DIRECCIÓN DE ESTADÍSTICA DE LA PROVINCIA

Curso de Capacitación

INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

Víctor Fabio Lazarte – Paula Lorena Naidicz

San Miguel de Tucumán. 16 – 18 de Noviembre de 2009.

Secretaría de Estado de Planeamiento – Gobierno de la Provincia de Tucumán

Etapas de un estudio estadístico

El proceso de analizar estadísticamente un fenómeno consta de tres etapas, que se pueden describir brevemente de la siguiente manera:

1) **Recolección de datos.**

   Los datos que serán la base del estudio estadístico deben ser recolectados adecuadamente, pues de ello depende la validez y el alcance de los resultados y conclusiones.

   Ejemplos de formas de recolección de datos que no brindan información confiable son las encuestas telefónicas y las encuestas callejeras, ya que dejan grupos de la población sin ser representados.

2) **La Estadística Descriptiva**

   Se ocupa de resumir los datos para ser presentados en forma óptima mediante:

   - Tablas de frecuencias
   - Cálculo de medidas de posición y de dispersión.
   - Realización de gráficos y diagramas.

3) **La Inferencia Estadística**

   Se ocupa de la interpretación y generalización de la información obtenida a partir de los datos.

   - Permite realizar estimaciones y predicciones.
   - Sirve para tomar decisiones con un cierto nivel de riesgo.

Observaciones:

1) Naturalmente, estas tres etapas del estudio estadístico están íntimamente relacionadas entre si.
2) En este curso **no** se estudiarán las técnicas para recolectar adecuadamente los datos, es decir que partiremos de un dado conjunto de datos.

Los datos son mediciones u observaciones de ciertas características de interés que llamaremos variables, estas se clasifican en dos tipos.

TIPOS DE VARIABLES

Variables Cualitativas o Categóricas:

Son aquellas variables que describen cualidades o atributos, pero no toman valores numéricos.

Ejemplos:

a) "Nivel de educación de un ciudadano"

   Categorías: {Sin instrucción, Prim. incompleta, Prim. completa, Sec. incompleta, Sec. completa, Univ. o superior incompleta, Univ. o superior completa}

b) "Situación laboral de un ciudadano"

   Categorías: {Empleado, Desempleado, Inactivo}

c) "Departamento de residencia de un ciudadano de Tucumán"

   Categorías: {Burruyacu, Capital, Chicligasta, Cruz Alta, ... }

d) "Tipos de medios de transporte según distancia que recorren"

   Categorías: {corta distancia, media distancia, larga distancia}

Se puede observar que en el ejemplo a) y d) las categorías tienen un orden natural, en este caso se dice que la variable tiene **Escala Ordinal**

Si las categorías no tienen ningún orden, como en el ejemplo b) y c), se dice que la variable tiene **Escala Nominal**

Variables Cuantitativas:

Son aquellas variables que toman valores numéricos.

Ejemplos:

a) "Cantidad de artículos fabricados en un día por una empresa"

Valores que puede tomar: {0, 1, 2, …, n}

b) "Peso, en gramos, de un recién nacido"

Valores que puede tomar: (0, 5000)

Se llama **Variable Cuantitativa Discreta** a las variables que sólo toman valores enteros. Ejemplo a).

Se llama **Variable Cuantitativa Continua** a las variables que pueden tomar cualquier valor numérico. Ejemplo b).

DISTRIBUCIONES DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA DE DATOS

Una distribución de frecuencias es una tabla donde se presenta la información resumida de la siguiente forma: En una columna están las clases (o categorías) y en otra columna la cantidad de datos en cada clase. La tabla sacrifica parte de la información a cambio de obtener mayor claridad.

Primer caso: Supongamos que se cuenta con datos de una variable Cualitativa.

En este caso sólo se necesita agrupar los datos en sus categorías.

Ejemplo: Se quiere estudiar el nivel de educación de las madres de recién nacidos en cierta ciudad, para ello se toma una muestra al azar de 80 madres y se les pregunta el nivel de instrucción alcanzado.

Los datos están codificados de la siguiente manera:

Sin instrucción=1, Primaria incompleta=2, Primaria completa=3, Sec. incompleta=4, Sec. completa=5, Univ. o superior incompleta= 6, Univ. o superior completa=7. Los datos se muestran en la siguiente tabla.

Tabla 1: Nivel de instrucción de 80 madres encuestadas.

3	5	3	1	4	3	7	5
5	3	4	3	3	6	2	5
1	4	4	4	4	5	4	4
3	2	3	3	3	4	6	5
3	2	2	3	4	3	3	7
3	3	3	5	4	5	6	4
2	6	4	5	3	5	3	4
3	4	3	2	4	5	4	4
3	7	4	4	3	5	5	6
2	5	2	2	5	7	5	5

Agrupando los datos según su categoría se obtiene la distribución de frecuencias para esta variable.

Tabla 2: Distribución de frecuencias del nivel de educación de la madre.

Nivel de educación de la madre	**Frecuencias**	**Porcentajes**	**Porcentajes acumulados**
Sin instrucción	2	2,5%	2,5%
Primaria incompleta	9	11,3%	13,8%
Primaria completa	23	28,8%	42,6%
Secundaria incompleta	20	25,0%	67,6%
Secundaria completa	17	21,3%	88,9%
Sup. o universitaria incompleta	5	6,3%	95,2%
Sup. o universitaria completa	4	5,0%	100%
Total	80	100,0%	

Su correspondiente representación gráfica se puede realizar mediante el *DIAGRAMA DE BARRAS.* Figura 1, o *DIAGRAMA CIRCULAR*. Figura 2.

Figura 1: Distribución de frecuencias del nivel de educación de la madre.

