

# Métodos cuantitativos aplicados al diseño de Políticas Públicas

Año 2021



MAESTRÍA EN ECONOMÍA PÚBLICA Y POLÍTICAS  
ECONÓMICAS SOCIALES Y REGIONALES

---

MGTER. MARIANA V. GONZALEZ

(mariana.gonzalez@unc.edu.ar)

# Unidad 1: El método estadístico. Organización y presentación de datos. Análisis descriptivo



## Objetivos específicos:

- Introducir al alumno en los conceptos básicos de la disciplina
- Capacitar al alumno para realizar análisis descriptivo e interpretar resultados de procesamiento con softwares estadísticos
- Promover la capacidad de transmitir resultados de un análisis estadístico en su futura actuación profesional

## Bibliografía:

- ESTADISTICA I: ciclo básico a distancia. Coord. Margarita Díaz "et al." Córdoba, Argentina, Asociación Cooperadora de la Facultad de Ciencias Económicas, 2015. Capítulos 1 y 2.
- BERENSON, Mark L. y LEVINE, David M. Estadística Básica en Administración. Conceptos y Aplicaciones. 6ª ed. México, D. F., Prentice Hall hispanoamericana, 1997, c1996. 943 p.

# Introducción



## Estadística

*Es la aplicación del método científico en el análisis de datos para facilitar la toma de decisiones ante situaciones de incertidumbre (Mood y Graybill, 1978).*

Método  
Estadístico

1. Planteo del problema. Hipótesis
2. Diseño y Recolección de datos
3. Organización y presentación de los datos
4. Análisis de datos
5. Resultados y conclusiones que intentan resolver el problema planteado.

Abarca la recolección, presentación y caracterización de la información para ayudar tanto en el análisis de datos como en el proceso de la toma de decisiones.

# Importancia de las estadísticas públicas y sociales



Los datos estadísticos constituyen una herramienta estratégica para el diseño de políticas públicas y sociales adecuadas, la toma de decisiones acertadas y el ejercicio de una gestión transparente, eficiente y responsable.

*"El Estado requiere información estadística útil para **diseñar** las políticas apropiadas, así como de la evidencia que permita **monitorear** y **evaluar** permanentemente los efectos de tales políticas" (Alicia Bárcena, Secretaria Ejecutiva de la CEPAL, Naciones Unidas, 2011)*

Enfrentar los retos del desarrollo exige cada vez más de información confiable y oportuna sobre las tendencias en la sociedad y la economía. Es por ello que los gobiernos han reconocido la importancia de producir regularmente dicha información y de utilizarla para el diseño de las políticas públicas. La calidad de las estadísticas nacionales está intrínsecamente relacionada con la capacidad de lograr una gestión pública basada en evidencia (*evidence based policy making*) y orientada a resultados (*results focused public management*)

# Importancia de las estadísticas públicas y sociales



No sólo la producción sino el uso de la información estadística son fundamentales para lograr políticas públicas efectivas y eficientes.

A través del uso adecuado de las estadísticas:

- los gobiernos pueden diseñar y medir el riesgo de sus políticas de corto y largo plazo,
- las empresas son capaces de dimensionar el tamaño de sus mercados y programar sus planes de inversión,
- la clase política puede rendir cuentas a sus electores,
- los ciudadanos pueden conocer la coyuntura socioeconómica,
- la sociedad puede orientar mejor sus decisiones colectivas,
- las instituciones académicas y centros de investigación pueden estudiar y hacer seguimiento a los diversos fenómenos económicos y sociales

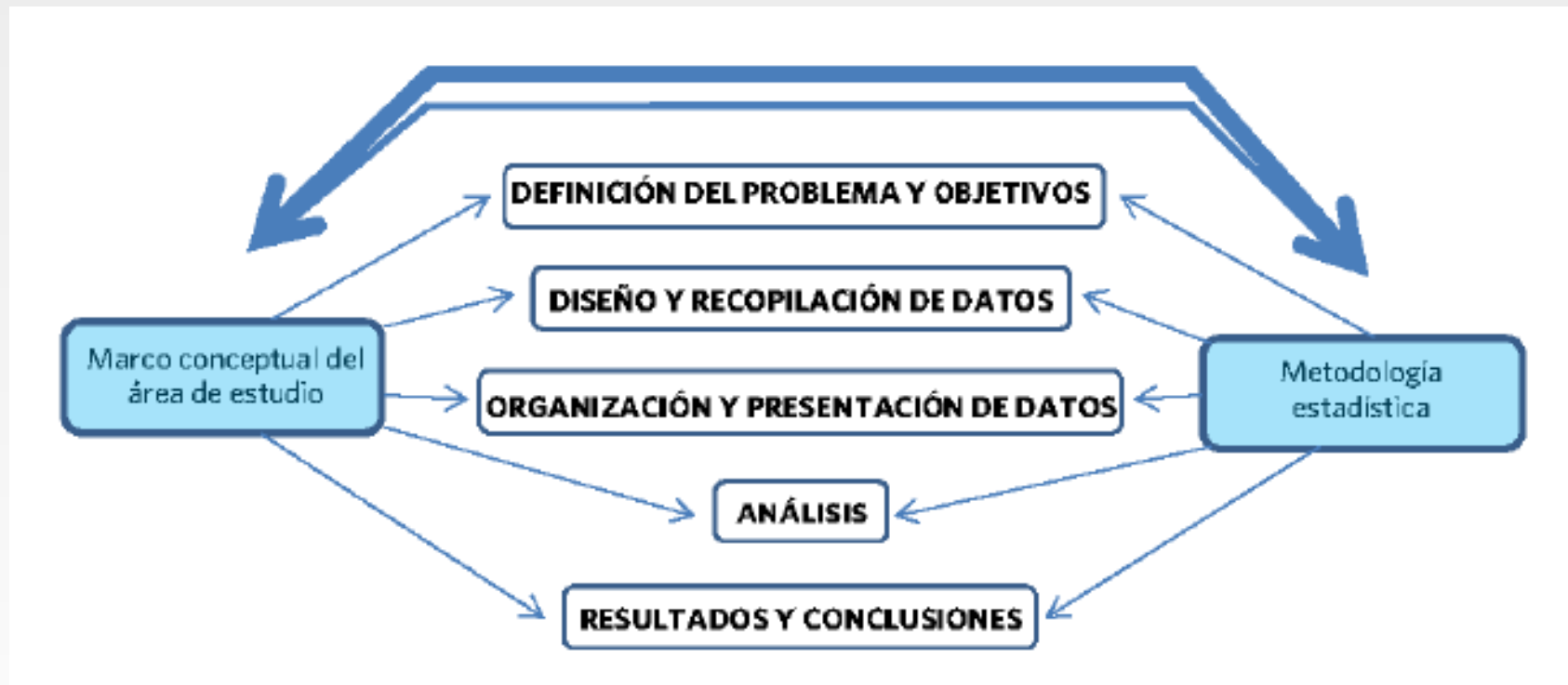
# Importancia de las estadísticas públicas y sociales



En la gestión de la política pública la **evaluación** es tan importante como el objetivo que se pretende alcanzar ya que permite analizar la planificación como también retroalimentar el proceso de ejecución y también avanzar más allá a través de la medición del impacto que ella genera en la comunidad.

1. **Diseño de la evaluación:** población, unidades de análisis, método de muestreo, variables a observar, fuentes de información, instrumentos a utilizar para el relevamiento.
2. **Relevamiento de la información:** se obtiene la información tal como lo señala el diseño.
3. **Procesamiento de la información:** softwares estadísticos, resúmenes descriptivos
4. **Análisis de la información:** técnicas estadísticas uni y multivariadas, herramientas inferenciales.

# Plan de investigación estadística



# Estadística descriptiva e inferencial



## **Estadística descriptiva**

Comprende aquellos métodos que incluyen la recolección, organización, presentación y análisis de un conjunto de datos con el fin de describir apropiadamente las diversas características del mismo.

