## Tema 2. Descrición Estatística de Datos Unidimensionais.

***Bibliografia***

1. Baró Llinás, B.B. "Cálculo de Probabilidades". Ed. Parramón. 1987.
2. Berenson e Levine. "Estadistica Basica en Administracion. Conceptos y aplicaciones" 1996.
3. Chao, L.L. "Estadistica para las ciencias administrativas". 1993
4. Cuadras, C. M.. "Problemas de probabilidades y estadística". Vols:1, 2. Ed. P.P.U. 1990-1991.
5. Freund, J.E. e outros "Estadistica para la administracion. Con enfoque moderno.5a/Ed". 1990
6. Hildebrand e Ott "Estadistica aplicada a la administracion y a la economia" 1997
7. Levin e Rubin "Estadistica para Administradores, 6ª edic.1996"
8. Peña Sánchez de Rivera, D. "Estadística. Modelos y Métodos. Fundamentos". Ed. Alianza Universidad. 1991.

### 2.0. Introducion.

O obxectivo de este tema é a caracterizacion de colectivos cuxos "indivíduos" tenen cualidades diferentes. A información que proporcionan os individuos debe ser ordenada e agrupada, isto determina unha primeira sintese ou reducion de informacion.

É de grande importancia a recollida de datos na determinacion da cualidade de todo o proceso estatístico que segue. A recopilacion dos datos é a primeira fase de todo estudo estatístico, e quizá a máis delicada, porque vai determinar a cualidade de todo o traballo. Os resultados serán falsos se os datos non representan ben á realidade, ainda que fagamos correctamente todo o proceso estatístico: ordenación, análise e interpretacion.

Os datos poden obter-se de duas formas:

a) Datos indirectos ou publicados
b) Datos directos.

Tamén admitimos a seguinte clasificacion.

a) Datos observados ou experimentais.
b) Datos de enquisas.

a) Datos Indirectos ou Publicados.

Moitas veces podemos ter ao noso dispor informacion xa recopilada por outros axentes, que podemos utilizar, como poden ser: Censos Oficiais, Anuários e Informes Ministeriais, Datos (económicos) de Sociedades Financeiras, Rexistos Administrativos e Bases de Datos en xeral. Cumpre consultar previamente a dcoumentacion existente ante posibilidade de estudos feitos sobre o mesmo tema de interese. Os datos recollidos por terceiros poden estar elaborados ou sen elaborar.

Sempre debemos mencionar a fonte de datos utilizada e a data de referéncia. A fonte de datos garante sobre a sua cualidade, a data de referéncia sobre a sua actualidade.

As Fontes de Datos ou Bancos de Datos son abondosas na actualidade xa que os médios de obtención dos mesmos son cada vez máis avanzados. Na Xestión e Administración Publicas existen organismos oficiais que teñen por mision recopilar e tratar a informacion relevante para facilitar o desenvolvemento dos entes públicos diversos. Eurostat, I.N.E., I.G.E son, entre outros, organismos que actúan dependendo dos governos comunitário, nacional e autónomico respeitivamente.

**Exercício**. Acceder ao servidor web do Instituto Nacional de Estatística, I.N.E. e navegar polas suas páxinas accedendo a outros servidores nacionais ou estranxeiros.

b) Datos Directos son os que recopilamos por nos mesmos e dos que somos responsables. Fundamentalmente se utilizan as enquisas para a sua obtención ou ben a experimentación nas C.C. experimentais. Debemos especificar como foron obtidos, en qué circunstáncias, e en que momento.

*Elaboracion de enquisas*. Cuando desexamos saber algo acerca da opinión, gustos, aspirancions ou simplesmente características da xente, o procedemento que pode seguir-se é preguntando á poboacion:

- Cuantos fillos ten?
- Qué partido político pensa vostede que gañará as vindeiras eleccions?

Os tipos de enquisa poden ser:

i) Entrevistas persoais.

ii) Cuestionários a cumplimentar (correo, etc.).

iii) Entrevistas telefónicas.

Todas as enquisas requiren de un cuestionário ou lista de preguntas previamente elaborado con o obxectivo de que nos proporcione a información desexada. A diferéncia máis importante entre eles estriba na maior ou menor presenza do entrevistador.

Vantaxes e desvantaxes dos tipos de entrevistas:

| Tipo de entrevistas | Vantaxes | Inconvenientes |
|---|---|---|
| 1. Entrevistas persoais | -O entrevistador pode orientar, informar, etc. | -Poden intimidar ou influir nas respostas |
| 2. Cuestionários a cumplimentar (correo, etc.) | -Máis baratos.<br>-Non hai influéncia do entrevistador | -Unha parte importante poden non remitir os cuestionários.<br>-Non podera ser orientado. |
| 3. Entrevistas telefónicas. | -Rápidas | -Deben ser breves. |

Parece evidente que en cada caso particular, teremos que optar por unha ou outra modalidade e mesmo combinalas. A elaboracion do cuestionário deberá perseguir o obxectivo de conseguir esencialmente a informaicon lque necesitamos, (os datos). A lista de preguntas a realizar constitue o cuestionário. O que nos interesa debe estar o suficientemente ben definido, como para que ao chegar a este ponto, non nos quede nengunha duda da informacion que necesitamos. Os tipos de preguntas poden ser:

Fechadas

- De respostas incompatibles: Cre que España debe permanecer na OTAN?

  1. □ SI 2. □ NON.

- De resposta múltiple: Indique por orde de preferéncia que tipo de establecimentos utiliza para as suas compras de alimentación.

1. □ Venda Ambulante 2. □ Tenda tradicional

3. □ Supermercado 4. □ Grandes superfícies

## 2.1. Poboación, mostra, caracteres, variables estatísticas.

Sempre que se realiza un estudo estatístico no que debemos obter informacion, diriximos este esforzo a un "conxunto" ou colectivo. Entenderemos por poboacion P cualquer colectivo do que nos interese obter información de algunha ou algunhas características propias de dita poboacion. Suporemos que a poboacion está constituida por unidades definidas de xeito que sexan perfectamente identificables. Isto é vital na definicion da poboacion porque ás veces non está claro cal é a unidade básica para obtermos os datos.

Exemplo:

a) Poboacion 1: constituida por os alunos matriculados nunha materia concreta.
b) Poboacion 2: constituida polos "vehículos de tracción motor".
c) Poboacion 3: clientes que acoden a un determinado estabelecemento.
d) Poboacion 4: constituida por cidadáns que acoden a unha Administración Pública.
e) Poboacion 5: constituida por empresas de un determinado sector económico como actividade principal.

Como podemos observar na vida real aparecen problemas de accesibilidade ás unidades que constituen unha poboacion. Ben porque a sua caracterizacion como unidade da poboacion sexa difícil ben porque conceptualmente non é doado o acceso á información que nos pode facilitar a unidade en cuestion. Pensemos nas empresas que pertencen a un sector económico como actividade principal. Nun momento dado pode mudar a situación de pertenza ao sector económico pola própria dinámica da empresa, porque diversifique a sua producion cara a outros sectores, etc. Ou os alunos matriculados nunha matéria pero non acoden moito a clase (traballan) e por tanto, no é posible localizalos como unidades informantes, tan doadamente.

