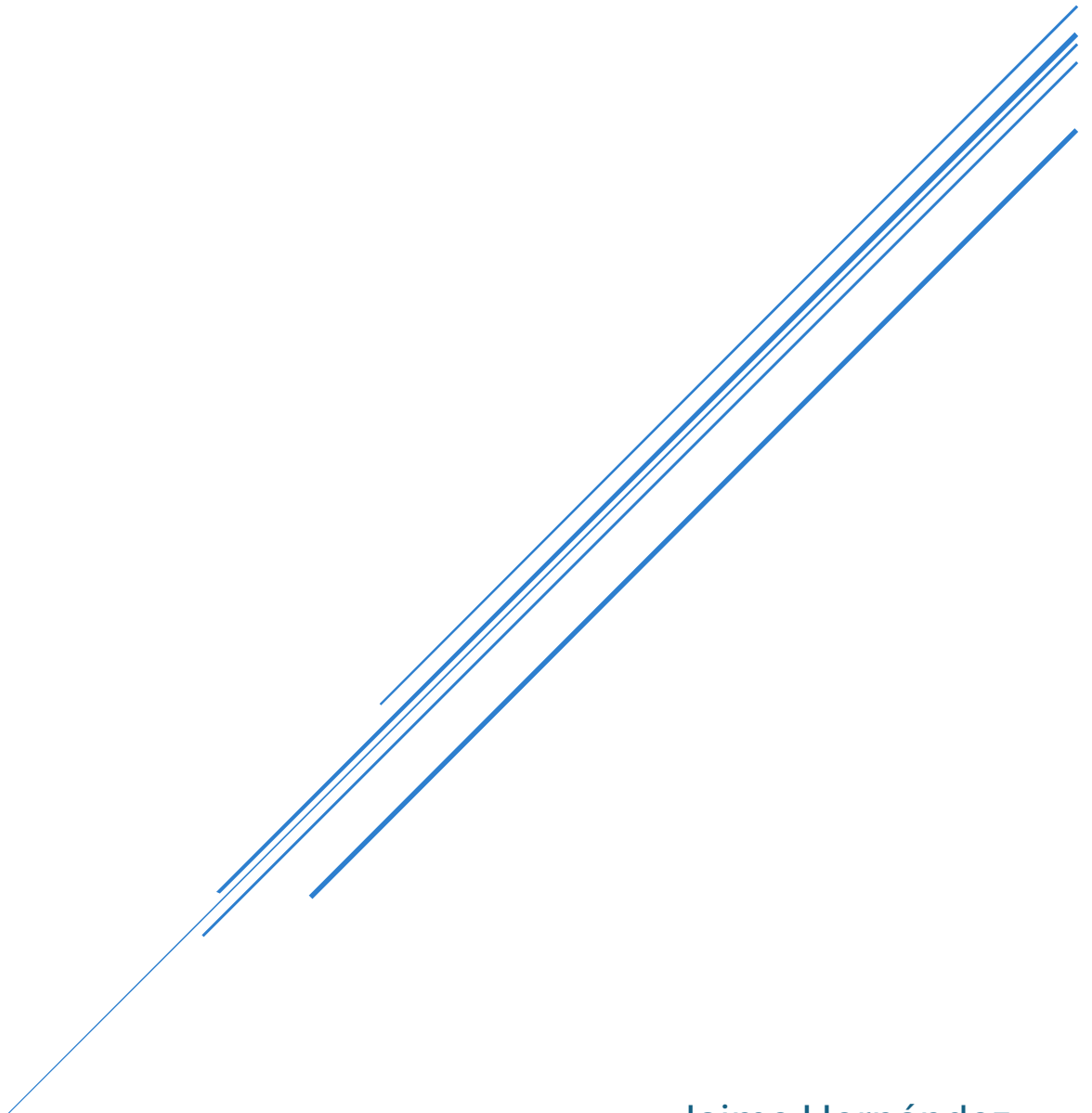


PRÁCTICA 3

Estadística



Jaime Hernández
Universidad de Alicante

Ejercicio 1:



El rendimiento –referido a capacidad de procesamiento– de los 60 clusters de los distintos departamentos de una gran empresa es el siguiente, medido en GFLOPS (109 operaciones en coma flotante por segundo):

Para analizar la distribución de la capacidad de procesamiento disponible en la empresa, se pide:

- a) Construye la tabla de frecuencias completa.
- b) Representa el histograma.
- c) Explica e interpreta los resultados obtenidos en los apartados anteriores.

