UNIVERSIDAD ORT
Uruguay
Facultad de Ingeniería
Bernard Wand - Polak

PROBABILIDAD Y ESTADÍSTICA APLICADA

NOTAS DE CLASE
DEL CURSO DE LA

Licenciatura en
Sistemas

FASCÍCULO 1

Prof. Orual Andina
Cátedra de Matemáticas

Año 2012

FASCÍCULO 1

UNIVERSIDAD ORT URUGUAY
LICENCIATURA EN SISTEMAS

ORUAL ANDINA

PROBABILIDAD Y ESTADÍSTICA APLICADA

NOTAS DE CLASE

PREFACIO

Esta publicación ha sido preparada para servir de apoyo al curso de Probabilidad y Estadística Aplicada, que se dicta en la Licenciatura en Sistemas de la Universidad ORT Uruguay. Se trata de un curso breve, limitado a cuatro horas docentes semanales durante un semestre lectivo. Por lo tanto, ha sido necesario condensar al máximo la presentación de los temas exigidos por el programa, tratando de lograr un adecuado balance entre el peso relativo de los aspectos teórico conceptuales y el de las indispensables aplicaciones prácticas.

Las notas correspondientes a todo el curso se presentan en dos fascículos. El primero de ellos contiene una introducción a la estadística descriptiva, números índices y distribuciones de probabilidad, mientras que el segundo está dedicado a presentar los elementos básicos del muestreo e inferencia estadística, prueba de hipótesis y relaciones de asociación o dependencia entre atributos y variables.

La presentación de cada tema se completa con un conjunto de ejercicios que permiten aplicar los conocimientos que se van adquiriendo durante el desarrollo de las clases. Los datos de los ejemplos que se incluyen, son en general de carácter ficticio y tienen una finalidad exclusivamente docente. En todos los casos se ha procurado simplificar al máximo los cálculos necesarios, para poder dedicar la mayor parte de la atención, a los propósitos esenciales de la ejercitación.

Se debe aclarar que la lectura de estas notas, no exime de asistir a clases, ni obviar por completo la consulta a la abundante y excelente bibliografía disponible sobre los temas que aquí se tratan. Una parte recomendada de dicha bibliografía, se cita como anexo al final del segundo fascículo.

ORUAL ANDINA

NOTA DE CLASE Nº 1

ESTADÍSTICA DESCRIPTIVA. Datos estadísticos. Atributos y variables. Variable discreta y variable continua. Presentación tabular y gráfica.

ESTADÍSTICA DESCRIPTIVA

La denominación de Estadística Descriptiva se refiere a la parte de la Estadística que tiene por finalidad la reducción de conjuntos de datos empíricos, a un pequeño número de características que concentran la información esencial suministrada por los datos. En general, la estadística descriptiva se aplica a conjuntos muy numerosos de datos, es decir cuando los procesos de resumen resultan ser indispensables para una cabal interpretación de la información que contienen.

1. DATO ESTADÍSTICO

La información numérica que se obtiene contando o midiendo una característica de interés se denomina: *dato estadístico*.

La forma de obtener medidas o recuentos, es decir, de recolectar los datos, es una de las partes más importantes del procedimiento estadístico, pero no será considerada aquí, admitiéndose en general, que la información ya se encuentra disponible para su procesamiento. No obstante, cabe señalar que solo pueden considerarse como datos estadísticos a las mediciones o recuentos que se obtienen mediante procedimientos sistemáticos de observación, ceñidos a rigurosas normas de control de calidad.

2. VARIABLES Y ATRIBUTOS

Podemos distinguir dos clases de datos estadísticos: cuantitativos y cualitativos. La denominación clásica que reciben los primeros es la de *variables*, mientras que a los segundos usualmente se les denomina *atributos*.

Una *variable* es una característica medible; por ejemplo, la altura, el peso, la edad, el salario, el número de animales por establecimiento. Un *atributo* es una característica no medible, por ejemplo: el estado civil, la nacionalidad, la raza.

3. VARIABLE DISCRETA Y VARIABLE CONTINUA

A continuación centraremos nuestro interés en el tratamiento de los datos estadísticos cuantitativos o de variable, cuando existe una sola variable de interés (conjuntos unidimensionales).

Si la variable solo puede tomar algunos valores aislados se denomina: *variable discreta.*

El número de personas por vivienda, es una variable discreta porque no puede tomar cualquier valor. La proporción con que puede presentarse cierta característica de interés en determinado conjunto de observaciones, el número de hijos por hogar, son algunos ejemplos de variable discreta. Hay valores que la variable no puede tomar.

Si en cambio, la variable puede tomar cualquier valor dentro de cierto intervalo de interés, se le denomina: *variable continua.*

El peso y la edad, son ejemplos de variable continuas. Tiene sentido pensar que, dentro de ciertos intervalos, la variable pueda observar cualquier valor, al menos desde el punto de vista teórico ya que desde el punto de vista práctico todas las variables se comportan como discretas, dependiendo la forma de expresión de los valores observados, del sistema y de las unidades de medida que para el caso se hayan definido.

4. PRESENTACIÓN DE LOS DATOS

Se consideran como ***datos originales*** a los datos tal cual se obtienen de la fuente de información, sin ningún tratamiento posterior. Para poder avanzar en el proceso de resumen, destinado a poner en evidencia las características relevantes del conjunto, es necesario proceder a su ordenamiento y posterior presentación tabular y gráfica. Una vez que estos datos se procesan con dichos fines, pasan a denominarse: ***datos agrupados***.

El tratamiento de los datos originales difiere según se trate de variable discreta o bien de variable continua.

4.1 Variable discreta

Simbología:

x_i : valores de la variable tal cual se recolectan (datos originales)

n : número de observaciones consideradas (tamaño del conjunto)

X_i : valores tabulados de la variable (datos agrupados)

m : número de clases, o de distintos valores de la variable $m \leq n$

n_i : frecuencia absolutas simples, o repeticiones de los distintos valores de la variable

h_i : frecuencia relativas simples. Se obtienen mediante el cociente: $h_i = \dfrac{n_i}{n}$

N_i : frecuencias absolutas acumuladas. Se obtienen acumulando las frecuencias absolutas simples, según los sucesivos valores de la variable

H_i : frecuencias relativas acumuladas. Se obtienen acumulando las frecuencias relativas simples, según los sucesivos valores de la variable.

Ejemplo 1: Como parte de una investigación, se pregunta a 10 informantes seleccionados[1], el número de personas que integran sus hogares. Las respuestas son las siguientes:

$x_1 = 3$ $x_2 = 5$ $x_3 = 3$ $x_4 = 4$ $x_5 = 4$

$x_6 = 2$ $x_7 = 3$ $x_8 = 5$ $x_9 = 4$ $x_{10} = 3$

Se pide organizar estos datos en un cuadro de distribución de frecuencias y presentarlos en forma gráfica.

No hay duda que se trata de una variable discreta, ya que sólo podría haber tomado los valores 1, 2, 3, 4, 5 etc. El cuadro de la distribución de frecuencias resultante es el siguiente:

[1] En este caso, como en la mayoría de los ejemplos planteados, los procedimientos se ilustran mediante el tratamiento de un pequeño número de observaciones, procurando evitar cálculos engorrosos y reducir al máximo otras dificultades de carácter práctico y operativo.

Cuadro 1

X_i	n_i	h_i	N_i	H_i
2	1	0,1	1	0,1
3	4	0,4	5	0,5
4	3	0,3	8	0,8
5	2	0,2	10	1,0
	10	1,0		

De acuerdo con la simbología elegida, se observa que n = 10 y m = 4. Deben destacarse las siguientes relaciones de interés, que se cumplen en todo cuadro de distribución de frecuencias:

$$\sum_{i=1}^{m} n_i = n \quad ; \qquad \sum_{i=1}^{m} h_i = 1 \quad ; \qquad N_m = n \quad ; \qquad H_m = 1 \quad ; \qquad H_i = \frac{N_i}{n}$$

En cuanto a la presentación gráfica, se obtienen los diagramas de frecuencias simples y acumuladas que se muestran en las figuras 1 y 2.

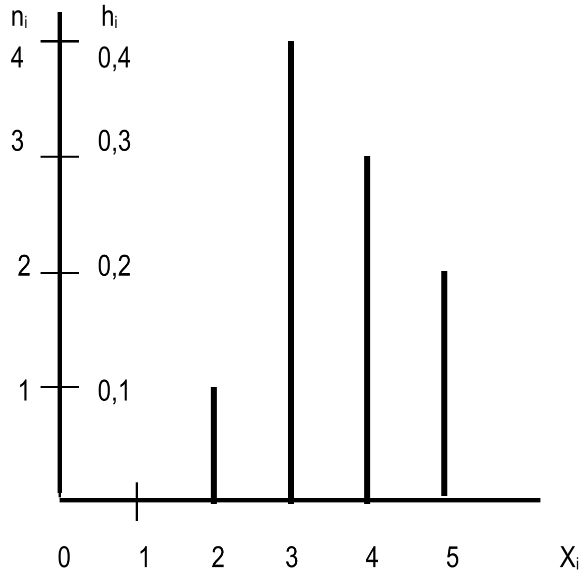
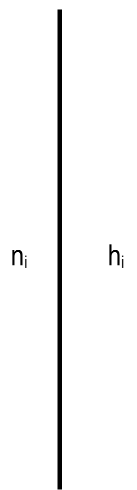


Figura 1 – Gráfico de frecuencias simples de variable discreta.



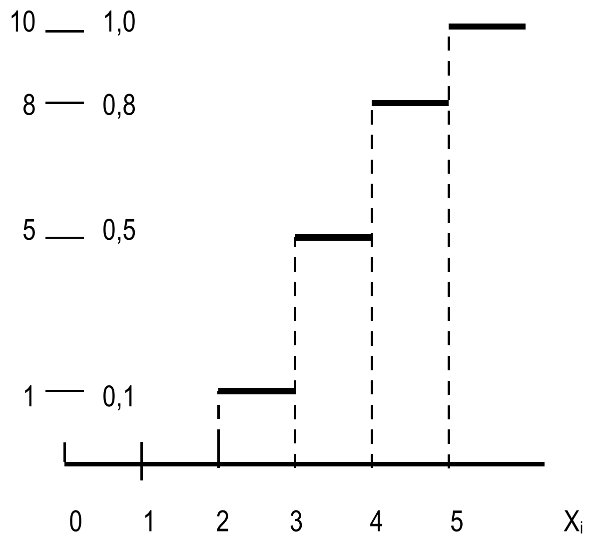


Figura 2 – Gráfico de frecuencias acumuladas de variable discreta.

Los gráficos tienen características discontinuas como cabe esperar, dada la naturaleza de la variable.

4.2 Variable continua

Simbología: Se mantiene la misma simbología utilizada en el caso de variable discreta, excepto:

X_i : representa ahora la marca de clase o centro de los intervalos de clase.

Además se definen:

X'_{i-1} - X'_i : intervalos de clase

c_i : amplitud de los intervalos de clase

Ejemplo 2: Se dispone de 25 observaciones, relativas a la suma de puntajes obtenidos por cada uno de los 25 postulantes en una prueba de admisión. Se pide organizar los datos en un cuadro de distribución de frecuencias y presentarlos gráficamente.

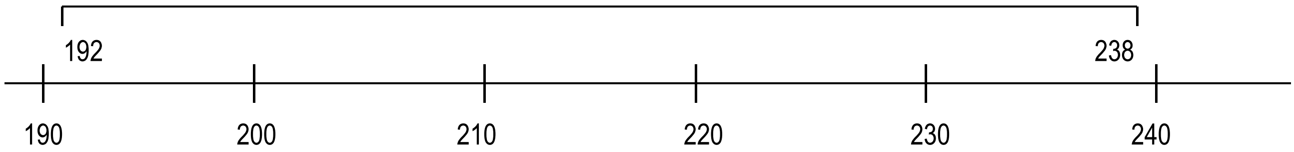
x_1 = 192	x_2 = 210	x_3 = 216	x_4 = 221	x_5 = 234
x_6 = 194	x_7 = 205	x_8 = 220	x_9 = 226	x_{10} = 233
x_{11} = 217	x_{12} = 228	x_{13} = 212	x_{14} = 223	x_{15} = 235
x_{16} = 204	x_{17} = 215	x_{18} = 225	x_{19} = 238	x_{20} = 213
x_{21} = 214	x_{22} = 227	x_{23} = 216	x_{23} = 230	x_{25} = 211

Obsérvese que para simplificar la presentación, las cifras han sido redondeadas a valores enteros. Sin embargo, tiene sentido pensar que los puntajes puedan tomar cualquier valor dentro de un cierto intervalo.

Corresponde entonces considerarlos como una variable de tipo continuo. Recordemos que desde el punto de vista práctico, todas las variables se comportan como discretas.

Para agrupar las observaciones en un cuadro de distribución de frecuencia, hay que preceder de la siguiente forma:

1) Determinar un intervalo que contenga el recorrido de la variable.



2) Dividir este intervalo en ***intervalos de clase***, a los efectos de agrupar la información original. La amplitud (c_i) de los intervalos de clase puede ser constante o variable. No es necesario utilizar normas preestablecidas fijas para la determinación del número de intervalos de clase, ni para determinar la amplitud de los mismos. Como el agrupamiento puede implicar cierta pérdida de información, es el propio usuario de los datos quien deberá decidirlo, en función del grado de detalle que en cada caso desea conservar. Se tomará la convención de que estos intervalos son abiertos por la izquierda (no incluyen el extremo inferior) y cerrados por la derecha (incluyen el extremo superior).

3) De manera que en el ejemplo (procediendo en forma completamente arbitraria) se comienza por determinar un intervalo que contenga el recorrido de la variable:]190 ; 240] . Junto a la ventaja de trabajar con valores sencillos, se asegura de esta forma que todos los datos tengan cabida en el cuadro, cosa que no ocurriría si el mismo se ajustara exactamente al intervalo] 192 ; 238] , limitado por los valores extremos de la distribución (192 quedaría fuera). Luego, se determina el número de intervalos de clase (4) y su amplitud (20 para el primero y 10 en los restantes). Al tabular los datos, se observa que hay valores cuya ubicación debe determinarse en función de la convención asumida, por ejemplo: 210. Si los intervalos de clase son abiertos por la izquierda y cerrados por la derecha, corresponde ubicar esta observación en el primer intervalo. Esta es la convención que utilizaremos en lo sucesivo, aunque considerar intervalos de clase cerrados por la izquierda y abiertos por la derecha no ocasionaría diferencias apreciables.

El cuadro de distribución de frecuencias resultante del proceso de tabulación, es el que se presenta a continuación:

Cuadro 2

X'_{i-1} - X'_i	X_i	n_i	h_i	N_i	H_i	n_i / c_i	h_i / c_i
190 – 210	200	5	0,20	5	0,20	0,25	0,010
210 – 220	215	9	0,36	14	0,56	0,90	0,036
220 – 230	225	7	0,28	21	0,84	0,70	0,028
230 – 240	235	4	0,16	25	1,00	0,40	0,016
		25	1,00				

Una vez que la información original ha sido tabulada, se pasa a admitir que la distribución de las observaciones dentro de cada intervalo de clase es homogénea, prescindiéndose de la información sobre su verdadera ubicación dentro del mismo. La representación gráfica de las frecuencias se realiza mediante rectángulos cuyas bases corresponden a los intervalos de clase utilizados. A los efectos que el área de cada rectángulo resulte numéricamente igual a la frecuencia que representa, es necesario determinar las alturas de los mismos mediante el cociente entre la frecuencia y la amplitud del intervalo correspondiente, según se observa en las dos últimas columnas del cuadro presentado.

Los gráficos que se obtienen (figuras 3 y 4) reciben respectivamente el nombre de ***histogramas de frecuencias*** para las frecuencias simples, y ***polígonos de frecuencias***, para las frecuencias acumuladas.

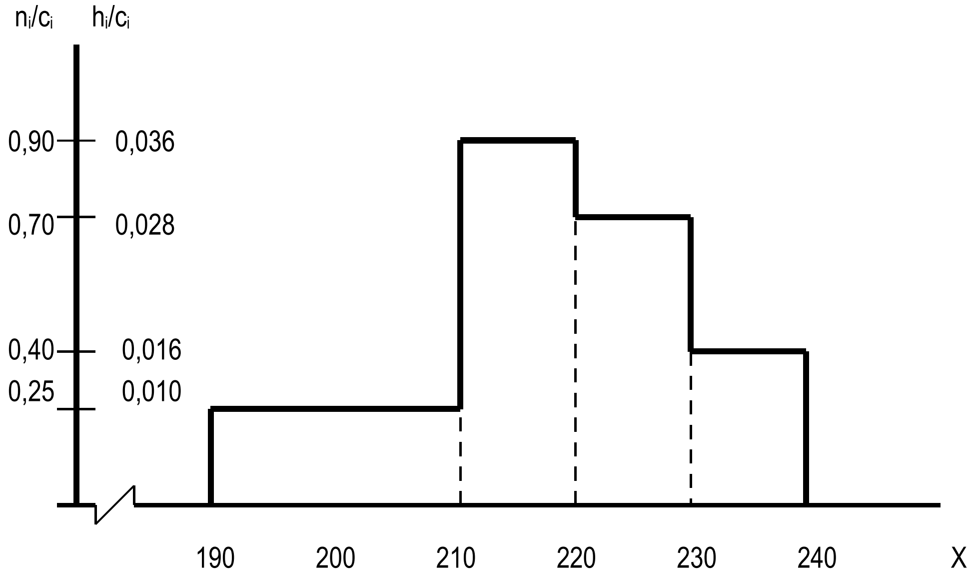


Figura 3 – Histograma de frecuencias de variable continua.

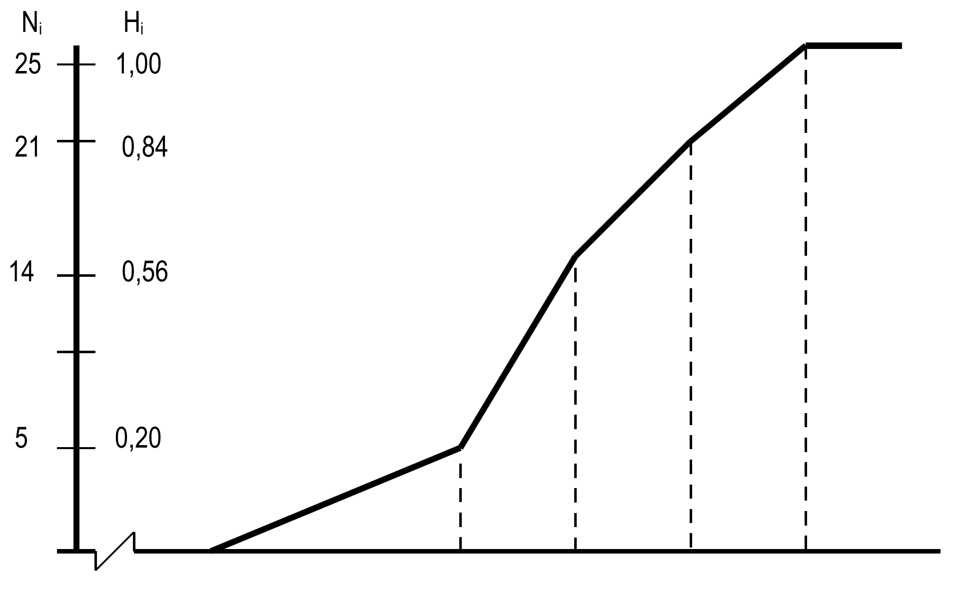


Figura 4 – Polígono de frecuencias acumuladas de variable continua.

Resulta evidente que la tarea de ordenar los datos y presentarlos adecuadamente en cuadros de distribución de frecuencias y en gráficos representativos, permite avanzar en el proceso de resumen, ya que algunas características relevantes de la distribución comienzan a percibirse en forma directa.

1. Se dispone de información numérica de las siguientes características: SEXO; ESTADO CIVIL; PULSACIONES POR MINUTO; NACIONALIDAD; PESO AL NACER; PROPORCION DE MUJERES EN UN CURSO; NÚMERO DE TELEVISORES POR HOGAR; ESTATURA; EDAD.

   Se pide clasificar estos datos en: Atributos; Variables Discretas y Variables Continuas.

2. Proponga un ejemplo de cada categoría.

3. Veinte viviendas presentan la siguiente información relativa al número de ocupantes:

3	1	5	4	1	4	2	3	5	2
2	2	0	1	3	3	2	4	3	2

3.1 Agrupe estos datos en un cuadro de distribución de frecuencias.
3.2 Construya el gráfico de frecuencias absolutas simples y el gráfico de frecuencias absolutas acumuladas.

4 Los pesos en Kg de veinte bultos depositados en una bodega, son los siguientes:

1.2	5.4	7.1	8.5	5.5	3.0	9.8	8.0	7.3	3.7
6.2	6.9	4.0	7.5	6.3	7.2	7.0	7.8	7.4	7.9

Adoptando la convención de que los intervalos de clase son abiertos por la izquierda, cerrados por la derecha, se pide:

4.1 Agrupar los datos en un cuadro de distribución de frecuencias, utilizando los siguientes intervalos de clase:

0 – 5   5 – 7   7 – 8   8 – 10

4.2 Construya el histograma de frecuencias absolutas y el gráfico de frecuencias absolutas acumuladas.

VALORES ESTADÍSTICOS DESCRIPTIVOS. Algunos valores estadísticos de posición. Media aritmética. Cuantiles. Moda. Media geométrica. Media armónica.

VALORES ESTADÍSTICOS DESCRIPTIVOS

Las características principales de una distribución de datos estadísticos son: la posición, la dispersión, la asimetría y el apuntamiento. En función de los objetivos del curso, en estas notas sólo se presentarán algunos valores estadísticos de posición y de dispersión. No obstante, cabe señalar que la asimetría y el apuntamiento también pueden cuantificarse mediante valores estadísticos especialmente definidos con ese fin[2].

ALGUNOS VALORES ESTADÍSTICOS DE POSICIÓN

Los valores estadísticos descriptivos que resumen la información relativa a la posición de un conjunto de datos con respecto a un origen, se denominan ***valores estadísticos de posición*** o bien, ***medidas de posición***. Por lo general señalan centros de la distribución, por lo que también reciben la denominación de ***medidas de la tendencia central.***

Teniendo en cuenta que todos los valores estadísticos descriptivos se pueden calcular a partir de los ***datos originales,*** o bien a partir de los ***datos agrupados***, se presentarán expresiones de las definiciones correspondientes, adecuadas a cada una de ambas situaciones.

También es posible, en el caso de que los datos se encuentren agrupados, calcular los distintos valores estadísticos a partir de las frecuencias absolutas, o bien a partir de las frecuencias relativas.

1. MEDIA ARITMÉTICA

Se define para:

i) datos originales: $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

ii) datos agrupados:

$$\overline{X} = \frac{\sum_{i=1}^{m} X_i\, n_i}{n} \quad y \quad \text{recordando que} \quad h_i = \frac{n_i}{n}$$

$$\overline{X} = \sum_{i=1}^{m} X_i\, h_i$$

[2] En parte de la bibliografía que se cita en el anexo, al final del segundo volumen, es posible encontrar información sobre cómo se definen y utilizan diversos coeficientes destinados a medir la asimetría y el apuntamiento en un conjunto de datos estadísticos.

Como se observa, en el caso de datos agrupados, cada valor de la variable (si es discreta) o marca de clase (si es continua) se encuentra ponderado por la frecuencia que presenta en la distribución. De ahí la denominación de ***media aritmética ponderada***, que esta expresión recibe habitualmente.

Ejemplo 3: Variable discreta

Considerando el ejemplo 1, de la Nota de Clase Nº 1 (Pág. 2) se tiene:

Para los datos originales:
$$\bar{x} = \frac{3+5+3+\cdots+4+3}{10} = 3{,}6$$

Para los datos agrupados (Cuadro 1, Pag. 3):
$$\bar{X} = \frac{2(1)+3(4)+4(3)+5(2)}{10} = 3{,}6$$

Al agrupar los valores de una variable discreta en un cuadro de la distribución de frecuencias no se pierde información. Consecuentemente, los resultados que se obtienen a partir de los datos tabulados, no difieren de los que se basan en los valores originalmente observados por la variable.

