# Descripción simultánea de varias variables con STATGRAPHICS

Ficheros empleados: AlumnosIndustriales.sf3, Rotura.sf3

## 1. Introducción

En muchas ocasiones nos interesará comparar varias variables, o comparar los valores de una variable en dos o más grupos de individuos. En esos casos es más interesante producir gráficos y resúmenes estadísticos conjuntamente, que faciliten esa comparación, que realizar el análisis univariante por separado de cada variable. Por ejemplo, nos interesará hacer diagramas box-plot de cada variable pero en un mismo gráfico.

## 2. Box-plot Múltiple

Nuestro objetivo es una representación gráfica que tenga los diagramas box-plot de varias variables. Este gráfico permitirá una mejor comparación de esas variables. El Statgraphics proporciona varios lugares para hacer box-plot múltiples.

## 2.1 Una variable que se subdivide en subgrupos

Nos interesa analizar cómo es la distribución de valores de una variable cuando el conjunto de datos lo subdividimos en subgrupos de acuerdo a algún criterio. Por ejemplo las alturas de un grupo de estudiantes en función de su sexo. Necesitamos tener dos columnas:

- Columna con los datos de la variable
- Columna con códigos que nos permita hacer los subgrupos. Por ejemplo, con la variable sexo, basta con que tenga valores 1 y 0. Estos valores son sólo para distinguir a los miembros de cada grupo, por lo que su valor es irrelevante. Podrían ser -1 y 1, 33 y 34, o incluso caracteres.

Hay varios lugares para hacer este tipo de Box-plots. El primer lugar es en Plot/Exploratory Plots/Multiple Boxand-Whisker Plot

<mark>沙</mark> s	STATGRAPHICS Plus - Untitled StatFolio								
File	Edit	Plot	Describe	Compare	Relate	Special	View	Window	Help
a		So	atterplots		🔹 🕨 🗖		जन्म विद्य	-	z   1594   15
_		E>	ploratory F	Plots	•	Box-and-	Whiske	r Plot	
	Multir	Bu	isiness Cha	rts	•	Multiple B	Box-anc	l-Whisker F	Plot
		-	1 1 1 1				1 1 1		

Veamos un ejemplo con el fichero AlumnosIndustriales.sf3. que tiene algunos datos de un conjunto de estudiantes de Ingeniería Industrial. Vamos a comparar la estatura de los chicos y de las chicas. Se nos pregunta entonces por el nombre de la variable que tiene los datos -.altura- y la variable que nos ayudará a formar los dos grupos chicos/chicas –sexo- (Level Codes)

Multiple Box-and	-Whisker Plot
altura dinero hermanos locomoción nacimiento peso residencia sexo tiempo Variables zapato	Data Altura Level codes Select:



donde chicos=1, chicas=0. Puede verse que los chicos son, en general, más altos que las chicas. Viendo el solapamiento de los bigotes de ambos box-plots podemos interpretar que, aproximadamente, sólo el 25% de las chicas más altas tienen estaturas comparables al 25% de los chicos más bajos.

Los Box-plots múltiples también se pueden hacer en Compare/Multiple Samples/Multiple-Sample Comparison.

🥖 STATGRAPHICS Plus - Untitled StatFolio							
File Edit Plot Describe	Compare Relate	Special View Window Help					
	Two Samples	🔸 🗖 📼 🔤 🚾 🚾 👞 🐛					
<u>– – – – –</u> – –	Multiple Samples	Multiple-Sample Comparison					
Hultiple Box-and-Wb	Analysis of Variar	nce 🕨 Comparison of Proportions					

Esta opción es mucha más flexible que la anterior, pues permite no solo comparar una variable en varios subgrupos como comparar varias variables diferentes (varias columnas). Se nos ofrecen dos posibilidades, donde la que nos interesa en este momento es la de Data and Code Columns:

	Multiple-Sample Comparison			
Similar a la anterior	Input     Multiple Data Columns	OK		
	Oata and Code Columns			
	C Sample Statistics	Help		