Tamén pode pasar que a poboacion estexa constituida por moitas unidades e por tanto inviable cualquer estudo que pretenda utilizar todas e cada unha de elas. Este tipo de estudo recibe un nome especial chamado Censo. Os censos son estudos exhaustivos de unha poboacion concreta. Hai censos de poboacion a respeito de diversas características. O realizar un censo non prevén de falta de información de certas unidades inaccesibles e ao final nunca se estudan realmente todas as unidades senon ás que se ten acceso. O conxunto de unidades da poboacion ás que se ten acceso informativo desde o ponto de vista da característica estudada constitúe o que se chama Marco. O marco é o conxunto de unidades accesibles da poboacion xunto coa informacion necesária que perfila as suas caraterísticas de acceso para poder obtermos informacion. Un bó marco é o ponto de partida para a selección de certas unidades nas que estudaremos as caractarísticas de interese, nun subconxunto dado de unidades da poboacion que recebe o nome de Mostra.

Consideramos a seguinte notacion:

Utilizaremos **N** para designar o número de unidades do Marco: $\{u_1, u_2, ..., u_N\}$.

Utilizaremos **n** para designar o número de unidades da Mostra: $\{u_1, u_2, ..., u_n\}$.

Utilizaremos $\mathbf{X_1, X_2,...}$ ou **X, Y, .....** para designar as distintas características das que estamos interesados, nunha poboacion dada.

**Exemplo**:

X = Actos Administrativos realizados no negociado de alunos da Facultade de Ciencias Sociais nun período de tempo dado (por exemplo, unha semana)

A característica X poderá tomar distintos valores: $x_1$ = validación, $x_2$ = matriculacion, $x_3$ = certificación académica, etc...

## 2.2. Distribución de Frecuencias. Representación Gráfica.

Seguindo con o exemplo derradeiro, estaremos interesados en saber sobre a característica X que desde agora chamaremos Variable Estatística. O seu valor non é sempre o mesmo, hai distintos e posibles actos administrativos. Por tanto, será de interese saber os posibles valores que pode tomar a variable xunto con o número de veces que aparece cada un dos valores. Este conxunto de información é o que se chama distribución de frecuéncias:

Valores da variable X = $x_i$ e frecuencia con que aparecen cada un dos valores $n_i$, i = 1, ..., k (k posibles valores distintos).

Adoita expresar-se en forma tabular.

| X | $x_1$ = validacion | $x_2$ = matriculacion | $x_3$ = certificación académica | $x_4$ = ..... |
|---|---|---|---|---|
| n frecuéncia | $n_1 = 3$ | $n_2 = 60$ | $n_3 = 1$ | $n_4$ = .... |

$n_i$ = número de veces que a variable X presenta o valor X = $x_i$ recebe o nome de FRECUENCIA ABSOLUTA.

Tipos de variables estatísticas. Segundo a información subministrada por unha unidade informativa cabe distinguirmos distintos tipos de variables:

a) Variables cualitativas, nominais, ou de atributos: non toman valores numéricos e descreben cualidades. Por exemplo, clasificarmos unha peza como aceitable ou defectuosa.
b) Variables cuantitativas discretas: toman únicamente valores enteros, corresponden en xeral ao contarmos o número de veces que ocorre un suceso. Por exemplo, número de compras de un produto nun mes.
c) Variables cuantitativas contínuas, ou de intervalo: toman valores nun intervalo, corresponden a medir magnitudes contínuas. Por exemplo,

tempo que tarda un administrado en recibir unha resolución a un recurso presentado.

A tabulacion como Distribución de Frecuéncias á que fixemos alusion con anterioridade é válida para todos os tipos de variables, pero o posterior tratamento da información que constitúe a taboa é diferente segundo teñamos un tipo ou outro, de variable estatística.

Este tipo de variables no que o orde en que se observan é irrelevante chaman-se **variables stock**. Cuando o orde dos datos sí ten importancia entón a variable recebe o nome de **série.** A maioría dos datos de tipo económico son de tipo série. Porque interesa predecir valores futuros da série a partir de valores observados no pasado. As séries reciben un tratamento diferente ás variables tipo stock que son as que trataremos en este tema.

Outras definicions nunha táboa de frecuéncias

Consideramos agora as seguintes definicions que complementan a presentación da distribución de frecuéncias en forma tabular:

**Frecuencia relativa do valor $x_i$**: $f_i = \frac{n_i}{n}$ sendo n o número total de datos $n = \sum_{i=1}^{k} n_i$

A frecuencia relativa pode vir dada en tanto por un ou en tanto por cento.

**Frecuencia absoluta acumulada do valor $x_h$**: $N_h = \sum_{i=1}^{h} n_i$

**Frecuencia relativa acumulada do valor $x_h$**: $F_h = \sum_{i=1}^{h} f_i$

As variables cualitativas ou cuantitativas discretas resumen na distribución de frecuéncias a súa información básica. Sen embargo, as variables contínuas utilizan previamente unha clasificacion dos seus valores en intervalos ou rangos de valores. E o que se chama **agrupación dos datos**. Porque a non agrupación de datos con moitos valores diferentes non nos permitiria visualizar a situación a respeito da variable. Este tipo de agrupamento tamén se utiliza con variables discretas con "moitos" valores diferentes. O proceso é como segue:

a) Arredondar os datos a dous ou, con moito tres cifras significativas elixindo as unidades para que cada observacion conteña dous ou tres díxitos, sen coma decimal.

b) Decidir o número k de clases a considerar. Este número debe ser entre 5 e 20. Unha regra frecuentemente utilizada é tomar r igual ao inteiro máis perto á raiz cuadrada de n, sendo n o número de datos, pero conven probarmos con distinto número de clases e escoller aquel que proporcione unha descricion máis clara.

c) Seleccionar os límites de clase que definen os intervalos, de xeito que as clases sexan da mesma lonxitude e cada observacion se clasifique sen ambigüidades nunha soa clase. Por exemplo a **clase i-ésima $[L_{i-1}, L_i)$**, onde o corchete indica que debemos considerar dentro da clase os datos igual ao extremo $L_{i-1}$, e o paréntese indica que non debemos considerar dentro de esta clase aos datos igual ao extremo $L_i$. O valor médio da clase que representará a todos os valores da mesma chama-se **Marca de Clase: $x_i$**. A lonxitude da clase chama-se **Amplitude da clase: $a_i$.**

d) Contarmos o número de observacions en cada clase, que chamaremos a frecuéncia de cada clase, e obter a frecuéncia relativa de cada clase dividindo aquela polo total de datos.

Despois do proceso inicial de tabulacion na táboa de frecuéncias procederemos a resumir a informacion da mesma mediante gráficos e medidas analíticas de resumo dos datos.