Obsérvese que la media aritmética como valor estadístico de resumen, puede tomar cualquier valor central en la distribución aún, como en este caso, un valor que la variable original no puede observar (3,6).

Ejemplo 4: Variable continua

Considerando el ejemplo 2, de la Nota de Clase Nº 1 (pag. 4) se tiene:

Para los datos originales:
$$\bar{x} = \frac{192+210+\cdots+230+221}{25} = 218{,}36$$

Para los datos agrupados (Cuadro 2, pag. 5):
$$\bar{X} = \frac{200(5)+215(9)+225(7)+235(4)}{25} = 218$$

Se observa una pequeña diferencia entre los resultados encontrados para la media aritmética en uno y otro caso. Ello no debe sorprender, ya que en este caso al tabular los datos se pierde parte de la información original, al admitir que dentro de los intervalos de clase la distribución es homogénea (lo que justifica tomar como representativas a las marcas de clase). También hay que considerar que estamos trabajando con un muy reducido número de observaciones. En conjuntos numerosos de datos, las diferencias suelen compensarse entre sí, restando toda significación a esta eventual discrepancia.

1.1 Propiedades

A continuación se enunciarán algunas propiedades importantes de la media aritmética, seguidas de breves demostraciones. Estas últimas se realizan considerando la definición para datos originales, dejándose como ejercicio las correspondientes comprobaciones para el caso de datos agrupados. Dado que las sumatorias se extienden a todo el recorrido de la variable, se omitirá en lo sucesivo, registrar en las expresiones los límites de las mismas.

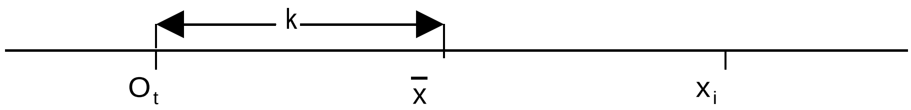
1.1.1 La suma de las desviaciones de las observaciones con respecto a la media aritmética es igual a cero. En otras palabras, la media aritmética configura el “centro de gravedad” de un conjunto de datos. Si las observaciones se sustituyen por pesos y en la posición de la media aritmética se ubica el fiel de una balanza, el sistema queda en equilibrio.

$$\sum\left(x_i - \bar{x}\right) = \sum x_i - \sum \bar{x} \quad = \quad \sum x_i - n\bar{x} \quad = \quad \sum x_i - \cancel{n}\sum \frac{x_i}{\cancel{n}} \quad = \quad 0$$

1.1.2 La suma de los cuadrados de las desviaciones de las observaciones con respecto a la media aritmética, es menor que la suma de los cuadrados de las desviaciones con respecto a cualquier otro origen de trabajo.

$$\sum\left(x_i - \bar{x}\right)^2 \qquad < \qquad \sum\left(x_i - O_t\right)^2$$

Gráfico auxiliar para la demostración:



Demostración:

$$\left(x_0 - O_t\right) \quad = \quad \left(x_i - \bar{x}\right) + k$$

Elevando al cuadrado ambos miembros y sumando a través de todo el recorrido de la variable:

$$\begin{aligned}
\sum\left(x_i - O_t\right)^2 &= \sum\left[\left(x_i - \bar{x}\right) + k\right]^2 \\
&= \sum\left[\left(x_i - \bar{x}\right)^2 + 2k\left(x_i - \bar{x}\right) + k^2\right] \\
&= \sum\left(x_i - \bar{x}\right)^2 + \underbrace{2k\sum\left(x_i - \bar{x}\right)}_{=0} + nk^2 \\
&= \sum\left(x_i - \bar{x}\right)^2 + nk^2
\end{aligned}$$

Siendo todos estos valores positivos al estar elevados al cuadrado, es evidente que:

$$\sum\left(x_i - \bar{x}\right)^2 \qquad < \qquad \sum\left(x_i - O_t\right)^2$$

1.1.3 La media aritmética de una variable más una constante, es la media aritmética de la variable más la constante.

$$y_i = x \quad + q$$

$$\overline{y} = \frac{\sum (x_i + q)}{n} = \frac{\sum x_i}{n} + \frac{\sum q}{n} = \overline{x} + \frac{nq}{n} = \overline{x} + q$$

1.1.4 La media aritmética de una variable por una constante, es igual a la media aritmética de la variable por la constante.

$$y_i = kx_i$$

$$\overline{y} = \frac{\sum (kx_i)}{n} = k\frac{\sum x_i}{n} = k\overline{x}$$

Las propiedades 3 y 4 pueden resumirse de la siguiente forma:

Si $y_i = kx_i + q$, entonces $\overline{y} = k\overline{x} + q$

1.1.5 Si dos subconjuntos de datos tienen medias:

$$\overline{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} \qquad \overline{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2}$$

la media aritmética del conjunto formado por ambos es:

$$\overline{\overline{x}} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

Dado que $\sum_{i=1}^{n_1} x_i = n_1\overline{x}_1$ y $\sum_{i=1}^{n_2} x_i = n_2\overline{x}_2$ la media general será

$$\overline{\overline{x}} = \frac{\sum_{i=1}^{n_1} x_{1i} + \sum_{i=1}^{n_2} x_{2i}}{n_1 + n_2} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

La propiedad se puede generalizar para r subconjuntos:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{r} n_i \bar{x}_i}{\sum_{i=1}^{r} n_i}$$

1.1.6 La media aritmética de la suma de dos variables es igual a la suma de las medias aritméticas de las variables.

$$z_i = x_i + y_i$$

$$\bar{z} = \frac{\sum (x_i + y_i)}{n} = \frac{\sum x_i + \sum y_i}{n} = \frac{\sum x_i}{n} + \frac{\sum y_i}{n} = \bar{x} + \bar{y}$$

La propiedad se puede generalizar para la media aritmética de la suma de r variables.

1.1.7 La media aritmética es un valor muy sensible a los valores extremos, unilateralmente muy alejados del resto del recorrido de la variable

En el caso de distribuciones marcadamente asimétricas, con presencia de valores extremos muy alejados unilateralmente, es necesario tener en cuenta que la media aritmética puede no ser un valor estadístico muy adecuado como medida de resumen.

Ejemplo 5: Los siguientes valores en $, corresponden a donaciones recibidas de seis contribuyentes a cierta colecta:

40; 40; 40; 40; 40; 1.000

La media aritmética es 200, debido a su sensibilidad al valor 1.000. Decir que los contribuyentes aportaron en promedio $ 200 resulta ser un resumen descriptivo poco adecuado.

A continuación se presentará un conjunto de valores estadísticos de posición, que se calculan con prescindencia de los valores extremos de la distribución. Estos valores en general, reciben la denominación de *cuantiles.*

2 CUANTILES

El conocimiento de un valor aislado de una variable adquiere significado, en la medida que resulte posible relacionarlo con la distribución de la que proviene. Por ejemplo, si solo se sabe que un puntaje alcanza el valor 75, no es posible formular ningún juicio de valor respecto a la posición que este puntaje ocupa con relación a los demás.

Para tener información descriptiva al respecto, se define una familia de valores estadísticos de posición, que en general se denominan *cuantiles*.

El cuantil *r de orden s* ($p_{\frac{r}{s}}$), es un valor estadístico que divide una distribución ordenada, de tal forma que la $\frac{r}{s}$ parte de la misma, son valores iguales o menores que él, el resto son valores iguales o mayores.

Lo usual es dividir la distribución en dos, en cuatro, en cinco, en diez o en cien partes iguales, originándose respectivamente los cuantiles denominados: mediana, cuartiles, quintiles, deciles y centiles (también llamados percentiles).

La mediana es el cuantil 1 de orden 2: ($P_{\frac{1}{2}}$)

Los cuartiles son: el cuantil 1 de orden cuatro ($P_{\frac{1}{4}}$) ; el cuantil 2 de orden cuatro ($P_{\frac{2}{4}}$) y el cuantil 3 de orden cuatro ($P_{\frac{3}{4}}$).

Los quintiles son: el cuantil 1 de orden cinco ($P_{\frac{1}{5}}$) ; el cuantil 2 de orden cinco ($P_{\frac{2}{5}}$) ; el cuantil 3 de orden cinco ($P_{\frac{3}{5}}$) y el cuantil 4 de orden cinco ($P_{\frac{4}{5}}$).

De similar forma se definen los deciles y los centiles. Por ejemplo: el cuantil 2 de orden 10 es el segundo decil ($P_{\frac{2}{10}}$). El cuantil 75 de orden cien es el centil (o percentil) 75 ($P_{\frac{75}{100}}$).

En la práctica, es generalmente aceptado considerar que la denominación de un cuantil en particular, se transfiere a todos los valores que se encuentran en el intervalo limitado por éste y el cuantil anterior (si existe). Por ejemplo: "Las observaciones que pertenecen al primer cuartil", son en realidad las menores o iguales a ese valor. "Las que pertenecen al segundo cuartil", son en realidad las que se encuentran comprendidas entre el segundo y el primer cuartil, etc.

A continuación se presentará con cierto detalle el caso específico de la mediana o cuantil 1 de orden 2 ($P_{\frac{1}{2}}$), que puede ser fácilmente generalizable a los demás cuantiles.

2.1 MEDIANA

La mediana, o sea el **cuantil** *1 de orden 2* ($P_{\frac{1}{2}}$), es un valor estadístico que divide una distribución ordenada, de tal forma que la $\frac{1}{2}$ parte de la misma, son valores iguales o menores que él, el resto son valores iguales o mayores.

2.1.1 Para datos originales

- n impar

Cuando el número de observaciones es impar, la mediana es el valor central de la distribución.

$$Me = x_{\frac{n+1}{2}}$$

Ejemplo 6: Sea la distribución

2 ; 2 ; 3 ; <u>3</u> ; 4 ; 4 ; 15

$$Me = x_4 = 3$$

- n par

Cuando el número de observaciones es par, hay dos valores centrales en la distribución. Para salvar la indecisión, se acepta tomar como mediana a la media aritmética de los dos valores centrales.

$$Me = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$$

Ejemplo 7: Sea la distribución:

2 ; 3 ; 5 ; 6 ; 6 ; 14

$$Me = \frac{1}{2}(x_3 + x_4) = \frac{5+6}{2} = 5{,}5$$

Debe notarse que, en ambos casos, los valores extremos no influyen en absoluto para el cálculo de la mediana.

2.1.2 Para datos agrupados

Dado el siguiente cuadro de distribución de frecuencias:

Cuadro 3

X_i	n_i	h_i	N_i	H_i	
X_1	n_1	h_1	N_1	H_1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
			N_{j-1}	H_{j-1}	Clase j-1
X_j	n_j	h_j	N_j	H_j	Clase j
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
X_m	n_m	h_m	N_m	H_m	

Se define ***clase j*** o ***clase mediana*** a la primera clase cuya frecuencia absoluta acumulada supere a *n / 2* , o cuya frecuencia relativa acumulada supere a ½ .

- Variable discreta

i) Cuando $N_{j-1} < \frac{n}{2} < N_j$, o bien $H_{j-1} < \frac{1}{2} < H_j$, entonces:

$$Me = X_j$$

En el gráfico de frecuencias acumuladas de la figura 5 se observa que el valor que cumple con la definición de mediana es X_j pues hasta él se acumula la mitad de las observaciones.

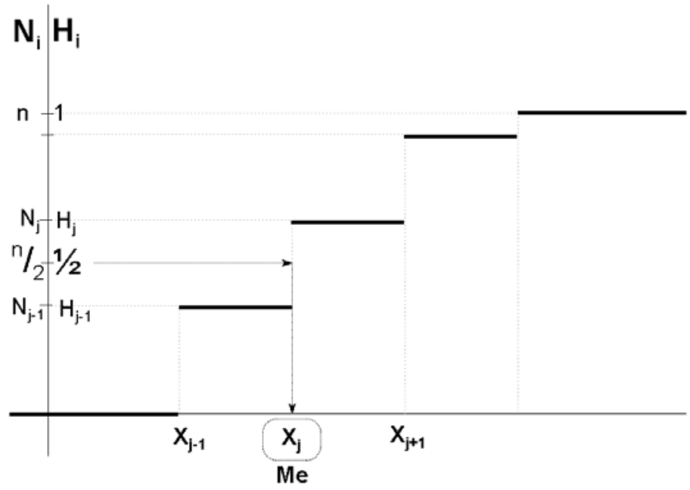


Figura 5 – Valor de la mediana en un gráfico de frecuencias acumuladas para una variable discreta.

Ejemplo 8: Sea el siguiente cuadro de distribución de frecuencias:

Cuadro 4

	X_i	n_i	h_i	N_i	H_i	
	2	20	0,20	20	0,20	
Clase j	3	50	0,50	70	0,70	$Me = x_2 = 3$
	4	20	0,20	90	0,90	
	5	10	0,10	100	1,00	
		100	1,00			

ii) Cuando $N_{j-1} = \dfrac{n}{2} < N_j$, o bien

$H_{j-1} = \dfrac{1}{2} < H_j$, entonces

los valores comprendidos en el intervalo $\left]X_{j-1}\,;X_j\right]$ cumplen con la condición de que hasta ellos se acumula la mitad del total de las observaciones, como se puede observar en el gráfico de la figura 6. Para salvar la indecisión, se acepta considerar como mediana al punto medio de dicho intervalo.

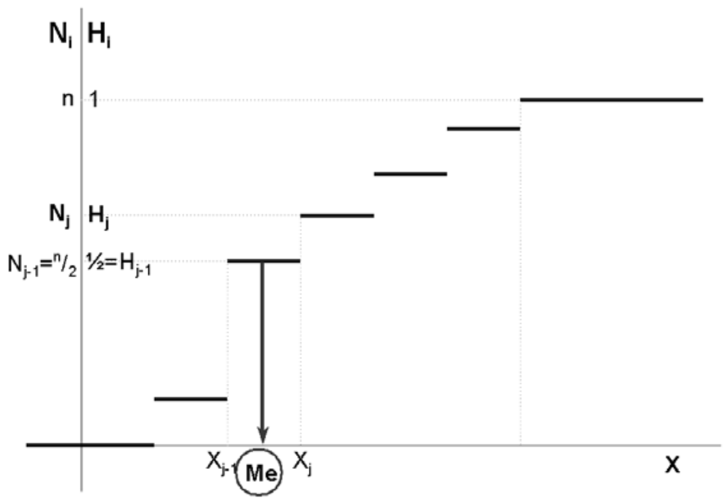


Figura 6 – Valor de la mediana en un gráfico de frecuencias acumuladas para variable discreta.

Ejemplo 9: Sea el siguiente cuadro de distribución de frecuencias:

Cuadro 5

	X_i	n_i	h_i	N_i	H_i
	3	40	0,2	40	0,2
	4	60	0,3	100	0,5
CLASE j	5	80	0,4	180	0,9
	6	20	0,1	200	1,0
		200	1,0		

$$Me = \frac{x_2 + x_3}{2} = \frac{4+5}{2} = 4,5$$

- Variable continua

La definición *de clase* j o *clase mediana*, es en todo similar a la expresada para variable discreta. Sólo es diferente el tipo de cuadro de distribución de frecuencias, a partir del cual se efectúa el cálculo de la mediana.

Cuadro 6

X'_{i-1} - X'_i	X_i	n_i	h_i	N_i	H_i	
X'_0 - X'_i	X'_i	n_i	h_i	N_i	H_i	
.	.	.	.	.	.	
.	.	.	.	.	.	
.	.	.	.	N_{j-1}	H_{j-1}	
X'_{j-1} - X'_j	X'_j	N_j	H_j	N_j	H_j	CLASE j
.	.	.	.	.	.	
.	.	.	.	.	.	
.	.	.	.	.	.	
X'_{m-1} - X'_m	X'_m	N_m	H_m	N_m	H_m	

i) Cuando $N_{j-1} < \frac{n}{2} < N_j$, o bien

$H_{j-1} < \frac{1}{2} < H_j$, entonces

Según se puede ver en el polígono de frecuencias, el valor de la mediana está comprendido dentro del intervalo de clase j. Más precisamente es el valor $X'_{j-1} + z$.

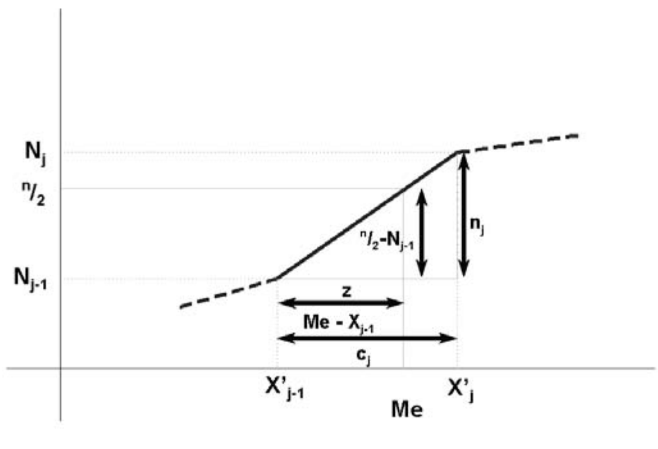


Figura 7 – Valor de la mediana en un gráfico de frecuencias acumuladas para variable continua.

El valor de z se puede determinar observando la relación existente entre los triángulos indicados en el gráfico de la figura 7. Utilizando las frecuencias absolutas acumuladas tendremos:

$$\frac{z}{\frac{n}{2} - N_{j-1}} = \frac{c_j}{N_j - N_{j-1}} \quad \text{, como} \quad N_j - N_{j-1} = n_j$$

Despejando z : $\quad z = c_j \dfrac{\frac{n}{2} - N_{j-1}}{n_j}$ Por lo tanto, el valor de la mediana es:

$$Me = X'_{j-1} + c_j \frac{\frac{n}{2} - N_{j-1}}{n_j}$$

Por razones análogas, utilizando las frecuencias relativas:

$$Me = X'_{j-1} + c_j \frac{\frac{1}{2} - H_{j-1}}{h_j}$$

Ejemplo 10: Sea el siguiente cuadro de distribución de frecuencias:

Cuadro 7

X'_{i-1} - X'_i	n_i	h_i	N_i	H_i	
20 - 40	20	0,20	20	0,2	
40 - 50	40	0,40	60	0,6	CLASE j
50 - 60	30	0,30	90	0,9	
60 - 120	10	0,10	100	1,0	
	100	1,00			

Utilizando las frecuencias absolutas:

$$Me = 40 + 10 \frac{50 - 20}{40} = 47,5$$

Utilizando las frecuencias relativas:

$$Me = 40 + 10\,\frac{\frac{1}{2} - 0{,}20}{0{,}40} = 47{,}5$$

ii) Cuando $N_{j-1} = \frac{1}{2} < N_j$, o bien

$H_{j-1} = \frac{1}{2} < H_j$, entonces

El cálculo de la mediana es mucho más simple. Según se observa en el gráfico de la figura 8, el valor hasta el que acumula la mitad de las observaciones, es X'_{j-1}.

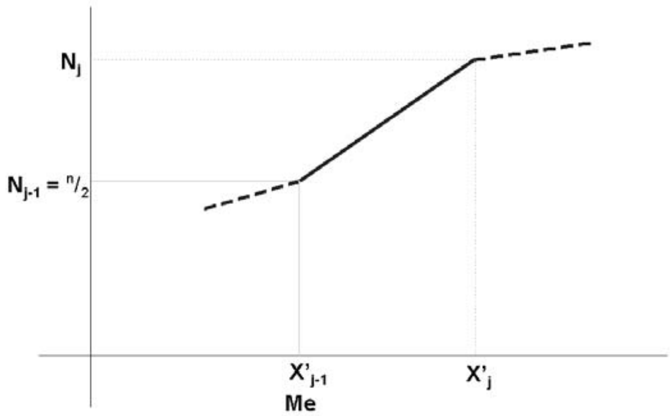


Figura 8 – Valor de la mediana en un gráfico de frecuencias acumuladas para variable continua.

Ejemplo 11: Sea el siguiente cuadro de distribución de frecuencias:

Cuadro 8

X'_{i-1} - X'_i	n_i	h_i	N_i	H_i	
5 - 10	100	0,1	100	0.1	
10 - 20	400	0,4	500	0,5	
20 - 50	300	0,3	800	0,8	CLASE j
50 - 100	200	0,2	1.000	1,0	
	1.000	1,0			

$Me = X'_2 = 20$

2.1.3 Propiedades de la mediana

La mediana posee una importante propiedad: la suma de los valores absolutos de las desviaciones tiene un valor mínimo, cuando se consideran las desviaciones con respecto a la mediana.

$$\sum |x_i - Me| \quad \leq \quad \sum |x_i - O_t|$$

Consideremos en primer término la suma S de los valores absolutos de las desviaciones de las observaciones con respecto a la mediana. Si elegimos un origen de trabajo (O_t) , situado una unidad antes que el valor de la mediana y consideramos las nuevas desviaciones en valor absoluto, resulta evidente que las N_{j-1} observaciones menores que la mediana, estarán una unidad más cerca del origen de trabajo que de la mediana, por lo que la suma de las nuevas desviaciones disminuirán el valor de S en N_{j-1} unidades.

Por otra parte, las desviaciones de las observaciones iguales o mayores que la mediana con respecto al origen de trabajo, aumentarán el valor de S en ($n - N_{j-1}$) unidades, de manera que:

$$\sum |x_i - O_t| \quad = \quad S - N_{j-1} + n - N_{j-1} \quad = \quad S + n - 2N_{j-1}$$

De acuerdo a la definición de mediana, se sabe que:

$$n \quad \geq \quad 2\,N_{j-1}$$

Por lo que se comprueba que:

$$\sum |x_i - O_t| \quad \geq \quad \sum |x_i - Me|$$

Si en cambio, el origen de trabajo estuviera situado una unidad después que la mediana, se tendría:

$$\sum |x_i - O_t| \quad = \quad S + N_j - \left(n - N_j\right) \quad = \quad S - n + 2N_j$$

Y como $2\,N_j \geq n$, también se verifica que:

$$\sum |x_i - O_t| \quad \geq \quad \sum |x_i - Me|$$

Resulta evidente que si se considera un origen de trabajo más de una unidad distante de la mediana, esta relación se cumplirá con mayor razón.

Ejemplo 12:

A lo largo de una misma carretera, existen cinco caseríos. Los cuatro últimos están situados a 6, 9,17 y 19 km de distancia del primero. En el primer caserío vive un 15 por ciento de los niños y en los demás: 40, 30,10 y 5 por ciento, respectivamente. En cuál caserío debe situarse una escuela para que el esfuerzo de los escolares de los cinco caseríos para concurrir a ella, sea mínimo.

Cuadro 9

	X_i	h_i	H_i	
	0	0,15	0,15	
Clase j	6	0,40	0,55	$Me = X_2 = 6$
	9	0,30	0,85	
	17	0,10	0,95	
	19	0.05	1.00	

La escuela deberá situarse a 6 km del primer caserío, es decir en el segundo caserío.

2.2 CUANTILES

Dado que la mediana es un caso particular de cuantil, se generalizarán los procedimientos que han permitido determinar las expresiones de cálculo que han sido presentadas anteriormente. Se recordará que:

El cuantil *r de orden s* ($P^{\frac{r}{s}}$), es un valor estadístico que divide una distribución ordenada, de tal forma que la $\frac{r}{s}$ parte de la misma, son valores iguales o menores que él, el resto son valores iguales o mayores.

La *clase j*, o clase que contiene el valor del cuantil, se define para el caso general como la primera clase cuya frecuencia absoluta acumulada supera a $\left(\frac{r}{s}n\right)$, o cuya frecuencia relativa acumulada supera a $\frac{r}{s}$.

