



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

1. INTRODUCCIÓN

Debemos diferenciar entre dos tipos de estadísticas:

Estadística teórica: Se ocupa de los aspectos formales y normativos

Estadística aplicada: Constituye la aplicación a un campo concreto. Ha tenido distintos nombres, entre ellos "Análisis de Datos"

2. LA INVESTIGACIÓN EN PSICOLOGÍA

S.XVII: El método científico se convierte en la fuente de conocimiento más utilizada. La Psicología utiliza el método científico para acercarse a su objeto de estudio: **La conducta**

El método científico

Definición: Consiste en dar razón sistemática, empírica y experimental de los fenómenos.

Características

Sistemático: Tiene etapas definidas (**sigue un sistema**)

Replicable: Los datos obtenidos pueden ser utilizados por otro investigador

Fases

1. Definición del Problema
2. Deducción de hipótesis contrastables
3. Establecimiento de un procedimiento de recogida de datos
4. Análisis de resultados obtenidos (**de este punto se ocupa la asignatura**)
5. Discusión de dichos resultados y búsqueda de soluciones
6. Elaboración de un informe de investigación

A continuación repito los puntos anteriores pero ejemplificados con el ejemplo de la página 17.

1. Posible influencia de las estrategias sobre el control del estrés en la hipertensión
2. El grupo que recibe terapia tendrá unos niveles menores en la hipertensión
3. Escoger 40 pacientes y dividirlos en 2 grupos de manera aleatoria
- 4 y 5: Se analizan los niveles de los 2 grupos para comprobar si realmente el grupo que ha recibido tratamiento tiene los niveles más bajos que el grupo que no lo ha recibido.
6. Elaboramos un informe para difundir los resultados.

3. CONCEPTO Y FUNCIONES DE LA ESTADÍSTICA: DESCRIPTIVA e INFERENCIAL

La estadística se ocupa de la sistematización, recogida, ordenación, y presentación de los datos referentes a un fenómeno que presenta variabilidad o incertidumbre para su estudio metódico, con objeto de hacer previsiones sobre los mismos, tomar decisiones u obtener conclusiones.

A partir de aquí obtenemos dos grandes áreas:



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

Estadística descriptiva

Se organizan y resumen conjuntos de observaciones procedentes de una muestra o de la población total, en forma cuantitativa.

Procedimientos

(no entro en detalle porque dice que los veremos en profundidad en los siguientes temas)

Para una variable

- Índices de tendencia central
- Estadísticos de variabilidad
- Estadísticos de asimetría

Para dos variables

- Coeficientes de correlación
- Ecuaciones de regresión

Estadística Inferencial

Se realizan inferencias acerca de una población basándose en los datos obtenidos a partir de una muestra.

Procedimientos : El cálculo de probabilidades

Conceptos importantes

Población: Es el conjunto de todos los elementos que cumplen una determinada característica objeto de estudio. (según el ejemplo de antes, la población serían todas las personas que padecen hipertensión)

Muestra: Es un subconjunto cualquiera de una población (las 40 personas escogidas)

Parámetro: Es una propiedad descriptiva (una medida) de una población. (para una encuesta electoral queremos conocer el porcentaje de la población que respondería "Sí" a favor de una ley; ese "sí" sería el parámetro en este caso)

Estadístico: Es una propiedad descriptiva (una medida) de una muestra. (es el porcentaje de la muestra que responde sí).

Las conclusiones obtenidas de una muestra solo servirán para el total de una población si la muestra es **representativa**. Para asegurarnos que la muestra es representativa se utilizan métodos de **muestreo probabilístico**.

Si queremos saber qué porcentaje de hombres españoles se declaran seguidores del Real Madrid y la muestra que escogemos está formada por mujeres, los resultados que obtengamos no van a ser representativos.

También existen las **muestras no probabilísticas** como por ejemplo la muestra de conveniencia o incidental (personas cercanas al investigador, por lo cual acceder a ellas es mucho más fácil)



4. MEDICIÓN Y ESCALAS DE MEDIDA

Medir la altura de una persona es algo sencillo, sin embargo en ocasiones la Psicología debe medir cosas más complicadas, como por ejemplo la timidez de una persona. Para ello existen numerosas herramientas que nos ayudan.

