

# R en estadística aplicada en la universidad de Alcalá

~~Imagina que lanzas un dado~~

Marcos Marvá, Fernando San Segundo

Departamento de Física y Matemáticas UAH

VIII Jornadas usuarios R  
Albacete 18.11.2016



LibreTICs

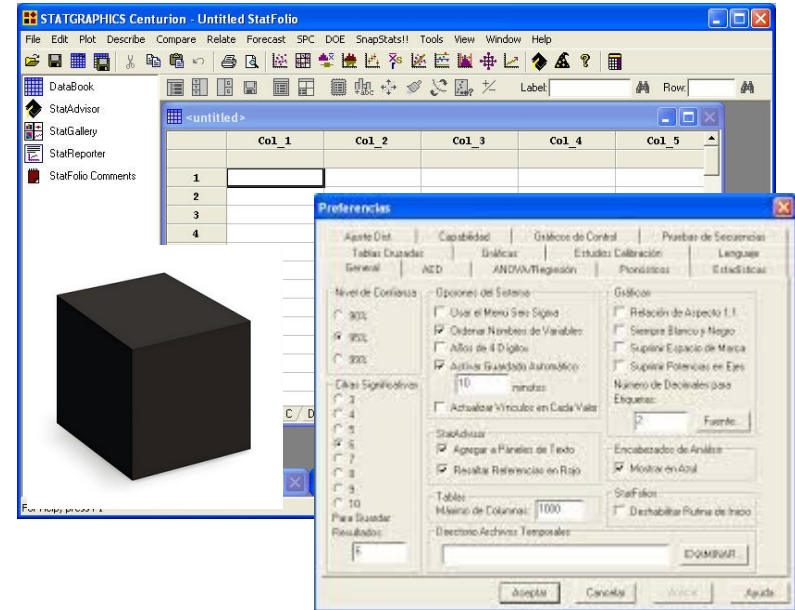


# ¿Qué hacíamos? Lo habitual\*

Clase magistral (grupo grande)

Ordenadores (grupo pequeño)

Problemas a mano (grupo pequeño)



$$\hat{p}_1 - \hat{p}_2 \mp Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\bar{X}_1 - \bar{X}_2 \mp Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\bar{X} \mp t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

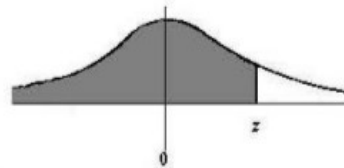
$$\left[ \frac{s_1^2/s_2^2}{F_{\alpha/2, n_1-1, n_2-1}}; \frac{s_1^2/s_2^2}{F_{1-\alpha/2, n_1-1, n_2-1}} \right]$$

$$\bar{X}_1 - \bar{X}_2 \mp Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

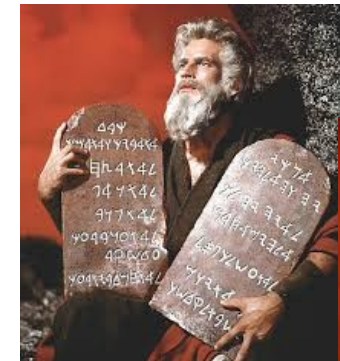
$$\bar{X}_1 - \bar{X}_2 \mp t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$



$$F_Z(z) = P\{Z \leq z\}$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131



\* en nuestro departamento

# ¿Qué queríamos?

- Computar la solución es parte de la solución

- Que la herramienta no sea el límite

- Integrar trabajo “a mano” y “a máquina”  
¡¡itablas no, menús fuera!!

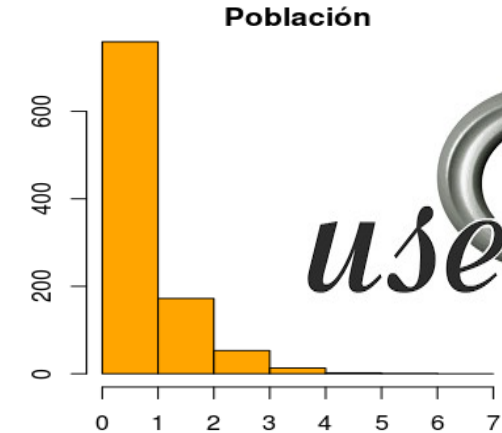
- Mejorar intuición en conceptos clave  
Experimental Reproducible Interactivo