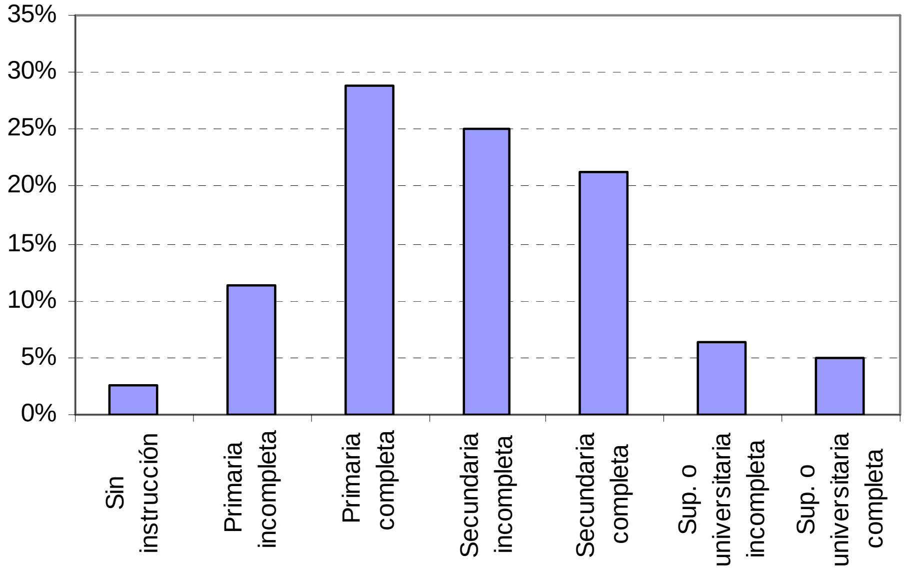
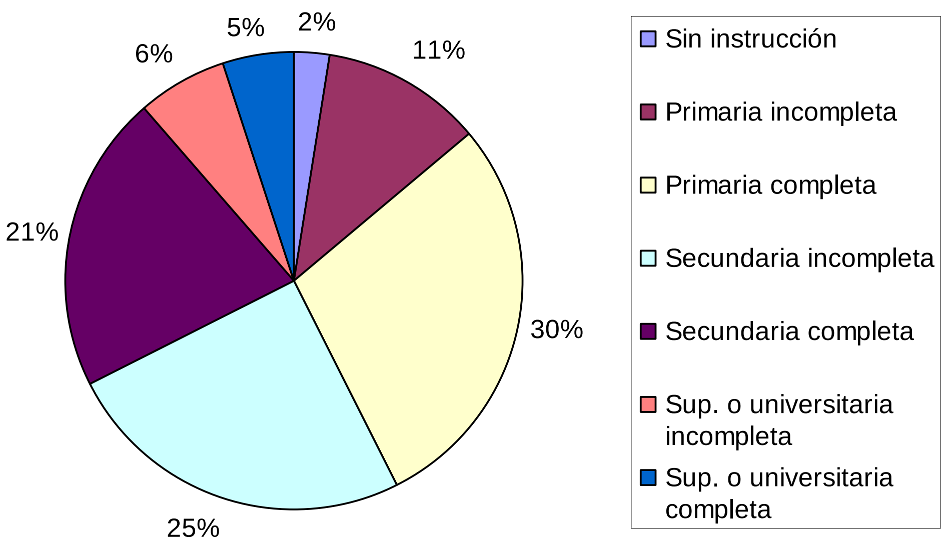


Figura 2: Distribución de frecuencias del nivel de educación de la madre.



<u>Segundo caso</u>: Datos provenientes de una variable Cuantitativa discreta con pocos valores

En este caso, similar a lo anterior, sólo se necesita agrupar los datos iguales.

Ejemplo: Se desea estudiar la distribución de la cantidad de miembros de una familia de cierta zona. Se toma una muestra al azar de 80 familias y se registra el número de miembros.

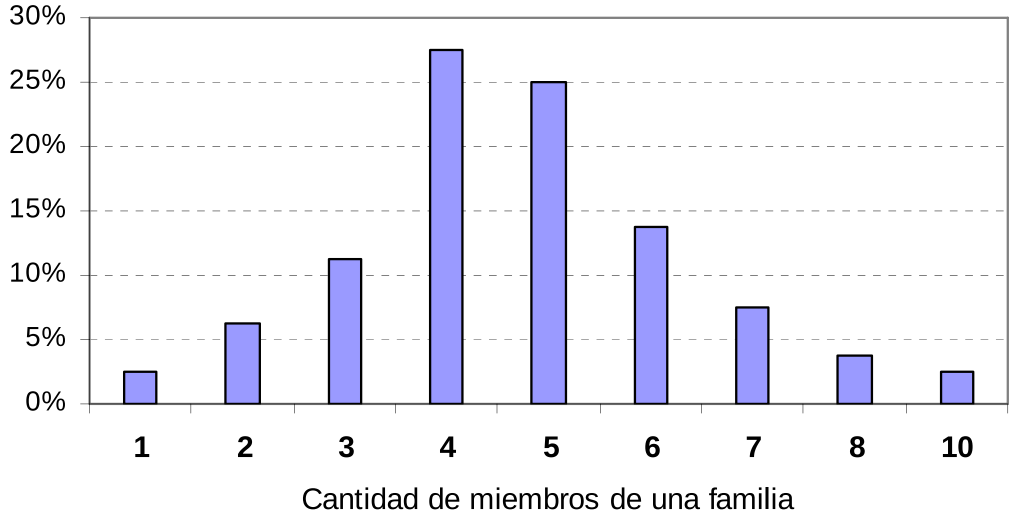
Tabla 3: Cantidad de miembros de las familias seleccionadas.

6	5	6	4	8	4	4	4
4	6	4	2	8	2	5	5
8	5	2	4	3	5	6	6
5	7	5	7	5	5	3	5
1	2	7	5	7	4	4	5
4	6	6	5	4	5	4	4
7	3	2	3	6	5	3	3
5	5	3	3	4	4	4	4
4	4	5	5	5	7	6	6
1	3	10	4	10	4	4	6

Tabla 4: Distribución de frecuencias de la cantidad de miembros de las familias de cierta zona.

