

TEMA 0. REPASO BÁSICO DE CONCEPTOS DE ESTADÍSTICA

0.1. Estadística descriptiva

- 0.1.1. Conceptos iniciales
- 0.1.2. Índices de tendencia central
- 0.1.3. Índices de posición
- 0.1.4. Medidas de variabilidad
- 0.1.5. Puntuaciones directas, diferenciales y típicas
- 0.1.6. Relación entre variables

0.2. Modelos continuos de probabilidad

- 0.2.1. Distribución normal
- 0.2.2. Distribución t de Student
- 0.2.3. Distribución χ^2 de Pearson
- 0.2.4. Distribución F de Fisher-Snedecor

TEMA 0. REPASO BÁSICO DE CONCEPTOS DE ESTADÍSTICA

La estadística es la rama de las matemáticas que se encarga del estudio de determinadas características en una población, recogiendo los datos, agrupándolos, organizándolos en tablas, representándolos gráficamente y analizándolos para sacar conclusiones de dicha población.

Podemos considerar dos grandes áreas: la estadística descriptiva y la estadística inferencial. Esta última es la que se tratará en el temario correspondiente a la asignatura de Diseños de investigación y análisis de datos. Por lo tanto, en este tema 0 nos centraremos en un repaso rápido de la descriptiva.





0.1. Estadística descriptiva

0.1.1. Conceptos iniciales

Se organizan y se resumen conjuntos de observaciones cuantificadas procedentes de una muestra o de la población total.

Población: Conjunto de todos los elementos que cumplen una determinada característica objeto de estudio.

Muestra: Es un subconjunto cualquiera de una población.

Parámetro: Es una propiedad descriptiva (una medida) de una población, Se denota con letras griegas.

Estadístico: Es una propiedad descriptiva (una medida) de una muestra. Se denota con letras latinas.

	MUESTRA	POBLACIÓN
MEDIA	\bar{X}	μ
VARIANZA SESGADA	$S_n^2 = S_x^2$	σ_x^2
CUASIVARIANZA	$S_{n-1}^2 = \hat{S}_x^2$	
DESVIACIÓN TÍPICA SESGADA	$S_n = S_x$	σ_x
CUASIDESVIACIÓN TÍPICA	$S_{n-1} = \hat{S}_x$	
PROPORCIÓN	p	π





VARIABLES

Una variable es un conjunto de valores resultantes de medir una característica de interés sobre cada elemento individual de una población o muestra. Se representan con letras latinas en mayúsculas.

Tipo de variable	Escala de medida	Características básicas	Relaciones válidas	Ejemplos
<u>CUALITATIVA:</u> - Dicotómica - Politémica	<u>Nominal</u>	Los números identifican y clasifican objetos	Relaciones del tipo "igual que" o "distinto que"	Sexo, estado civil, raza, diagnóstico clínico
<u>CUASICUANTITATIVA</u>	<u>Ordinal</u>	Además, los números indican las posiciones relativas de los objetos	Además, relaciones del tipo "mayor que" o "menor que"	Dureza, posición en el ranking ATP, grado de satisfacción.
<u>CUANTITATIVA:</u> - Discreta - Continua	<u>Intervalo</u>	Además, hay una unidad de medición común	Además, igualdad o desigualdad de diferencias	Temperatura en grados centígrados, inteligencia
	<u>Razón</u>	Además, el punto cero es absoluto	Además, igualdad o desigualdad de razones	Longitud, peso, altura, tiempo de reacción



0.1.2. Índices de tendencia central.

Dentro de los índices de tendencia central, tenemos la moda, la mediana y la media. En este repaso solo vamos a hacer hincapié en la media.

MEDIA

Se define como la suma de todas las puntuaciones de la distribución, dividida por el total de los casos u observaciones.

$$\bar{X} = \frac{\sum n_i X_i}{n}$$

donde:

n = Número total de observaciones

X_i = Cada uno de los valores de la variable o el punto medio si tenemos intervalos

n_i = Frecuencia absoluta de valor X o del intervalo



Vamos a calcular la media con la distribución de frecuencias de los alumnos que han aprobado el examen de Introducción al Análisis de datos de la UNED.

X	n _i	n _i · X _i
5	5	25
6	7	42
7	8	56
8	7	56
9	1	9
10	1	10
	n=29	198

$$\bar{X} = \frac{\sum n_i \cdot X_i}{n}$$

$$\bar{X} = \frac{198}{29} = 6,83$$

0.1.3. Índices de posición

Son los índices que permiten establecer la posición relativa de una puntuación respecto al grupo de puntuaciones y se suelen llamar cuantiles.

En este repaso lo que nos interesa es entender qué información nos facilita este tipo de índices, el cálculo como tal no lo necesitamos de cara a la asignatura de Diseños de investigación.

Dentro de los cuantiles podemos tener: percentiles, cuartiles y deciles.