# ¿Qué hacemos ahora?

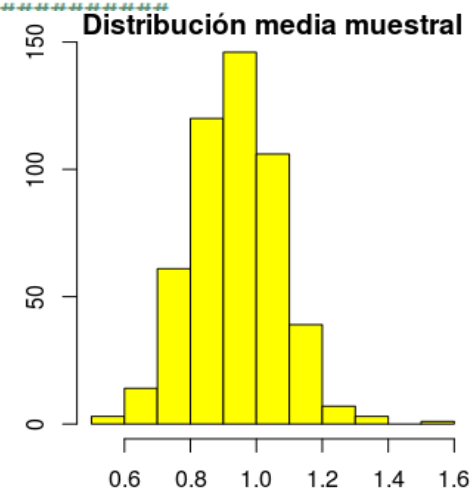
En el aula, grupo grande

- Simulaciones (respaldo [libro](#))

```
#####  
Marv  
cisco Javier Baron  
#####  
  
# clean memory and set graphical parameters  
rm(list=ls())  
par(mfrow = c(1,1), mar=c(4,2,1,4))  
  
#####  
# You are at home waiting for a book you bought in the Internet  
# The delivery arrives at any time withing the next hour,  
# X = "number of minutes before the delivery arrives"  
#####  
  
# define and plot the population  
pop1 <- seq(from = 0, to = 60, length.out = 60)  
hist(pop1, breaks = 0:60, xlab = "Minuto", main = "")  
  
# sample the population and represent the sample  
# to get an insight of how long costumers wait  
sample.size <- 10  
(sample1 <- sample(pop1, size = sample.size, replace =  
hist(sample1, xlab = "Minuto", main = "", breaks = 0:60))  
  
# function to automate the process  
plotSample = function(sample.size = n){
```



*use* **R!**



# ¿Qué hacemos ahora?

## En el aula, grupo grande

- Simulaciones (respaldo [libro](#))

## Aula de ordenadores, grupo pequeño

- Tutoriales/docs reproducibles

### 1.4. Tablas de contingencia relativas en R.

Vamos a ver cómo utilizar R para obtener las tablas relativas que hemos discutido en la página [476](#) del libro. Concretamente, vamos a ver cómo reproducir los resultados del Ejemplo [12.1.6](#) en el que se analizaba la tabla de contingencia correspondiente a una prueba diagnóstica, que hemos usado varias veces en el libro. Empezamos con la tabla de datos básica :

```
(tablaObservada = matrix( c(192, 4, 158, 9646), nrow= 2))  
  
##      [,1] [,2]  
## [1,] 192 158  
## [2,]   4 9646
```

Ponemos nombre a las filas y columnas:

```
colnames(tablaObservada) = c("Enfermos", "Sanos")  
rownames(tablaObservada) = c("Positivo", "Negativo" )  
tablaObservada  
  
##           Enfermos Sanos  
## Positivo      192   158  
## Negativo         4  9646
```

y ya estamos listos para pasar a los valores marginales. Los añadimos a la tabla pero, además, calculamos la suma total:

```
(tablaObservadaMarg = addmargins(tablaObservada))  
  
##           Enfermos Sanos Sum  
## Positivo      192   158  350  
## Negativo         4  9646  9650  
## Sum             196  9804 10000  
  
(n = sum(tablaObservada) )  
  
## [1] 10000
```

# ¿Qué hacemos ahora?

## En el aula, grupo grande

- Simulaciones (respaldo [libro](#))

## Aula de ordenadores, grupo pequeño

- Tutoriales/docs reproducibles
- Plantillas

```
1 #####
2 # www.postdata-statistics.com
3 # POSTDATA. Introducción a la Estadística
4 # Tutorial-06.
5 #
6 # Fichero de instrucciones R para calcular un intervalo de confianza (1-alfa) para la
7 #
8 #   DESVIACION TIPICA de una población normal N(mu,sigma),
9 #
10 # a partir de una muestra de tamaño n. ste fichero usa los estadísticos de una muestra,
11 # previamente calculados (numero de datos, media muestral, etc.)
12 #####
13
14 rm(list=ls()) #limpieza inicial
15
16 # ATENCIÓN: Para usar este fichero
17 # la población debe ser (al menos aprox.) normal
18 # EN OTROS CASOS NO USES ESTE FICHERO!!
19 # ASEGURATE DE HABER ENTENDIDO ESTAS INSTRUCCIONES
20
21 # Introducimos el valor de la desviación típica muestral,
22 s =
23
24 # el tamaño de la muestra,
25 n =
26
27 # y el nivel de confianza deseado.
28 nc =
29
30 #####
31 #NO CAMBIES NADA DE AQUI PARA ABAJO
32 #####
33 # Calculamos alfa
34 alfa = 1 - nc
35
36 # y los grados de libertad:
37 (k= n - 1)
38
39 # Calculamos los valores críticos necesarios:
40 (chiAlfa2 = qchisq(1 - (alfa/2), df=k))
41 (chiUnoMenosAlfa2 = qchisq(alfa/2, df=k))
42
43 #Para la varianza, el intervalo de confianza sera
44 (intervaloVar = s^2 * k / c(chiAlfa2, chiUnoMenosAlfa2))
45
46 # Y para la desviación típica el intervalo de confianza es este:
47 (intervaloS = s * sqrt(k / c(chiAlfa2, chiUnoMenosAlfa2)))
48
```

