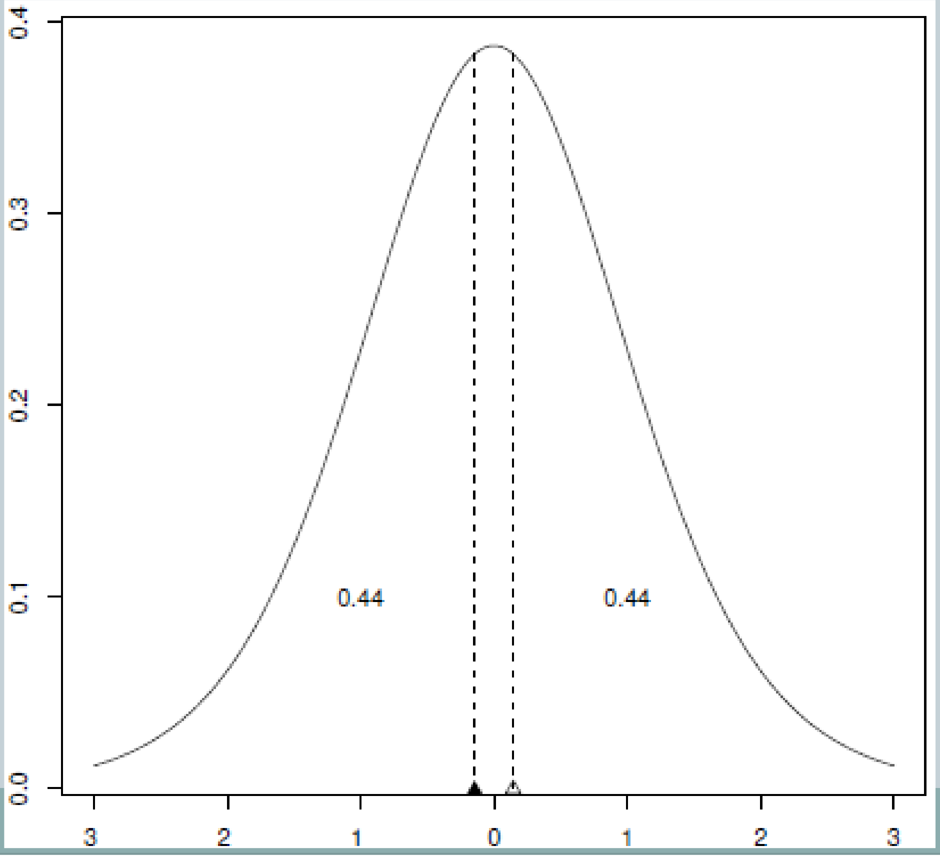
- Valor del estadística T y región de rechazo para α=0.05



Test de Hipótesis bilateral

- A partir de los datos de la muestra:

  p-value = P (obteniendo estas observaciones o cualquier otra más alejada de Ho | Ho verdadera)



Regla de decisión

- Rechazamos H0 cuando p-valor ≤ α

nivel de significación deseado α=0.05

Ejemplo:

```
...
t = -0.1448, df = 9, p-value = 0.888 ≰ α = 0.05
...
```

El test es claramente no significante ⟶ no podemos rechazar H0

Test de hipótesis para μ

>names(t.test(...)) ->...$conf.int ... $statistic ... $p.value #Parámetros del objeto

- >t.test(..., mu = ..., alt = ..., conf.level = ...)