La opción *Data and Code Columns* permite usar datos en el mismo formato que antes: una variable con los datos y otra con la información para hacer subgrupos. Si seleccionamos esta opción llegamos a una ventana similar a la que vimos antes donde nos preguntan los nombres de las variables

Multiple-Samp	le Comparison
altura dinero hermanos locomocion nacimiento peso residencia sexo tiempo Vanables zapato	Data altura Level codes Sexo (Select:)

Para visualizar el box-plot tenemos que seleccionarlo en las opciones gráficas, como se muestra a continuación

Hultiple-Sample Comparison					
[] [] [] [] [] [] [] [] [] [] [] [] [] [	触 Row:	艚			
Analysis Ammary					
Graphical Options	×	1		Box-and-V	Whisker Plot
nt variable: al				2001 1110	
			<sup>200</sup> F		
Number of bservations			Ē		
Number of levels: 2 🔽 Means Plot			190 -		
■ Box-and-Whisker Plot			<u>8</u> 180 E		
The StatAdv.sor			<b>2</b>		
			Tel 170 -		- <u>+</u>
This projedure comp			100		
the semples. The Rite Residuals versus Observation			100 -		
are any significant di			150 È		-
Multiple Rarge Tests w 🗌 🗌 Analysi (Means (ANOM) Plot				0	1
different frywhich o OK 3 el All	Help	Ŀ		s	exo

Cuando, como en el ejemplo de las estaturas de chicos y chicas, sólo tenemos dos subgrupos, los box-plot también pueden hacerse en:

<mark>沙</mark> STATGRAPHICS Plus - Untitled StatFolio									
File Edit Plot Describe	Compare	Relate	Specia	View	Window	Help			
2 🗖 📷 📟 😎	۱.	Two-Sample Comparison							
_ <b>_</b>	Multiple	Samples	•	Paired-Sample Comparison					
Analysis of Variance					Hypothesis Tests				
	Samp	ple-Size De	etermination						

A continuación hay que especificar que nuestros datos, la variable altura, están en una columna (Data Column) y que los dos subgrupos se forman con la variable sexo (Code Column)

Two-Sample Con	nparison			X
altura dinero hermanos locomocion nacimiento peso residencia <u>sexo</u> tiempo Variables zapato	Data: Sample Coo Select:)	ura de: ko		
Sort		iput Two Data C Data and Co	olumns ode Columns	
OK	Cancel	Delete	Transform	Help

Entonces seleccionamos la opción gráfica de Box-Plot

📲 Two-Sample Comparison - altura & sexo	
1 ivsistedummary for altura 1 ivsistedummary for altura 1 ivsistedummary for altura C Frequency Histogram 2 ivsistedummary for altura	Box-and-Whisker Plot
Sample 1: 32 values ranging from 15. Sample 2: 63 values ranging from 165 Quantile Plot	
The StatAcrisor Quapile Quantile Plot This procedure is designed to com calculate various statistics and gra OK 3 cel	
run several tests to decermine whether there are subscitairy significan differences between the two samples.	150 E

#### 2.2 Variables en columnas diferentes

Si tenemos varias columnas, de una misma magnitud o de magnitudes diferentes, podemos también hacer box-plot múltiples de esas variables en un mismo gráfico. Como ejemplo usaremos los datos del fichero rotura.sf3. Este fichero contiene la tensión de rotura (presión ejercida en el momento en que se produce la rotura) de un conjunto de piezas idénticas con el objetivo de probar la resistencia del material empleado. El fichero contiene 10 variables, LAB1 a LAB10, donde cada una muestra las tensiones de roturas obtenidas en 10 laboratorios diferentes. En cada laboratorio se rompieron 100 piezas diferentes. y se anotaron las correspondientes tensiones de rotura. Vamos a ver el box-plot de las 10 variables. En primer lugar vamos a

STATGRAPHICS Plus - Untitled StatFolio							
File Edit Plot Describ	e Compare Rela	ite Special	View Wi	ndow	Help		
	Two Sample:	; •		<b>-</b> 1 💀	2   🖂   🔜   🛌   1		
<b></b>	2 Multiple Sam	ples 🕨	Multiple-:	Sample	Comparison		
Multiple Box-and-	Analysis of \	ariance 🕨	Comparis	son of I	Proportions		

donde ahora seleccionamos que tenemos varias variables:

1ultiple-Sample Comparison	×
 Input Multiple Data Columns	OK
O Data and Code Columns	Cancel
Sample Statistics	Help

e introducimos los nombres de las variables



a continuación seleccionamos la opción gráfica de box-plot.



