

Programa de Doctorado Formación en la Sociedad del Conocimiento

Seminario de doctorado

13 y 14 de marzo de 2014



Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Dra. M^a José Rodríguez Conde

Dra. Susana Olmos Migueláñez

Dr. Fernando Martínez Abad

Grupo de Evaluación Educativa y Orientación (GE2O)
GRupo de investigación en InterAcción y eLearning (GRIAL)
Instituto Universitario de Ciencias de la Educación (IUCE)
Universidad de Salamanca

Salamanca, 9 de mayo de 2014



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL



GE₂O



Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Estadística descriptiva

Gráficos de representación de datos

Coeficientes de correlación

Regresión lineal simple y múltiple

Inferencia estadística

Estadística descriptiva

ESTADÍSTICA DESCRIPTIVA

Cálculo de los **estadísticos** de una **muestra**.

Tendencia central

- Media
- Mediana
- Moda

\bar{X}
 Mdn
 Mo

Dispersión

- Varianza
- Desviación típica

S_x^2
 S_x

Forma

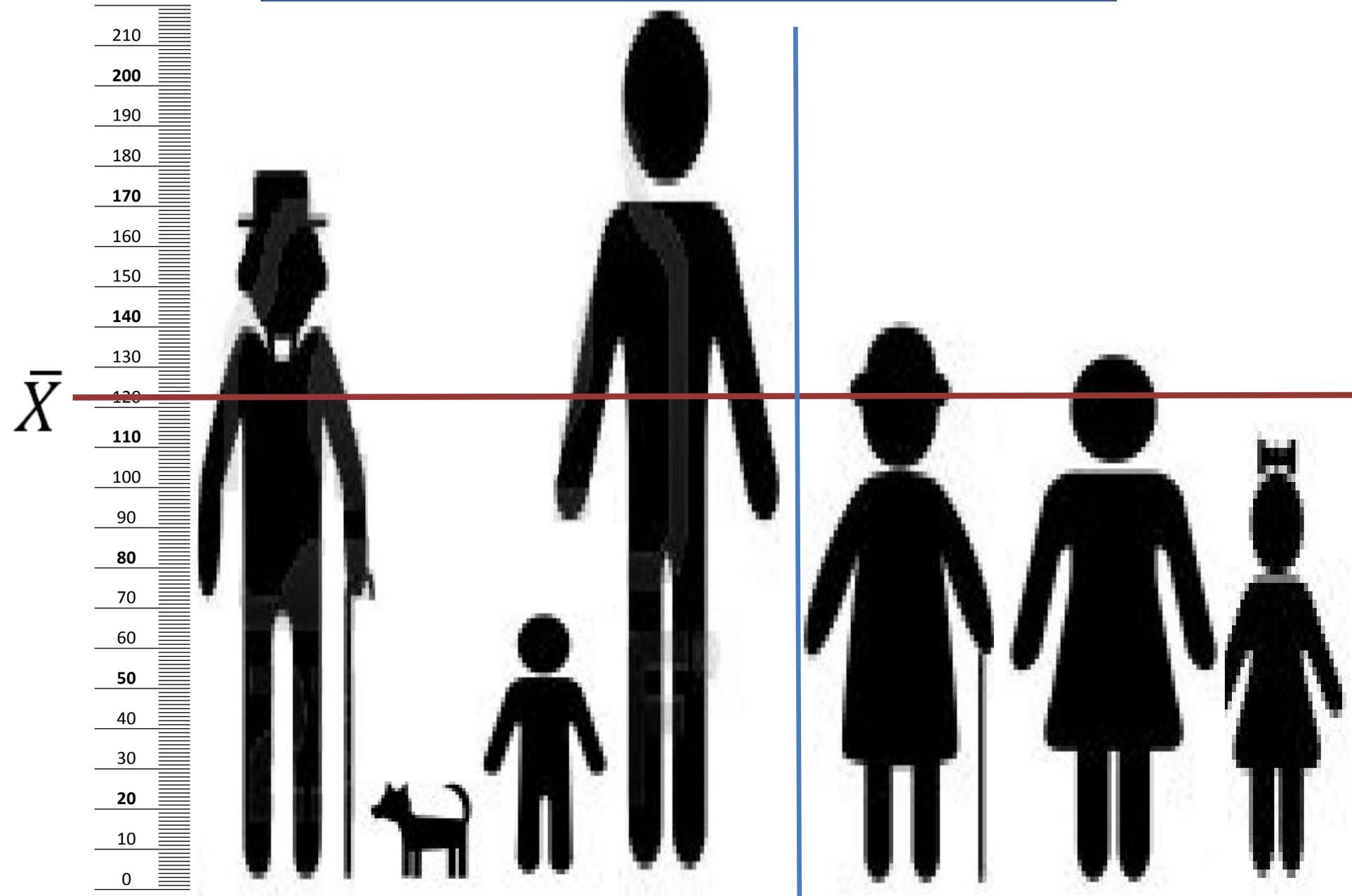
- Asimetría
- Curtosis

As
 $Curt$

Tendencia central

- **Media:** Indica el valor promedio de todas las puntuaciones de una variable.
- **Mediana:** Puntuación que ocupa la posición central en una variable.
- **Moda:** Puntuación más repetida.