Medición: Es el proceso por el cual se asignan números a objetos o características según determinadas reglas.

Característica: Cualquier propiedad de objetos o personas que deseamos estudiar.

Modalidad: Distintas formas de presentarse esta característica.

Por lo tanto...

Se asigna un número a cada una de las modalidades de una característica.

Ejemplo: A las dos modalidades de la característica sexo (hombre y mujer) le asignamos los números 1 y 2.

Tipos de escala de medida

Según las relaciones que puedan verificarse empíricamente entre las modalidades de las características, distinguimos 4 tipos de escalas:

1) Escala Nominal

Asignación de números o símbolos. Por lo tanto la única relación que obtenemos es la de **igualdad** o **desigualdad**.

Ejemplo: Católicos practicantes (1) frente a católicos no practicantes (2). No tiene sentido hacer relaciones como $1+1=2$. Esta escala solo nos dice si un sujeto es diferente o igual a otro.

2) Escala Ordinal

Como su propio nombre indica, aparte de decirnos la igualdad o la diferenciación, también nos indica el **orden** de las posiciones de lo que está clasificado. Nos permite establecer relaciones del tipo "mayor que" o "menor que".

Ejemplo: Estatus socioeconómico, bajo (1), medio (2), alto (3). Los números no solo indican diferencia de modalidades sino también orden entre ellas.

3) Escala de intervalo

Además de las características de la escala nominal (igualdad o desigualdad) y de la escala ordinal (orden), la escala de intervalo se identifica porque cada número es una **unidad de medición**, podemos sumarlos o restarlos obteniendo resultados coherentes (a diferencia de la escala nominal).

Es importante señalar que en la escala de intervalo el origen es **arbitrario**



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

Ejemplo: En el conocido test de inteligencia, la escala comienza en 0, pero este 0 es arbitrario ya que si alguien lo obtiene no significa que no tenga absolutamente nada de inteligencia. Esta arbitrariedad indica que aunque una persona obtenga un coeficiente de 80 y otra 160, no significa que la segunda persona sea el doble de inteligente que la primera.

4) Escala de razón

Además de todas las características de las 3 escalas diferentes, supera a las demás en que se puede establecer un punto de origen verdadero de valor 0 (a diferencia del 0 arbitrario de la escala de intervalo). Es lo que se conoce como valor **absoluto**.

Ejemplo: La altura y el peso..

En la página 25 del libro tenemos un cuadro resumen que explica todo lo anterior de manera bastante esquemática.

5. VARIABLES: CLASIFICACIÓN Y NOTACIÓN

Variable: Representación numérica de una característica que presenta más de una modalidad (valor) de un conjunto determinado. Si una característica tiene una única modalidad se llama **constante**.

Según el nivel de medición tenemos tantas variables como escalas (variable nominal, ordinal, de intervalo, de razón).

Tipos de variables

A) Nominal (Cualitativa)

Dicotómica: 2 categorías (sexo)

Politómica: + de 2 categorías (nacionalidad)

B) Ordinal (Cuasicuantitativa)

C) De intervalos, razón (cuantitativa)

Discreta: Valores fijos (número de hijos). Entre 1 y 2 hijos no pueden haber 1,5 hijos.

Continua: Puede haber medidas intermedias entre los valores. Por ejemplo el peso (entre 89 y 90 podemos tener 89,5).



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

Aparte de esta clasificación existen más clasificaciones:

- **Variable independiente:** Cualquier suceso que sospechamos es la causa de otro
- **Variable dependiente:** Medida utilizada para estudiar la variable independiente
- **Variable extraña:** Aquellas que pueden influir sobre la variable dependiente pero no nos interesa estudiar sus efectos.)