Exemplo:

## Distribucions de frecuencia agrupada

Supoñamos que medimos a estatura dos habitantes de unha vivenda e obtemos os seguientes resultados (cm):

| Habitante | Estatura | Habitante | Estatura | Habitante | Estatura |
|---|---|---|---|---|---|
| Habitante 1 | 1,15 | Habitante 11 | 1,53 | Habitante 21 | 1,21 |
| Habitante 2 | 1,48 | Habitante 12 | 1,16 | Habitante 22 | 1,59 |
| Habitante 3 | 1,57 | Habitante 13 | 1,60 | Habitante 23 | 1,86 |
| Habitante 4 | 1,71 | Habitante 14 | 1,81 | Habitante 24 | 1,52 |

| | | | | | |
|---|---|---|---|---|---|
| Habitante 5 | 1,92 | Habitante 15 | 1,98 | Habitante 25 | 1,48 |
| Habitante 6 | 1,39 | Habitante 16 | 1,20 | Habitante 26 | 1,37 |
| Habitante 7 | 1,40 | Habitante 17 | 1,42 | Habitante 27 | 1,16 |
| Habitante 8 | 1,64 | Habitante 18 | 1,45 | Habitante 28 | 1,73 |
| Habitante 9 | 1,77 | Habitante 19 | 1,20 | Habitante 29 | 1,62 |
| Habitante 10 | 1,49 | Habitante 20 | 1,98 | Habitante 30 | 1,01 |

Se presentáramos esta información nunha taboa de frecuencia obteriamos unha taboa de 30 liñas (unha para cada valor), cada un de eles con unha frecuencia absoluta de 1 e con unha frecuencia relativa do 3,3%. Esta taboa nos aportaría escasa información

No seu lugar, preferimos agrupar os datos por intervalos, con o que a información queda máis resumida (se perde, por tanto, algo de información), pero é máis manexable e informativa:

| Estatura | Frecuencias absolutas | | Frecuencias relativas | |
|---|---|---|---|---|
| Cm | Simple | Acumulada | Simple | Acumulada |
| | | | | |
| 1,01 - 1,10 | 1 | 1 | 3,3% | 3,3% |
| 1,11 - 1,20 | 3 | 4 | 10,0% | 13,3% |
| 1,21 - 1,30 | 3 | 7 | 10,0% | 23,3% |
| 1,31 - 1,40 | 2 | 9 | 6,6% | 30,0% |
| 1,41 - 1,50 | 6 | 15 | 20,0% | 50,0% |
| 1,51 - 1,60 | 4 | 19 | 13,3% | 63,3% |
| 1,61 - 1,70 | 3 | 22 | 10,0% | 73,3% |
| 1,71 - 1,80 | 3 | 25 | 10,0% | 83,3% |
| 1,81 - 1,90 | 2 | 27 | 6,6% | 90,0% |
| 1,91 - 2,00 | 3 | 30 | 10,0% | 100,0% |

O número de tramos nos que se agrupa a información é unha decisión que debe tomar o analista: a regra é que mentres máis tramos se utilicen menos información se perde, pero pode que menos representativa e informativa sexa a taboa. Podemos observar que na táboa de riba non se segue a notacion [¿, ¿) que se indicaba con anterioridade pero este tipo de táboas non son as que predominan na Bibliografía en xeral. O máis habitual é precisamente a notacion corchete-paréntese.

Un primeiro paso despois da tabulacion é a representacion gráfica dunha distribucion de frecuéncias. Os procedementos e gráficos son diversos, pero todos pretenden presentar a informacion de forma visual de xeito que se capte o esencial da variable estatística. Dependendo do tipo de variable (cuantitativa ou cualitativa) utilizaremos tipos de gráficos específicos.

Diagramas de Tronco-e-Follas

Constitúen un procedemento semi-gráfico de presentarmos a informacion para variables cuantitativas, que é especialmente útil cuando o número total de datos é pequeño (menor que 50). Os principios para construilo son:

a) Arredondar os datos a dous ou tres cifras significativas, expresando-os en unidades convenientes.
b) Dispoñelos nunha tabela con duas colunas separadas por unha liña como segue:
    1. Para datos con dous díxitos, escrever á esquerda da liña os díxitos das decenas –que forma o tronco- e á direita as unidades, que serán as follas. Por exemplo, 87 escreve-se 8 | 7.
    2. Para datos con tres díxitos o tronco estará formado polos díxitos das centenas e decenas, que se escreverán á esquerda, separados das unidades. Por exemplo 127 será 12 | 7.
c) Cada tronco define unha clase, escrevendo-se só unha vez. O número de follas representa a frecuéncia de dita clase.

Exemplo.

Supoñamos os seguintes datos recollidos en cm:

11.357, 12.542, 11.384, 12.431, 14.212, 15.213, 13.300, 11.300, 17.206, 12.710, 13.455, 16.143, 12.162, 12.721, 13.420, 14.698.

Os datos arredondados expresados en mm:

114, 125, 114, 124, 142, 152, 133, 113, 172, 127, 135, 161, 122, 127, 134, 147.

Diagrama de tronco e follas, datos en mm:

| | |
|---|---|
| 11 | 443 |
| 12 | 54727 |
| 13 | 354 |
| 14 | 27 |
| 15 | 2 |

| | |
|---|---|
| 16 | 1 |
| 17 | 2 |
| decenas | Unidades |

Diagrama de Pareto.

Este diagrama se utiliza para representar datos cualitativos e construe-se como segue:

1) Se ordenan as categorias ou clases pola súa frecuéncia relativa de aparicion.
2) Cada categoria representa-se por un rectángulo cuxa altura é a sua frecuéncia relativa.

No gráfico de abaixo observamos a variable Nivel educativo (anos de escolarizacion) de unha determinada poboacion. Fonte de datos "Datos de empleados" do SPSS

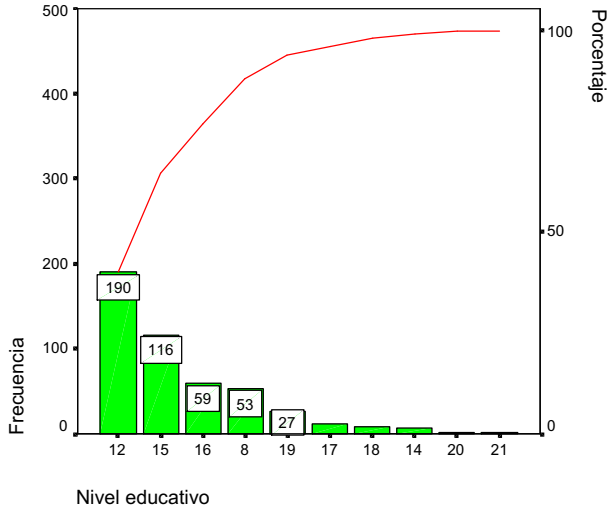



Diagrama de Barras.

Para datos de variables discretas, e en xeral para distribucions de frecuéncias de datos sen agrupar, se utiliza o diagrama de barras. Este diagrama representa os valores da variable no eixo de abscisas levantando en cada ponto unha barra de altura igual á frecuencia absoluta ou relativa (en tanto por un o e tanto por cento).

