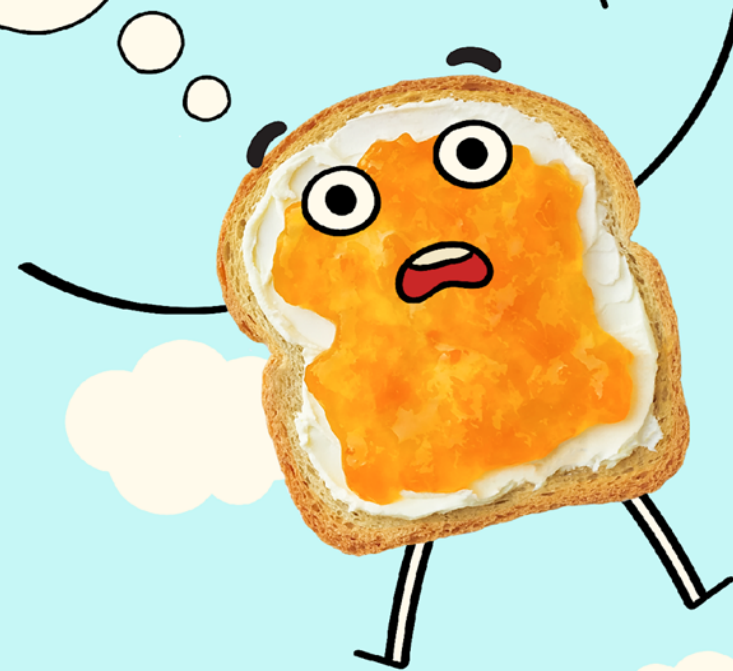


Quarta Jornada de divulgació del món de l'estadística, 4 i 7 de març 2024

Per a què serveix L'ESTADÍSTICA?

ESTADÍSTICA ADREÇADA A BATXILLERAT



UNIVERSITAT DE
BARCELONA



LaUB
divulga



Guadalupe Gómez Melis



Concepció Arenas Sola



Leire Garmendia Bergés



Yovaninna Alarcón Soto



Andrea Toloba López-Egea



Mireia Besalú Mayol



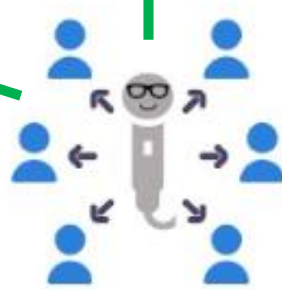
Antonio Miñarro Alonso



Nuria Pérez Álvarez



Cristian Tebé Cordoní

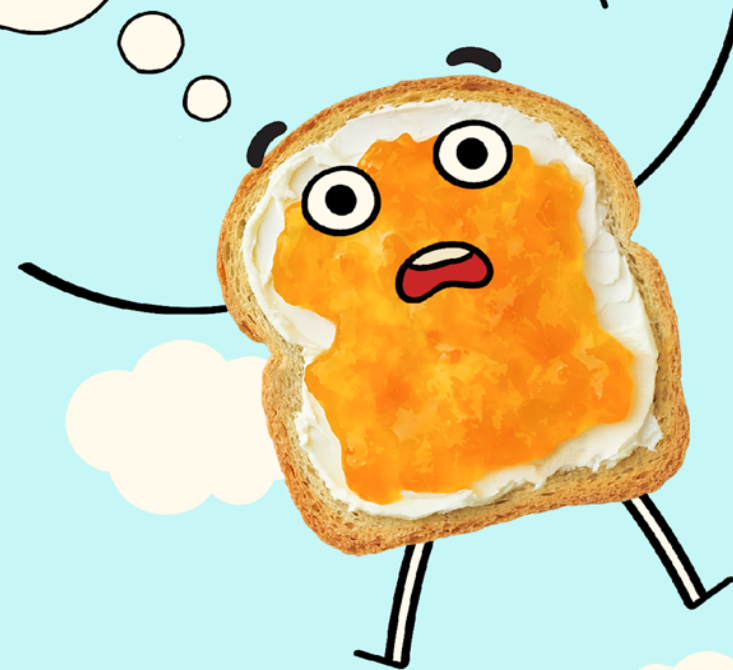
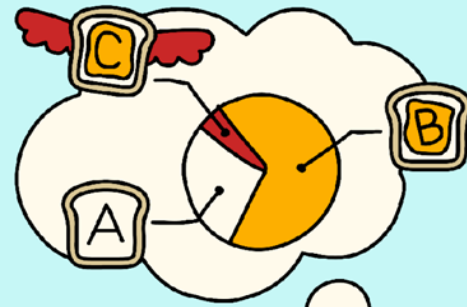


GRBIO Divulga

Quarta Jornada de divulgació del món de l'estadística, 4 i 7 de març 2024

Per a què serveix L'ESTADÍSTICA?

**ESTADÍSTICA ADREÇADA A
BATXILLERAT**
Part II: Nuria Pérez Álvarez



UNIVERSITAT DE
BARCELONA



LaUB
divulga

Estadística Descriptiva Bàsica



Year	Rochester	Vancouver
2015	17.7	0.46
2014	17.7	3.12
2013	19.9	1.43
2012	31.4	6.1
2011	33.8	0.38
2010	29.3	16.95
2009	15.2	6
2008	29.8	8.53
2007	11.1	7.48

Statistical summary for 'Returns 1' (Rochester):

- Min: 4
- Mean: 28.577
- Std. Deviation: 10.577

Menu options:

- AVERAGE
- AVERAGEA
- AVERAGEIF
- AVERAGEIFS



Eines de computació



Eines:

Excel

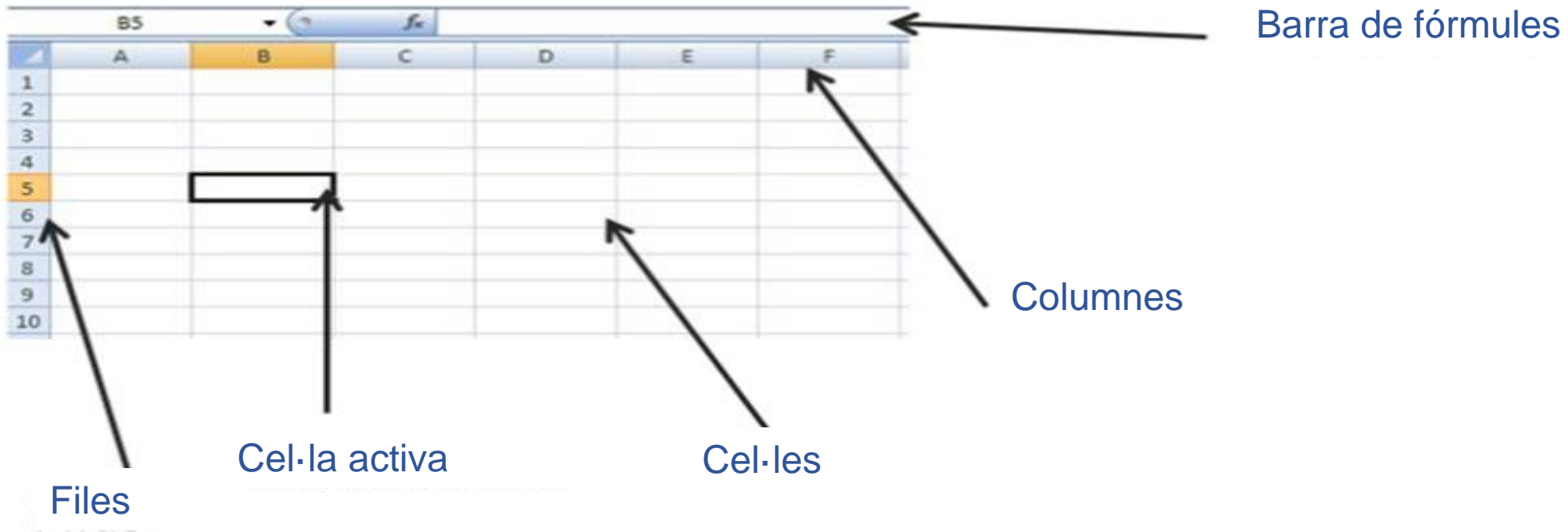
Full de càlcul de Google

Paquets estadístics i d'anàlisi de dades (gratuït): R

Focus:

Excel, extrapolable fàcilment a altres full de càlcul

Components d'una fulla de càlcul d'Excel



Les fórmules són operacions que Excel farà amb uns determinats valors que indicarem ja sigui directament o per posició a les diferents cel·les.