En las siguientes capturas se muestra como se ha creado la tabla de frecuencia y el histograma:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Rendimiento	Numeric	8	2		None	None	8	Right	Unknown
2	Clusters	Numeric	8	2		None	None	8	Right	Unknown

	 Rendimiento	 Clusters
1	30	1
2	31	1
3	61	17
4	91	30
5	121	3
6	161	4
7	181	2
8	211	2

Recode into Different Variables

Clusters

Numeric Variable -> Output Variable:
Rendimiento --> rendimiento_intervalo

Output Variable
Name:
rendimiento_intervalo
Label:

Change

Old and New Values...

If... (optional case selection condition)

OK Paste Reset Cancel Help

Recode into Different Variables: Old and New Values

Old Value

☒ Value:

☐ System-missing

☐ System- or user-missing

☐ Range:

through

☐ Range, LOWEST through value:

☒ Range, value through HIGHEST:

☐ All other values

New Value

☒ Value:

☐ System-missing

☐ Copy old value(s)

Old -> New:

31 thru 60 --> 'De 31 a 60'
61 thru 90 --> 'De 61 a 90'
91 thru 120 --> 'De 91 a 120'
121 thru 150 --> 'De 121 a 150'
151 thru 180 --> 'De 151 a 180'
181 thru 210 --> 'De 181 a 210'
210 thru Highest --> 'Mas de 210'

Add Change Remove

☒ Output variables are strings Width: 18

☐ Convert numeric strings to numbers ('5'-->5)

Continue Cancel Help

Clusters	rendimiento_intervalo
1	Menos de 31
1	Menos de 31
17	De 61 a 90
30	De 91 a 120
3	De 121 a 150
4	De 151 a 180
2	De 181 a 210
2	Mas de 210

Rendimiento

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	30	1	12,5	12,5	12,5
	31	1	12,5	12,5	25,0
	61	1	12,5	12,5	37,5
	91	1	12,5	12,5	50,0
	121	1	12,5	12,5	62,5
	161	1	12,5	12,5	75,0
	181	1	12,5	12,5	87,5
	211	1	12,5	12,5	100,0
Total		8	100,0	100,0	

Chart Builder

Variables: Rendimiento, Clusters, rendimiento_intervalo

Chart preview uses example data

Simple Bar Mean of Clusters by Rendimiento

Mean Clusters

Rendimiento

No categories (scale variable)

Filter?

Gallery Basic Elements Groups/Point ID Titles/Footnotes

Choose from:

Favorites

Bar

Line

Area

Pie/Polar

Scatter/Dot

Histogram

High-Low

Boxplot

Dual Axes

OK Paste Reset Cancel Help

Element Properties Chart Appearance Options

Edit Properties of:

Bar1

X-Axis1 (Bar1)

Y-Axis1 (Bar1)

Title 1

Axis Label: Mean Clusters

Scale Range

Variable: Clusters

Automatic Custom

Minimum ☒ 0

Maximum ☒ 0

Major Increment ☒ 0

Origin ☒ 0

Scale Type

Type: Linear

Base: 10

Exponent: 0,5

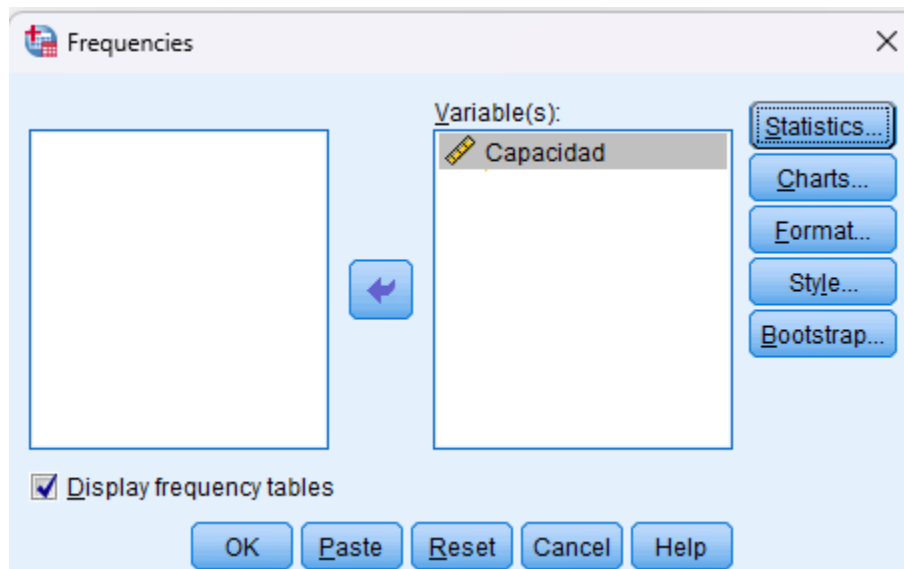
La distribución de los clusters en la empresa muestra que la mayoría se concentra en el rango de 91 a 120 GFlops, lo que sugiere que esta es la capacidad de procesamiento predominante. Hay pocos clusters en los extremos de rendimiento, es decir, en los rangos de menos de 31 y más de 210 GFlops, lo que indica que estos valores extremos son inusuales en la infraestructura de la empresa. En general, esto implica que la capacidad de procesamiento está centrada en valores moderados, probablemente optimizados para las necesidades operativas habituales, mientras que las capacidades extremas se encuentran en menor cantidad, quizás reservadas para tareas o departamentos específicos que requieren procesamiento muy bajo o excepcionalmente alto.

