### Práctica 1. Estadística descriptiva

Departamento de Estadística Universidad de Oviedo



Precuencias y porcentajes







### Cargar el programa R y el paquete Rcommander

- Iniciamos el programa R.
- Cargamos el paquete RCommader. Dos opciones:
  - **1** Menú *Paquetes*  $\rightarrow$  *Cargar paquete*  $\rightarrow$  Seleccionamos **Rcmdr**.



Escribimos library(Rcmdr) en la consola y pulsamos retorno de carro.



### Cargar Rcommander

Sy, Coupen de detes, "He hys conjunts de detes actives"   Editar conjunts de detes   Visualian conjunts de Teotran de indoucciónes	edatos Modelo: «His hay modelo activo»
ferfana fe individures	
entana de resultados	Eecutar
enajes	
1) NOTA: Versión de R Commander 1.5-3: Fri Oct 23 18:26:09 2009	*



El fichero acero.rda se encuentra en el Campus Virtual. Hay que haberlo descargado previamente.

### Cargar la base de datos acero.rda



Fichero Editar	Datos Estadísticos Gráficas Modelos Distribuc	iones Herramientas Ayuda
R Conjunto de	Nuevo conjunto de datos Cargar conjunto de datos	isualizar conjunto de datos Modelo: <no hay<="" th=""></no>
	Fusconar conjuntos de datos Importar datos Conjunto de datos en paquetes Conjunto de datos activo Modificar variables del conjunto de datos activo	desde archivo de texto, portapapeles o URL desde datos SPSS desde datos Minitab desde datos STATA

> load("/home/emilio/clases/acero.rda")

NOTA: El conjunto de datos acero tiene 117 filas y 20 columnas.

#### • Para visualizar la base de datos:

MQth	Help					R Com	mander					
Fiche	ero Editar	Datos B	stadístic	os G	Gráficas	Modelos Dist	ribuciones	Herr	amientas Ay	/uda		
R	conjunto de	datos:	acero	Ed	litar conju	unto de datos	Visualizar (	conju	nto de datos	Modelo	: mode	lo.maximo.o
Vent	tana de inst	rucciones						Т				
	orary(r	elimp,	pos=4	1)			acero					
	CODSUM	o pr t	hc pr	<i>cc</i>	pr ca	nr galv	l pr gal	1 77 2	nr nint	linea	avoria	hora r
	00110 011		be pr		priod	prigary	r prigai		priprine	TTHEA	averad	noru r
1	135.31	1 68	40	830	0	57	9 14	101	0	1	S1	1
2	84.08	2 4	43	903	58	61	1 16	535	717	1	No	2
3	131.61	5 72	70	572	36	98:	2 19	96	243	1	No	3
4	90.46	0 50	31	694	122	89	6 15	568	0	1	No	4
5	120.04	3 93	65 1	054	157	40	3 14	180	0	1	No	5
6	153.67	8 92	81 1	003	172	60	5 15	525	473	1	Si	6
7	99.08	9 32	23 1	118	0	64	3 14	124	732	1	No	7
8	226.37	5 104	90 1	077	179	73	7 13	333	93	1	No	8
0	140 06	8 73	Q/ 1	204	167	5.8	n d	1.54	217	1	Mo	1

Aparece una ventana con los datos disponibles. Moviendo el cursor hacia la izquierda o hacia abajo podemos recorrer toda la

base de datos.

E								A		Α.	
2	84.08	- 443	90.8	5.8	611	1886	717	- A	2.4		
5			5/1	iene	20 va	riables	(colur	nnas	31	11	
3	10.40	5031	- 624	122	- 030	1300	feered.	× 1	<ul> <li>41</li> </ul>	н	
2	120.04	9165	1054		487	1.4.60			57	м	
0	323.68	92.85			697		473	- A		ы	
0.0								5	- 22	2	
0		10400						λ.	_ 0 *	λ	
0	140.07	7394	1204	167	510	904	247		1-		
83		8454	- 22 C		141	1324		- 14	- 25	2.	
EE		8705	JISDC	ne a	e ma	observ	acione	s (Q	ias)	н	
BBI		51.78	0		62.6			- 8	-3*	н	
BEI	57.64	0544	0		54	942		- n	5.	м	
13		93.52	0		r		561		- A*	3.	
1000		4818							7.4	2	

### Variables de la base de datos acero

- Consumo Consumo energético de la empresa (Megavatios/hora).
- In the Producción del tren de bandas calientes (Toneladas de acero).
- In the pr.cc Producción de colada continua (Toneladas de acero).
- I pr.ca Producción del convertidor de acero (Toneladas de acero).
- pr.galv1 Producción de galvanizado de tipo I (Tns. de acero).
- I pr.galv2 Producción de galvanizado de tipo II (Tns. de acero).
- Int Producción de chapa pintada (Tns. de acero).
- Iinea Línea de producción empleada (A o B).
- Uturno Turno de mañana (M), tarde (T), noche (N).
- 🔟 temperatura Temperatura del sistema: Alta, Media y Baja.
- 🔟 pres.aver Presencia de averías: hubo Averías (A), no hubo averías (NoA).
- 1 nun.aver Número de averías detectadas.
- sistema Activación de un sistema de detección de sobrecalientamiento: encendido (ON), apagado (OFF).