Les funcions i la seva utilitat per descriure una variable



- ❖ Funcions: fórmules predefinides per Excel que operen amb un o diversos valors
- ❖ Arguments: valors sobre els quals operen les funcions . Generalment són números, depenent de la funció poden ser també textos, valors lògics, fórmules o fins i tot altres funcions
- ❖ Gran varietat de funcions, extensible amb complements (ex: mòdul anàlisi estadístic de dades)

Inserir una funció a Excel

❖ Col·locar-nos a la cel·la on volem que aparegui el resultat de la funció.

❖ Inserir una funció:

- ❖ Si no coneixem la denominació de la funció, podem utilitzar l'assistent d'Excel
- ❖ Seleccionar directament la funció entre els diferents tipus en cas de conèixer la seva denominació

❖ Nota:

La funció que seleccionem està explicada (arguments requerits, definició, ...)

The screenshot shows the Microsoft Excel interface with the 'Insert Function' dialog box open. The dialog box is titled 'Insert Function' and has a search bar with the text 'Go' and a 'Go' button. Below the search bar, there is a dropdown menu for 'Or select a category' set to 'Most Recently Used'. The 'Select a function' list shows 'VLOOKUP' selected. Below the list, the function signature 'VLOOKUP(lookup_value;table_array;col_index_num;range_lookup)' and a brief description are visible. At the bottom, there is a 'Help on this function' link and 'OK' and 'Cancel' buttons.

Mòdul d'anàlisi de dades d'Excel

Excel també permet obtenir diferents mesures estadístiques (i altres anàlisis estadístiques) mitjançant un complement (no instal·lat per defecte) que és el **Mòdul d'Anàlisi de Dades**.

Instal·lació:

Arxiu-> Opcions -> Complementos, seleccionar Eines per anàlisi i clicar al botó Acceptar...

S'instal·larà un nou conjunt d'icones a la cinta de Dades

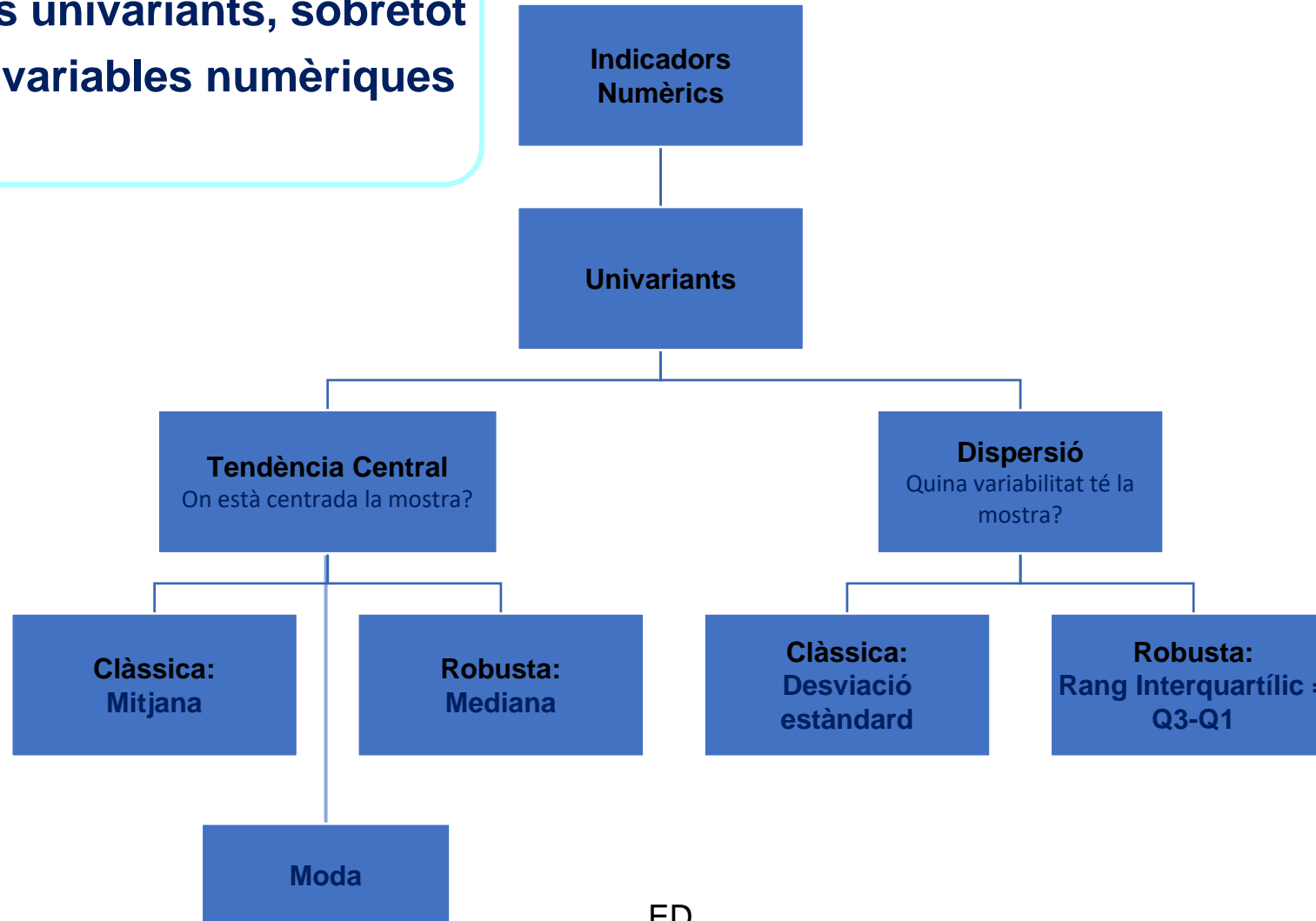


OPCIONAL. No és necessari per al que hem de fer

Eines numériques

Eines numèriques per la descriptiva univariant

👍 Els indicadors univariants, sobretot tenen sentit per variables numèriques contínues



Càlcul dels indicadors clàssics

Mitjana

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

👉 **n**: longitud de la mostra
 x_i : observació i-èsima

Variància

$$Var_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$Var_x = \frac{1}{n} \left[\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right]$$

Desviació tipus o estàndard

$$SD_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$SD_x = \sqrt{\frac{1}{n} \left[\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right]}$$

👉 La variància i la desviació tipus són mesures de la variabilitat de les dades. La primera és el quadrat de la segona.

