

preunab 

TIPOS DE DATOS Y SU TABULACIÓN

Clase # 24

Universidad Andrés Bello

CONCEPTOS BÁSICOS DE ESTADÍSTICA

- **Estadística:** Ciencia que estudia los métodos y técnicas para obtener datos provenientes de fenómenos aleatorios, para organizarlos, presentarlos e interpretarlos en el contexto donde surgen, con el objeto de construir una comprensión del fenómeno que les dio origen.

Muchas veces el propósito de un estudio estadístico es construir conocimiento con un interés científico explicativo, mientras que otras veces es para tomar decisiones en el ámbito económico, político, comercial, etc.

CONCEPTOS BÁSICOS DE ESTADÍSTICA

Población: Conjunto de todos los sujetos cuyo estudio interesa. También se puede comprender una población como el conjunto de todas las medidas o datos a estudiar.

Ejemplos:

- 1 Usuarios de Internet en Chile.
- 2 Accidentes de tránsito en la región Metropolitana.
- 3 Microempresas del sector agroindustrial.
- 4 Sandías producidas en Paine.

Muestra: Subconjunto de la población que se extrae con el objeto de estudiar las características poblacionales.

Algunas ventajas de trabajar con muestras:

- Es más rápido, ya que se estudian menos sujetos..
- Es más barato, ya que al ser menos sujetos en estudio, se requieren menos recursos.
- Si la muestra es representativa, se obtienen resultados muy cercanos a la realidad poblacional.
- Al ser menos sujetos involucrados, se les puede estudiar detalladamente.

Algunas ventajas de trabajar con muestras:

- Es más rápido, ya que se estudian menos sujetos..
- Es más barato, ya que al ser menos sujetos en estudio, se requieren menos recursos.
- Si la muestra es representativa, se obtienen resultados muy cercanos a la realidad poblacional.
- Al ser menos sujetos involucrados, se les puede estudiar detalladamente.

Algunas Desventajas de trabajar con muestras:

- Todo trabajo con muestras está sujeto a incertidumbre (error). Es imposible escapar de este fenómeno.
- El trabajo con muestras requiere personal especializado.
- Si la muestra no está bien seleccionada, se puede llegar a resultados erróneos.
- El trabajo con muestras requiere técnicas estadísticas muy especializadas.

Tamaño de Muestra

Es el número de elementos de la población que se seleccionan para estudiar. Se representa por la letra n minúscula.

El tamaño de la muestra varía desde un individuo, hasta la población total. Sin embargo, un estudio se encarece al aumentar el tamaño de la muestra. Por esta razón, siempre ha de tratarse de trabajar con la menor muestra posible, con la consideración de que, a menor tamaño de muestra, mayor es el error o incertidumbre de los resultados. Existen métodos para el cálculo del tamaño óptimo de la muestra.

Variable

Característica de los sujetos en estudio, que adopta distintas medidas, valores o categorías entre los componentes de una muestra o población.

Ejemplos:

- Color de automóvil preferido por las mujeres.
- Número de hijos por familia.
- Sistema de previsión social de las personas.
- Grado de acuerdo de los ciudadanos con la reforma a la Constitución Política de Chile.

Datos cualitativos

Son aquellos que expresan cualidades de los sujetos en estudio.
Ejemplo: medio de transporte para llegar a su trabajo de los trabajadores del sector comercio. Los valores de la variable pueden ser: Metro, Transantiago, taxi, bicicleta, etc.

Datos cuantitativos

Son aquellos que expresan características numerables de los sujetos en estudio.

Ejemplo: tiempo, en minutos, que una muestra de personas emplea para desplazarse desde su casa al trabajo o lugar de estudio. Los valores de la variable pueden ser: 35 min, 18 min, 50 min, etc.

Variable Nominal

Corresponde a la simple clasificación de los individuos de una muestra o población en distintas categorías mutuamente excluyentes. Por esta razón es también llamada variable categórica.

Ejemplos:

- Situación civil de las personas: soltero, casado, viudo, otro.
- Comuna de residencia: Quilleco, Conchalí, Parinacota, etc.