PERCENTILES

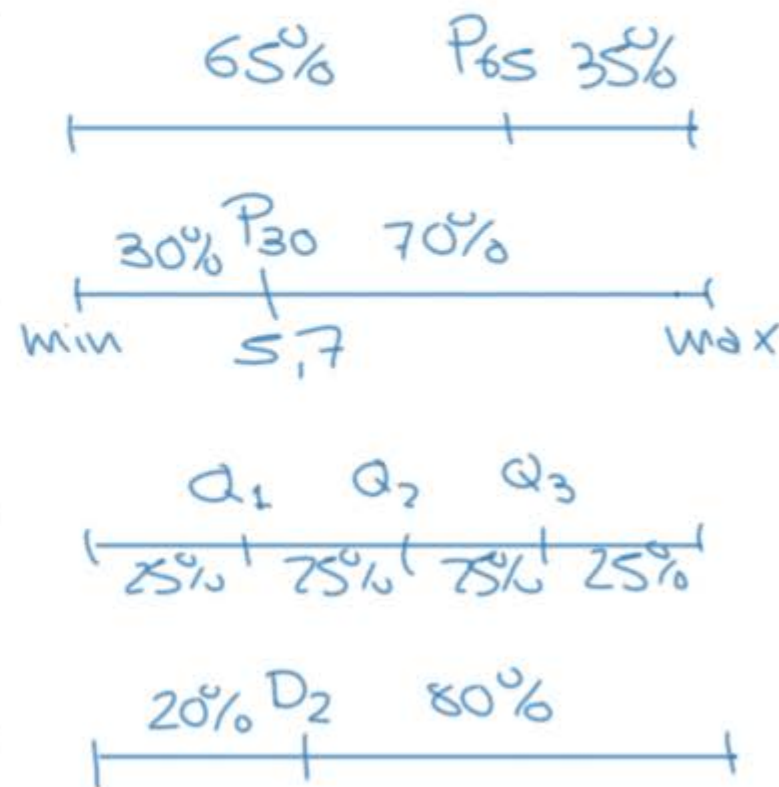
Son los 99 valores que dividen en 100 partes iguales la distribución de frecuencias de la variable.

CUARTILES

Son los 3 valores que dividen en 4 partes secciones, cada una conteniendo el 25% de las observaciones.

DECILES

Son los 9 valores que dividen en 10 partes iguales la distribución de frecuencias de la variable.



0.1.4. Medidas de variabilidad

Las medidas de variabilidad nos informan sobre en qué medida las puntuaciones en una determinada variable están próximas o alejadas del índice de tendencia central.

VARIANZA SESGADA

Se representa por S_x^2 y se define como el promedio de los cuadrados de las desviaciones de las puntuaciones con respecto a la media. Siempre es positiva.

$$S_n^2 = S_x^2 = \frac{\sum n_i \cdot X_i^2}{n} - \bar{X}^2$$

DESVIACIÓN TÍPICA SESGADA

Se representa por S_x y es la raíz cuadrada positiva de la varianza.

$$S_n = S_x = \sqrt{S_x^2}$$

CUASIVARIANZA

Se representa por S_{n-1}^2 y también se la conoce como varianza insesgada. Se calcula a través de la siguiente fórmula.

$$S_{n-1}^2 = \frac{\sum n_i (X_i - \bar{X})^2}{n - 1} = \hat{S}_x^2$$

CUASIDESVIACIÓN TÍPICA

Se representa por S_{n-1} y también se la conoce como desviación típica insesgada. Se calcula a través de la siguiente fórmula.

$$S_{n-1} = \sqrt{\frac{\sum n_i (X_i - \bar{X})^2}{n - 1}} = \hat{S}_x$$



Vamos a calcularlas con la distribución de frecuencias de los alumnos que han aprobado el examen de Introducción al Análisis de datos de la UNED.

$$\bar{x} = 6,83$$

X	n _i	n _i · X _i ²	n _i · (X _i - \bar{x}) ²
5	5	125	16,74
6	7	252	4,82
7	8	392	0,23
8	7	448	9,58
9	1	81	4,71
10	1	100	10,05
	n=29	1398	46,13

$$S_n^2 = S_x^2 = \frac{\sum n_i \cdot X_i^2}{n} - \bar{x}^2 = \frac{1398}{29} - 6,83^2 = 1,58$$

$$S_n = S_x = \sqrt{S_x^2} = \sqrt{1,58} = 1,26$$

S_x^2 → S_x
 elevar al cuadrado

otra fórmula:

$$S_{n-1}^2 = \frac{\sum n_i (X_i - \bar{x})^2}{n-1} = \frac{46,13}{29-1} = 1,65$$

$$S_{n-1} = \sqrt{1,65} = 1,28$$

0.1.5. Puntuaciones directas, diferenciales y típicas

Las puntuaciones directas son con las que tratamos habitualmente (puntuaciones de un sujeto en un test, etc.), pero la comparación de las puntuaciones directas de un mismo sujeto en dos variables distintas puede llevarnos a confusión, ya que las puntuaciones directas nos ofrecen muy poca información. De hecho, conocida una puntuación directa no sabemos si se trata de un valor alto o bajo porque esto depende del promedio del grupo.

Una posible solución es trabajar con puntuaciones diferenciales. Si a una puntuación directa le restamos la media de su grupo obtenemos una puntuación diferencial.

$$x_i = X_i - \bar{X}$$

Aportan más información ya que podemos saber si la puntuación coincide con la media, si es inferior o es superior a ella.

Estas puntuaciones tienen dos propiedades:

- La media es 0
- La varianza de las puntuaciones diferenciales es igual a la varianza de

las puntuaciones directas.



Sin embargo, dos puntuaciones diferenciales idénticas pueden tener un significado muy diferente en función de la media y de la varianza de las distribuciones de las que proceden. Para eliminar este inconveniente se utilizan las puntuaciones típicas. Las puntuaciones típicas van más allá y nos permiten no sólo comparar las puntuaciones de un sujeto en dos variables distintas, sino también comparar dos sujetos distintos en dos pruebas o variables distintas.

Una puntuación típica se calcula con la siguiente fórmula:

$$z_x = \frac{x_i}{S_X} = \frac{X_i - \bar{X}}{S_X}$$

Al proceso de obtener las puntuaciones típicas se llama tipificación. Una puntuación típica indica el número de desviaciones típicas que se aparta de la media una determinada puntuación.

Sus propiedades son:

- La media es igual a 0.
- La varianza es igual a 1