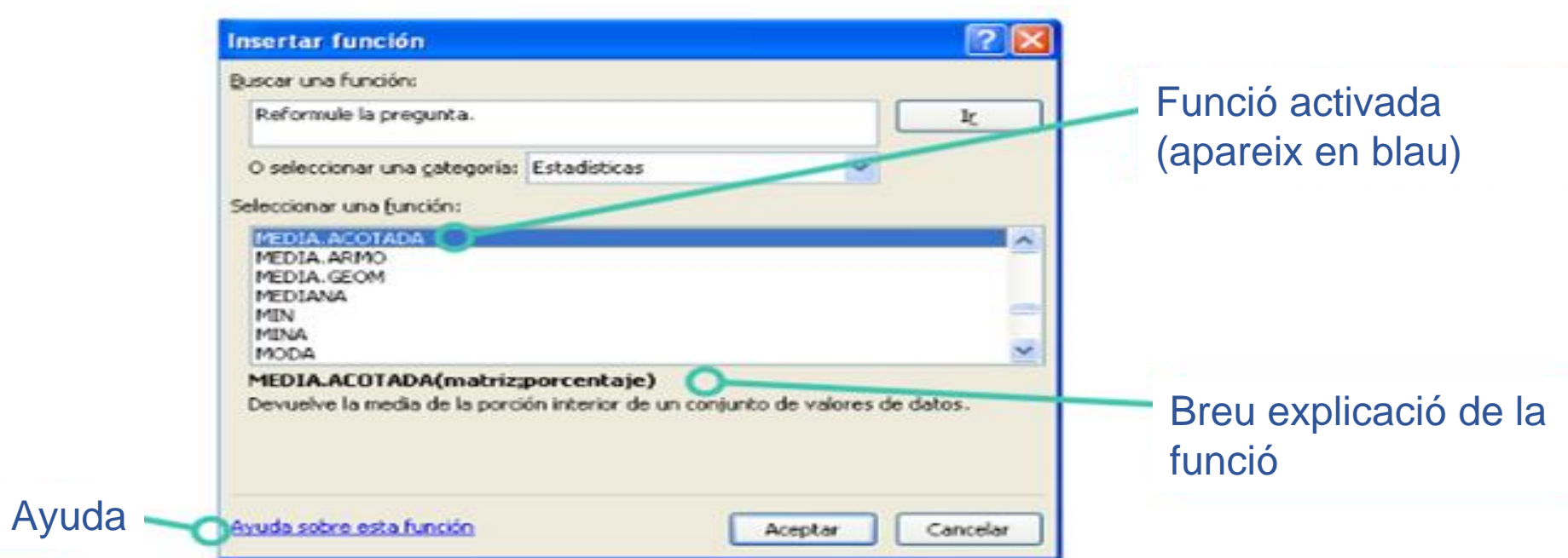
👉 Expressió de la variància i la desviació tipus (SD, de l'anglès: Standard Deviation) sense corregir. La versió corregida s'obté fent la divisió per (n-1) enlloc de n.

Mitjana (I)

Podríem introduir “mitjana” (en l’idioma de l’Excel) al requadre superior i pressionar “Acceptar” o buscar-la dins del menú en el tipus “estadístiques”.

Seleccionem l’opció desitjada.

ATENCIÓ! en activar una funció en el quadre “seleccionar una funció”, l’Excel mostra una breu explicació de la funció. Si no és suficient, podem utilitzar l’opció “ajuda sobre aquesta funció” on trobarem més informació sobre la funció.



Mitjana (II)

Gran varietat de funcions que ofereix Excel;
Idioma (Promedio, Average,...)

Arguments de la funció:



The dialog box is titled "Argumentos de función" and contains the following information:

- Function name: PROMEDIO
- Argument 1: "Número1" with the value "D5:D211" and a list of numbers: {8|7|7|8|7|8|7|8|8|7|7|8|9|8|8|...}
- Argument 2: "Número2" with the value "número"
- Result: "= 7,251219512"
- Description: "Devuelve el promedio (media aritmética) de los argumentos, los cuales pueden ser números, nombres, matrices o referencias que contengan números."
- Usage: "Número1: número1;número2;... son entre 1 y 255 argumentos numéricos de los que se desea obtener el promedio."
- Formula result: "Resultado de la fórmula = 7,251219512"
- Buttons: "Aceptar" and "Cancelar"
- Link: "Ayuda sobre esta función"



f_x	=PROMEDIO(D5:D211)	
C	D	E

Estadístics de totes les dades mitjançant les funcions estadístiques

Podem calcular mesures estadístiques de totes les dades mitjançant les funcions estadístiques. Des de Fórmules -> Més Funcions -> Estadístiques seleccionarem la funció PROMEDIO() que calcula la **mitjana aritmètica** d'un conjunt de dades.

Variància

VAR.P() (o **VARP()**) calcula la variància (també coneguda com a variància NO corregida)

VAR.S()(o VAR()) calcula la variància corregida (també coneguda com quasi-variància),

VARA() i VARPA() que tenen en compte valors de text i lògics (a nosaltres no ens interessen).

Desviació tipus o estàndard

Anàlogament:

DESVEST.P() (o **DESVESTP()**)

DESVEST(),

DESVESTA() i DESVESTPA()

que són les arrels quadrades de les anteriors.



Càlcul dels indicadors robusts

Quartil 1

$$\text{Posició Q1} = (n+1)/4$$

$$n=243$$

$$\text{Q1} = X_{(61)}$$

Mediana (Quartil 2)

$$\text{Posició Q2} = (n+1)/2$$

$$\text{Q2} = X_{(122)}$$

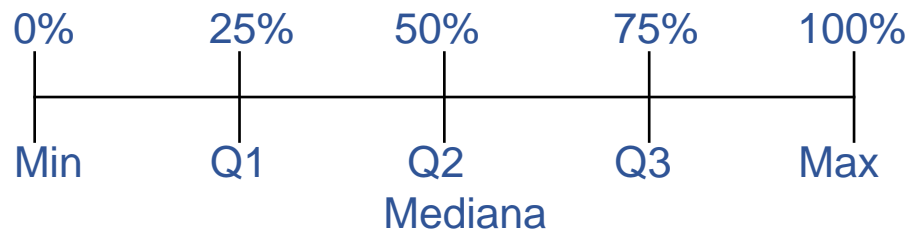
Quartil 3

$$\text{Posició Q3} = 3 \cdot (n+1)/4$$

$$\text{Q3} = X_{(183)}$$

Rang Interquartílic (IQR de l'anglès Interquartile Range)

$$\text{IQR} = \text{Q3} - \text{Q1}$$



👉 La mediana (Q2) és el valor de la mostra que deixa el 50% de les observacions per sota i l'altre 50% per sobre. El Q1 és el valor que deixa el 25% de les observacions per sota i el Q3 és el que deixa el 75% de les observacions per sota

👉 Si el càlcul de la posició dóna un nombre enter, llavors el quartil corresponent serà el nombre que ocupi aquella posició en la llista de valors endreçats (de menor a major).

👉 Si el càlcul dóna un nombre no enter, es ponderaran els valors corresponents a la posició entera anterior i posterior segons la part decimal de la posició.

Altres estadístics de resum que es poden calcular

En el cas del quartil, a més d'indicar les cel·les de les dades és necessari indicar el número de quartil que volem.

Així es reflecteix en la sintaxis després de l'àrea amb un “;”
(=CUARTIL(AA5:AA211;1)).



Mitjana i Mediana

Vocabulari:

Llengua	Concepte 1	Concepte 2
Català	Mitjana	Mediana
Castellà	Media	Mediana
Anglès	Mean	Median

NOTA: [Mediana o mitjana?](#)

Exemple, en el sou a Catalunya (euros anuals el 2021):