Exemplo:

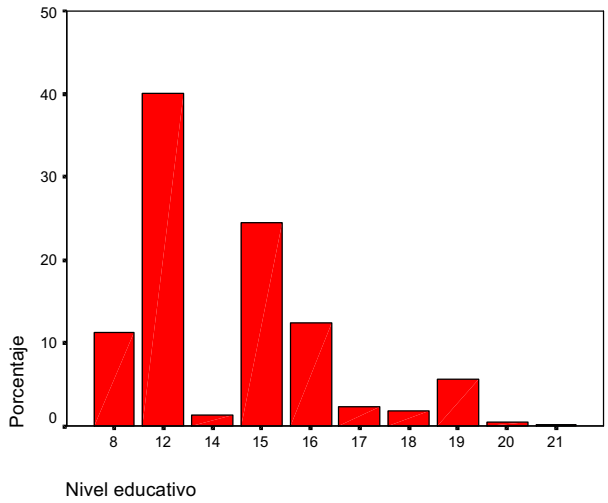
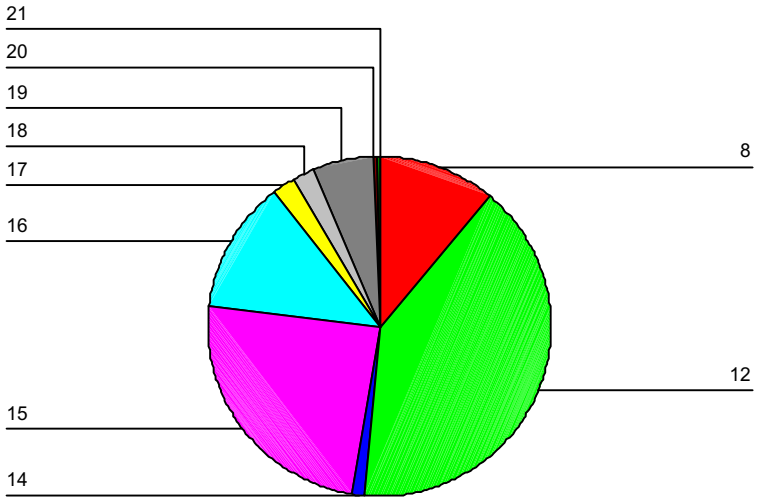


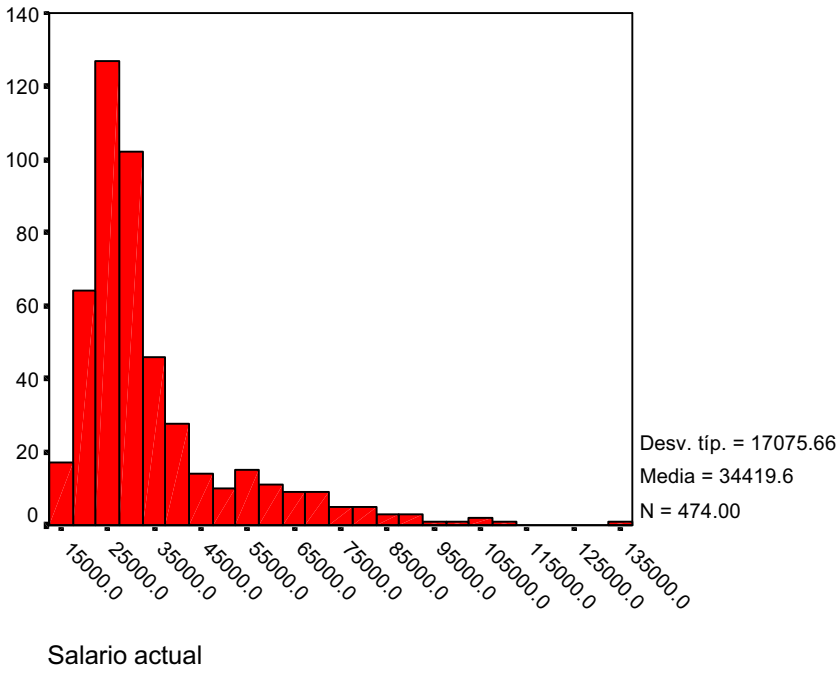

Diagrama de sectores ou tarta.




Histograma:

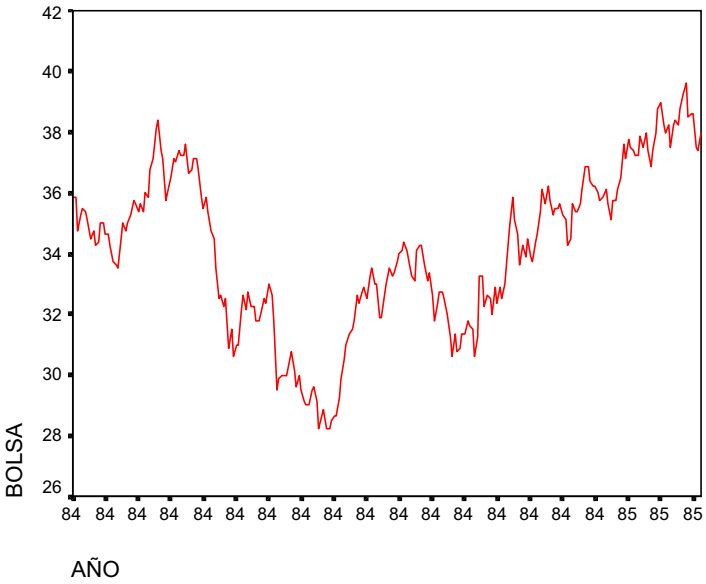
A representacion gráfica máis frecuente para datos agrupados é o histograma. Un histograma é un conxunto de rectángulos, cada un dos cuais representa un intervalo de agrupación ou clase. As suas bases son iguais á amplitude do intervalo, e as alturas se determinan de xeito que a sua área sexa proporcional á frecuéncia de cada clase.

Exemplo: Variable Salário Actual da fonte de datos xa indicada con anterioridade.




Gáfico Temporal. Representa a evolución duna variable que se observa temporalmente.

Un exemplo é o gráfico onde se representan os valores dos índices da Bolsa




## 2.3. Medidas de Posición, Dispersión, Forma e Concentración.

Medidas de Posición.

As medidas de posición nos facilitan información sobre a serie de datos que estamos a analisar. Estas medidas permiten coñecer diversas características de esta serie de datos.

As **medidas de posición** son de dous tipos:

**a) Medidas de posición central**: informan sobre os valores médios da serie de datos.

**b) Medidas de posición non centrales**: informan de como se distribúe o resto dos valores da serie.

## a) Medidas de posición central

As principais medidas de posición central son as seguintes:

**1.- Media**: é o valor medio ponderado da serie de datos. Se poden calcular diversos tipos de media, sendo as máis utilizadas:

**a) Media aritmética:** se calcula multiplicando cada valor polo número de veces que se repite. A suma de todos estes produtos se divide polo total de datos da mostra:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n}$$ para datos agrupados

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$ para datos sen agrupar

**b) Media xeométrica:** se eleva cada valor ao número de veces que se ten repetido. Se multiplican todo estes resultados e ao produto fiinal se lle calcula a raíz "n" (sendo "n" o total de datos da mostra).

$$x_G = (\prod_{i=1}^{k} x_i^{n_i})^{1/n}$$

Segundo o tipo de datos que se analice será máis apropriado utilizar a media aritmética ou a media xeométrica.

A media Xeométrica adoita utilizar-se en series de datos como tipos de interese anuais, inflación, etc., onde o valor de cada ano ten un efecto multiplicativo sobre os dos anos derradeiros. En todo caso, a média aritmética é a medida de posición central máis utilizada.

O máis positivo da média é que no seu cálculo se utilizan todos os valores da serie, polo que non se perde nenguna información.