- Variable discreta

i)
$$\left.\begin{array}{l} \text{Si} \quad N_{j-1} < \frac{r}{s}n < N_j \\ \text{o} \quad H_{j-1} < \frac{r}{s} < H_j \end{array}\right\} \quad P^{\frac{r}{s}}$$

ii)
$$\left.\begin{array}{l} \text{Si} \quad N_{j-1} = \frac{r}{s}n < N_j \\ \text{o} \quad H_{j-1} = \frac{r}{s} < H_j \end{array}\right\} \quad P^{\frac{r}{s}} = \frac{X_{j-1} + X_j}{2}$$

Ejemplo 13: Dada la siguiente distribución de variable discreta, se pide calcular el cuantil 3 de orden 10 ($P_{\frac{3}{10}}$), y el cuantil 75 de orden 100 ($P_{\frac{75}{100}}$)

Cuadro 10

X_i	n_i	h_i	N_i	H_i	
2	20	0.10	20	0.10	
3	40	0.20	60	0.30	
4	70	0.35	130	0.65	Clase j para $P_{\frac{3}{10}}$
5	50	0.25	180	0.90	Clase j para $P_{\frac{75}{100}}$
6	20	0.10	200	1.00	
	200	1.00			

$$\frac{3}{10}200 = 60 \qquad P_{\frac{3}{10}} = \frac{X_{j-1} + X_j}{2} = \frac{4+3}{2} = 3,5$$

$$\frac{75}{100}200 = 150 \qquad P_{\frac{75}{100}} = X_j = 5$$

- Variable continua

$$P_{\frac{r}{s}} = X'_{j-1} + c_j \frac{\frac{r}{s}n - N_{j-1}}{n_j} = X'_{j-1} + c_j \frac{\frac{r}{s} - H_{j-1}}{h_j}$$

Ejemplo14: En la siguiente tabla de frecuencia, calcular el primer cuartil ($P_{\frac{1}{4}}$)

Cuadro 11

$X'_{i-1} - X'_i$	X_i	n_i	h_i	N_i	H_i	
20 – 30	25	10	0.0625	10	0.0625	
30 – 40	35	50	0.3125	60	0.3750	Clase j
40 – 60	50	70	0.4375	130	0.8125	
60 – 100	80	30	0.1875	160	1.0000	
		160	1.0000			

$$\frac{1}{4}160 = 40 \quad \rightarrow \quad \begin{array}{l} P^{\frac{1}{4}} = 30 + 10\dfrac{40-10}{50} = 36 \\ P^{\frac{1}{4}} = 30 + 10\dfrac{0,25-0,0625}{0,3125} = 36 \end{array}$$

3. MODA

- Variable discreta

En el caso de variable discreta la moda (o *modo)* es en general, el valor de la variable al que le corresponde la mayor frecuencia. Sin embargo, se tiene un valor modal, cada vez que un valor de la variable se presenta con una frecuencia mayor que las que presentan los valores contiguos, es decir:

Md = X_j si se cumple que: $n_{j-1} < n_j > n_{j+1}$
$h_{j-1} < h_j > h_{j+1}$

De acuerdo con esta definición, en una distribución se pueden presentar una, dos o más modas. Esto hace que las distribuciones reciban el nombre de unimodales, o bien multiimodales (bimodales etc.). En distribuciones multimodales, al valor que presenta el mayor máximo se le denomina moda absoluta.

Ejemplo 15:

Dado el siguiente cuadro y el correspondiente gráfico de frecuencias de la figura 9:

Cuadro 12

X_i	n_i	h_i
0	2	0,04
1	9	0,18
2	4	0,08
3	7	0,14
4	20	0,40
5	8	0,16
	50	1,00

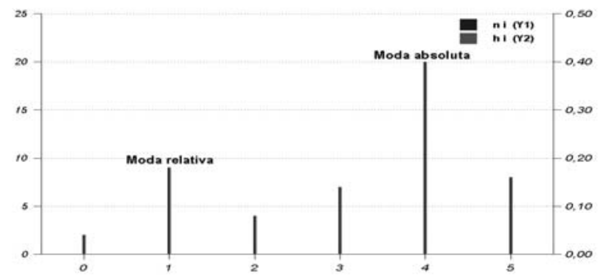


Figura 9. Gráfico de frecuencias simples para los valores del Cuadro 9.

Se observa que los valores 1 y 4 son las modas de la distribución. En efecto, sus frecuencias absolutas cumplen: 2 < 9 > 4 y 7 < 20 > 8. El valor 4 es la moda absoluta de la distribución.

- Variable continua

Cuando la información proviene de una variable de tipo continuo agrupada en intervalos de clase de igual amplitud, si se presenta un intervalo cuya frecuencia es mayor que las de los valores contiguos, éste será un intervalo modal. Por razones prácticas y operativas admitiremos aquí que la moda será la marca de clase de de dicho intervalo. El valor de la moda puede precisarse mejor mediante fórmulas de cálculo que se desarrollan en función de las frecuencias contiguas, pero en razón de las limitaciones que tiene el uso de la moda como medida de centralización, no las tendremos en cuenta en estas notas.

Ejemplo 16: Sea el siguiente cuadro y el correspondiente histograma de frecuencias de la figura 10:

Cuadro 13

$X'_{i-1} - X'_i$	X_i	n_i	n_i/c_i
4 – 6	5	20	10
6 – 8	7	40	20
8 – 10	9	30	15
10 – 12	11	20	10
		110	

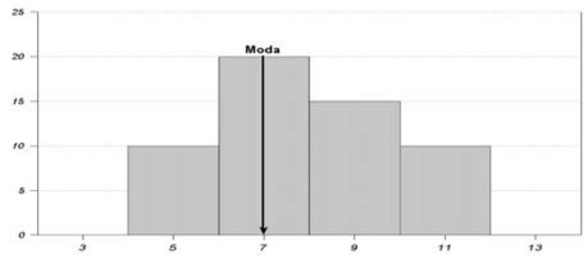


Figura 10. Histograma de frecuencias para los valores del Cuadro 10.

Dada la simplificación acordada, la moda será la marca de clase del segundo intervalo:

Md = 7 ya que: 20 < 40 > 30

4. MEDIA GEOMÉTRICA

Se define para:

i) datos originales:

$$m_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

ii) datos agrupados:

$$M_g = \sqrt[n]{X_1^{n_1} \cdot X_2^{n_2} \cdot \ldots \cdot X_m^{n_m}} = \sqrt[n]{\prod_{i=1}^{m} X_i^{n_i}}$$

A efectos de evitar dificultades de cálculo, originadas al trabajar con cantidades de gran tamaño, se puede utilizar una transformación algebraica de la formula presentada, que consiste en tomar logaritmos de ambos miembros de la igualdad:

$$\log M_g = \frac{\sum_{i=1}^{m} n_i \log X_i}{n} = \sum_{i=1}^{m} h_i \log X_i$$

Como se observa, el logaritmo de la media geométrica es la media aritmética de los logaritmos de los valores de la variable.

Ejemplo 17: Una variable representa la población de una localidad en dos fechas. Se pide determinar la población media en el período. Sea $x_1 = 6{,}5$ y $x_2 = 8{,}0$ estando ambas cantidades expresadas en miles.

La forma en que crece la población es exponencial, ya que la tasas de crecimiento se aplica a cantidades a su vez, crecientes. Por lo tanto, no corresponde en este caso estimar la población media a través de la media aritmética, que resulta adecuada cuando el crecimiento es de tipo rectilíneo.

La media geométrica resulta ser: $m_g = \sqrt{6{,}5 \times 8{,}0} = 7{,}21$

Como ocurre con generalidad, el valor de la media geométrica se encuentra por debajo del valor de la media aritmética que para el caso resulta ser 7,25.

En el gráfico de la figura 11 puede apreciarse la diferencia.

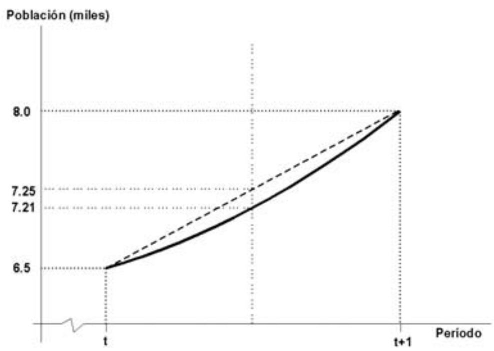


Figura 11. Gráfico comparativo de los valores media aritmética y media geométrica del Ejemplo 14.

Ejemplo 18: Dada la siguiente distribución de frecuencias, se pide calcular la media geométrica:

Cuadro 14

X_i	n_i	$\log X_i$	$n_i \log X_i$
1	2	0	0
3	5	0,477	2,385
5	10	0,699	6,990
7	5	0,845	4,225
9	3	0,954	2,862
	25		16,462

$$\log M_g = \frac{16{,}462}{25} = 0{,}658 \quad \rightarrow \quad M_g = \text{antilog}\, 0{,}658 = 4{,}55$$

Es necesario mencionar una limitación importante para el uso de la media geométrica: es muy sensible a valores pequeños, anulándose si al menos uno de los valores es cero. Adicionalmente, si la variable toma valores negativos, no es posible utilizar logaritmos.

5. MEDIA ARMÓNICA

Se define para:

i) datos originales:

$$m_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

ii) datos agrupados:

$$M_h = \frac{n}{\sum_{i=1}^{m} \frac{n_i}{X_i}} = \frac{1}{\sum \frac{h_i}{X_i}}$$

Como se observa, la media armónica es el valor inverso de la media aritmética de los valores inversos de la variable.

Ejemplo 19: En tres períodos sucesivos se gastaron \$300 por período en adquirir cierto producto cuyo precio unitario fue de \$15, \$20 y \$30 respectivamente. Se desea saber cuál fue el costo promedio del producto.

Corresponde calcular una media armónica simple entre los precios unitarios, es decir:

$$m_h = \frac{3}{\frac{1}{15} + \frac{1}{20} + \frac{1}{30}} = 20$$

Esto es así porque el costo promedio debe obtenerse como cociente entre el costo total y la cantidad total comprada.

$$\frac{900}{\frac{300}{15} + \frac{300}{20} + \frac{300}{30}} = \frac{900}{300\left(\frac{1}{15} + \frac{1}{20} + \frac{1}{30}\right)} = \frac{3}{\frac{1}{15} + \frac{1}{20} + \frac{1}{30}}$$

Si en cada período se gastaran distintas cantidades de dinero, por ejemplo \$300, \$ 500 y \$900 respectivamente, habría que calcular una media armónica ponderada:

$$M_h = \frac{1700}{\frac{300}{15} + \frac{500}{20} + \frac{900}{30}} = 22,67$$

La media armónica también se aplica al cálculo de velocidades medias, cuando los datos son diferentes espacios recorridos a distintas velocidades.

Ejemplo 20: Un vehículo recorre los primeros 50 Km. de su recorrido a 60 Km./h, y los siguientes 50 Km. a 80 Km./h ¿Cuál es la velocidad media del vehículo?

$$m_h = \frac{2}{\frac{1}{60}+\frac{1}{80}} = 68,57$$

y no 70 Km./h como podría suponerse al emplear erróneamente la media aritmética.

1. Se dispone de la siguiente información proveniente de un cuadro de distribución de frecuencias de variable discreta:

X_i	n_i	h_i	N_i	H_i
2			2	
3	3			
4				
5	5		17	
6			20	

1.1 Complete la tabla y construya el gráfico de frecuencias relativas simples y el de frecuencias relativas acumuladas.

1.2 Calcule la media aritmética (*4.20*), la mediana (*4*), el primer cuartil (*3.5*) el tercer cuartil (*5*), la moda (*4*) y la media geométrica (*4.02*).

2. Ciento veinte puntajes se agrupan en una tabla de distribución de frecuencias con cuatro intervalos de igual amplitud, de la que se conocen los siguientes datos:

$X_2 = 60$ $X_3 = 70$ $n_1 = 8$ $n_3 = 60$ $N_2 = 40$

2.1 Complete la tabla y construya el histograma de frecuencias relativas y el polígono de frecuencias relativas acumuladas.

2.2 Calcule la media aritmética (*67.67*), el primer quintil (*60*), la moda (*70*) y la media geométrica (*67.17*).

2.3 Calcule el percentil 75 (*73.33*) y estime el número de puntajes iguales o superiores a ese valor (*30*).

2.4 Si se adicionan 5 puntos a cada puntaje, ¿cuál sería el valor de la nueva media (*72.67*)? ¿Y si se aumentan un 20 por ciento (*81.20*)?

NOTA DE CLASE Nº 3

VALORES ESTADÍSTICOS DE DISPERSIÓN. Recorrido. Desviación media. Varianza. Desviación estándar. Coeficiente de variación.

ALGUNOS VALORES ESTADÍSTICOS DE DISPERSIÓN

La idea de dispersión se relaciona con la mayor o menor concentración de los datos alrededor de un valor central, generalmente la media aritmética. A continuación se presentan algunas medidas cuyo fin es cuantificar esta dispersión o variabilidad.

1. RECORRIDO

El *recorrido* o *rango* es una medida de dispersión que depende exclusivamente de los valores extremos observados por la variable. [3]

$$L = x_n - x_1$$

Al no tener en cuenta los valores intermedios de la distribución, resulta ser un valor estadístico de aplicación muy restringida.

2. DESVIACIÓN MEDIA

Una adecuada medida de la variabilidad debe tener en cuenta a todos los valores de la variable. Si la dispersión se relaciona con la media aritmética, considerar un promedio de todas las desviaciones con respecto a ella parecería muy indicado. Pero en razón de una de las propiedades de la media aritmética (1.1), sabemos que:

$$\sum \left(x_i - \bar{x} \right) = 0 \qquad \text{, o que:}$$

$$\sum \left(X_i - \bar{X} \right) n_i = 0$$

por lo que habrá que recurrir a un artificio para evitar que la suma de las desviaciones se haga cero.

Si se toman las desviaciones en valor absoluto se obtiene la llamada desviación media:[4]

$$d_m = \frac{\sum \left| x_i - \bar{x} \right|}{n} \qquad \text{o bien:} \qquad D_m = \frac{\sum \left| X_i - \bar{X} \right| n_i}{n} = \sum \left| X_i - \bar{X} \right| h_i$$

---

[3] Para eliminar la influencia de los valores extremos, a veces se utiliza el *recorrido interdecil* (diferencia entre el noveno y el primer decil) y, más frecuentemente, el *recorrido intercuartil*, que se define como la diferencia entre el tercer y el primer cuartil, la que comprende el 50% central de la distribución.

$$L^{\frac{1}{4}} = P^{\frac{3}{4}} - P^{\frac{1}{4}}$$

Del recorrido intercuartílico se obtiene el *recorrido semi–intercuartílico*, dividiéndolo el primero entre dos, o sea que resulta la mitad del intervalo que contiene el 50% central de la distribución.

$$Q = \frac{P^{\frac{3}{4}} - P^{\frac{1}{4}}}{2}$$

[4] Si para calcular las desviaciones, se usa la mediana en lugar de la media aritmética, se obtiene otra medida de dispersión que, como sabemos por la propiedad demostrada en 2.3, resultará una desviación media mínima.

3. VARIANZA

Otro artificio para evitar que la suma de las desviaciones de las observaciones con respecto a la media aritmética sea siempre igual a cero, consiste en elevar al cuadrado estas desviaciones. Se obtiene así la más importante de las medidas de dispersión: la **varianza**.[5]

i) Datos originales: $s^2 = \dfrac{\sum (x_i - \overline{x})^2}{n}$, o bien

ii) Datos agrupados: $S^2 = \dfrac{\sum \left(X_i - \overline{X}\right)^2 n_i}{n} = \sum \left(X_i - \overline{X}\right)^2 h_i$

Para la ejercitación, la aplicación directa de esta definición suele no resultar muy conveniente. Por lo general, se prefiere utilizar como fórmula de cálculo, una sencilla transformación algebraica de la misma.

Para datos originales esta transformación consiste en:

$$
\begin{aligned}
s^2 &= \frac{\sum (x_i - \overline{x})^2}{n} = \frac{\sum \left(x_i^2 - 2\overline{x}\,x_i + \overline{x}^2\right)}{n} = \\
&= \frac{\sum x_i^2}{n} - 2\overline{x}\sum x_i + \frac{n\,\overline{x}^2}{n} = \frac{\sum x_i^2}{n} - 2\overline{x}^2 + \overline{x}^2 = \\
&= \frac{\sum x_i^2}{n} - \overline{x}^2
\end{aligned}
$$

Igualmente, para datos agrupados se obtiene:

$$S^2 = \frac{\sum X_i^2 n_i}{n} - \overline{X}^2 = \sum X_i^2 h_i - \overline{X}^2$$ [6]

3.1 *Propiedades*

3.1.1 La varianza de una variable más una constante, es igual a la varianza de la variable.

$y_i = x_i + q$

$$s_y^2 = \frac{\sum \left[(x_i + q) - (\overline{x} - q)\right]^2}{n} = \frac{\sum [x_i + q - \overline{x} - q]^2}{n} = \frac{\sum (x_i - \overline{x})^2}{n} = s_x^2$$

3.1.2 La varianza de una variable por una constante, es igual al cuadrado de la constante por la varianza de la variable.

$y_i = k\,x_i$

$$s_y^2 = \frac{\sum (k\,x_i - k\,\overline{x})^2}{n} = \frac{\sum k^2 (x_i - \overline{x})^2}{n} = \frac{k^2 \sum (x_i - \overline{x})^2}{n} = k^2 s_x^2$$

---

[5] En razón de lo demostrado para la media aritmética en 1.2, la varianza es un promedio de desvíos cuadráticos mínimos.

[6] La demostración se deja como ejercicio.

4. DESVIACIÓN ESTÁNDAR

La varianza se expresa en unidades que son el cuadrado de las unidades de la variable original. Para muchos fines resulta necesario definir una medida de dispersión que se exprese en las mismas unidades que las observaciones, como la **desviación estándar** o **desviación típica**, que se define como la raíz cuadrada positiva de la varianza:

i) Datos originales: $$s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

ii) Datos agrupados: $$S = \sqrt{\frac{\sum X_i^2 n_i}{n} - \bar{X}^2} = \sqrt{\sum X_i^2 h_i - \bar{X}^2}$$

Resulta obvio que las propiedades enunciadas para la varianza se cumplen también para la desviación estándar, aplicando raíz cuadrada a ambos miembros de las igualdades propuestas.

Ejemplo 21. En el siguiente cuadro de distribución de frecuencias se puede calcular la desviación estándar:

Cuadro 15

X_i	n_i	$X_i n_i$	$X_i^2 n_i$
5	1	5	25
6	4	24	144
7	10	70	490
8	4	32	256
9	1	9	81
	20	140	996

$$S^2 = \frac{996}{20} - 7^2 = 0{,}80$$

$$S = \sqrt{0{,}8} = 0{,}89$$

5. COEFICIENTE DE VARIACIÓN

Cuando las variables de dos o más distribuciones se expresan en unidades de distinta denominación (metros, toneladas, años, etc.), no es posible utilizar las medidas de dispersión presentadas hasta ahora, para comparar directamente las medidas de variabilidad que presentan, puesto que las mismas también se expresan en función de las unidades originales.

Esta dificultad puede superarse mediante el cálculo del ***coeficiente de variación***, que se define como:

$$CV = \frac{s}{\overline{x}}$$

Debe mencionarse que el coeficiente de variación también tiene utilidad para comparar la variabilidad de distribuciones de variables que se expresan en las mismas unidades. Especialmente en el caso de las medias aritméticas se diferencien en forma muy apreciable.

Asimismo debe tenerse en cuenta que dos variables pueden presentar igual dispersión o variablidad en términos absolutos, pero diferente variabilidad relativa medida por el coeficiente de variabilidad.

Ejemplo 22. La media aritmética de las alturas de un grupo de estudiantes universitarios es 1,68m, con una desviación estándar de 0,252m. Por otra parte, su peso medio es de 70 Kg. con una desviación estándar de 7 Kg. ¿Qué medida presenta mayor variabilidad: la altura o el peso?

$$CV_h = \frac{0,252}{1,68} = 0,15$$ Es decir: el 15 por ciento.

$$CV_p = \frac{7}{70} = 0,10$$ Es decir: el 10 por ciento.

En base a estos resultados, es posible afirmar que, en términos relativos, es mayor la variabilidad de las alturas que la variabilidad de los pesos.

1. De una distribución de variable continua, con cinco intervalos de clase de igual amplitud, se sabe que:

   - $n_1 = n_5$ y $n_2 = n_4$ es decir que es SIMÉTRICA
   - $\bar{X} = 25$ $n_2 - n_1 = 10$ $n_3 = 10$ $n = 110$
   - El extremo inferior del primer intervalo de clase es 12,5

Se pide: 1.1 Completar la tabla.
1.2 Calcular la moda (*20 y 30*), la varianza (*50*) y el coeficiente de variación (*0.28*).

2. En una distribución simétrica, de variable continua, con cinco intervalos de igual amplitud, se conocen los siguientes datos:
   $X_1 = 1$ $X_5 = 9$ $n_1 = 4$ $n_3 - n_5 = 2$ $N_4 = 16$

Se pide: 2.1 Reconstruir la tabla.
2.2 Calcular la desviación estándar (*2.76*).
2.3 ¿Cuál sería el valor de la desviación estándar si a cada valor de la variable se le adicionaran 2 unidades (*2.76*)? ¿Y si se aumentaran un 2 por ciento (*2.82)*?

3. Los datos provenientes de una variable de tipo continuo, se agrupan en la siguiente tabla:

X'_{i-1} - X'_i	h_i
10 – 20	0.20
20 – 30	
30 – 40	
40 – 50	0.12

Si se sabe que la media aritmética es igual a 29.4, se pide calcular la varianza (*88.64*) y estimar el primer quintil (20) y el percentil 95 (45.833).

4. Cincuenta observaciones provenientes de una variable continua, se agrupan en un cuadro de distribución de frecuencias con cuatro intervalos de clase de igual amplitud. Se pide reconstruir la tabla sabiendo que:

   $X_2 = 50$ $n_1 = 4$ $N_2 = 20$ $n_3 = 25$ $\bar{X} = 62.4$

4.1 Calcule la media geométrica (*60.*23) y la varianza (*238.24*).
4.2 Determine los valores de la variable entre los que se encuentra el 50 por ciento central de la distribución (*50.625 y 74.00*).

1. Las alturas de 198 personas (en metros), se agrupan en el siguiente cuadro de distribución de frecuencias:

X'_{i-1}-X'_i	X_i	n_i
1,45 – 1,55	1,50	74
1,55 – 1,65	1,60	99
1,65 – 1,75	1,70	25

Con posterioridad, se conoce la existencia de 2 alturas que no han sido tabuladas: 1,55 y 1,65. Se pide incluir esas observaciones en el cuadro, teniendo en cuenta que los intervalos de clase son abiertos por la izquierda y cerrados por la derecha.
Una vez completado el cuadro:

1.1 Calcule la media aritmética y la varianza (1.575m; 0.0044m^2).
1.2 Si el instrumento de medida utilizado era imperfecto, registrando 2 cm de menos en cada medición: corrija los valores de la media aritmética y de la varianza, utilizando las propiedades de ambos valores estadísticos (1.595m; 0.0044m^2).
1.3 Determine entre qué valores de la variable están comprendidas las 120 alturas centrales de la distribución (1.5033m; 1.6350m).
1.4 Construya el histograma de frecuencias relativas.

2. La siguiente información se refiere al número años de duración del matrimonio de 104 personas.

Número de años de matrimonio	Frecuencias absolutas
0 - 5	23
5 - 10	26
10 - 20	40
20 - 40	15

A efectos de asignar categorías teniendo en cuenta la duración de los matrimonios, se decide considerar:

Matrimonio corto : menor que el primer cuartil
Matrimonio medio: del primer al tercer cuartil
Matrimonio largo : mayor que el tercer cuartel

Se pide: 2.1 Calcular los cuartiles mencionados (5.5769; 17.2500).
2.2 Estimar el número de matrimonios que contendrá cada categoría (26; 52; 26).