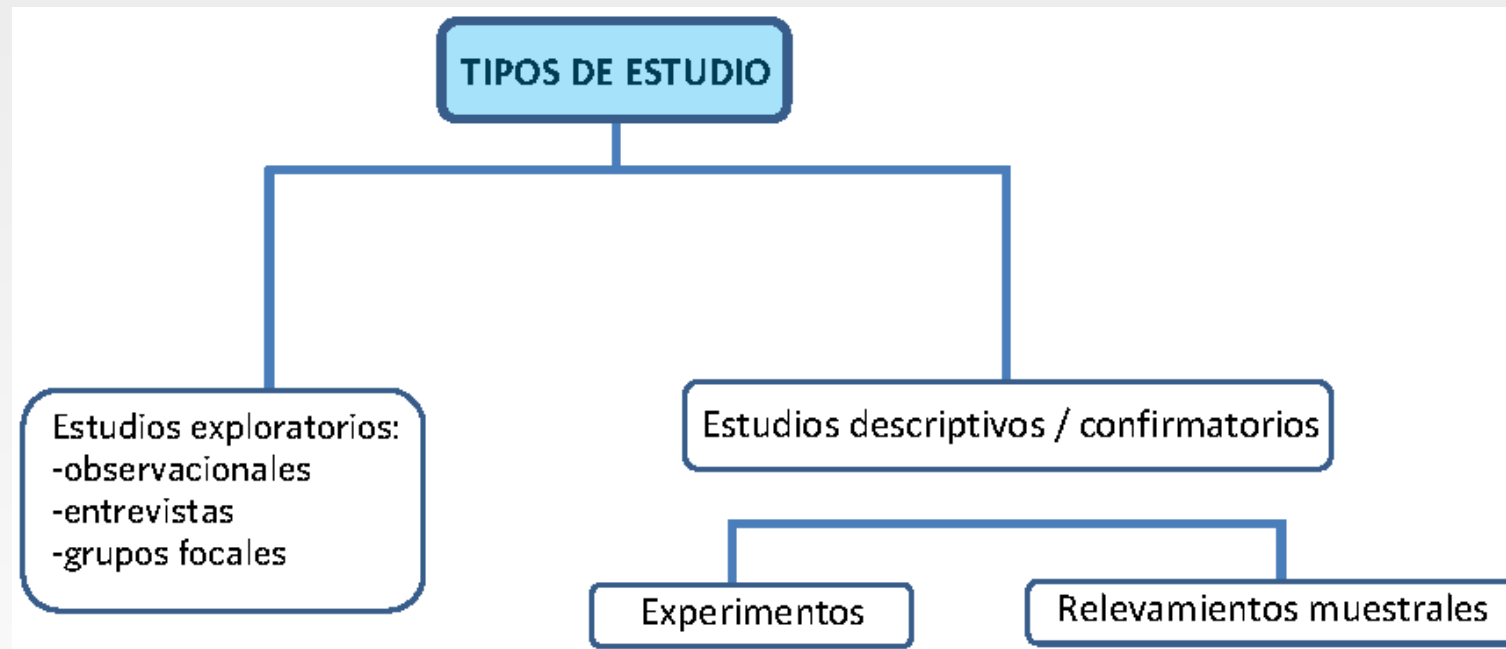


## **Estadística inferencial**

Comprende aquellos métodos que permiten hacer generalizaciones con respecto a la población con base en información proporcionada por una muestra aleatoria, con un grado de incertidumbre cuantificable.



# Tipos de estudio



◦ Fuente: Estadística I. FCE. UNC

# Tipos de estudio



- **Estudios exploratorios:** se basan en casos seleccionados discrecionalmente y suelen ser el paso inicial de estudios posteriores.
- **Estudios descriptivos / confirmatorios:** en estos estudios el objetivo es hacer inferencia respecto a la población.

**Experimentos:** estudios organizados y controlados por el investigador.

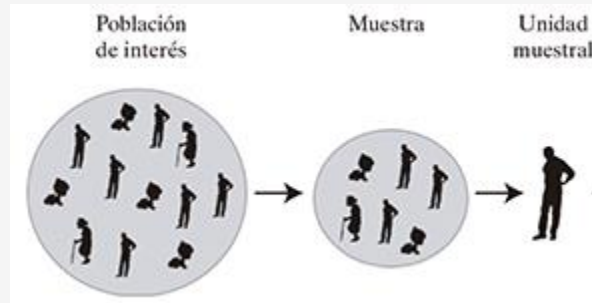
**Relevamientos muestrales:** el investigador estudia una población bien definida extrayendo de ella un subconjunto de unidades de observación denominado muestra.



Los datos pueden recopilarse

- Estudios observacionales
- Diseño de experimentos
- Encuestas
- Entrevistas

# Conceptos básicos



**Unidad estadística:** cada uno de los elementos de la población o de la muestra sobre la cual se busca la información.

**Población o Universo:** conjunto de todos los elementos o individuos bajo estudio.

**Muestra:** parte de la población que se selecciona para el análisis, debe ser representativa.

**Muestra aleatoria:** es aquella en la que todos los elementos de la población tienen una probabilidad conocida de ser seleccionados.

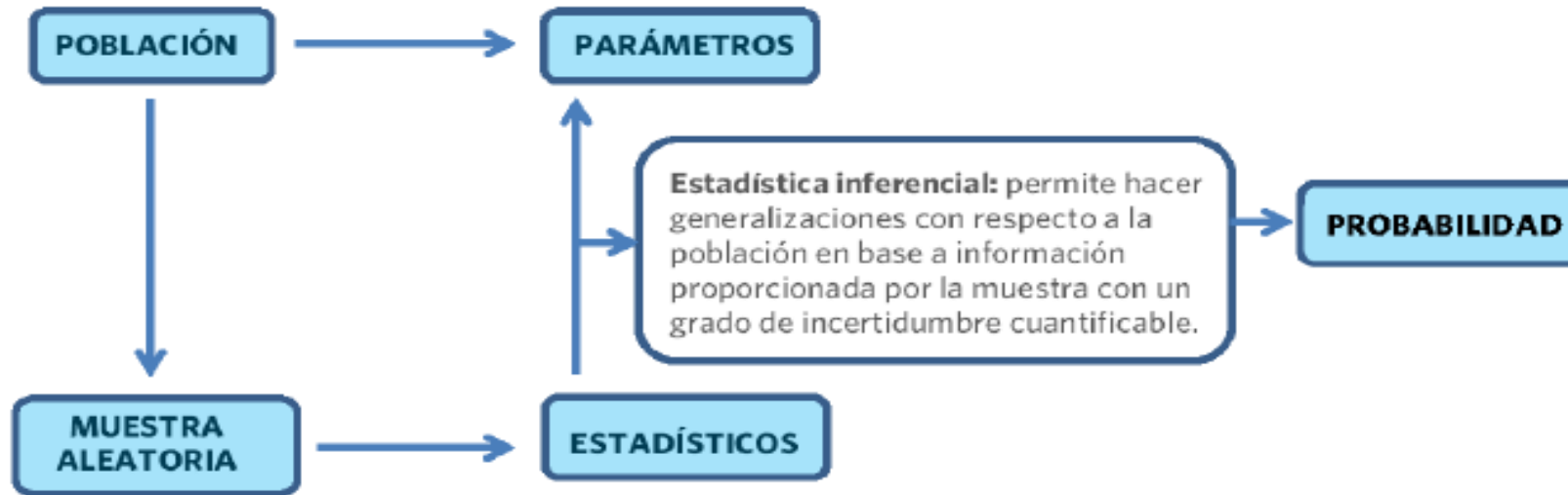
**Parámetro:** es una medida que se calcula para describir una característica de la población.

**Estadístico:** es una medida que se calcula para describir una característica de una muestra de la población.

# Conceptos básicos



**Estadística descriptiva:** resume y describe al conjunto de datos, calculando medidas (parámetros o estadísticos según corresponda).



Fuente: Estadística I. FCE. UNC

# Fuentes de datos

---



## Fuentes primarias

- Observación
- Experimentación
- Encuestas

## Fuentes secundarias

- Publicaciones
- Investigaciones anteriores

<https://www.indec.gob.ar>

<http://estadisticas.cepal.org/cepalstat/Portada.html>

<https://estadistica.cba.gov.ar>

# Ejemplo 1



A los fines de identificar con mayor precisión los factores tanto estructurales como funcionales que afectan el normal desenvolvimiento del sector industrial, la Unión Industrial de Córdoba, durante el año 2006, realizó un estudio sobre la problemática integral que afecta al sector.

Con tales propósitos se llevó a cabo una encuesta -dirigida a gerentes y propietarios de establecimientos industriales- en Capital e Interior de la Provincia de Córdoba, que incluyó a 150 establecimientos industriales y que tuvo como objetivo principal reunir la información necesaria para elaborar un diagnóstico integral de todos aquellos problemas que obstan o retardan la operación del sector en los mercados nacionales y/o externos. Los establecimientos fueron seleccionados en forma aleatoria.

Con los datos de la muestra se realizó un análisis descriptivo de cada una de las características relevadas y de algunas relaciones entre variables. Además, otro objetivo del estudio consistió en estimar el porcentaje de empresas que exportan, para lo cual se utilizaron métodos inferenciales como la estimación de parámetros por intervalos.

En una segunda etapa la Unión Industrial solicita un estudio sobre un conjunto específico de empresas, seleccionadas por ciertas características particulares para analizar las formas de gerenciamiento de las mismas a través de entrevistas a personas claves dentro de la empresa.

## Analicemos:

- Tipo de estudio
- Técnica de recolección de datos
- Población
- Muestra
- U. estadística
- Estadística descriptiva e inferencial
- Porcentaje de empresas que exportan en la muestra y en la población.

# Concepto de variable



Se desean conocer las características de las empresas industriales de la Provincia de Córdoba . Se seleccionaron al azar 150 empresas.

¿ Qué características de las empresas vamos a analizar?

- Rama de actividad
- Cantidad de empleados
- Nivel de ventas brutas (en \$)



Variable de interés: características  
que deben relevarse para responder a  
los objetivos del estudio



Dato: la realización  
de una respuesta particular

# Bases de datos. Softwares



Nº	Rama de actividad	Cantidad de empleados	Ventas brutas (en millones de \$)
1	Alimentos y Bebidas	154	4,5
2	Papel e Impresiones	27	2,5
3	Productos Metálicos	130	10,7
...	...	...	...

**InfoStat**  
SOFTWARE ESTADÍSTICO

QUIENES SOMOS | CURSOS | COMPRAR | CONTÁCTENOS | VERSIÓN ESTUDIANTIL | DESCARGAS

**Infostat**

InfoStat es un software para análisis estadístico de aplicación general desarrollado bajo la plataforma Windows.

Cubre tanto las necesidades elementales para la obtención de estadísticas descriptivas y gráficos para el análisis exploratorio, como métodos avanzados de modelación estadística y análisis multivariado. Una de sus fortalezas es la sencillez de su interfaz combinada con capacidades profesionales para el análisis estadístico y el manejo de datos. Debido al origen universitario, el programa tiene muchas facilidades para la enseñanza de la estadística que no son fáciles de encontrar en otros programas similares. La versión en español es muy valorada por los usuarios, especialmente por los estudiantes. Una propiedad casi única entre el software estadístico es la habilidad de InfoStat de conectarse con R, una plataforma de desarrollo de algoritmos estadísticos de dominio público de gran crecimiento. InfoStat se conecta con R de dos maneras: mediante un intérprete integrado que permite ejecutar script de R sin salir del ambiente de trabajo de InfoStat y mediante el desarrollo de aplicaciones utilizando el motor de cálculo de R pero con la interfaz amigable que los usuarios esperan. Ese es el caso de la inclusión de modelos lineales mixtos y generalizados mixtos en InfoStat. Estos han sido siempre modelos difíciles de especificar por su complejidad, pero la interfaz lograda en InfoStat nos ha permitido incorporar sus contenidos en cursos de postgrado y capacitaciones a técnicos de empresas que de otra manera hubieran sido imposibles de abordar.