Sen embargo, presenta o problema de que o seu valor (tanto eo caso da media aritmética como xeométrica) se pode ver moi influido por valores extremos, que se aparten en exceso do resto da serie. Estes valores anómalos poderían condicionar en grande medida o valor da media, perdendo ésta representatividade.

**2.- Mediana**: é o valor da serie de datos que se sitúa xustamente no centro da mostra (un 50% de valores son inferiores e outro 50% son superiores).
Non presenta o problema de estar influido polos valores extremos, pero en troca non utiliza no seu cálculo toda a información da serie de datos (non pondera cada valor polo número de veces que se ten repetido).

Cálculo da Mediana. Consideramos dous casos:

a) Para datos sen agrupar poden darse, tamén, duas situacións.
   - ⇒ Hai un número impar de datos, en este caso a mediana coincide con o dato que ocupa o lugar n/2 + 1
   - ⇒ Hai un número par de datos, en este caso a mediana é o promedio dos dous datos centrais.

b) Para datos agrupados en intervalos se utiliza unha fórmula de aproximación.
   1. O primeiro paso é localizar o intervalo onde estará a Mediana que é o intervalo correspondente ao primeiro valor de $N_i$ que supera n/2.
      Supoñamos que este intevalo é $[L_{i-1},L_i)$. Vamos a por-lle o nome de Intervalo Mediano
   2. O segundo paso é aplicar a fórmula seguinte:

$$M_e = L_{i-1} + \frac{n/2 - N_{i-1}}{n_i} a_i$$

Sendo:
$n_i$ a frecuéncia absoluta do intervalo mediano.
$a_i$ a amplitude do intervalo mediano
$N_{i-1}$ a frecuéncia absoluta acumulada do intervalo anterior ao intervalo mediano.
$L_{i-1}$ o extremo inferior do intervalo mediano.

**3.- Moda**: é o valor que máis se repite na mostra.

**Exemplo:** vamos a utilizar a taboa de distribución de frecuencias con os datos da estatura dos alumnos seguinte:

| Variable | Frecuencias absolutas | | Frecuencias relativas | |
|---|---|---|---|---|
| (Marca de clase) | Simple | Acumulada | Simple | Acumulada |
| | | | | |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |

| | | | | |
|---|---|---|---|---|
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

Vamos calcular os valores das distintas posicions centrais:

**1.- Media aritmética:**

$$Xm = \frac{(1,20*1) + (1,21*4) + (1,22 * 4) + (1,23 * 2) + ..... + (1,29 * 3) + (1,30 * 3)}{30}$$

Logo:

Xm = 1,253

Por tanto, a estatura media de este grupo de alunos é de 1,253 cm.

**2.- Media xeométrica:**

X = ((1,20^ 1) * (1,21^4) * (1,22^ 4) * .....* (1,29^3)* (1,30^3)) ^ (1/30)

Logo:

Xm = 1,253

En este exemplo a media aritmética e a media xeométrica coinciden, pero non ten sempre por qué ser así.

**3.- Mediana:**

A mediana de esta mostra é 1,26 cm, xa que por debaixo está o 50% dos valores e por riba o outro 50%. Isto se pode ver ao analisarmos a coluna de frecuencias relativas acumuladas.

En este exemplo, como o valor 1,26 se repite en 3 ocasions, a mediana se situaría exactamente entre o primeiro e o segundo valor de este grupo, xa que entre estes dous valores se encontra a división entre o 50% inferior e o 50% superior.

**4.- Moda:**

Hai 3 valores que se repiten en 4 ocasions: o 1,21, o 1,22 e o 1,28, por tanto esta serie conta con 3 modas.

## Medidas de posición non centrais

As medidas de posición non centrais permiten coñecer outros puntos característicos da distribución que non son os valores centrais. Entre outros indicadores, adoitan utilizar-se unha serie de valores que dividen a mostra en tramos iguais:

**Cuartís**: son 3 valores que distribuen a serie de datos, ordenada de xeito crecente ou decrecente, en cuatro tramos iguais, nos que cada un de eles concentra o 25% dos resultados.

**Decís**: son 9 valores que distribuen a serie de datos, ordenada de xeito crecente ou decrecente, en dez tramos iguais, nos que cada un de eles concentra o 10% dos resultados.

**Percentiles**: son 99 valores que distribuen a serie de datos, ordenada de forma crecente o decrecente, en cen tramos iguais, nos que cada un de eles concentra o 1% dos resultados.

**Exemplo**: Vamos calcular os cuartis da serie de datos referidos á estatura de un grupo de alumnos. Os decís e centís se calculan de igual xeito, ainda que faría falta distribucions con maior número de datos.

| Variable | Frecuencias absolutas | | Frecuencias relativas | |
|---|---|---|---|---|
| (Valor) | Simple | Acumulada | Simple | Acumulada |
| x | x | x | x | x |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

**1º cuartil:** es el valor 1,22 cm, ya que por debajo suya se situa el 25% de la frecuencia (tal como se puede ver en la columna de la frecuencia relativa acumulada).

**2º cuartil:** es el valor 1,26 cm, ya que entre este valor y el 1º cuartil se situa otro 25% de la frecuencia.

**3º cuartil:** es el valor 1,28 cm, ya que entre este valor y el 2º cuartil se sitúa otro 25% de la frecuencia. Además, por encima suya queda el restante 25% de la frecuencia.

**Calculo de un Percentil, Cuartil, Decil**. Vamos a ilustrar o proceso con o cálculo do percentil 30:

1. O primeiro paso é localizar o intervalo onde estará o percentil 30 $P_{30}$ que é o intervalo correspondente ao primeiro valor de $N_i$ que supera (nx30)/100 = 0.30n
   Supoñamos que este intevalo é $[L_{i-1},L_i)$. Vamos a por-lle o nome de *"Intervalo $P_{30}$"*
2. O segundo paso é aplicar a fórmula seguinte:

$$P_{30} = L_{i-1} + \frac{0.30n - N_{i-1}}{n_i} a_i$$

Sendo:
$n_i$ a frecuéncia absoluta do *Intervalo $P_{30}$*.
$a_i$ a amplitude do *Intervalo $P_{30}$*
$N_{i-1}$ a frecuéncia absoluta acumulada do intervalo anterior ao *Intervalo $P_{30}$*.
$L_{i-1}$ o extremo inferior do *Intervalo $P_{30}$*.

## ***Medidas de Dispersión.***

Estuda a distribución dos valores da variable, analisando se estes se atopan máis ou menos concentrados, ou máis ou menos dispersos.

Existen diversas **medidas de dispersión**, entre as máis utilizadas podemos destacar as seguintes:

**1.- Rango**: mide a amplitude dos valores da mostra e se calcula por diferencia entre o valor máis elevado e o valor máis baixo.

**2.- Varianza**: Mide a distancia existente entre os valores da variable e a media. Se calcula como sumatorio das difrencias ao cuadrado entre cada valor e a media, multiplicadas polo número de veces que se ten repetido cada valor. O sumatorio obtido se divide polo tamaño da mostra.

$$S^2_x = \frac{\Sigma (x_i - \bar{x}m)^2 * n_i}{n}$$

A varianza sempre será maior que cero. Mentres máis se aproxima a cero, máis concentrados están os valores da variable alrededor da media. Polo contrario, mentres maior sexa a varianza, máis dispersos están.

