

Ejemplo desarrollado para el uso del facilitador R

Vicente Manzano-Arrondo, 2017

Previo

1. Este documento es un ejemplo desarrollado. No explica con detalles el funcionamiento del facilitador. Para ello, habría que leer su manual.
2. Antes de utilizar el facilitador, hay que tener cargado R en el ordenador. No es necesaria ninguna librería extra, ni RCommander ni RStudio.
3. Crea una carpeta de trabajo donde se encuentran tus datos en código texto. Los datos están separados por comas. La primera fila contiene el nombre de las variables.
4. En esa carpeta de trabajo copia el archivo DAD.R cuyo enlace lo tienes en la web del facilitador.
5. La web del facilitador es
<http://asignatura.us.es/dadpsico/r/index.htm>
6. Este ejemplo utiliza los datos del anexo, que también encontrarás en la web del facilitador.
7. Para que funcionen los ejemplos, antes has tenido que rellenar el recuadro "Archivo de datos" del facilitador, al menos haber incluido los datos con el botón "Examinar" y haber escrito o pegado la ruta completa de tu carpeta de trabajo en el recuadro "Carpeta de trabajo:". A su vez, has pulsado el botón "Estudiar contenido". Todo ello habrá funcionado bien, si aparecen los nombres de las variables en las listas de variables de las diferentes utilidades del facilitador. En la ventana de resultados debería aparecer algo como:

```
# 25 casos y 5 variables: Sexo, Vive, Placer, Depende, Utilidades.  
# Carpeta y datos de trabajo:  
setwd("c:/Users/Vicente/Desktop/DAD/r/ejemplo")  
datos = read.table("datosEjemplo.csv", header=TRUE, sep=";", fill=TRUE, strip.white=TRUE,  
blank.lines.skip=TRUE)  
attach(datos)  
source("DAD.R")  
# Archivo de salida:  
sink("salida.txt", append=F, split=T)
```

Objetivo 1:

Estimar el número de utilidades medio que se utilizan en la población

Variable implicada: Utilidades.

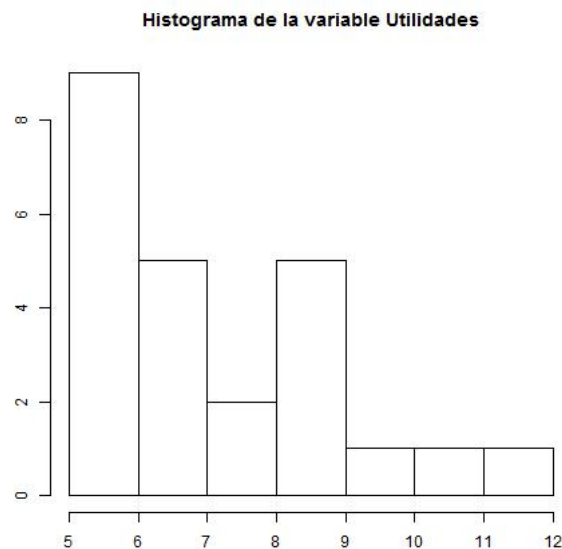
Opciones de análisis: - UNITARIO -

1. Comienza el proceso con nuevo archivo de salida ("salida.txt")
2. Selecciona la variable Utilidades del listado de variables
3. Selecciona todas las opciones del recuadro "Tabla de frecuencias"
4. Selecciona "Media y más" del recuadro de medidas numéricas
5. Selecciona "Representación gráfica", pon el título que quieras y escoge la opción "Histograma"
6. En el recuadro inferencia:
 1. Deja la seguridad al 95%
 2. Escoge "media"
 3. Selecciona "Construir intervalo..."
7. Pulsa botón "Insertar código".
8. Captura el código generado y ejecútalo en R (*este punto ya no se indica más, se sobreentiende en el resto de ejemplos*)