2.3 Construir el histograma de frecuencias absolutas.
2.4 Calcular la media geométrica y la varianza (9.3791; 74.9403).

3. Los pesos en Kg. de un lote de artículos, se agrupan en un cuadro de distribución de frecuencias con 7 intervalos de igual amplitud. Se sabe que la distribución es simétrica y que además, el extremo inferior del quinto intervalo de clase es 73. Los demás datos conocidos son:

$X_1=10$ $n_1=1$ $N_4=70$ $n_2=15$ $\sum_{i=5}^{7} h_i = \frac{4}{11}$

3.1 Reconstruya todas las columnas del cuadro.
3.2 Estime entre qué valores de la variable se encuentra ubicado el 80 por ciento central de la distribución (31; 97).
3.3 Calcule el coeficiente de variación (0.366).
3.4 Si se descubre que la balanza utilizada tenía un desperfecto que agregó 0.5 Kg en cada pesaje, corrija el coeficiente de variación calculado en 3.3, aplicando propiedades (0.369).

NOTA DE CLASE Nº 4

NÚMEROS INDICE.

La construcción de números índice, proporciona una respuesta generalmente aceptada para el problema de resumir, con diversos fines, un conjunto de datos estadísticos relativos a una variable de interés y describir su evolución a través del tiempo.
Al tener en cuenta el orden temporal en que se producen las observaciones, el tema de los números índice se inscribe íntegramente dentro de las denominadas series cronológicas, o históricas, o de tiempo.
La realidad se encuentra llena de ejemplos sencillos que muestran la utilidad de los números índice para comparar variaciones de precios de cantidades, de valores globales, etc. producidas en cierto lapso en un artículo o en un conjunto limitado de ellos. En otras oportunidades sin embargo, hay que resumir una gran cantidad de datos provenientes de fuentes de información sumamente complejas, como los precios de los artículos del consumo, los salarios, el volumen físico de la producción global, etc.
La sola mención de estos últimos ejemplos resultarían razones suficientes como para comprender que nos encontramos frente a un problema delicado, con profundas implicancias en los ámbitos políticos, sociales y económicos y que debe encararse con gran cuidado metodológico y operativo.
No obstante, hay que adelantar que no se trata de un tema con grandes requerimientos teóricos. La comprensión de sus fundamentos, así como la de sus limitaciones, son fácilmente accesibles para aquellos que se encuentran familiarizados con los conceptos de la Estadística Descriptiva y con la utilización de la misma de acuerdo a sus finalidades prácticas.

1. SELECCION DEL PERIODO BASE

Todos los índices requieren la selección de un período base como punto de referencia. Por razones que más adelante serán aclaradas, no es conveniente elegir como base un período caracterizado por marcadas irregularidades ni que se encuentre demasiado alejado en el tiempo.
Ambas condiciones representan importantes limitaciones a la hora de tomar decisiones, ya que no es fácil determinar cuándo un período es suficientemente “regular” como para constituir la base de un índice y ofrece muchas dudas definir el período a partir del cual la base se consideraría demasiado “alejada”.

2. INDICES SIMPLES O RELATIVOS

Un índice simple es la relación expresada como porcentaje entre el valor de una característica en el período de interés con respecto al valor de la misma característica en el período base.
En los comentarios siguientes dedicaremos especial atención a los precios, cantidades y valores globales como características relevantes del campo económico, sin embargo los mismos pueden generalizarse sin esfuerzos a otras áreas de interés particular.

- Precios

De acuerdo a la definición precedentemente anotada, si el precio de un artículo i, en el período t, se simboliza por p_{it}, y en el período base por p_{io}, tendremos que el índice de precios del artículo i del período t con respecto al período base será:

$$p_{i\,0}^{\,t} = \frac{p_{it}}{p_{i0}}$$

Resulta evidente que, por tratarse del resultado de un cociente entre cantidades expresadas en las mismas unidades, todo índice es un número sin unidades, es decir un “relativo”.
Como los índices siempre se expresan como tanto por ciento es frecuente omitir la multiplicación por cien en las fórmulas para el cálculo.
Resulta obvio que en la base el índice siempre vale 100 por tratarse del resultado de un cociente entre idénticas cantidades, multiplicado por 100.

Ejemplo 23: Supongamos que se compara el precio unitario de cierto artículo i, que es de \$ 1.500 en el período t, con el del período base que era de \$ 500. El índice simple de precio es:

$$p_{i\,0}^{\,t} = \frac{\$\,1.500}{\$\quad 500} = 3 \quad \text{, o sea 300\%}$$

o Cantidades

Análogamente, si la cantidad de un artículo i en el período t se simboliza por q_{it}, y en el período base por q_{io}, el índice simple de cantidades o "relativo" de cantidades del período t con respecto al período base será:

$$q_{i\,0}^{\,t} = \frac{q_{it}}{q_{i0}}$$

Ejemplo 24: En el período t se producen 150.000 unidades del mismo producto i, mientras que en el período base la producción fue de 200.000. El índice simple de cantidades es:

$$q_{i\,0}^{\,t} = \frac{150.000}{200.000} = 0{,}75 \quad \text{, o sea 75\%}$$

o Valor Global

Del mismo modo, si el valor global de cierta producción i (precio unitario por cantidad producida) en el período t se simboliza por $p_{it}\,q_{it}$, y en el período base por $p_{i0}\,q_{i0}$, el índice simple de valor global o "relativo" de valor global del período t con relación al período base será:

$$v_{i\,0}^{\,t} = \frac{p_{it}\cdot q_{it}}{p_{i0}\cdot q_{i0}}$$

Ejemplo 25: Con los datos anteriormente considerados, el índice simple de valor global del período t con respecto al período base es:

$$v_{i\,0}^{\,t} = \frac{1.500 \times 150.000}{500 \times 150.000} = 2{,}25 \quad \text{, o sea 225\%}$$

2.1. PROPIEDADES DE LOS ÍNDICES SIMPLES O "RELATIVOS"

2.1.1 Reversibilidad temporal

Dados los valores de una característica en los períodos t_2 con relación a t_1, así como del período t_1 con relación a t_2, la propiedad establece que:

$$I_{t_1}^{t_2} \cdot I_{t_2}^{t_1} = 1 \qquad \text{o sea que:}$$

$$I_{t_1}^{t_2} = \frac{1}{I_{t_2}^{t_1}} \qquad y \qquad I_{t_2}^{t_1} = \frac{1}{I_{t_1}^{t_2}}$$

Es decir que cuando los períodos se intercambian, sus correspondientes relativos son recíprocos entre sí. En el ejemplo que se venía considerando:

$$p_o = 500 \qquad y \qquad p_t = 1.500$$

Los relativos son:

$$p_0^t = \frac{1.500}{500} \quad y \quad p_t^0 = \frac{500}{1.500}$$

Resulta evidente que:

$$p_0^t \, . \, p_t^0 = \frac{1.500}{500} \cdot \frac{500}{1.500} = 1$$

2.1.2 Reversibilidad de factores

Considerando que p es el precio de un artículo en cierto período t y que q es la cantidad (producida, consumida etc.) en el mismo período. El producto $p\,.\,q$ es lo que llamamos valor global.

La propiedad establece que si consideramos un relativo de precios $\frac{p_{t_1}}{p_{t_2}}$ y otro de cantidades $\frac{q_{t_1}}{q_{t_2}}$, el producto de ambos es igual al relativo de valores globales:

$$v_{t_2}^{t_1} = \frac{p_{t_1}}{p_{t_2}} \frac{q_{t_1}}{q_{t_2}} = \frac{p_{t_1} \cdot q_{t_1}}{p_{t_2} \cdot q_{t_2}}$$

2.1.3 Circular

Dados tres períodos, t_1, t_2 y t_3, la propiedad establece que: $I_{t_1}^{t_2} \cdot I_{t_2}^{t_3} \cdot I_{t_3}^{t_1} = 1$

Es decir que si el precio de un artículo fuera en los tres períodos de $p_1, p_2 \; y \; p_3$

$$\frac{p_{t_2}}{p_{t_1}} \cdot \frac{p_{t_3}}{p_{t_2}} \cdot \frac{p_{t_1}}{p_{t_3}} = 1$$ lo que resulta evidente.

2.1.4 Cíclica

Si para dos períodos: t_2 y t_3, los índices simples son iguales con respecto a otro período t_1, es decir que:

$$I_{t_1}^{t_2} = I_{t_1}^{t_3}$$

El valor de la característica de interés es igual en t_1 y en t_3.

3. CAMBIO PORCENTUAL

Con relativa frecuencia se expresa la variación de un índice en términos absolutos. Por ejemplo si un índice pasa de 100 a 200 se dice que “aumentó 100 puntos”. Obsérvese que si en el período siguiente el mismo índice pasa de 200 a 300 también experimenta un “aumento de 100 puntos” pero la ***tendencia*** ahora es muy

diferente. En el primer caso se verifica un aumento del 100%, mientras que en el segundo el mismo es del 50%.
Para que la información sobre la variación adquiera un mayor significado, es conveniente expresarla siempre como tanto por ciento en relación al valor que se desea utilizar como referencia. Para ello se deberá aplicar la siguiente fórmula, denominada de “cambio porcentual” (cuando multiplicado por cien).

$$\Delta = \left(\frac{I_0^t}{I_0^{t-x}} - 1 \right)$$

Esta expresión surge del sencillo razonamiento de considerar que, para modificar una cantidad en cierto porcentaje de sí mismo (Δ), hay que realizar la siguiente operación:

$$I_0^{t-x} + \Delta\, I_0^{t-x} = I_0^t$$

siendo I_0^{t-x} un factor común, tenemos:

$$I_0^{t-x}(1+\Delta) = I_0^t$$

O sea que, para modificar una cantidad en un cierto porcentaje Δ, hay que multiplicarla por (1 + Δ), siendo Δ un número positivo o negativo.

Despejando:

$$1+\Delta = \frac{I_0^t}{I_0^{t-x}} \qquad \Delta = \left(\frac{I_0^t}{I_0^{t-x}} - 1 \right)$$

, con lo que llegamos a la fórmula del "cambio porcentual".

Volviendo al ejemplo considerado tenemos:

$$I_0^{t_0} = 100 \qquad I_0^{t_1} = 200 \qquad I_0^{t_2} = 300$$

t_0 t_1 t_2

$$\Delta_1 = \left(\frac{200}{100} - 1 \right) = 1$$

, que expresado como porcentaje es 100 %

$$\Delta_2 = \left(\frac{300}{200} - 1 \right) = 0{,}5$$

, o sea 50 %

Resulta interesante destacar que la variación total experimentada por el índice entre t_0 y t_2 no puede obtenerse por adición de las variaciones parciales, lo que llevaría a concluir que es de 150%. En realidad esa variación es de:

$$\Delta_T = \left(\frac{300}{100} - 1 \right) = 2$$

, o sea 200 %

Representación gráfica

Siempre apoyándonos en el mismo ejemplo, veamos primeramente una forma de representación que es frecuente encontrar pero que, como veremos, resulta inadecuada.

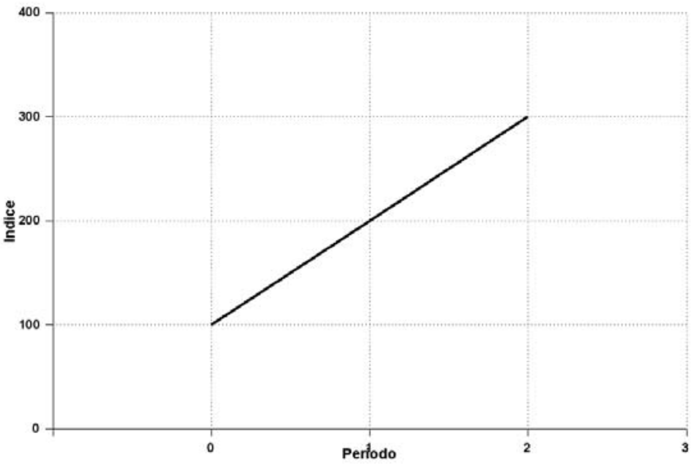


Figura 12

Obsérvese que el gráfico muestra que el índice tiene una tendencia al alza con idéntica pendiente en ambos períodos considerados. Sin embargo ya hemos calculado que en el segundo período se produjo una importante desaceleración en el ritmo de crecimiento, pasando del 100 % al 50 % . La escala natural utilizada no permite representar adecuadamente las variaciones realmente experimentadas.
En cambio, si utilizamos un gráfico semilogarítmico, representando en el eje de las ordenadas los logaritmos de los índices, tendremos:

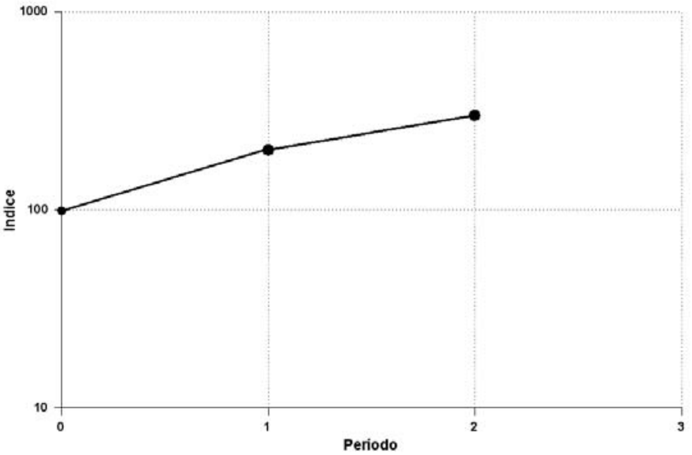


Figura 13

Esta representación, que supone una evolución de tipo exponencial para el índice, resulta perfectamente apropiada para describir la situación planteada. Obsérvese que la pendiente de la recta entre t_1 y t_2 muestra claramente que se enlentece el crecimiento del índice, en concordancia con los cálculos efectuados.
Consecuentemente, siempre es preferible representar a los números índice mediante gráficos de tipo semilogarítmico.

4. LA TASA PROMEDIO DE VARIACION

Vamos a suponer que disponemos de valores de un índice separados en el tiempo (n-r) unidades de período (años, meses, etc.) y que deseamos estimar la tasa promedio de variación por unidad de período (i).

$$I_0^{t_r}(1+i)^{n-r} = I_0^{t_n} \quad \Rightarrow \quad (1+i)^{n-r} = \frac{I_0^{t_n}}{I_0^{t_r}} \quad \Rightarrow$$

$$1+i = \sqrt[n-r]{\frac{I_0^{t_n}}{I_0^{t_r}}} \quad \Rightarrow \quad i = \sqrt[n-r]{\frac{I_0^{t_n}}{I_0^{t_r}}} - 1$$

Aplicando logaritmos tendríamos :

$$i = \text{antilog}\left(\frac{\log I_0^{t_n} - \log I_0^{t_r}}{n-r}\right) - 1$$

Ejemplo 26: En diciembre de 2001 el índice de los precios de cierto artículo alcanzó a 134,9. En diciembre de 2002, el mismo índice fue 164,8. Supongamos que deseamos calcular la tasa promedio de variación mensual para el período.

$134,9\ (1 + i)^{12} = 164,8$

$$(1+i)^{12} = \frac{164,8}{134,9} \qquad (1+i) = \sqrt[12]{\frac{164,8}{134,9}} = 1,0168$$

i ≈ 0,0168 o sea 1,68%

También utilizando logarítmos:

$$12 \log (1 + i) = \log\frac{164,8}{134,9}$$

log (1 + i) = 0,007245…

(1 + i) = 1,0168…→ i ≈ 0,0168 o sea 1,68%

5. BASE FIJA , BASE VARIABLE

Como se ha visto, el valor que se toma como denominador para el cálculo de un número índice, se denomina base. Supongamos que disponemos del precio de un bien en los siguientes períodos:

t_1 :500 t_2 :1.000 t_3 :1.050 t_4 :1.250

Tomando como base el precio en el período t_1 tendríamos los siguientes índices simples para los períodos considerados:

$$p_{t_1}^{t_1} = \frac{500}{500} \times 100 = 100 \qquad p_{t_1}^{t_2} = \frac{1.000}{500} \times 100 = 200$$

$$p_{t_1}^{t_3} = \frac{1.050}{500} \times 100 = 210 \qquad p_{t_1}^{t_4} = \frac{1.250}{500} \times 100 = 250$$

Si nos interesa cambiar de base, es decir, transformar los índices anteriores que tienen base en t_1, en índices que tengan base, por ejemplo, en t_2, bastará con dividirlos por el valor del índice en ese período y multiplicarlos por 100.

$$p_{t_2}^{t_1} = \frac{100}{200} \times 100 = 50 \qquad p_{t_2}^{t_2} = \frac{200}{200} \times 100 = 100$$

$$p_{t_2}^{t_3} = \frac{210}{200} \times 100 = 105 \qquad p_{t_2}^{t_4} = \frac{250}{200} \times 100 = 125$$

Este cambio de base aritmético obtenido mediante una transformación lineal, no debe confundirse con el cambio de base que requiere una investigación tendiente a determinar los pesos relativos de las características de interés en el caso de índices ponderados.
También puede interesar representar el relativo de precio de cada período, con respecto al período inmediatamente anterior en cuyo caso tendríamos:

$$p_{t_1}^{t_2} = \frac{1.000}{500} \times 100 = 200 \quad p_{t_2}^{t_3} = \frac{1.050}{1.000} \times 100 = 105 \quad p_{t_3}^{t_4} = \frac{1.250}{1.050} \times 100 = 119$$

El relativo para un período dado con respecto a otro considerado como base, puede expresarse en términos de estos relativos de la siguiente forma:

$$p_{t_1}^{t_4} = \frac{1.000}{500} \times \frac{1.050}{1.000} \times \frac{1.250}{1.050} \times 100 = \frac{1.250}{500} \times 100 = 250$$

Como se observa, el índice se obtiene mediante un encadenamiento de relativos, lo que justifica que reciba el nombre de cadena de relativos con respecto a la base elegida, o índices eslabonados.
Obviamente, estas consideraciones se extienden a los casos de cantidades o valores globales, así como a otras características de interés.

6. INDICES PONDERADOS

Hemos visto que, para describir la evolución de una característica relacionada con cierto bien a través del tiempo, podríamos construir un índice simple. Por ejemplo, si estamos interesados en las variaciones del precio de la leche, observaríamos ese precio en períodos regulares y efectuaríamos los cocientes correspondientes utilizando como denominador el precio del período elegido como base, que puede ser fijo o variable.
Muchas veces se desea conocer lo que ha pasado con los precios de un conjunto de bienes. Es decir de la leche, el pan, la carne, los huevos, etc.
Claro es que podríamos construir índices simples para cada uno de ellos, obteniendo un conjunto de relativos que difícilmente nos darían una idea general de la evolución, en función de las diferentes maneras con que estos precios varían.
Una forma de resumir la información considerada, es calcular un valor estadístico de la tendencia central de los diversos relativos de precios calculados, por ejemplo la media aritmética, o la media geométrica, o la media armónica, otorgándole un cierto peso o ponderación a cada uno de los relativos intervinientes.
Como hemos visto, los relativos resultan ser valores homogéneos al estar independizados de las unidades de medida de los diferentes artículos.
Lo que hacemos al calcular un número índice es sustituir un conjunto generalmente muy numeroso de observaciones correspondientes a un cierto período, por una sola cifra que describe, resume o caracteriza a la totalidad de la distribución.
El valor estadístico aceptado con generalidad a esos efectos es la media aritmética en cuya definición se apoyan las principales fórmulas para el cálculo de índices ponderados.

6.1 La media aritmética ponderada, como número índice

Vamos a definir un vector $\overrightarrow{X} = (X_1 \, X_2 \cdots X_m)$, constituido por variables, al que se asocia un vector de ponderaciones : $\overrightarrow{N} = (n_1 \, n_2 \cdots n_m)$, como indica la siguiente tabla:

X_i	n_i
X_1	n_1
X_2	n_2
…	…
X_m	n_m

Como se recuerda de Estadística Descriptiva, la media aritmética ponderada se define para el conjunto como:

$$\overline{X} = \frac{\sum_{i=1}^{m} X_i n_i}{\sum_{i=1}^{m} n_i}$$

En la fórmula se distinguen dos componentes: la variable X_i, y la ponderación n_i. Precisemos la definición de ambas de acuerdo con nuestro interés.
Definiremos la variable X_i como un cociente entre variables referidas a ciertos períodos de interés, es decir:

$$X_i = \frac{X_{it}}{X_{i0}}$$

Los precios, las cantidades, los valores globales -entre otros– son características medibles -esto es, variables– y a estos cocientes les hemos llamado índices simples o relativos.
La variable X_i asume las siguientes expresiones según se considere como razón entre precios, entre cantidades o entre valores globales:

$$X_i = \frac{X_{it}}{X_{i0}} = \begin{cases} \text{Precios} & : \dfrac{p_{it}}{p_{i0}} \\ \text{Cantidades} & : \dfrac{q_{it}}{q_{i0}} \\ \text{Valores Globales} & : \dfrac{p_{it}\, q_{it}}{p_{i0}\, q_{i0}} \end{cases}$$

En cuanto a la ponderación n_i, existen diferentes maneras de definirla. Entre ellas presentaremos a continuación las empleadas por Laspeyres y por Paasche en sus fórmulas tan ampliamente difundidas:

n_i	LASPEYRES	PAASCHE
Precios	$p_{i0}\, q_{i0}$	$p_{i0}\, q_{it}$

Cantidades	$p_{i0}\ q_{i0}$	$p_{it}\ q_{i0}$
Valores Globales	$p_{i0}\ q_{i0}$	$p_{i0}\ q_{i0}$

6.2 Fórmulas para el cálculo de Índices ponderados

Reemplazando en la definición de media aritmética ponderada y omitiendo registrar el subíndice i con respecto al que se efectúan las operaciones, para simplificar la notación, se obtiene:

INDICE	PRECIOS	CANTIDADES
LASPEYRES	$_L P_0^t = \dfrac{\sum \frac{p_t}{p_0} p_0 q_0}{\sum p_0 q_0} = \dfrac{\sum p_t q_0}{\sum p_0 q_0}$	$_L Q_0^t = \dfrac{\sum \frac{q_t}{q_0} p_0 q_0}{\sum p_0 q_0} = \dfrac{\sum p_0 q_t}{\sum p_0 q_0}$
PAASCHE	$_P P_0^t = \dfrac{\sum \frac{p_t}{p_0} p_0 q_t}{\sum p_0 q_t} = \dfrac{\sum p_t q_t}{\sum p_0 q_t}$	$_P Q_0^t = \dfrac{\sum \frac{q_t}{q_0} p_t q_0}{\sum p_t q_0} = \dfrac{\sum p_t q_t}{\sum p_t q_0}$
	VALOR GLOBAL	
	$V_0^t = \dfrac{\sum \frac{p_t q_t}{p_0 q_0} p_0 q_0}{\sum p_0 q_0} = \dfrac{\sum p_t q_t}{\sum p_0 q_0}$	

Como se observa, hay una única expresión para el índice de Valores Globales, dado que tanto Laspeyres como Paasche utilizan la misma ponderación.

Entre las diversas fórmulas existentes para el cálculo de números índice vamos a presentar además las que se deben a Fisher. Estas fórmulas se obtienen como media geométrica entre los índices de Laspeyres y de Paasche.

INDICE DE PRECIOS DE FISHER	INDICE DE CANTIDADES DE FISHER
$_F P_0^t = \sqrt{_L P_0^t \ _P P_0^t}$	$_F Q_0^t = \sqrt{_L Q_0^t \ _P Q_0^t}$

Dado que las fórmulas de Laspeyres, Paasche y Fisher son las más conocidas y utilizadas, serán las únicas que presentaremos aquí. No obstante, la variedad es prácticamente ilimitada, ya que se pueden considerar promedios y ponderaciones de tipos muy diferentes.
En la bibliografía especializada[7] se pueden citar, entre otras, las fórmulas de Marshall–Edgworth, Walsh, Keynes y Sidgwick–Drobisch.

