

3 Inferencia simple

Student's t-tests

- Prueba hipotesis sobre las medias de poblaciones normales
- Muestras independientes

$$H_0 : \mu_1 = \mu_2$$

Varianzas identicas

- estadistico

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s\sqrt{1/n_1 + 1/n_2}}$$

100(1 - α)% confidence interval

$$\bar{y}_1 - \bar{y}_2 \pm t_{\alpha, n_1+n_2-2} s \sqrt{n_1^{-1} + n_2^{-1}}$$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- Welch's test

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}.$$

La hipótesis nula sigue t Student con g.l. $\nu = \left(\frac{c}{n_1 - 1} + \frac{(1 - c)^2}{n_2 - 1} \right)^{-1}$

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}.$$

- Test por pares

$$d_i = y_{1i} - y_{2i} \quad \text{Diferencias por pares}$$

$$H_0 : \mu' = 0.$$

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

$$100(1 - \alpha)\% \text{ confidence interval} \quad \bar{d} \pm t_{\alpha, n-1} s/\sqrt{n}$$

$$P(t \leq t_{\alpha, n-1}) = 1 - \alpha/2.$$

Tests no parametricos

- Wilcoxon Mann-Whitney rank sum
 - Calcula los rankings de las muestras
 - Aplica el t-statistic a los rankings

– Por pares:

$$d_i = y_{1i} - y_{2i}$$

- rank the absolute differences $|d_i|$
- El estadístico es la suma de los rangos de las diferencias positivas

Test de Independencia en tablas de contingencia

- Tabla de contingencia

Table 3.6: The general $r \times c$ table.

		y			
		1	...	c	
1		n_{11}	...	n_{1c}	$n_{1\cdot}$
2		n_{21}	...	n_{2c}	$n_{2\cdot}$
\vdots		\vdots	...	\vdots	\vdots
x		\vdots	...	\vdots	\vdots
r		n_{r1}	...	n_{rc}	$n_{r\cdot}$
		$n_{\cdot 1}$...	$n_{\cdot c}$	n

- Hipotesis nula: variables independientes

chi-squared test.

$$E_{jk} = n_{j.}n_{.k}/n.$$

estadístico

$$X^2 = \sum_{j=1}^r \sum_{k=1}^c \frac{(n_{jk} - E_{jk})^2}{E_{jk}}.$$

χ^2 -distribution with $(r - 1)(c - 1)$ degrees of freedom

- McNemar test

Table 3.7: Frequencies in matched samples data.

		Sample 1	
		present	absent
Sample 2	present	a	b
	absent	c	d

- Hipotesis nula: igual probabilidad de presencia

$$X^2 = \frac{(c - b)^2}{c + b}$$

χ^2 -distribution with a single degree of freedom.

Caso 1

Shortly after metric units of length were officially introduced in Australia in the 1970s, each of a group of 44 students was asked to guess, to the nearest metre, the width of the lecture hall in which they were sitting. Another group of 69 students in the same room was asked to guess the width in feet, to the nearest foot. The data were collected by Professor T. Lewis, and are given here in Table 3.1, which is taken from Hand et al. (1994). The main question is whether estimation in feet and in metres gives different results.

```
data("roomwidth", package = "HSAUR2")  
View(roomwidth)
```

- Conversion de los datos

```
> convert <- ifelse(roomwidth$unit == "feet", 1, 3.28)
```

- Sumario por unidad de medida

```
tapply(roomwidth$width * convert, roomwidth$unit, summary)
```

```
tapply(roomwidth$width * convert, roomwidth$unit, sd)
```

- **visualizacion**

```
> layout(matrix(c(1,2,1,3), nrow = 2, ncol = 2, byrow = FALSE))
> boxplot(l(width * convert) ~ unit, data = roomwidth,
+ ylab = "Estimated width (feet)",
+ varwidth = TRUE, names = c("Estimates in feet",
+ "Estimates in metres (converted to feet)"))
> feet <- roomwidth$unit == "feet"
> qqnorm(roomwidth$width[feet],
+ ylab = "Estimated width (feet)")
> qqline(roomwidth$width[feet])
> qqnorm(roomwidth$width[!feet],
+ ylab = "Estimated width (metres)")
> qqline(roomwidth$width[!feet])
```