En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
# Descriptivos
cat("\nTabla de descriptivos:", "\n")
Variable = list("Utilidades")
.noUsar=rbind(fDescribe(Utilidades,med=T))
cbind(Variable,.noUsar)
#Tablas de frecuencias
cat("\nTabla de frecuencias de Utilidades:", "\n")
fTabFrec (Utilidades,acu=T,cal=T,zet=T,int=T)
#Representaciones gráficas
sink()
jpeg("imagen1.jpg")
hist(Utilidades,main="Histograma de la variable Utilidades",ylab="",xlab="")
dev.off()
cat("\nSe ha creado el archivo de imagen \"imagen1.jpg\" con la variable Utilidades\"","\n","\n")
sink("salida.txt",append=T)
#Inferencia estadística
cat("\nTabla de inferencia:", "\n")
Variable = list("Utilidades")
.noUsar=rbind(inf.Mediana(Utilidades,int=T,seg=0.95))
cbind(Variable,.noUsar)
```

Se ha generado la siguiente representación gráfica:



Y también la siguiente información en el archivo "salida.txt":

```
Tabla de descriptivos:
      Variable      Media DTip DAcó Mín Máx n
[1,] "Utilidades"  7.42  1.96  59   5  12  24

Tabla de frecuencias de Utilidades:
      f      p Fa  Pa  xf  x2f      Z Linf Lsup
5      5  20.8  5  21  25  125 -1.235  4.6 37.1
6      4  16.7  9  38  24  144 -0.724  1.8 31.6
7      5  20.8 14  58  35  245 -0.213  4.6 37.1
8      2   8.3 16  67  16  128  0.298  0.0 19.4
9      5  20.8 21  88  45  405  0.809  4.6 37.1
10     1   4.2 22  92  10  100  1.321  0.0 12.2
11     1   4.2 23  96  11  121  1.832  0.0 12.2
12     1   4.2 24 100  12  144  2.343  0.0 12.2
Tot  24 100.0 NA  NA 178 1412      NA   NA  NA
```

Tabla de inferencia:

	Variable	Media	Dtip	n	Conf	Linf	Lsup
[1,]	"Utilidades"	7.42	1.96	24	0.95	6.62	8.22

Hemos aprovechado la ocasión para mostrar más resultados, aunque realmente lo que directamente interesa es la "Tabla de inferencia", donde puede verse que la media de la muestra es 7,42 y que se estima una media poblacional entre 6,62 y 8,22 mediante una confianza del 95%.

Objetivo 2:

Comparar las distribuciones de Utilidades, Depende y Placer

De paso, pediremos una prueba de hipótesis para proporciones

Variable implicada: Placer, Depende y Utilidades.

Opciones de análisis: - UNITARIO -

1. Comienza el proceso, con nuevo archivo de salida ("s2.txt")
2. Selecciona las tres variables
3. No pidas tabla de frecuencias
4. Selecciona las tres casillas de medidas numéricas
5. En "Representación gráfica", con el título que quieras, escoge "Caja y patillas múltiple"
6. En "Inferencia":
 1. Estadístico: proporción, del valor 6
 2. Sin intervalo de confianza
 3. En "Inferir si se puede..." el valor de proporción es .2
7. Pulsa "Insertar código"

En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
# Archivo de salida:
sink("s2.txt", append=F, split=T)
# Descriptivos
cat("\nTabla de descriptivos:", "\n")
Variable = list("Placer", "Depende", "Utilidades")
.noUsar=rbind(fDescribe(Placer, mod=T, mdn=T, med=T))
.noUsar=rbind(.noUsar, fDescribe(Depende, mod=T, mdn=T, med=T))
.noUsar=rbind(.noUsar, fDescribe(Utilidades, mod=T, mdn=T, med=T))
cbind(Variable, .noUsar)
#Representaciones gráficas
sink()
jpeg("imagen1.jpg")
x=list(Placer, Depende, Utilidades)
boxplot(x, main="Placer, Depende y Utilidades")
dev.off()
cat("\nSe ha creado el archivo de imagen \"imagen1.jpg\" \n", "\n")
sink("s2.txt", append=T)
#Inferencia estadística
cat("\nTabla de inferencia: [valor= 6] \n", "\n")
Variable = list("Placer", "Depende", "Utilidades")
.noUsar=rbind(inf.Prop(Placer, val=6, hip=T, pi=.2))
.noUsar=rbind(.noUsar, inf.Prop(Depende, val=6, hip=T, pi=.2))
.noUsar=rbind(.noUsar, inf.Prop(Utilidades, val=6, hip=T, pi=.2))
cbind(Variable, .noUsar)
```

