Nombre del alumno: Nathasha Vanesa Aguilar Méndez

Nombre del profesor: Rosario Gómez Lujano

Nombre del trabajo: Ensayo “Prueba de hipótesis”, mapa conceptual “Prueba de hipótesis con dos muestras” y ejercicios

Materia: Estadística inferencial

Grado: 4° cuatrimestre

Grupo: único

Pichucalco, Chiapas a 05 de Diciembre de 2020.

Introducción

La mayoría conoce el término de hipótesis desde hace mucho tiempo, podemos decir que es la tentativa de explicación de algún fenómeno o problema que puede ser corroborado mediante observación o experimentación. Los científicos necesitan proponer hipótesis como posibles explicaciones al problema que desean resolver.

Sin embrago en nuestra materia de estadística inferencial vamos a centrarnos en un tema específico que es: la prueba de hipótesis, que es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos.

De igual manera se abarcara dentro del tema puntos como : hipótesis nula y alternativa, error tipo I y II, contraste de hipótesis bilateral para media, hipótesis y prueba de hipótesis, procedimiento sistemático para prueba hipótesis, prueba para proporciones.

ENSAYO

Prueba de hipótesis con una muestra

Justificación de la prueba de hipótesis

Veamos que la prueba de hipótesis es un método esencial para la toma de decisiones. La decisión relaciona la elección entre dos enunciados competitivos y mutuamente excluyentes, respecto de uno o más parámetros de la población. Los enunciados competitivos se conocen como hipótesis nula y alternativa, respectivamente.

Con base a lo anterior, es necesario señalar los atributos principales que debe poseer una hipótesis:

Debe hacer referencia a una situación real.

Las variables que se presentan en su planteamiento deben ser precisas, comprensibles y concretas.

La relación entre las variables debe ser clara, verosímil y lógica.

Los términos y las relaciones planteadas deben ser observables y medibles.

Las variables deben estar relacionadas con técnicas disponibles para probarlas.

Hipótesis nula y alternativa

La hipótesis nula, representada por H_o, es la afirmación sobre una o más características de poblaciones que al inicio se supone cierta. Indica que un parámetro de población (tal como la media, la desviación estándar, etc.) es igual a un valor hipotético. La hipótesis nula suele ser una afirmación inicial que se basa en análisis previos o en conocimiento especializado.

La hipótesis alternativa, representada por H_1, es la afirmación contradictoria a H_o, y ésta es la hipótesis del investigador. Indica que un parámetro de población es más pequeño, más grande o diferente del valor hipotético de la hipótesis nula. La hipótesis alternativa es lo que usted podría pensar que es cierto o espera probar que es cierto.

Error tipo I y II

El error tipo I, también llamado alfa (α), se comete al rechazar la hipótesis nula (H0) siendo esta verdadera. Así, la probabilidad de cometer un error de tipo I es α, que es el nivel de significación que hemos establecido para nuestra prueba de hipótesis.

Si por ejemplo el α que habíamos establecido es de 0.05, esto indicaría que estamos dispuestos a aceptar una probabilidad del 5% de equivocarnos al rechazar la hipótesis nula.

El error tipo II o beta (β), se comete al aceptar la hipótesis nula (H0) siendo esta falsa. Es decir, la probabilidad de cometer un error tipo II es beta (β), y depende de la potencia de la prueba (1-β).

Para reducir el riesgo de cometer un error tipo II, podemos optar por asegurarnos de que la prueba tiene suficiente potencia. Para ello, deberemos asegurarnos de que el tamaño de la muestra es lo suficientemente grande como para detectar una diferencia cuando ésta realmente exista.

Contraste de hipótesis bilateral para media

El contraste bilateral sitúa la región de rechazo en los dos extremos (colas) de la distribución muestral. En cambio, el contraste unilateral sitúa la región de rechazo en uno de los dos extremos (colas) de la distribución muestral. El contraste bilateral (o de dos colas) se utiliza cuando la Hipótesis Alternativa asigna al parámetro cualquier valor diferente al establecido en la Hipótesis Nula.

Hipótesis y prueba de hipótesis

Una hipótesis se define como una afirmación transitoria que debe ser sometida a prueba. La inferencia estadística propone un procedimiento para llevar a cabo la prueba de las hipótesis. Propone, primero, enunciarlas formalmente y luego contrastarlas con la evidencia de los datos. Son los datos, entonces, con su coro de características, los que dirán si una hipótesis es falsa o verdadera.

Este procedimiento se realiza considerando a los parámetros, que ya sabemos corresponden al universo, como los objetos para los cuales se enuncian las hipótesis. Dicho de otro modo, una hipótesis se enuncia para una característica del universo o población y se origina en la observación del comportamiento de la misma característica en un grupo restringido o muestra.

Como ya sabemos que una prueba de hipótesis es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos.