**NOTICIAS**

Versión Estudiantil (no puedo llenar formulario)  
...

**Versión 2017**  
Nueva forma de comunicarnos con R

**Actualización 2014**  
...

[Noticias](#)

**CALENDARIO**

<https://www.infostat.com.ar/>



# Clasificación de variables



Las variables se clasifican como muestra el gráfico siguiente:



**Variables cualitativas o categóricas:** están definidas por las clases o categorías que las componen. No pueden medirse. Ejemplo: sector de actividad

**Variables cuantitativas o numéricas:** pueden expresarse como cantidades numéricas.

**Discretas:** surgen de un proceso de conteo. Ejemplo: cantidad de empleados

**Continuas:** surgen de un proceso de medición. Ejemplo: ventas brutas (en \$)

# Análisis unidimensional

Escala  
nominal

Organización y presentación de los datos

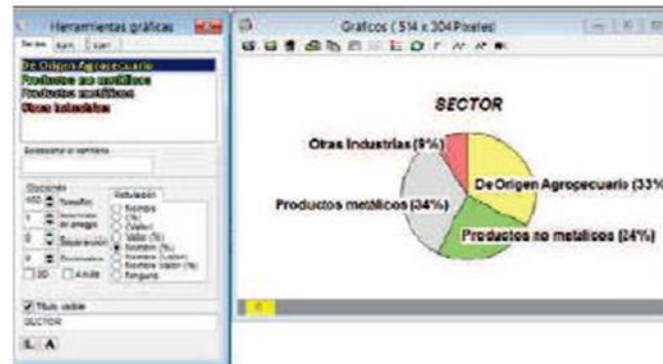
Distribuciones Unidimensionales  
(Variable categórica)

SECTOR DE LA ECONOMÍA	FRECUENCIA ABSOLUTA	PORCENTAJE
De Origen Agropecuario	49	33,1
Productos no metálicos	36	24,3
Productos metálicos	50	33,8
Otras Industrias	13	8,8
Total	148	100,0

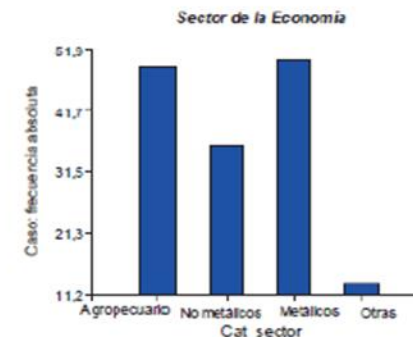
**Frecuencias absolutas:** cantidad de veces que se repite el valor de la variable.

**Frecuencias relativas:** proporción de veces que se repite el valor de la variable.

Sector de la economía



Sector de la economía



# Análisis unidimensional

Escala  
ordinal

SUFICIENCIA DE LA OFERTA DE MANO DE OBRA CALIFICADA	FRECUENCIA ABSOLUTA	PORCENTAJE	FRECUENCIA ABSOLUTA ACUM.	PORCENTAJE ACUM.
Suficiente	50	33,8	50	33,8
Medianamente suficiente	42	28,4	92	62,2
Escasamente suficiente	39	26,4	131	88,5
Insuficiente	17	11,5	148	100,0
Total	148	100,0		

$$131/148 \times 100$$

Oferta de Mano de Obra Calificada

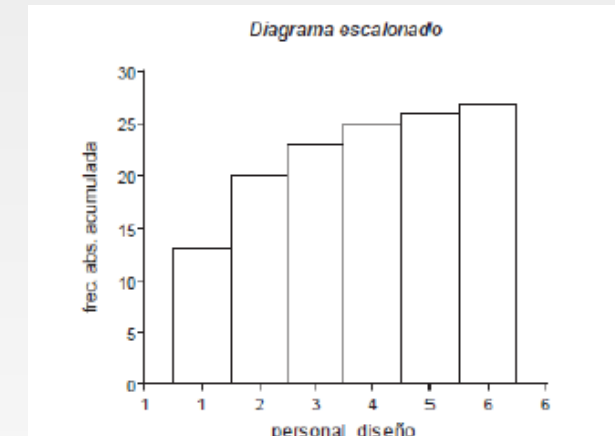
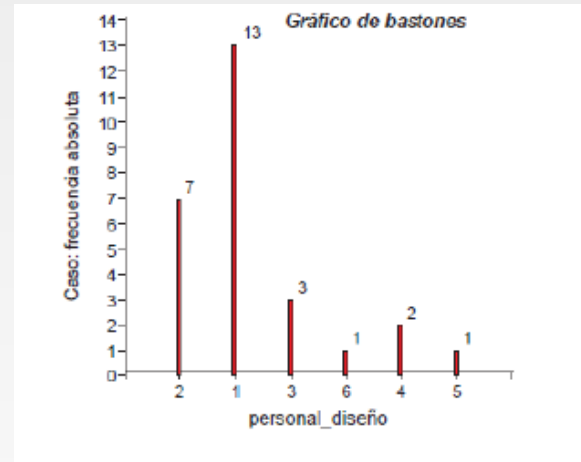


# Análisis unidimensional

(Variable numérica discreta)

Tabla 1.3. Cantidad de empleados área Diseño e Ingeniería industrial

EMPLEADOS QUE REALIZAN LA ACTIVIDAD $x_i$	FRECUENCIA ( $n_i$ )	FRECUENCIA RELATIVA ( $h_i$ )	FRECUENCIA ABSOLUTA ACUM. ( $N_i$ )	FRECUENCIA RELATIVA ACUM. ( $H_i$ )
1	13	0.46	17	0.46
2	7	0.25	20	0.71
3	3	0.11	23	0.82
4	2	0.07	25	0.89
5	2	0.07	27	0.96
6	1	0.04	28	1.00
Total	28	1.00		



Escala de razón

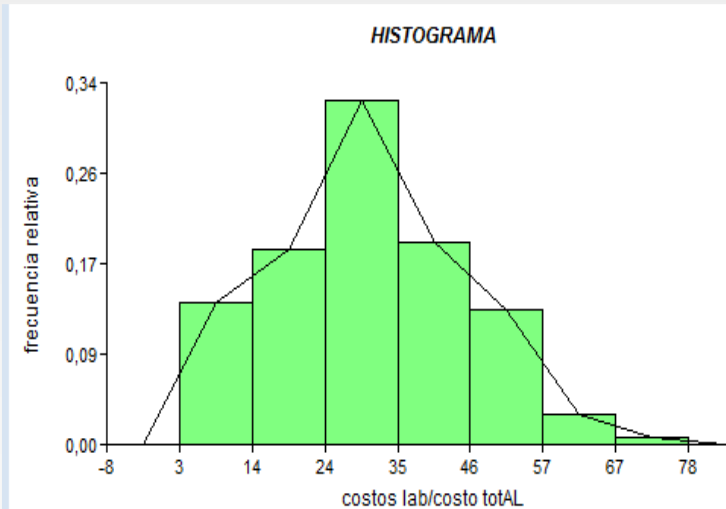
# Análisis unidimensional

(Variable numérica continua)

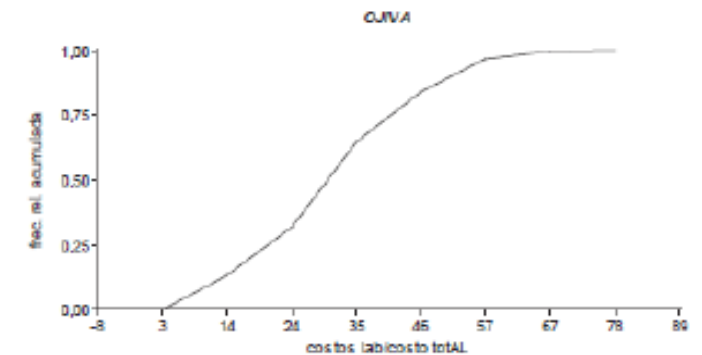
Porcentaje de costo laboral

PORCENTAJE DE COSTO LABORAL $y_i - y_i$	FREC. ABSOLUTA ( $n_i$ )	PORCENTAJE ( $h_i$ )	FREC. ABSOLUTA ACUM. $N_i$	PORCENTAJE ACUM. $H_i$
2 - 13	18	12,8	18	12,8
13 - 24	27	19,1	45	31,9
24 - 35	37	26,2	82	58,2
35 - 46	36	25,5	118	83,7
46 - 57	18	12,8	136	96,5
57 - 68	4	2,8	140	99,3
68 - 79	1	0,7	141	100,0
	141	100,0		

Escala de razón



Ojiva de base menor para el de frecuencias acumuladas Laboral



# Tablas bidimensionales o de contingencia



Se utilizan cuando interesa analizar dos variables categóricas o numéricas (discretas o continuas organizadas en intervalos)

Supongamos que contamos con  $n$  individuos clasificados de acuerdo a dos criterios que llamaremos A y B. Se sabe que A tiene  $r$  categorías y la variable B tiene  $s$  categorías. Con esta información podemos confeccionar una tabla como la siguiente:

	$B_1$	$B_2$	....	$B_s$	Total $n_{i.}$
$A_1$	$n_{11}$	$n_{12}$	....	$n_{1s}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	....	$n_{2s}$	$n_{2.}$
...	...	....	...	....	...
$A_r$	$n_{r1}$	$n_{r1}$	....	$n_{rs}$	$n_{r.}$
Total $n_{.j}$	$n_{.1}$	$n_{.1}$	.....	$n_{.s}$	$n$