**3.- Desviación típica**: Se calcula como raíz cuadrada da varianza.

**4.- Coeficiente de varización (de Pearson)**: se calcula como cociente entre a desviación típica e a media.

**Exemplo:** vamos a utilizar a variable de datos da estatura dos alunos de unha clase e vamos calcular as suas medidas de dispersión.

| Variable | Frecuencias absolutas | | Frecuencias relativas | |
|---|---|---|---|---|
| (Marca de clase) | Simple | Acumulada | Simple | Acumulada |
| | | | | |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

**1.- Rango:** Diferencia entre o maior valor da mostra (1,30) e o menor valor (1,20). Logo o rango de esta mostra é 10 cm.

**2.- Varianza:** recordemos que a media de esta mostra é 1,253. Logo, aplicamos a fórmula:

$$S^2_x = \frac{((1{,}20-1{,}253)^2 * 1) + ((1{,}21-1{,}253)^2 * 4) + ((1{,}22-1{,}253)^2 * 4) + \ldots + ((1{,}30-1{,}253)^2 * 3)}{30}$$

Por tanto, a varianza é 0,0010

**3.- Desviación típica:** é a raíz cuadrada da varianza.

$$\sigma = (S^2_x)^{(1/2)}$$

Portanto:

$$\sigma = (0{,}010)^{(1/2)} = 0{,}0320$$

**4.- Coeficiente de variación de Pearson**: se calcula como cociente entre a desviación típica e a media da mostra. Sempre que a méda sexa diferente de cero.

Cv = 0,0320 / 1,253

portan

to,

Cv = 0,0255

O interese do coeficiente de variación é que ao ser unha percentaxe permite comparar o nivel de dispersión de duas mostras. Isto non acontece coa desvación típica, xa que ven expresada nas mesmas unidades que os datos da variable.

> Por exemplo, para compararmos o nivel de dispersión de unha serie de datos da altura de os alunos de unha clase e outra serie con o peso de ditos alunos, non se pode utilizar as desviacions típicas (unha ven expresada en cm e a outra en kg). En troca, os seus coeficientes de variación son ambos percentaxes, polo que sí se poden comparar.

As **medidas de forma** permiten coñecer que forma ten a curva que representa a serie de datos da mostra. En concreto, podemos estudar as seguintes características da curva:

**a) Concentración**: mide si os valores da variable están máis ou menos uniformemente repartidos ao longo da mostra.

**b) Asimetría**: mide se a curva ten unha forma simétrica, é dicir, se a respeito do centro da mesma (centro de simetría) os segmentos de curva que fican a direita e esquerda son similares.

**c) Curtose:** mide se os valores da distribución están máis ou menos concentrados ao redor dos valores medios da mostra.

**a) Concentración**

Para medir o nivel de concentración de unha distribucón de frecuencia adoita utilizar-se distintos indicadores, entre eles o **Indice de Gini**.
Este índice se calcula aplicando a seguinte fórmula:

$$IG = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

(i toma valores entre 1 y k-1)

Onde pi mide a percentaxe de individuos da mostra que presentan un valor igual ou inferior ao de xi. Isto é, a frecuéncia Relativa Acumulada en % $\mathbf{F_i}$

n1 + n2 + n3 + ... + ni

---------- x 100

n

Mentres que qi se calcula aplicando a seguinte fórmula:

$$qi = \frac{(X1*n1) + (X2*n2) + ... + (Xi*ni)}{(X1*n1) + (X2*n2) + ... + (Xn*nn)} \times 100$$

O **Indice Gini** (IG) pode tomar valores entre 0 e 1:

**IG = 0** : concentración mínima. La muestra está unifomemente repartida ao longo de todo o seu rango.

**IG = 1** : concentración máxima. Un só valor da mostra acumula o 100% dos resultados.

**Exemplo**: vamos calcular o Indice Gini de unha serie de datos cos salários dos empregados de unha empresa (millons pesetas).

| Sueldos | Empleados (Frecuencias absolutas) | | Frecuencias relativas | |
|---|---|---|---|---|
| (Millones) | Simple | Acumulada | Simple | Acumulada |
| Marca de clase | | | | |
| 3,5 | 10 | 10 | 25,0% | 25,0% |
| 4,5 | 12 | 22 | 30,0% | 55,0% |
| 6,0 | 8 | 30 | 20,0% | 75,0% |
| 8,0 | 5 | 35 | 12,5% | 87,5% |
| 10,0 | 3 | 38 | 7,5% | 95,0% |
| 15,0 | 1 | 39 | 2,5% | 97,5% |
| 20,0 | 1 | 40 | 2,5% | 100,0% |

Calculamos os valores que necesitamos para aplicar a fórmula del Indice de Gini:

| Xi | ni | □ ni | pi | Xi * ni | □ Xi * ni | qi | pi - qi |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| 3,5 | 10 | 10 | 25,0 | 35,0 | 35,0 | 13,6 | 10,83 |
| 4,5 | 12 | 22 | 55,0 | 54,0 | 89,0 | 34,6 | 18,97 |
| 6,0 | 8 | 30 | 75,0 | 48,0 | 147,0 | 57,2 | 19,53 |
| 8,0 | 5 | 35 | 87,5 | 40,0 | 187,0 | 72,8 | 15,84 |
| 10,0 | 3 | 38 | 95,0 | 30,0 | 217,0 | 84,4 | 11,19 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 15,0 | 1 | 39 | 97,5 | 15,0 | 232,0 | 90,3 | 7,62 |
| 25,0 | 1 | 40 | 100,0 | 25,0 | 257,0 | 100,0 | 0 |
| | | | | | | | |

Por tanto:

IG = 83,99 / 435,0 = 0,19

**Un Indice Gini de 0,19** indica que a mostra está bastante uniformemente repartida, quer dicir, o seu nivel de concentración non é excesivamente alto.

**Exemplo**: Agora vamos a analisar novamente a mostra anterior, pero considerando que hai máis persoal da empresa que cobra o salário máximo, o que leva maior concentración da renda nunhas poucas persoas.

| Salários | Empregados (Frecuencias absolutas) | | Frecuencias relativas | |
|---|---|---|---|---|
| (Millons) | Simple | Acumulada | Simple | Acumulada |
| | | | | |
| 3,5 | 10 | 10 | 25,0% | 25,0% |
| 4,5 | 10 | 20 | 25,0% | 50,0% |
| 6,0 | 8 | 28 | 20,0% | 70,0% |
| 8,0 | 5 | 33 | 12,5% | 82,5% |
| 10,0 | 3 | 36 | 7,5% | 90,0% |
| 15,0 | 0 | 36 | 0,0% | 90,0% |
| 20,0 | 4 | 40 | 10,0% | 100,0% |