Una prueba de hipótesis examina dos hipótesis opuestas sobre una población: la hipótesis nula y la hipótesis alternativa. La hipótesis nula es el enunciado que se probará. Por lo general, la hipótesis nula es un enunciado de que "no hay efecto" o "no hay diferencia". La hipótesis alternativa es el enunciado que se desea poder concluir que es verdadero de acuerdo con la evidencia proporcionada por los datos de la muestra.

Con base en los datos de muestra, la prueba determina si se puede rechazar la hipótesis nula. Utiliza el valor p para tomar esa decisión. Si el valor p es menor que el nivel de significancia (denotado como α o alfa), entonces puede rechazar la hipótesis nula.

Primer paso: Plantear la hipótesis

La prueba de hipótesis formula dos hipótesis estadísticas que deben anunciarse explícitamente: hipótesis nula y alternativa. La primera, se designa por el símbolo H0. Esta hipótesis se conoce también como la hipótesis de no diferencia, ya que es una proposición de conformidad con (o sin diferencia respecto a) Características que se suponen ciertas en la población de interés. Esta hipótesis siempre se opone a la hipótesis del investigador.

La segunda, identificada mediante el símbolo H1, es una proposición que se creerá cierta si los datos de la muestra siguieren que llevan al rechazo de la H0 es falsa. Por lo general, la H1 y la hipótesis de investigación son la misma y, de hecho; se utilizan los dos términos indistintamente.

En general H0, esta se establece con el propósito expreso de ser rechazada. Si no se rechaza, esto no necesariamente significa que es verdadera, se dirá que los datos sobre los cuales se basa la prueba no proporcionan evidencia suficiente que cause el rechazo. Por el contrario, si se rechaza se concluye que los datos disponibles no son compatibles con la H0, pero sirven como apoyo a alguna otra hipótesis. Rechazarla entonces, sugiere que la hipótesis alternativa puede ser verdadera.

Segundo paso: Seleccionar el nivel de significancia.

Nivel de significancia: Probabilidad de rechazar la hipótesis nula cuando es verdadera. Se le denota mediante la letra griega α, también es denominada como nivel de riesgo, este término es más adecuado ya que se corre el riesgo de rechazar la hipótesis nula, cuando en realidad es verdadera. Este nivel está bajo el control de la persona que realiza la prueba.

Si suponemos que la hipótesis planteada es verdadera, entonces, el nivel de significación indicará la probabilidad de no aceptarla, es decir, estén fuera de área de aceptación. El nivel de confianza (1-α), indica la probabilidad de aceptar la hipótesis planteada, cuando es verdadera en la población.

La distribución de muestreo de la estadística de prueba se divide en dos regiones, una región de rechazo (conocida como región crítica) y una región de no rechazo (aceptación). Si la estadística de prueba cae dentro de la región de aceptación, no se puede rechazar la hipótesis nula. La región de rechazo puede considerarse como el conjunto de valores de la estadística de prueba que no tienen posibilidad de presentarse si la hipótesis nula es verdadera. Por otro lado, estos valores no son tan improbables de presentarse si la hipótesis nula es falsa. El valor crítico separa la región de no rechazo de la de rechazo.

Las pruebas de proporciones son adecuadas cuando los datos que se están analizando constan de cuentas o frecuencias de elementos de dos o más clases. El objetivo de estas pruebas es evaluar las afirmaciones con respecto a una proporción (o Porcentaje) de población. Las pruebas se basan en la premisa de que una proporción muestral (es decir, x ocurrencias en n observaciones, o x/n) será igual a la proporción verdadera de la población si se toman márgenes o tolerancias para la variabilidad muestral. Las pruebas suelen enfocarse en la diferencia entre un número esperado de ocurrencias, suponiendo que una afirmación es verdadera, y el número observado realmente. La diferencia se compara con la variabilidad prescrita mediante una distribución de muestreo que tiene como base el supuesto de que es realmente verdadera.

En muchos aspectos, las pruebas de proporciones se parecen a las pruebas de medias, excepto que, en el caso de las primeras, los datos muestrales se consideran como cuentas en lugar de como mediciones. Por ejemplo, las pruebas para medias y proporciones se pueden utilizar para evaluar afirmaciones con respecto a:

- Un parámetro de población único (prueba de una muestra)

-La igualdad de parámetros de dos poblaciones (prueba de dos muestras), y

-La igualdad de parámetros de más de dos poblaciones (prueba de k muestras). Además, para tamaños grandes de muestras, la distribución de muestreo adecuada para pruebas de proporciones de una y dos muestras es aproximadamente normal, justo como sucede en el caso de pruebas de medias de una y dos muestras.

Bibliografía

SUÁREZ, Mario, (2012), Interaprendizaje de Probabilidades y Estadística Inferencial con Excel, Winstats y Graph, Primera Edición. Imprenta M & V, Ibarra, Ecuador.

Hernández, R. (2006) Metodología de la Investigación. Cuarta Edición. México. McGraw-Hill Interamericana.

Daniel, Wayne (2005). Bioestadística: base para el análisis de las ciencias de la salud. 4a ed., México: Limusa-Wiley.