$i$ : fila  
 $j$ : columna  
 $n_{ij}$ :  
frecuencia  
conjunta  
 $n_{i.}$  y  $n_{.j}$ :  
frecuencias  
marginales

# Tablas bidimensionales o de contingencia



## Distribuciones Bidimensionales – Frecuencias conjuntas y marginales

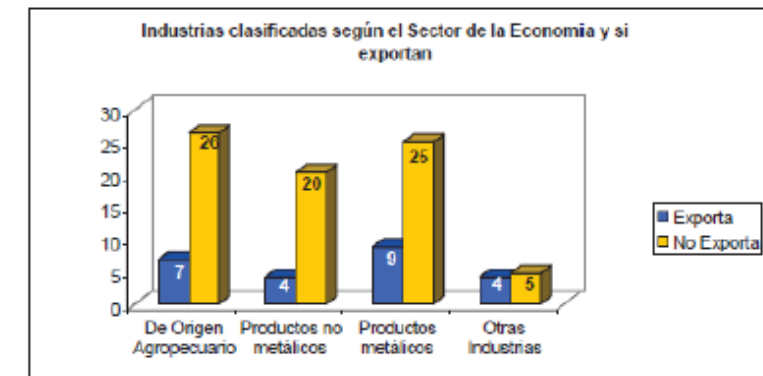
Tabla de contingencia Sector de la economía y Exporta

SECTOR DE LA ECONOMÍA	EXPORTA		TOTAL
	Si	No	
De Origen Agropecuario	10	39	49
Productos no metálicos	6	30	36
Productos metálicos	13	37	50
Otras Industrias	6	7	13
Totales	35	113	148

Frecuencias  
conjuntas

Frecuencias  
marginales

Análisis conjunto de Sector de la economía y Exporta.



# Tablas bidimensionales o de contingencia



## Distribuciones Bidimensionales – Frecuencias condicionadas

¿Participa o ha participado en emprendimientos conjuntos con otras empresas?

PARTICIPA	FRECUENCIA ABSOLUTA	PORCENTAJE
SI	16	10,8
NO	132	89,2
TOTAL	148	100,0

A las 16 empresas que contestaron afirmativamente, se les pidió que indique la localización de las empresas con las que se asociaron.

Localización de otras empresas con las que han participado en emprendimientos

LOCALIZACIÓN DE OTRAS EMPRESAS	FRECUENCIA ABSOLUTA	PORCENTAJE
Locales (100 km)	10	62,5
De la provincia	3	18,75
Nacionales	3	18,75
TOTAL	16	100,00

Participación en emprendimientos conjuntos con otras empresas y su localización





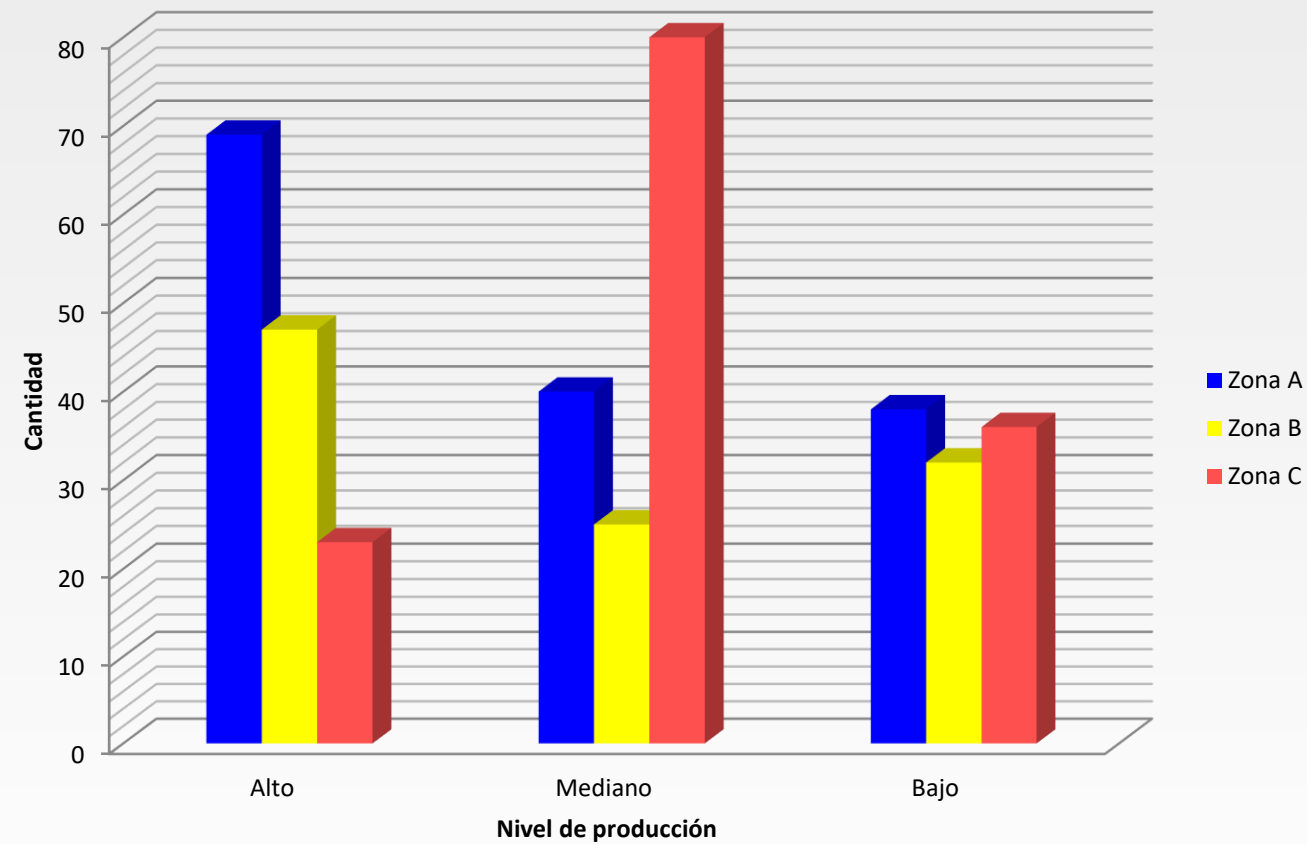
# Ejemplo 2



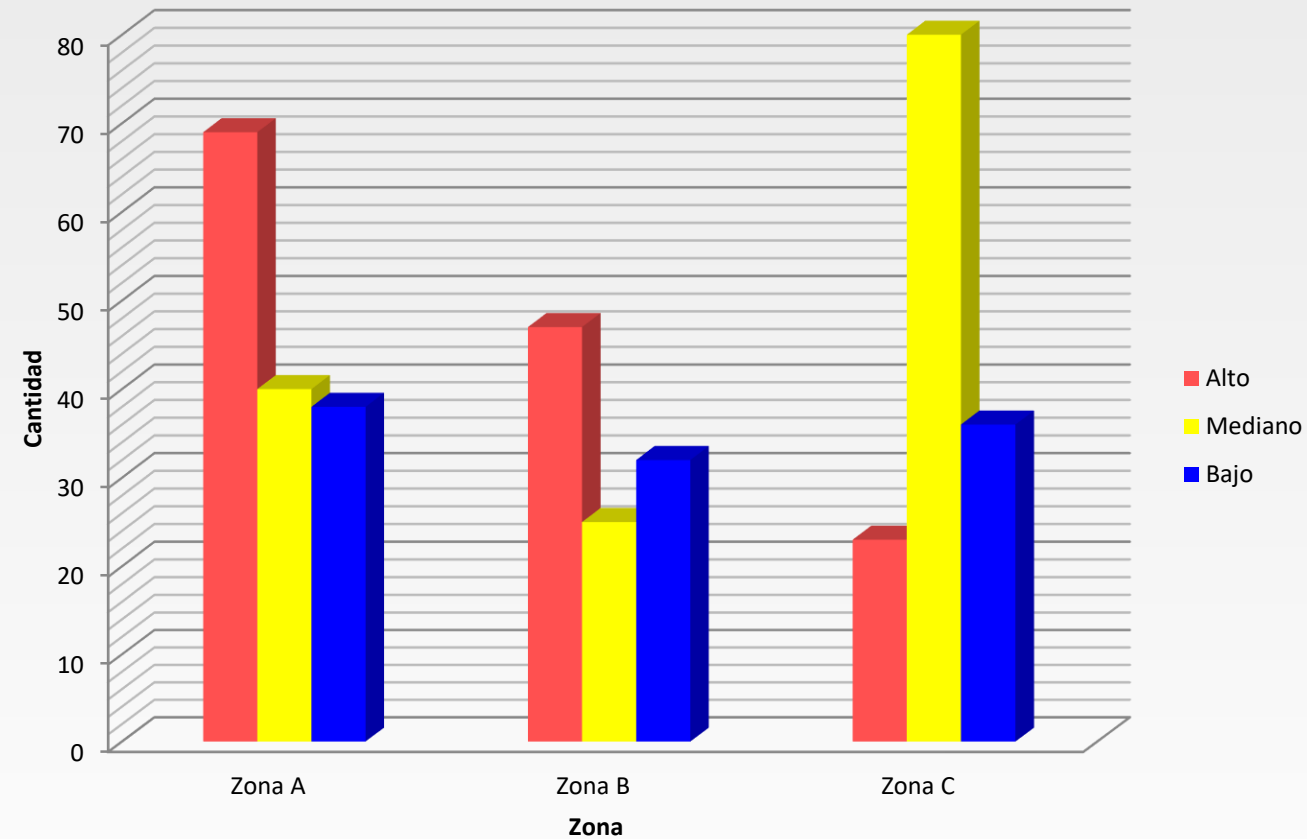
Los establecimientos agropecuarios de la Provincia de Córdoba han sido clasificados de acuerdo a sus niveles de producción en altos, medianos y bajos y de acuerdo a la zona geográfica donde se encuentran ubicados. Los resultados se muestran en la siguiente tabla:

<b>Nivel de pcción.</b>	<b>Zona A</b>	<b>Zona B</b>	<b>Zona C</b>	<b>Total</b>
Alto	69	47	23	139
Mediano	40	25	80	145
Bajo	38	32	36	106
Total	147	104	139	390

# Gráfico de barras agrupadas



# Gráfico de barras agrupadas



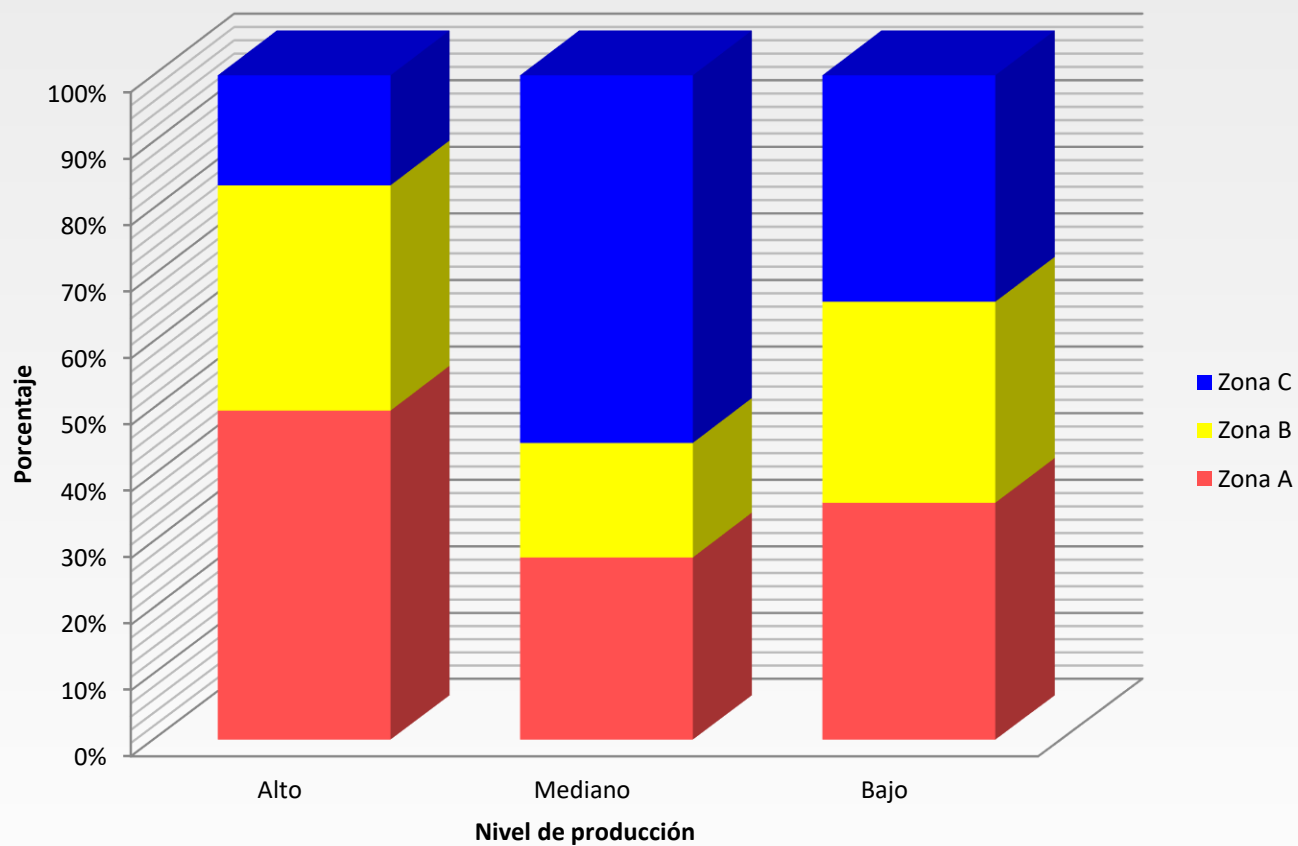
# Tablas bidimensionales o de contingencia



Porcentajes por fila  $n_{ij} / n_{i.}$

<b>Zona</b> <b>Nivel de pcción.</b>	<b>Zona A</b>	<b>Zona B</b>	<b>Zona C</b>	<b>Total</b>
Alto	49,6	33,8	16,5	100,0
Mediano	27,6	17,2	55,2	100,0
Bajo	35,8	30,2	34,0	100,0
Total	37,7	26,7	35,6	100,0

# Gráfico de barras porcentuales



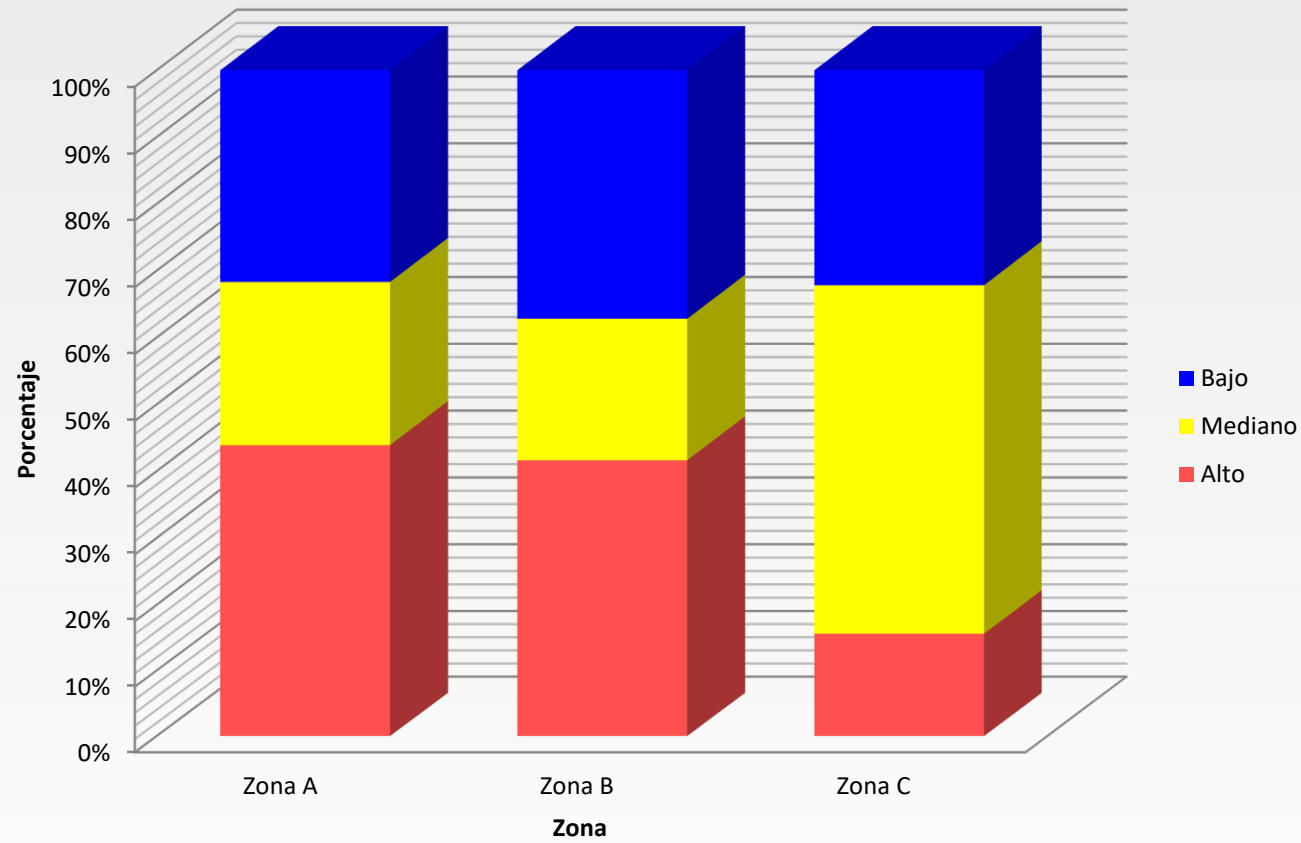
# Tablas bidimensionales o de contingencia



Porcentajes por columna  $n_{ij} / n_{.j}$

<b>Zona</b> <b>Nivel de pcción.</b>	<b>Zona A</b>	<b>Zona B</b>	<b>Zona C</b>	<b>Total</b>
Alto	46,9	45,2	16,5	35,6
Mediano	27,2	24,0	57,6	37,2
Bajo	25,9	30,8	25,9	27,2
Total	100,0	100,0	100,0	100,0

