

Temas de Estadística Práctica

Antonio Roldán Martínez

Proyecto <http://www.hojamat.es/>

Tema 3

Medidas típicas. Índices

[Clases de puntuaciones](#)

[Índices de posición](#)

[Números índices](#)

Clases de puntuaciones

Las medidas directas que se efectúan sobre una muestra no siempre informan claramente de algunos hechos o propiedades que permanecen ocultos, y que un cambio de escala o el uso de una medida derivada puede destacarlos. Este es el objeto de esta sesión teórica: el uso de medidas derivadas e índices que faciliten el conocimiento de hechos no percibidos en la medida inicial.

Medidas o puntuaciones típicas

Medida directa

Llamaremos *medida directa* en cualquier estudio o experimento, a aquella que se ha obtenido directamente sobre los objetos, individuos o entidades con los instrumentos usuales de medida.

Así, son medidas directas: la estatura en cm., la edad en años, la producción de una fábrica en toneladas, etc.

Sobre esta medida directa, mediante operaciones matemáticas o de ordenación, se pueden establecer otras

medidas *derivadas* que informen del mismo fenómeno destacando otros aspectos. Por ejemplo: la producción de la fábrica en pesetas constantes de 1990, la estatura de una persona en comparación con la del año pasado, la edad comparada con el resto de su colectivo, etc.

La medida directa tiene el defecto de no informarnos sobre la posición o nivel que ese dato tiene dentro de su grupo.

Medida diferencial

Dada una medida directa X , llamaremos **medida diferencial** x a su diferencia con la media del grupo:

$$x = X - \bar{X}$$

Si recuerdas la propiedad de la media

$$\sum X = N\bar{X}$$

comprenderás que la suma de las medidas diferenciales será igual a cero y que, además, unas serán positivas y otras negativas

$$\sum x = 0$$

La consecuencia es que **la media de las medidas diferenciales siempre es cero**, y se puede demostrar que la desviación típica s de las medidas diferenciales es la misma que la de las medidas directas

La medida diferencial nos informa sobre lo cerca o lejos que se encuentra un dato respecto a la media. Es, por tanto, representativa de la situación del individuo dentro de su grupo, pero no nos permite evaluar si esa distancia es importante o no.

En realidad, es una simple traslación.

Medida típica Z

Si se divide una medida diferencial entre la desviación típica del grupo, se obtiene la *medida o puntuación típica Z*:

$$Z = \frac{X - \bar{X}}{\sigma}$$

Esta medida es muy importante, pues permite comparar dos colectivos distintos, debido a la siguiente propiedad:

La media de las puntuaciones Z siempre es cero y su desviación típica siempre es 1

De esta forma, mediante Z, las medidas obtenidas por cualquier sujeto en variables diferentes, siempre tendrán media 0 y desviación 1, con lo que Z mide **el verdadero nivel** dentro de cada grupo, al haber eliminado los parámetros de centro y dispersión. Es como si dos conjuntos los redujéramos a la misma escala para poderlos comparar.

Las medidas Z comprendidas entre -2 y 2 suponen como mínimo el 75% de los datos. Así, puntuaciones superiores a 2 o inferiores a -2 son extraordinarias, en el sentido de que lo probable es lo contrario.

De igual forma, entre -3 y 3 están contenidos al menos el 89% de los datos. Puntuaciones más alejadas que 3 y -3 se consideran improbables.

Otras medidas

Hemos visto que la puntuación Z suele estar entre -3 y 3 y, por tanto, puede ser positiva o negativa y, en general, con decimales. Para simplificar su lectura, especialmente en Psicología y Ciencias de la Educación, se han introducido otras medias convencionales. Las más importantes son:

Escala T

La puntuación T se obtiene multiplicando Z por 10 y después sumando 50:

$$T = 10Z + 50$$

Escriba aquí la ecuación.

Lo normal es que una puntuación T oscile entre 20 y 80 puntos. Las medidas más extremas son improbables.

Índices de posición

Cuantiles

Si recordamos que la mediana es un punto convencional en una distribución de frecuencias caracterizado por la propiedad de tener a la mitad de los datos inferior a él, ¿no podríamos considerar también un punto que tuviera sólo un 25% inferior? ¿o el punto que sólo tiene el 10% de los datos superior a él? Estas consideraciones han llevado a la idea de **cuantil**:

Diremos que un número es el **cuantil de orden p** en una distribución de frecuencias si el porcentaje de datos inferiores a él es igual a **p** (y los superiores 100-p).

Por ejemplo, el cuantil C_{85} será un punto que cumple que el 85% de los datos sea inferior o igual a él.

Cálculo con datos aislados

Si los datos están aislados, se calcula el **p%** del total de datos y se va contando de menor a mayor hasta llegar a ese número de datos:

En el conjunto 1 1 1 2 2 3 3 3 3 4 4 5 6 6 7 8 9 9 de 18 datos ¿dónde se encuentra el cuantil C_{60} ? Calculamos el 60% de 18, que son 10,8 datos, redondeando, 11 datos. Contamos, pues, los 11 primeros y llegamos al 4, luego **el número 4 es el cuantil del 60%**.

Cálculo con datos agrupados

Su fórmula es similar a la de la mediana:

$$C_p = L_i + \frac{N \cdot p - n_{ant}}{n_{interv}} \cdot Ampl.$$

En la fórmula los símbolos se interpretan así:

$N \cdot p$ es el producto del número de datos por la proporción que define el cuantil (si es porcentaje, se deberá después dividir entre 100)

$Ampl.$ es la amplitud del intervalo contiene la frecuencia acumulada **p**

n_{ant} es la frecuencia acumulada anterior al intervalo del cuantil

$n_{interv.}$ es la frecuencia absoluta de dicho intervalo.