Se ha generado la siguiente información en el archivo "s2.txt":

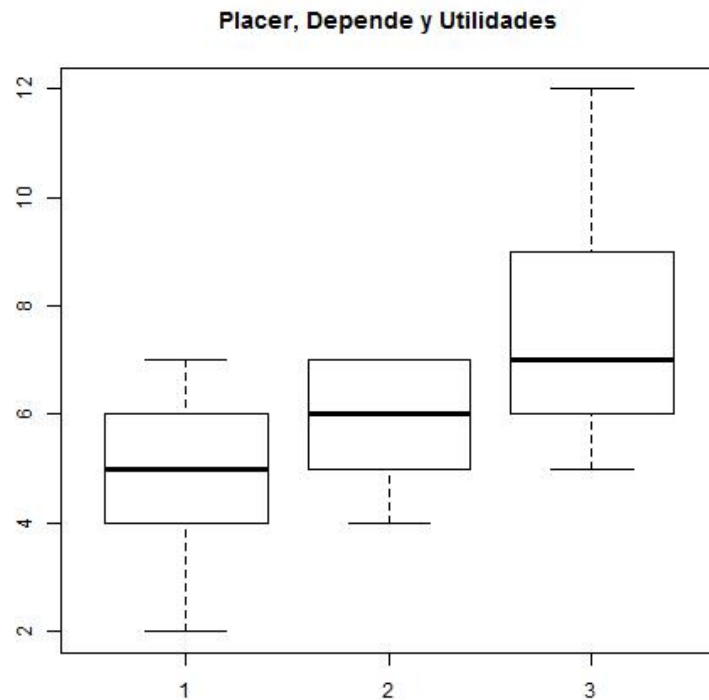
Tabla de descriptivos:

	Variable	Moda	PModa	Mediana	MDM	Media	DTip	DAco	Mín	Máx	n
[1,]	"Placer"	4	24	5	1	5.04	1.34	55	2	7	25
[2,]	"Depende"	6	32	6	1	5.76	0.99	67	4	7	25
[3,]	"Utilidades"	5	21	7	2	7.42	1.96	59	5	12	24

Tabla de inferencia: [valor= 6]

	Variable	Frec	Prop	n	pi	Z	p2col
[1,]	"Placer"	6	0.24	25	0.2	0.468	0.6396
[2,]	"Depende"	8	0.32	25	0.2	1.286	0.1984
[3,]	"Utilidades"	4	0.16	25	0.2	-0.546	0.5854

Y el siguiente gráfico en el archivo imagen1.jpg



Objetivo 3:

Estudiar la relación entre Sexo y Vive

Variable implicada: Sexo y Vive

Opciones de análisis: - CUANTÍA RELACIONES -

1. Selecciona "V de Cramer"
2. En filas: Sexo; en columnas: Vive.
3. Deja la seguridad al 95%
4. Selecciona "Incluir procesos de cálculo"
5. Pulsa "Insertar código"

