ENRIC RUIZ MORILLAS

# ESTADÍSTICA APLICADA A EXPERIMENTOS Y MEDICIONES

DIAZ DE SANTOS

# Índice

1. Experimento y medición .......... 1
2. Frecuencia y probabilidad .......... 3
3. Características teóricas de las variables aleatorias .......... 25
4. Distribuciones discretas .......... 39
5. Distribuciones continuas .......... 47
6. Distribuciones muestrales .......... 61
7. Estimación de parámetros poblacionales .......... 77
8. Comparación de poblaciones .......... 89
9. Bibliografía .......... 99

# 1

# Experimento y medición

Podría definirse un experimento como un proceso que genera un conjunto de datos.

El experimento está formado por una serie de condiciones fijadas por el experimentador para la obtención de unos resultados mediante una medición.

Sin embargo, los resultados obtenidos siempre están afectados en mayor o menor grado por condiciones no controladas y por errores en la medida. A consecuencia de ello surgen una serie de fenómenos aleatorios que introducen desviaciones en el fenómeno que se quiere observar.

Una manera de conocer la influencia de los fenómenos aleatorios consiste en repetir el experimento varias veces. De esta forma se descubren regularidades dentro de la variación de los resultados, que permiten determinar el efecto de los fenómenos aleatorios sobre los fenómenos a observar.

La *Teoría de las probabilidades* estudia estas regularidades.

No es factible obtener *todos* los resultados posibles de un experimento dado. En la práctica solo se obtiene un conjunto limitado de estos. La *Estadística* permite caracterizar el conjunto de todos los resultados posibles de un experimento a partir del conjunto de datos experimentales obtenidos.

# 2

# Frecuencia y probabilidad

Los resultados de un experimento pueden ser caracterizados cualitativamente y cuantitativamente. A las características cualitativas de los resultados se las llama *acontecimientos* o *sucesos*, y a las cuantitativas *magnitudes o variables aleatorias*.

## Sucesos y frecuencia

Dado un experimento, podemos distinguir tres tipos de sucesos: un suceso se denomina *cierto* si se produce necesariamente en el experimento, *imposible* si el suceso no puede tener lugar, y *aleatorio* si puede ocurrir o no.

Consideremos un experimento en el que se obtiene como un resultado posible el suceso A. Al repetir el experimento $n$ veces el suceso A puede producirse $m$ veces ($m \leq n$). La relación entre el número de veces que este suceso se produce y el número de experimentos realizados se denomina *frecuencia*.

Si designamos la frecuencia del suceso A por $P^*(A)$ se puede escribir:

$$P^*(A) = \frac{m}{n}$$

## Propiedades de la frecuencia de los sucesos

La frecuencia de cualquier suceso se sitúa entre 0 y 1.

$$0 \leq P^*(A) \leq 1$$

La frecuencia de un suceso imposible siempre es igual a cero, y la frecuencia de un suceso cierto es siempre igual a uno.

Dos o más sucesos son *incompatibles* o *mutuamente excluyentes* en el experimento dado si la obtención de uno de ellos excluye la obtención de los otros.

Dados el suceso A y el suceso B, incompatibles entre sí, el suceso que consiste en que tenga lugar el suceso A o el suceso B se denomina *suceso compuesto* A o B.

*Teorema de adición de frecuencias* – La frecuencia de obtención del suceso compuesto A o B es igual a la suma de las frecuencias de cada suceso.

$$P^*(A\ o\ B) = P^*(A) + P^*(B)$$

Diremos que el suceso compuesto A o B es el resultado de la *adición* de los sucesos A y B. Por lo que también lo representaremos como el suceso A + B.

Por lo tanto, este teorema también puede escribirse así:

$$P^*\ (A + B) = P^*\ (A) + P^*\ (B)$$

En el caso de que el experimento tenga como resultado posible el suceso A y además el suceso B, puede calcularse otra frecuencia que informa sobre la influencia del suceso B en la obtención del suceso A. Esta frecuencia se denomina *frecuencia condicional* del suceso A con respecto al suceso B, y se designa como $P^*(A/B)$.

Si al efectuar *n* experimentos el suceso B ha tenido lugar *l* veces y el suceso A ha tenido lugar junto al suceso B *k* veces, entonces:

$$P^*(A/B) = \frac{k}{l}$$

Como:

$$P^*(A\ y\ B) = \frac{k}{n}$$

$$P^*(B) = \frac{l}{n}$$

Entonces:

$$P^*(A/B) = \frac{P^*(A\ y\ B)}{P^*(B)}$$

Análogamente puede calcularse la frecuencia condicional del suceso B con respecto al suceso A:

$$P^*(B/A) = \frac{P^*(A\ y\ B)}{P^*(A)}$$

Expresiones que permiten enunciar el siguiente teorema:

*Teorema de multiplicación de frecuencias* – La frecuencia de la obtención conjunta de los sucesos A y B es igual al producto de la frecuencia de uno de ellos por la frecuencia condicional del otro con respecto al primero.

$$P^*(A\ y\ B) = P^*(A)\, P^*(B/A) = P^*(B)\, P^*(A/B)$$

Diremos que el suceso en el que se obtienen de manera conjunta los sucesos A y B es el resultado de la *multiplicación* de los sucesos A y B. Por lo que también lo representaremos como el suceso AB.