Cantidad de miembros	Frecuencias	Porcentajes	% acumulado
1	2	2,5%	2,5%
2	5	6,3%	8,8%
3	9	11,3%	20,0%
4	22	27,5%	47,5%
5	20	25,0%	72,5%
6	11	13,8%	86,3%
7	6	7,5%	93,8%
8	3	3,8%	97,5%
10	2	2,5%	100,0%
Total	**80**	**100,0%**	

Figura 3: Distribución de frecuencias de las cantidad de miembros de las familias de cierta zona.



<u>Tercer Caso</u>: Observaciones de una variable Cuantitativa Continua (o Discreta con muchos valores)

En este caso existe la necesidad de construir las clases. Para ello se subdivide el rango de datos en subintervalos, según la siguiente regla

Regla práctica

1. Selección del número de clases $\quad n^{o} \text{ de clases} = 1.75\sqrt[3]{n}$ donde n es la cantidad de datos.

2. Selección de la amplitud de clase $\quad \text{Amp de clase} = \dfrac{\text{máximo} - \text{mínimo}}{n^{o} \text{ de clases}}$.

3. Elegir intervalos semiabiertos y disjuntos que cubran todo el rango de datos.
4. Elegir los límites de los intervalos como números sencillos, es decir, fáciles de leer y recordar.

Ejemplo. Se desea estudiar el peso de los recién nacidos en un cierto periodo en una ciudad. Se toma una muestra al azar de 80 recién nacidos y se registra su peso al nacer, los datos se muestran en la siguiente tabla:

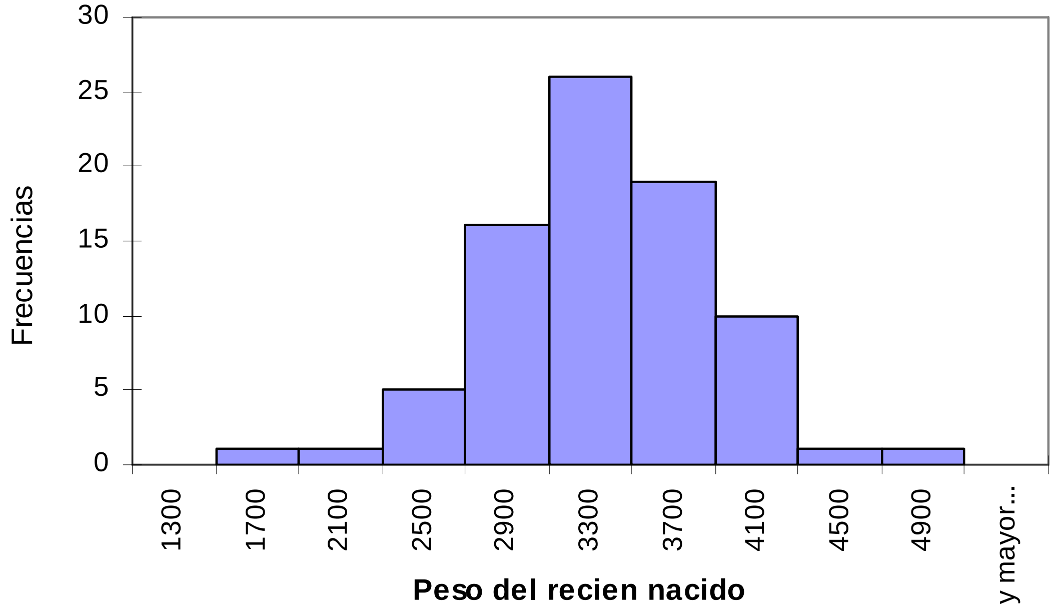
Tabla 5: Pesos de los recién nacidos muestreados.

3300	3220	3060	2600	3150	3400	3200	2830
3200	3450	3150	2250	4600	3300	3000	3200
2975	3500	3480	3700	3000	3150	2600	3400
2650	4170	3650	2780	3665	4000	3550	3050
2800	3100	3140	2800	3800	2900	2900	2300
3900	3000	2900	3100	2500	3800	2270	3400
2570	3250	2570	3300	3000	2650	3440	3950
2900	3800	3200	1390	1920	3150	3100	3800
3900	3410	3300	2415	3800	3550	2900	3670
3900	3250	3500	3400	3420	2750	3400	3450

Tabla 6: Distribución de frecuencias del peso de los recién nacidos.

Clase	Frecuencias	Porcentajes	% Acumulados
901 a 1300	0	0,0%	0,0%
1301 a 1700	1	1,3%	1,3%
1701 a 2100	1	1,3%	2,5%
2101 a 2500	5	6,3%	8,8%
2501 a 2900	16	20,0%	28,8%
2901 a 3300	26	32,5%	61,3%
3301 a 3700	19	23,8%	85,0%
3701 a 4100	10	12,5%	97,5%
4101 a 4500	1	1,3%	98,8%
4501 a 4900	1	1,3%	100,0%
Total	**80**	**100,0%**	

Figura 4: Distribución de frecuencias del peso de los recién nacidos en un cierto periodo de una cierta ciudad.



Observación:

La cantidad de clases resultantes puede diferir del número de clases calculada con la fórmula, en una clase más o una menos.

MEDIDAS DE POSICIÓN Y DISPERSIÓN

(Solo para datos de variables cuantitativas)

El conjunto de datos también se puede resumir en unos pocos valores numéricos que sirven para describir ciertas características específicas de la muestra.

MEDIDAS DE POSICIÓN:

Resumen la información referida de la posición de la muestra. Las medidas de posición más utilizadas son: Media, Mediana, Moda y Cuartiles, las tres primeras, media, mediana y moda son medidas de tendencia central.

Supongamos que los datos observados son: $x_1, x_2, ..., x_n$.

Media: es el promedio de los datos $\overline{x} = \dfrac{x_1 + x_2 + ... + x_n}{n}$

Ejemplo 1: Las notas de un alumno son: 10, 8, 7, 9, 10, 7, 6.