Obviamente, si se utilizan fórmulas diferentes para el cálculo de un mismo índice los resultados no tienen por qué ser iguales. Aún utilizando una misma fórmula los resultados podrían diferir significativamente, ya que dependen de los recursos utilizados, del momento en que se concreta el relevamiento de los datos, y del sistema de crítica y elaboración de la información. En el caso de que se construyeran índices paralelos

7 FISHER, Irving. “The Making of Index Numbers”1922. El autor describe 134 fórmulas distintas para el cálculo de números índice.

debería tomarse como algo natural que los resultados difieran, aunque las tendencias deberían guardar razonable similitud.

Ejemplo 27: Los datos siguientes corresponden a cantidades consumidas de diversos artículos y sus respectivos precios en los períodos indicados de un año dado.

Artículo	Unidad	Enero		Junio	
		P	q	p	q
Cardigan	Unidad	43,0	4	48,4	3
Camisa	Unidad	8,0	3	8,0	4
Calcetines	Par	3,5	1	3,6	2
Tela	Metro	4,0	3	5,0	2
Zapatos	Par	46,0	1	55,0	1

Se desea calcular el índice de precios de Laspeyres, de Paasche y de Fisher para junio con base en enero.

$p_e\, q_e$	$p_j\, q_e$	$p_j\, q_j$	$p_e\, q_j$
172.0	193.6	145.2	129.0
24.0	24.0	32.0	32.0
3.5	3.6	7.2	7.0
12.0	15.0	10.0	8.0
46.0	55.0	55.0	46.0
257.5	291.2	249.4	222.0

$${}_{L}P_{e}^{j} = \frac{291{,}2}{257{,}5} \times 100 = 113{,}1\% \quad ; \quad {}_{P}P_{e}^{j} = \frac{249{,}4}{222{,}0} \times 100 = 112{,}3\% \quad ; \quad {}_{F}P_{e}^{j} = \sqrt{113{,}1 \times 112{,}3} = 112{,}7\%$$

6.3 Examen de las formulas presentadas

Desde el punto de vista teórico sería ventajoso que las fórmulas para el cálculo de índices ponderados cumplieran con las mismas propiedades señaladas para el caso de los índices simples o relativos. Sin embargo, el hecho de no cumplirlas no significa una razón suficiente como para invalidar, ni tan siquiera limitar su utilización, en especial si reúnen otros méritos de tipo práctico y operativo o de adecuación a la información disponible.

En cuanto a la propiedad de "reversibilidad temporal" se puede comprobar fácilmente que las fórmulas de Laspeyres y de Paasche no tienen por qué cumplirla. En cambio las de Fisher satisfacen este requisito, como puede observarse a continuación para el caso de precios:

$${}_{F}P_{t_1}^{t_2} = \sqrt{{}_{L}P_{t_1}^{t_2} \cdot {}_{P}P_{t_1}^{t_2}} = \sqrt{\frac{\sum p_{t_2} \cdot q_{t_1}}{\sum p_{t_1} \cdot q_{t_1}} \cdot \frac{\sum p_{t_2} \cdot q_{t_2}}{\sum p_{t_1} \cdot q_{t_2}}}$$

$${}_{F}P_{t_2}^{t_1} = \sqrt{{}_{L}P_{t_2}^{t_1} \cdot {}_{P}P_{t_2}^{t_1}} = \sqrt{\frac{\sum p_{t_1} \cdot q_{t_2}}{\sum p_{t_2} \cdot q_{t_2}} \cdot \frac{\sum p_{t_1} \cdot q_{t_1}}{\sum p_{t_2} \cdot q_{t_1}}}$$

$$ {}_{F}P_{t_1}^{t_2} \quad {}_{F}P_{t_2}^{t_1} = 1 $$

Lo mismo ocurre en el caso del índice de cantidades de Fisher.

Con respecto a la segunda propiedad vista, la de "reversibilidad de factores", resulta inmediata la comprobación de que solo las fórmulas de Fisher las cumplen:

$$ {}_{F}P_0^t \cdot {}_{F}Q_0^t = \sqrt{\frac{\sum p_t \cdot q_0}{\sum p_0 \cdot q_0} \cdot \frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_t}} \cdot \sqrt{\frac{\sum p_0 \cdot q_t}{\sum p_0 \cdot q_0} \cdot \frac{\sum p_t \cdot q_t}{\sum p_t \cdot q_0}} = $$

$$ = \sqrt{\frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_0} \cdot \frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_0}} = \frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_0} = V_0^t $$

No obstante, cundo la información disponible lo permite, se puede calcular un índice de precios mediante Paasche y un índice de cantidades mediante Laspeyres, en cuyo caso se verifica que:

$$ {}_{P}P_0^t \cdot {}_{L}Q_0^t = \frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_t} \cdot \frac{\sum p_0 \cdot q_t}{\sum p_0 \cdot q_0} = \frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_0} = V_0^t $$

Es decir que, mediante una combinación de índices es posible cumplir con el criterio de reversibilidad de factores. Análogamente, es posible demostrar que:

$$ {}_{L}P_0^t \cdot {}_{P}Q_0^t = V_0^t $$

La propiedad "cíclica" (cuando un índice toma el mismo valor en dos períodos diferentes, en los que todos los valores de la característica de interés son iguales) se cumple para todo índice en el cual las ponderaciones son fijas a lo largo del tiempo, como es el caso de los índices de Laspeyres calculados sobre una base constante. El índice de Paasche utiliza ponderaciones variables y por lo tanto no tiene por qué cumplir esta propiedad, lo mismo que el de Fisher, que lo incluye en su cálculo.

La propiedad "circular" vista para los índices simples o relativos, no se cumple para ninguna de las fórmulas consideradas, lo que no resulta ningún obstáculo para que el propio Fisher denomine a sus fórmulas como "ideales" por ser las que cumplen con un mayor número de propiedades.

Para la elección de la fórmula de cálculo, juega un papel decisivo la información disponible. Obsérvese por ejemplo que, para el cálculo del índice de los precios del consumo, se requiere determinar una canasta básica de bienes y servicios mediante una investigación muy compleja y costosa. Con los recursos disponibles en el país, sería muy difícil disponer de las cantidades disponibles en cada período, para poder utilizar la fórmula de Paasche. En cambio, tratándose del comercio exterior, en que las cantidades de los bienes importados y exportados están disponibles período a período, el obstáculo señalado sería inexistente.

7. CAMBIO DE BASE

La elección del período base debe tratarse con bastante atención. Supongamos que en el período t_1 el precio de un bien es 5 y en el período t_2 es 10. Si construimos un índice simple tomando como base t_1, el resultado será 200. En cambio si el período base es t_2, el índice será 50.

Sicológicamente, estos resultados -ambos correctos- no son equivalentes, dado que existe la tendencia natural a considerar que el período base es un período “normal”.
La situación de anormalidad extrema es, en muchas oportunidades, fácil de detectar. Es claro que, por ejemplo, a nadie se le ocurriría situar la base de un índice de productos agropecuarios en un período caracterizado por una gran sequía o, en el otro extremo, por condiciones extraordinariamente favorables para los cultivos.
En cambio, la situación de normalidad es bastante más difícil de determinar, en especial cuando se carece de la perspectiva histórica necesaria, o no existe información suficiente.
En los casos en que una decisión sobre el particular resulta más problemática, es cuando la construcción de la base de un índice requiere de importantes recursos (consumo, salarios, etc.). Muchas veces la investigación necesaria se hace cuando los recursos están disponibles, o cuando la base anterior se considera obsoleta.
Una posibilidad muchas veces explotada, es la de situar la base de un índice en un promedio calculado entre varios períodos, lo que obviamente “suaviza” su posición con respecto a las observaciones extremas.
La otra recomendación clásica sobre el período base –además de la de “normalidad”– es la que se refiere a su ubicación no muy alejada en el tiempo. Este requisito adquiere real significado en el caso de índices ponderados (IPC, IMS, etc.).
Una de las razones está dada por el hecho que, en los índices de base fija con ponderaciones constantes (por ejemplo el índice de los precios del consumo, que utiliza la fórmula de Laspeyres), presuponen que hay una “canasta” de bienes y servicios cuya composición se determina en el período base que permanece constante a través del tiempo.
Se sabe que, a la larga, no es así. En cuanto a los bienes de consumo hay que considerar que las preferencias de los consumidores cambia. Hay bienes que pasan a ser sustituidos por otros o simplemente desaparecen del mercado. Esto obviamente va afectando las ponderaciones, que luego de un período prolongado deben ser revisadas y corregidas mediante una nueva investigación que actualice la “canasta”, determinando una nueva base para el índice.
Otro argumento importante está dado por el comportamiento de la distribución de los relativos cuyo promedio es el índice. Se puede comprobar que, en el corto plazo, las frecuencias de las variaciones presentan una concentración importante alrededor del promedio. Pero a medida que el tiempo transcurre la dispersión aumenta y las variaciones promedio se van haciendo menos frecuentes, lo que debilita las posibilidades del índice como valor estadístico de resumen.
A la larga, los precios se comportan de manera bien diferente, resultando imprescindible actualizar la base luego de un tiempo razonable.
El plazo a partir del cual una base puede considerarse obsoleta, no puede establecerse con generalidad, porque depende de la variabilidad de la característica investigada. En el caso del Índice de los Precios del Consumo, no es conveniente utilizar una misma base por más de 10 años.

8. EMPALME DE LAS SERIES

Como hemos visto, después de calcular durante cierto tiempo -de 5 a 10 años dependiendo de las características de interés– un índice con respecto a una misma base fija, es necesario cambiar de base. Se obtiene entonces un nuevo índice que es necesario “empalmar” con el anterior a efectos de mantener la continuidad de la serie. La situación que se presenta se ilustra con el siguiente ejemplo:
Ejemplo 28:

Período	Índice		Índice empalmado	
	Base t_1 = 100	Base t_5 = 100	Hacia atrás	Hacia adelante
t_1	100		94	
t_2	102		96	
t_3	104		98	
t_4	103		97	
t_5	106	100	100	106
t_6		103	103	109

Para recalcular la serie “hacia atrás” se establece la proporción:

para t_4 $\qquad \frac{x}{100} = \frac{103}{106} \quad \Rightarrow \quad x = 97$

para t_3 $\qquad \frac{x}{100} = \frac{104}{106} \quad \Rightarrow \quad x = 98$, y así sucesivamente.

Para empalmar la serie "hacia delante" el procedimiento es análogo.

9. DEFLACTACION

Las ventas comerciales, los sueldos y salarios, las utilidades, los presupuestos fiscales, varían período a período por diferentes causas. Una de ellas es el aumento de los precios.
El procedimiento que tiende a eliminar el efecto del aumento de precios, con el fin de permitir comparaciones en términos reales, recibe comúnmente el nombre de deflactación.
Para deflactar un valor nominal, se requiere disponer de un índice deflactor, es decir, un indicador de las alteraciones de los precios directamente relacionado con dicho valor.
Hay que aclarar que esto frecuentemente no es posible, ya que los índices de que se dispone, son índices "generales" que en realidad no sirven bien a los usuarios, con fines específicos de deflactación.
Es necesario comprender que se debe contar con el mayor número posible de índices específicos, a fin de evitar el uso indiscriminado de ciertos indicadores -el IPC por ejemplo– para emplearlos como deflactores en situaciones que poco o nada tienen que ver con ellos.
La deflactación consiste en dividir los montos monetarios nominales por el índice de precios elegido como deflactor. De esta forma se obtendrá un valor expresado en unidades monetarias que tienen un poder adquisitivo del año que es la base del índice deflactor.

$$\frac{\text{valor}_t}{\text{valor}_0} = \frac{P_0^t}{100} \quad \Rightarrow \quad \text{valor}_0 = \frac{\text{valor}_t}{P_0^t} \times 100$$

Ejemplo29: El monto de los sueldos y salarios pagados en cierta economía entre 1994 y 1998, en miles de millones de unidades monetarias son los siguientes:

1994	1995	1996	1997	1998
325,5	579,8	985,4	1.236,5	1.659,8

En esa economía, se dispone de un índice de los precios del consumo con base en 1998 = 100 , que presenta los siguientes valores en el período considerado:

1994	1995	1996	1997	1998
21,5	38,0	63,2	79,4	100,0

Supongamos que deseamos expresar la serie de sueldos y salarios en valores de 1994, usando dicho índice como deflactor.
El primer paso consiste en realizar un cambio aritmético de la base, llevándola a 1994 = 100.

1994	1995	1996	1997	1998
100,0	176,7	294,0	369,3	465,1

Efectuando los cocientes correspondientes y multiplicando por cien, se obtiene la serie en valores de 1994.

1994	1995	1996	1997	1998
325,5	328,1	326,0	334,8	356,9

10. ALGUNAS APLICACIONES DE INTERÉS

11.1 El índice de los precios del consumo (IPC)[8]

La simple inspección de la fórmula del índice de precios elaborada por Laspeyres, que es la que comúnmente se utiliza para el cálculo del IPC, permite observar que representa la relación entre los precios de una canasta de bienes y servicios en un período t con relación a los precios de un período base, a los que se adquiere una misma canasta de bienes y servicios del consumo.

$$ {}_{L}P_{0}^{t} = \frac{\sum \frac{p_t}{p_0} p_0 q_0}{\sum p_0 q_0} = \frac{\sum p_t q_0}{\sum p_0 q_0} $$

En la práctica se utiliza la siguiente expresión equivalente:

$$ {}_{L}P_{0}^{t} = \sum w_0 \cdot \frac{p_t}{p_0} \qquad \text{donde} \qquad w_0 = \frac{p_0 \cdot q_0}{\sum p_0 \cdot q_0} $$

Como se observa w_0 es la importancia relativa de cada artículo en el gasto total, en el período base.

La canasta de bienes y servicios consumida en promedio por los hogares, se determina mediante una encuesta por muestreo a los hogares representativos de los diferentes estratos socioeconómicos, llevada a cabo en el área de referencia, en el período que se elige como base. Hasta el presente las encuestas de gastos e ingresos de los hogares llevadas a cabo en el país con dicho objetivo (entre otros), han tenido frecuencia aproximadamente decenal, pero existe la opinión generalizada de que en función de la rapidez de los cambios que se vienen experimentando en el consumo de los hogares, en el futuro será necesario reducir dicho período, o bien llegar a disponer de información continua sobre el particular.

La encuesta, tendiente a captar la información sobre todos los bienes y servicios consumidos por los hogares, se lleva a cabo durante todo un año, a efectos de contemplar las variaciones estacionales producidas en el consumo. Una vez que la encuesta haya sido completada, se procede a seleccionar los bienes y servicios representativos y a determinar las ponderaciones y los diferentes niveles de clasificación.

La recolección de los precios de todos los bienes que componen la canasta, debe efectuarse en forma continua, en una muestra representativa de los puestos de venta del área de referencia. El número de precios necesarios para cada artículo se determina en función de la variabilidad observada en el período base.

10.2. El índice de salarios (IMS)

El objetivo del IMS calculado en nuestro país, es estimar las variaciones experimentadas en los ingresos corrientes de los trabajadores comprendidos en los grandes sectores de las actividades públicas y privadas,

[8] Ver la metodología detallada del IPC en www.ine.gub.uy

que son remunerados exclusivamente en dinero, excluyendo los regímenes de ocupación estacional o zafral, los trabajadores rurales y los ingresos por pasividades.

Las fuentes de información para el cálculo del índice son por un lado las empresas privadas y por otro, las unidades ejecutoras del gobierno central, las intendencias municipales y las empresas públicas. Cada una de las unidades informantes seleccionadas brinda información mes a mes de todos los trabajadores pertenecientes a las categorías elegidas como representativas.

Tanto en el sector público como en el privado las unidades informantes se seleccionan en forma deliberada, en función del monto total de las remuneraciones pagadas.

Para el cálculo del índice medio de salarios también se utiliza la fórmula de Laspeyres. El precio utilizado es la remuneración por hora ordinaria, en cada categoría y en cada empresa u organismo. El precio promedio en cada categoría se obtiene promediando la remuneración ordinaria por hora a través de todas las empresas u organismos de la muestra, en cada rama o inciso.

10.3. El índice de salarios reales (ISR)[9]

Deflactando el índice medio de salarios por el índice de los precios del consumo, se obtiene el índice de salarios reales.

$$ISR_0^t = \frac{IMS_0^t}{IPC_0^t}$$

Obsérvese que si el IMS y el IPC no se encuentran expresados en la misma base, será necesario proceder al cambio de base de al menos uno de ellos, para que el ISR resulte expresado en una base común a ambos.

Este índice expresa el poder adquisitivo de los salarios. Si el numerador crece más que el denominador, se estará ganando poder adquisitivo. Si en cambio el denominador es el que crece más, se estará perdiendo poder adquisitivo.

Ejemplo 30:

Se dispone de los siguientes valores del índice de salarios (IMS) y del índice de precios de los artículos del consumo (IPC) (Datos ficticios). Se pide calcular para los mismos meses, los índices de salarios reales con base en diciembre de 1999=100.

	IMS Base dic 1999=100	IPC Base junio de 2000=100
Dic 1999	100.0	90.5
Junio 2000	105.0	100.0
Dic 2000	110.0	118.3

[9] Ver la metodología detallada del IMS en www.ine.gub.uy

Luego de expresar la serie del IPC con base en 1999=100 (100.0; 110.5; 130,7) Se realiza el cociente $ISR_{99}^{t} = \dfrac{IMS_{99}^{t}}{IPC_{99}^{t}}$; obteniéndose la serie pedida: (100.0; 95.0; 84.2).

LABORATORIO 5

1. Se dispone de la siguiente información sobre precios y cantidades de tres artículos:

	1997		1999	
ARTÍCULO	p	q	p	q
A	1	3	1.5	3
B	2	2	8	2
C	3	1	2	

1.1 Calcule el Indice de Precios para 1999 con base en 1997=100, utilizando la fórmula de Laspeyres (225).
1.2 Determine el cambio porcentual de 1999 con relación a 1997 (125%).
1.3 Determine la tasa promedio de variación anual entre 1997 y 1999 (50%).

1.4 Bajo el supuesto de que esta tasa promedio se mantenga, estime el Indice para el año 2005 (2562.89).

1.5 Si se sabe que el cambio porcentual de 1997 con relación a 1996 había sido del 20 por ciento, determine cuánto varió el Indice de 1999 con relación a 1996 (170%).

2. Cierto índice de producción es 147. Se desea elevarlo en un 60 por ciento en el término de 10 años. Si a los 6 años, el cálculo del mismo índice revela que alcanzó un valor de 191. Se pide:

   2.1 Determinar a que tasa promedio de variación anual ha crecido la producción en esos 6 años (4.46%).

   2.2 A qué tasa promedio anual debería crecer en el lapso que falta para los 10 años, a efectos de cumplir con el plan original (5.34%).

3 Un índice calculado en diciembre de 1995 es 210.4. Este valor revela un aumento de 14.6 por ciento con respecto a diciembre de 1992, y una disminución de 14.6 por ciento con respecto a diciembre de 1993.

   3.1 ¿Cuál es el cambio porcentual experimentado por el índice en diciembre de 1993 con respecto a diciembre de 1992? (34.2%).

   3.2 ¿Qué tasa promedio de variación anual, se registró entre diciembre de 1992 y diciembre de 1995? (4.6%).

4 Como la base de un índice se encuentra demasiado alejada en el tiempo, se realizan los trabajos necesarios para situarla en 1996=100. A continuación se presenta la serie del índice con la base antigua y con la base nueva:

	1994	1995	1996	1997	1998	1999
Base antigua	240	270	300			
Base nueva			100	120	125	170

   4.1 Presente la serie empalmada: de acuerdo a la nueva base y de acuerdo a la base antigua.

   4.2 Sitúe la base en 1994 y presente la serie resultante. (100 112.5 125 150 156.3 212.5)

5. La siguiente información corresponde a los índices de salarios y de precios de un país:

AÑO	MES	Indice Medio de Salarios (IMS) Base dic. 1994=100	Indice de los precios del Consumo (IPC) Base dic. 1983=100
1994	dic.	100	15.485
1995	dic.	113	16.959
1996	dic.	233	31.274
1997	dic.	408	52.508

Determine el Índice de Salario Reales en las mismas fechas, con base en dic. de 1994=100, y calcule la variación que experimentó en diciembre de 1997 con relación a diciembre de 1995. ¿Qué significa en términos del poder adquisitivo de los asalariados? (100 103.2 115.4 120.3) (Ganancia del 16.6%)

LABORATORIO 6

1. Se dispone de las siguientes series de Índice Medio de Salarios (IMS) y de Índice de los Precios del Consumo (IPC):

ÍNDICE	Dic-02	Ene-03	Feb-03	Mar-03	Abr-03
IMS	100.00	101.41	101.73	102.21	102.56
IPC	170.15	173.33	175.68	177.86	179.55

1.1 Si el Gobierno espera un crecimiento del 19% para el IPC durante el año 2003: ¿Cuál debería ser desde mayo la variación media mensual del IPC, para cumplir con dicha meta? (1,514%)

1.2 ¿Cuánto debería crecer en total desde abril de 2003 hasta fin de año el IMS para mantener el mismo poder adquisitivo de diciembre de 2002, si se supone que se cumple la meta de 1.1? (16,03%)

2. El Índice de Salarios de un país, con base en diciembre de 1998=100, toma los siguientes valores (datos ficticios):

Dic. 1998	Ene. 1999	Feb. 1999	Mar. 1999	Abr. 1999
100.0	105.2	107.1	110.4	110.6

Se sabe que el Índice de los Precios del Consumo, con base en febrero de 1999=100, había subido en los mismos meses a una tasa constante del 2 por ciento mensual.

2.1 Construya la serie del Índice de Precios del Consumo, utilizando una sola cifra decimal. (96.1; 98.0; 100.0; 102.0; 104.0)

2.2 Determine la serie del Índice de Salarios Reales para esos meses, con base en diciembre de 1998=100. (100.0; 103.1; 102.9; 104.1; 102.2)

2.3 ¿Cuál ha sido en términos porcentuales, la variación del poder adquisitivo de los salarios de abril de 1999 con relación a diciembre de 1998? (2.2 %)

3. Dadas las siguientes series de índices de precios del consumo (IPC) y de salarios (IMS):

PERÍODO	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
IMS	100	110	120	150	200	210	215	220	225	230	240
IPC	90	100	105	125	160	175	180	200	210	215	220

3.1 Determine la serie del índice de salarios reales (ISR) para esos años, con base en 1990=100. (*100.0 99.0 102.9 108.0 112. 108.0 107.5 99.0 96.4 96.3 98.2*)

3.2 Si en 2001 con relación a 2000, el ISR creció un 5 por ciento, y en 2002 bajó un 3 por ciento con relación a 2001 ¿cuál habrá sido el valor de ese índice en 2002, con base en 1990=100? (100.0)

NOTA DE CLASE Nº 5

INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES. Modelos matemáticos determinísticos y no determinísticos. Condiciones para la asignación de probabilidades. Definiciones de probabilidad. Leyes básicas para el cálculo. Algunas aplicaciones elementales. Teorema de la probabilidad total. Teorema de Bayes.

INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES

La observación cuidadosa del comportamiento del mundo real, y la reproducción intencionada de situaciones empíricas con fines experimentales, permiten comprobar diferentes tipos de "regularidades" en los resultados que se producen a partir de condiciones iniciales esencialmente iguales. Estas "regularidades" hacen posible la construcción de modelos matemáticos, que pueden ser utilizados para el trabajo, como herramientas capaces de efectuar predicciones razonables.

1. MODELOS MATEMÁTICOS DETERMINÍSTICOS

Supongamos que al repetir un experimento en condiciones esencialmente iguales, se logran aproximadamente los mismos resultados. Dado un cierto estado inicial conocido, se podrá entonces predecir el estado final con razonable certeza.
Los modelos matemáticos construidos para explicar estas situaciones empíricas reciben el nombre de *MODELOS MATEMATICOS DETERMINISTICOS*. Entre otros ejemplos de este tipo de modelos, se puede citar el caso de las leyes de Física. Por ejemplo: un modelo como: *Fuerza* = *Masa* $\times$ *Aceleración*, permite efectuar predicciones satisfactoriamente comprobables a través de la experimentación.

