Prácticas de Estadística con *R*

Contrastes de hipótesis

En esta segunda parte de la práctica vamos a resolver con *R* algunos contrastes de hipótesis. Resolveremos algunos de los que hemos resuelto en la pizarra y otros del libro «Fundamentos de Estadística», de Daniel Peña, porque en ambos casos tenemos las soluciones disponibles. No podemos resolver otros de los ejercicios que hemos hecho en la pizarra porque no disponemos de las muestras de las poblaciones, que es lo que necesita *R*, en el enunciado se proporcionaba la información mediante unos estadísticos que eran suficientes para hacer los contrastes.

Si hacemos la búsqueda `help.search("test")` aparece, entre otra información:

`ansari.test(stats)`	`Ansari-Bradley Test`
`bartlett.test(stats)`	`Bartlett Test of Homogeneity of Variances`
`binom.test(stats)`	`Exact Binomial Test`
`Box.test(stats)`	`Box-Pierce and Ljung-Box Tests`
`chisq.test(stats)`	`Pearson's Chi-squared Test for Count Data`
`cor.test(stats)`	`Test for Association/Correlation Between Paired Samples`
`fisher.test(stats)`	`Fisher's Exact Test for Count Data`
`fligner.test(stats)`	`Fligner-Killeen Test of Homogeneity of Variances`
`friedman.test(stats)`	`Friedman Rank Sum Test`
`kruskal.test(stats)`	`Kruskal-Wallis Rank Sum Test`
`ks.test(stats)`	`Kolmogorov-Smirnov Tests`
`mantelhaen.test(stats)`	`Cochran-Mantel-Haenszel Chi-Squared Test for Count Data`
`mauchly.test(stats)`	`Mauchly's Test of Sphericity`

mcnemar.test(stats)	McNemar's Chi-squared Test for Count Data
mood.test(stats)	Mood Two-Sample Test of Scale
oneway.test(stats)	Test for Equal Means in a One-Way Layout
pairwise.prop.test(stats)	Pairwise comparisons for proportions
pairwise.t.test(stats)	Pairwise t tests
pairwise.wilcox.test(stats)	Pairwise Wilcoxon rank sum tests
power.anova.test(stats)	Power calculations for balanced one-way analysis of variance tests
power.prop.test(stats)	Power calculations two sample test for proportions
power.t.test(stats)	Power calculations for one and two sample t tests
PP.test(stats)	Phillips-Perron Test for Unit Roots
print.power.htest(stats)	Print method for power calculation object
prop.test(stats)	Test of Equal or Given Proportions
prop.trend.test(stats)	Test for trend in proportions
quade.test(stats)	Quade Test
shapiro.test(stats)	Shapiro-Wilk Normality Test
stats-deprecated(stats)	Deprecated Functions in Stats package
t.test(stats)	Student's t-Test
var.test(stats)	F Test to Compare Two Variances
wilcox.test(stats)	Wilcoxon Rank Sum and Signed Rank Tests
cox.zph(survival)	Test the Proportional Hazards Assumption of a Cox Regression
plot.cox.zph(survival)	Graphical Test of Proportional Hazards
survdiff(survival)	Test Survival Curve Differences
survival-internal(survival)	Internal survival functions
survobrien(survival)	O'Brien's Test for Association of a Single Variable with Survival

Entre paréntesis nos dice el paquete en que está cada función. Si fuese un paquete de los que no se carga por defecto al abrir *R*, tendríamos que cargarlo con `library()`. Los contrastes que están subrayados son los que nos interesan ahora.

De una población

Proporción

Ejercicio 2 de la Hoja 1: Cierta medicina, en tabletas, ha sido comprobada eficaz en el alivio de una alergia en más del 60% de los pacientes. El fabricante ha desarrollado una versión soluble del producto y desear comprobar si la medicina en esta forma es igual de eficaz. Se toma una muestra de 40 personas que tienen la alergia, el nuevo producto alivió a 32 de ellos. ¿Hay suficiente evidencia para sugerir que la introducción de la versión soluble ha alterado la eficacia de la medicina? Realiza el contraste relevante usando $\alpha = 0.01$? Encuentra el nivel crítico (p-valor).