Ejemplo de la hipertensión visto anteriormente:

- Variable independiente: Tipo de tratamiento
- Valor 1 de la variable independiente: Tratamiento estándar
- Valor 2 de la variable independiente: Tratamiento estándar más terapia
- Variable dependiente: Hipertensión arterial
- Variable extraña: Obesidad (sabemos que la obesidad puede influir en la hipertensión arterial pero no nos interesa estudiarla en este caso)

Para referirnos a un valor cualquiera de la variable **X** se utiliza el subíndice **i** (**X_i**), siendo **n** el número de elementos que componen la muestra:

X_i siendo $i = 1, 2, 3 \dots n$

6. DISTRIBUCIÓN DE FRECUENCIAS

Una distribución de frecuencias es una representación de la relación entre un conjunto de medidas exhaustivas y mutuamente influyentes y la frecuencia de cada una de ellas.

Funciones:

- Ofrecer la información necesaria para realizar representaciones gráficas
- Facilitar los datos para obtener los estadísticos muestrales

En este punto es necesario leer y comprender los gráficos de las páginas 29 y 30. Son tan sencillos que ni siquiera es necesario sacar la calculadora. De momento lo que explica se comprende con la sola lectura. Ya tendremos tiempo de que todo se complique.

Conceptos

Frecuencia absoluta (n_i): Número de veces que se repite cada uno de los valores de una variable. La suma de todas las frecuencias absolutas representa el total de la muestra (n)

Proporción o frecuencia relativa (p_i): Cociente entre la frecuencia absoluta de cada valor de la variable (n_i) y el número total de observaciones (n). $p_i = n_i / n$

Porcentaje (P_i): Valor de la frecuencia relativa (p_i) multiplicado por cien: $P_i = p_i \times 100$

Frecuencia absoluta acumulada (n_a): Número de veces que se repita cada modalidad o cualquiera de las modalidades inferiores.



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

Proporción acumulada o frecuencia relativa acumulada (p_a): Cociente entre la frecuencia absoluta acumulada de cada clase y el total de observaciones. $p_a = n_a / n$

Porcentaje acumulado (P_a): Valor de la frecuencia relativa acumulada multiplicado por cien. $P_a = p_a \times 100$

Más conceptos

Si la variable cuantitativa es reducida (número de hijos) no hay problema, la utilizamos tal cual. Pero si ocurre que puede ser muy amplia (edad) debemos recurrir a la agrupación en **intervalos**.

La **amplitud** de los intervalos representa al número de unidades que tenemos dentro de cada intervalo. No es lo mismo tener una amplitud de intervalos de 10 en 10 (11-20) (21-30) (31-40) que de 5 en 5 (15-19) (20-24) (25-29).

Obviamente al establecer intervalos siempre se pierde información. Por lo tanto hay que buscar equilibrio entre la precisión que buscamos y lo cómodo que sea trabajar con esos datos (si no estableciésemos intervalos podríamos encontrarnos con tablas de 100 filas)

Aunque tengamos un intervalo de edad (25-35) (36-45) no significa que no pueda haber personas con 35,5 años. Por lo tanto los intervalos (25-35) y (36-45) reciben el nombre de **límites informados o aparentes**. (aparentan ser lo que no son) ya que en realidad los límites verdaderos son (25,5-35,5) y (35,5-45,5). El **límite superior exacto** del primer intervalo (35,5) coincide con el **límite inferior exacto** del siguiente (35,5)

Cálculo de límites exactos

Límite exacto = Valor informado (el que nos han dado) $\pm 0,5 \times I$

I = Instrumento de medida. (en el caso de las edades $I = 1$)

Punto Medio: Es la semisuma del límite superior e inferior del intervalo. Se calcula a través de los límites informados o de los límites exactos.

El libro asume que hemos nacido con el concepto de la "semisuma" escrito en nuestros genes ya que no explica en qué consiste. Quizá soy la única persona a la que no le han otorgado este don de la naturaleza, pero lo explicaré por si hay alguien más en mi situación:

Con un poco de sentido común el punto medio se haya sumando los dos valores del intervalo y dividiendo el resultado entre 2

Tomemos el intervalo (26-35) = $26 + 35 = 61/2 = 30,5$.

Ya sabemos lo que es la semisuma!

