

PROPUESTA TALLER APLICACIONES DE LA REGRESIÓN LINEAL SIMPLE Y MULTIPLE

LEONEL DELGADO ERASO

lederudenar@hotmail.com

Universidad de Nariño.

1 PRESENTACION

El análisis de regresión es una de las técnicas de uso más frecuente para analizar datos multivariantes. Su atractivo y utilidad general son el resultado del proceso conceptualmente lógico de usar una ecuación para expresar una relación entre una variable de interés (respuesta o variable dependiente) y un conjunto de variables predictoras relacionadas (variables independientes). El análisis de regresión tiene también interés teórico, por las elegantes matemáticas que se usan y por una teoría estadística bien desarrollada.

La propuesta en realizar aplicaciones de la regresión lineal simple y múltiple tiene como objetivo motivar a los docentes de educación media en el uso de esta técnica en situaciones de su entorno. Se podría pensar en como relacionar el rendimiento académico de los estudiantes con el número de horas dedicada al estudio en jornada contraria, nivel académico de los padres, salario familiar y otras variables de interés. También se pretende incentivar a los docentes en el campo de la investigación

En campo del conocimiento existen varios niveles de investigación, en la educación media y con la teoría de la regresión, introduciremos a los docentes a la búsqueda de herramientas necesarias para realizar investigación dentro del aula.

En la actualidad existen varios paquetes estadísticos que cumplen con la función de ejecutar y analizar datos, sin descuidar la comprensión de los conceptos básicos. El paquete estadístico PSPP (software libre), permite ingresar los datos y obtener los estadísticos necesarios para evaluar la regresión haciendo fácil su interpretación a usuarios novicios en la estadística.

La propuesta contempla tres jornadas:

- Fundamentación teórica: Conceptos básicos, Supuestos de la regresión
- Ejemplos de aplicación desarrollados en calculadora, Excel, P SPP
- Taller de aplicación en PSPP.

Lógicamente, lejos de solucionar todas las inquietudes que puedan surgir respecto a la materia, se trata de despertar interés para posibles cursos más especializados sobre el tema de la regresión en general

MARCO TEORICO

REGRESIÓN

La regresión es una técnica estadística que surge de la necesidad de evaluar el grado de relación entre un grupo de variables denominadas independientes o explicativas (X_1, X_2, \dots, X_k) y una variable de interés denominada dependiente o respuesta (Y)

El objetivo principal se puede resumir en construir un modelo de regresión *con fines explicativos* basado en evaluar cómo afecta el cambio en unas características determinadas en las variables explicativas sobre la variable respuesta; o elaborar un *modelo con fines predictivos* para intentar estimar o aproximar el valor de una característica variable respuesta en función de los valores que pueden tomar en conjunto otra serie de características de las variables explicativas.

Existen varias opciones para estimar un modelo de regresión, entre los que destacan por su facilidad de aplicación e interpretación, el modelo de regresión lineal y el modelo de regresión logística. Teniendo en cuenta el tipo de variable que deseemos estimar (variable dependiente o respuesta) aplicaremos un modelo de regresión u otro. Simplificando, cuando la variable dependiente es una variable continua, el modelo de regresión más frecuentemente utilizado es la regresión lineal, mientras que cuando la variable de interés es dicotómica (es decir, toma dos valores como sí/no, hombre/mujer) se utiliza la regresión logística.

Para poder clarificar los conceptos básicos y los supuestos del análisis de regresión, encontramos que es útil comenzar con una sola variable independiente, la cual simplemente denotaremos como X . El problema más simple (no necesariamente el más trivial) se presenta cuando dada una muestra de n individuos obtenemos para cada valor de (X) un valor de (Y) obteniendo n pares que podemos denotar $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, donde los subíndices ahora hacen referencia a observaciones de individuos más que a diferentes variables,

de tal manera que los pares obtenidos se deben de considerar como puntos del espacio bidimensional

ASPECTOS BÁSICOS QUE DEBEN DE SER RESUELTOS.

Existen dos aspectos en la regresión que deben de ser analizados.

1. ¿Qué tipo de modelo matemático es el más apropiado? ¿una línea recta, una parábola, una función logarítmica?
2. Partiendo de un modelo específico ¿Cómo podemos obtener un mejor ajuste? Es decir si nuestro modelo es una línea recta ¿Cómo podemos obtener la mejor recta que se aproxime más a la realidad?