Nota promedio $\overline{x} = \dfrac{10 + 8 + 7 + 9 + 10 + 7 + 6}{7} = 8.14$

En este caso la media es un buen representante del rendimiento académico del alumno.

Ejemplo 2: En el ejemplo de la cantidad de miembros de una familia de cierta ciudad.

La cantidad promedio de miembros de una familia es $\overline{x} = 4.7$.

Existen algunos casos en donde la media no es un buen representante del conjunto de datos.

Ejemplo 3: Una empresa tiene 5 empleados, sus sueldos son: 1000, 1000, 1000, 1000, 10000.

Entonces la media es $\overline{x} = 2800$.

El dueño de la empresa diría orgulloso *¡El sueldo promedio de mi empresa es de 2800 al mes!!*

Pero claramente, en este caso la media no es un buen representante de este conjunto de datos, dado que hay 4 empleados que ganan sólo 1000. En general cuando la muestra tenga valores alejados o presente asimetría, la media no será un buen representante del conjunto de datos.

Media para datos agrupados en clases $\overline{x} = \sum_{i=1}^{k} x_i f_r(x_i)$, donde x_i es el centro de cada intervalo y k es la cantidad de clases

Es una forma aproximada, por lo tanto, si se cuenta con los datos originales conviene calcular la media a partir de ellos.

Una medida de posición alternativa a la media sería la *Mediana*.

<u>Definición</u>:

La mediana es un valor de la muestra que, en cierto modo, divide al lote de datos, ordenado, en 2 partes iguales, es decir:

- Si la cantidad de datos es impar, la mediana es el valor central de la muestra ordenada

- Si la cantidad de datos es par, la mediana es el promedio de los dos datos centrales de la muestra ordenada

Ejemplo

1) Para n impar.

Notas de un alumno A: 10, 8, 7, 9, 10, 7, 6.

Datos ordenados: 6, 7, 7, **8**, 9, 10, 10.

El dato que está en el centro es 8 por lo tanto la Mediana de las notas es 8. Notación: $\tilde{x} = 8$

2) Para n par

Notas de un alumno B: 10, 8, 7, 9, 10, 7, 6, 9.

Datos ordenados: 6, 7, 7, **8**, **9**, 9, 10, 10.

En este caso quedan dos datos en el medio de la muestra ordenada. Por lo tanto la mediana será el promedio de los dos datos centrales. $\tilde{x} = \frac{8+9}{2} = 8{,}5$.

En el ejemplo de la empresa con 5 empleados la mediana sería igual a $ 1000, por lo tanto sería un mejor representante del conjunto de datos.

Observación:

La mediana no es afectada por valores alejados ni es afectada por asimetría.

Cuartiles: Para los cuartiles se utiliza básicamente la misma idea que la mediana. En cierta forma, dividen el lote de datos ordenado en cuatro partes iguales.

Existen diferentes formas de calcular los cuartiles, una de ellas es tomando la primera mitad de la muestra ordenada (es decir de la mediana para abajo) y se calcula la mediana de este conjunto que será el primer cuartil, para el tercer cuartil se trabaja igual con los datos de la segunda mitad.

En el ejemplo 1 de la página anterior

Primer cuartil: $Q_1 = 7$,   Segundo cuartil: $Q_3 = 9.5$

En el ejemplo 2 de la página anterior

Primer cuartil: $Q_1 = 7$,   Segundo cuartil: $Q_3 = 9.5$.

Moda: Es el dato más frecuente (si es que este existe)

En los ejemplos de las notas de los alumnos no existe la moda.

En el ejemplo de la empresa con 5 empleados la moda sería igual a $ 1000, por lo tanto sería un buen representante del conjunto de datos.

En el ejemplo de la cantidad de miembros de las familias de cierta zona la moda será igual a 4 miembros, ya que hay 22 familias con esa cantidad de miembros. (Ver tabla 4)

MEDIDAS DE DISPERSIÓN:

Ejemplo 1:

Notas del alumno Juan: 6, 6, 8, 10, 10.

Notas del alumno Pedro: 8, 8, 8, 8, 8.

Observe que la nota promedio de los dos alumnos es 8, sin embargo claramente su desempeño no es igual ¿Cómo los comparo? ¿Cuál es la diferencia?

Una medida de dispersión es una medida de cuan alejados están los datos del centro de la distribución, ya sea que se tome como centro a la media o a la mediana de los datos.

Varianza muestral

Definición: La varianza es el promedio de los desvíos al cuadrado, es decir, se mide la distancia de cada dato a la media, se la eleva al cuadrado y se las promedia, en fórmulas.

$$\text{Varianza} = \hat{\sigma}^2 = \frac{(x_1 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n}$$

Esta medida de dispersión tiene la desventaja de estar expresada en unidades al cuadrado, por ello se define la *Desviación Estándar* como la raíz cuadrada positiva de la varianza, es decir:

$$\text{Desviación Estándar} = \hat{\sigma} = \sqrt{\text{Varianza}}$$

Desviación estándar corregida = $s = \sqrt{\frac{(x_1 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}}$

Ejemplo: Notas del alumno Juan: 6, 6, 8, 10, 10.

La media es 8 por lo tanto los desvíos son: -2, -2, 0, 2, 2, los desvíos al cuadrado son: 4, 4, 0, 4, 4.

Así la varianza será $\hat{\sigma}^2 = \frac{16}{5} = 3.2$ y la Desviación Estándar será $\hat{\sigma} = \sqrt{3.2} = 1.79$

Para el alumno Pedro naturalmente la varianza y la desviación estándar serán iguales a cero.

Observación: Ninguna de estas medidas se calculan a manualmente, todas se las calcula con Excel o cualquier calculadora científica.

Coeficiente de variación

El coeficiente de variación es una medida de la magnitud de la dispersión en relación a la media.

$$\text{Coeficiente de variación} = \frac{\text{Desviación estándar}}{\text{Media}}$$

Observemos que es adimensional, por lo tanto también sirve para comparar datos de magnitudes diferentes.