Ejercicio 2: De la misma empresa, se ha contabilizado la capacidad de almacenamiento de los clusters, medida en GB, obteniendo

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Capacidad	Numeric	8	0		None	None	8	Right	Unknown
2										
3										

Primero hemos insertado todos los elementos:

	Capacidad
1	730
2	600
3	680
4	590
5	620
6	760
7	830
8	610
9	800
10	790
11	600



Ejercicio 3: Las dos tablas siguientes muestran datos sobre peticiones http diarias sobre un servidor web. La primera tabla representa las peticiones de URLs internas y la segunda las externas.

Colocamos los datos correspondientes, en las siguientes capturas se presentan los pasos para hacer un análisis de los datos:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Internas	Numeric	8	0		None	None	8	Right	Unknown
2	Exetmas	Numeric	8	0		None	None	8	Right	Unknown

Internas	Exetmas
345	324
634	39
456	519
32	21
666	984
9	720
671	921
754	452
399	52
621	410
43	61
333	197
71	317
374	470

Descriptives

Variable(s):

- Internas
- Exetrnas

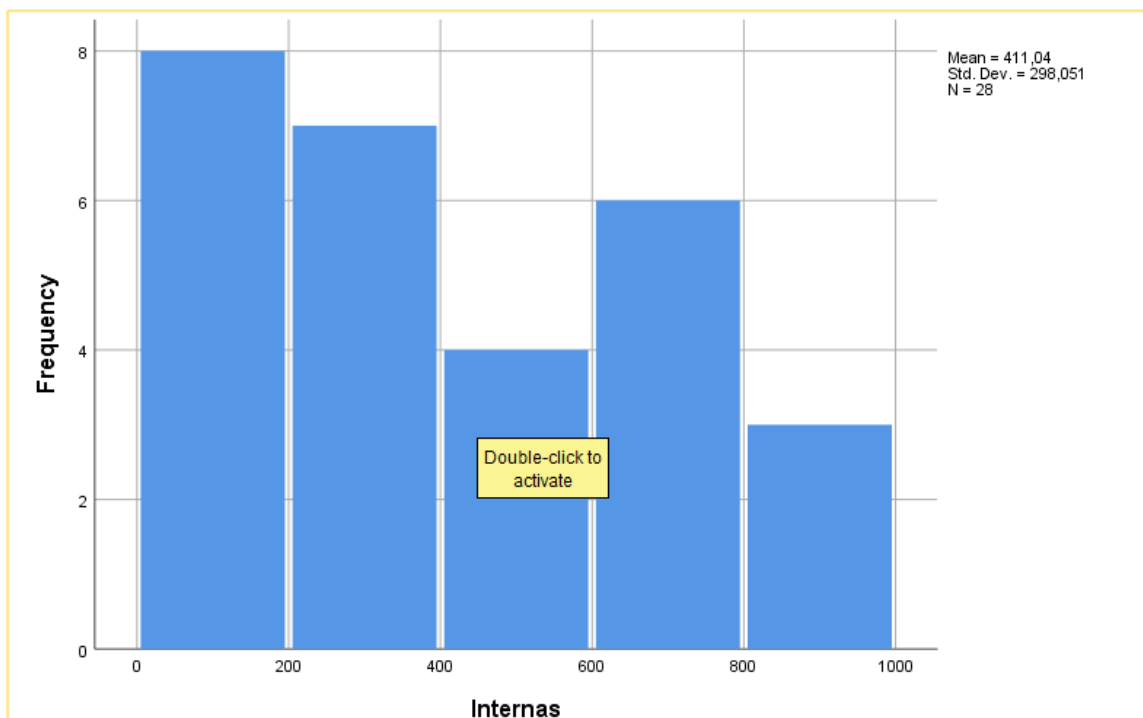
☐ Save standardized values as variables

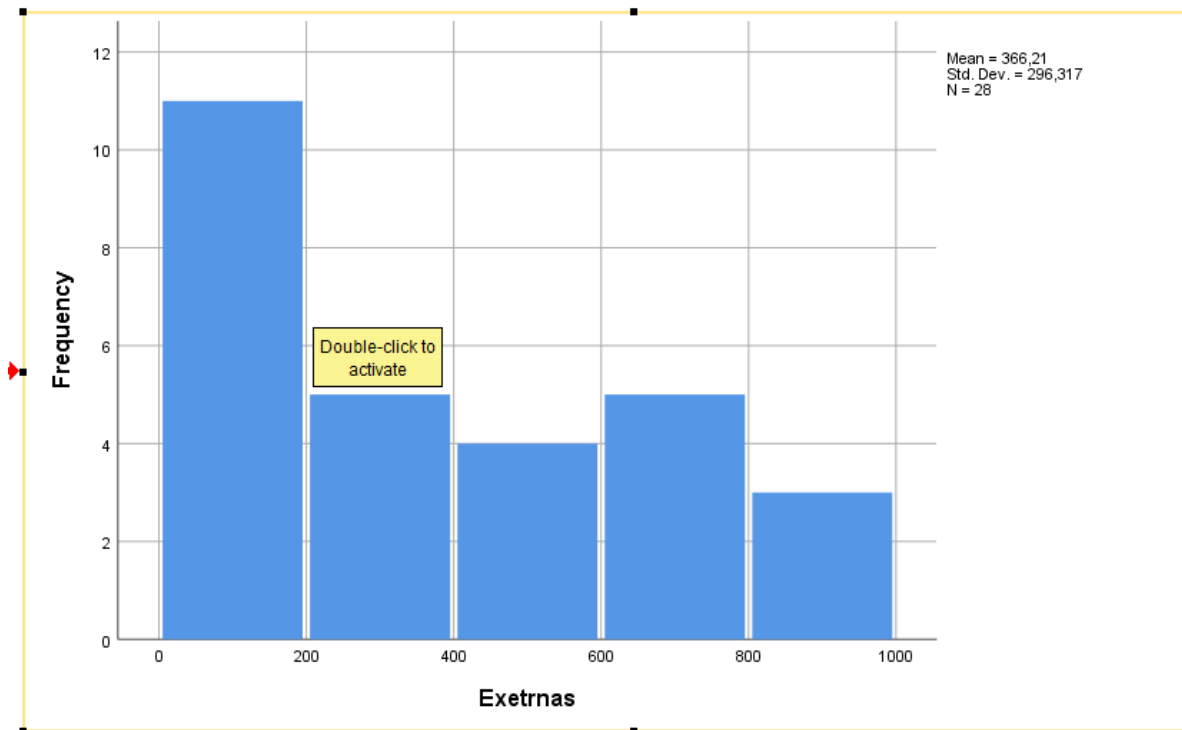
Options... Style... Bootstrap...

OK Paste Reset Cancel Help

Descriptive Statistics