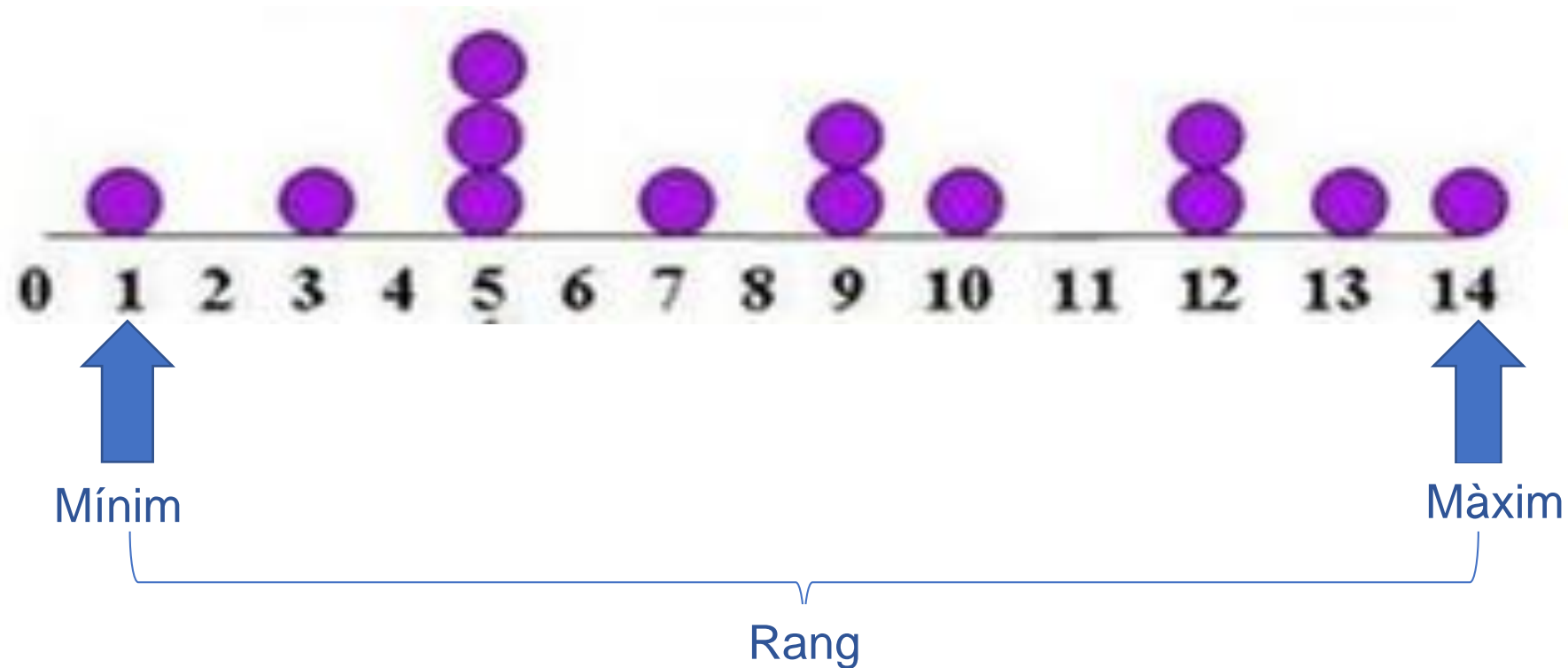
Mitjana	28.145,02
Percentil 10	10.898,33
Primer quartil	16.355,41
Mediana	23.962,36
Tercer quartil	35.367,64
Percentil 90	49.635,94

Màxim, Mínim i Rang

Màxim: Límit superior o valor més gran observat

Mínim: Límit inferior o valor més petit observat

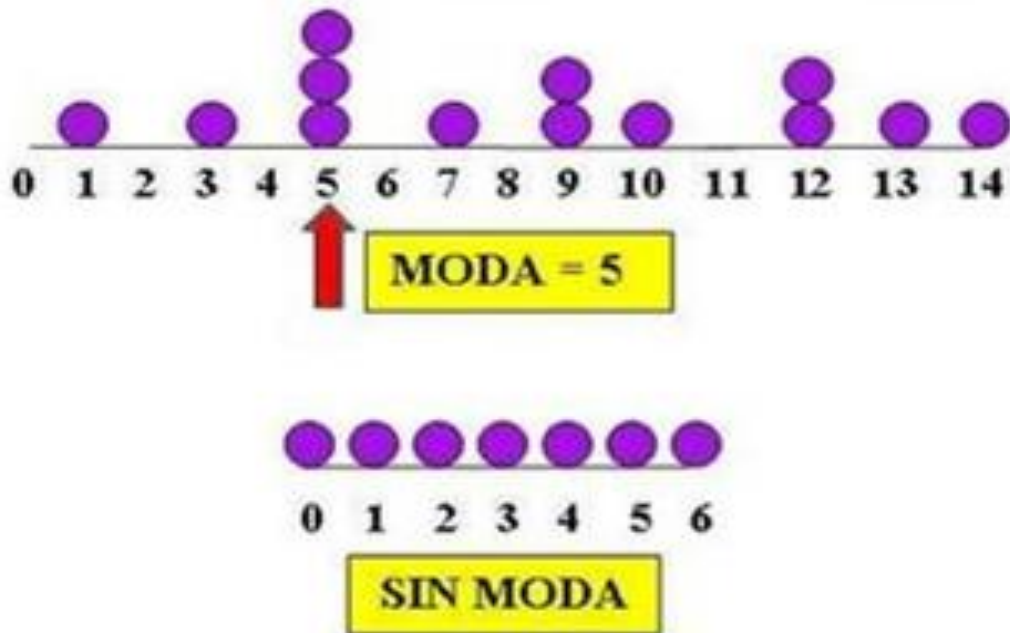
Rang: Diferència entre el valor més gran i el més petit



Moda

Valor que té una freqüència més gran.

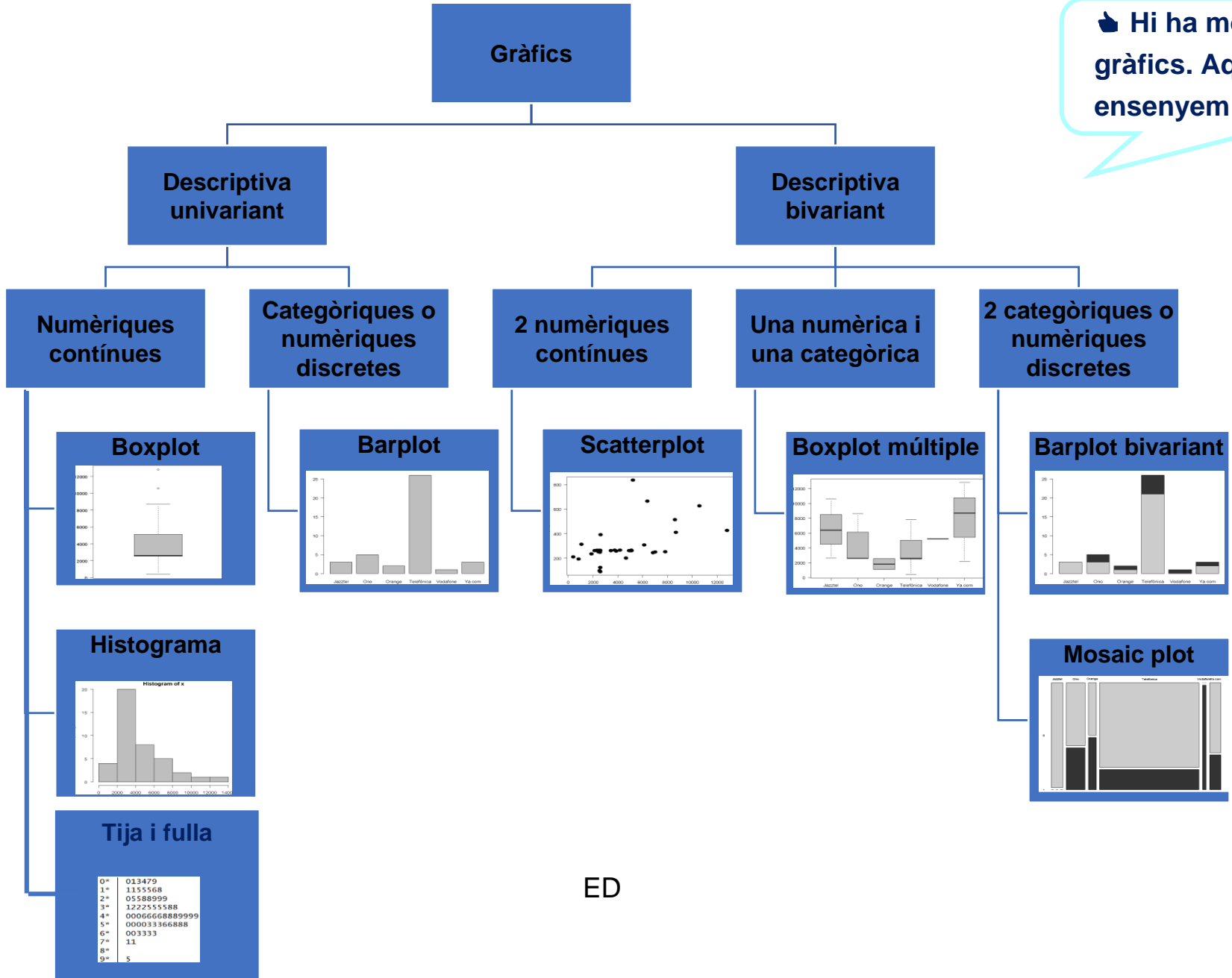
Quan la variable discreta pot prendre molts valors diferents o és contínua, llavors es parla de classe o interval modal, que és la classe que té una freqüència més gran.