# ¿Qué hacemos ahora?

## En el aula, grupo grande

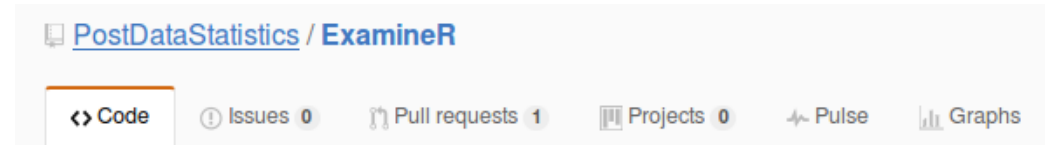
- Simulaciones (respaldo [libro](#))

## Aula de ordenadores, grupo pequeño

- Tutoriales/docs reproducibles
- Plantillas

## (Auto)evaluación <sup>(1)</sup>

- [Github repo ExamineR](#)  
157 preguntas “aleatorias”



A repository of statistics quizz questions, generated with the Exams package of R.

### Thesaurus

- Language: en - English, es - Spanish
- Question type: num - numeric, mult - multiple choice

### 00 Rounding numbers

- 000001 Round a number between 0 and 10 to 3-5 significant digits (en - es - num)
- 000002 Round a number between 0.01 and 1 to 3-5 significant digits (en - es - num)
- 000003 Round a number between 100 and 100000 to 3-5 significant digits (en - es - num)
- 000004 Round a large number to 3-6 significant digits (en - es - num)

### 01 Descriptive statistics

#### 0101 Central values

- 01010101 Mean (en - es - num)
- 01010301 Median (en - es - num)

#### 0102 Dispersal

- 01020101 Population variance (en - es - num)
- 01020201 Population standard deviation (en - es - num)
- 01020301 Sample variance (en - es - num)
- 01020401 Sample standard deviation (en - es - num)

(1) basado en exams

A. Zeileis, B. Gruen, F. Leisch, N. Umlauf, D. Ernst

# ¿Qué hacemos ahora?

## En el aula, grupo grande

- Simulaciones (respaldo [libro](#))

## Aula de ordenadores, grupo pequeño

- Tutoriales/docs reproducibles
- Plantillas

## (Auto)evaluación <sup>(1)</sup>

- [Github repo ExamineR](#)  
157 preguntas “aleatorias”
- Script StudentR:  
¡un estudiante, un examen!

StudentR An script aided by the exams R package <https://github.com/PostDataStatistics/ExamineR>

---

Nº: 1 Bayes , Thomas (Presbyterian Minister)

---

1. Round the number  
3.03625416010618  
to 3 significant figures.
2. Find the **median** of this set of numbers:  
5, 0, 7, 10, 5, 0, 10, 5, 11, 11, 7, 6, 6.  
Round the result to 4 significant digits.
3. Find the **(population) variance** of this set of numbers:  
3, 12, 0, 10, 12, 1, 6, 0, 1.  
Round the result to 4 significant digits.

StudentR An script aided by the exams R package <https://github.com/PostDataStatistics/ExamineR>

---

Nº: 2 Cox , Gertrude

---

1. Round the number  
3.11451492831111  
to 3 significant figures.
2. Find the **median** of this set of numbers:  
6, 10, 2, 11, 3, 9, 5, 10, 7.  
Round the result to 4 significant digits.
3. Find the **(population) standard deviation** of this set of numbers:  
1, 0, 11, 9, 8, 4, 4, 2, 5.  
Round the result to 4 significant digits.

(1) basado en exams

A. Zeileis, B. Gruen, F. Leisch, N. Umlauf, D. Ernst



# Resultados – conclusiones – ideas

## Estudiantes

- Pocas dificultades con los scripts
  - Muchos usan plantillas como menús
  - Bastantes comparten/retocan scripts
  - Algunos buscan/usan librerías “avanzadas”

## Profesores

Resulta natural usar scripts a las clases “magistrales”

**simulaciones** ↔ **imagen** ↔ **refuerza intuición**

- Compartir código con estudiantes (errores reproducibles)
- Clases más satisfactorias

## Ahora mismo....y más allá

- Asignar medida a los resultados
- Trabajo reproducible (coordinado, cuando es posible)
- TFG con R
- Píldora R

# Gracias por vuestra atención!!

## Enlaces

- Material cuestionarios en repo GitHub
- Libro, tutoriales (R, Python), blog,...
- Material para las clases y otros...

## Inspiraciones



Confessions of a converted lecturer  
Eric Mazur (physician, Harvard University) – Peer instruction



**TED**Talks  
**EDUCATION**

Material de **Fco Javier Barón** - UMA