2. MODELOS MATEMÁTICOS NO DETERMINÍSTICOS

Es evidente, que los modelos determinísticos no pueden ser aplicados a los experimentos en que los resultados tienen un comportamiento tan irregular, que resulta imposible predecir uno en particular con razonable certeza. Estos experimentos, reciben el nombre de *EXPERIMENTOS ALEATORIOS*
Un ***experimento aleatorio***, cumple con las siguientes condiciones:

- Se puede repetir en condiciones esencialmente iguales.
- Se conocen todos los resultados posibles.
- Un resultado individual no se puede predecir con certeza.

Resulta claro que las mencionadas condiciones se cumplen para los experimentos aleatorios clásicos, relacionados con los juegos de azar; entre otros:

"Lanzar una moneda"
"Lanzar un dado"
"Extraer una carta"

La preferencia por este tipo de ejemplos, se debe a su sencillez y a su capacidad para representar en forma resumida, infinidad de otros ejemplos de la realidad que también se encuentran afectados por esa fuerza aparentemente oculta y misteriosa, a la que comúnmente se llama "el azar".

Sin embargo, los resultados de este tipo de experimentos también pueden ser explicados a través de modelos matemáticos. Estos modelos reciben en general la denominación de *MODELOS MATEMÁTICOS NO DETERMÍNISTICOS, PROBABILÍSTICOS O ESTOCÁSTICOS*. El conjunto de los modelos matemáticos que se construyen para explicar los experimentos aleatorios y sus resultados se denomina: ***"El Cálculo de Probabilidades"***.

3. ALGUNAS DEFINICIONES DE INTERÉS PARA EL CÁLCULO DE PROBABILIDADES.

3.1 Espacio muestral: [10]

Es el conjunto de todos los resultados posibles de un experimento aleatorio.

Deberán cumplirse las siguientes condiciones:

i) Todos los elementos del espacio muestral representarán resultados del experimento aleatorio.

[10]Nos referimos con exclusividad a Espacios Muestrales Finitos, aquellos en que sus elementos pueden contarse. Más precisamente: $\Omega \text{ finito} \Leftrightarrow \Omega \approx \emptyset, \text{ o } \Omega \approx \mathrm{N}_n = \{1, 2, \ldots n\}$ donde N_n: segmento de los números naturales (enteros positivos).

ii) Los resultados del experimento aleatorio estarán representados una sola vez en el espacio muestral.

Ejemplo: Experimento aleatorio: "Lanzar un dado y observar la cara superior".

Espacio muestral: Ω = {1, 2, 3, 4, 5, 6}

Resulta evidente que el conjunto cumple con las condiciones señaladas por lo que es un espacio muestral asociado al experimento aleatorio propuesto.

Por comodidad utilizaremos el mismo ejemplo para ilustrar los puntos presentados a continuación.

3.2 Sucesos o eventos:

Son todos los subconjuntos del espacio muestral. Se dice que ocurre el suceso A si el resultado del experimento aleatorio corresponde a un elemento del subconjunto A.

Ejemplo: "Obtener par": A_1 = {2, 3, 4}

"Obtener un 6": A_2 = {6}

Si A es el suceso de interés, su presentación será un "éxito", si no se presenta será un "fracaso".

3.3 Clasificación general de sucesos o eventos:

i) Suceso simple o elemental:

Es un subconjunto del espacio muestral, constituido por un único elemento.

Ejemplo: "Obtener 1" A = {1}

ii) Suceso imposible:

Es el conjunto vacío como subconjunto de Ω. Recordemos que el conjunto vacío es subconjunto de todo conjunto y es único.

Ejemplo: "Obtener un 7" $\varnothing = \{ \cdot \}$

iii) Suceso cierto o seguro:

Es el espacio muestral como subconjunto de sí mismo.

Ejemplo: "Obtener número menor igual que 6"

A = {1, 2, 3, 4, 5, 6} = Ω

iv) Suceso contrario:

Es el conjunto complementario de A, o sea A^c.

Ejemplo: "Obtener par" A = {2, 4, 6}

Suceso contrario A^c ={1, 3, 5}

v) Suceso compuesto:

Dados dos o más sucesos, es la ocurrencia simultánea de todos y cada uno de ellos. Corresponde a la operación de intersección de conjuntos.

Ejemplo: "Obtener par y al menos 3 simultáneamente"

A = {2, 4, 6} B = {3, 4, 5, 6}

$A \cap B$ = {4, 6}

Obsérvese que la conjunción "y" se asocia a la intersección de conjuntos.

vi) Sucesos mutuamente excluyentes o incompatibles:

Dos o más sucesos son incompatibles, cuando no pueden ocurrir simultáneamente.

Ejemplo: "Obtener par y obtener impar simultáneamente".

A = {2, 4, 6} B = {1, 3, 5}

$A \cap B = \emptyset$

vii) Suceso total:

Dados dos o más sucesos, es la ocurrencia de al menos uno de ellos. Corresponde a la operación de unión de conjuntos.

Ejemplo: "Obtener par o al menos 3"

A = {2, 4, 6} B = {3, 4, 5, 6}

$A \cup B = \{2, 3, 4, 5, 6\}$

Obsérvese que la conjunción "o" se asocia a la unión de conjuntos.

4. CONDICIONES PARA LA ASIGNACION DE PROBABILIDADES

Dado un espacio muestral $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, si hacemos corresponder a cada suceso simple de Ω un número real $p(\omega i)$, de tal forma que se cumplan las siguientes condiciones:

4.1 $p(\omega_i) \geq 0$

4.2 $\sum_{i=1}^{n} p(\omega_i) = 1$

Se dice que $p(\omega_i)$ es la probabilidad del suceso simple ω_i. Por lo tanto: $0 \leq p(\omega_i) \leq 1$.

5. DEFINICIONES DE PROBABILIDAD

5.1 "A posteriori":

Si la sucesión de las frecuencias relativas obtenidas para valores crecientes del número de ensayos converge a un límite, ese límite es la probabilidad de éxito en un solo ensayo.

Ejemplo: Determinar la probabilidad de germinación de una semilla.

Se requiere obtener información sobre las frecuencias relativas con que se presenta el suceso de interés. Si n(A) es el número de germinaciones obtenidas en $n(\Omega)$ ensayos, se tiene para valores crecientes de n:

$$\left[\frac{n(A)}{n(\Omega)}\right]_1 = h_1 \quad ; \quad \left[\frac{n(A)}{n(\Omega)}\right]_2 = h_2 \quad ; \cdots ; \quad \left[\frac{n(A)}{n(\Omega)}\right]_m = h_m$$

Si se verifica la existencia de un cierto valor al que convergen frecuencias relativas, a medida que aumente el número de ensayos, ese valor es la probabilidad de éxito en un solo ensayo.

Si observamos que las frecuencias relativas convergen a 0,80, diremos que la probabilidad de germinación de una semilla es 0,80, o bien el ochenta por ciento.

Para la determinación de la probabilidad en este caso, se requiere entonces apoyo imprescindible en la experiencia. A los efectos prácticos bastará con considerar como mejor aproximación de la probabilidad, la frecuencia relativa obtenida para el mayor número de ensayos realizados.

5.2 "A priori":

Si un suceso A de interés puede presentarse de n(A) formas, y dejar de presentarse de n(A^c) formas, y si cualquiera de esas n(Ω) formas son igualmente posibles de aparecer, la probabilidad de ocurrencia de A es:

$$P(A)=\frac{n(A)}{n(\Omega)}=\frac{\textit{Número de casos favorables}}{\textit{Número de casos posibles}}$$

Para determinar la probabilidad de un suceso A en el caso de los sucesos equiposibles o equiprobables, no es entonces necesario recoger datos acerca de como se distribuyen las frecuencias, bastará con efectuar el cociente entre el número de casos favorables y el número de casos posibles.

Ejemplo: Determinar la probabilidad de obtener un 6 en el lanzamiento de un dado correctamente construido.

A = {6} n(A) = 1

Ω = {1, 2, 3, 4, 5, 6} n(Ω) = 6

$$p(A)=\frac{n(A)}{n(\Omega)}=\frac{1}{6}$$

De todos modos si recogiéramos datos sobre las frecuencias relativas mediante un creciente número de ensayos, veríamos que las mismas tenderían al valor límite 1/6, por lo que la primera definición resulta ser de carácter general.

6. DOS LEYES BÁSICAS PARA EL CÁLCULO DE PROBABILIDADES

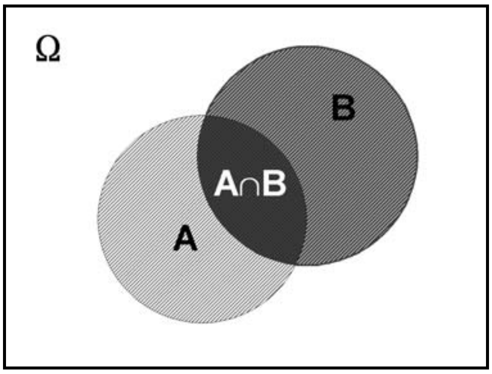
6.1 Ley de suma

Dados dos sucesos A y B; la probabilidad de que ocurra A o B o ambos es:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

La comprobación de esta igualdad es inmediata, aplicando la definición de probabilidad a priori y conocimientos elementales de teoría de conjuntos.

$$\begin{aligned}P(A\cup B) &= \frac{n(A\cup B)}{n(\Omega)}\\ &= \frac{n(A)+n(B)-n(A\cap B)}{n(\Omega)}\\ &= \frac{n(A)}{n(\Omega)}+\frac{n(B)}{n(\Omega)}-\frac{n(A\cap B)}{n(\Omega)}\\ &= P(A)+P(B)-P(A\cap B)\end{aligned}$$



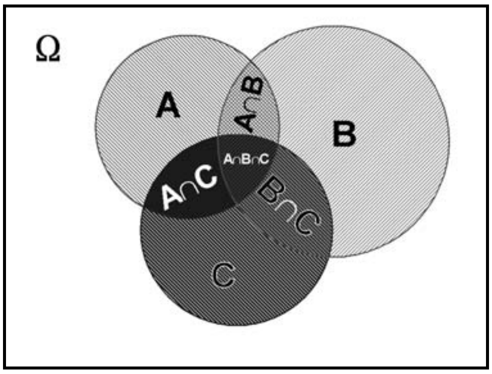
Ejemplo: Calcule la probabilidad de obtener Rey o Copa en una extracción al azar, de un mazo de 48 cartas españolas.

A: “Obtener Rey” ; B: “Obtener Copa”

$$P(A\cup B)=\frac{4}{48}+\frac{12}{48}-\frac{1}{48}=\frac{15}{48}=\frac{5}{16}$$

La extensión de esta ley a tres sucesos A, B y C es la siguiente:

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$



Ejemplo: En una población, los adultos mayores declaran informarse a través de los diferentes medios de comunicación en las siguientes proporciones:

Televisión: 70%

Radio: 40%

Diarios: 10%

Los que ven televisión y escuchan la radio, son el 21%. Los que ven televisión y leen el diario, son el 2,8%. Los que escuchan la radio y leen el diario son el 9%. Sólo el 2% de la población de adultos mayores, se informa mediante los tres medios simultáneamente.

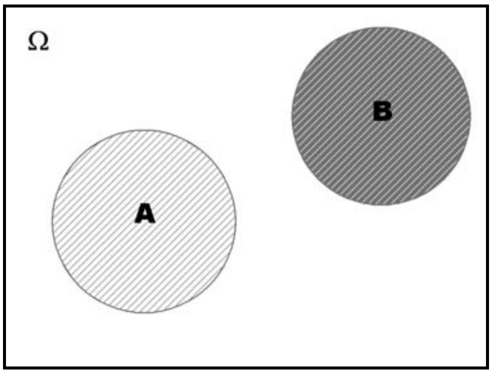
Si se elige al azar una persona de esta población: ¿cuál es la probabilidad de que vea la televisión, escuche la radio o lea el diario?

$$\begin{aligned}P(T\cup R\cup D)&=P(T)+P(R)+P(D)-\left[P(T\cap R)+P(T\cap D)+P(R\cap D)\right]+P(T\cap R\cap D)\\&=0,7+0,4+0,1-\left[0,21+0,028+0,09\right]+0,02=0,892\end{aligned}$$

Es fácil generalizar esta expresión para *n* sucesos teniendo en cuenta que si *n* es par el último término (suceso compuesto) es negativo, mientras que si *n* es impar, este término es positivo.

Un caso particular se presenta cuando los sucesos son mutuamente excluyentes, o sea que los conjuntos son disjuntos. La ley de suma se reduce a:

$$P(A\cup B)=P(A)+P(B)$$



Esta expresión permite proporcionar una explicación razonable a la afirmación de que el suceso imposible tiene probabilidad cero.

En efecto, partiendo de que la probabilidad del suceso cierto o seguro es igual a 1, dado que:

$P(\Omega) = P(\omega_1) + P(\omega_2) + \ldots + P(\omega_n) = 1$, como Ω y $\varnothing$ representan sucesos complementarios, es decir mutuamente excluyentes,

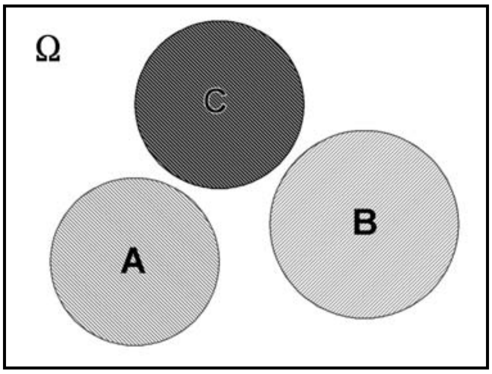
$\varnothing \cup \Omega = \Omega$ $P(\varnothing \cup \Omega) = P(\Omega)$ $P(\varnothing) + P(\Omega) = P(\Omega)$

Esto sólo puede ocurrir si $P(\varnothing) = 0$.

La extensión de la ley de suma a tres sucesos mutuamente excluyentes es la siguiente:

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$

La generalización a *n* sucesos no ofrece dificultades.



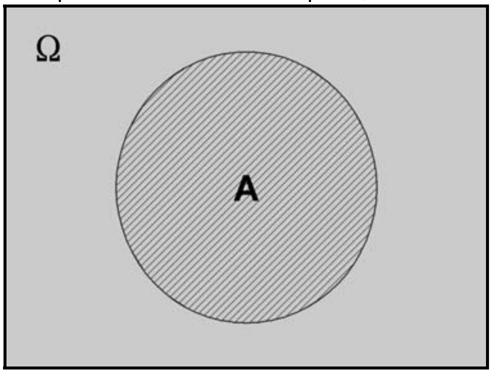
6.2 Ley de multiplicación

Dados los sucesos A y B, la probabilidad de que ocurra A y B simultáneamente es:

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$$

Para brindar una explicación razonable de la validez de esta expresión, comencemos por suponer que hay un suceso A, subconjunto de Ω.

Aplicando la definición a priori, resulta evidente que:



$P(A) = \dfrac{n(A)}{n(\Omega)}$, o bien:

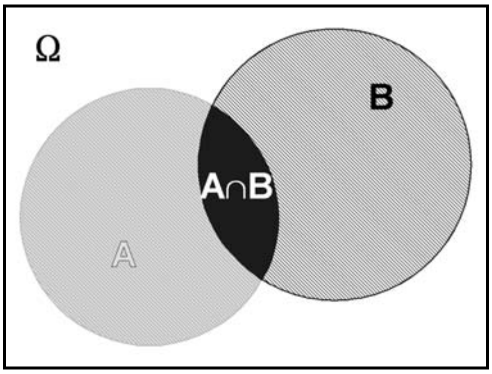
$$P(A) = \frac{n(A \cap \Omega)}{n(\Omega)}.$$

El suceso A ocurrirá "dado Ω". Hay allí una condición que se puede expresar como A/Ω.

$$P(A/\Omega) = P(A) = \frac{n(A \cap \Omega)}{n(\Omega)}$$

Ahora supongamos que se define un nuevo suceso B también subconjunto de Ω. Por analogía se puede expresar:

$$P(A/B) = \frac{n(A \cap \Omega)}{n(B)}$$



y dividiendo numerador y denominador del segundo miembro por n(Ω) se tiene:

$$P(A/B) = \frac{\dfrac{n(A \cap \Omega)}{n(\Omega)}}{\dfrac{n(B)}{n(\Omega)}} = \frac{P(A \cap B)}{P(B)}$$

Si la probabilidad de A dado B es igual a la probabilidad de la intersección de A con B sobre la probabilidad de la condición, que es B, también:

$P(A \cap B) = P(A) \cdot P(A/B)$, o lo que es lo mismo: $P(A \cap B) = P(B) \cdot P(B/A)$

<u>Ejemplo:</u> Calcule la probabilidad de obtener Rey y Copa en una extracción al azar, de un mazo de 48 cartas españolas.

A: "Obtener Rey" ; B: "Obtener Copa"

$$P(A \cap B) = \frac{4}{48} \cdot \frac{1}{4} = \frac{1}{48}$$

Nótese que la probabilidad de obtener copa, dado que se obtuvo rey es ¼.

Es importante señalar que si existe "independencia probabilística" entre A y B, es decir que A y B son sucesos independientes la probabilidad de B dado A es igual a la probabilidad de B y que la probabilidad de A dado B es igual a la probabilidad de A

$$P(B/A) = P(B) \qquad \text{y} \qquad P(A/B) = P(A)$$

6.2.1 Sucesos independientes:

Dos sucesos son independientes cuando la presentación de cualquiera de ellos no afecta la probabilidad de presentación del otro.

En este caso, resulta obvio que:

$$P(A \cap B) = P(A) \cdot P(B)$$

Ejemplo: "Se realizan dos lanzamientos de una moneda correcta".

Definiendo los sucesos:

A: "Cara en el primer lanzamiento"

B: "Cara en el segundo lanzamiento"

La probabilidad de obtener cara en el primer lanzamiento y cara en el segundo será:

$$P(A \cap B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Como se observa, existe independencia probabilística entre A y B.

La extensión de la ley de multiplicación a tres sucesos es la siguiente:

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P\left[C/(A \cap B)\right]$$

Si A, B y C son independientes:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

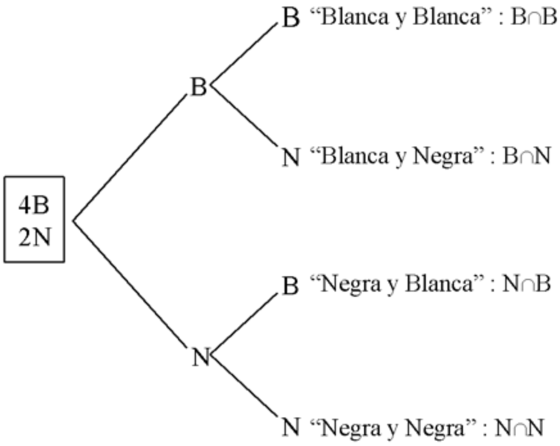
La generalización para *n* sucesos no ofrece dificultades.

7. ALGUNAS APLICACIONES ELEMENTALES

Ejemplo: Una urna contiene 4 bolillas blancas y 2 negras. Si se efectúan 2 extracciones al azar se pide describir un espacio muestral asociado al experimento y calcular la probabilidad de obtener al menos una bolilla blanca si las extracciones se realizan "con reposición".

Debemos asumir que efectuar extracciones al azar, implica no otorgar privilegios a ninguna de las bolillas, lo que en su oportunidad se definirá como "muestreo aleatorio simple".

Nos apoyaremos en el llamado "diagrama del árbol", que permite visualizar el experimento en todas sus etapas y efectuar una asignación ordenada de las probabilidades.

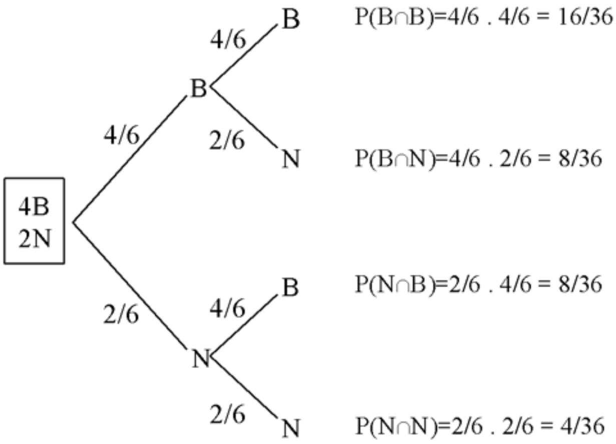


Utilizando una notación resumida es posible presentar el espacio muestral como:

$\Omega = \{BB\,;\,BN\,;\,NB\,;\,NN\}$

De acuerdo con la definición de probabilidad en el caso de los sucesos equiposibles o equiprobables:

$\left(\dfrac{\textit{Número de casos favorables}}{\textit{Número de casos posibles}}\right)$ y a las leyes básicas presentadas se tiene:



Obsérvese que hemos aplicado:

- La definición de probabilidad para el caso de los sucesos equiposibles o equiprobables $\left(\dfrac{\textit{Número de casos favorables}}{\textit{Número de casos posibles}}\right)$.
- La ley de multiplicación para el cálculo de probabilidad: $P(A \cap B) = P(A) \cdot P(A/B)$ en su expresión reducida: $P(A \cap B) = P(A) \cdot P(B)$, dado que el muestreo "con reposición" asegura independencia probabilística. Es decir que $P(B/A) = P(B)$.

"Al menos una bolilla blanca" se satisface mediante los tres primeros resultados listados en el "diagrama del árbol". Es decir que el "éxito" se obtiene si ocurre el primer resultado "o" el segundo "o" el tercero.

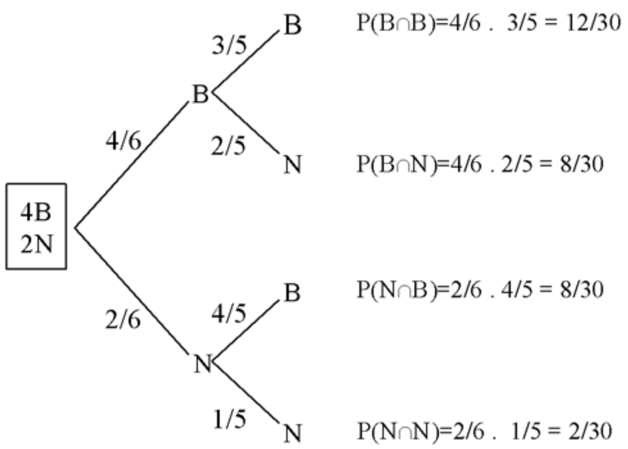
$$P\left[(B\cap B)\cup(B\cap N)\cup(N\cap B)\right] = P(B\cap B)+P(B\cap N)+P(N\cap B)$$
$$=\frac{16}{36}+\frac{8}{36}+\frac{8}{36}=\frac{32}{36}=\frac{8}{9}$$

Aquí hemos aplicado la ley de suma para el cálculo de probabilidades, en su expresión reducida para el caso de sucesos mutuamente excluyentes. Es obvio que si se presenta uno de los tres sucesos, se excluye la posibilidad de que simultáneamente pueda ocurrir cualquiera de los demás.