👉 Gràficament es determina mirant al diagrama de barres (si la variable és discreta) o a l'histograma el valor màxim. Quan el diagrama de barres o l'histograma tenen diversos màxims locals, aleshores es parla de variables o distribucions estadístiques bimodals o trimodals, etc.

Eines gràfiques

Hi ha molts tipus de gràfics. Aquí només ensenyem els més habituals



Gràfics per la descripció univariant de variables numèriques

❖ Diagrama de tija i fulles

❖ Boxplot

❖ Histograma

❖ Altres

<https://zenodo.org/records/5060267>

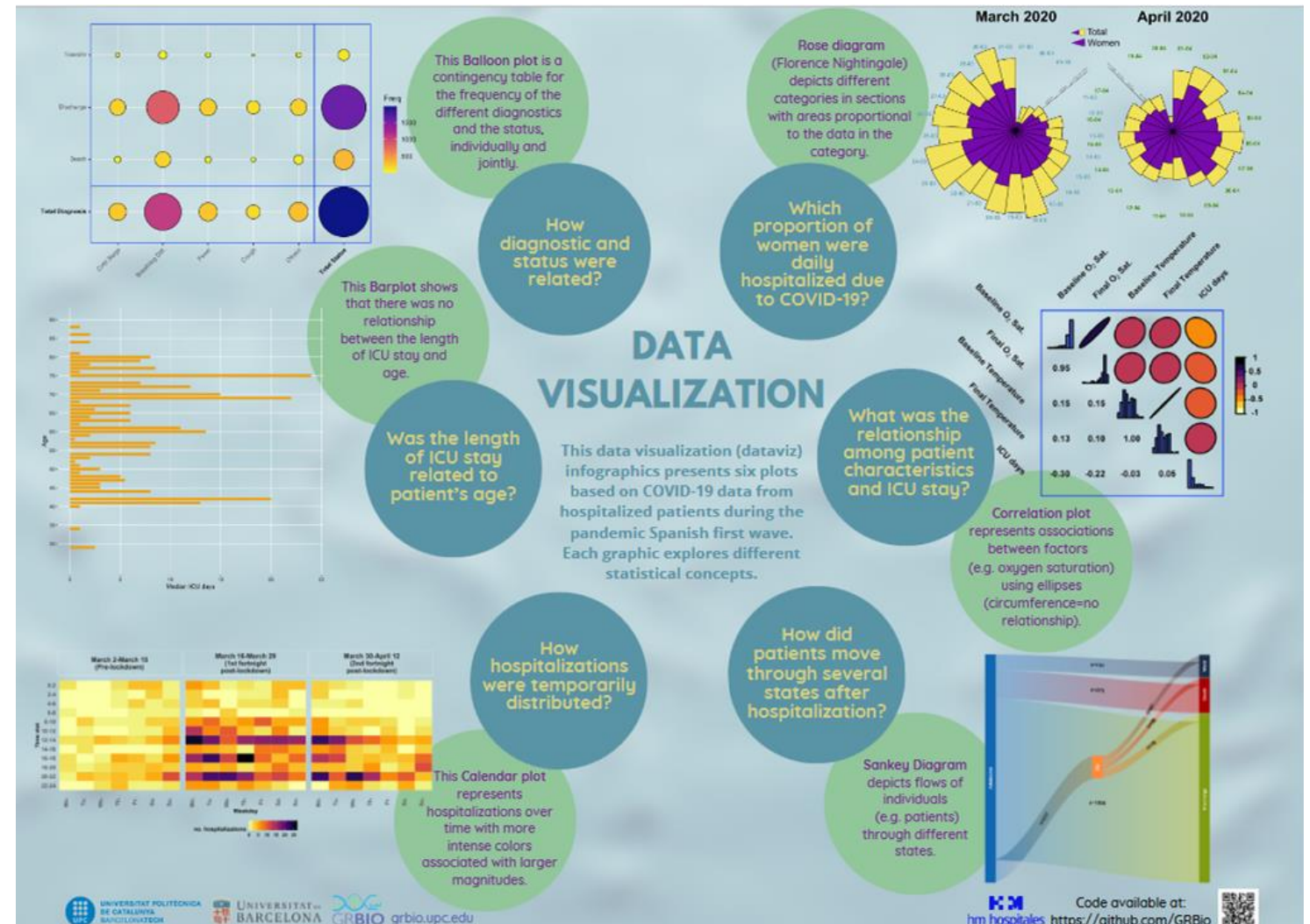


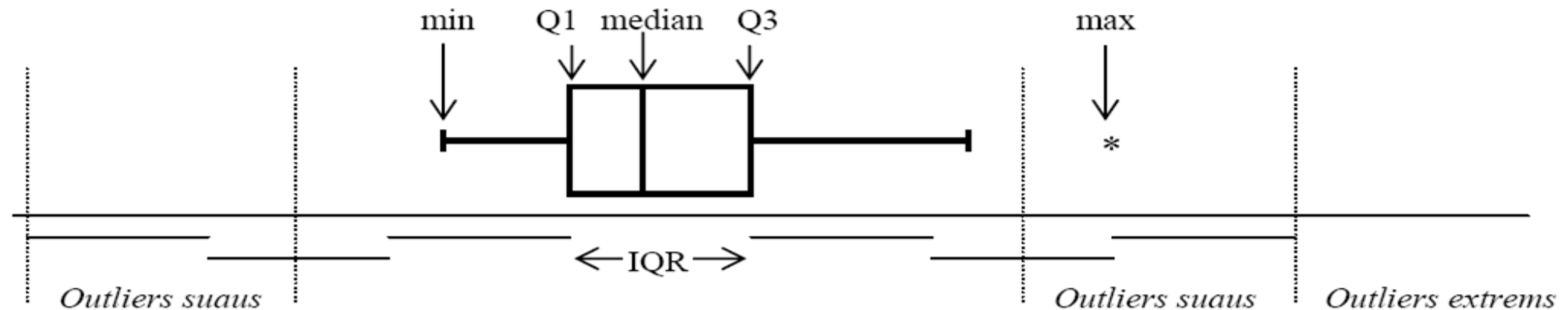
Diagrama de tija i fulla (stem and leaf)

- ❖ Tukey (1977)
- ❖ Proporciona una ordenació de totes les dades en intervals sense perdre els valors originals que i visualitzant la distribució.
- ❖ Per a dibuixar-lo: cada valor se separa en dues parts: el/s primer/s dígit/s, que rep/reben el nom de **tija** (stem), i l'últim dígit, anomenat **fulla** (leaf).
- ❖ Quan les dades han estat mesurades de forma molt precisa pot ser necessari arrodonir-les prèviament.