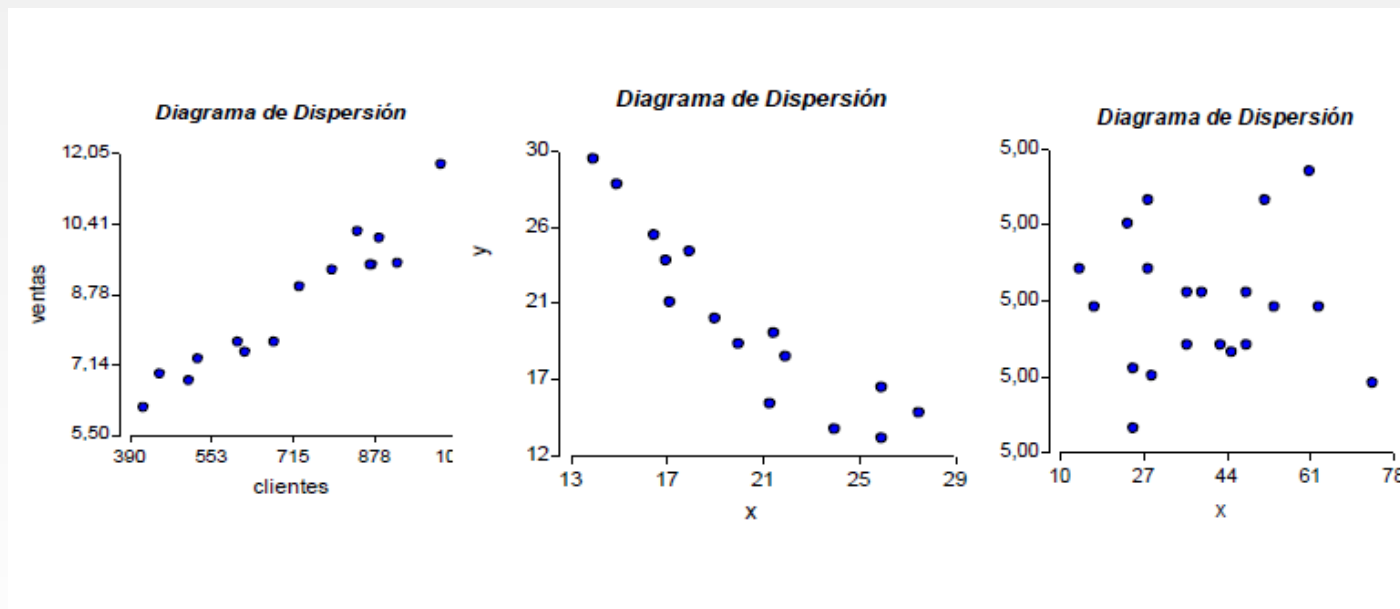
# Gráfico de barras porcentuales



# Diagrama de dispersión



El diagrama de dispersión es un gráfico bidimensional donde cada valor es graficado con sus coordenadas (x,y). Se utiliza cuando se analizan dos variables numéricas.





# Distribuciones unidimensionales



Para el análisis descriptivo de una variable podemos utilizar tres tipos de **medidas descriptivas**:

Medidas de posición: permiten determinar la ubicación de los datos a lo largo del eje real.

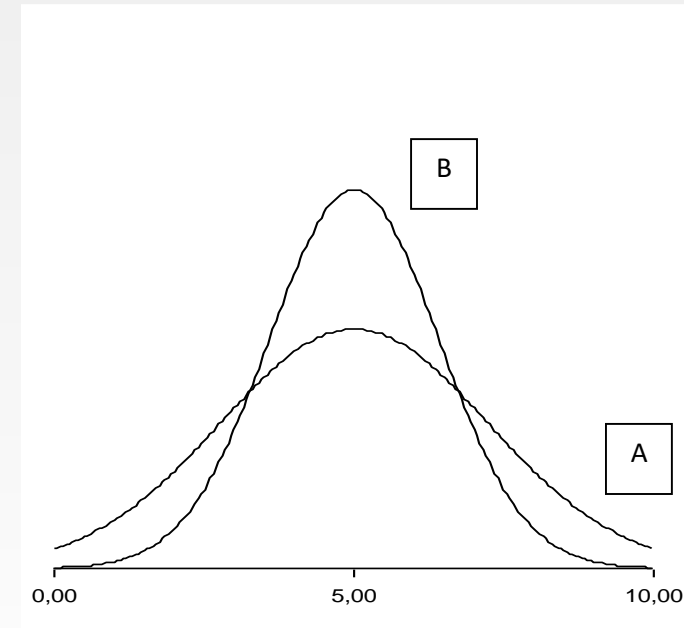
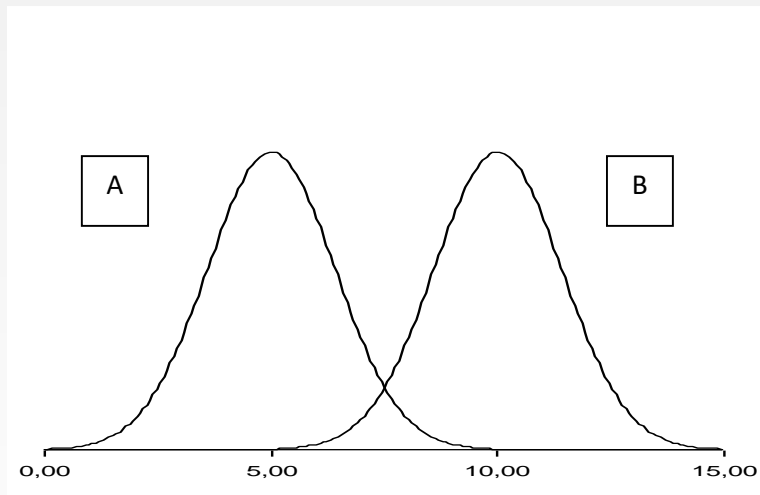
Medidas de dispersión: se refieren a la dispersión o concentración del conjunto de datos alrededor de las medidas de posición. Si los valores observados de una distribución están muy concentrados alrededor de un valor central esta medida es representativa del conjunto, pero si estos valores están muy dispersos la medida central pierde representatividad.

Medidas de forma: representan la deformación horizontal (simetría) y vertical (curtosis) del conjunto de datos.

# Analicemos...



Analice los gráficos que se exponen a continuación. ¿Qué diferencias observa?



# Medidas de posición

---



## Medidas de tendencia central



- Media aritmética
- Mediana
- Modo



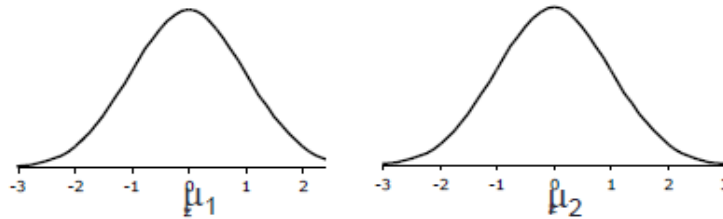
## Medidas de tendencia no central

- Cuartiles

# Medidas de posición



## Análisis de datos



## Medidas de posición de tendencia central

### Media aritmética

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

- Todo conjunto de escala medible tiene una media
- Un conjunto de datos solo tiene una media
- En su cálculo, se incluyen todos los valores de la variable, por lo cual es sensible a los valores extremos y puede No ser representativa

# Medidas de posición



## Propiedades de la Media Aritmética

$$M(c) = c$$

$$M(x + c) = M(x) + c$$

$$M(c.x) = c.M(x)$$

$$M(x) = \frac{n_1 M(x_1) + n_2 M(x_2) + \dots + n_h M(x_h)}{n}$$

$$\sum_{i=1}^k (x_i - M(x)) n_i = 0$$

$$\sum_{i=1}^k (x_i - M(x))^2 n_i = \text{Mínimo}$$

# Medidas de posición



Cuando en un conjunto de datos hay valores muy pequeños o muy altos (valores extremos) la media aritmética no es representativa de conjunto de datos y es necesario utilizar otra medida, como la mediana.

La **mediana** (Me) de una muestra es el valor de la variable que divide a un conjunto de datos ordenado en dos grupos iguales, es decir que no más de la mitad de las observaciones estarán por debajo de la mediana y no más de la mitad de las observaciones estarán por encima de la misma.

Para obtener la Me:

- Se ordenan los valores de menor a mayor

- Se determina el punto de posicionamiento de la Me  $\longrightarrow \frac{n + 1}{2}$

# Medidas de posición



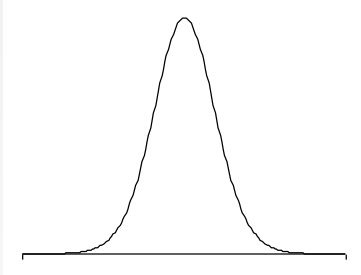
La **moda** ( $M_o$ ) es el valor de la variable que se repite mayor cantidad de veces, es decir, el más frecuente.

- No se ve afectado por valores extremos.
- Algunos conjuntos de datos no tienen un modo, otros tienen dos o más valores modales.
- No tiene propiedades matemáticas.
- Es fácil de calcular.
- Suele ser la única medida de obtener en un conjunto de datos categóricos.

# Forma de una distribución



Una primera inspección de la simetría de la distribución la podemos realizar comparando los valores de las medidas de posición: media aritmética, mediana y moda.



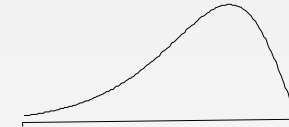
**Simétrica**

**Media= Me= Mo**



**Asimétrica derecha o con  
sesgo positivo**

**Media > Me > Mo**



**Asimétrica izquierda o con  
sesgo negativo**

**Media < Me < Mo**



# Medidas de posición



## Cuartiles

Son medidas descriptivas que dividen los datos ordenados en cuartos.

El Primer Cuartil ( $Q_1$ ) es un valor de la variable que divide el 25 % de los valores más bajos del 75 % de los valores restantes.

Me



El Tercer Cuartil ( $Q_3$ ) es un valor de la variable que divide el 25 % de los valores más altos del 75 % de los valores restantes.