En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
sink("s3.txt",append=F)
# Cuantificación de relaciones
cat("\n- Cuantificación de relaciones -","\n")
# Chi cuadrado y V de Cramer:
cat("\n* Chi cuadrado y V de Cramer:\n","\n")
cat("Frecuencias observadas","\n")
tabcon(Sexo,Vive,cel="fo",mar=T)
cat("\nFrecuencias esperadas","\n")
tabcon(Sexo,Vive,cel="fe",mar=T)
cat("\nComponentes de la chi cuadrado","\n")
tabcon(Sexo,Vive,cel="ji",mar=T)
cat("\n")
tabcon(Sexo,Vive,cel="0",mar=F,ji=T)
```

En R se han generado los siguientes resultados:

- Cuantificación de relaciones -

* Chi cuadrado y V de Cramer:

Frecuencias observadas

x	y			
	Costa	Rural	Urbano	Sum
Hombre	1	0	4	5
Mujer	9	5	6	20
Sum	10	5	10	25

Frecuencias esperadas

x	y			
	Costa	Rural	Urbano	Sum
Hombre	2	1	2	5
Mujer	8	4	8	20
Sum	10	5	10	25

Componentes de la chi cuadrado

x	y			
	Costa	Rural	Urbano	Sum
Hombre	0.500	1.000	2.000	3.500
Mujer	0.125	0.250	0.500	0.875
Sum	0.625	1.250	2.500	4.375

	Chi2	V	p
[1,]	4.38	0.42	0.1122

Objetivo 4:

Estudiar la relación entre Utilidades y Sexo

Variable implicada: Sexo y Utilidades

Opciones de análisis: - DESCRIPCIÓN RELACIONES -

1. Selecciona "Caja y patillas de niveles"
2. Caja: Utilidades, agrupadas por Sexo.
3. Pulsa "Insertar código"

Opciones de análisis: - CUANTÍA RELACIONES -

4. Selecciona "t de Student", opción "Transversal"
5. Medida: Utilidades; grupos: Sexo
6. Pulsa "Insertar código"

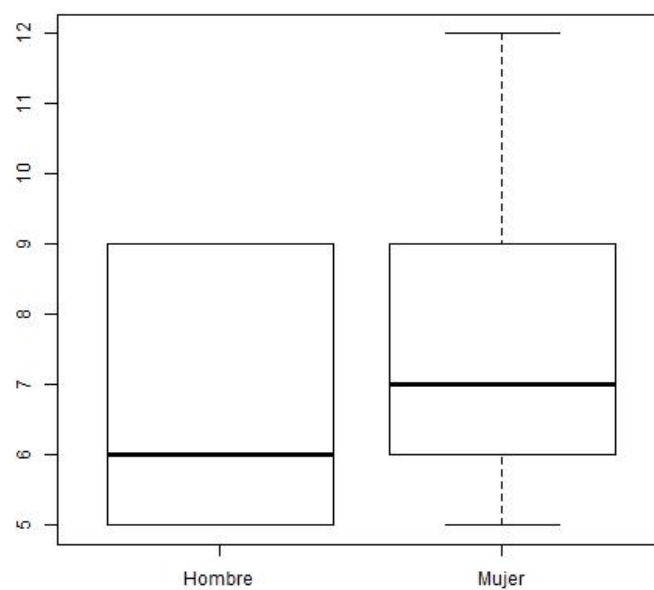
En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
# Diagrama de caja y patillas por niveles
jpeg("imagen1.jpg")
boxplot(Utilidades~Sexo,main="")
dev.off()
# Cuantificación de relaciones
cat("\n- Cuantificación de relaciones -","\n")
# t de Student para estrategia transversal
cat("* t de Student con Utilidades según niveles de Sexo","\n")
student(Utilidades,Sexo,par=F,seg=.95)
```

En R se han generado los siguientes resultados:

```
- Cuantificación de relaciones -  
* t de Student con Utilidades según niveles de Sexo  
$Estadísticos  
      Hombre Mujer  
n         5.00 19.00  
media     6.80  7.58  
dtipo     1.83  1.96  
dacot     92.00 58.00  
  
$Test  
      t    gl    p  
t -0.76  6.19 0.4756
```