0*	013479
1*	1155568
2*	05588999
3*	1222555588
4*	000666688889999
5*	000033366888
6*	003333
7*	11
8*	
9*	5

Diagrama de caixa (Boxplot)

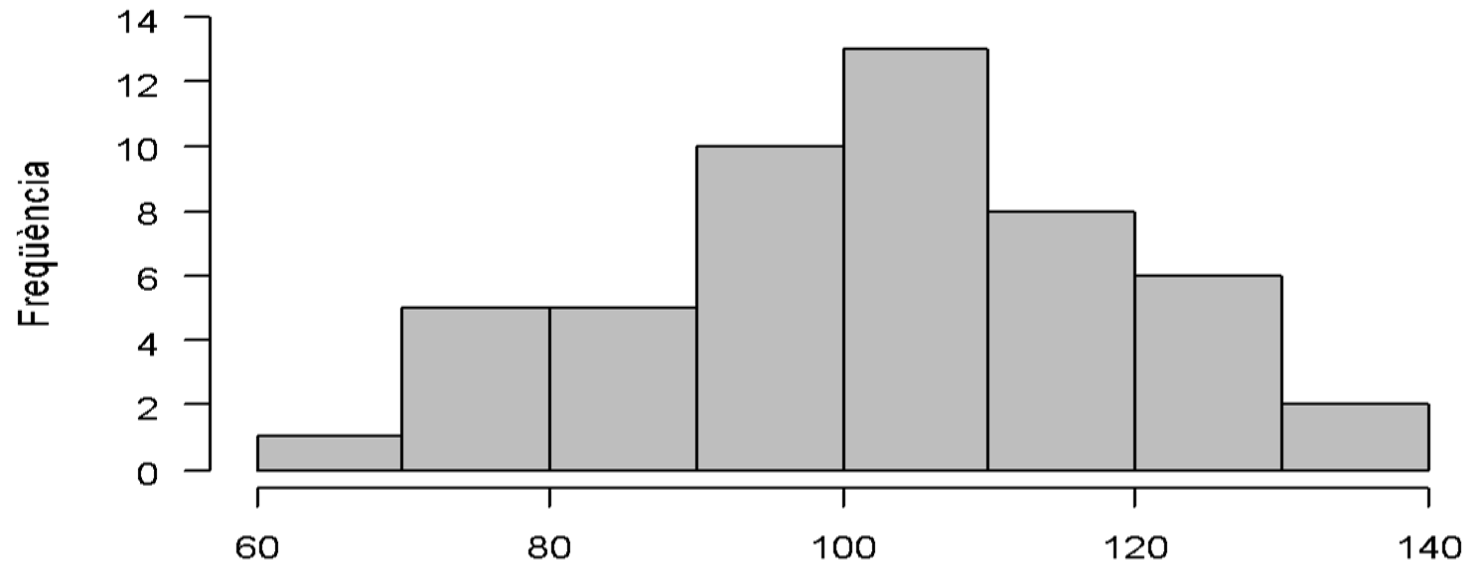
- ❖ Representa els indicadors robustos i els outliers (valors extrems o anòmals)
- ❖ Elements:
 - ❖ **Caixa.** Delimitada pel Q1 i pel Q3 i té una línia interior que representa la mediana.
 - ❖ **Bigotis.** surten des de la caixa i tenen una longitud de 1.5 vegades el IQR (IQR=longitud de la caixa). En cas de que el mínim de la mostra sigui major que el final del bigoti esquerra, aquest bigoti només arribarà fins al mínim. De forma anàloga es procedeix amb l'altra bigoti.
 - ❖ **Outliers.** Són els punts que queden més enllà dels bigotis del box-plot. Es consideren dades extremes o anòmales.



- ❖ Nota: No és habitual, però alguns softwares representen també la mitjana al boxplot.

Histograma

- ❖ Representa la **distribució** de la variable estudiada
- ❖ L'**eix horitzontal** conté els **valors** de la variable i l'**eix vertical** les **freqüències** (o proporcions)
- ❖ La **superfície de cada barra** és proporcional a la freqüència dels valors representats



👍 En aquest histograma, entre 90 i 100 hi ha 10 valors

Gràfics en Excel

Plantilles (Templates)

Veure: “Calculadora - Estadística descriptiva básica.xlsx”

Materials

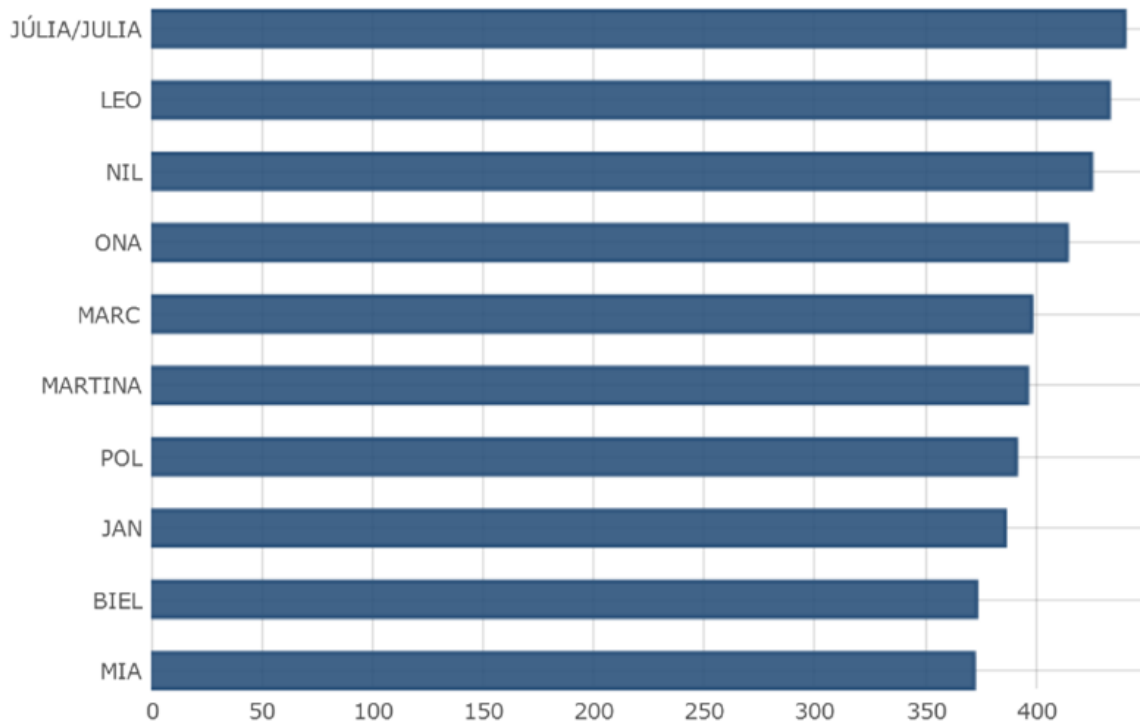
- ❖ Institut d'Estadística de Catalunya - IDESCAT
- ❖ Institut Nacional d'Estadística - INE

- ❖ Dades
- ❖ Exemples per explicar aquests conceptes a classe

- ❖ www.idescat.cat accedir a l'apartat Serveis i a la secció “Per al sector educatiu”
<https://www.idescat.cat/serveis/>
 - ❖ Visites a l'IDESCAT:
 - ❖ Sessions presencials (o virtuals –en format més breu) on s'expliquen les funcions i els serveis que ofereix l'IDESCAT
 - ❖ Sessions monogràfiques
 - ❖ 45-60 minuts
 - ❖ Aprofundir en temes específics que siguin d'interès. Per exemple, com es fa una enquesta, quines són les principals fonts demogràfiques o com són les macromagnituds de l'economia catalana.

