

NOMBRE DE ESTUDIANTE:

DARIANA ALEJANDRA VILLANUEVA CONSILCO

DOCENTE:

DR. CULEBRO RICALDI JOSE MIGUEL

**MATERIA;
DISEÑO EXPERIMENTAL**

**TEMA: ESTADÍSTICA DESCRIPTIVA Y LA
INFERENCIAL**

CRRERA: MEDICINA HUMANA

SEMESTRE: 4°

FECHA: 07 DE JUNIO DEL 2020

INTRODUCCIÓN

La definición de estadística ha adoptado muchas formas a lo largo de la historia. Hoy en día, se puede establecer como la ciencia que permite extraer información de los datos, así como la medición, el control, y la comunicación de la incertidumbre. Constituye, por tanto, la herramienta esencial para controlar el curso de los avances científicos y sociales.

La ciencia estadística ha ido aumentando el rango de técnicas, métodos y teorías que aglutinar, pero posiblemente la división más primaria que se puede hacer de la misma es la que distingue los campos de estadística descriptiva y estadística inferencial.

La estadística descriptiva es el conjunto de métodos estadísticos que describen y/o caracterizan un grupo de datos. La estadística inferencial busca deducir y sacar conclusiones acerca de situaciones generales mas allá del conjunto de datos obtenidos.

TIPOS DE DATOS

Se llama variable estadística o carácter a cada una de las características que pueden estudiarse de la población. Las variables estadísticas pueden ser de dos tipos:

- Cualitativas: son aquellas en la que los resultados posibles no son valores numéricos. Por ejemplo: color del pelo, tipo de ropa preferida, lugar de veraneo, etc.
- Cuantitativas: aquellas cuyo resultado es un número. A su vez, las hay de dos tipos:
- Cuantitativas discretas: cuando se toman valores aislados. Por ejemplo: número de amigos de tu pandilla, número de veces que vas al cine al mes, número de coches que tiene tu familia.
- Cuantitativas continuas: cuando, entre dos valores cualesquiera, puede haber valores intermedios. Es decir, se toman todos los valores de un determinado intervalo. Por ejemplo: peso de las personas, nivel sobre el mar en que se encuentra tu ciudad, medida del perímetro torácico.

ESTADISTICA DESCRIPTIVA

La estadística descriptiva tiene como objetivo resumir la información contenida en los datos de la forma más sencilla y presentable posible, obteniendo así los parámetros que distinguen las características de un conjunto de datos (lo que se conoce como estadísticos). Pertenecen al ámbito de la estadística descriptiva las tablas de frecuencias, a partir de las cuales se obtienen los estadísticos:

- Medidas de centralización: la media en todas sus variantes (aritmética, geométrica, ponderada), la moda y la mediana
- Medidas de dispersión: la varianza, la desviación típica (raíz cuadrada de la varianza) y el rango
- Medidas de tendencia central: los cuantiles y sus desgloses (percentiles, cuartiles, deciles, etc.)
- Medidas de forma: los coeficientes de asimetría y curtosis
- Medidas de concentración: el coeficiente Gini, a partir del cual se obtiene la curva de Lorenz

INFERENCIA ESTADÍSTICA

La inferencia estadística es el conjunto de métodos que permiten inducir, a través de una muestra estadística, el comportamiento de una determinada población. La inferencia estadística, estudia entonces como, a través de la aplicación de dichos métodos sobre los datos de una muestra, se pueden extraer conclusiones sobre los parámetros de la población de datos. De la misma manera estudia también el grado de fiabilidad de los resultados extraídos del estudio.

“Para entender el concepto es importante entender tres conceptos:

- Inferencia: Inferir significa, literalmente, extraer juicios o conclusiones a partir de ciertos supuestos, sean estos generales o particulares.
- Población: Una población de datos, es el conjunto total de datos que existen sobre un variable.
- Muestra estadística: Una muestra es una parte de la población de datos.
- Teniendo claro a lo que nos referimos con el concepto de inferir, una de las dudas fundamentales recae en el hecho de elegir una muestra en lugar de una población.” (LOPEZ, 2019)

Normalmente, en estadística, se trabaja con muestras debido a la gran cantidad de datos que tiene una población. Por ejemplo, si queremos sacar conclusiones, esto es, inferir, los resultados de las elecciones generales, es imposible preguntar a toda la población del país. Para solventar ese problema se escoge una muestra variada y representativa. Gracias a la cual se puedan extraer una estimación del resultado final. Escoger una muestra adecuada corre a cargo de las distintas técnicas de muestreo.

La otra gran rama de la estadística es la estadística descriptiva.

Métodos de la inferencia estadística: Los métodos y técnicas de la inferencia estadística se pueden dividir en dos: métodos de estimación de parámetros y métodos de contraste de hipótesis.

- Métodos de estimación de parámetros: Se encarga de asignar un valor al parámetro o al conjunto de parámetros que caracterizan el campo sujeto a estudio. Claro que al ser una estimación existe cierto error. Para obtener estimaciones adaptadas a esa realidad, se crean intervalos de confianza.

- Métodos de contraste de hipótesis: Su objetivo es comprobar si una estimación corresponde con los valores poblacionales. En todo contraste de hipótesis existen dos supuestos. La hipótesis nula (H_0) que recoge la idea de que un valor tiene un valor predeterminado. Si se rechaza la hipótesis nula (H_0), entonces se acepta la hipótesis alternativa (H_1).