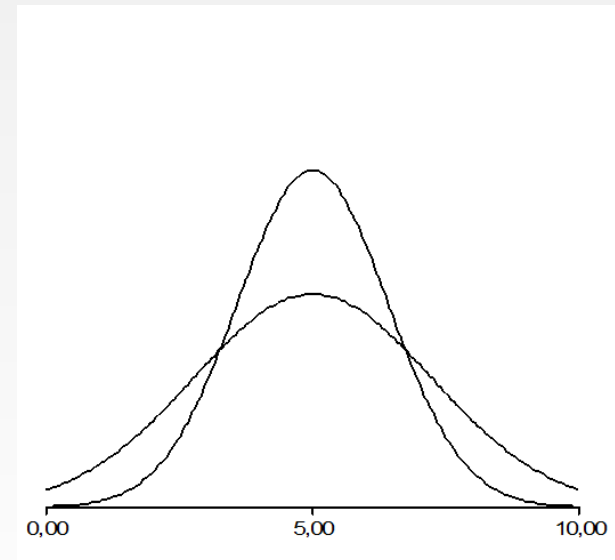


# Medidas de dispersión



Una **medida de dispersión** es un número que nos indica el grado de concentración de un conjunto de datos en torno a un valor particular, generalmente de tendencia central. Si el resultado es pequeño, entonces los valores son homogéneos o existe poca dispersión.

Llevando esto a un extremo, piense.....  
¿en qué caso la dispersión sería cero?



# Medidas de dispersión



## Varianza

Es la media aritmética del cuadrado de las desviaciones de cada observación respecto a su media.

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

Al calcularse como los desvíos al cuadrado de la variable respecto a la media está elevada a una magnitud superior que la variable original, por eso al interpretarla no resulta útil.



Desviación estándar o típica

# Medidas de dispersión



## Desviación estándar

- Se calcula con respecto a la media aritmética
- Cuanto mayor sea la dispersión, mayor será el valor de la varianza y desviación estándar

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

## Coeficiente de variación

$$CV(x) = \frac{S}{\bar{x}}$$

Dos conjuntos de datos son comparables a través de este coeficiente. Se trata de una medida adimensional de dispersión

Es una medida de dispersión relativa, refleja la desviación estándar como porcentaje de la media.

# Medidas de dispersión



## Rango

Es la diferencia entre el valor máximo y mínimo que asume la variable en estudio.

$$R = x_{\max} - x_{\min}$$

- Es fácil de calcular.
- Está basado únicamente en el valor más bajo y más alto de un grupo de datos.
- Al no considerar todos los valores de la variable es afectado por los valores extremos y puede no ser una medida adecuada de dispersión.
- Es útil cuando se trabaja con pocos datos

## Rango intercaartil

Se calcula como la diferencia entre el tercer cuartil y el primero y refleja la variación que existe en el 50 % central de los datos.

$$RI = Q_3 - Q_1$$

- No está afectado por los valores extremos.
- Al igual que el recorrido total de la variable, es una medida limitada, en cuanto a que no considera la totalidad de los datos, pero nos permite tener una primera aproximación de la variabilidad en el centro de la distribución.

# Ejemplo 3



La Secretaría de la Pequeña y Mediana Empresa y Desarrollo Regional (SEPyME) del Ministerio de Industria de la Nación ejecuta una serie de programas y líneas de trabajo que focalizan su atención en facilitar a las PyMEs el acceso al crédito. Uno de estos programas es el de Sociedades de Garantías Recíprocas.

Las Sociedades de Garantía Recíproca (SGR) son sociedades comerciales que tienen por objeto facilitar el acceso al crédito a las pymes a través del otorgamiento de garantías para el cumplimiento de sus obligaciones.

Han surgido como respuesta a los problemas que deben afrontar las pymes en sus relaciones de negocios con sectores de mayor tamaño o envergadura, Las Sociedades de Garantía Recíproca tienen por objeto otorgar garantías líquidas a sus socios partícipes (pymes) para mejorar sus condiciones de acceso al crédito.

Para evaluar la efectividad del programa se tomó una muestra al azar de 100 empresas en el año 2009 (antes del programa) y en los años siguientes.

# Ejemplo 3. Actividad

---



Analice:

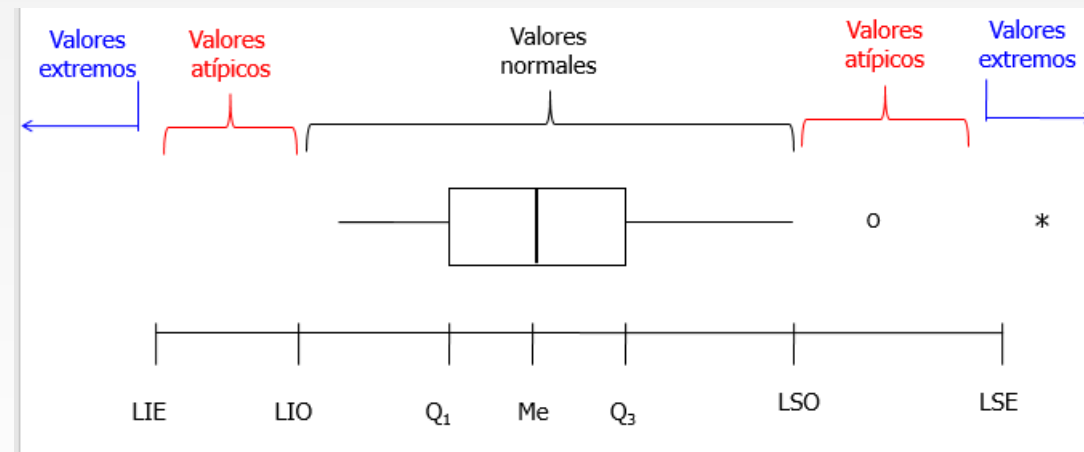
- Tipo de estudio
- Población
- Muestra
- Unidad estadística
- Tipo de datos: serie de tiempo, de corte transversal, datos de panel? (leer el archivo “Econometría y datos económicos”)
- Clasificar las variables de la base de datos: media del salario (en miles de \$); antigüedad de la empresa (en años); cantidad de personas ocupadas, rama de actividad, acceso al crédito (sí, no), situación frente al BCRA (1, 2 o 3), ventas brutas (miles de \$), monto del crédito otorgado (miles de \$).

# Diagrama de Caja y Brazos (Box plot)



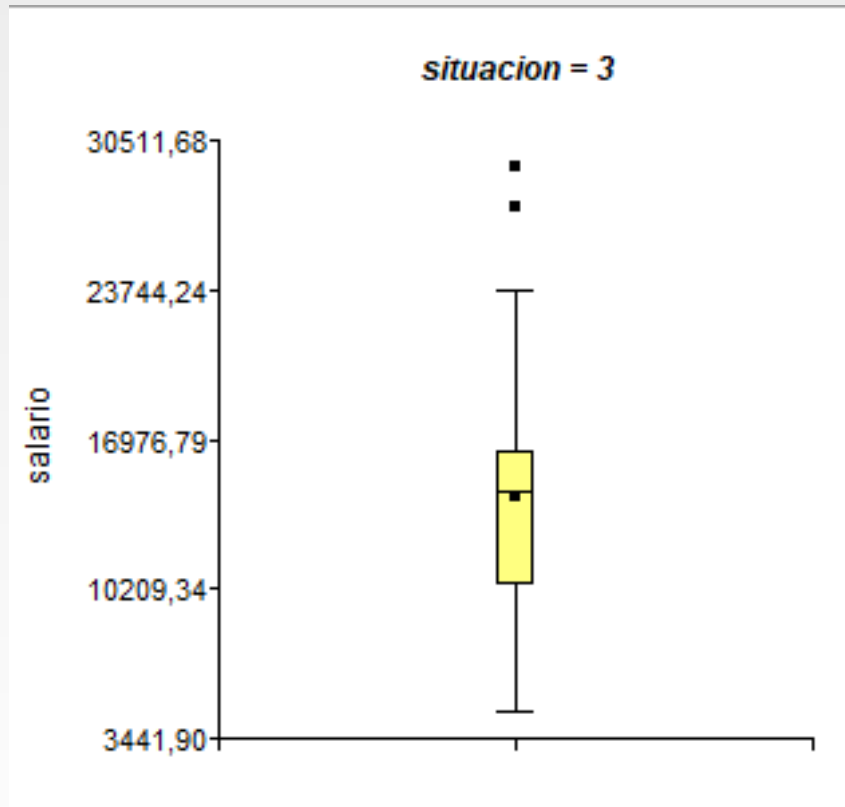
El **diagrama de Caja y Brazo** da una visión general del comportamiento de un conjunto de datos mediante el resumen de 5 medidas descriptivas ( $Me$ ,  $Q_1$ ,  $Q_3$ ,  $x_{\min}$ ,  $x_{\max}$ ) y permite identificar cualquier valor extremo.

Con los cuartiles se forma una caja y cuyos lados son el cuartil 1 y 3 y el cuartil 2 se marca en el centro, esta caja representa la distribución el 50 % central de los datos. Para construir los brazos se utiliza una regla empírica basada en el Recorrido Inter cuartílico, definiendo tres límites.