En este caso obteríamos os seguintes datos:

| Xi | Ni | □ ni | pi | Xi * ni | □ Xi * ni | qi | pi - qi |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| 3,5 | 10 | 10 | 25,0 | 35 | 35 | 11,7 | 13,26 |
| 4,5 | 10 | 20 | 50,0 | 45 | 80 | 26,8 | 23,15 |
| 6,0 | 8 | 28 | 70,0 | 48 | 128 | 43,0 | 27,05 |
| 8,0 | 5 | 33 | 82,5 | 40 | 168 | 56,4 | 26,12 |
| 10,0 | 3 | 36 | 90,0 | 30 | 198 | 66,4 | 23,56 |
| 15,0 | 0 | 36 | 90,0 | 0 | 198 | 66,4 | 23,56 |
| 25,0 | 4 | 40 | 100,0 | 100 | 298 | 100,0 | 0,00 |
| | | | | | | | |
| □ pi (entre 1 y n-1) = | | | 407,5 | | □ (pi - qi) (entre 1 y n-1 | | 136,69 |

) =

O **Indice Gini** sería:

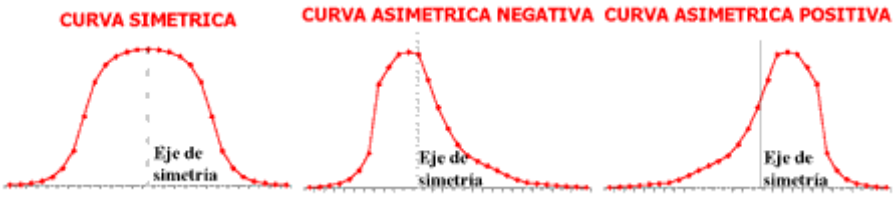
IG = 136,69 / 407,5 = 0,34

O Indice Gini se ten elevado considerablemente, reflectindo a maior concentración de rendas que temos comentado.

Curva de LORENTZ. Consiste en representarmos gráficamente os pares ($p_i$, $q_i$) nun eixo cartesiano XY.

## b) Asimetría

Temos comentado que o conceito de asimetría se refire a se a curva que forman os valores da serie presenta a mesma forma á esquerda e direita de un valor central (media aritmética)




Para medir o nivel de asimetría se utiliza o chamado **Coeficiente de Asimetría de Fisher**, que ven definido:

$$g_1 = \frac{(1/n) * \Sigma (x_i - xm)^3 * n_i}{((1/n) * \Sigma (x_i - xm)^2 * n_i))^{3/2}}$$

Os resultados poden ser os seguintes:

**g1 = 0** (distribución simétrica; existe a mesma concentración de valores á direita e á esquerda da media)

**g1 > 0** (distribución asimétrica positiva; existe maior concentración de valores á direita da media que á sua esquerda)

**g1 < 0** (distribución asimétrica negativa; existe maior concentración de valores á esquerda da media que á sua direita)

**Exemplo**: Vamos calcular o Coefiente de Asimetría de Fisher da serie de datos referidos á estatura de un grupo de alunos:

| Variable | Frecuencias absolutas | | Frecuencias relativas | |
|---|---|---|---|---|
| (Valor) | Simple | Acumulada | Simple | Acumulada |
| x | x | x | x | x |
| 1,20 | 1 | 1 | 3,3% | 3,3% |

| | | | | |
|---|---|---|---|---|
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

Recordemos que la media de esta muestra es 1,253

| □((xi - x)^3)*ni | □((xi - x)^2)*ni |
|---|---|
| | |
| 0,000110 | 0,030467 |

Logo:

$$g1 = \frac{(1/30) * 0,000110}{(1/30) * (0,030467)^{(3/2)}} = -0,1586$$

Por tanto o **Coeficiente de Fisher de Simetría** de esta mostra é -0,1586, o que significa que presenta unha distribución asimétrica negativa (se concentran máis valores á esquerda da media que á sua direita).

## ) Curtose

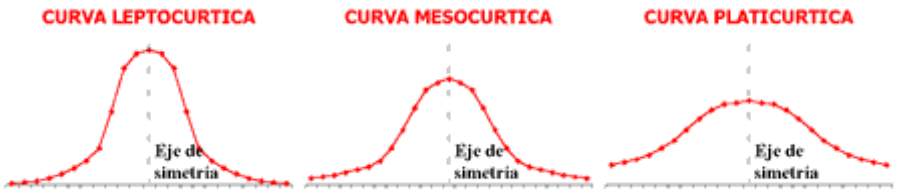
O **Coeficiente de Curtose** analisa o grau de concentración que presentan os valores ao redor da zona central da distribución.

Se definen 3 tipos de distribucions segundo o seu grau de curtose:

**Distribución mesocúrtica:** presenta un grau de concentración medio ao redor dos valores centrais da variable (o mesmo que presenta unha distribución normal).

**Distribución leptocúrtica**: presenta un elevado grau de concentración ao redor dos valores centrais da variable.

**Distribución platicúrtica:** presenta un reducido grau de concentración ao redor dos valores centrais da variable.




O **Coeficiente de Curtose** ven definido pola siguiente fórmula:

$$g_2 = \frac{(1/n) * \Sigma (x_i - xm)^4 * n_i}{((1/n) * \Sigma (x_i - xm)^2 * n_i))^2} - 3$$

Os resultados poden ser os seguintes:

**g2 = 0 (distribución mesocúrtica)**.

**g2 > 0 (distribución leptocúrtica)**.

**g2 < 0 (distribución platicúrtica)**.

**Exemplo**: Vamos calcular o Coefiente de Curtose da serie de datos referidos á estatura de un grupo de alumnos:

| Variable | Frecuencias absolutas | | Frecuencias relativas | |
|---|---|---|---|---|
| (Valor) | Simple | Acumulada | Simple | Acumulada |
| x | x | x | x | x |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

Recordemos que a media de esta mostra é 1,253

| □((xi - xm)^4)*ni | □((xi - xm)^2)*ni |
|---|---|
| | |
| 0,00004967 | 0,03046667 |

Logo:

$$g2 = \frac{(1/30) * 0{,}00004967}{((1/30) * (0{,}03046667))^2} - 3 = -1{,}39$$

Por tanto, o **Coeficiente de Curtose** de esta mostra é -1,39, o que significa que se trata de unha distribución platicúrtica, isto é, con unha reducida concentración ao redor dos valores centrais da distribución.

## 2.4. Momentos dunha distribución de frecuencias.

Os momentos duna distribución de frecuencias determinan completamente a mesma.

***Definición:*** Dada unha variable estatística X con valores $x_i$ e frecuencias relativas $f_i$ (i = 1 ... k) se define o momento de orde r a respeito de un valor c como:

$$\sum_{i=1}^{k} (x_i - c)^r f_i$$

Se o valor c = média da variable entón falamos de momento central de orde r.

Se o valor c = 0 entón falamos de momento respeito á orixe de orde r.

Exemplos:

a) O momento respeito á orixe de orde 1 é a média
b) O momento central de orde 2 é a varianza
c) As medidas de simetría e curtose se poden poñer en función dos momentos

Exercícios Tema 2.

**Exercício 1.** Clasificar as seguientes variables:
1. Preferencias políticas (esquerda, direita o centro).
2. Marcas de cervexa.
3. Velocidade en Km/h.
4. O peso en Kg.
5. Signo do zodiaco.
6. Nivel educativo (primário secundário, superior).
7. Anos de estudos completados.
8. Tipo de ensino (privado o público).
9. Número de empregados de unha empresa.
10. A temperatura de un enfermo en graos Celsius.
11. A clase social (baixa, media ou alta).