### Ordene la variable Presencia de averías (pres.aver).

#### En la ventana Rscript escribimos:

#### sort(acero\$pres.aver)

#### Con el cursor en esa línea, pulsamos Ejecutar.

R Commander - e	8
Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda	
🕞 Conjunto de datos: 🔟 acero 📝 Editar conjunto de datos 🔯 Visualizar conjunto de datos Modelo: 🗴 <no activo<="" hay="" modelo="" td=""><td>&gt;</td></no>	>
R Script R Markdown	
sort(acero\$pres.aver)	
Salida	
> sort(acero\$pres.aver) [1] A A A A A A A A A A A A A A A A A A A	

#### La categoría A representa avería y la NoA no avería.

(Universidad de Oviedo)

### Halle la distribución de frecuencias de la variable pres.aver.



```
counts:
pres.aver
A NoA
28 89
percentages:
pres.aver
A NoA
```

```
23.93 76.07
```

Así, se ha obtenido el número de casos de cada modalidad y el porcentaje que representan dentro de la muestra. La moda es el dato que más se repite; en este caso, la modalidad NoA.

(Universidad de Oviedo)

### Ordene la variable *Número de averías* (num.aver) y obtenga su tabla de frecuencias.

#### En la ventana *Rscript* escribimos: sort(acero\$num.aver) y pulsamos *Ejecutar*. table(acero\$num.aver) y pulsamos *Ejecutar*.



# Convierta la variable *Número de averías* (num.aver) en una variable de tipo factor con el nombre Num.aver

Es una variable numérica.

#### 2 Datos

Modificar variables del conjunto...

➡Convertir variable numérica en factor



Seleccionar la variable num.aver
 Utilizar números
 Nuevo nombre:
 Num.aver
 Aceptar

Convertir variables	numéricas en factores — + ×
Variables (seleccione una o más) NOX NOX proce pr.cc pr.cc	Niveles del factor Asignar nombres a los niveles Utilizar números
Nuevo nombre o prefijo para varia	ables múltiples: Num.aver

#### > acero\$Num.aver <- as.factor(acero\$num.aver)</pre>

# Obtenga la distribución de frecuencias de la nueva variable Num.aver.

- Estadísticos
  - Resúmenes
    - Distribución de frecuencias
- Seleccionar la variable Num.aver
   ➡Aceptamos



Q	Distribuciones de frecuencias - + ×
Variables (seleccion linea Num.aver pres.aver sistema temperatura turno Test Chianadra	use a máit
🔞 Ayuda	Reinidar Aplicar Cancelar Aceptar

cou	nts	:				
Num	.av	er				
0	1	2	3	4		
89	2	9	9	8		
per	cen	tag	es:			
Num	.av	er				
	0		1	2	3	4
76.	07	1.	71	7.69	7.69	6.84

Un gráfico erróneo es garantía de fracaso.



http://imgs.xkcd.com/comics/convincing.png

# Represente gráficamente la distribución de la variable pres.aver mediante una gráfica de barras.



#### Obtenga el gráfico de barras de la variable naverias

- Es variable numérica, hay que pasarla a factor.
- Trabajaremos con la variable Num.aver.







### Represente gráficamente la distribución de la variable averias mediante un gráfico de sectores.

Gráficas

- Gráfica de sectores
  - Seleccionar la variable pres.aver
    - Aceptamos

#### No se recomienda este tipo de gráficos.



https://en.wikipedia.org/wiki/File:Piecharts.svg

(Universidad de Oviedo)

Práctica 1

#### Obtenga el histograma de la variable consumo.





(Universidad de Oviedo)

Práctica 1

#### Obtenga el diagrama de cajas de la variable consumo.





### Obtenga los diagramas de cajas de consumo según la temperatura.

#### Gráficas

- Diagrama de caja...
  - Seleccionar la variable **consumo** 
    - Pulsamos Gráfica por grupos...





# Calcule la media, desviación típica y percentiles de la variable *Número de averías* (num.aver).



mean	sd	IQR	0%	25%	50%	75%	100%	n
0.6752137	1.292078	0	0	0	0	0	4	117

Los resultados nos indican que la media es de 0.67 averías por hora, con una desviación típica de 1.29. El número de averías varía desde 0 hasta 4, y al menos el 75% de la observaciones no presentaron averías. En total disponemos de 117 observaciones.