Por lo tanto, este teorema también puede escribirse así:

$$P^*(AB) = P^*(A)\, P^*(B/A) = P^*(B)\, P^*(A/B)$$

## Variables aleatorias

A cada resultado posible de un experimento se le puede asociar un número real mediante la función denominada *variable aleatoria*.

$$X \ : S \longrightarrow \mathbb{R}$$

$$w \longrightarrow X(w) = x \ \text{(o abreviadamente } X = x\text{)}$$

X es la función variable aleatoria. Los valores que puede tomar la variable aleatoria, como consecuencia del experimento, se llaman *valores posibles* y se representan por x.

S se denomina *espacio muestral* y está formado por todos los resultados posibles del experimento. Los subconjuntos de S son los sucesos.

## Propiedades de las variables aleatorias

- *Suma de variables aleatorias* – La suma de dos variables aleatorias es otra variable aleatoria.

  Así, los valores posibles $Z(w) = z$ de la variable aleatoria Z definida como $Z = X + Y$, donde X e Y son dos variables aleatorias, son $Z(w) = X(w) + Y(w)$; es decir, $z = x + y$.
- *Producto de variables aleatorias* – El producto de dos variables aleatorias es otra variable aleatoria.
- *Producto de una variable aleatoria por un número real* – El producto de una variable aleatoria por un número real es otra variable aleatoria.
- *Función de una variable aleatoria* – Si X es una variable aleatoria y g es una función real de variable real, entonces $Y = g(X)$ es una variable aleatoria.

## Función estadística de distribución

Para toda variable aleatoria se puede definir en el espacio muestral el suceso que representa el cumplimiento de la desigualdad $X < x$. Entonces se puede determinar la frecuencia de este suceso para diferentes valores de x, lo cual permite definir la que se denomina *función estadística de distribución* de la variable aleatoria X:

$$F^*(x) = P^*(X < x)$$

Consideremos la realización de *n* experimentos y que la variable aleatoria X ha tomado los valores posibles $x_1, x_2, \ldots, x_n$, ordenados de forma creciente.

Para todo valor de x en el intervalo $x_k < x \leq x_{k+1}$ el número de experimentos en los que se ha cumplido la desigualdad $X < x$ es k.

Por tanto:

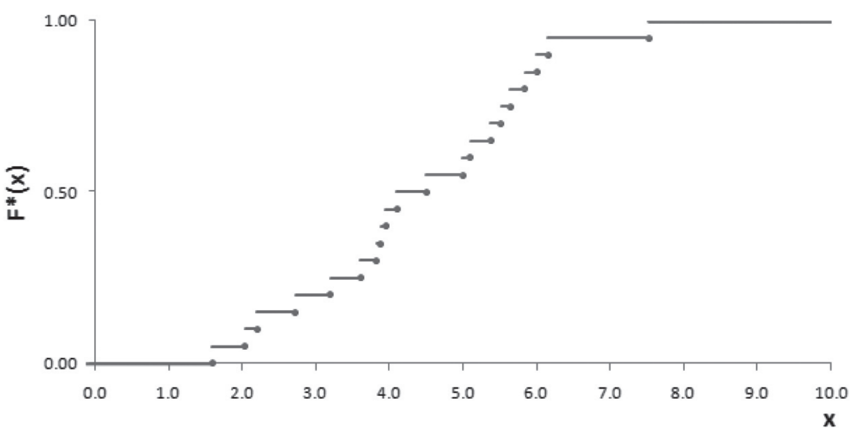
$$F^*(x) = \frac{k}{n} \quad si \quad x_k < x \leq x_{k+1} \quad (k = 1, \dots, n-1)$$

La función estadística de distribución así definida es una función escalonada.

**Ejemplo**. Consideremos los resultados de los 20 experimentos mostrados en la tabla.

| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 5.4 | 2.7 | 3.8 | 6.2 | 7.5 | 4.1 | 5.1 | 3.2 | 1.6 | 5.0 |
| i | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $x_i$ | 5.8 | 4.0 | 3.6 | 2.2 | 5.6 | 4.5 | 2.0 | 5.5 | 3.9 | 6.0 |

La función estadística de distribución se representa en la siguiente figura:



## Valor medio

La *característica experimental* más simple de la variable aleatoria X es su *media o valor medio estadístico*:

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

La media proporciona información sobre la localización de los resultados del experimento; es decir, del valor alrededor del cual se sitúan.