Si dudamos de cómo utilizar la función `binom.test()`, podemos consultar la ayuda con `help(binom.test)`. Depende de cómo enunciemos las hipótesis nula y alternativa. Para hacer un contraste bilateral:

```
binom.test(32, 40, p = 0.6, alternative = "two.sided", conf.level = 0.99)

        Exact binomial test

data:  32 and 40
number of successes = 32, number of trials = 40, p-value = 0.009415
alternative hypothesis: true probability of success is not equal to 0.6
99 percent confidence interval:
 0.5945808 0.9318154
sample estimates:
probability of success
                   0.8
```

Para hacer los contrastes unilaterales:

```
binom.test(32, 40, p = 0.6, alternative = "less", conf.level = 0.99)

        Exact binomial test

data:  32 and 40
number of successes = 32, number of trials = 40, p-value = 0.998
alternative hypothesis: true probability of success is less than 0.6
99 percent confidence interval:
 0.0000000 0.9232752
sample estimates:
probability of success
                   0.8

binom.test(32, 40, p = 0.6, alternative = "greater", conf.level = 0.99)

        Exact binomial test

data:  32 and 40
number of successes = 32, number of trials = 40, p-value = 0.006065
alternative hypothesis: true probability of success is greater than 0.6
99 percent confidence interval:
 0.6145578 1.0000000
sample estimates:
probability of success
                   0.8
```

Como 32/40 = 0.8, podríamos haber pensado en hacer sólo el último de los contrastes. Como el nivel crítico es menor que el nivel de significación, tomamos la decisión de que no hay evidencia para suponer que no se cumple la hipótesis nula.

Para ver cómo hacer contrastes sobre la media y sobre la varianza de una población normal, utilizamos el siguiente ejercicio del libro de Peña.

Ejemplo 10.4: Se espera que la resistencia en kg/cm_2 de cierto material suministrado por un proveedor se distribuya normalmente, con media 220 y desviación típica 7.75. Se toma una muestra de 9 elementos y se obtiene: 203, 229, 215, 220, 223, 233, 208, 228, 209. Se pide:

1. Contrastar la hipótesis $\mu = 220$ y σ cualquiera.
2. Contrastar la hipótesis $\sigma = 7.75$ y μ cualquiera.

Introducimos la muestra en una variable:

```
x <- c(203, 229, 215, 220, 223, 233, 208, 228, 209)
```

Media de una población normal (bilateral)

```
t.test(x, y = NULL,
        alternative = "two.sided",
        mu = 220, paired = FALSE,
        var.equal = FALSE, conf.level = 0.95)

        One Sample t-test

data:  x
t = -0.3801, df = 8, p-value = 0.7138
alternative hypothesis: true mean is not equal to 220
95 percent confidence interval:
 210.5774 226.7560
sample estimates:
mean of x
 218.6667
```

Varianza de una población normal (unilateral)

No he encontrado este contraste implementado, así que podríamos implementarlo nosotros fácilmente.

Calculamos el estadístico del contraste:

```
D <- (length(x)-1)*var(x)/7.75^2

14.7513
```

Calculamos el nivel crítico (p-valor) de la distribución, es decir, el área que queda a la derecha de D, teniendo en cuenta que el estadístico del contraste sigue una Ji-cuadrado.

```
pchisq(D, length(x)-1, ncp=0, lower.tail = FALSE)

0.06416552
```

Aceptamos la hipótesis nula si el nivel de significación es 0.05.

De dos poblaciones

Diferencia de proporciones

Ejemplo 10.5 del libro de Peña: La proporción de defectos de un lote de n_1 = 100 unidades del proveedor A es 0.04, mientras que en un lote de n_2 = 150 unidades de B han aparecido 0.07. Estudiar si hay evidencia suficiente de diferencias entre los proveedores.

