

Índice

1. Introducción	1
2. Ejercicio N°1 - Estadística Descriptiva	1
2.1. Medidas de tendencia central	2
Bibliografía	3

TRABAJO FINAL DE ESTADÍSTICA



MAESTRÍA EN CIENCIA DE DATOS

Alumno:

Alejandro Chaín

Docentes:

Rodrigo Del Rosso

Ezequiel Nuske

4 de Junio del 2022

1. Introducción

El presente trabajo práctico se encuadra dentro del programa de la materia de Estadística de la maestría en Ciencia de Datos de la Facultad de Ingeniería de la Universidad Austral. El objetivo de este trabajo es la aplicación de los conceptos teóricos y prácticos desarrollados durante la cursada de la materia.

2. Ejercicio N°1 - Estadística Descriptiva

Para el abordaje de este punto se utiliza como base de datos a la Encuesta Permanente de Hogares a nivel de individuos correspondiente al cuarto trimestre de 2021, esta encuesta es relevada por el instituto de estadísticas y Censo (INDEC), tiene una frecuencia trimestral y cubre 31 aglomerados urbanos y un área urbano-rural.

Para la descarga de esta base de datos se emplea la biblioteca de EPH (Kozlowski et al., 2020). La población que se tomará en cuenta serán todos los individuos encuestados que se encuentren con una condición de actividad de “Ocupados” y que hayan declarado haber recibido algún tipo de ingreso laboral. En total, esta población consta de 17,243 individuos de los cuales se tomará una muestra de 1.000 individuos de manera aleatoria, sin reposición y asignándole a cada individuo de la población la misma probabilidad de ser seleccionado (distribución uniforme de probabilidad).

Las variables que se tienen en cuenta para este análisis son tanto del tipo cuantitativo como cualitativo, a continuación se realiza una breve descripción de las mismas.

Variables Cualitativas:

- Sexo: Esta variable indica si el individuo es Hombre o Mujer.
- Calificación: Esta variable indica el nivel de calificación del individuo, esta variable puede ser “No calificados,” “Operativos,” “Técnicos” o “Profesionales.”
- Nivel educativo: Esta variable indica el nivel educativo alcanzado por el individuo, este puede ser “Sin instrucción,” “Primario Completo,” “Secundario Completo” o “Superior Universitario Completo,”

Variables Cuantitativas:

- Edad: Esta variable numérica discreta indica la edad del individuo al momento del relevamiento.
- Horas de trabajo semanal: Esta variable numérica indica la cantidad de horas semanales que trabaja el individuo en su ocupación laboral.

- Ingreso total: Esta variable numérica discreta indica el ingreso total mensual en pesos que recibió la persona, este está compuesto por el ingreso laboral proveniente del desarrollo de su ocupación sumado a los ingresos no laborales percibidos.

Para realizar un análisis estadístico descriptivo desde un enfoque general de la muestra se pueden analizar tanto medidas de tendencia central, de variabilidad, de sesgo y de curtosis (Chao & Castaño, 1993) de las variables cuantitativas de los individuos muestreados.

2.1. Medidas de tendencia central

Una medida de tendencia central es aquel número que se toma como orientación para referirnos a un conjunto de datos. Dentro de las medidas de tendencia central, también conocidas como medidas de posición, se pueden encontrar:

- La Media aritmética (1), esta medida de posición es la que cuenta con mayor popularidad, esta medida calcula el centro físico del conjunto de datos y esta definida como la suma de los valores observados de una variable dividido por el total de observaciones.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

- La Mediana (2), esta medida de posición indica el valor que divide un conjunto de observaciones ordenadas respecto de la magnitud de los valores, de tal manera que el número de datos por encima de la mediana sea igual al número de datos por debajo de la misma.

$$X_m = \begin{cases} X_{(\frac{n+1}{2})} & \text{si } n \text{ es un numero impar} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2})+1}}{2} & \text{si } n \text{ es un numero par} \end{cases} \quad (2)$$

- La Moda es una medida de posición que indica el valor que se da con mayor frecuencia en una sucesión de datos. En un conjunto de datos puede haber una moda, más de un moda (multimodal) o puede no haber ninguna moda.

Mediante el cálculo de estas medidas de posición para las variables cualitativas de la muestra se puede obtener una descripción sencilla y simplificada de los datos. En el cuadro se puede ver un resumen de estas medidas de posición para las variables de *Edad*, *Años de Educación*, *Horas de Trabajo Semanal* e *Ingreso Total de los individuos*.

Cuadro 1: Medidas de posición

	Edad	Horas de trabajo semanal	Ingreso total
Media	41.143	70.623	55622.65
Mediana	40.000	80.000	46000.00
Moda	35.000	80.000	60000.00

Bibliografía

- Chao, L. L., & Castaño, J. M. (1993). *Estadísticas para las ciencias administrativas* (3a. ed). McGraw-Hill.
- Kozlowski, D., Tiscornia, P., Weksler, G., Rosati, G., & Shokida, N. (2020). *Eph: Argentina's permanent household survey data and manipulation utilities*. <https://doi.org/10.5281/zenodo.3462677>