**Exercício 2**. Clasifique as variables que aparecen no seguiente cuestionário.
1. ¿Cál es su idade?
2. Estado civil:
   (a) Solteiro (b) Casado (c) Separado (d) Divorciado (e) Viuvo
3. ¿Canto tempo emprega para desplazar-se ao seu traballo?
4. Tamaño do seu municipio de residéncia:
   (a) Municipio pequeno (menos de 2.000 habitantes)
   (b) Municipio mediano (de 2.000 a 10.000 hab.)
   (c) Municipio grande (de 10.000 a 50.000 hab.)
   (d) Cidade pequena (de 50.000 a 100.000 hab.)
   (e) Cidade grande (más de 100.000 hab.)
5. ¿Está afiliado á seguridade social?

**Exercicio 3.** No seguinte conxunto de datos, se proporcionan os pesos (redondeados a libras) de nenos nacidos en certo intervalo de tempo:
4, 8, 4, 6, 8, 6, 7, 7, 7, 8, 10, 9, 7, 6, 10, 8, 5, 9, 6, 3, 7, 6, 4, 7, 6, 9, 7, 4, 7, 6, 8, 8, 9, 11, 8, 7, 10, 8, 5, 7, 7, 6, 5, 10, 8, 9, 7, 5, 6, 5.

1. Construir unha distribución de frecuéncia de estes pesos.
2. Encontrar as frecuéncias relativas.
3. Encontrar as frecuéncias acumuladas.
4. Encontrar as frecuéncias relativas acumuladas.
5. Dibuxar un gráfico de barras con os datos do apartado 1.
7. Calcular as medidas de tendéncia central.
8. Calcular as medidas de dispersión.
9. Encontrar el percentil 24. Qué indica?

**Ejercicio 4**. Como quedan afectadas a média e a varianza cando facemos o seguinte cambio de variable, Y=a + bX?

**Exercicio 5**. Un automóvil ten percorrido catro traxectos $t_1$, $t_2$, $t_3$, e $t_4$ con uns consumos médios respeitivos de 10, 8, 6 e 9 litros/100km. Determinar o consumo médio no global dos catro traxectos.
a) Se as distáncias percorridas en cada traxecto son 10, 20, 50 e 30 km., respeitivamente.

b) Se os litros de combustible consumidos en cada traxecto son 1, 2, 3 e 2 respeitivamente

**Exercício 6.** Ten-se observado o tamaño de 100 arquivos de texto armazenados no disco duro de un ordenador, anotando-se a seguinte distribucion de frecuéncias:

| Tamaño en Kb | 0-32 | 32-64 | 64-128 | 128-256 | 256-512 | 512-1024 |
|---|---|---|---|---|---|---|
| Nº de arquivos | 10 | 35 | 25 | 15 | 10 | 5 |

a) Completar a táboa de frecuéncias e construir o histograma.
b) Un operador informático quere repartir estes 100 arquivos en dous directórios, de xeito que o primeiro de eles contería aos de tamaño menor ou igual que a média aritmética e o outro aos restantes. Determinar entre qué valores estaria o número de arquivos que contería o primeiro directorio.
c) Cál é o valor no que debería basear-se para facer reparto de arquivos de xeito que nos dous directorios houbese igual número de eles?
d) Se desexamos borrar o 20% dos arquivos de menor tamaño, cal sería o tamaño do maior arquivo borrado?

**Exercício 7**. Vou pola estrada e observo que adianto tantos coches como me adiantan a min. Iso queres dicir que a miña velocidade é a velocidade .....

**Exercício 8**. Unha empresa agrícola ten 5 fincas, dedicadas á producion de trigo. As producions e rendementos obtidos son os seguintes:

| Finca | A | B | C | D | E |
|---|---|---|---|---|---|
| Producion (Qm) | 2500 | 3000 | 4000 | 6000 | 7000 |
| Rendemento (Qm/Ha) | 10 | 20 | 25 | 15 | 14 |

Calcular o rendemento médio por Ha para o conxunto das fincas utilizando a producion como criterio de ponderación.

**Exercício 9**. En tres empresas do sector transportes dedicadas á explotación de liñas regulares de viaxeiros se dan as seguintes cifras de producion total e produtividade média pro empregado. (Esta última magnitude se define como o cociente entre producion total e o número de empregados).

| Empresa | A | B | C |
|---|---|---|---|
| Producion (millons de viaxeiros-Km) | 100 | 150 | 300 |
| Produtividade média por empregado (en millons de viaxeiros.Km) | 0.50 | 0.60 | 0.75 |

Calcular a produtividade média por empregado para o conxunto das tres empresas:
a) Mediante unha média Harmónica
b) Mediante unha média aritmética.

**Exercício 10.** Nunha rexion económica ten-se observado, durante un período de dez anos, os seguintes incrementos anuais nos índices de salários e de prezos ao consumo

| Ano | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Incremento no índice de salários (%) | 7 | 10 | 8 | 14 | -4 | 3 | 4 | 8 | -1 | 10 |
| Incremento no índice de prezos ao consumo (%) | 6 | 9 | 4 | 10 | -2 | 3 | 4 | 9 | -2 | 6 |

a) Achar os incrementos médios anuais acumulativos, para o total do período, dos índices de salarios e de prezos ao consumo.
b) O incremento médio anual acumulativo, para o total do período, do índice de salários reais.
c) Os incrementos médios anuais acumulativos dos índices de salarios, de prezos ao consumo e de salarios reais, para o 1º e 2º quinquenios respeitivamente.

Nota. Se define unha taxa de variación porcentual ou incremento porcentual duna magnitude S en dous períodos h e h-1 como: $t_h = \frac{S_h - S_{h-1}}{S_{h-1}} x100$ (Ezequiel Uriel y Manuel Muñiz, Ed AC, pax. 64)

**Exercício 11.** Dada a seguinte Táboa:

| $X_i$ (Núm. De horas de estudo diarias) | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $n_i$ | 5 | 15 | 20 | 8 | 2 |

a) Achar as médias: (a) harmónica, (b) xeométrica, (c) aritmética.
b) Comprobar a relacion $H \leq X \leq A$

**Exercício 12.** Realizada unha enquisa por mostrase entre fumadores temos os seguintes resultados:

| Nº pitillos diarios | 4.5-9.5 | 9.5-14.5 | 14.5-19.5 | 19.5-24.5 | 24.5-29.5 |
|---|---|---|---|---|---|
| Nº individuos | 10 | 15 | 25 | 18 | 22 |

1. O número médio de pitillos fumados por indivíduo e dia.
2. A desviación típica.
3. O coeficiente de Variación.
4. A percentaxe de individuos que fuman entre 12 e 22 pitillos diarios ambos os dous incluidos. (supoñer a distribución uniforme dentro de cada intervalo)
5. O primeiro cuartil.
6. O valor máis frecuente da variable.

**Exercício 13.** De un sector económico teñen-se os seguintes datos sobre as empresas que o componen:

| Volume de vendas en millons de pts. | 50-100 | 100-200 | 200-500 | 500-1000 | 1000-2000 | 2000-5000 |
|---|---|---|---|---|---|---|
| Nº de empresas | 30 | 25 | 40 | 50 | 25 | 30 |

Calcular o índice de concentración do sector.