



Nombre de alumnos:

Diego Armando Gómez Domínguez

Nombre del profesor:

Rosario Gomez

Nombre del trabajo:

Ensayo

Materia:

Estadística Descriptiva

Grado:

2 Cuatrimestral

Grupo:

“U”

Ixtacomitan, Chiapas a 10 de abril de 2021.

Introducción

En relación a las experiencias de todo tipo de variables muchos de estos casos son muy Aceptables, ya que en base de las variables son de tal función, ayudan a realizar y a responder cada una de ellas.

Que si los conceptos de la asociación de las variables pueden ser de gran ayuda explicando en este ensayo, ya que hay que plantear y saber cual es la relación entre partes del programa como varianza, correlación y etc. Por eso doy un pequeño paso en darle la observación de cada investigación que se hizo para dar un buen desarrollo a los alumnos o profesores.

Estadístico de asociación entre variable

La asociación de este tema es muy buena a darla a conocer ya que en un concepto de variables que permita la existencia de algún tipo de asociación entre dos o más que variables representa de algún tipo de tendencia o patrón de emparejamiento entre los distintos valores de esas variables. El análisis estadístico de la asociación (relación, covarianza, correlación).

Entre variables representa una parte básica del análisis de datos en cuanto que muchas de las preguntas e hipótesis que se plantean en los estudios que se lleva a cabo en la práctica implica analizar la existencia de relación entre variables. En el 2005 ofrecen otra propuesta general de lo que significa la asociación entre dos variables difiere en función de los valores de la otra.

Un ejemplo de la presencia de asociación entre 2 variables, una, la puntuación en un test de aptitud lingüística y la otra, la variable sexo. Con conjuntos de datos de estas dos variables, la diferencia existe entre las distribuciones de frecuencias de la variable “aptitud lingüística” condicionada al variable sexo.

La asociación entre variables no debe entenderse como una cuestión de todo o nada, sino como un continuo que iría desde la ausencia de relación, al máximo de relación entre las variables. Este grado se plasmaría en una relación determinista, el caso en que a partir del valor de un sujeto cualquiera en una variable, se puede afirmar su valor en la otra variable.

Midiendo la asociación entre 2 variables.

En este momento el caso de dos variables categóricas, que sería el estado de ánimo y vivir en residencia, para evaluar ambas variables están relacionadas y hay que observar si la distribución de los valores de una de las variables, en todo caso hay que comparar las distribuciones condicionales de una de las dos variables agrupadas en función de los valores de la otras. Si no hay relación entre variables estas distribuciones deberían ser iguales. El análisis gráfico de la asociación entre 2 variables categóricas puede instituir a partir de un diagrama de barras conjunto de frecuencias absolutas si nos fijamos en la forma de las distribuciones condicionadas, pues si se elimina el efecto del posible distinto tamaño de los subgrupos.

Existen diferentes índices estadísticos orientados a resumir de forma cuantitativa la asociación entre dos variables categóricas, aquí nos vamos a centrar en los dos siguientes:

El índice: X^2 toma el valor 0 cuando dos variables son independientes, siendo mayor que 0 cuando exista asociación entre ellas, tanto mayor cuanto más intensa sea esa correlación. No tiene un límite máximo, lo cual supone una dificultad a nivel interpretativo. Diferente intensidad de asociación en 3 partes de variables basado en la visualización de las distribuciones condicionales mediante diagrama de caja y bigotes. El coeficiente de correlación puede utilizarse para medir el grado de

relación de dos variables siempre y cuando ambas sean cuantitativas, Muestra, población. El coeficiente de regresión, indica el número de unidades en que se modifica la variable dependiente "Y" por efecto del cambio de la variable independiente "X" o viceversa en una unidad de medida.

El análisis estadístico de la asociación, regresión lineal la primera forma de regresión fue el método de los mínimos cuadrados que fue publicada por Legendre en 1805, Gauss publicó un trabajo de donde desarrollaba de manera más profunda el método de los mínimos cuadrados.

El método de regresión se utilizó por primera vez en el estudio de variables antropométricas; al comparar la estatura de padres e hijos, donde resultó que los hijos cuyos padres tenían una estatura muy superior al valor medio, tendían a igualarse a éste, mientras que aquellos cuyos padres eran muy bajos tendían a reducir su diferencia respecto a la estatura media; es decir, "regresaban" al promedio. La constatación empírica de esta propiedad se vio reforzada más tarde con la justificación teórica de ese fenómeno.

El término lineal se emplea para distinguirlo del resto de técnicas de regresión, que emplean modelos basados en cualquier clase de función matemática. Los modelos lineales son una explicación simplificada de la realidad, mucho más ágiles y con un soporte teórico mucho más extenso por parte de la matemática y la estadística. Pero bien, como se ha dicho, se puede usar el término lineal para distinguir modelos basados en cualquier clase de aplicación.