HT CI

Test de hipótesis para μ

- Ejemplo: Protección del aislamiento menos de 10 mm

  H0 : $\mu \geq 10$

  H1 : $\mu < 10$

faltaria conf.level

```
>t.test(insulation, mu=10, alt="less")
```

unilateral

One Sample t-test

data: insulation

t = -0.1448, df = 9, p-value = 0.444

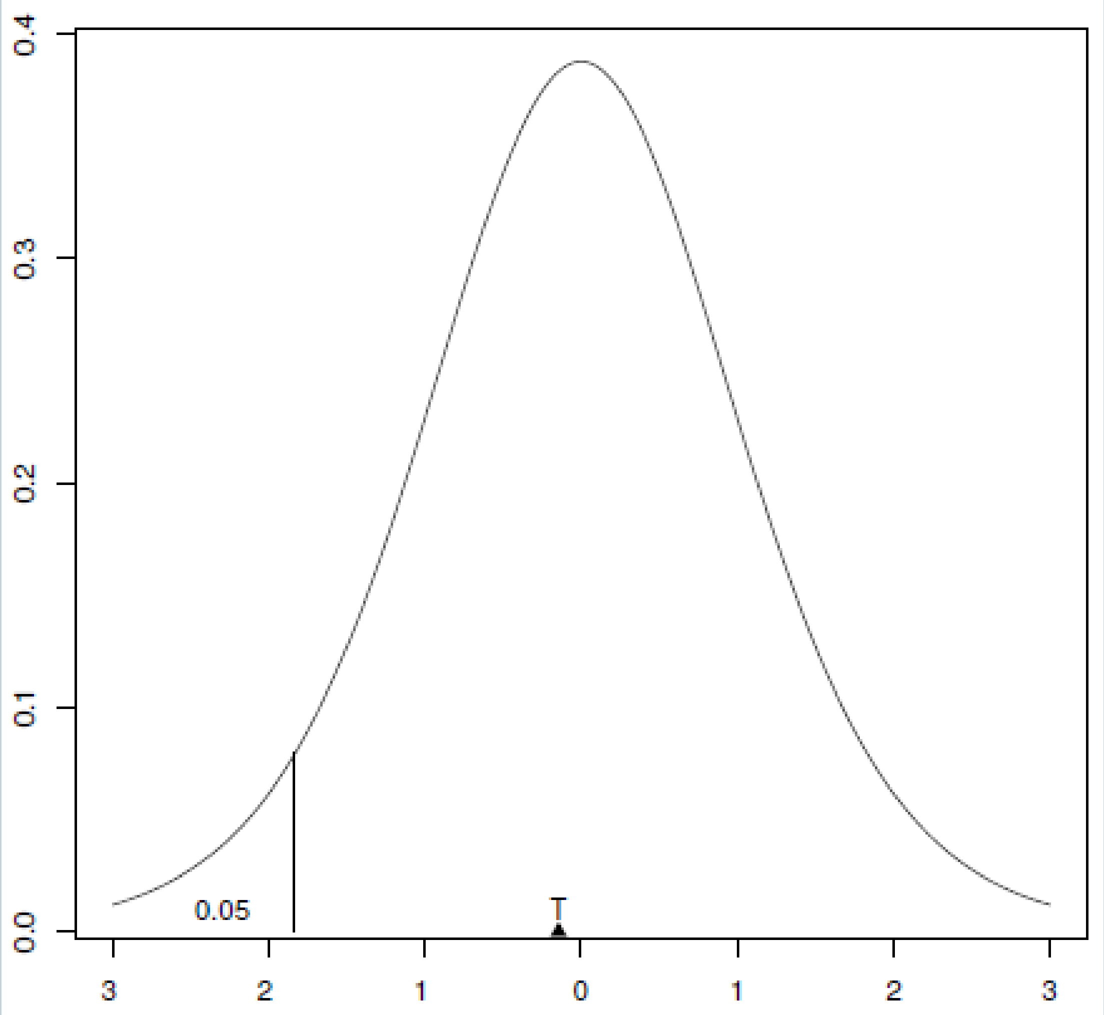
se acepta H0 no demuestras nada, no se puede aceptar H1 > 0,05

alternative hypothesis: true mean is less than 10

95 percent confidence interval:

-Inf 10.11656

Test de hipótesis para μ



Test de hipótesis para π (prop)

- Ejemplo de proporciones
  - Proporción de individuos a favor o en contra de algo.
  - Proporción de piezas defectuosas
  - …
- Población p desconocida, de una muestra $\hat{p}$ -> CI y HT
- Ejemplo:
- En una muestra de 100 partes … hay 3 defectuosos: $\hat{p} = \frac{3}{100}$
- En otra muestra … 5 defectuosos: $\hat{p} = \frac{5}{100}$

IC usando prop.test()

- En una muestra aleatoria de 500 familias que poseen una *smart-TV* en una ciudad, 340 están suscritas a Netflix. Calcule el intervalo de confianza con un 97% de confianza de la proporción de las familias suscritas a Netflix en la ciudad.

```
>prop.test(x=340, n=500, conf.level=0.97)
1-sample proportions test with continuity correction
data: 340 out of 500, null probability 0.5
X-squared = 64.082, df = 1, p-value = 1.193e-15
alternative hypothesis: true p is not equal to 0.5
97 percent confidence interval:
0.6322022 0.7243649
sample estimates:
p
0.68
```

puedes aceptar H1 y rechazar H0

Bilateral HT(p)

- Un proveedor de piezas asegura que la proporción de piezas defectuosas es del 2%. Se muestrearon 500 partes, hubo 17 defectuosas.

  H0 : p = 0.02

  H1 : p ≠ 0.02 significación por defecto = 0,05

  >prop.test(x=17,n=500,p=0.02,alt="two.sided")

  1-sample proportions test with continuity correction

  data: 17 out of 500, null probability 0.02

  X-squared = 4.3112, df = 1, p-value = 0.03786

  alternative hypothesis: true p is not equal to 0.02

  ... se rechaza H0 < 0,05

Bilateral HT(p)

- La prueba es significativa a 5% ⇒ rechazamos H0 ⇒

  ⇒ Proporción de defectuoso significativamente diferente del 2%.
- La prueba NO es significativa al 1% ⇒ no podemos rechazar H0.

Unilateral HT(p)

- Un proveedor de piezas asegura que la proporción de piezas defectuosas es menos del 2%. Se muestrearon 500 partes, hubo 17 defectuosas.

  H0 : p ≤ 0.02

  H1 : p > 0.02

  ```
  >prop.test(x=17,n=500,p=0.02,alt="greater")
  1-sample proportions test with continuity correction
  data: 17 out of 500, null probability 0.02
  X-squared = 4.3112, df = 1, p-value = 0.01893
  alternative hypothesis: true p is greater than 0.02
  ...
  ```

Ejercicio 1

- Una empresa de fabricación de piezas de engranaje ha tomado datos del diámetro de diez de ellas (suponer que sigue una distribución Normal):

Diámetros: 15.7   15.4   15.9   16.1   16.7   15.8   16.3   16.4   15.7   16

(a) Construir un intervalo de confianza para la media de la característica en estudio para un nivel de confianza del 90%.

(b) Contrastar la hipótesis de que la media de los diámetros de las piezas es menor que 16.2 a un nivel de significación α del 5%.

Ejercicio 2

- El servicio médico de una consultora informática decidió hacer un estudio acerca del estado dental de los empleados de la misma en una de sus fábricas, tomando con tal fin una muestra piloto de 160 individuos y observando que 37 de ellos presentaban algún diente cariado.

(a) Calcular el intervalo de confianza al 99% de la proporción de empleados sin problemas dentales.

(b) Contrastar la hipótesis de que la proporción de empleados con dientes cariados es igual a un 10% a un nivel de significación α del 1%.