Dentro de las variables nominales hay un caso especial llamado variable dicotómica, que resulta cuando la variable tiene dos valores mutuamente excluyentes. Es decir, uno de los valores es el contrario lógico del otro.

Ejemplos:

- Estado civil: casado, no casado.
- Disponibilidad de Internet en el hogar: Sí, No.

Variable Ordinal

Clasifica los individuos en distintas categorías que tienen un orden de precedencia o graduación desde un mínimo a un máximo.

Ejemplos:

- Grado de interés de los estudiantes en las ciencias: Alto, Medio, Bajo.
- Frecuencia de uso de Internet: Siempre, Generalmente, A veces, Nunca.

Variable Discreta

Es aquella que tiene valores representados por números enteros. En general, se obtienen por conteo, es decir contando.

Ejemplos:

- Número de hijos por familia: 0, 1, 2, 3, ... hijos.
- Número de integrantes del grupo familiar: 1, 2, 3, ... integrantes.

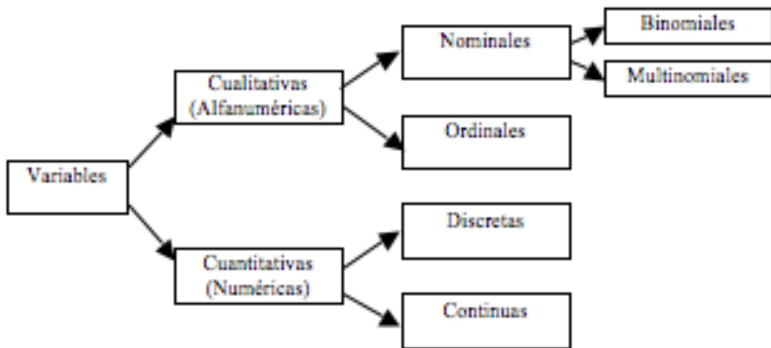
Variable Continua

Es aquella que tiene valores representados por números reales. Generalmente se obtienen por medición de la característica con algún instrumento.

Ejemplos:

- Peso al nacer de una muestra de recién nacidos: 4,320; 3,740; 2,860 Kg, etc.
- Interés mensual cobrado por casas comerciales: 2,6 %; 4,15 %; 3,45 %; etc.

Resumen Tipos de Variables



Tabulación de Datos Cualitativos

- Primera columna: aquí se ubican los valores de la variable. Si la variable es nominal, no requiere un orden determinado. Si es ordinal, se ordena de mayor a menor o viceversa.
- Segunda columna: se ubica la frecuencia absoluta. Esto es, el número de veces que aparece cada valor de la variable en los datos observados.
- Tercera columna: Porcentaje. Se calcula qué % representa cada frecuencia absoluta respecto de la frecuencia total o tamaño de la muestra n .

X	N° casos	%
Categoría A		
Categoría B		
Categoría C		
Categoría D		
Total	n	100

Ejemplo

Comuna de residencia de una muestra de trabajadores del sector comercio.

Comuna	Nº de casos	%
Santiago Centro	31	24,8
Independencia	17	13,6
San Miguel	26	20,8
Conchalí	14	11,2
Ñuñoa	29	23,2
Peñalolén	8	6,4
Total	125	100

Esta tabla muestra una variable de tipo nominal, por cuanto sus valores, que son alfanuméricos, no tienen un orden determinado.

Ejemplo

En Santiago Centro la frecuencia porcentuada es igual a:

$$\frac{31}{125} \cdot 100 = 24,8 \%$$

Interpretación:

- De acuerdo a la tabla podemos enunciar que en la muestra estudiada el 24,8 % de los trabajadores encuestados tiene residencia en la comuna de Santiago Centro, mientras que el 20,8 % de ellos residen en la comuna de San Miguel.
- La comuna con menos residentes, es Peñalolén, con solo un 6,4 % de la muestra.