Puede verse que los resultados de los 10 laboratorios son muy similares. A partir de los bigotes del tercer cuartil y de los atípicos se aprecia que, en general, las distribuciones son asimétricas positivas. Por tanto, esos valores no son propiamente atípicos, sino la cola de la distribución por la derecha.

Si sólo tuviésemos dos variables, las podemos comparar con Box-plots en:

🛃 STATGRAPHICS Plus - Untitled StatFolio								
File Edit Plot Describe	Compare	Relate	Special	View	Window	Help		
	Two Sar	mples	•	Two-	Sample Co	mparison		
<u>– – – – –</u> –	Multiple	Samples	•	Paire	d-Sample (	Comparison		
Two-Sample Compar	Analysis	s of Varia	nce 🕨	Нурс	thesis Tes	ts		
				Samr	ble-Size De	termination		

A continuación habría que especificar los datos, que en nuestro caso serían los datos de los laboratorios que quisiésemos comparar.



y haríamos el box-plot como se hizo anteriormente



# 3. Dos histogramas superpuestos

Si queremos comparar dos histogramas, es útil colocarlos en la misma figura. Esta opción gráfica se hace en Compare/Two Samples/Two-sample Comparison, Tanto si tenemos una variable dividida en dos grupos, como dos variables en dos columnas diferentes

STATGRAPHICS Plus - Untitled StatFolio										
File	Edit	Plot	Describe	Compare	Relate	Special	View	Window	Help	
3	a 🗖 📷 📟 📼				Two Samples 🔹 🕨			Two-Sample Comparison		
<u>– – – – – – –</u>				Multiple Samples 💦 🕨			Paired-Sample Comparison			
	Two-9	Samp	e Compai	Analysis of Variance 🕨			Hypothesis Tests			
						74	Sample-Size Determination			

Vamos a aplicarlo al caso de las estaturas de chicos y chicas. El Statgraphics nos pregunta el formato de nuestros datos: dos columnas de datos, o una columna con datos y otra con códigos. En nuestro caso la opción a seleccionar es la segunda:

altura dinero hermanos locomocion	Data:	ltura		
nacimiento peso residencia sexo tiempo Variables zapato	Sample C Select:)	ode: exo		
Sort		Input O Two Data ( O Data and C	Columns Code Columns	
ОК	Cancel	Delete	Transform	Help

y seleccionamos entonces los histogramas. El resultado es el siguiente



vemos que para las chicas salen muy pocas clases. Podemos aumnetar el número de clases colocándonos en la ventana del histograma y pulsando el botón derecho del ratón. En Pane Options podemos cambiar el número de clases.



Con este número de clases se aprecia mejor que dentro de cada grupo, las alturas son bastante simétricas, y que los chicos tienen estaturas más altas.

Si queremos hacer dos hostogramas con las tensiones de rotura de dos laboratorios (fichero rotura.sf3) tenemos que introducir los siguientes datos



donde puede apreciarse que las distribuciuo es en los laboratorios 1 y 2 son muy similares: unimodales, con el mismo intervalo modal y con fuerte asimetría positiva.

### 4. Medidas características

En general, el análisis de los datos se comienza mediante gráficos. Posteriormente buscaremos medidas características que nos cuantifiquen los aspectos que más nos interesen. El Statgraphics proporciona medidas características de varias variables en casi los mismos lugares en los que proporciona Boxplots múttiples. Por ejemplo, si queremos conocer las medias y las asimetrías de las tensiones de rotura de los 10 laboratorios (fichero rotura.sf3) podemos ir a Compare/Multiple Samples/Multiple Sample Comparison. Allí seleccionamos Tabular Options y Summary Statistics