Matemáticamente una línea recta puede ser descrita por la siguiente ecuación:

$$y = \beta_0 + \beta_1 x$$

Observe que en la anterior ecuación hemos usado letras minúsculas para las variables dependientes e independientes, para enfatizar que estamos tratando las variables matemáticamente más que en el contexto estadístico. Los símbolos β_0 y β_1 son constantes para la recta dada y representan el intercepto y la pendiente respectivamente.

SUPUESTOS ESTADÍSTICOS PARA EL MODELO DE REGRESIÓN SIMPLE.

Suponga que tentativamente asumimos un modelo lineal para determinar la relación entre las variables dependientes e independientes, no tenemos ningún problema en decidir que tenemos un buen ajuste cuando cada pareja de puntos pasa por la recta del modelo. Desafortunadamente, esta situación casi nunca acontece y la recta del modelo de regresión no puede predecir exactamente el valor de la variable dependiente para un valor determinado de la variable independiente. En efecto esto sucede debido a que se toma una muestra de la población, lo cual requiere estimar parámetros desconocidos de la población.

Estos parámetros son: β_0 y β_1 los cuales deben de ser estimados de manera que se aproximen lo más posible a la realidad. El método más apropiado para lograr este fin es el que se conoce como estimación de parámetros por mínimos cuadrados.

Para hacer inferencias sobre la población por medio de una muestra, necesitamos tener en cuenta los siguientes supuestos: Existencia, independencia, linealidad, homoscedasticidad y normalidad.

MEDIDA DE BONDAD DE AJUSTE PARA LA ECUACIÓN DE LA REGRESIÓN

La ecuación tendría un ajuste perfecto si el valor obtenido para predecir la variable dependiente fuera exactamente igual al de la variable dependiente, pero en la práctica esto no ocurre, por lo que es necesario implementar un coeficiente de determinación el cual permite calcular el porcentaje de efectividad del modelo matemático obtenido por la regresión. El coeficiente de regresión está dado por:

$$r^2 = \frac{SCR}{SCT} \times 100$$

$$r^2 = \left[1 - \frac{SCE}{SCT} \right] \times 100$$

Si extremos la raíz cuadrada del coeficiente de determinación obtenemos el coeficiente de correlación con el signo de la pendiente, este indica como dos variables están asociadas linealmente.

ANÁLISIS DE VARIANZA.

El método del análisis de varianza y el análisis de regresión están fuertemente relacionados, se puede decir que los problemas que son tratados por el análisis de varianza tienen una estructura de regresión, este hecho conlleva a que los resultados obtenidos en la tabla de la ANOVA puedan ser usados como resultados obtenidos por la regresión. El análisis de varianza mediante la tabla de ANOVA, conduce a un resumen de resultados, donde se puede obtener algunos estimados de varianza para resolver o dar respuesta a algunas inquietudes sobre la inferencia del análisis de regresión. En el caso de la regresión lineal simple se debe de resolver:

1. Existe relación entre las variables X y Y . Esto conduce a probar si el estimado de la pendiente $\hat{\beta}_1 = 0$ (no importa el resultado obtenido por la estimación de mínimos cuadrados, se debe hacer esta prueba)
2. ¿Qué tan fuerte es la relación entre las dos variables en caso de que esta exista?
3. ¿Es el modelo de regresión lineal simple el apropiado para describir la relación entre las variables?

MÉTODOLOGÍA

- Fundamentación teórica: Conceptos básicos, Supuestos de la regresión
- Ejemplos de aplicación desarrollados en calculadora, Excel, PSPP
- Taller de aplicación en PSPP.

En la primer sesión de fundamentación teórica, mediante clase magistral se explicaran los conceptos Básicos de regresión, modelo, correlación, parámetros, métodos de estimación de parámetros por mínimos cuadrados, varianza.

En la segunda sesión, mediante ejemplosprácticos se explicara los procedimientos pertinentes para la regresión simple con el uso de calculadora, Excel y el paquete estadístico PSPP, También se realizara ejemplos de regresión múltiple.

Para la tercera sesión se hará taller de aplicación en el paquete PSPP

BIBLIOGRAFIA

Montgomery D. Peck E. Vining G. Introducción al análisis de regresión lineal Editorial continental México 2002

Thompson, M. Selección de variables en la regresión múltiple Part II Chosen procedures , computations and examples Rev. 46,1978, pp. 129-146

Weisberg S Applied Linear regression, segunda edition, Wiley, Nueva York, 1985