Dispersión



Dispersión

Altura

175 cm

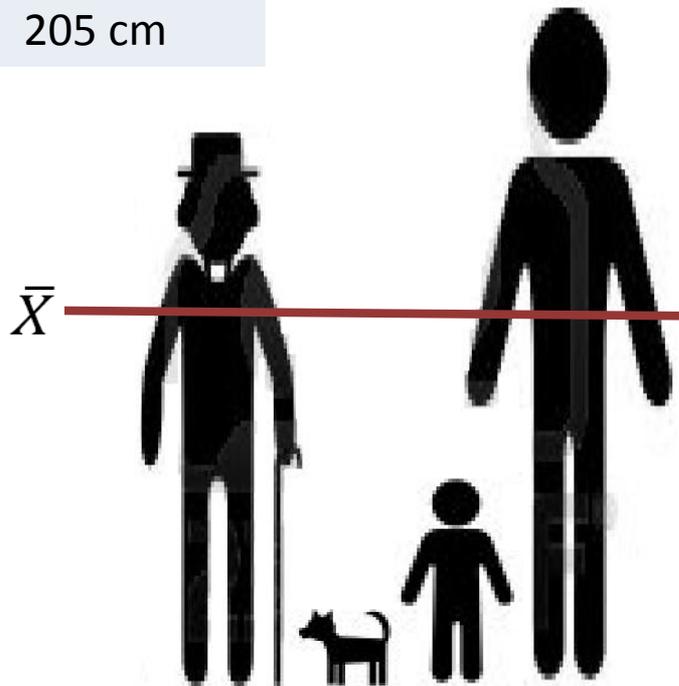
25 cm

60 cm

205 cm

$$\bar{X} = 116.25$$

$$S_x = 87.21$$



Altura

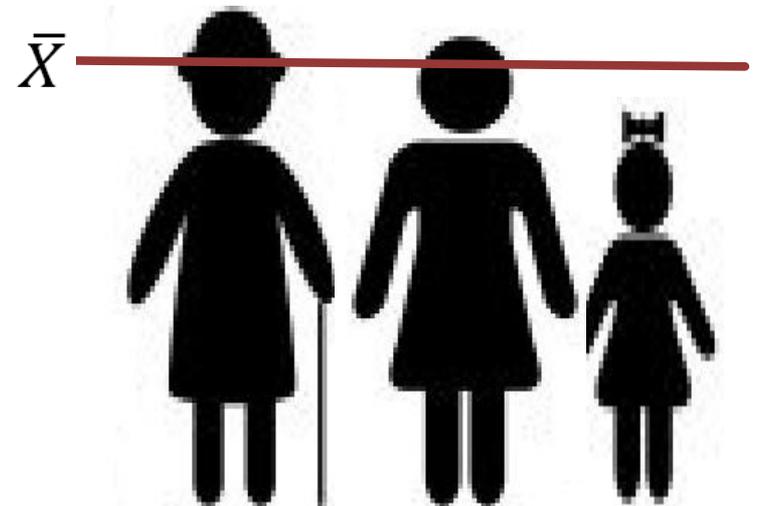
132 cm

125 cm

91.75 cm

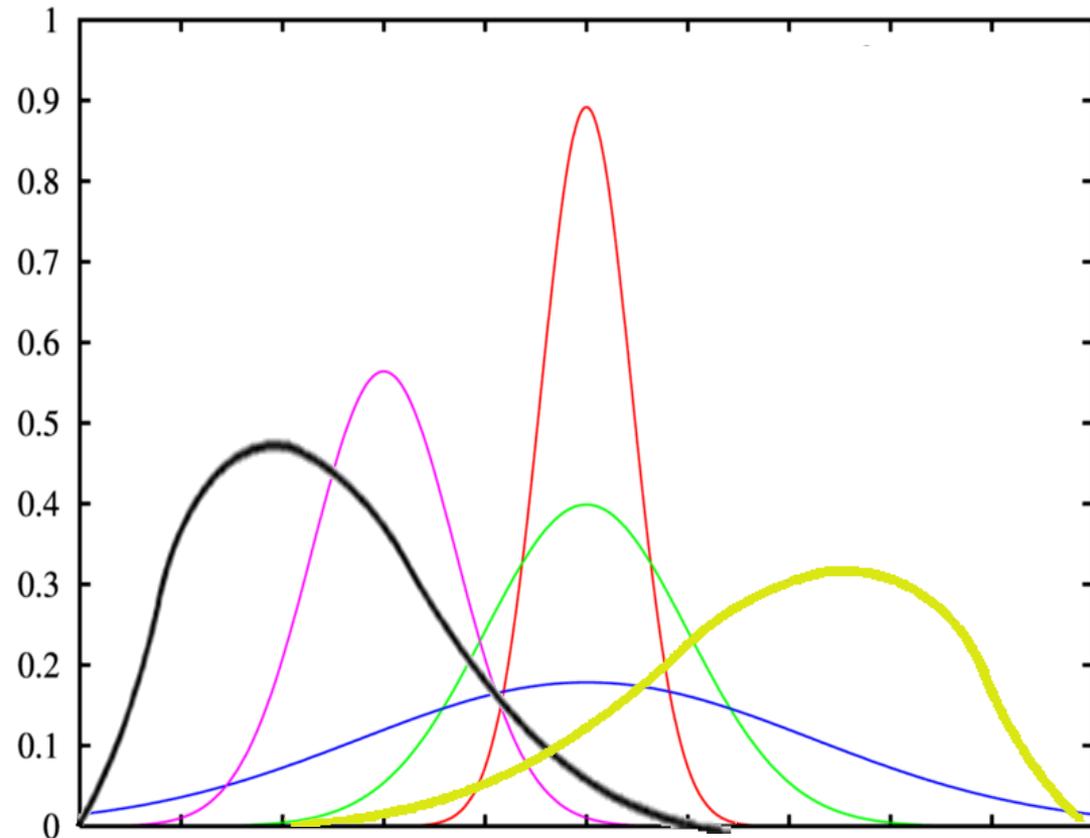
$$\bar{X} = 116.25$$

$$S_x = 21.50$$



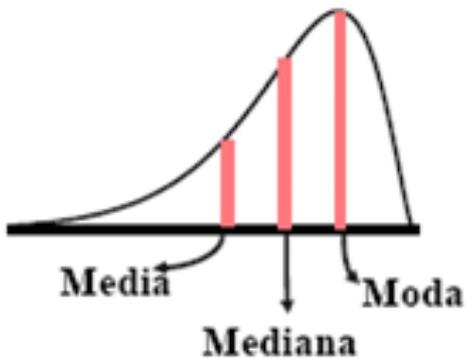
Forma

- **Asimetría:** Mide el grado de simetría que posee la distribución.
- **Curtois:** Indica el grado de agrupación de las puntuaciones en torno a la media.



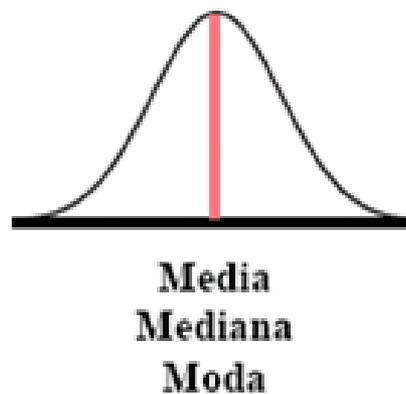
Forma

ASIMETRÍA



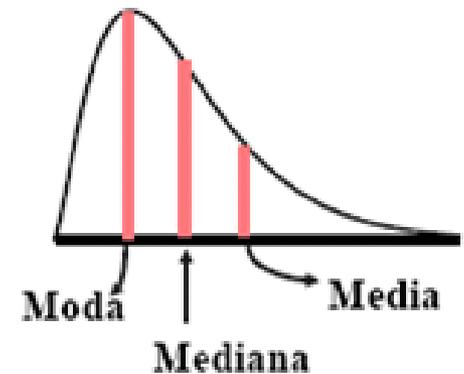
Asimétrica negativa

$As < 0$



Simétrica

$As = 0$

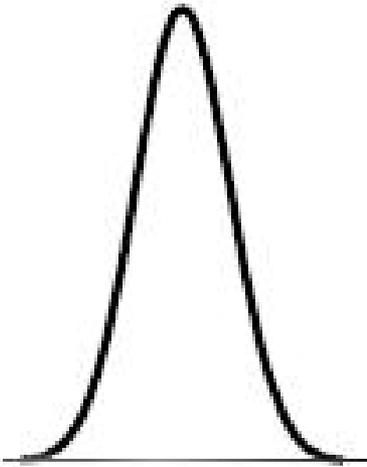


Asimétrica positiva

$As > 0$

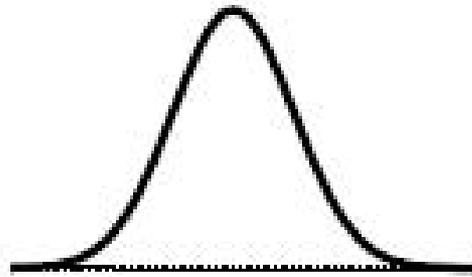
Forma

CURTOSIS



Leptocúrtica

Curt > 0



Mesocúrtica

Curt = 0



Platicúrtica

Curt > 0

Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Estadística descriptiva

Gráficos de representación de datos

Coeficientes de correlación

Regresión lineal simple y múltiple

Inferencia estadística

Diagrama de sectores

- ❖ El **diagrama de sectores** se emplea principalmente en la representación de **variables cualitativas**, aunque también es posible representar variables ordinales y cuantitativas discretas.
- ❖ El **tamaño de los sectores** puede representar **frecuencias absolutas o porcentajes**.