La misma probabilidad puede ser calculada a través de un camino más corto, utilizando lo conocido para el caso de sucesos contrarios.

$$P\left[(B\cap B)\cup(B\cap N)\cup(N\cap B)\right] = 1-P(N\cap N)$$
$$=1-\frac{4}{36}=\frac{32}{36}=\frac{8}{9}$$

La resolución de este mismo ejercicio, para el caso de que el muestreo se realice "sin reposición" no ofrece dificultades, teniendo en cuenta que la probabilidad de la bolilla obtenida en la segunda extracción está condicionada por lo ocurrido en la primera extracción. Es decir que en este caso no existe independencia probabilística.



$$P\left[(B\cap B)\cup(B\cap N)\cup(N\cap B)\right] = \frac{12}{30}+\frac{8}{30}+\frac{8}{30}$$
$$=\frac{28}{30}=\frac{14}{15}$$

El mismo resultado se puede obtener mediante el suceso contrario:

$$P\left[(B\cap B)\cup(B\cap N)\cup(N\cap B)\right]=1-\frac{2}{30}=\frac{28}{30}=\frac{14}{15}$$

8. TEOREMA DE LA PROBABILIDAD TOTAL

En ciertas oportunidades, la probabilidad de presentación de un suceso A no se conoce y sólo puede ser calculada a partir de las probabilidades condicionales que lo asocian con otros sucesos:

Ejemplo: Existen tres máquinas (B_1 , B_2 y B_3) que producen un mismo artículo. La proporción de artículos defectuosos producidos por cada máquina en un período prolongado (probabilidad a posteriori) es la siguiente:

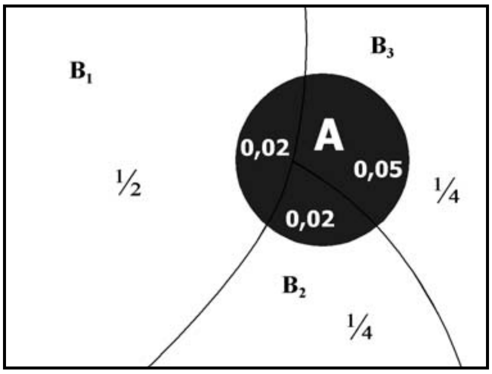
P (A/B_1) = 0,02

P (A/B_2) = 0,02

P (A/B_3) = 0,05

A/B_i significa obtener defectuoso dado que proviene de la máquina B_i.

También se sabe que la máquina B_1 produce la mitad de la producción diaria, mientras que B_1 y B_2, una cuarta parte cada una.



Si la producción de un día se junta, y se elige un artículo al azar: ¿Cuál es la probabilidad de que sea defectuoso?

Como se observa los sucesos B_1, B_2 y B_3 constituyen una partición de Ω. Es decir que se cumple:

◊ $B_i \cap B_j = \varnothing \quad \forall\, i \neq j$

◊ $\bigcup_{i=1}^{n} B_i = \Omega$

◊ $P(B_i) > 0 \quad \forall\, i$

Esto quiere decir que si se elige un artículo al azar, éste proviene de una y sólo una de las máquinas B_i, entonces:

$$A = A \cap B_1 \cup A \cap B_2 \cup A \cap B_3$$

De acuerdo con la ley de suma:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

y de acuerdo con la ley de multiplicación:

$$P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + P(B_3) \cdot P(A/B_3)$$

Generalizando para n sucesos se tiene la expresión del teorema de la probabilidad total:

$$\boxed{P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)}$$

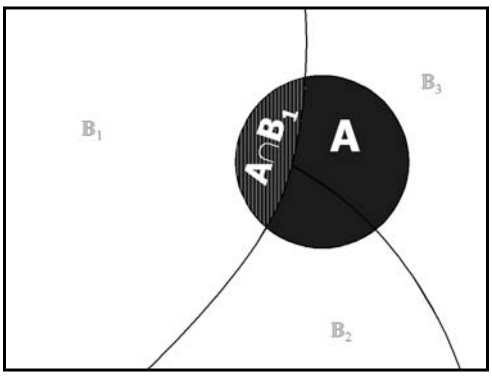
que aplicado al ejemplo permite calcular la probabilidad pedida:

$$P(A) = \frac{1}{2} \cdot 0,02 + \frac{1}{4} \cdot 0,02 + \frac{1}{4} \cdot 0,05 = 0,0275$$

8. TEOREMA DE BAYES

Apoyándonos en el mismo ejemplo, se propone la siguiente situación:
Del total de la producción de cierto día, se elige un artículo al azar, que resulta ser defectuoso: ¿Cuál es la probabilidad de que haya sido producido por la máquina B_1?

El teorema de Bayes o de las "causas" permite resolver este tipo de situaciones, donde se conoce el estado final y se quiere determinar la probabilidad de un cierto estado inicial.



Teniendo en cuenta que la probabilidad de A es conocida, o bien se puede determinar aplicando el teorema de la probabilidad total, y que la condición ahora es A, se obtiene la expresión final del teorema de Bayes:

$$\boxed{P\left(B_j/A\right)=\frac{P\left(A\cap B_j\right)}{P(A)}=\frac{P\left(B_j\right)\cdot P\left(A/B_j\right)}{\sum_{i=1}^{n}P\left(B_i\right)\cdot P\left(A/B_i\right)}}$$

En el ejemplo:

$$P\left(B_1/A\right)=\frac{\frac{1}{2}\cdot 0,02}{0,0275}=0,364$$

Con el mismo razonamiento podríamos calcular la probabilidad de que el artículo defectuoso provenga de la máquina B_2:

$$P\left(B_2/A\right)=\frac{\frac{1}{4}\cdot 0,02}{0,0275}=0,182$$

O bien la probabilidad de que provenga de la máquina B_3:

$$P\left(B_3/A\right)=\frac{\frac{1}{4}\cdot 0,05}{0,0275}=0,454$$

Como se observa, si al elegir un artículo al azar, éste resulta ser defectuoso, existe una mayor probabilidad de que provenga de la máquina B_3.

LABORATORIO 7

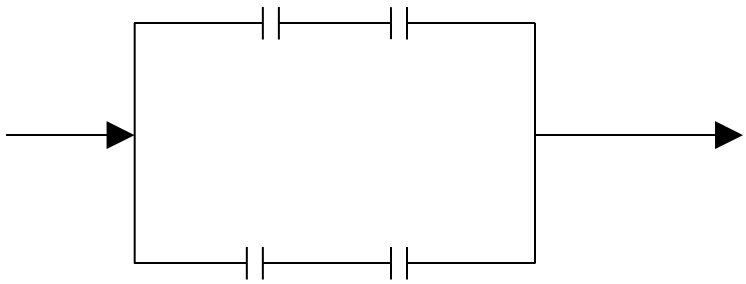
1. De una urna que contiene seis bolillas rojas y cuatro negras, se extraen dos al azar. Se pide:
   1.1 Construir una espacio muestral asociado al experimento.
   1.2 Calcular la probabilidad de obtener al menos una bolilla roja sí:
      1.2.1 La primera bolilla se repone antes de la segunda extracción. (0.8400)
      1.2.2 La segunda bolilla se extrae sin reponer la primera. (0.8667)

2. Dos cajas contienen cada una, dos bolillas rojas y dos negras. Determine en qué caso se tiene mayor probabilidad de obtener exactamente dos bolillas rojas:
   2.1 Sacando una bolilla de cada caja. (0.2500)
   2.2 Juntando las ocho y sacando dos sin reposición. (0.2143)

3. De una urna que contiene tres fichas rojas, dos verdes y dos negras, se extraen al azar y sin reposición dos fichas. Calcule la probabilidad de que las dos fichas sean del mismo color. (5/21)

4. Suponga que A y B son sucesos independientes asociados a un experimento aleatorio. Si la probabilidad de que A o B ocurra es igual a 0.6, mientras que la probabilidad de que A ocurra es igual a 0.4, determine la probabilidad de que B ocurra. (1/3)

5. La probabilidad de que cada uno de los relés que se indican en la figura se encuentre cerrado es p. Si todos los relés funcionan independientemente: ¿Cuál es la probabilidad de que la corriente pase de izquierda a derecha. ($2p^2-p^4$)



6. Una persona posee una casa en Montevideo y otra en San Luis. Admitiendo que un año cualquiera, la probabilidad de que entren a robar en la casa de Montevideo, es 0.01, mientras que en San Luis esa probabilidad es 0.05, se pide calcular la probabilidad de que en un año cualquiera:

   6.1 Roben en ambas casas. (0.0005)
   6.2 Roben en una u otra, o en ambas simultáneamente. (0.0595)
   6.3 No roben en ninguna de las casas. (0.9405)

7. La probabilidad de que una puerta se encuentre cerrada es ½. La llave de la puerta es una de 12 llaves que se encuentran en un estante. Si una persona toma al azar una llave del estante y va a abrir la puerta: ¿Cuál es la probabilidad de que pueda abrirla sin necesidad de volver a buscar más llaves? (0.5417)

8. En una localidad del interior del país se leen los diarios A y B. Una encuesta recientemente realizada permitió saber que:
   - El 70 por ciento de la población lee el diario A, y el 50 por ciento lee B.
   - Los que leen A y B son el 25 por ciento de la población.

   Calcule la probabilidad de elegir una persona al azar en esa localidad que:

   8.1 Lea A, lea B, o ambos simultáneamente. (0.95)
   8.2 No lea ni A, ni B, ni ambos simultáneamente. (0.05)

8.3 Lea A, sabiendo que lee B. (0.50)
8.4 Lea B, sabiendo que lee A. (0.3571)

9. Una urna contiene cuatro bolillas rojas y una negra. Una segunda urna contiene dos bolillas blancas. Se sacan al azar tres bolillas de la primera urna y se colocan en la segunda. Luego se sacan cuatro bolillas también al azar, de la segunda y se colocan en la primera. Encuentre la probabilidad de que la bolilla negra esté en la primera urna luego de ambos intercambios. (0.4+0.48=0.88)

10. Sean A y B dos sucesos asociados a un experimento aleatorio. Suponga que P(A)=0.4; mientras que P(AUB)=0.7 ; y que P(B)= p.
    10.1 ¿Para qué elección de p; A y B son mutuamente excluyentes? (p=0.3)
    10.2 ¿Para qué elección de p; A y B son independientes? (p=0.5)

NOTA DE CLASE Nº 6

VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD. Variable aleatoria y probabilidad inducida. Variables aleatorias discretas. Bernoulli. Binomial. Hipergeométrica. Variables aleatorias continuas. Normal. Anexo 6.1.

1. VARIABLE ALEATORIA Y PROBABILIDAD INDUCIDA.

Estos conceptos, de suma importancia para el Cálculo de Probabilidades, serán presentados en el ANEXO 6.1.

2. VARIABLES ALEATORIAS DISCRETAS.

En general, X es una variable aleatoria discreta si su recorrido es contable, es decir finito o numerable.

$$\Omega = \{X_1, \ldots, X_n\}$$

Función de cuantía : Se denomina función de cuantía , a la probabilidad P(X) de que X asuma un valor particular X_i de su recorrido.

Función de distribución acumulativa: $F(X_i)$ representa la probabilidad acumulada hasta el valor de X_i.

Ejemplo 1:

El experimento aleatorio consiste en lanzar dos veces una moneda correcta. I_i es una variable indicadora de la ausencia o de la presencia del suceso de interés (0;1). La variable aleatoria X, se define como: $I_1 + I_2$. Se pide determinar: dominio y recorrido de X, función de cuantía y función de distribución acumulativa.

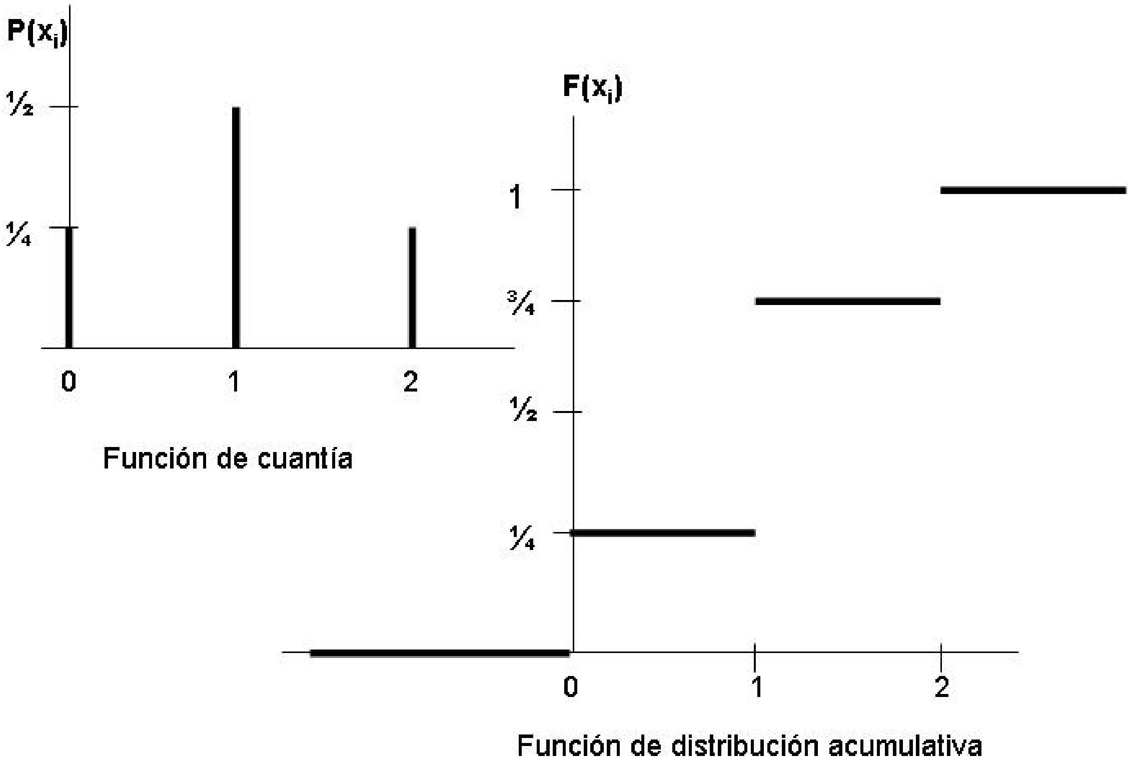
Dominio de X: $\Omega = \{cc; cs; sc; ss\}$

Recorrido de X: $\Omega_1 = \{0;1;2\}$

$\omega_i \in \Omega$	$P(\omega_i)$	I_1	I_2	$X_i=I_1+I_2$	$P(X_i)$
cc	¼	1	1	2	¼
cs	¼	1	0	1	¼
sc	¼	0	1	1	¼
ss	¼	0	0	0	¼

Resumiendo:

X_i	$P(X_i)$	$F(X_i)$
0	$^1/_4$	$^1/_4$
1	$^2/_4$	$^3/_4$
2	$^1/_4$	$^4/_4$



Media aritmética y varianza teórica.

Dada una variable aleatoria X y su función de cuantía, o sea los valores de X_i del recorrido de la variable aleatoria X y las probabilidades correspondientes a esos valores $P(X_i)$, se define la media aritmética teórica o *esperanza de* X como:

$$E(X) = \sum_{i=1}^{n} X_i . P(X_i)$$

Mientras que la varianza teórica se define como:

$$V(X) = E\,[X_i - E(X_i)]^2 = E(X_i^2) - [\,E(X_i)]^2 = \sum_{i=1}^{n} X_i^2 .\ P(X_i) - [\,E(X_i)]^2$$

Ambos valores estadísticos, así como algunas de sus propiedades más notables, ya fueron suficientemente discutidos en la Nota de Clase Nº1, correspondiente a "Estadística Descriptiva". La novedad consiste en utilizar como ponderadores de los valores de X_i , las probabilidades asociadas a estos valores $[P(X_i)]$ en lugar de las frecuencias relativas $[h_i]$. Sin embargo, la conocida relación existente entre ambos conceptos (las frecuencias relativas son valores empíricos de la probabilidad) exime de mayores justificaciones al respecto.

ALGUNAS DISTRIBUCIONES DISCRETAS

2.1 BERNOULLI (Bipuntual)

Consideremos un experimento aleatorio con dos resultados posibles y una transformación del espacio muestral directamente asociado mediante una variable I que indica la ausencia (fracaso) o la presencia (éxito) del suceso de interés.

I = 0 Si el suceso de interés no se presenta (fracaso)
I = 1 Si el suceso de interés se presenta (éxito)

Es claro que la variable I indica el número de veces que puede presentarse el suceso de interés en el único ensayo efectuado. En cuanto a la probabilidad asociada a cada uno de estos resultados se denominará:

q : probabilidad de fracaso
p : probabilidad de éxito

Obviamente, en función de las condiciones oportunamente presentadas para la asignación de probabilidades: $p+q=1$

La media aritmética (*esperanza*) y la varianza de I serán :

$$E(I) = \sum_{i=1}^{2} I_i \, P(I_i) = 0\, q + 1\, p = p$$

$$V(I) = \sum_{i=1}^{2} I_i^2 \, P(I_i) - [\, E(I_i)\,]^2 = 0^2 q + 1^2 p - p^2 = p - p^2 = p\,(1-p) = p\, q$$

La notación resumida que emplearemos para identificar esta distribución de probabilidad es:
$I \sim b\,(p)$

2.2 BINOMIAL

Consideremos ahora un experimento aleatorio con dos resultados posibles, que se repite n veces, con probabilidad de éxito (p) constante. Definiremos una variable indicadora X de la siguiente forma:

$X = I_1 + I_2 + \ldots\ldots\ldots\ldots + I_n$
Donde cada $I_i \sim b\,(p)$

Obsérvese que como p es constante, X se define como la suma de n variables aleatorias *independientes*[11] con distribución de Bernoulli. X_i puede tomar entonces los valores: 0, 1, 2,…… r…….hasta n.

La media aritmética (*esperanza*) y la varianza de X serán :

$$E(X) = E(I_1 + I_2 + \ldots + I_n) = E(I_1) + E(I_2) + \ldots E(I_n)$$

Dado que la *esperanza* de una variable con distribución de Bernoulli es la probabilidad de éxito p, se tiene:

$$E(X) = p + p + \ldots\ldots\ldots\ldots + p = n\,.p$$

[11] Es obvio que la presentación del suceso 0, o bien del suceso 1 en el primer ensayo, no afecta en absoluto la probabilidad del suceso que pueda presentarse en el segundo ensayo, y así sucesivamente. Existe entonces lo que oportunamente se ha definido como “independencia probabilística” entre sucesos.

En tanto, la varianza de X_i es:

$$V(X) = V(I_1 + I_2 + \ldots\ldots\ldots\ldots\ldots\ldots + I_n)$$

Dado que existe independencia probabilística, la varianza de una suma es igual a la suma de las varianzas. Entonces:

$$V(X) = V(I_1)+V(I_2) + \ldots\ldots\ldots\ldots\ldots\ldots+V(I_n)$$

Recordando que para variables con distribución de Bernoulli, la varianza es p.q, se tiene:

$$V(X) = n.\ p.q$$

La notación resumida que emplearemos para identificar esta distribución de probabilidad es la siguiente:

$$X \sim B\ (n;\ p)$$

Como veremos a continuación, si la variable X tiene distribución Binomial, el cálculo de la probabilidad de obtener *r* éxitos en *n* ensayos, se simplifica notablemente.

Ejemplo 2:

Un experimento aleatorio que tiene solamente dos resultados posibles en cada ensayo, se repite dos veces. Si la probabilidad de éxito (*p*) es constante, es evidente que la variable que indica el número de éxitos que se pueden presentar en los dos ensayos, tiene distribución Binomial.

$X=I_1+I_2$ $\qquad\qquad$ $X \sim B\ (n=2;\ p)$

En la siguiente tabla se presenta el conjunto de todos los resultados posibles y las probabilidades asociadas. Las mismas se calculan mediante la aplicación de las leyes básicas de probabilidad: ley de multiplicación (sucesos independientes) y ley de suma (sucesos mutuamente excluyentes):

I_1	I_2	$X_i=I_1+I_2$	$P(X_i)$
0	0	0	q^2
0	1	1	qp
1	0	1	qp
1	1	2	p^2

X_i	$P(X_i)$
0	q^2
1	$2qp$
2	p^2

Si recordamos que: $(q+p)^2 = q^2 + 2qp + p^2$, se puede observar que existe una estricta correspondencia entre la probabilidad de obtener *0 éxito*, *1 éxito*, 2 *éxitos* -en dos ensayos- con los sucesivos términos del desarrollo del binomio $(q+p)^2$. También se puede observar que la probabilidad de obtener exactamente *r éxitos* en los dos ensayos está dada por el término del desarrollo del binomio, en que el exponente de *p* es precisamente *r.*

Ejemplo 3:

Un experimento aleatorio que tiene solamente dos resultados posibles en cada ensayo, se repite tres veces. Si la probabilidad de éxito (*p*) es constante, la variable que indica el número de éxitos que se pueden presentar en los tres ensayos, tiene distribución Binomial.

$X = I_1 + I_2 + I_3$ $\qquad$ $X \sim B(n=3;\ p)$

En la siguiente tabla se presenta el conjunto de todos los resultados posibles y las probabilidades asociadas. Las mismas se calculan mediante la aplicación de las leyes básicas de probabilidad: ley de multiplicación (sucesos independientes) y ley de suma (sucesos mutuamente excluyentes):

I_1	I_2	I_3	$X_i=I_1+I_2+I_3$	$P(X_i)$
0	0	0	0	q^3
0	0	1	1	q^2p
0	1	0	1	q^2p
0	1	1	2	qp^2
1	0	0	1	q^2p
1	0	1	2	qp^2
1	1	0	2	qp^2
1	1	1	3	p^3

X_i	$P(X_i)$
0	q^3
1	$3q^2p$
2	$3qp^2$
3	p^3

Si recordamos que: $(q+p)^3 = q^3 + 3q^2p + 3q\,p^2 + p^3$, se puede observar que existe una estricta correspondencia entre la probabilidad de obtener 0 éxito, 1 éxito, 2 éxitos, 3 éxitos -en tres ensayos- con los sucesivos términos del desarrollo del binomio $(q+p)^3$. También se puede observar que la probabilidad de obtener exactamente *r* éxitos en los dos ensayos está dada por el término del desarrollo del binomio, en que el exponente de *p* es precisamente *r.*

Esta correspondencia, se puede generalizar a n ensayos, es decir que los sucesivos términos del desarrollo del binomio:

$$(q+p)^n = q^n + C_1^n\, q^{n-1} p + C_2^n\, q^{n-2} p^2 + \ldots\ldots\ldots + C_r^n\, q^{n-r} p^r + \ldots\ldots\ldots + p^n$$

estarán dando las probabilidades de obtener *0 éxito*, 1 *éxito*, … , *r éxitos*, …, hasta *n éxitos* en *n* ensayos. La probabilidad de obtener exactamente *r éxitos* en *n ensayos* estará dada por el término del desarrollo del binomio, en el que el exponente de *p* es precisamente *r.*

$P(X_i=r) = C_r^n\, q^{n-r} p^r$.

Se recordará que $\quad C_r^n = \dfrac{n!}{r!(n-r)!} = \dfrac{n(n-1)(n-2)\ldots\ldots\ldots\ldots(n-r+1)}{r!}$

Y que $\quad C_r^n = C_{n-r}^n \qquad\qquad 0! = 1$

Ejemplo 4:

Un experimento cosiste en efectuar 5 extracciones con reposición, de una urna que contiene 8 bolillas blancas y 6 negras. Se pide:

4.1 Identificar la distribución de la variable que describe el número de bolillas blancas en la muestra.

4.2 Utilizando esa distribución, calcular la probabilidad de obtener exactamente dos bolillas del mismo color.

4.1 Como se trata de ensayos repetidos con dos resultados posibles y con *p* constante (extracciones con reposición): X ~ B (n=5; *p*= $^8/_{14}$).

4.2 $P(X_i=2) = C_2^5\ (^8/_{14})^2\ (^6/_{14})^3 = 0{,}26$

$P(X_i=3) = C_3^5\ (^8/_{14})^3\ (^6/_{14})^2 = 0{,}34$

Como el suceso de interés es obtener dos blancas o tres blancas:

$P(X_i=2) + P(X_i=3) = 0{,}60$

Obsérvese que el suceso "exactamente dos negras" es equivalente al suceso "exactamente tres blancas" en las cinco extracciones efectuadas. Se calcula la probabilidad de este último suceso, dado que se pide utilizar la distribución de probabilidad de la variable X , que describe el número de bolillas <u>blancas</u> en la muestra.

2.3 HIPERGEOMÉTRICA

Consideremos ahora un experimento aleatorio también con dos resultados posibles, que se repite n veces, pero con probabilidad de éxito (p) variable. La distribución de probabilidad asociada a este tipo de experimentos y en general al muestreo sin reposición[12] se denomina: Distribución Hipergeométrica.