Y también el siguiente gráfico:



Objetivo 5:

Relación entre Depende y Vive

Variable implicada: Depende y Vive

Opciones de análisis: - DESCRIPCIÓN RELACIONES -

1. Selecciona "Caja y patillas de niveles"
2. Caja: Depende, agrupadas por Vive.
3. Pulsa "Insertar código"

Opciones de análisis: - CUANTÍA RELACIONES -

4. Selecciona "Análisis de la varianza"
5. Dependiente: Depende; Independiente: Vive
6. Pulsa "Insertar código"

En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
# Diagrama de caja y patillas por niveles
jpeg("imagen2.jpg")
boxplot(Depende~Vive,main="")
dev.off()
# Cuantificación de relaciones
cat("\n- Cuantificación de relaciones -","\n")
# Análisis de la varianza:
cat("\n* Análisis de la varianza","\n")
cat("\nTabla de estadísticos de Depende según Vive","\n")
tabla(Depende,Vive)
cat("\nTabla de componentes de la varianza","\n")
summary(aov(Depende~Vive))
```

En R se han generado los siguientes resultados:

- Cuantificación de relaciones -

* Análisis de la varianza

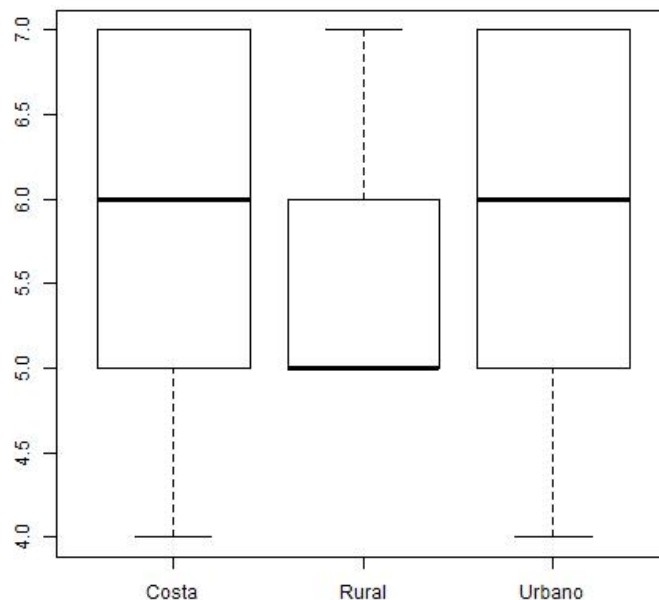
Tabla de estadísticos de Depende según Vive

	Costa	Rural	Urbano
n	10.00	5.0	10.00
media	5.80	5.6	5.80
dtipo	1.08	0.8	0.98
dacot	73.00	87.0	67.00

Tabla de componentes de la varianza

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Vive	2	0.16	0.080	0.072	0.931
Residuals	22	24.40	1.109		

Y también el siguiente gráfico:



Objetivo 6:

Relación entre Placer, Depende y Utilidades

Variable implicada: Utilidades, Depende y Placer

Opciones de análisis: - DESCRIPCIÓN RELACIONES -

1. Selecciona "Diagrama de dispersión"
2. Variables: Utilidades y Depende
3. Pulsa "Insertar código"
4. Repite 2+3 para Utilidades - Placer y Depende - Placer

Opciones de análisis: - CUANTÍA RELACIONES -

5. Selecciona "r de Pearson"
6. Escribe Placer, Depende, Utilidades
7. Pulsa "Insertar código"

En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
# Diagrama de dispersión
jpeg("imagen1.jpg")
plot(Utilidades,Depende,main="")
dev.off()
# Diagrama de dispersión
jpeg("imagen2.jpg")
plot(Utilidades,Placer,main="")
dev.off()
# Diagrama de dispersión
jpeg("imagen3.jpg")
plot(Depende,Placer,main="")
dev.off()
# Cuantificación de relaciones
cat("\n- Cuantificación de relaciones -","\n")
# Correlación lineal simple de Pearson:
cat("\n* Matriz de correlaciones","\n")
lista=cbind(Placer,Depende,Utilidades)
matrizr(lista)
```