Observemos que $\frac{1}{n}$ representa el incremento de la función estadística de distribución en cada uno de los valores posibles tomados por la variable aleatoria X en los *n* experimentos. Por tanto, también podemos escribir:

$$\bar{x} = \sum_{i=1}^{n} x_i \ \Delta F^*(x_i)$$

## Dispersión

Una manera de caracterizar la dispersión alrededor de la media de los valores de la variable aleatoria X en los experimentos realizados es mediante el valor medio del cuadrado de la diferencia entre cada uno de los valores obtenidos y el valor medio de la variable aleatoria X. A esta magnitud la denominaremos *dispersión estadística:*

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

$$D_x^* = \sum_{i=1}^{n} (x_i - \bar{x})^2 \ \Delta F^*(x_i)$$

Habitualmente, para caracterizar la dispersión alrededor de la media de los valores de la variable aleatoria X se utiliza la *desviación* *estándar*, magnitud que tiene la misma dimensión que los valores que puede tomar la variable aleatoria X:

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

La desviación estándar informa del intervalo alrededor de la media en el que se sitúan los resultados del experimento.

**Ejemplo**. En el ejemplo anterior la media y la desviación estándar de los resultados experimentales son:

$$\bar{x} = 4.4$$
$$s_x = 1.6$$

Puede observarse que aproximadamente 2/3 partes de los resultados se encuentran dentro del intervalo $(\bar{x} - s_x,\ \bar{x} + s_x)$ y que, en este caso, la totalidad de los resultados se encuentra dentro del intervalo $(\bar{x} - 2s_x,\ \bar{x} + 2s_x)$.

## Correlación

En un experimento en el que se obtienen dos tipos de resultados, a los que les corresponden las variables aleatorias X e Y, puede suceder que exista alguna dependencia entre los dos tipos de resultados. Por ello, para caracterizar el grado de dependencia entre las variables aleatorias en una serie dada de *n* experimentos, se introduce la *covarianza* entre X e Y:

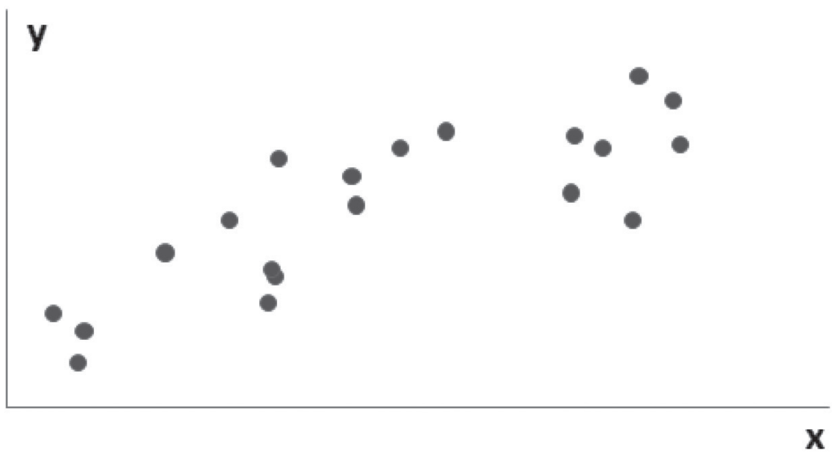
$$Cov(X,Y) = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Si la covarianza entre X e Y es próxima a cero en una serie de experimentos suficientemente grande, entonces las variables aleatorias X e Y pueden considerarse independientes. Una covarianza entre X e Y distinta de cero indica que entre X e Y existe una relación.

**Ejemplo**. Consideremos los resultados de los 20 experimentos mostrados en la tabla.

| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 2.3 | 8.2 | 9.8 | 1.1 | 5.0 | 5.7 | 3.9 | 3.8 | 3.2 | 1.0 |
| $y_i$ | 5.4 | 7.6 | 9.2 | 2.7 | 8.2 | 9.1 | 4.6 | 3.7 | 6.6 | 1.5 |
| i | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $x_i$ | 4.0 | 5.1 | 0.7 | 3.9 | 9.7 | 8.3 | 9.2 | 8.7 | 9.1 | 6.4 |
| $y_i$ | 8.8 | 7.1 | 3.3 | 4.8 | 10.8 | 9.6 | 11.7 | 9.2 | 6.6 | 9.7 |

Los pares de resultados de estos 20 experimentos se representan en la siguiente gráfica, denominada *gráfica de dispersión*.




La covarianza entre X e Y es:

$$Cov(X,Y) = 7.2$$

Otra característica que permite evaluar el grado de asociación entre dos variables aleatorias es el *coeficiente de correlación*:

$$r = \frac{Cov(X,Y)}{s_x\, s_y}$$