- **t-Test varianzas iguales**

```
>t.test(l(width * convert) ~ unit, data = roomwidth,  
+ var.equal = TRUE)
```

- **t-test varianzas distintas**

```
>t.test(l(width * convert) ~ unit, data = roomwidth,  
+ var.equal = FALSE)
```

- **Wilcoxon**

```
>wilcox.test(l(width * convert) ~ unit, data = roomwidth, conf.int =  
TRUE)
```

Caso 2

In a design study for a device to generate electricity from wave power at sea, experiments were carried out on scale models in a wave tank to establish how the choice of mooring method for the system affected the bending stress produced in part of the device. The wave tank could simulate a wide range of sea states and the model system was subjected to the same sample of sea states with each of two mooring methods, one of which was considerably cheaper than the other. The resulting data (from Hand et al., 1994, giving root mean square bending moment in Newton metres) are shown in Table 3.2. The question of interest is whether bending stress differs for the two mooring methods.

```
data("water", package = "HSAUR2")
```

```
View(waves)
```

- **visualizacion**

```
> mooringdiff <- waves$method1 - waves$method2
> layout(matrix(1:2, ncol = 2))
> boxplot(mooringdiff, ylab = "Differences (Newton metres)",
+ main = "Boxplot")
> abline(h = 0, lty = 2)
> qqnorm(mooringdiff, ylab = "Differences (Newton metres)")
> qqline(mooringdiff)
```

- T-test

```
t.test(mooringdiff)
```

- Wilcoxon

```
wilcox.test(mooringdiff)
```

Caso 3

The data shown in [Table 3.3](#) were collected in an investigation of environmental causes of disease and are taken from Hand et al. (1994). They show the annual mortality per 100,000 for males, averaged over the years 1958–1964, and the calcium concentration (in parts per million) in the drinking water for 61 large towns in England and Wales. The higher the calcium concentration, the harder the water. Towns at least as far north as Derby are identified in the table. Here there are several questions that might be of interest including: are mortality and water hardness related, and do either or both variables differ between northern and southern towns?

```
data("water", package = "HSAUR2")
```

```
View(water)
```


- **visualizacion**

```
> nf <- layout(matrix(c(2, 0, 1, 3), 2, 2, byrow = TRUE),  
+ c(2, 1), c(1, 2), TRUE)  
> psymb <- as.numeric(water$location)  
> plot(mortality ~ hardness, data = water, pch = psymb)  
> abline(lm(mortality ~ hardness, data = water))  
> legend("topright", legend = levels(water$location),  
+ pch = c(1,2), bty = "n")  
> hist(water$hardness)  
> boxplot(water$mortality)
```

- Test sobre la correlacion de las variables

`cor.test(~ mortality + hardness, data = water)`

$$r / \sqrt{(1 - r^2) / (n - 2)}$$

Hipotesis nula: correlacion cero

Caso 4

The two-way contingency table in [Table 3.4](#) shows the number of piston-ring failures in each of three legs of four steam-driven compressors located in the same building (Haberman, 1973). The compressors have identical design and are oriented in the same way. The question of interest is whether the two categorical variables (compressor and leg) are independent.

```
data("pistonrings", package = "HSAUR2")
```

```
View(pistonrings)
```

- Ji-cuadrado test

```
chisq.test(pistonrings)
```

```
chisq.test(pistonrings)$residuals
```

```
> library("vcd")  
> assoc(pistonrings)
```

Caso 5

The data in [Table 3.5](#) (taken from Agresti, 1996) arise from a sample of juveniles convicted of felony in Florida in 1987. Matched pairs were formed using criteria such as age and the number of previous offences. For each pair, one subject was handled in the juvenile court and the other was transferred to the adult court. Whether or not the juvenile was rearrested by the end of 1988 was then noted. Here the question of interest is whether the true proportions rearrested were identical for the adult and juvenile court assignments?

```
data("rearrests", package = "HSAUR2")
```

```
View(rearrests)
```

```
mcnemar.test(rearrests, correct = FALSE)
```

```
binom.test(rearrests[2], n = sum(rearrests[c(2,3)]))
```