Importante:

- Estas tres medidas toman como centro a la media, por lo tanto están asociadas a ella.
- Si el lote de datos es simétrico y no tiene valores alejados, utilizaremos a la media y la desviación estándar o s para describir el lote de datos.

Una medida de dispersión asociada a la mediana es el Rango Intercuartil.

El Rango intercuartil es simplemente la diferencia entre el primer y el tercer cuartil:

$$RI = Q_3 - Q_1$$

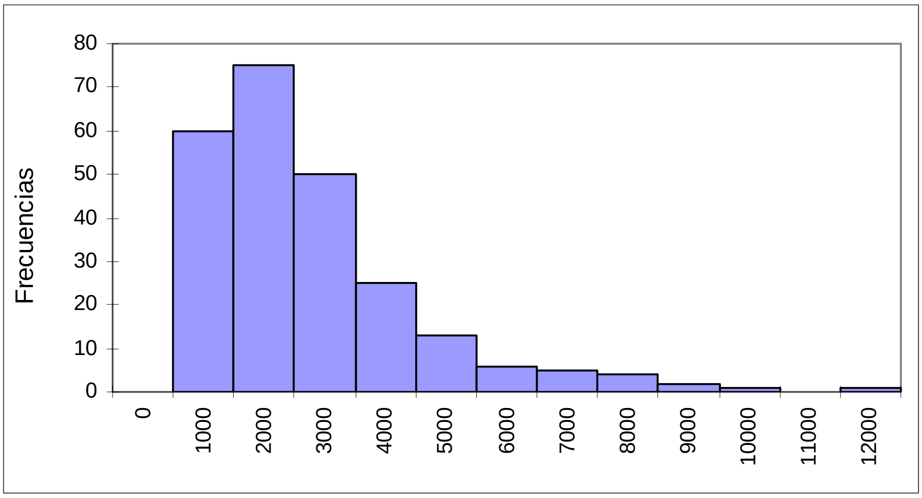
Cuando se presenta la estadística descriptiva de un conjunto de datos se debe reportar una medida de posición con una de dispersión asociada, entonces:

- Si el conjunto de datos es simétrico y no tiene valores alejados se recomienda utilizar a la media como medida de posición con la desviación estándar como medida de dispersión.
- Si el conjunto de datos es asimétrico o tiene valores alejados se recomienda utilizar a la mediana como medida de posición con el rango intercuartil como medida de dispersión.

¿Cómo analizo simetría?

Por lo general la asimetría de un conjunto de datos se ve gráficamente, pero también es posible calcular algún coeficiente de asimetría, ejemplo $\left|\frac{\bar{x}-\tilde{x}}{\hat{\sigma}}\right| 100\%$.

Figura 5: Ejemplo de una distribución asimétrica positiva



Valores alejados: Estos son valores observados que se apartan demasiado del resto de la muestra. Para detectarlos se puede utilizar la siguiente regla:

- Si un valor x_i de la muestra es menor que $Q_1 – 1.5(Q_3 – Q_1)$, entonces x_i es alejado por defecto

- Si un valor x_i de la muestra es mayor que $Q_3 + 1.5(Q_3 – Q_1)$, entonces x_i es alejado por exceso

No significa que haya que descartar ese dato, significa que hay que estudiar ese caso y usar medidas que no se vean afectadas por valores alejados.

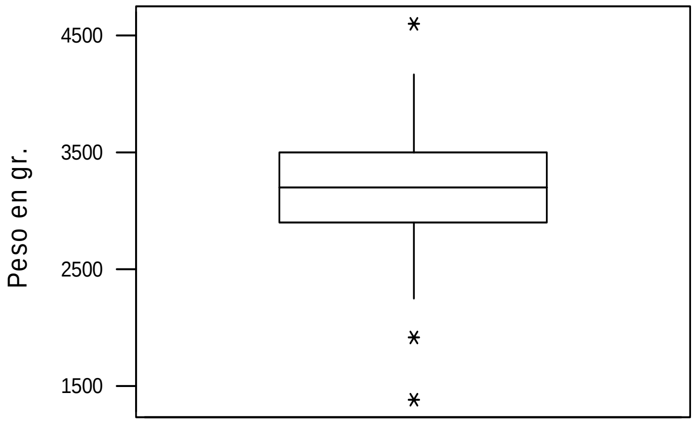
Diagrama de tipo caja:

Es una representación gráfica muy importante que se construye a partir de ciertas medidas resumen.

- Se deben calcular mínimo y máximo, mediana, Q_1 , Q_3 y Detectar valores alejados

Ejemplo: En el ejemplo de los pesos de los recién nacidos el diagrama de tipo caja para esos datos se presenta en la siguiente figura

Figura 5: Peso en gramos de los recién nacidos en un cierto periodo de una cierta ciudad.



- Se nota la presencia de 3 valores alejados.
- Se utilizan para describir una muestra y también para comparar dos o más muestras.
- Con éste diagrama se puede visualizar claramente rango, posición, dispersión, presencia de valores alejados y forma de la distribución de los datos.

Distribuciones de Frecuencias Conjuntas

Dado una unidad experimental podemos observar o medir más de una variable simultáneamente. Por lo general se necesitan estudiar todas las variables a la vez, no analizar a cada una por separado. A este tipo de estudio en Estadística se le llama análisis multivariado, en caso de 2 variables se llama análisis bidimensional.

Por ejemplo: X = "peso de una persona"

Y = "altura de una persona"

<u>Distribuciones de frecuencia bivariadas y su representación gráfica</u>

Básicamente se trata de estudiar la relación entre dos variables. Pueden presentarse tres casos.

1er caso: X e Y ambas variables cualitativas. En este caso la distribución conjunta se llama "tabla de contingencia"

<u>Ejemplo</u>: Se desea estudiar si existe o no relación entre los hábitos de fumar de una persona y el hecho de padecer problemas de hipertensión. Es decir que se sospecha que el fumar aumenta el riesgo de tener problemas de hipertensión.

Se definen las siguientes variables.

X = “Condición de Fumador”.