Definiremos una variable X de la siguiente forma:

$$X = I_1 + I_2 + \ldots\ldots\ldots\ldots + I_n$$

Obsérvese que X se define como la suma de n variables aleatorias con distribución de Bernoulli. X puede tomar entonces los valores: 0, 1, 2,……r…….hasta n. Sin embargo en este caso, estas variables indicadoras no son independientes.

Se puede demostrar que la media aritmética (*esperanza*) y la varianza de X son :

$$E(X) = n\ p$$

$$V(X) = n\,p\,q\ \frac{N-n}{N-1}$$

[12] Cuando el tamaño de la población (*N*) es suficientemente grande en relación al tamaño de la muestra (*n*), se puede admitir que desde el punto de vista práctico, existe independencia . Por lo tanto es posible aproximar la Hipergeométrica mediante la Binomial.

Al factor $\dfrac{N-n}{N-1}$ se le denomina "Factor de corrección por población finita". Obsérvese que cuando el tamaño de la población (*N*) es suficientemente grande en relación al tamaño de la muestra (*n*), es decir cuando $^n/_N$ es cercano o inferior al 5 por ciento, este factor no difiere significativamente de uno.

La notación resumida que emplearemos para identificar esta distribución de probabilidad es la siguiente:

$$X \sim H\ (N;n;\ p)$$

Como veremos a continuación, si la variable X tiene distribución Hipergeométrica, el cálculo de la probabilidad de obtener *r* éxitos en *n* ensayos, también se simplifica notablemente.

La expresión que permite calcular la probabilidad de obtener exactamente r éxitos en n ensayos es la siguiente:

$$P(X_i{=}r) = \frac{C_r^a C_{n-r}^b}{C_n^N}$$

Donde: r : número de éxitos
a : número de elementos que en la población poseen el atributo de interés.
b : número de elementos que en la población no poseen el atributo de interés.
n : tamaño de la muestra
N: tamaño de la población (a+b)

La comprobación de la utilidad de esta expresión para calcular probabilidades en el caso de muestreo sin reposición, se ilustrará con el ejemplo siguiente:

Ejemplo 5.

Dada una población constituida por dos bolillas blancas y dos negras (N=4), se efectúan dos extracciones sin reposición (n=2). El suceso de interés (éxito) es obtener bolilla blanca (a=2) (b=2).
A efectos de que las bolillas sean distinguibles, se asignará un número a cada bolilla. Las blancas se numerarán del 1 al 2 y las negras del 3 al 4.
A continuación construiremos el espacio muestral directamente asociado al experimento, el espacio muestral inducido por una transformación que asigna a cada suceso elemental de Ω los valores 0 o 1, teniendo en cuenta si está ausente o presente el atributo de interés, y las probabilidades correspondientes:

$\omega_i \in \Omega$	$\omega_i \in \Omega_1$	$X_i = I_1 + I_2$	$P(X_i)$
1; 2	1; 1	2	$^1/_{12}$
1; 3	1; 0	1	$^1/_{12}$
1; 4	1; 0	1	$^1/_{12}$
2; 1	1; 1	2	$^1/_{12}$

2; 3	1; 0	1	$^1/_{12}$
2; 4	1; 0	1	$^1/_{12}$
3; 1	0; 1	1	$^1/_{12}$
3; 2	0; 1	1	$^1/_{12}$
3; 4	0; 0	0	$^1/_{12}$
4; 1	0; 1	1	$^1/_{12}$
4; 2	0; 1	1	$^1/_{12}$
4; 3	0; 0	0	$^1/_{12}$

Resumiendo:

X_i	$P(X_i)$
0	$^2/_{12}$
1	$^8/_{12}$
2	$^2/_{12}$

Utilizando la expresión propuesta para el cálculo inmediato de las mismas probabilidades, se comprueban idénticos resultados:

$$P(X_i = 0) = \frac{C_0^2 C_{2-0}^2}{C_2^4} = {}^1/_6$$

$$P(X_i = 1) = \frac{C_1^2 C_{2-1}^2}{C_2^4} = {}^4/_6$$

$$P(X_i = 2) = \frac{C_2^2 C_{2-2}^2}{C_2^4} = {}^1/_6$$

Ejemplo 6.

De una urna que contiene 10 bolillas blancas y 6 negras, se realizan 6 extracciones sin reposición. Se pide:

6.1 Identificar la distribución de probabilidad de la variable que describe el número de bolillas blancas en la muestra y mediante esa distribución calcular la probabilidad de obtener:
6.2 Al menos una bolilla blanca.
6.3 Dos bolillas del mismo color.

6.1 $X_i \sim H(N= 16;n=6;\ p={}^{10}/_{16})$

6.2 $P(X_i \geq 1) = 1- P(X_i =0) = 1- \dfrac{C_0^{10} C_{6-0}^6}{C_6^{16}} = 0{,}9999$

6.3 $P(X_i =2) + P(X_i =4) = \dfrac{C_2^{10} C_{6-2}^6}{C_6^{16}} + \dfrac{C_4^{10} C_{6-4}^6}{C_6^{16}} = 0{,}4776$

1. VARIABLES ALEATORIAS CONTINUAS

En general, X será una variable aleatoria continua si su recorrido no es finito ni numerable, es decir que puede tomar todos los valores dentro de cierto intervalo considerado.

Función de densidad: La probabilidad asociada a un intervalo de amplitud infinitesimal estará representada por un área infinitesimal de altura f(x) y de base dx.

Función de distribución acumulativa: F(x) representa la probabilidad acumulada hasta el valor *x*.

Media aritmética y varianza teórica:

Dada una función de densidad f(x) se define la media aritmética teórica o esperanza de x como:

$$E(x)= \int_{\lambda_1}^{\lambda_2} x\, f(x)\, dx$$

Obsérvese que conceptualmente esta expresión no difiere de las empleadas anteriormente como definiciones de media aritmética ponderada. La novedad es que, dado que se trata de una suma de infinitos términos, es necesario calcular la integral definida entre λ_1 y λ_2 (que pueden ser valores finitos o infinitos). Estos infinitos términos se componen de valores de la variable x ponderados por las probabilidades asociadas f(x) dx.

Por otra parte la varianza teórica se define como:

$$V(x)= \int_{\lambda_1}^{\lambda_2} x^2 f(x)\, dx - [E(x)]^2$$

Se puede apreciar que esta expresión tampoco difiere sustancialmente de la que se empleara anteriormente como definición de varianza: "La media de los cuadrados, menos el cuadrado de la media".

ALGUNAS DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

3.1 NORMAL

La función de densidad denominada "distribución normal" o "campana de Gauss" tiene la siguiente expresión:

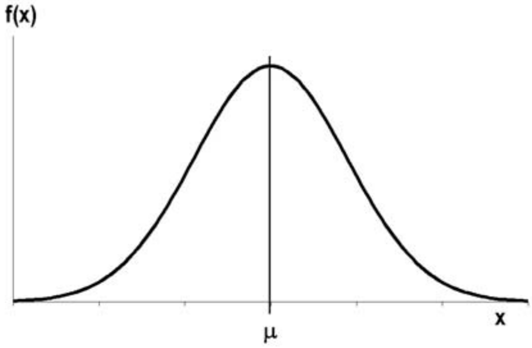
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde e : base de los logaritmos neperianos (2,718..)

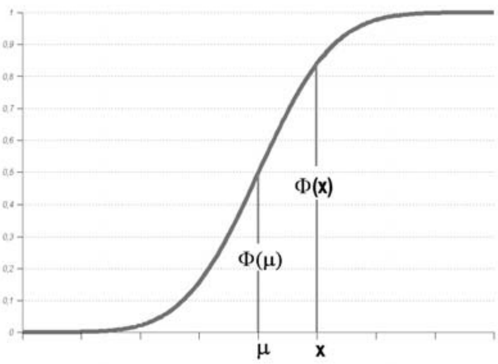
x : valores de la variable

μ : media aritmética poblacional

σ^2 : varianza poblacional



Función de densidad



Función de distribución acumulativa

Dividiendo ambos miembros de la igualdad por σ , se tiene ahora a f(x) expresada en términos de la desviación estándar, es decir:

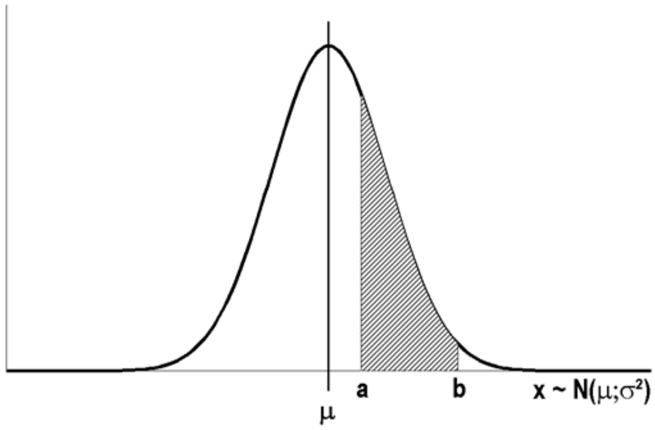
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La notación resumida que emplearemos para identificar esta distribución de probabilidad será la siguiente:

$$X \sim N(\mu ; \sigma^2)$$

Supongamos que una variable x tiene distribución normal, con media μ y varianza σ^2 . Se quiere calcular la probabilidad de que x se encuentre comprendida entre *a* y *b*, es decir:

$$P(a \leq x \leq b) = \int_a^b f(x)\, dx$$



Esta integral definida (suma de infinitos términos entre *a* y *b*) tiene dificultades importantes para ser calculada. El procedimiento que se utiliza para simplificar los aspectos operativos, es el de efectuar un cambio de variable definiendo:

$$z = \frac{x-\mu}{\sigma}$$

Sustituyendo en

f(x) = $\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ se tiene f(z) = $\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

A esta última expresión se le denomina: curva normal estándar, o tipificada, o reducida. El procedimiento simplificado consiste en pasar mediante el mencionado cambio de variable, de cualquier distribución de la familia de distribuciones normales a la distribución normal estándar

X~ N (μ ; σ^2) $\Rightarrow$ Z~ N (μ_z ; ${\sigma^2}_z$)

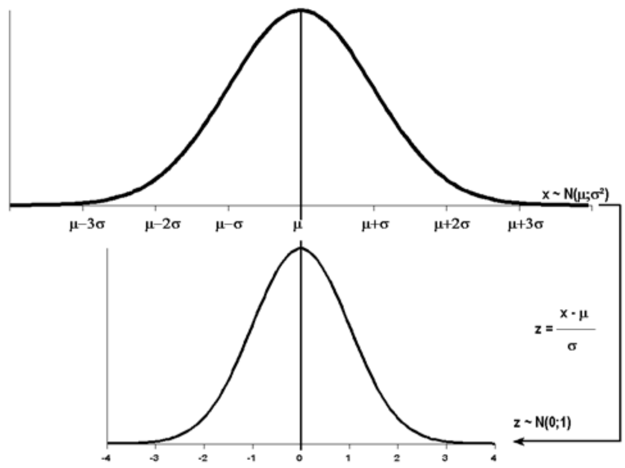
Veremos a continuación cuál es la media aritmética y cuál es la varianza de z:

$\mu_z = \mu[\frac{x-\mu}{\sigma}] = \frac{1}{\sigma}\mu[x-\mu] = \frac{1}{\sigma}[\mu(x)-\mu]$ y como $\mu(x) = \mu$ μ_z =0

${\sigma^2}_z = \sigma^2[\frac{x-\mu}{\sigma}] = \frac{1}{\sigma^2}\{\sigma^2[x-\mu]\} = \frac{1}{\sigma^2}\{\sigma^2\} = 1$ De manera que:

Z~ N (μ_z =0; ${\sigma^2}_z$ =1)

Utilizando esta transformación será posible calcular probabilidades bajo cualquier curva normal. Las áreas correspondientes a z se encuentran calculadas y tabuladas, de manera que se pueden leer directamente (Tabla 1, pag. 11 del manual de Tablas y Fórmulas del curso).



En el ejemplo propuesto la probabilidad entre a y b se encontrará transformando estos valores de la siguiente forma:

$$P(a \leq x \leq b) = P(z_1 = \frac{a-\mu}{\sigma} \leq z \leq z_2 = \frac{b-\mu}{\sigma})$$

Ejemplo 7.

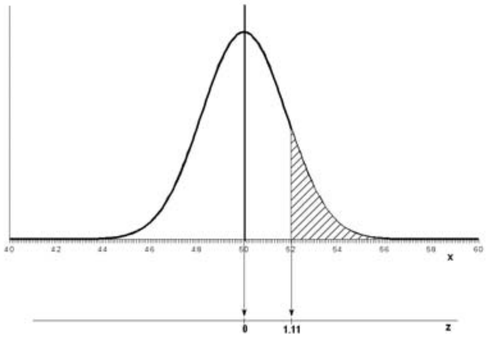
Asumiendo que la talla de los niños al nacer tienen una distribución normal con media 50 cm. y una desviación estándar de 1,8 cm. Se pide calcular la probabilidad de que un niño al nacer tenga una talla:

7.1 Igual o superior a 52 cm.
7.2 Comprendida entre 48 y 51 cm.

Además se pide encontrar:

7.3 los valores de X entre las cuales se encuentra el 95 por ciento central de las tallas de los niños al nacer:

7.1 $X \sim N(\mu = 50; \sigma^2 = 1{,}8^2)$

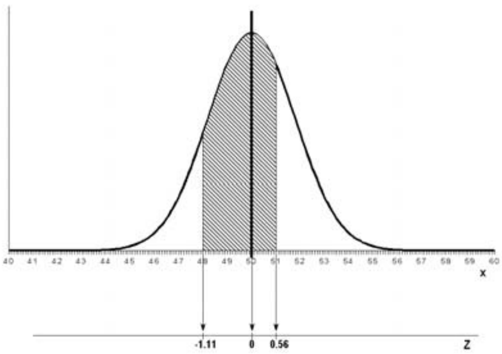


$$z = \frac{52-50}{1,8} = 1,11$$

$P(x \geq 52) = P(z \geq 1,11) = 0,1335$

7.2

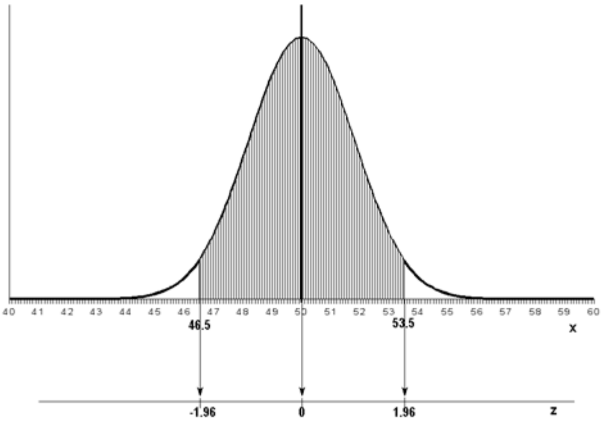
$$z_1 = \frac{48-50}{1,8} = -1,11$$



$$z_2 = \frac{51-50}{1,8} = 0,56$$

$$P(48 \leq x \leq 51) = P(-1,11 \leq z \leq 0,56) = 0,3665 + 0,2123 = 0,5788$$

7.3



$$z = \frac{x-\mu}{\sigma} \Rightarrow x = \mu + z\sigma$$

x_1 = 50 - 1,96 . 1,8 = 46,5 x_2 = 50 + 1,96 . 1,8 = 53,5

$$P(-1,96 \leq z \leq 1,96) = P(46,5 \leq x \leq 53,5) = 0,95$$

ANEXO 6.1

VARIABLE ALEATORIA Y PROBABILIDAD INDUCIDA.

La introducción a estos conceptos, se ilustrará mediante el siguiente ejemplo:

Experimento Aleatorio: "Lanzar una moneda correcta"
Un *espacio muestral* [13]directamente asociado a este experimento es:

$\Omega = \{c, s\}$ Donde c : cara y s : sello (no cara)

Ahora bien, es posible aplicar a este espacio muestral (Ω) una regla de transformación (función o aplicación) que asigne a cada suceso elemental un número real. Si el suceso de interés es "obtener cara", podríamos definir una regla que tenga en cuenta si el atributo "cara" se encuentra ausente o presente, en cada uno de los resultados posibles. Se podría generar entonces, un nuevo espacio muestral (Ω_1), inducido por la referida transformación de tal forma que a cada elemento de Ω le corresponda un elemento de Ω_1. Simbolizando por X a esta función, tendríamos:

$X(\Omega)=\Omega_1=\{0,1\}$ $X(s)=\{0\}$ $X(c)=\{1\}$

El espacio muestral Ω se transforma según X en un nuevo espacio muestral Ω_1, que puede o no tener el mismo número de elementos. Ω es el *dominio* de X, mientras que Ω_1 es el *recorrido* de X.

Esta función X recibe generalmente en la literatura especializada, el nombre de "variable aleatoria". Hay que tener en cuenta que X será una variable aleatoria solo si la transformación inversa (X^{-1}) de sucesos de la *clase completamente aditiva de sucesos*[14] $X(a_1)$ inducida por la transformación, nos conduce como contraimagen a un suceso de la clase completamente aditiva original $X(a)$. Es decir:

$X^{-1}(0)=\{s\}$ y $X^{-1}(1)=\{c\}$

También existirá una probabilidad inducida por la transformación, de tal forma que:

$P(0)=P\left[X^{-1}(0)\right]=P(s)$ y $P(1)=P\left[X^{-1}(1)\right]=P(c)$

En lugar del espacio muestral Ω directamente asociado al experimento aleatorio es posible utilizar entonces, este nuevo espacio muestral Ω_1, inducido a partir del espacio muestral Ω, mediante la transformación o función X denominada "variable aleatoria".

LABORATORIO 8

1. Se sabe que la probabilidad de obtener un artículo defectuoso a partir de cierto proceso manufacturero, es constante y del 10 por ciento.
   Si se extrae una muestra aleatoria de 5 de esos artículos para el control de la calidad, se pide:

[13] Espacio muestral : Se define como el conjunto de todos los resultados posibles del experimento aleatorio.

[14] Clase completamente aditiva de sucesos: En teoría de conjuntos, se emplea la denominación "clase" para referirse a un conjunto de conjuntos, es decir a un conjunto cuyos elementos sean también conjuntos. En cambio, en teoría de probabilidades en lugar de "clase de conjuntos" se emplea la expresión "clase de sucesos". En el ejemplo empleado, el espacio muestral directamente asociado al experimento aleatorio es: $\Omega = \{c, s\}$. Los sucesos de la clase son:

ϕ ; $\{c\}$; $\{s\}$; Ω

Cualquier agregado o colección de estos sucesos es una clase. La clase completamente aditiva de sucesos es la que contiene a todos los sucesos de Ω, es decir: $a=\{\phi;\{c\};\{s\};\Omega\}$. La generalización a otros experimentos aleatorios, no ofrece dificultades.

1.1 Identificar la distribución de la variable aleatoria que indica el número de artículos defectuosos en la muestra.

1.2 Utilizando esa distribución, calcular la probabilidad de obtener:

1.2.1 Ningún artículo defectuoso.(0,5905)

1.2.2 Al menos dos defectuosos. (0,0814)

1.2.3 No más de un defectuoso. (0,9186)

2. Un muchacho parado en una esquina, lanza una moneda. Si sale cara camina una cuadra al este; si sale sello camina una cuadra al oeste. En cada esquina, repite la operación. Si en total realiza seis lanzamientos se pide:

2.1 Identificar la distribución de probabilidad de la variable aleatoria que describe el número de caras en la muestra.

2.2 Determine la probabilidad de que, realizados los seis lanzamientos se encuentre:

2.2.1 En el punto de partida. (0,3125)

2.2.2 A dos cuadras del punto de partida. (0,4688)

2.2.3 A cuatro cuadras del punto de partida. (0,1875)

3. Una urna contiene seis bolillas blancas y cinco negras. Si se realizan cinco extracciones al azar y sin reposición, se pide:

3.1 Identificar la distribución de la variable aleatoria que indica el número de bolillas blancas en la muestra.

3.2 Utilizando esa distribución, calcular la probabilidad de obtener:

3.2.1 Ninguna bolilla blanca. (0,0022)

3.2.2 No más de una bolilla blanca. (0,0671)

3.2.3 Dos bolillas del mismo color.(0,7576)

4. En una urna hay 10 bolillas blancas y 20 negras. Se efectúan 12 extracciones al azar y sin reposición.

Utilizando la distribución de probabilidad de la variable aleatoria que indica el número de bolillas blancas en la muestra, determine la probabilidad de que luego de las 12 extracciones, quede en la urna el mismo número de bolillas blancas, que de negras. (0,0194)

5. Una urna contiene ocho bolillas blancas y seis negras. El experimento aleatorio consiste en extraer cinco bolillas al azar. Identificando en cada caso la distribución de probabilidad de la variable que describe el número de bolillas blancas en la muestra, utilice esa distribución para calcular la probabilidad de obtener:

- Al menos una bolilla blanca
- Dos bolillas del mismo color

5.1 Si las extracciones se realizan con reposición. (0.9855; 0.5997)

5.2 Si las extracciones se realizan sin reposición. (0.9970; 0.6993)

LABORATORIO 9

1 Se sabe que los valores de cierta variable pueden ser ajustados razonablemente por una distribución normal, con media aritmética 140 y desviación estándar 25. Se pide calcular la probabilidad de encontrar en dicha población, Valores:

1.1 Superiores a 170. (0,1151)
1.2 Inferiores a 98. (0,0465)
1.3 Comprendidos entre 100 y 150. (0,6006)
1.4 Comprendidos entre 180 y 190. (0,0320)
1.5 Comprendidos entre 100 y 130. (0,2898)

También se pide determinar los valores de la variable entre los que se ubica el 90 por ciento central de la distribución. (98,88; 181,13)

2. Las puntuaciones de un examen se distribuyen en forma aproximadamente normal, con media aritmética 67 y desviación estándar 12. Se pide determinar:
   2.1 El porcentaje de exámenes con 79 o más puntos. (0,1587)
   2.2 El puntaje bajo el cual quedó el 15.87 por ciento de los exámenes. (55)

3. En un test aplicado al personal de cierta empresa, se supone que los puntajes siguen una distribución normal, con media 70 y varianza 100. Los 270 empleados con puntajes de 61 a 79, tuvieron la calificación B. Estime cuántos empleados se examinaron en total. (427)

4. Se dispone de 5.422 resultados de cierto test. Sabiendo que los puntajes se pueden aproximar mediante una distribución normal, con media aritmética 75 y desviación estándar 7, se pide calcular:

   4.1 La probabilidad de que un test extraído al azar tenga entre 61 y 68 puntos. (0,1359)
   4.2 El número de tests que obtuvieron entre 61 y 68 puntos. (737)
   4.3 Entre qué valores queda ubicado el 95 por ciento central de la distribución. (61,28; 88,72)

5. La altura de los niños en cierta edad, tiene distribución aproximadamente normal con media aritmética 140 cm y varianza 100 cm^2. ¿Cuál es la probabilidad de que un niño elegido al azar mida:

   5.1 más de 150 cm de altura?. (0,1587)
   5.2 Entre 145 y 150 cm de altura?. (0,1498)
   5.3 ¿Cuánto mide el niño más alto del 20 por ciento más bajo de la distribución? (131,6)

6. Un profesor otorga las calificaciones de manera que la distribución puede considerarse como razonablemente normal. Ordenadas de la calificación más baja a la más alta, se tiene un 10 por ciento de calificaciones A, 20 por ciento B, 40 por ciento C, 20 por ciento D, y 10 por ciento E. La media aritmética es 68. Si el límite entre C y D es 78, se pide calcular la desviación estándar. (19,23)

Educando para la vida

Cuareim 1451 Tel. 2902 15 05 Fax 2908 13 70
info@ort.edu.uy - www.ort.edu.uy

Abril 2012