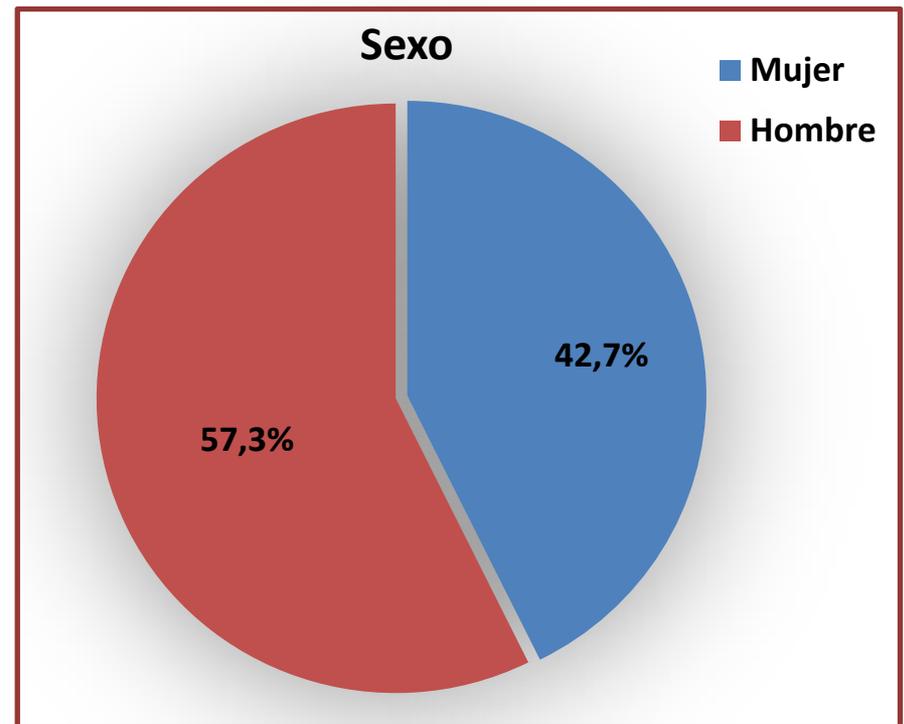
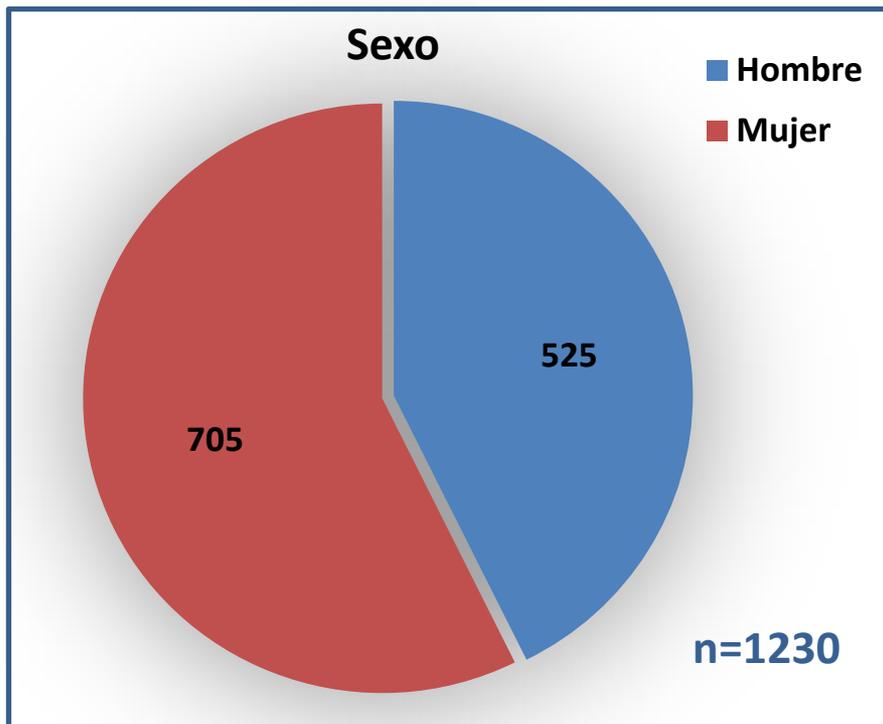
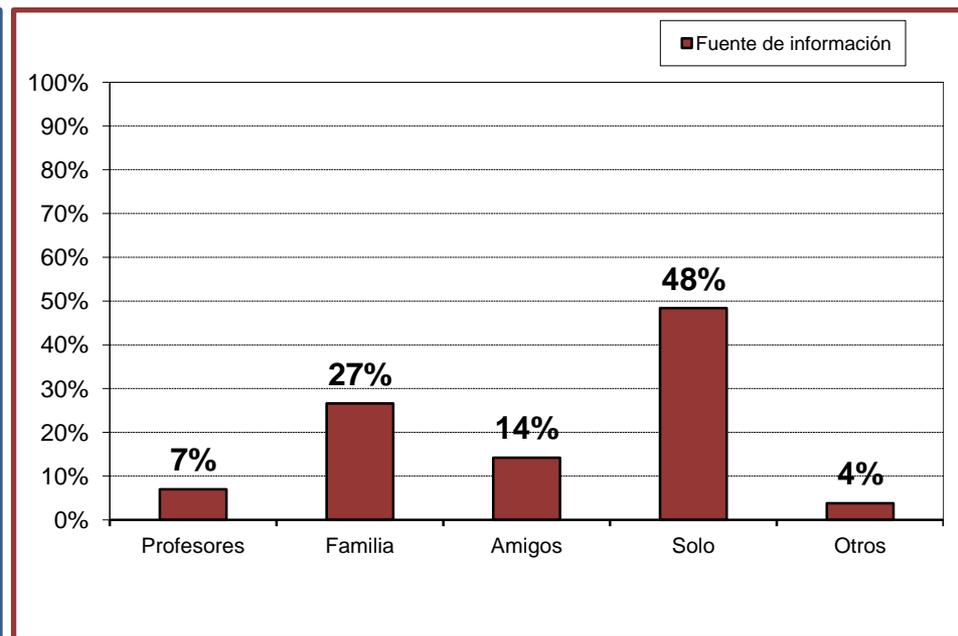
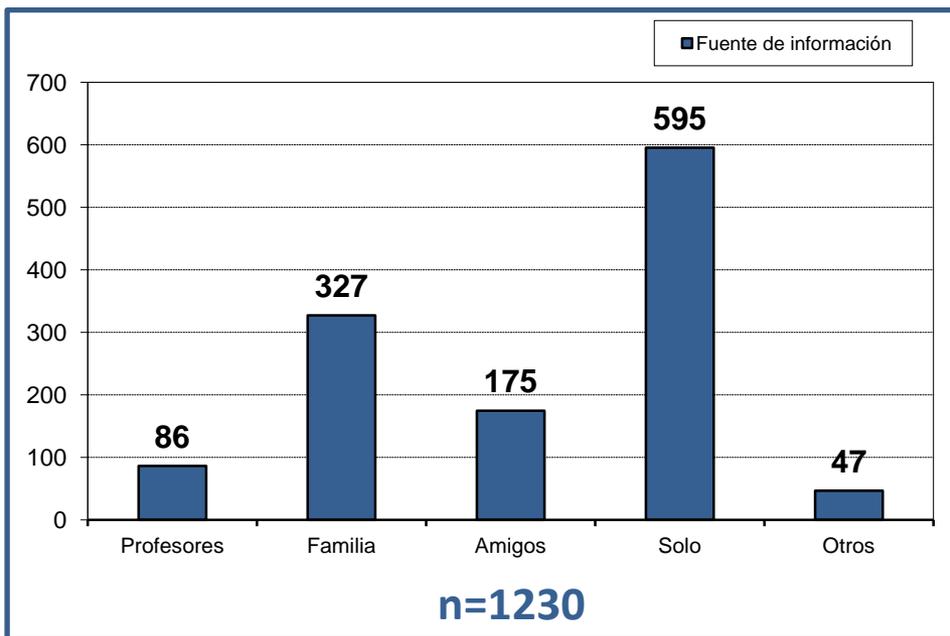


Diagrama de barras

- ❖ El **diagrama de barras** se emplea en la representación de **variables cualitativas, ordinales y cuantitativas discretas**.
- ❖ La **altura de las barras** puede representar **frecuencias absolutas o porcentajes**, y cada barra representa una de las categorías o puntuaciones de la variable.



Histograma y curva de densidad

- ❖ El **histograma** se emplea en la representación de **variables cuantitativas continuas**. A diferencia del diagrama de barras, el histograma presenta las barras unidas.
- ❖ La **Curva de densidad** puede formar un gráfico por separado o acompañar al histograma en un gráfico conjunto.
 - La altura de la curva representa la frecuencia de observaciones para una puntuación x_i determinada.
- ❖ El **eje vertical (y)** representa **frecuencias absolutas o porcentajes**, y el **eje horizontal** la escala de la variable.