Categorías = {No fuma, Fuma Moderadamente, Fuma en Exceso}

Y = “Antecedentes de hipertensión”.

Categorías = {Si tiene hipertensión, No tiene hipertensión}

En el análisis bivariado los datos están de a pares, es decir, a cada persona se le asigna el valor de X y de Y. En este caso tendrían la siguiente forma.

Tabla 7: Datos de condición de fumador y condición de hipertenso.

Persona	**X (condición de fumador)**	**Y (condición de hipertenso)**
1	No F.	Si tiene
2	No F.	No tiene
3	No F.	No tiene
4	F. Exc.	No tiene
5	No F.	No tiene
6	No F.	No tiene
7	F. Mod.	Si tiene
8	F. Exc.	Si tiene
9	F. Mod.	Si tiene
10	F. Mod.	Si tiene
11	F. Exc.	Si tiene
12	No F.	Si tiene
13	F. Mod.	No tiene
14	No F.	No tiene
15	F. Mod.	No tiene
16	No F.	No tiene
…	…	…

Construyendo la correspondiente distribución de frecuencias tendríamos:

Tabla 8: Distribución de frecuencias conjunta de condición de fumador vs. condición de hipertenso.

Y / X	Si tiene	No tiene	**Total**
No fuma	21	48	**69**
Fuma Mod.	36	26	**62**
Fuma exc.	30	19	**49**
Total	**87**	**93**	**180**

Distribuciones marginales: Son los totales por filas y columnas, estos corresponden a las distribuciones de las variables X e Y respectivamente.

Estudio de porcentajes:

Se pueden calcular porcentajes sobre total de filas o total de columnas. Sirve para estudiar la influencia de una variable sobre la otra.

En este ejemplo, para saber en que sentido calcular los porcentajes tenemos que preguntarnos que variable influye sobre cual, digamos:

¿el hecho de fumar puede influir en tener o no hipertensión? o por el contrario, ¿tener o no hipertensión, influye en la decisión de fumar? Parece que la primera es más razonable.

Una vez resuelto este problema calculamos los porcentajes sobre los totales de las categorías de la posible variable influyente, en este caso sería conveniente calcular porcentajes sobre el total de filas. Tabla 9.

No siempre es fácil identificar cuál es la variable influyente, incluso a veces es útil calcular los porcentajes en los dos sentidos.

Tabla 9: Distribución conjunta de porcentajes sobre total de filas

X \ Y	Si tiene	No tiene	**Total**
No fuma	30 %	70 %	**100 %**
Fuma Mod.	58 %	42 %	**100 %**
Fuma exc.	61 %	39 %	**100 %**

Se puede observar que en el grupo de no fumadores el porcentaje de hipertensos es del 30 % mientras que mientras que en los grupos de fumadores y fumadores en exceso este aumenta a 58% y 61% respectivamente. Por lo que los datos indican que si hay relación entre fumar e hipertensión.

Representación Gráfica: Diagramas de barras o Diagrama Circular

Figura 6: Distribución de personas con y sin hipertensión
según su condición de fumador

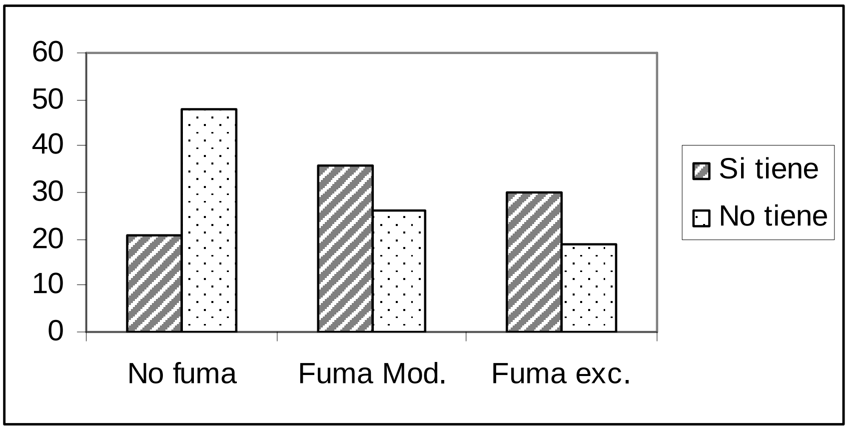
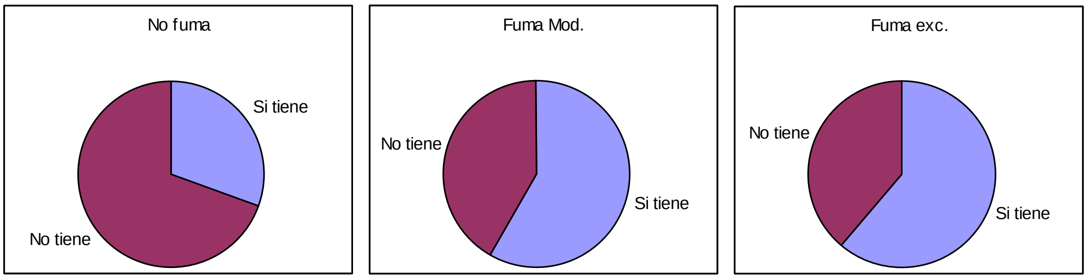


Figura 7: Distribución de personas con y sin hipertensión según su condición de fumador



2do caso: X Cualitativa e Y Cuantitativa

Ejemplo: Se desea estudiar si las personas de sexo masculino tienen mayores niveles de colesterol en la sangre que las de sexo femenino. Se definen las siguientes variables

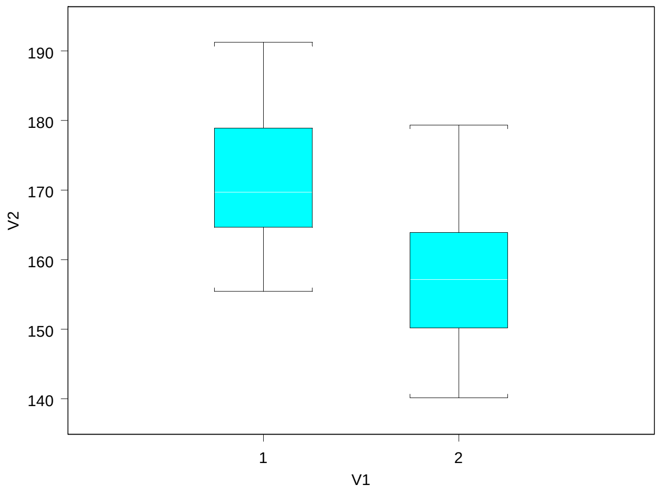
X = "Sexo de una persona", Categorías = {Masculino, Femenino}

Y = "Nivel de colesterol de la persona"

Comúnmente a la cualitativa se le llama variable agrupadora.