	N Statistic	Range Statistic	Minimum Statistic	Maximum Statistic	Sum Statistic	Mean Statistic Std. Error		Std. Deviation Statistic	Variance Statistic
Internas	28	934	9	943	11509	411,04	56,326	298,051	88834,480
Exetrnas	28	963	21	984	10254	366,21	55,999	296,317	87803,952
Valid N (listwise)	28								





Al analizar las distribuciones de peticiones HTTP internas y externas mediante medidas de tendencia central, se puede observar si existen diferencias en el número promedio de peticiones diarias entre ambos tipos de URLs, destacando las medias y medianas como indicadores clave. La dispersión, evaluada mediante la desviación estándar y representada en los boxplots, permite identificar cuál de las dos distribuciones presenta mayor variabilidad, lo que indicaría fluctuaciones más pronunciadas en la cantidad de peticiones. Los histogramas y boxplots también revelan la forma de las distribuciones, mostrando si son simétricas o si tienen algún sesgo, lo que puede indicar concentraciones de peticiones en ciertos rangos. Finalmente, las conclusiones generales ayudarán a determinar si existe una mayor cantidad de peticiones en URLs internas o externas y si una de las distribuciones tiene patrones específicos, como variabilidad alta o presencia de valores atípicos, proporcionando una visión completa del comportamiento de las peticiones en el servidor web.