# Diagrama de Caja y Brazos



Box plot de la variable salario para empresas de la rama Agricultura, ganadería, caza, silvicultura y pesca en situación 3 (con problemas para atender normalmente la totalidad de sus compromisos financieros) frente al BCRA. Año 2009

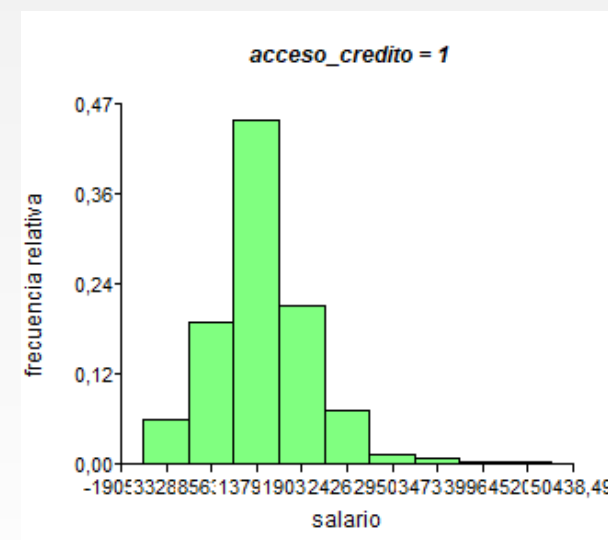
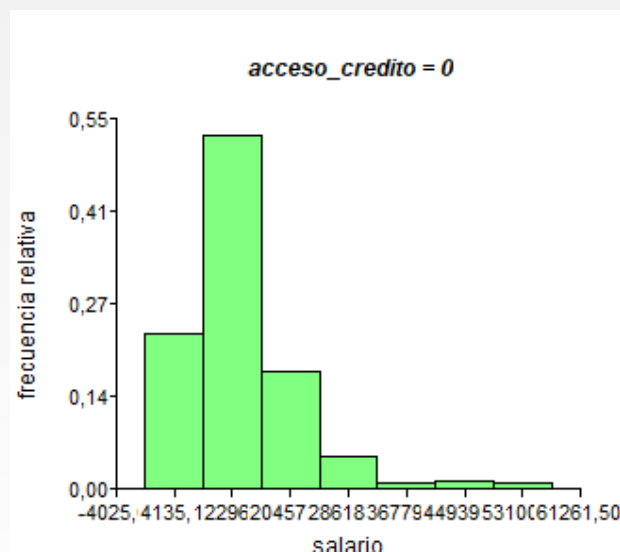


# Para interpretar

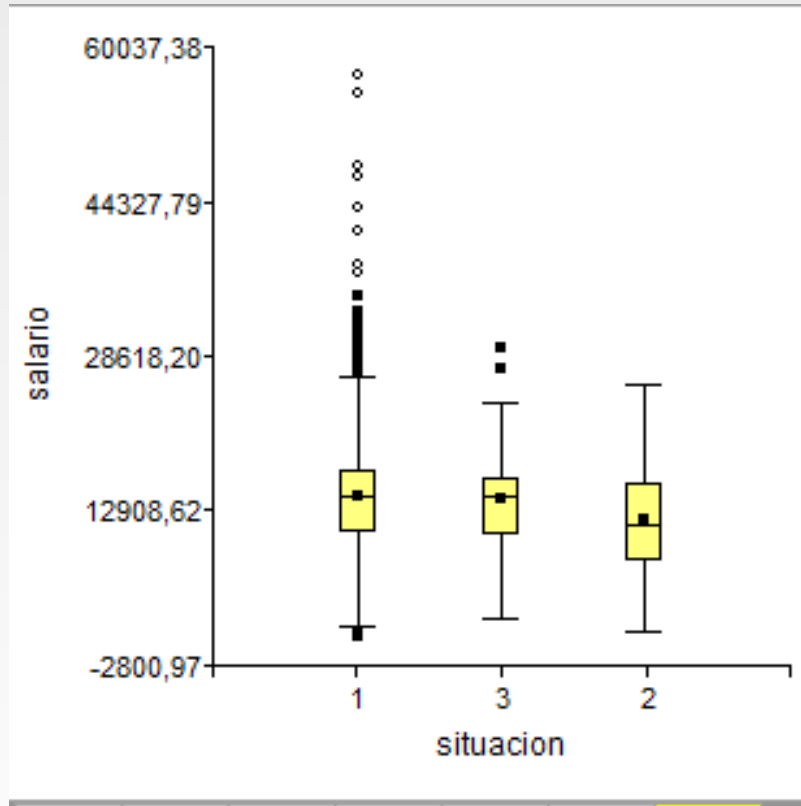
Variable salario para empresas de la rama Agricultura, ganadería, caza, silvicultura y pesca por situación según haya tenido acceso al crédito (1) o no (0). Año 2009

## Medidas resumen

acceso credito	Variable	n	Media	D.E.	Var(n-1)	CV	Mín	Máx	Mediana	Q1	Q3
0	salario	248	13770,57	8236,87	67845947,14	59,81	55,32	57181,09	13361,70	8461,70	16369,50
1	salario	604	14556,92	5581,49	31153029,96	38,34	711,49	47821,28	14605,87	11276,60	17050,21



# Para interpretar



Box plot de la variable salario para empresas de la rama Agricultura, ganadería, caza, silvicultura y pesca por situación frente al BCRA. Año 2009

1. Situación normal
2. Con seguimiento (hasta 90 días de atraso)
3. Con problemas (hasta 180 días de atraso)

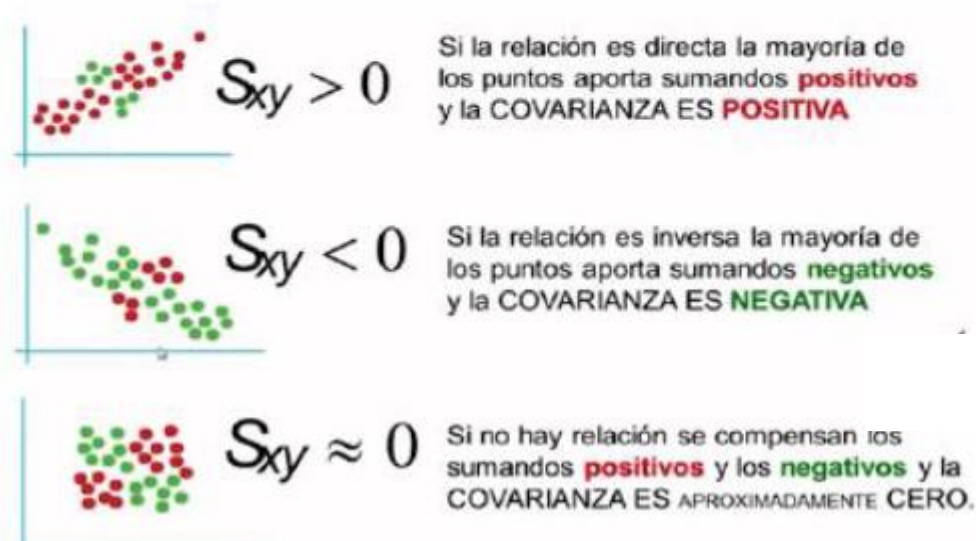
# Análisis bivariado



$$S_{xy} = \text{Cov}(x, y)$$

## Covarianza

Es una medida de dispersión conjunta entre dos variables cuantitativas que indica la presencia de asociación lineal directa o inversa entre las mismas.



# Análisis bivariado



## Coeficiente de correlación:

Mide la intensidad de la asociación lineal entre las variables. Puede asumir valores entre -1 y 1 y se calcula dividiendo la covarianza por el producto de las desviaciones estándar.

$$r = \frac{Cov(x,y)}{S_x S_y}$$

$r = 1$  Correlación directa perfecta

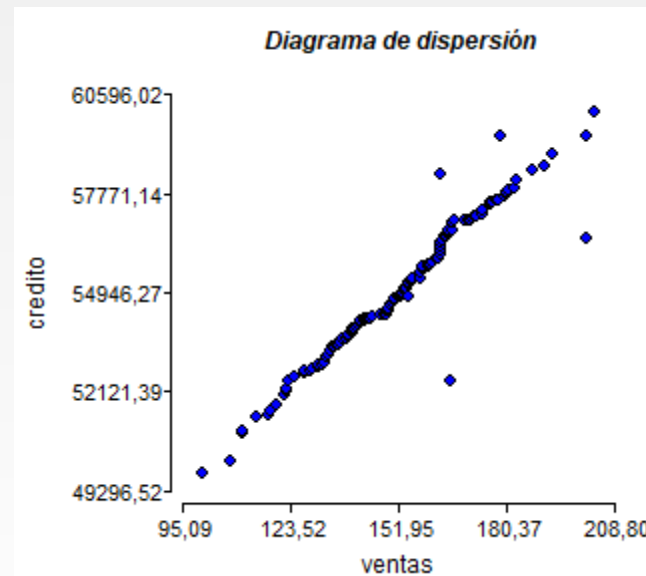
$r = 0$  Ausencia de relación lineal

$r = -1$  Correlación inversa perfecta

# Análisis bivariado



Se quiere estudiar si hay relación entre el monto del crédito obtenido (miles de \$) y el valor de las ventas brutas (miles de \$) para las empresas del sector Suministro de agua, cloacas, gestión de residuos y recuperación de materiales y saneamiento publico. Año 2009



$$\text{Cov}(x,y)=38795,58$$
$$r = 0,96$$

# Ejemplo 3. Actividad



Uno de los objetivos del programa de SGR es incrementar el acceso al crédito de las empresas. Analice en base a la información siguiente si ese objetivo se ha cumplido para las empresas manufactureras, teniendo en cuenta que se implementó en los años 2010 y 2011. Elaboren un informe de 1 página trabajando en los grupos indicados en el aula virtual (enviar hasta el 21/03/21 por Tareas del Aula Virtual).

## Medidas resumen

year	Variable	n	Media	D.E.	CV	Mín	Máx	Mediana	Q1	Q3
2009	credito	11493	8500,07	46540,50	547,53	200,00	2593866,41	1230,64	502,13	3916,60
2010	credito	12040	8589,44	42230,52	491,66	200,41	2087025,77	1230,64	492,53	3981,66
2011	credito	13438	9616,58	50490,90	525,04	200,32	2235744,24	1319,06	520,19	4380,72

La categoría 0 indica que no accedió al crédito y la 1 que si

## Tablas de contingencia

Frecuencias absolutas  
En columnas: acceso\_credito

año	0	1	Total
2009	24005	11493	35498
2010	24068	12040	36108
2011	23455	13438	36893
Total	71528	36971	108499