PRUEBAS ESTADÍSTICAS

Las pruebas estadísticas forman parte de la teoría de decisión, a partir de la información que extraemos de la muestra estimamos características generales de la población de referencia. Existen tres tipos de pruebas estadísticas:

- 1) Pruebas de conformidad: en las que se comprueba si una estimación coincide con un valor teórico. Por ejemplo, queremos comprobar si la proporción de recurrencia de una úlcera duodenal al tomar cierto fármaco es inferior al 10%.
- 2) Pruebas de homogeneidad: comparan poblacionalmente dos o más grupos; supongamos que nos interesa comprobar si la proporción de recurrencia de la úlcera duodenal con un nuevo fármaco es igual a la proporción de recurrencia en pacientes tratados con otro fármaco.
- 3) Pruebas de relación: evalúan la relación entre variables.

Los contrastes de hipótesis o tests de hipótesis permiten comprobar si la información muestral concuerda con la hipótesis estadística formulada, nos permiten cuantificar hasta qué punto los resultados de un estudio particular dependen de la variabilidad de la muestra.

La hipótesis que se contrasta se denomina hipótesis nula y se denota por H_0 , se puede interpretar como la hipótesis que normalmente sería aceptada mientras los datos no indiquen lo contrario. Rechazar la hipótesis nula supone asumir una hipótesis complementaria, la hipótesis alternativa (H_1), como correcta. Para realizar un contraste de hipótesis debemos definir la hipótesis nula y la alternativa y definir una medida, el estadístico de contraste, que permite cuantificar la magnitud de la diferencia entre la información que proporciona la muestra y la hipótesis H_0 . Se pueden cometer dos tipos de errores:

- 1) Error tipo I: rechazamos la hipótesis nula cuando es cierta.
- 2) Error tipo II: no rechazamos la hipótesis nula cuando es falsa.

En la práctica no es posible saber si estamos cometiendo un error tipo I o un error tipo II, pero existen ciertas recomendaciones para disminuir dichos errores. Por ejemplo, para disminuir el error tipo I deberíamos depurar la base de datos para evitar posibles outliers o valores extremos que puedan producir resultados significativos, utilizar un nivel de significación pequeño y disponer de una teoría que guíe las pruebas. Para reducir el error tipo II es aconsejable incrementar el tamaño muestral, estimar la potencia estadística o incrementar el tamaño del efecto a detectar.

Es necesario establecer a priori el nivel de significación (α) que se define como la probabilidad de cometer un error tipo I, normalmente se elige un valor pequeño, el 5% o el 1%. El valor del nivel de significación divide en dos regiones el conjunto de posibles valores del estadístico de contraste:

- 1) Zona de rechazo (con probabilidad α , bajo H_0).
- 2) Zona de aceptación (con probabilidad $1-\alpha$, bajo H_0).

Cuando analizamos la muestra obtendremos la significación del contraste, que se representa con la letra p , es un indicador de la discrepancia entre la hipótesis nula y los datos muestrales, de forma que cuanto más se acerque a cero tenemos mayor evidencia en contra de la hipótesis nula (si p es menor que el nivel de significación rechazaremos H_0). Debemos tener en cuenta que la significación estadística depende de la magnitud de la diferencia que queremos probar, cuanto mayor sea esta diferencia más sencillo será demostrar que es significativa. Al mismo tiempo depende también del tamaño muestral, cuanto más grande sea el número de observaciones más sencillo es detectar diferencias.

MODELOS DE REGRESIÓN

Un modelo de regresión es un modelo matemático que busca determinar la relación entre una variable dependiente (Y) con respecto a otras variables llamadas explicativas o independientes (X). El modelo de regresión se suele utilizar en las Ciencias Sociales con el fin de determinar si existe o no relación causal entre una variable dependiente (Y) y un conjunto de otras variables explicativas (X). Asimismo, el modelo busca determinar cuál será el impacto sobre la variable Y ante un cambio en las variables explicativas (X).

Así por ejemplo, un economista podría estar interesado en determinar la relación entre el ingreso de los trabajadores y su nivel de educación. Para esto, podría llevar a cabo un modelo de regresión en el cual la variable independiente (Y) será el ingreso del trabajador. En cuanto a las variables explicativas (X), se deben incluir todas aquellas que podrían explicar el ingreso entre las que se encuentran por supuesto la educación, la experiencia, la educación de los padres, etc. Forma del modelo de regresión.

El modelo de regresión simple tiene la siguiente forma:

$$Y = A + BX + u$$

Y= variable dependiente o endógena

X= variable independiente o explicativa

A, B = parámetros fijos y desconocidos

u= término de error que recoge todos los demás factores que afectan Y pero que no están incluidos en el modelo. También puede captar los errores de estimación de la variable dependiente. No observable. Luego, el objetivo del modelo de regresión será estimar los valores de A y B a partir de una muestra.

Significado de las variables

El parámetro B debería reflejar cuál es el impacto de un cambio de X sobre la variable Y, cuando el resto de las variables explicativas se mantienen constantes (*ceteris paribus*). El parámetro A en tanto, no afecta en nada la relación entre Y y X, sólo se trata de una normalización en donde se supone que el valor promedio de u será cero.