Histograma y curva de densidad

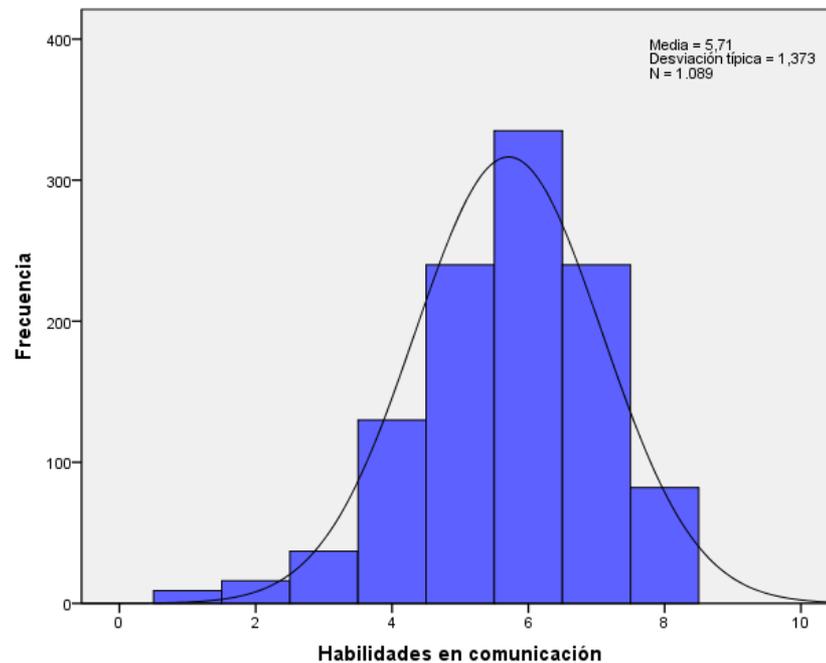
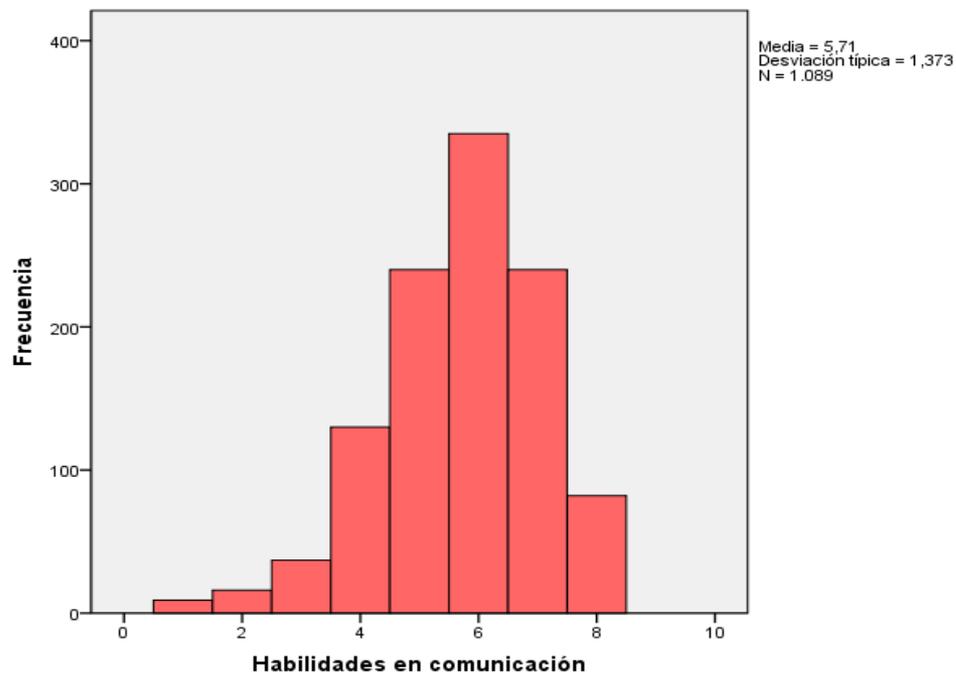
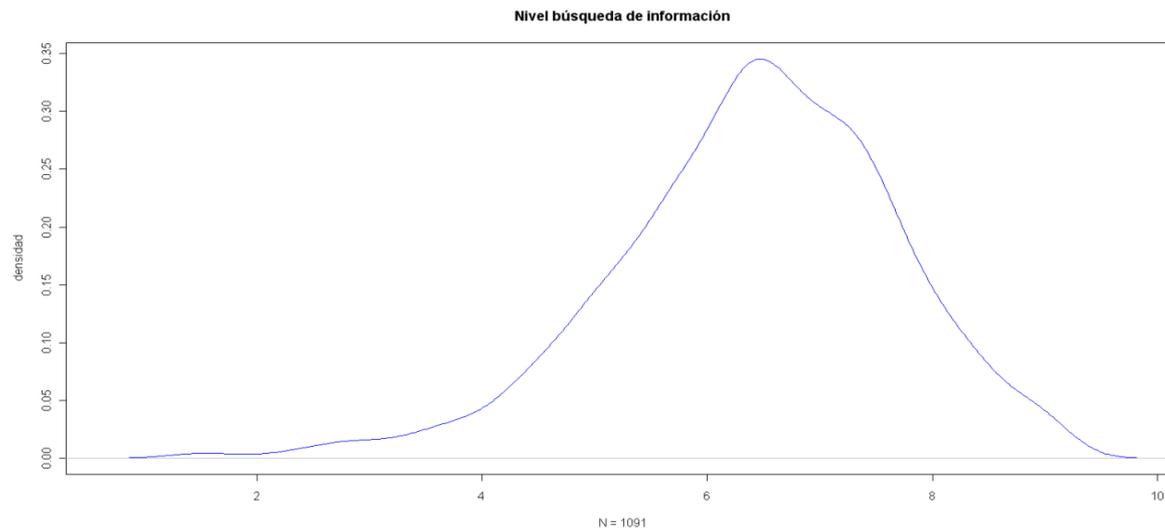
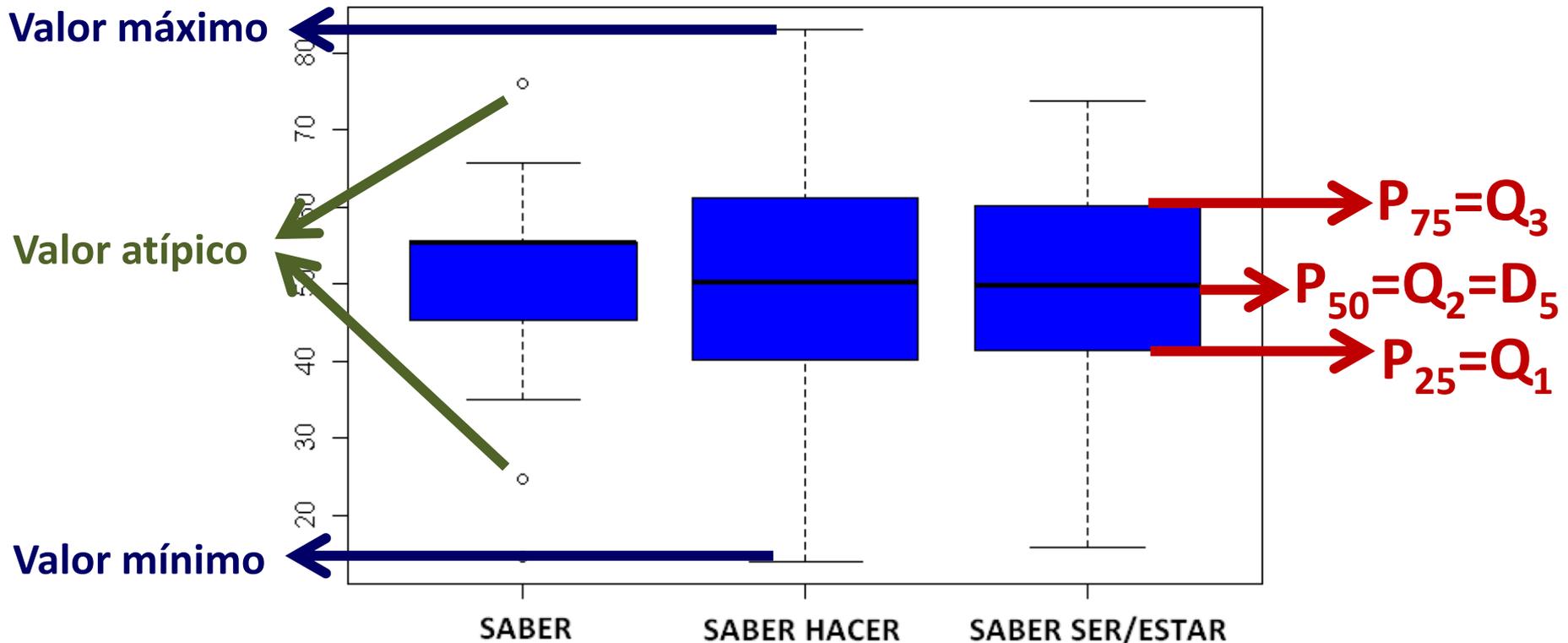


Diagrama de cajas

- ❖ El diagrama de cajas se emplea en la representación de variables continuas.

POSTEST. Búsqueda de información



Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Estadística descriptiva

Gráficos de representación de datos

Coeficientes de correlación

Regresión lineal simple y múltiple

Inferencia estadística

Coeficiente de correlación

El **coeficiente de correlación** mide la relación entre dos variables.

Indica el **tipo de relación** entre dos variables y la **intensidad de la relación**.

TIPO DE RELACIÓN:

DIRECTA	→	Valores positivos
INVERSA	→	Valores negativos
NULA	→	Cero

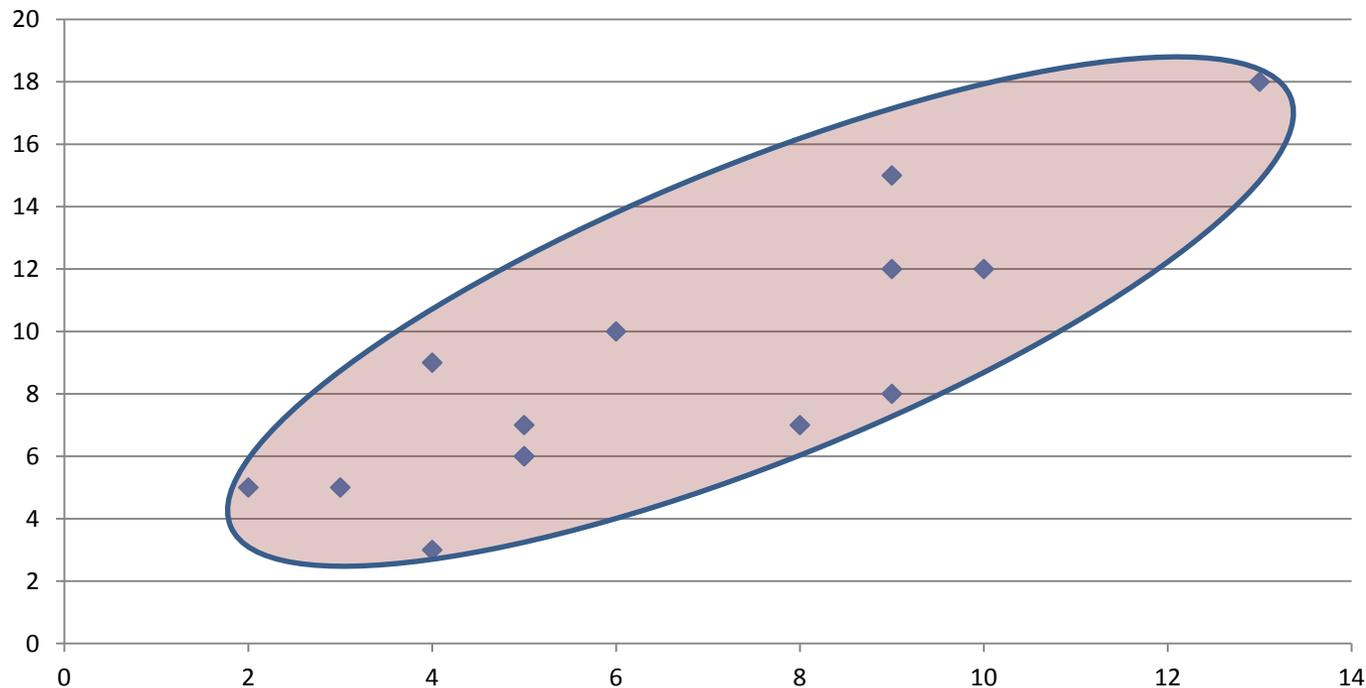
INTENSIDAD DE LA RELACIÓN:

ALTA	→	Valores cercanos a 1 o -1
BAJA	→	Valores cercanos a 0

Coeficiente de correlación

RELACIÓN DIRECTA

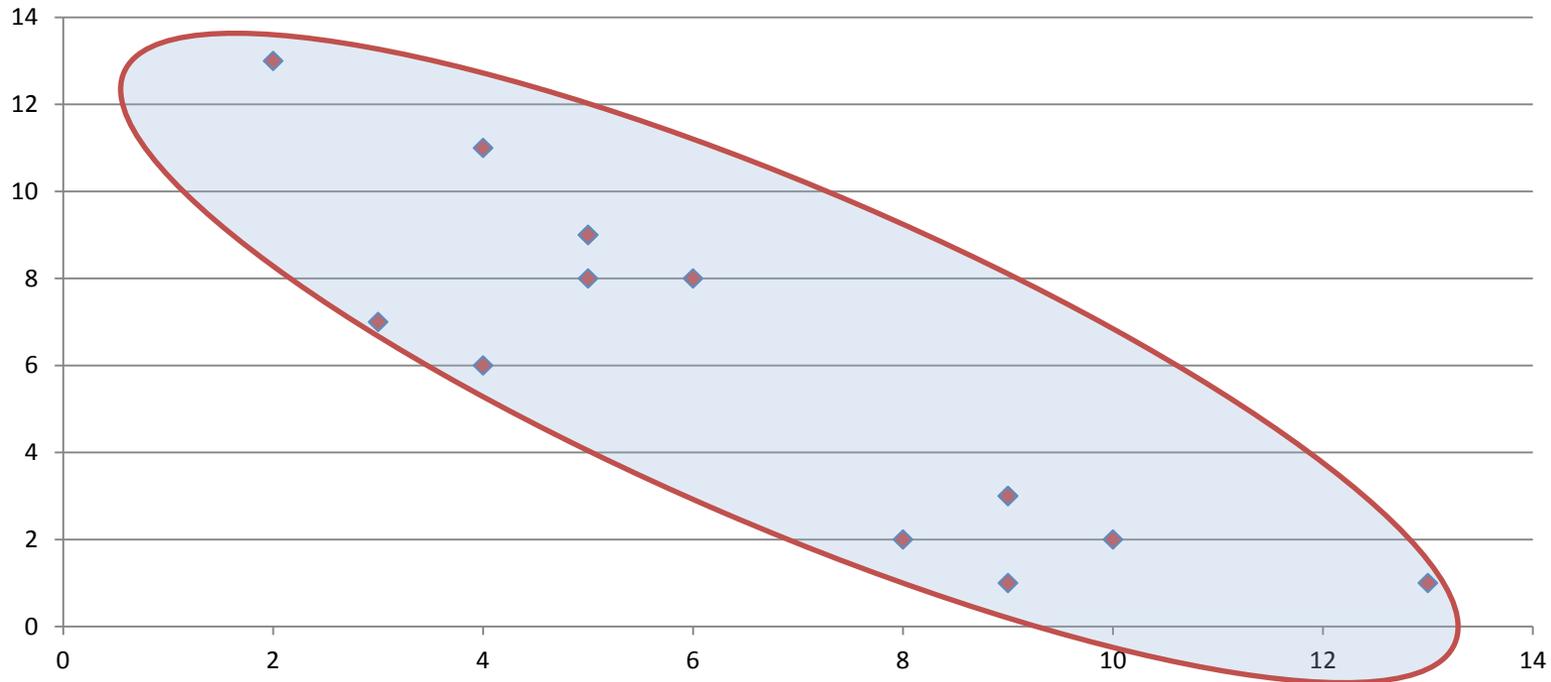
Tendencia de que a valores altos en la primera variable (X), corresponden valores altos de la segunda (Y), y viceversa.



Coeficiente de correlación

RELACIÓN INVERSA

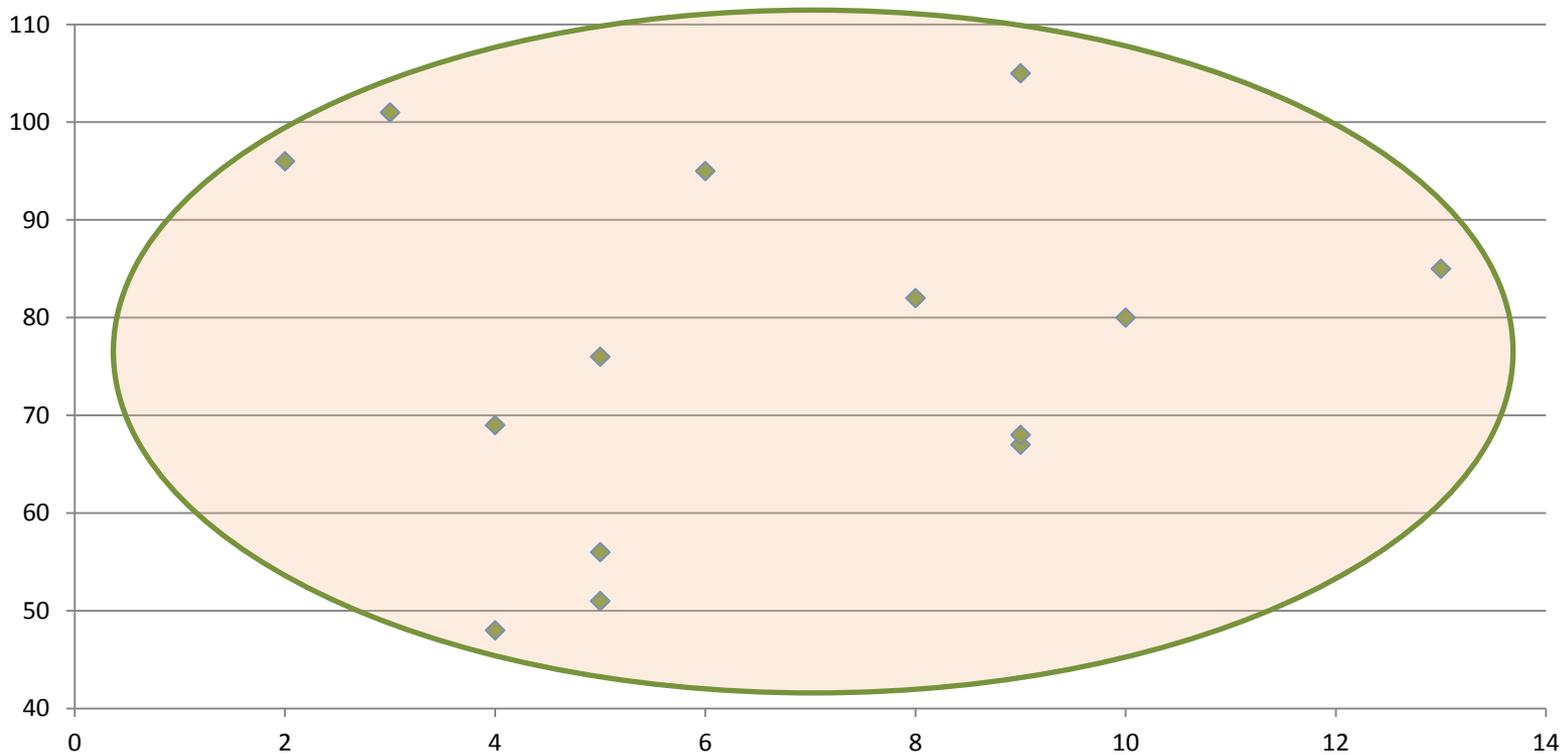
Tendencia de que a valores altos en la primera variable (X), corresponden valores bajos de la segunda (Y), y viceversa.



Coeficiente de correlación

RELACIÓN NULA

No se observa ninguna tendencia lineal.



Coeficiente de correlación en función del tipo de variables

	CUANTITATIVA	ORDINAL	CUALITATIVA POLITÓMICA	CUALITATIVA DICOTÓMICA
CUANTITATIVA	PEARSON (r_{xy})	SPEARMAN (r_s)		BISERIAL PUNTUAL (r_{bp})
ORDINAL		SPEARMAN (r_s)		
CUALITATIVA POLITÓMICA			COEFICIENTE DE CONTINGENCIA	COEFICIENTE DE CONTINGENCIA
CUALITATIVA DICOTÓMICA				COEFICIENTE FI

Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Estadística descriptiva

Gráficos de representación de datos

Coeficientes de correlación

Regresión lineal simple y múltiple

Inferencia estadística

Ecuación de regresión

$$y' = a + bx \longrightarrow y_i = a + bx + \varepsilon$$

y' = Valores pronosticados en y para un valor concreto en x

a = Ordenada en el origen
(punto de corte de la recta con el eje y)