<u>Contraste 1</u>

No he encontrado el contraste que se ha visto en clase, que se puede implementar como sigue.

Se estima la proporción utilizando las dos muestras:

```
p <- (0.04*100+0.07*150)/(100+150)
```

```
0.058
```

Se calcula el valor del estadístico del contraste

```
D <- (0.04-0.07)/sqrt(p*(1-p)/100+p*(1-p)/150)
```

```
-0.9941626
```

Se calcula el nivel crítico del contraste (unilateral)

```
pnorm(D, mean=0, sd=1, lower.tail = TRUE)
```

```
0.1600719
```

<u>Contraste 2</u> (Obsérvese que utiliza un método distinto, aparece una χ^2_1)

Miramos la ayuda de la siguiente función, y luego la aplicamos:

```
?prop.test
```

Description:

'prop.test' can be used for testing the null that the proportions (probabilities of success) in several groups are the same, or that they equal certain given values.

```
prop.test(c(0.04*100, 0.07*150), c(100, 150),
          p = NULL, alternative = "less")

       2-sample test for equality of proportions with continuity correction

data:  c(0.04 * 100, 0.07 * 150) out of c(100, 150)
X-squared = 0.5155, df = 1, p-value = 0.2364
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000  0.02537730
sample estimates:
prop 1 prop 2
```

```
  0.04   0.07
```

Para contrastar las varianzas y las medias se utiliza el mismo ejercicio:

Ejemplo 10.7 de Peña: Se desea comparar la muestra del ejemplo 10.4 con la obtenida para otro proveedor: 221, 207, 185, 203, 187, 190, 195, 204, 212. ¿Puede admitirse que ambas muestras provienen de la misma población?

```
x <- c(203, 229, 215, 220, 223, 233, 208, 228, 209) ## Ejemplo 10.4
y <- c(221, 207, 185, 203, 187, 190, 195, 204, 212)
```

Diferencia de medias (dos muestras independientes)

Una vez que sabemos si podemos asumir igualdad de varianzas o no, podemos hacer el contraste de las medias (bilateral, como en el libro).

```
t.test(x, y, alternative ="two.sided",
           mu = 0, paired = FALSE, var.equal = TRUE,
           conf.level = 0.95)


        Two Sample t-test

data:  x and y
t = 3.4049, df = 16, p-value = 0.003623
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  6.876871 29.567573
sample estimates:
mean of x mean of y
 218.6667  200.4444
```

Diferencia de medias (dos muestras pareadas)

Ejemplo 10.6 del libro de Peña: Para comparar la velocidad de dos ordenadores A y B se mide el tiempo que invierten en realizar operaciones de una cierta clase definida. Se toma una muestra de cinco operaciones de esta clase, que son realizadas por ambos. Los resultados, en milisegundos, son: A = (110, 125, 141, 113, 182); B = (102, 120, 135, 114, 175). Analizar si hay diferencias: (a) teniendo en cuenta que los datos están apareados; (b) considerando muestras independientes.

```
A <- c(110, 125, 141, 113, 182)
B <- c(102, 120, 135, 114, 175)

t.test(A, B, alternative ="two.sided",
           mu = 0, paired = TRUE, var.equal = TRUE,
           conf.level = 0.95)


        Paired t-test
```

```
data:  A and B
t = 3.1623, df = 4, p-value = 0.03411
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.6100548 9.3899452
sample estimates:
mean of the differences
                      5
```

Cociente de varianzas (de poblaciones normales independientes)

