

Clase 14-03-18 INTRODUCCIÓN ESTADÍSTICA BÁSICA

Desarrollo curso:

- Levantamiento
- Análisis datos
- Representación de datos

En la clase de hoy trabajaremos con las manzanas del precenso del 2011.

Concepto (lo que quiero medir) → Dimensión → Indicadores → variables

Ejemplo:

Concepto: Compasión

Dimensiones: Compasión con humanos/ Compasión con animales

Indicadores: Humanos → Visita hospitales en Navidad/ Abraza a alguien que llora

Animales → Regresa pajaritos a los nidos/ Salva ballenas

I. Definición del problema a trabajar.

Cuando me enfrento a un problema de investigación, lo primero es definir un problema, problematizar una temática x: ¿Qué quiero relevar acerca de alguna temática?

La información con la que trabajan los planificadores urbanos está sujeta a un contexto espacial, ya sea una unidad política administrativa, una manzana, etc. tenemos que cuestionar la **escala** y ver si el trabajo se va a desarrollar un **levantamiento a partir de fuentes primarias o secundarias**.

Ejemplo: Necesito saber cuál es el nivel de hacinamiento de la población en el área metropolitana de Santiago.

Concepto → Hacinamiento

Indicador → Cantidad de personas que viven en la vivienda/ Tamaño de la vivienda (cantidad de recintos habitables)

Fuente primaria: Toda información que levanto específicamente para mi estudio (ej. Encuesta)

Ejemplo: Si quiero saber cuántos lápices hay en cada casa, esa información no la encuentro en el censo, así que tendré que levantar yo mismo esa información.

Fuente secundaria: Fuentes a las que se puede acceder a través de instrumentos que ya existen y que puedo adecuar a mi investigación (ej. Censo)

Promedio: Medida que me permite saber cuál es el punto central de una frecuencia a partir de la distribución de los datos estadísticos. Ese promedio está modulado por el número de unidades de la muestra. ¿Por qué es tan relevante? Porque es un punto de referencia, una caracterización de un espacio determinado o de una muestra determinada.

Falacia ecológica: tipo de falacia o error en la argumentación basado en la mala interpretación de datos estadísticos, en el que se infiere la naturaleza de los individuos a partir de las estadísticas agregadas del grupo al que dichos individuos pertenecen. Esta falacia se da a partir del supuesto de que todos los miembros de un grupo muestran las mismas características del grupo.

Ejemplo: Peñalolén tiene familias con ingresos muy altos y otras con ingresos muy bajos, entonces cuando promedio los sueldos me da una cifra media con alta dispersión que no representa a nadie.

Desviación estándar: Dentro de una muestra qué tanto varían los datos desde el promedio.

Ejemplo: Promedio sueldo comuna \$250.000. Su desviación estándar es de \$10.000. Esto significa que es un territorio bastante homogéneo, ya que la mayor parte de la muestra se ubica entre los valores \$240.000 y \$260.000 (valores no tan marginales).

II. Explorar base de datos

Diccionario de datos/metadatos: significado de las abreviaciones

Variable cuantitativa: Todas aquellas expresadas con números

Variable cualitativa: Todas aquellas expresadas en cualidades por percepción

Nivel de datos de la planilla: manzana. Cada una de las filas es una manzana

Código INE: Código compuesto territorialmente.

región (13) / provincia (4) / comuna (02) / distrito censal (02) / urbano (1)-rural (2) / zona censal (001) / manzana (018) → **13402021001018**

Si tomo la manzana y la corto en el quinto código podré saber a qué comuna pertenece.

III. Breve análisis estadístico

Determinar el número de viviendas efectivas

Vivienda efectiva: Viviendas que efectivamente serán censadas al momento de contabilizar a la población → presentes + ausentes

Instrucciones actividad.

1. Abrir planilla Excel
2. Ir al final de la planilla, la columna AO está vacía. Escribir el nombre del campo "VIVEFEC" (=G2+H2)
3. Cuando se pincha la casilla aparece un cuadradito verde en su esquina inferior derecha. Hacer doble click. Inmediatamente se hace el cálculo para toda la columna.
4. Ir a la celda AN 59.066. Inicio → Autosuma → Suma [=SUMA(AN2:AN59065)] → Enter.
Si arrastro el cuadrado verde de la esquina inferior derecha hacia la derecha, replica la operación a la columna de al lado y así obtengo la sumatoria de las viviendas efectivas
5. ¿Cómo calcular el número de personas por vivienda?
Población/ Cantidad de viviendas (=AN59066/AO59066)
6. Promedio y desviación estándar
Promedio → Ir al final de la columna de la población
Inicio → Autosuma → Promedio [=PROMEDIO(AN2:AN59065)]
Replicar hacia el lado para obtener promedio de vivienda por manzana
Desviación estándar → Mientras más grande sea la muestra, más representativo será el promedio; el problema es que nosotros representamos el espacio, y en él la aleatoriedad no existe porque existe el proceso de autocorrelación espacial (las cosas que están cerca tienden a parecerse)
Excel no trae la opción de desviación estándar, pero se calcula de la siguiente manera:
= DESVEST.P(AN2:AN59065) y replicamos para el lado.
7. Conclusión: las desviaciones son casi el doble que los promedios, entonces hay una alta heterogeneidad en los datos.

IV. Módulo 2: generalización (bajar de escala: región → comuna)

1. Elegir dos comunas contiguas
2. Excel funciona con el código de programación visualbasic
Función left. [=izquierda → función → (celda; número caracteres que estoy robando)
3. Crear campo "CUT" [=izquierda (b2;5)]

Doble click cuadrado verde.

4. ¿Cómo saber el código de las comunas?
Poner en internet "Código único territorial" (CUT) Subdere
(Maipú: 13119 Cerrillos: 13102)
5. Seleccionar AP59065
Apretar ctrl + shift (al mismo tiempo) → Arrastrar a la izquierda y después hacia arriba y seleccionar toda la base de datos.
6. Insertar → Tabla dinámica → Nueva hoja de cálculo
Forma rápida y eficiente de generalizar cualquier cálculo
Sirve para extraer todas las estadísticas que se me ocurran
7. Buscar el campo CUT. Arrastrar a Filas. Así el sistema se da cuenta de que las comunas se repiten en los registros, entonces generaliza todo el resto en un solo registro.
8. Tomar TOTPOB y arrastrarlo a Valores. Población comunal para cada comuna de la Región Metropolitana. El sistema tomó todos los datos de las manzanas en las filas y los juntó por comuna.

Tabla dinámica: instrumento que permite sacar estadísticas de manera súper rápida.

9. Gráfico → Seleccionar las etiquetas (ctrl) que quiero graficar → Insertar

Encargo para la próxima clase:

Escoger dos comunas.

Calcular tres variables cuantitativas (ojalá no población ni vivienda) con las cuales se pueda definir algún relato (definir un objetivo de lo que se está mostrando. Ejemplo: Mostrar qué comuna tiene más recintos comerciales/ diferencias- semejanzas entre las comunas)

Sacar la suma, el promedio y la desviación estándar y crear un gráfico y un breve relato de la relación que se forme entre las dos comunas (1/2 párrafos)