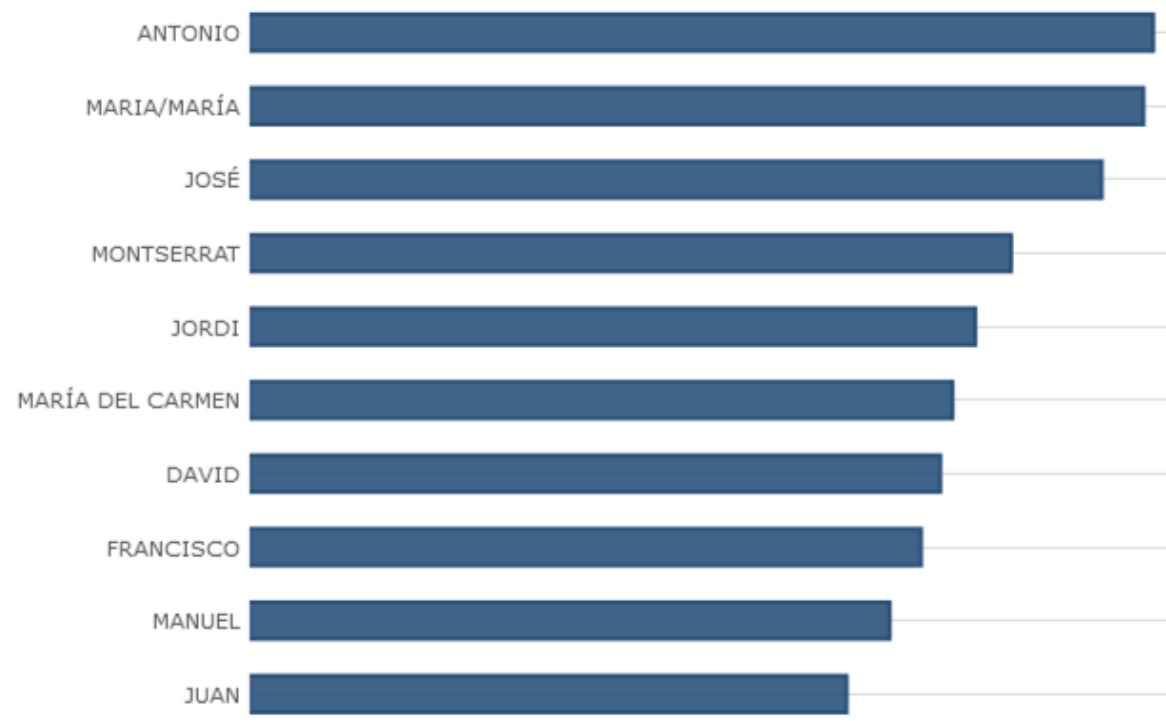
Noms més freqüents dels nadons i de la població

Noms més freqüents dels nadons. Catalunya. 2022

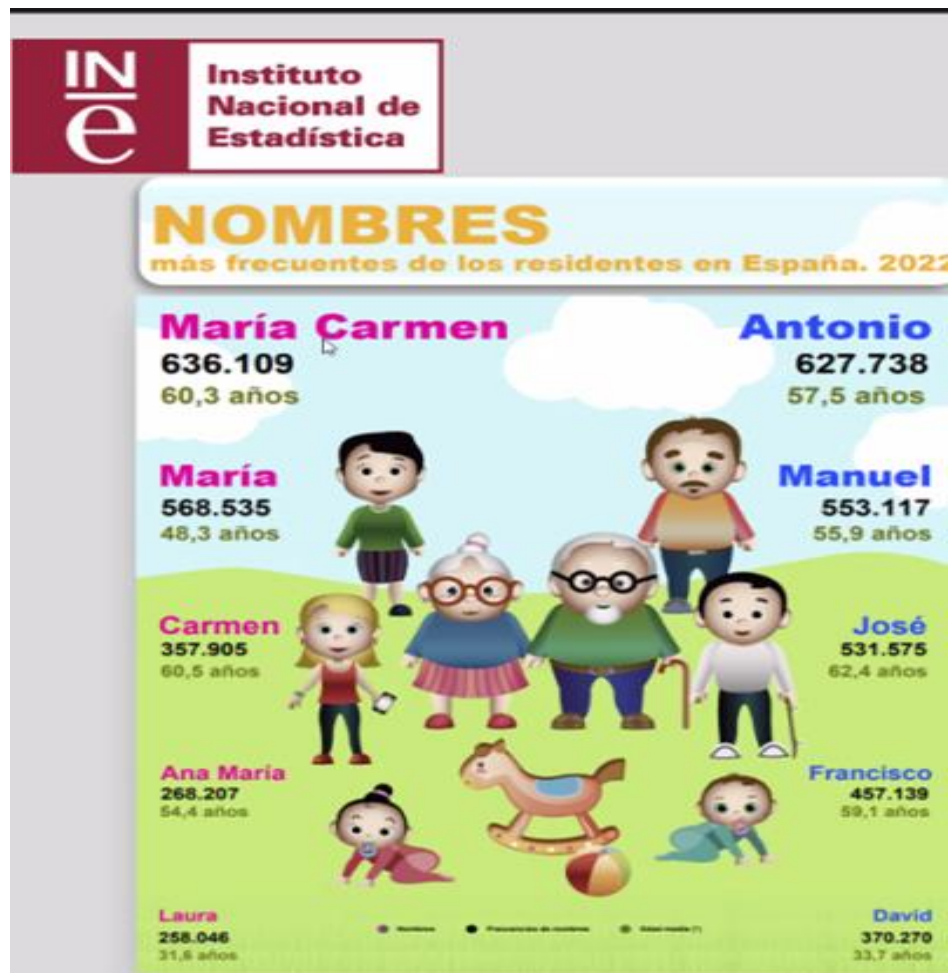


Font: Idescat, a partir de l'estadística de naixements.

Noms més freqüents de les persones. Catalunya. 2022

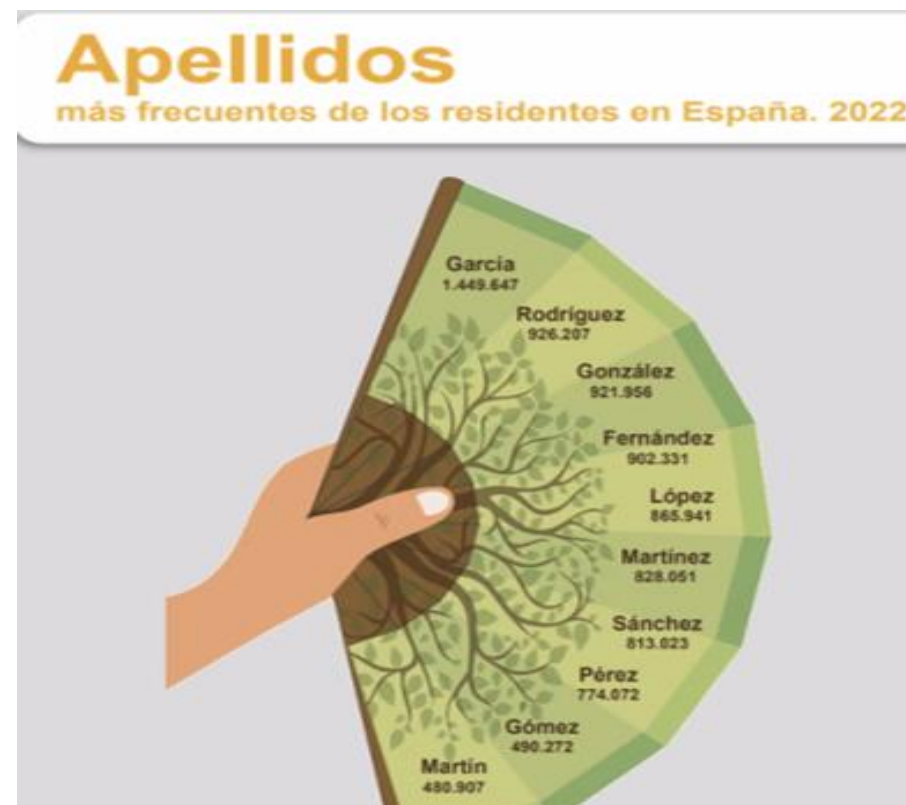


Font: Idescat, a partir del Padró continu de l'INE.



Es pot fer una classificació a la classe dels noms i cognoms i comparar-ho amb el que surt a l'INE.

Atenció a la mitjana d'edat!



Vídeo INE: Si España fuese un pueblo de 100 habitantes

<https://www.youtube.com/watch?v=LJpFD1pLmFA>

Si España fuese un pueblo de 100 habitantes...