Donde es la perturbación aleatoria que recoge todos aquellos factores de la realidad no controlables u observables y que por tanto se asocian con el azar, y es la que confiere al modelo su carácter estocástico. En el caso más sencillo, con una sola variable explícita, el hiperplano es una recta: El problema de la regresión consiste en elegir unos valores determinados para los parámetros desconocidos, de modo que la ecuación quede completamente especificada. Para ello se necesita un conjunto de observaciones. En una observación i -ésima ($i = 1, \dots, I$) cualquiera, se registra el comportamiento simultáneo de la variable dependiente y las variables explícitas (las perturbaciones aleatorias se suponen no observables).

El método más común de determinar si existe asociación lineal entre dos variables cuantitativas continuas es el Análisis de Correlación de Pearson. Con este método se obtiene el Coeficiente de Correlación de Pearson, usualmente representado por la letra R . Como suele utilizarse una muestra, lo que se obtiene en realidad es un estimado del coeficiente de correlación poblacional.

Dos aspectos importantes del coeficiente de correlación son su magnitud y su signo. La magnitud refleja la intensidad de la asociación entre las dos variables; el valor absoluto de la magnitud puede variar entre cero y uno. Valores cercanos a

cero indican que las variables no están asociadas, es decir, que el valor de una variable es independiente del valor de la otra.

El signo, por su parte, refleja cómo están asociados los valores de ambas variables. Si el signo es positivo indica que a valores altos de una variable corresponden valores altos de la otra, o a valores bajos de una variable corresponden valores bajos de la otra. Si el signo es negativo, indica que a valores altos de una variable corresponden valores bajos de la otra. Es decir, el signo positivo indica que los valores de ambas variables cambian en el mismo sentido, mientras que el signo negativo indica que cambian en sentido contrario. En la fórmula se observa que las unidades de ambas variables aparecen en el numerador y denominador, por lo tanto, se anulan aritméticamente, por lo que el coeficiente de correlación no tiene unidades de medición.

Coeficiente de contingencia χ^2 Estudia la diferencia entre las frecuencias observadas, n_{ij} , y las frecuencias esperadas, e_{ij} , $\chi^2 = \sum_{k=1}^K \sum_{j=1}^p (e_{ij} - n_{ij})^2 / e_{ij}$ donde $e_{ij} = n_{i.} \cdot n_{.j} / n$. Cuando hay independencia estadística, o no hay asociación entre los atributos, se tiene que $n_{ij} = e_{ij}$, resultando $\chi^2 = 0$. El coeficiente puede reescribirse de forma que facilite su cálculo, $\chi^2 = n \sum_{j=1}^p \sum_{i=1}^2 (n_{ij} / n_{i.} - n_{.j} / n)^2 / n_{.j}$, donde, si el coeficiente vale cero es que hay independencia entre las variables, y cuanto más grande salga, mayor asociación hay entre las variables. El mayor inconveniente que tiene este coeficiente es que es proporcional al número de observaciones, y por tanto no tiene una cota, por lo que no es muy adecuado su uso.

Los valores que puede tomar d no están acotados a un rango, pudiendo ser tanto positivos como negativos. Si las dos variables consideradas son independientes entonces d será igual a 0, mientras que cuanto mayor sea la asociación entre ellas, mayor será el valor de d en términos absolutos. Cohen sugiere las siguientes normas interpretativas, aunque el propio autor afirma que se deben utilizar sólo en el caso que no se tenga ningún criterio sustantivo que sirva de base

Conclusión

En el tema ocho se estudiaron el Coeficiente de correlación lineal, que es un coeficiente que nos indica el grado de asociación entre dos variables cuantitativas. En este tema vamos a estudiar coeficientes para determinar el grado de asociación de variables cualitativas, es decir, aquellas variables cuyas modalidades no se pueden cuantificar. Recordemos que a las variables cualitativas también se las denomina atributos, por ello a veces se denomina a este estudio, estudio de asociación entre atributos. Se denomina tabla de contingencia a la tabla de doble entrada que se representa para una distribución bidimensional de atributos. En realidad, es exactamente una tabla de correlación donde en la primera fila y columna, en lugar de aparecer las cuantificaciones numéricas de las variables, se escriben las definiciones de cada una de las modalidades asociadas a los atributos. Debe notarse que el concepto de independencia que se estudió para variables cuantitativas, es el mismo para atributos, y que por tanto el concepto de asociación también coincide solo que los coeficientes que se utilizan para medir la asociación son distintos.