Conclusión

Para cerrar con nuestro tema es importante recordar que Los estadísticos inferenciales utilizan una muestra aleatoria de datos tomada de una población para describir y hacer inferencias acerca de la población.

Los estadísticos inferenciales son valiosos cuando no es conveniente o posible examinar cada miembro de una población entera. Por ejemplo, no resulta práctico medir el diámetro de cada clavo fabricado en una acería, pero usted puede medir los diámetros de una muestra aleatoria representativa de los clavos y utilizar esa información para hacer generalizaciones sobre los diámetros de todos los clavos producidos.

PRUEBAS DE HIPÓTESIS CON DOS MUESTRAS Y VARIAS MUESTRAS DE DATOS NUMÉRICOS

Distribuciones normales y t de Student.

Distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño

Prueba de una y dos colas

En las "pruebas bilaterales o de dos colas" se comparan dos muestras para saber si difieren entre sí, sin preguntarse cuál de ellas tiene mayor estadístico

Si se pretende evaluar qué muestra tiene el estadístico mayor (sesgo positivo) se realiza una "prueba unilateral o de una cola"

Regresión y correlación

Regresión: se evalúa la dependencia de una variable "efecto" con respecto a varias variables independientes "causas"

Correlación: se evalúa la concordancia entre dos variables independientes (ambas juegan un papel simétrico)

Correlación por ajustes de una recta con el criterio de mínimos cuadrados

La recta de correlación pasa por el punto (x, y) siendo x, y las medias de los datos xi e yi. Aplicando la ecuación punto-pendiente de una recta podemos obtener la ecuación de la recta de mínimos cuadrados: y= y + B(x- x)

Errores de la pendiente y ordenada en el origen de la recta de regresión

La recta de regresión calculada se utiliza para estimar la concentración de las muestras problema por interpolación, y quizá también para estimar el límite de detección del método A es una constante y no participa en la suma de cuadrados.

Regresión lineal

En la primera etapa, meramente descriptiva, se utiliza el ajuste por mínimos cuadrados para hallar la ecuación de la recta que se ajuste mejor a los datos "recta de regresión".

En la primera etapa, meramente descriptiva, se utiliza el ajuste por mínimos cuadrados para hallar la ecuación de la recta que se ajuste mejor a los datos "recta de regresión".

Vertiente descriptiva o correlación

Etapas

Cálculo de la recta de regresión por el ajuste de mínimos cuadrados

Descomposición de la suma de cuadrados

Valoración de la bondad del ajuste por el coeficiente "r2

Vertiente inferencial o regresión

La regresión lineal pretende deducir la relación lineal entre una variable dependiente, y otras independientes que la condicionan

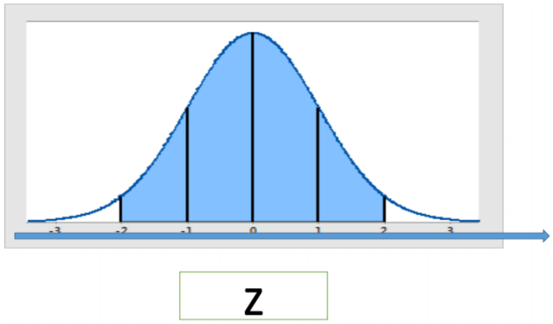
El modelo exige que los datos de la población cumplan los supuestos de linealidad, homocedasticidad, independencia entre las observaciones y normalidad de las distribuciones condicionadas de la variable "y"

Ejercicios:

1.- Determine el valor de Z, cuando la media es cinco, sabiendo que la desviación estándar es 2 y x=15

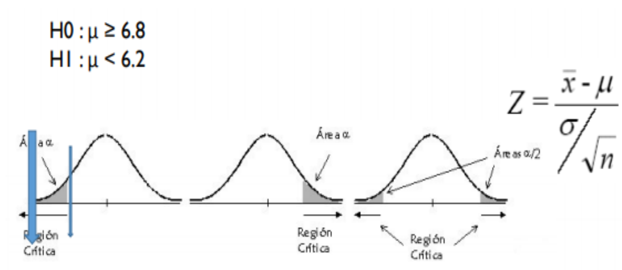
$$Z = \frac{x - M}{\sigma}$$

$$Z = \frac{15 - 5}{2} = 5$$



2.- Una encuesta nacional determino que los estudiantes de secundaria miraban en promedio 6.8 películas al mes, con una desviación estándar de 0.5 horas. Una muestra aleatoria de 36 estudiantes revelo que la cantidad media de película que vieron fue de 6.2. Con un nivel de significancia de 0.05, ¿Se puede concluir que los estudiantes universitarios ven menos películas que los estudiantes de secundaria?

Datos del problema:



Media teórica=6.8

Películas n= 36

Media muestral=6.2

Películas Desviación estándar=0.5

Horas Nivel de significancia=0.05

Nivel de confianza=95%

$$Z=\frac{6.2-6.8}{0.5/\sqrt{36}}=\frac{-0.6}{0.5/6}=-7.5$$