Tabulación de Datos Discretos

Número de hijos por matrimonio:

Nº de hijos	Nº de casos	%
0	6	24,0
1	11	44,0
2	5	20,0
3	1	4,0
4	2	8,0
Total	25	100

Se trata este caso de una variable numérica discreta, por cuanto se ordenan en la primera columna los distintos valores de la variable (0, 1, 2, 3, 4 y 5) de menor a mayor, cada uno con su correspondiente frecuencia absoluta.

Para cero hijos la frecuencia porcentuada es:

$$\frac{6}{25} \cdot 100 = 24\%$$

Para más de 2 hijos la frecuencia porcentuada es:

$$\frac{1 + 2}{25} \cdot 100 = 8\%$$

Interpretación:

- De acuerdo a la tabla, el 24% de las familias encuestadas no tienen hijos, mientras que el 76% tiene entre 1 y 4 hijos.
- Es destacable el hecho de que solo el 12% de los encuestados tienen más de 2 hijos.

Tabulación de Datos Continuos

Cuando los datos están medidos en una escala numérica continua, la construcción de tablas para presentar los datos se hace mediante la partición del recorrido de los valores de la variable en intervalos. La tabla se construye de modo que sus intervalos son abiertos por la derecha, de modo que el límite inferior pertenece al intervalo, pero el superior no. En la tabla adjunta, el primer intervalo corresponde a $[150, 155[$, intervalo que incluye el 150 y excluye el 155, que pasa a quedar en el segundo intervalo. Estatura de 40 personas, en centímetros.

Estatura (cm)	Nº de casos	%
150 - 155	3	7,5
155 - 160	11	27,5
160 - 165	13	32,5
165 - 170	9	22,5
170 - 175	4	10,0
Total	40	100

Para menos de 155*cm.* de estatura la frecuencia porcentuada es:

$$\frac{3}{40} \cdot 100 = 7,5\%$$

Interpretación:

Para la interpretación de este tipo de tablas se debe tener en cuenta que la variable es numérica continua, por lo que no es posible establecer la frecuencia de un valor puntual, como por ejemplo 154*cm.*, sino que siempre se debe interpretar en términos de intervalos.

- Según la tabla, el 7,5% de la muestra mide menos de 155 cm, el 60% mide entre 155 y menos de 165 cm, mientras que el 32,5% mide al menos 165 cm.
- Etc.

Marca de Clase

En una estadística de variable agrupada en intervalos, muchas veces se requiere un valor que represente a todos los valores posibles de cada intervalo. Este valor se denomina Marca de clase y se simboliza como X_m .

Cada intervalo tiene como marca de clase al punto medio de cada intervalo. Esto es:

$$X_m = \frac{Lim_{inf} + Lim_{sup}}{2}$$

Estatura (cm)	Xm	Nº de casos	%
150 - 155	152,5	3	7,5
155 - 160	157,5	11	27,5
160 - 165	162,5	13	32,5
165 - 170	167,5	9	22,5
170 - 175	172,5	4	10,0
Total	-	40	100

Tabulación Bivariada

Las tablas bivariadas agrupan los datos provenientes de la medición conjunta de dos variables en un mismo sujeto de una muestra. Por ejemplo, a una misma persona se le puede medir peso y estatura, edad y estado civil, etc.

Ejemplo: La tabla siguiente muestra la distribución de una muestra de estudiantes encuestados según si están o no interesados en seguir estudios superiores. Se puede observar que ambas variables, Interés y Género son de tipo nominal dicotómicas.

Sexo	Interés por seguir estudios superiores		Total
	Si	No	
Mujer	587	47	634
Hombre	402	66	468
Total	989	113	1.102

Según la tabla, la muestra consideró un total de 634 mujeres, las que representan el:

$$\frac{634}{1102} \cdot 100 = 57,5 \%$$

De acuerdo a la tabla, 989 encuestados tienen interés por seguir estudios superiores, los que representan de la muestra el:

$$\frac{989}{1102} \cdot 100 = 89,7 \%$$

Interpretación:

- Las mujeres encuestadas representan el 57,5% de la muestra estudiada.
- El 89,7% de los encuestados tienen interés en seguir estudios superiores.