### Calcule los principales estadísticos descriptivos de la variable consumo.

Estadísticos ● ➡Resúmenes ➡Resúmenes numéricos	Note of the second se
● Seleccionar la variable <b>consumo</b>	Resúmenes numéricos     ×       Variables (seleccione una numéricos)     possumo
➡Aceptamos	prica       prica     >       Outantiles of cuartiles (0, 25, 5, .75, 1       Respirat     Cancelar       Ayuda

mean sd IQR 0% 25% 50% 75% 100% n 135.6771 56.90756 83.39 17.5 99.09 135.1 182.48 290.72 117

El consumo medio se sitúa en torno a ¿ ? megavatios-hora, con una desviación típica de ¿ ? MWh. El consumo mínimo desciende hasta ¿ ? y el máximo asciende hasta ¿ ?. El 25% de los casos analizados consumen ¿ ? MWh o menos; el 50%, menos de ¿ ?; y un ¿ %? consume más de 182.48.

# Calcule los principales estadísticos descriptivos de la variable consumo para cada nivel de la temperatura.

Estadísticos ➡Resúmenes ➡Resúmenes numéricos

Seleccione la variable consumo Pinche *Resumir por grupos* Seleccionar temperatura OK (en la ventana de Grupos) OK (en Resúmenes numéricos)





	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
Alta	103.5239	51.13951	55.195	17.50	66.4700	99.75	121.6650	234.39	46
Baja	135.5716	46.97305	49.580	39.72	113.1625	135.51	162.7425	246.54	38
Media	180.6182	44.06330	35.190	68.30	165.5600	187.36	200.7500	290.72	33

#### Estos son las calificaciones del examen de prácticas de laboratorio de Estadística del año pasado. Determine los principales estadísticos.

 mean
 sd
 IQR
 0%
 25%
 50%
 75%
 100%
 n
 NA

 6.728571
 2.420019
 3.025
 0.85
 5.7625
 7.05
 8.7875
 10
 70
 21

### Cree una nueva variable denominada coste, si éste vale 2.34 veces cada unidad del consumo.

#### Datos Modificar variables del con-Fichero Editar Dates Estadisticos Gráficas Modelos Distribuciones Herramientas Avuda Conjunto Visualizar conjunto de datos Modelo: 1 junto de datos activo R Script R Ma filenane="/hi Importar datos rimero/2015/Presentaciones/Practical/dibujos15/orcajaco width=7, he atorewr, he' library(abind, Conjunto de datos en paquetes Calcular una nueva varilibrary(e1971, Conjunto de datos activo sunSannery(act Modificar variables del conjunto de datos ac nunSunnary(acc acero s- withinfacero. Grupoconsumo «· Recode(consumo, 'lo:100able Convertir variable numérica en factor... Segmentar variable numérica. Reordenar niveles de factor... width-7, height-7, pointsize-12, units-'in', res-72) Eliminar variables del conjunto de datos library(abind, pos=17) Nombre de la nueva variable coste Expresión a calcular:

coste Expresión a calcular: 2.34 \* consumo ➡Aceptamos

Variables actuales (doble clic CO CO2 consumo COV linea [factor] N2O	ara envira a la expresión)	
Nombre de la puesa variable coste	Expresión a calcular 2.34 * consumo	
🔞 Ayuda	🥱 Reiniciar 🛛 🥐 Aplicar 🗮 Cancelar 📝 Aceptar	

> acero\$coste <- with(acero, 2.34 \* consumo)

#### Cree una nueva variable GrupoConsumo

 $Grupoconsumo = \begin{cases} Bajo & \text{si consumo} \le 100\\ Medio & \text{si } 100 < \text{consumo} \le 200\\ Alto & \text{si consumo} > 200 \end{cases}$ 

#### Datos

Modificar variables del conjunto de datos activo...

Recodificar variables

 Variable a recodificar consumo Nueva variable: Grupoconsumo Introducir directrices: Io:100= "Bajo" 100:200 = "Medio" 200:hi = "Alto"



	Recodificar variables
Variables a recodific	ar (elige una o más)
0	
02	
onsumo	
ov	
inea H2O	
120	•
luevo nombre o pre	fijo para variables múltiples recodificada: Grupoconsumo
Convertir cada r	weva variable en factor
ntroducir directrice	s de regodificación
o:109= "Baio"	
00:200 = "Medio"	
00:hi = "Alto	
•	
🔞 Ayuda	🦘 Reiniciar 🛛 🥐 Aplicar 🛛 💥 Cancelar 🚽 Aceptar

```
> acero <- within(acero, {
+ Grupoconsumo <- Recode(consumo, 'lo:100= "Bajo"; 100:200 = "Medio"; 200:hi = "Alto', as.factor
+ ))</pre>
```

# Cree una nueva base de datos con los registros que corresponden a la tempertura Alta



> acero.temp\_alta <- subset(acero, subset=temperatura == "Alta") NOTA: El conjunto de datos acero.temp\_alta tiene 46 filas y 23 columnas.



#### Si queremos activar otras bases:



@ Seleccionato d	edatos – + $\times$
Conjuntos de datos (sel	eccionar uno)
acero	
acero.temp_alta	
	$\nabla$
🗱 Cancelar	🚽 Aceptar