L_i representa el límite inferior verdadero del intervalo del cuantil.

Cuartiles

Los cuantiles que dividen a la distribución **en cuatro partes iguales**, es decir, C_{25} , C_{50} y C_{75} , reciben el nombre de **cuartiles**, y se representan por

Q1 o primer cuartil es el número que deja inferiores a él un 25% de los datos.

Q2 o segundo cuartil o mediana es el número que deja inferiores a él un 50% de los datos.

Q3 o tercer cuartil es el número que deja inferiores a él un 75% de los datos.

Su cálculo se efectúa como en los demás cuantiles.

Los cuartiles son muy interesantes como alternativa a las medidas de variabilidad y asimetría:

La variabilidad de una distribución (ver sesión 2) se puede medir también mediante el

Semiintervalo intercuartílico

Esta medida se obtiene mediante la fórmula

$$I = \frac{Q_3 - Q_1}{2}$$

y es una buena alternativa a la desviación típica si los datos son de tipo ordinal, si existen intervalos abiertos o si la distribución es muy asimétrica.

Deciles

Se suelen definir **9 deciles D1, D2, ... D9**, que son los puntos que dividen al intervalo en **diez partes iguales**, correspondientes a los cuantiles de porcentajes 10%, 20%, ...90% respectivamente.

Percentiles (o centiles)

Similares a los anteriores, **P1, P2, P3,P99**, son 99 números que dividen la distribución en 100 partes iguales. Son muy usados en Ciencias del Comportamiento.

Quintiles

Son menos usados: equivalen a los percentiles 20, 40, 60 y 80, y dividen el conjunto en cinco partes iguales.

Rango percentil

Es la medida inversa del **percentil**. Dada una medida concreta, como puede ser la calificación de una alumna en Música, su rango percentil equivale al **percentil más cercano** a esa calificación. Un alumno que tenga rango percentil de 78 es aquel en el que el 78% de sus compañeros tiene una puntuación inferior a él.

Para calcular el rango percentil de una medida cualquiera en datos aislados basta contar los inferiores a él, dividir ese número entre el total, multiplicar por 100 y redondear:

Si los datos están agrupados, la fórmula para calcular el rango percentil es:

$$RP = F_{ant.} + \frac{X - L_i}{Ampl.} \cdot f_{int.},$$

donde **RP** es el rango percentil, **X** la medida directa, **$f_{ant.}$** la frecuencia relativa anterior al intervalo que contiene a X, **L_i** el límite inferior verdadero de ese intervalo, **Ampl.** la amplitud del mismo y **$f_{int.}$** su frecuencia relativa

Números índices

Estudio de series temporales o espaciales

En muchos estudios estadísticos nos encontramos con series de muchos datos consecutivos, que suelen estar repartidos temporalmente (por días, meses o años) o, más raramente, de forma espacial (los distintos grupos de 2º de ESO que existen en un Centro), de los que deseamos averiguar qué variación suponen unos respecto a otros. Por ejemplo:

¿Con 100 ptas. del año 1980, qué compraría hoy?

¿Cómo ha ido creciendo la población de mi pueblo en estos últimos años?

¿Ha habido algún progreso en este grupo de alumnos desde Octubre hasta Mayo?

Los números índices se usan por varias razones:

Permiten comparar unos valores con otros en un serie de forma porcentual.

Añaden claridad a algunos procesos.

Con ellos se pueden comparar series de distinta índole.

Su propiedad multiplicativa simplifica algunos cálculos.

Índice simple de base fija

Un término de la serie se identifica (convencionalmente) con el número 1, o el 100%. Diremos que este valor y_0 posee el índice 1. Para el resto de valores se define el índice como **el cociente entre su propio valor y_i y el valor y_0 identificado como de índice 1**. Eventualmente,

multiplicaremos por 100 si deseamos expresarlo como porcentaje, pero en los cálculos se deja así.

$$I_0 = \frac{y_i}{y_0} \cdot 100$$

Índice simple de base variable (o en cadena)

Tiene la misma definición que el anterior, pero en lugar de elegir un valor y_0 como base, en el cociente se toma el término anterior y_{i-1} .

$$I_i = \frac{y_i}{y_{i-1}} \cdot 100$$

La propiedad fundamental es la siguiente:

El índice definido entre dos términos de una serie no consecutivos equivale al producto de todos los índices en cadena intermedios.

Así, si el coste de vida en el 1993 fue en un país del 3%, en 1994 del 5% y en 1995 del 2%, los índices en cadena serían 1.03, 1.05 y 1.02 respectivamente. Para calcular el incremento habido entre 1993 y 1996, deberíamos multiplicar los tres índices, y nos resultaría 1,10313, es decir, un 10,31%. Si hubiéramos sumado $3+5+2 = 10$, se hubiera producido un error.

Los índices en cadena son multiplicativos y no aditivos.

Índice compuesto

Cuando se desea comparar la evolución de varios conjuntos a la vez, se definen *índices compuestos*,

obtenidos combinando los índices simples. En general es una operación compleja (recuérdese el índice de coste de la vida). Una técnica sencilla es sustituir los múltiples valores de cada término por su media ponderada.