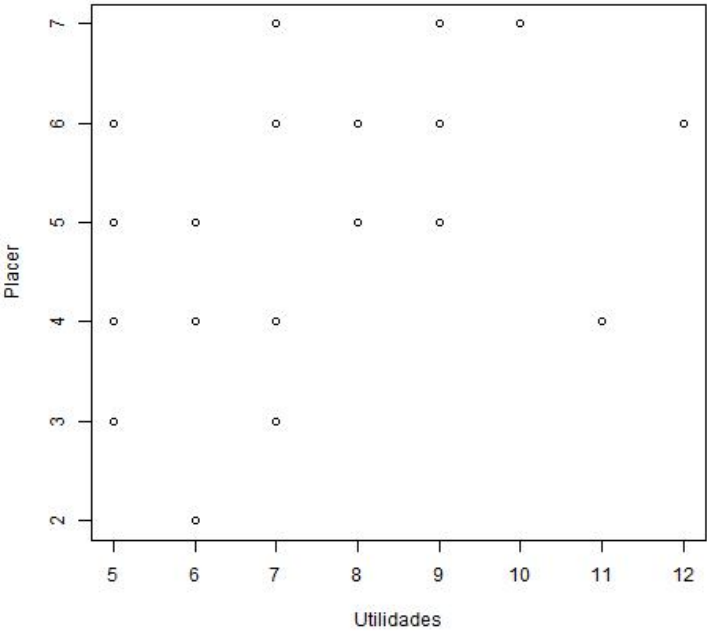
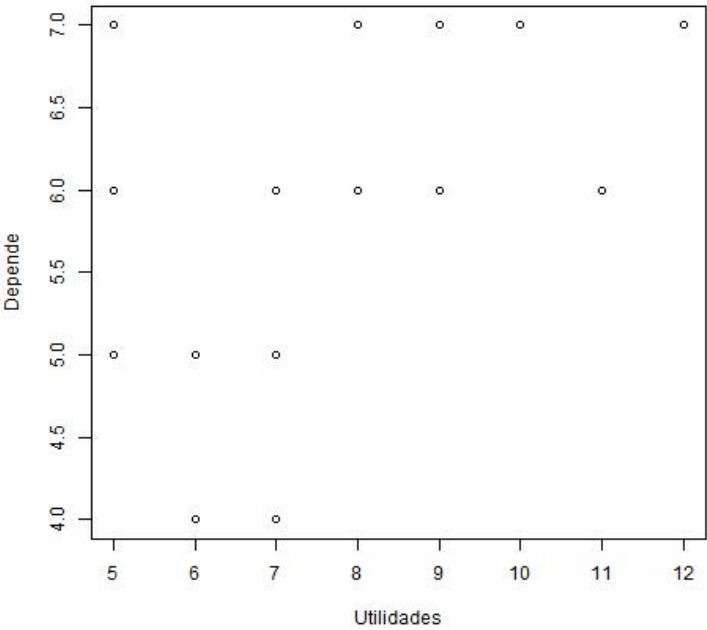
En R se han generado los siguientes resultados:

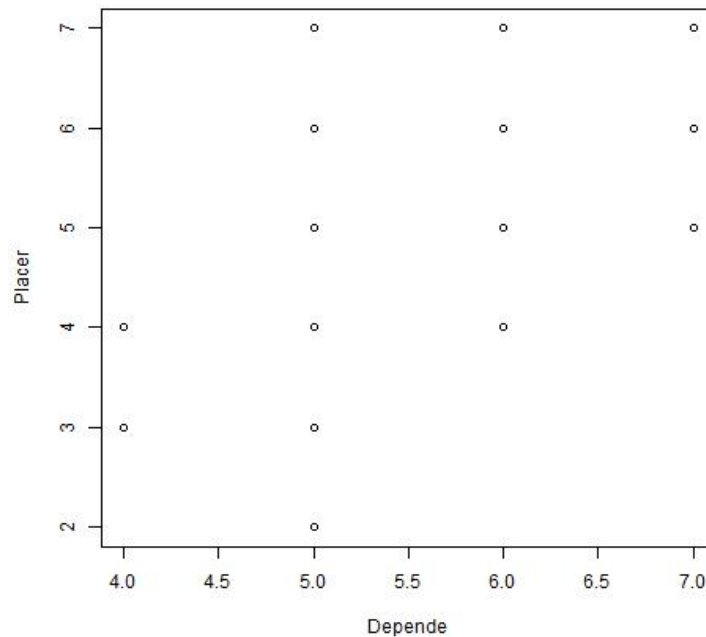
- Cuantificación de relaciones -

```
* Matriz de correlaciones
$Correlaciones
      Placer Depende Utilidades
Placer   1.000   0.579    0.383
Depende   0.579   1.000    0.538
Utilidades 0.383   0.538    1.000

$Riesgos
      Placer Depende Utilidades
Placer   1.0000   0.0024    0.0650
Depende   0.0024   1.0000    0.0067
Utilidades 0.0650   0.0067    1.0000
```


Y también los siguientes gráficos, cuya utilidad es muy discutible (pocos valores) pero muestra las posibilidades de ejecución de R a través del facilitador:





Objetivo 7:

Selección y generación aleatorias

No hay variables implicadas, si se requiere cargar ningún archivo de datos

Opciones de análisis: - MUESTREO -

1. Selecciona "Seleccionar una muestra aleatoria"
2. $N = 1000$, $n = 60$
3. Escoge ordenar los datos de la muestra
4. Selecciona "Generar una matriz aleatoria de datos"
5. 10 filas y 7 columnas
6. Distribución de éxito/fracaso con $p=.8$
7. Pulsa "Insertar código"

En el primer caso, supongamos que tenemos el listado de 1000 personas y que hemos de entrevistar a una muestra aleatoria de 60. En el segundo, estamos simulando una distribución de éxitos (representados por 1) y fracasos (representados por 0) para estimar proporciones, por ejemplo, siendo sensiblemente más probable tener éxito ($p=.8$).

En la ventana de códigos (resultados) se ha insertado el siguiente contenido:

```
# Selección de una muestra aleatoria
cat("Muestreo con N=1000 y n=60","\n")
muestra(rep(1:1000),n=60,ord=T)
# Generación de una matriz aleatoria de datos
# con 10 filas y 7 columnas.
cat("Matriz Bernoulli (éxito/fracaso) de 70 datos con probabilidad de éxito .8","\n")
matrix(data=simEF(n=70,p=.8),nrow=10,ncol=7)
```

En R se han generado los siguientes resultados:

```
Muestreo con N=1000 y n=60
[1] 10 63 99 119 133 162 167 175 207 215 315 320 343 354 364 372 395 400 413
419 456 466 483 484 528 529 554
```

[28] 565 604 611 613 620 627 634 653 655 658 662 672 674 694 697 704 723 732 738
 756 775 795 803 805 812 815 819
 [55] 830 888 932 938 966 990

Matriz Bernoulli (éxito/fracaso) de 70 datos con probabilidad de éxito .8

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	1	1	1	1	1	1	1
[2,]	1	1	1	1	1	1	0
[3,]	0	0	0	1	1	1	1
[4,]	1	0	0	1	1	1	1
[5,]	1	0	1	0	1	0	1
[6,]	1	0	1	0	1	1	1
[7,]	0	1	1	1	1	1	1
[8,]	1	1	0	1	1	1	1
[9,]	1	1	1	0	1	1	1
[10,]	1	1	1	1	1	0	1