Intervalo abierto: Son los intervalos finales que se utilizan para no tener frecuencia nula.
Ejemplo de edad: (76 años o más)



7. REPRESENTACIONES GRÁFICAS

Imprescindible acompañar estos apuntes con el libro para comprender los conceptos gráficos.

Un gráfico es una forma rápida de visualizar un conjunto de datos o distribución de frecuencias. El sistema de coordenadas más habitual está formado por:

Eje vertical (ordenada o eje Y)

Eje horizontal (abscisa o eje X)

Origen: Punto donde se juntan ambos ejes

Representación gráfica de una variable

a) Diagrama de barras

Se utiliza para variables nominales, ordinales y cuantitativas discretas.

- En el eje X se colocan los valores y en el Y las frecuencias.

Sobre cada valor de la variable se dibuja un rectángulo o barra perpendicular cuya altura debe ser igual a la frecuencia.

Diagrama de barras acumulativo

Variedad del diagrama de barras que se utiliza en variables ordinales y cuantitativas discretas. En el eje X se sitúan los valores de la variable y en el eje Y las frecuencias acumuladas.

b) Diagrama de sectores

Se utiliza para variables cualitativas y cuasicuantitativas

- Representación en forma de círculo. El círculo se divide en secciones cuya superficie es proporcional a la frecuencia de la modalidad correspondiente. Para determinar el ángulo de los sectores de cada modalidad se multiplica la frecuencia relativa por 360, que es el número de grados de una circunferencia.

c) Pictograma

Se utiliza para variables cualitativas

- Son dibujos alusivos al objeto de estudio. Son escalas que deben ser proporcionales al valor que representan.

d) Histograma (parecido al diagrama de barras pero no hay espacio entre los rectángulos)

Se utiliza para variables cuantitativas continuas con datos agrupados en intervalos

- En el eje X se colocan los límites exactos de los intervalos y en el eje Y la frecuencia.



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

e) Polígono de frecuencias

Se utiliza para variables discretas

- La forma que obtenemos es una línea poligonal cuya figura se obtiene a partir de una diagrama de barras o de un histograma.
- También se puede utilizar en frecuencias acumuladas (figura página 41)

Representación gráfica de dos variables

a) Diagrama de barras conjunto

Se utiliza cuando al menos una de las dos variables es cualitativa

- Los datos se organizan en una tabla de doble entrada. Se sitúan los valores de una variable en las filas y los valores de la otra variable en las columnas.
- Una vez construida la tabla, se procede a dibujar los gráficos por columnas

b) Diagrama de dispersión o nube de puntos

Se utiliza en el caso de dos variables cuantitativas

- Una variable se sitúa en el eje X y la otra en el eje Y. Para cada par de datos se localiza la intersección entre ambas variables y se marca con un punto.
- Es una manera rápida de hallar relaciones **lineales** entre las dos variables.

8. PROPIEDADES DE LA DISTRIBUCIÓN DE FRECUENCIAS

a) Tendencia central

Se refiere al lugar donde se centra una distribución particular en la escala de valores.

b) Variabilidad

Se refiere al grado de concentración de las observaciones en torno al promedio.

Una distribución de frecuencias será:

Homogénea (tiene poca variabilidad). Si los valores de distribución están cercanos al promedio. (cuando hay pocos rectángulos alrededor del central)

Heterogénea (tiene mucha variabilidad). Si los valores se dispersan mucho en torno al promedio. (cuando hay muchos rectángulos alrededor del central)

c) Asimetría o sesgo (con calma que es muy fácil pero cuesta un poco de entender al principio)

Se refiere al grado en que los datos se reparten equilibradamente por encima y por debajo de la tendencia central.



CAPITULO 1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

Diremos que una distribución será **simétrica** cuando al dividirla en dos a la altura de la media, las dos mitades se superponen. Si esto no ocurre decimos que se produce **asimetría**. Hay dos tipos de asimetría:

Asimetría positiva: Cuando la mayor concentración de puntuaciones se produce en la parte baja de la escala.

Asimetría negativa: Cuando la mayor parte de puntuaciones se sitúa en la parte alta de la escala.