Si España fuese un pueblo de 100 habitantes

IN EDifusion
3.23K subscribers
Subscribe

78K views · 11 years ago

343 likes · Share · Save

- ¿Cómo sería el Mundo si tuviera solo 100 Personas? Alex Explica 2.6M views · 6 years ago
- La Historia del Agua Embotellada eidemonionegro42 1.6M views · 13 years ago
- La DEMOGRAFÍA EN ESPAÑA (Explicación y conceptos... Aprende con Ana 9.1K views · 3 years ago
- Comunidades Autónomas ser como (Spain Slander) Antequera_25 184K views · 1 year ago
- ¿Y si el MUNDO tuviera 100 HABITANTES? ¿Y si? 466K views · 3 years ago
- España se está quedando sin población en zonas rurales CNN en Español 329K views · 1 year ago
- Si yo fuera rico inspanishplease 227K views · 13 years ago

Curiositats i infografies de l'INE

Instituto Nacional de Estadística

Explica
te ayudamos con la Estadística

- ¿es Explica?
- Primeros pasos
- Estadísticas oficiales
- Estadística y mucho más
- Trabaja con nosotros
- Un poco de historia
- Competición Estadística
- Red de centros de enseñanza al INE



Bienvenido al portal divulgativo del Instituto Nacional de Estadística

En este portal puedes encontrar videos, presentaciones, test, curiosidades que te mostrarán la actividad y la ciencia Estadística de una forma más sencilla. Te proponemos algunos ejemplos

Infografías

Te mostramos obra forma más visual de presentar los datos estadísticos



Competición estadística 2022-23

Inscripciones a partir del 1 de octubre. Ya puedes consultar las bases



El Anuario Estadístico de España.

Testigo de nuestra historia, desde 1855 el Anuario ha reflejado los cambios en la sociedad española



03 - www.ine.es - Accesibilidad - Aviso legal - Avenida de Manzanas, 50-52 - 28050 - Madrid - España

Lectura:

¿Quina és la diferència entre mitjana, mediana i moda?

<https://www.latercera.com/que-pasa/noticia/cual-es-la-diferencia-entre-media-mediana-y-moda/526128/>

¿Cuál es la diferencia entre media, mediana y moda?

<https://www.latercera.com/que-pasa/noticia/cual-es-la-diferencia-entre-media-mediana-y-moda/526128/>

18 feb 2019 05:01 PM. Tiempo de lectura: 2 minutos

Los tres corresponden a las medidas de tendencia central utilizadas en análisis descriptivos. La localización de datos juega un rol fundamental en el análisis.

Estos tres estadísticos corresponden a las medidas de tendencia central más comúnmente utilizadas en análisis descriptivos. La diferencia entre estos radica en como localizar el centro de los datos. Por un lado, la Media corresponde al punto de equilibrio que toma en cuenta la ubicación y el peso de cada dato. Mientras que la Mediana solo toma en cuenta la ubicación, correspondiendo al percentil 50, es decir, divide la muestra en dos grupos con igual cantidad de datos. Por otro lado, la Moda representa el punto de mayor concentración de datos en una muestra, pudiéndose incluso obtener varias modas.

Dependiendo de la naturaleza de la variable en estudio se decide cuál de estos tres utilizar. En el caso de variables cualitativas nominales, el único estadístico de tendencia central interpretable es la Moda, que en cuyo caso corresponde al dato con mayor frecuencia. Si la variable es cualitativa ordinal, generalmente se utiliza la Moda, aunque en algunos casos, dependiendo del contexto, es posible utilizar la Mediana. Con respecto a las variables cuantitativas, se suele usar la Media, como medida de tendencia central, pero el error que comúnmente se comete es no tener en cuenta la forma de la distribución de la muestra.

En distribuciones asimétricas la media se ve fuertemente influenciada por datos extremos, en cambio la Mediana es más robusta a la presencia de este tipo de datos. Por lo tanto, en estas situaciones es preferible utilizar la Mediana. Un ejemplo claro de distribución asimétrica corresponde a la variable salario mensual en Chile, ya que hay una mayor cantidad de personas con sueldos "bajos" en comparación con la cantidad de personas con sueldos "altos". De este modo si se obtiene la media de esta variable se sobreestima el centro, ya que su valor es \$554.493 (según el INE 2017), mientras que la mediana corresponde a \$379.673, evidenciándose una clara diferencia entre estos índices. La media no representa, en esta situación, a la mayor cantidad de la población, por su parte, la mediana nos indica que el 50% de la población chilena recibió en 2017 un sueldo menor o igual \$379.673.

Finalmente, cuando los datos corresponden a una muestra la Media se suele denominar "promedio", la cual tiene la misma interpretación como si se trabajara con toda la población.

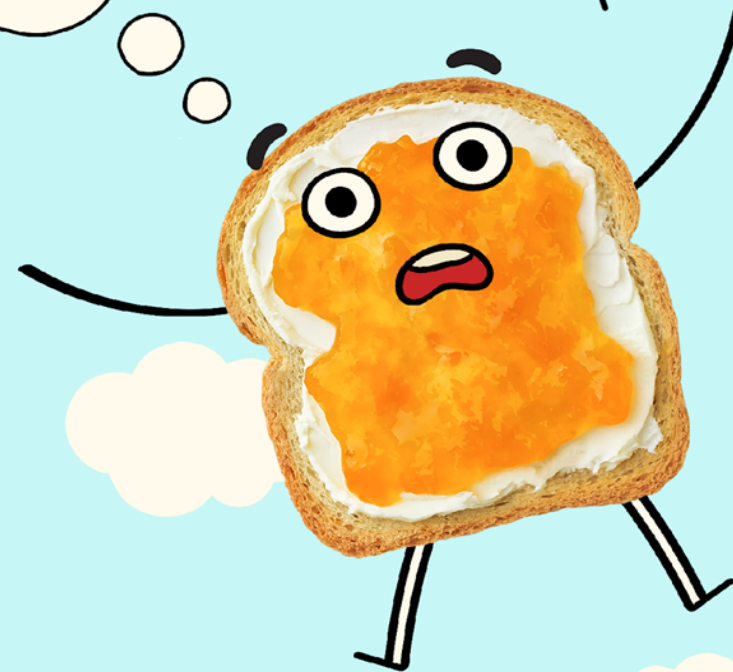
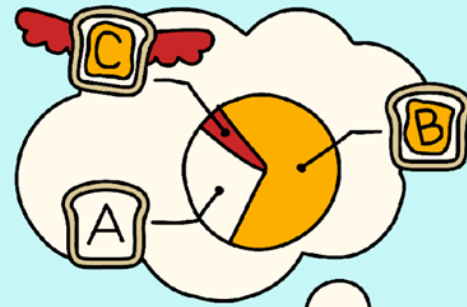
*Jonathan Acosta Salazar, Daira Velandia Muñoz y Francisco Barrera Navia, académicos del Instituto de Estadística de la Pontificia Universidad Católica de Valparaíso.

Quarta Jornada de divulgació del món de l'estadística, 4 i 7 de març 2024

Per a què serveix L'ESTADÍSTICA?

ESTADÍSTICA ADREÇADA A BATXILLERAT

Activitats part II



UNIVERSITAT DE
BARCELONA



LaUB
divulga

La resposta correcta s'indica en vermell

En cap cas, la mitjana i la mediana poden ser iguals

1. **Fals**
2. Cert

La moda en estadística és...

1. El vestit més exitós vist entre els/les assistents a la gala dels Oscar
2. El valor més gran
3. El valor més petit
4. **El valor que presenta una freqüència més gran**



Quins estadístics informen sobre la tendència central?

1. Variància, mitjana, mediana
2. Moda, mitjana, desviació típica
3. **Moda, mitjana, mediana**
4. Variància, desviació típica, rang interquartílic

Quina és l'afirmació falsa?

1. **La desviació estàndard només es pot calcular per a variables categòriques**
2. La desviació estàndard i la desviació típica són el mateix
3. La desviació estàndard està en les mateixes unitats que les dades
4. La desviació estàndard és l'arrel quadrada de la variància