Representación Gráfica: Una forma adecuada de representación gráfica para comparar el nivel de colesterol en estos dos grupos son los diagramas de tipo cajas.

Figura 8: Niveles de colesterol por sexo. El grupo 1 es de los varones y el grupo 2 de las mujeres.



3er caso: X e Y Cuantitativas

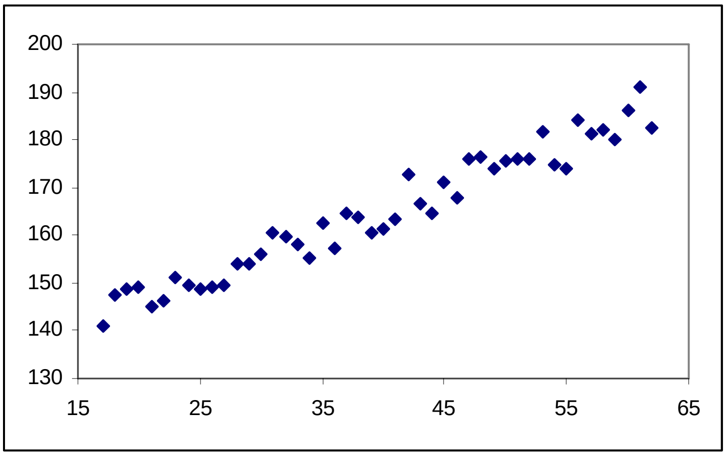
Ejemplo: Se desea estudiar si la tensión arterial sistólica cambia con la edad, se considera una muestra de 46 varones de edades entre 16 y 64 años y se mide su tensión arterial sistólica entre otras cosas. Se definen las siguientes variables:

X = “Edad”

Y = “Tensión arterial sistólica”

En este caso de tener dos variables cuantitativas el gráfico adecuado se llama *Diagrama de Dispersión* en el mismo se muestran los pares de datos para cada persona. Figura 9.

Figura 9: Tensión arterial sistólica por edad de 46 personas de sexo masculino.



En este diagrama se observa que la tensión arterial sistólica aumenta en promedio con la edad. Para estudiar con mayor precisión esta relación se puede ajustar una relación lineal, es decir una recta.

SERIES TEMPORALES

Serie temporal: Una serie temporal es un conjunto de observaciones de una variable en un cierto periodo de tiempo, en donde cada dato tiene su ubicación en el tiempo.

En general se de hacer predicciones sobre esa variable, teniendo en cuenta sus características históricas o del pasado.

Ejemplos de serie temporales:

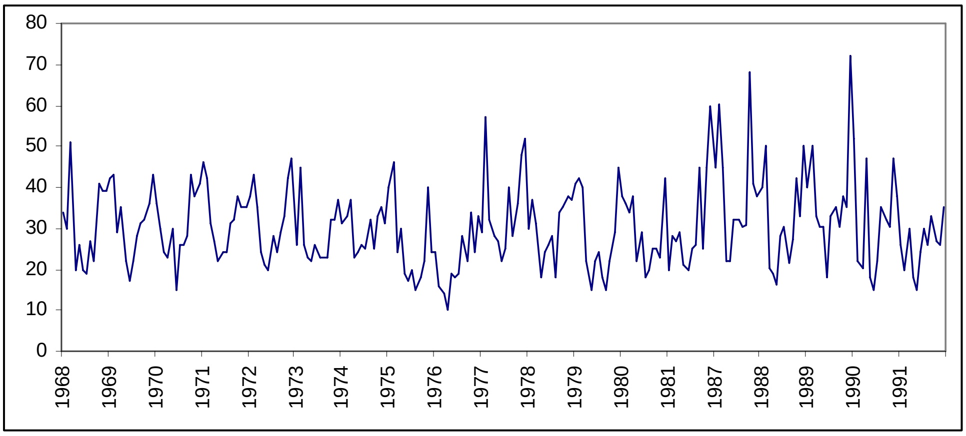
• Velocidades máximas mensuales del viento en una estación meteorológica.

• Cantidad de lluvia caída al día durante el último trimestre.

• Volumen de ventas mensuales durante los últimos 3 años.

Los datos tienen la forma y_t donde y es el valor de la variable y t es la ubicación en el tiempo.

Figura 10: Velocidades máximas mensuales del viento en Tucumán en el periodo 1968 – 1991.



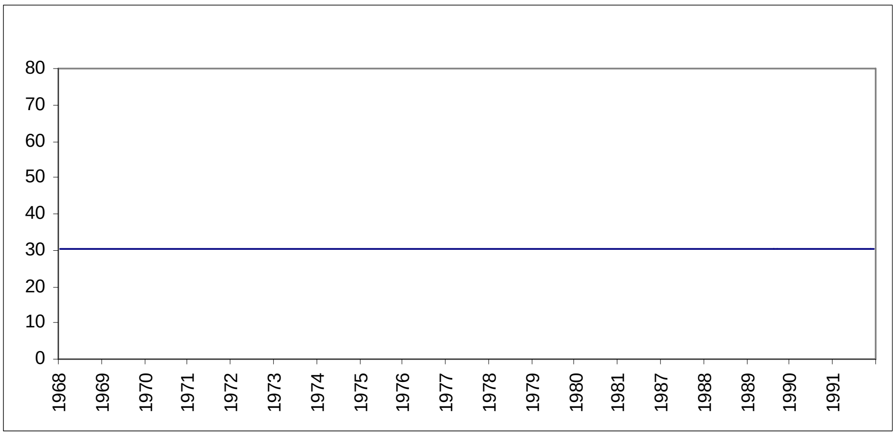
Componentes de una serie temporal:

Un modo adecuado de visualizar una serie temporal es considerar que la misma es el resultado de la suma (o combinación) de varias componentes: Tendencia, Ciclos, Estacionalidad y Componente residual.