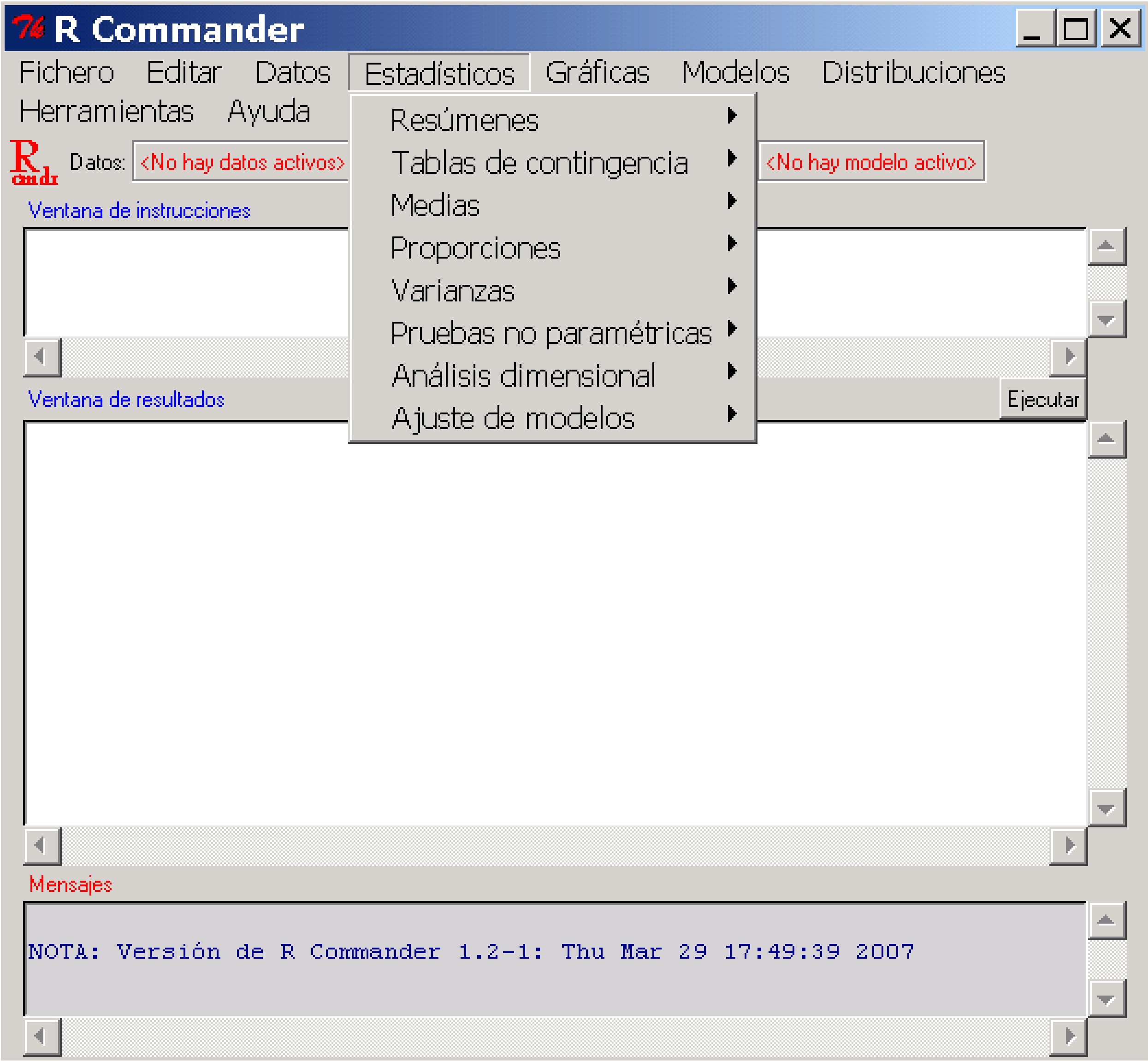
```
var.test(x,y)

        F test to compare two variances

data:  x and y
F = 0.7533, num df = 8, denom df = 8, p-value = 0.6982
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1699109 3.3393930
sample estimates:
ratio of variances
          0.753259
```

Apéndice: Entorno gráfico

Todos estos contrastes pueden ser hechos también utilizando el entorno gráfico de R que provee el paquete *Rcmdr*, del que se habló en la práctica anterior.



Universidad Complutense de Madrid
∟ Facultad de Ciencias Económicas y Empresariales
∟ Departamento de Estadística e Investigación Operativa II
∟ David Casado de Lucas

20 de febrero del 2012