Si nos posicionamos en la ventana de resultados y pulsamos el botón derecho de ratón, podemos seleccionar Pane Options, y allí escoger las medidas características que nos interesen:

		Pane Option: Analysis Opt	ons			
		Undo Cut <b>Copy</b> Paste		Ctrl+Z Ctrl+X Ctrl+C Ctrl+V		
		Print Print Preview	·	F4 Shift+F3		
Multiple-Sam	ple Comparison					
	· 監 🕸 🖄 🖼 ?	<u>*_</u> Lbi:	Row:	譜		
Summary Stati	stics					
	Count	Average	Skewness	Summary Statis	t <mark>.</mark> cs Options	
				🔽 Average	🥅 Min.	🔽 Skewness
LAB1	100	4,7659	1,12998	🗌 Median	Max.	🔲 Std. Skewness
LAB3	100	5,2422	1,35638	E Mada	E Paras	- Kuntania
LAB4	100	5,0333	2,51899	I Mode	i nange	I NUITUSIS
LAB5	100	5,5892	2,81372	🔲 Geo. Mean	🔲 Lower Quartile	🔲 Std. Kurtosis
LAB6	100	5,1124	3,1749	□ Variance	Upper Quartile	Coeff of Var
LABS	100	5,451	2,17343 0.997454	j vanarice	) Opper quartie	
LAB9	100	5,2225	3,14587	📃 Std. Deviation	ı 📃 İnterquartile Range	🔲 Sum
LAB10	100	4,9422	1,20103	🔲 Std. Error		
Total	1000	5,20971	2,42129	ОК	Cancel All	Help
						1

Puede verse que, efectivamente, las distribuciones son muy asimétricas

Información de las medidas características

En el caso de las alturas de chicas y chicos tenemos el siguiente resultado:

📑 Multiple-Sa	imple Comparison									
	🔁 🛒 🍄 💱 🖾 🗳 🛀	] сы: [	Bow:							
Summary Statistics for altura										
sexo	Count	Average	Median	5	ummary Statistics	Dptions		×		
0	32	165,313	165,0	1	✓ Average	🔲 Min.	🔽 Skewness			
1	63	179,349	180,0	. 1	✔ Median	🔲 Max.	🔲 Std. Skewness			
Total	95	174,621	177,0	ſ	Mode	Range	🗖 Kurtosis			
sexo	Variance	Lower quartile	Upper quartile	e ľ	Geo. Mean	🔽 Lower Quartile	🔲 Std. Kurtosis			
0	19,6411	161,5	168,5	1	<ul> <li>Variance</li> </ul>	🔽 Upper Quartile	Coeff. of Var.			
1	25,36	177,0	182,0	ſ	Std. Deviation	🔽 Interquartile Range	🗖 Sum			
Total	67,6847	168,0	180,0	_ 1	Std. Error					
sexo	Interquartile :	range Skewness			ОК	Cancel All	Help			
0	7,0	0,23093								
1 	5,0	-0,293724					-	-		
Total	12,0	-0,302876			120		L L	-		

Los resultados muestran que en media, los chicos son 179.349-165.313=14.036 cm. más altos que las chicas. Llama la atención que el 50% de los chicos mide (declara medir) más de 180 cm. La varianza de los chicos es mayor que la de las chicas, pero su rango intercuartílico es menor. Ambas distribuciones son bastante simétricas, pues los coeficientes de asimetría son pequeños.

Opciones similares podemos conseguir en Describe/Numeric Data/Multiple-Variable Analysis.

File	Edit	Plot	Describe	Compare	Relate	Speci	al View	Window	Help	
a		<b>       </b>	Numeri	c Data		÷	One-Va	ariable Ana	lysis	
_			Catego	rical Data		•	Multiple	e-Variable i	Analysis	
	[wo-9	Sampl	Distribu	itions		- <b>F</b>	Subset	Analysis		
			Life Da	ta		•	Row-W	/ise Statist	ics	
N N			Hypoth	esis Tests.			Power	Transform	ations	
<b>1</b> 2	1	- ~	Sample	-Size Deter	mination.		Statisti	cal Tolerar	nce Limits.	