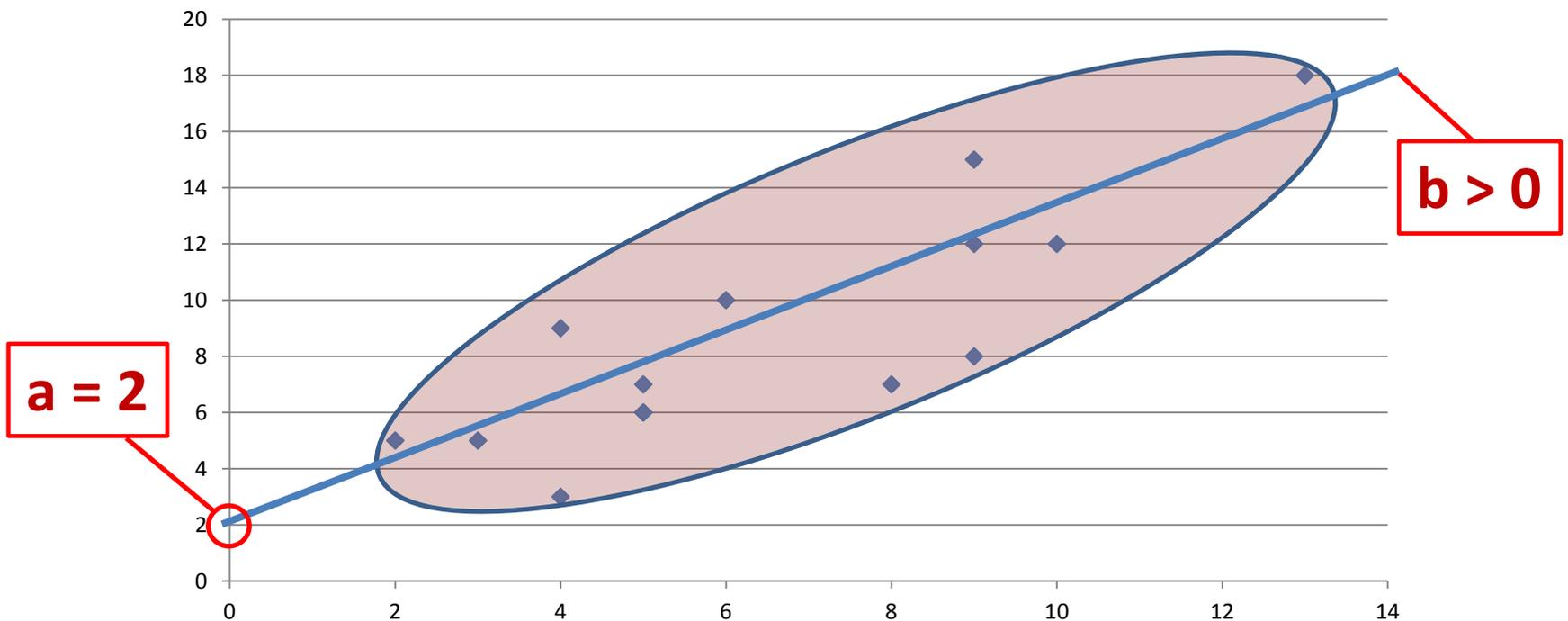
b = Pendiente de la recta de regresión

ε = Error de la predicción
(diferencia entre el valor pronosticado y el real para un sujeto)

$$b = \frac{S_{xy}}{S_x^2} \qquad a = \bar{Y} - b\bar{X}$$

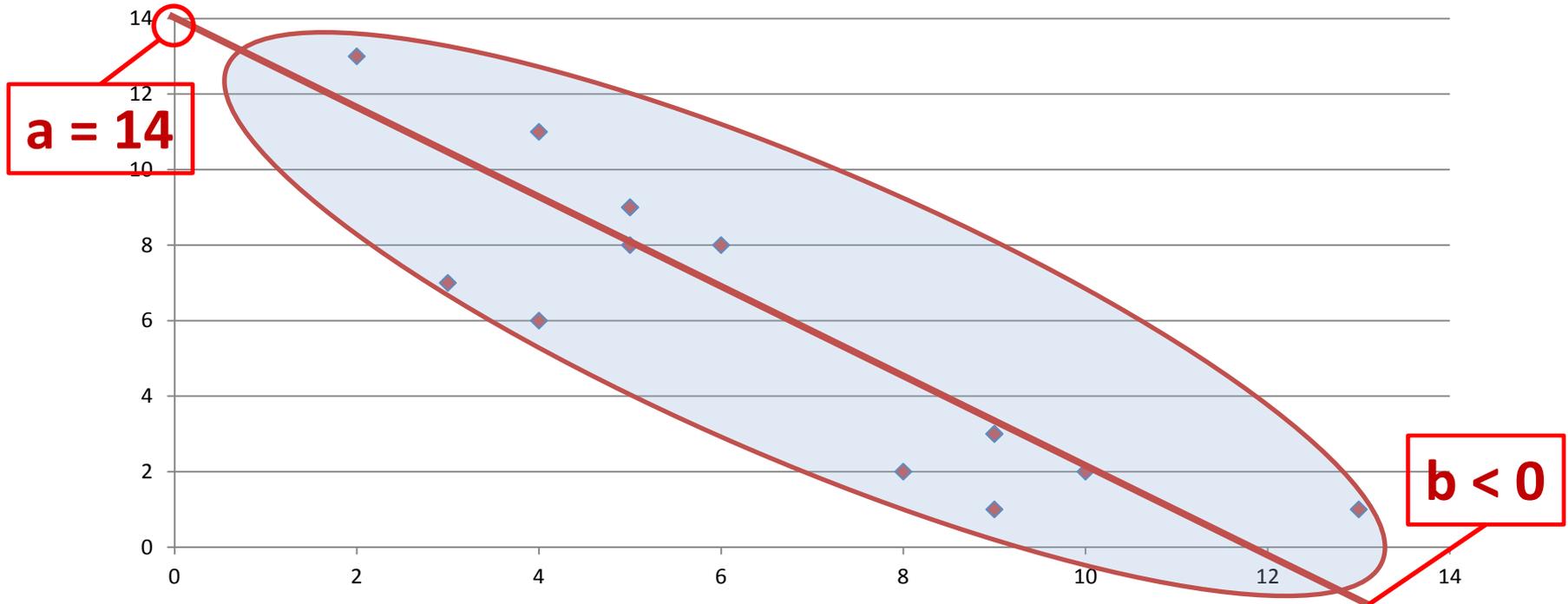
Recta de regresión

RECTA CRECIENTE: En este caso, a medida que aumenta el valor de x , aumenta el de y . Así, el valor de b (pendiente) es **positivo**, al igual que el del coeficiente de correlación.



Recta de regresión

RECTA DECRECIENTE: A medida que aumenta el valor de x , el valor de y disminuye. Así, el valor de b (pendiente) es **negativo**, al igual que el del coeficiente de correlación.



Bondad de ajuste de la ecuación de regresión

Para medir la bondad de ajuste de la ecuación de regresión, esto es, el error cometido al predecir los valores de y a partir de un valor x_i , se emplea el **coeficiente de determinación (R^2)**.

$$R^2 = (r_{xy})^2$$

El coeficiente de determinación tiene **valores entre 0 y 1**. Indica la **proporción de variabilidad explicada** por el modelo.

Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Estadística descriptiva

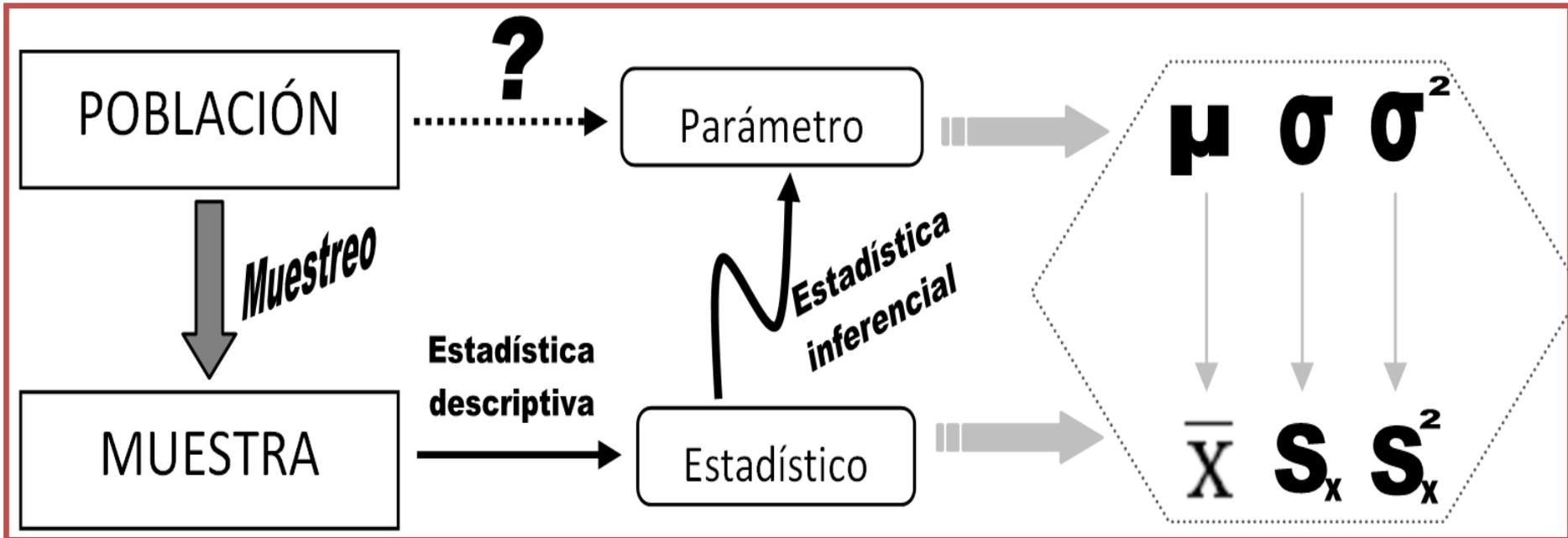
Gráficos de representación de datos

Coeficientes de correlación

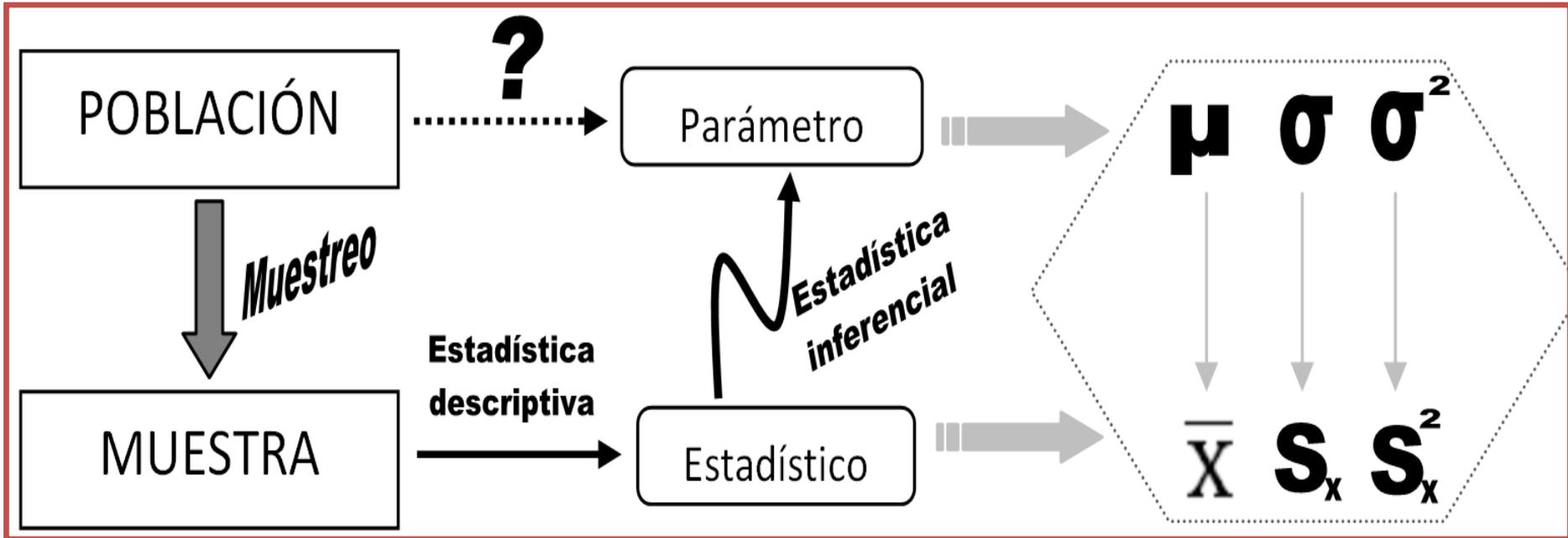
Regresión lineal simple y múltiple

Inferencia estadística

El proceso de inferencia Estadística



Intervalos de confianza



Gracias a la inferencia estadística, partiendo de unas **condiciones específicas** de las **distribuciones** de las variables, podré afirmar que, partiendo de una **muestra** determinada y con un **error** concreto, el valor de **un parámetro poblacional** se encuentra dentro de un **intervalo** dado.

Intervalos de confianza

Suponiendo que la **muestra** es **representativa** y que la variable es **normal** y **homocedástica**:

$$\bar{X} = 10$$

$$S_x = 1$$

Error= 5%

$$\mu = (8.04; 11.96)$$

Error= 1%

$$\mu = (7.42; 12.58)$$

$$\bar{X} = 15$$

$$S_x = 3$$

Error= 5%

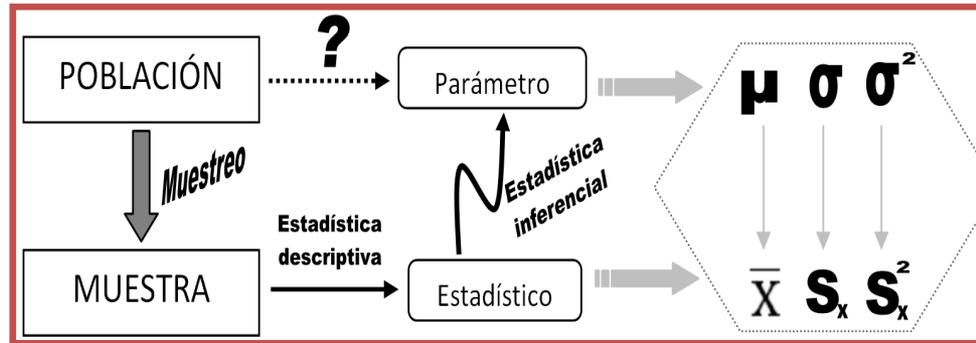
$$\mu = (12.06; 17.94)$$

Error= 1%

$$\mu = (11.13; 18.87)$$

La hipótesis estadística

La inferencia estadística se basa en la formulación de **hipótesis acerca de los parámetros**, y consiste en contrastar si esa hipótesis es cierta o falsa.



$$\bar{X} = 10$$

$$S_x = 1$$

Error= 5%

$$\mu = (8.04; 11.96)$$

¿Se puede afirmar, **con un error del 5%**, que la **media poblacional** de esta variable es 7.5?

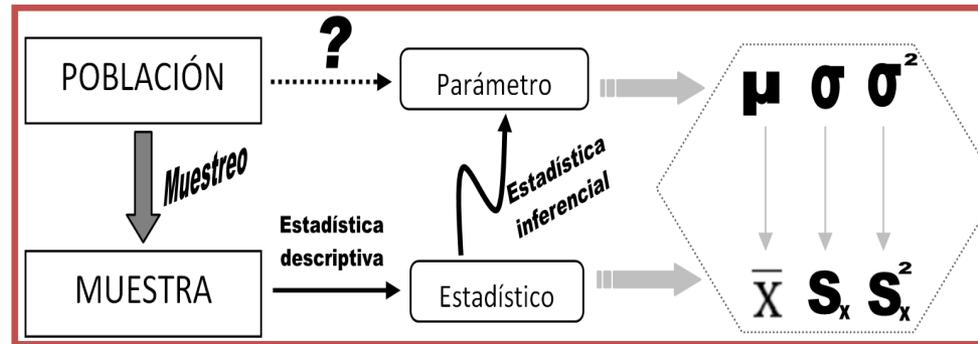
$$H_0: \mu = 7.5$$

$$H_1: \mu \neq 7.5$$

**NO SE PUEDE AFIRMAR!!!
RECHAZO H_0**

La hipótesis estadística

La inferencia estadística se basa en la formulación de **hipótesis acerca de los parámetros**, y consiste en contrastar si esa hipótesis es cierta o falsa.



$$\bar{X} = 10$$

$$S_x = 1$$

Error= 1%

¿Y **con un error del 1%**, se puede afirmar que la media poblacional es 7.5?

$$H_0: \mu = 7.5$$

$$H_1: \mu \neq 7.5$$

**SÍ SE PUEDE AFIRMAR!!!
NO RECHAZO H_0**

$$\mu = (7.42; 12.58)$$

La hipótesis estadística

Se observa mejor este procedimiento de manera gráfica:

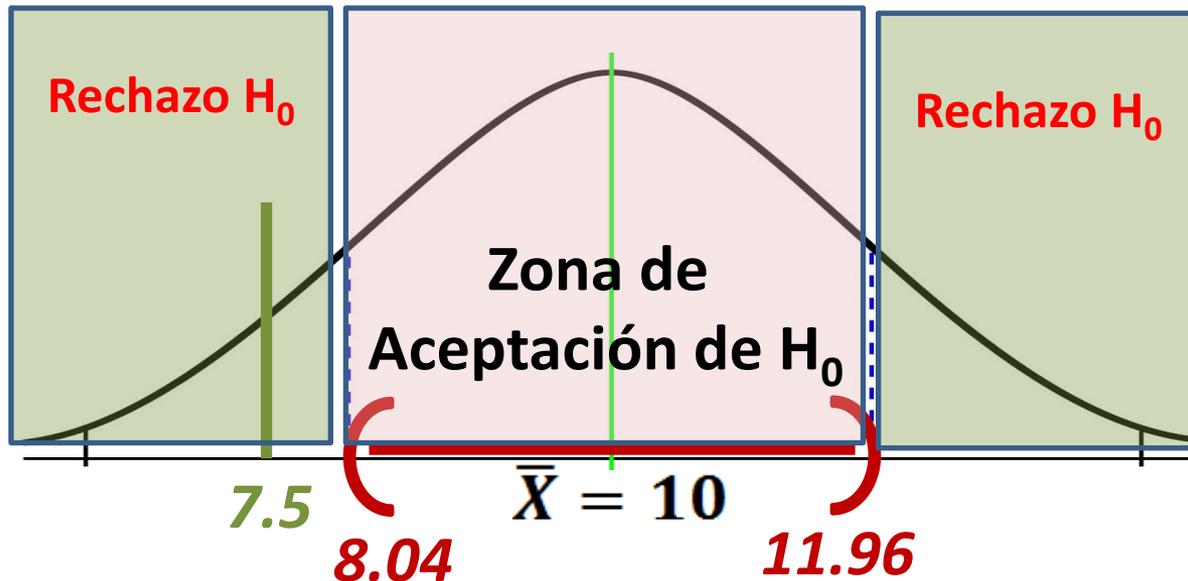
$$\bar{X} = 10$$

$$S_x = 1$$

Error= 5%

$$H_0: \mu = 7.5$$

$$H_1: \mu \neq 7.5$$



**ZONA DE RECHAZO!!!
RECHAZO H_0**

La hipótesis estadística

Se observa mejor este procedimiento de manera gráfica:

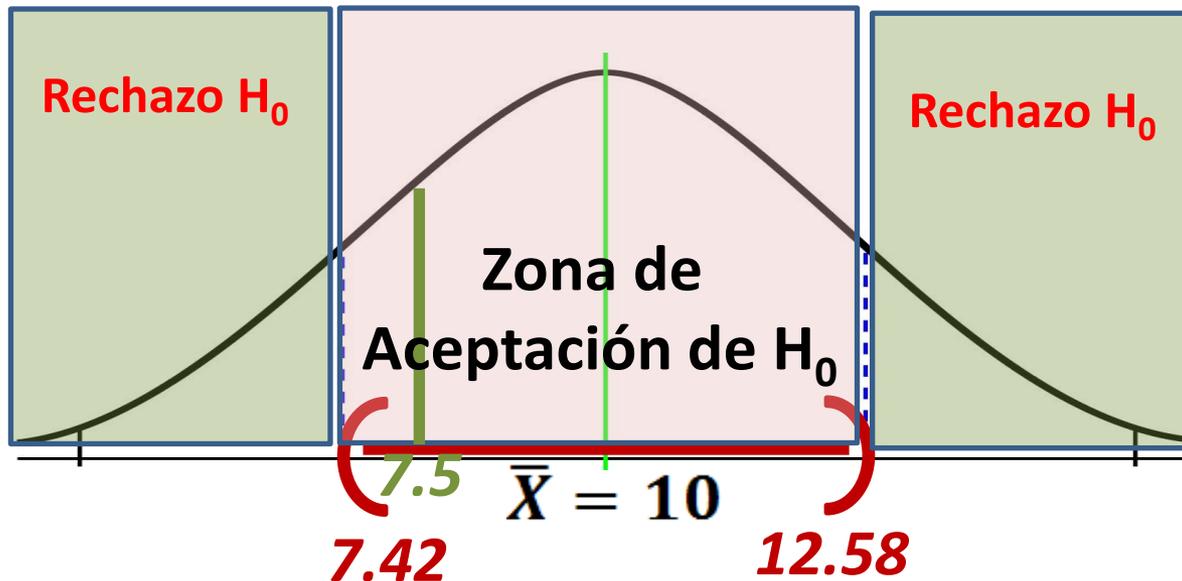
$$\bar{X} = 10$$

$$S_x = 1$$

Error= 1%

$$H_0: \mu = 7.5$$

$$H_1: \mu \neq 7.5$$



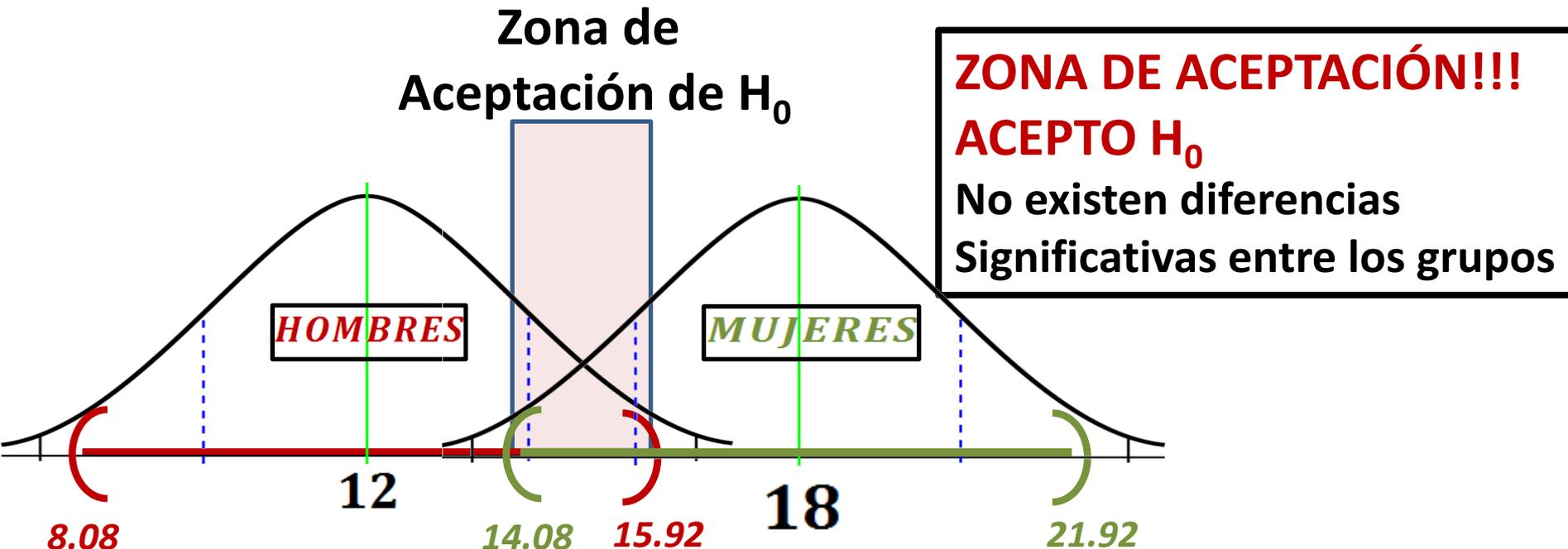
**ZONA DE ACEPTACIÓN!!!
ACEPTO H_0**

La hipótesis estadística

¿Y si lo que quiero es comparar grupos? ¿Existen diferencias significativas entre hombres y mujeres en la variable X?:

$$\begin{array}{l} \bar{X}_{\text{hombres}} = 12 \\ \bar{X}_{\text{mujeres}} = 18 \end{array} \quad s_x = 2 \quad \xrightarrow{\text{Error} = 5\%}$$

$$\begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{array}$$



Condiciones para la inferencia

- **MUESTREO**

- La **muestra** seleccionada debe ser **representativa** de la población de la que procede.

- **NORMALIDAD**

- Las variables sobre las que se va a realizar la inferencia deben ajustarse a la **distribución normal**.

- **HOMOCEDASTICIDAD**

- En el caso de que se contrasten varias variables, estas deben poseer **varianzas semejantes**.

Normalidad de las variables

- **Distribución de la variable sobre la que se aplica inferencia similar a la distribución de la curva normal**

- CURTOSIS MESOCÚRTICA ($Curt=0$)
- SIMETRÍA ($As=0$) → Media=Moda=Mediana

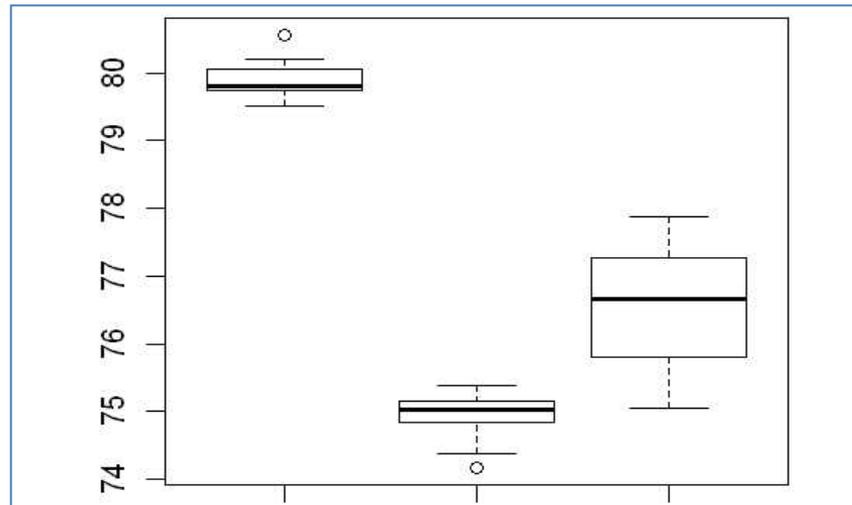
» Prueba de Kolmogorov-Smirnov

- H_0 : La variable se ajusta a la curva normal → $p \geq \alpha$
- H_1 : La variable **no** se ajusta a la curva normal → $p < \alpha$

Homocedasticidad de las variables

- **Varianzas similares de las variables sobre las que se contrastan parámetros**

– Estudio de la dispersión de los datos (Diagrama de cajas)



» Prueba de Levene

- H_0 : Las variables tienen varianzas iguales
- H_1 : Las variables **no** tienen varianzas iguales

→ $p \geq \alpha$

→ $p < \alpha$

Condiciones para la inferencia

- En el caso del cumplimiento de las condiciones emplearemos las **técnicas paramétricas** (basadas en la distribución normal de las variables).
- En el caso de que no se cumplan las condiciones emplearemos **técnicas no paramétricas** (basadas en otras distribuciones).

Estadística descriptiva

ESTADÍSTICA INFERENCIAL

Cálculo de los **parámetros** de una **población**.

PARAMÉTRICA

- Prueba T para la media
- Prueba T para grupos independientes
- Prueba T para grupos relacionados
- ANOVA

NO PARAMÉTRICA

- U de Mann-Whitney para 2 grupos independientes
- W de Wilcoxon para 2 grupos relacionados
- H de Kruskal-Wallis para k grupos independientes
- Prueba de Friedman para k grupos relacionados

Programa de Doctorado Formación en la Sociedad del Conocimiento

Seminario de doctorado

13 y 14 de marzo de 2014



Análisis de datos en CCSS:

introducción al análisis descriptivo e inferencial

Dra. M^a José Rodríguez Conde

Dra. Susana Olmos Migueláñez

Dr. Fernando Martínez Abad

Grupo de Evaluación Educativa y Orientación (GE2O)
GRupo de investigación en InterAcción y eLearning (GRIAL)
Instituto Universitario de Ciencias de la Educación (IUCE)
Universidad de Salamanca

Salamanca, 9 de mayo de 2014



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL



GE₂O