La tendencia: Es una componente de la serie temporal que refleja su evolución a largo plazo. Puede ser de naturaleza estacionaria o constante, de naturaleza lineal, de naturaleza parabólica, de naturaleza exponencial, etc.

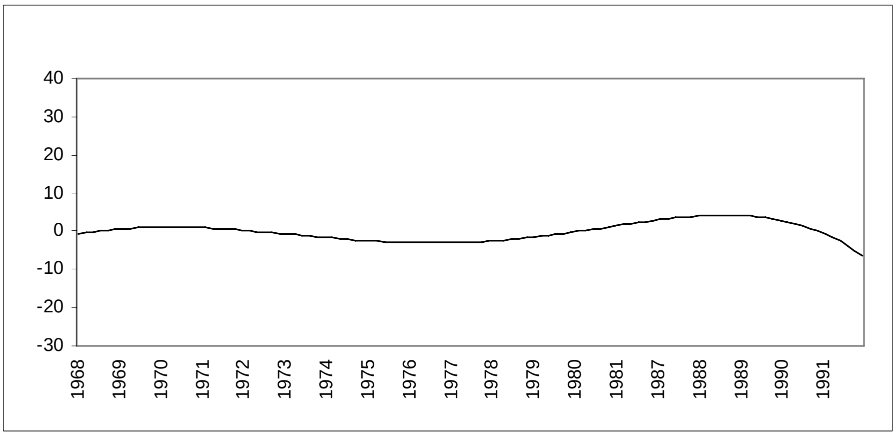
En nuestro ejemplo la tendencia es constante por lo que se puede estimar simplemente con la media de los datos.

Figura 11: Componente de tendencia de las velocidades máximas mensuales del viento en Tucumán en el periodo 1968 – 1991.



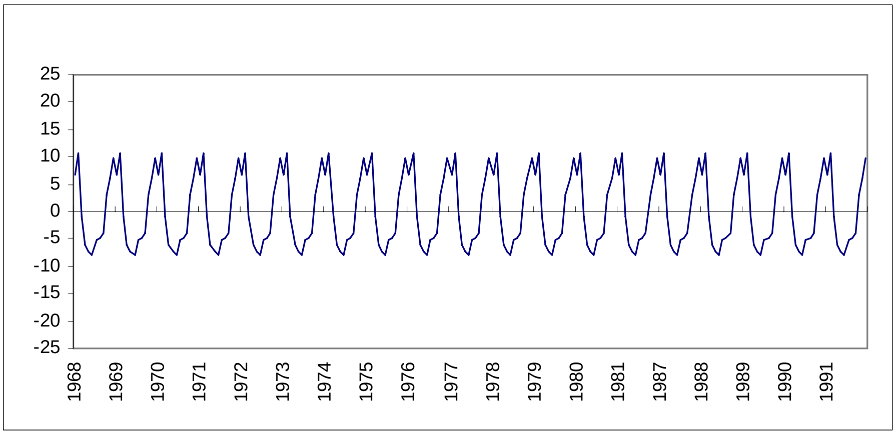
Los Ciclos: Es una componente de la serie que recoge oscilaciones periódicas de amplitud superior a un año. Estas oscilaciones periódicas por lo general no son regulares.

Figura 12: Componente cíclica de las velocidades máximas mensuales del viento en Tucumán en el periodo 1968 – 1991.



La estacionalidad: Es una componente de la serie que recoge oscilaciones que se producen alrededor de la tendencia, de forma repetitiva y en períodos iguales o inferiores a un año.

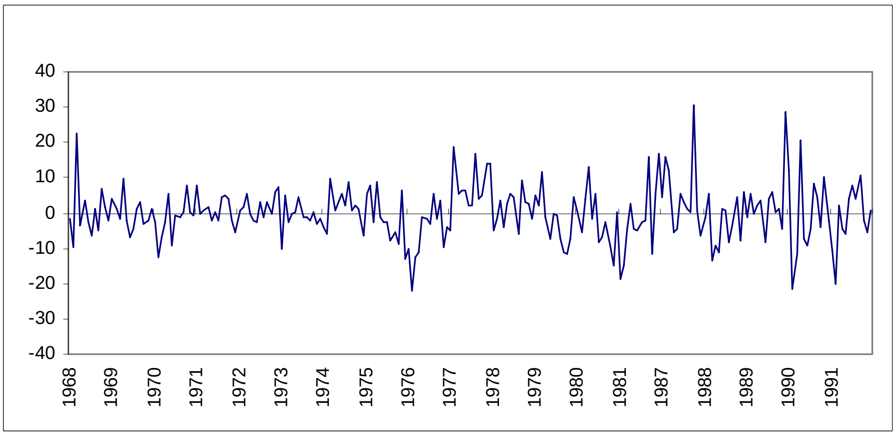
Figura 13: Componente estacional de las velocidades máximas mensuales del viento en Tucumán en el periodo 1968 – 1991.



Su nombre proviene de las estaciones climatológicas: primavera, verano, otoño e invierno

La componente residual: Es una componente de la serie que recoge movimientos provocados por factores imprevisibles. También reciben el nombre de variaciones irregulares o erráticas.

Figura 14: Componente residual de las velocidades máximas mensuales del viento en Tucumán en el periodo 1968 – 1991.



Dirección de Estadística de la Provincia de Tucumán – Noviembre de 2009.