

Análisis Paramétricos y No Paramétricos

A continuación se describirán brevemente los análisis paramétricos y no paramétricos más utilizados tanto en la investigación académica como en la de negocios.

Paramétricos	No Paramétricos
<ul style="list-style-type: none">• Coeficiente de Correlación de Pearson y Regresión Lineal• Prueba t• Prueba de la Diferencia de Proporciones• Análisis de Varianza	<ul style="list-style-type: none">• Chi Cuadrada y las Tabulaciones Cruzadas• Coeficientes de Spearman y Kendall

Siguiendo el análisis sugerido por Hernández Sampieri et al. (2010), daremos una definición de cada prueba y la analizaremos en función de la forma en que se establecen las hipótesis, el tipo y número de variables que se utilizan, el nivel de medición de esas variables y cómo se interpretan los resultados.

Coeficiente de Correlación de Pearson

Análisis de correlación paramétrico para variables continuas.

Hipótesis a probar: analiza la relación entre dos variables.

Variables: dos variables. La correlación se hace de dos en dos variables.

Análisis Paramétricos y No Paramétricos

Nivel de medición de las variables: de intervalo o de razón.

Interpretación: el coeficiente r de Pearson varía entre -1 y 1. -1 significa una relación negativa perfecta, 0 que no existe ninguna correlación entre las variables y 1 que existe una correlación positiva perfecta.

Regresión Lineal

Es el modelo matemático en que se demuestra el efecto de una variable sobre otra. Basándose en la r de Pearson se puede estimar la puntuación de una variable en base al valor de otra. A mayor correlación, mayor capacidad de predicción de la variables.

Hipótesis a probar: de tipo correlacional y causal entre las variables.

Variables: al menos dos. Al menos una variable dependiente y al menos una variable independiente.

Nivel de medición de las variables: de intervalo o de razón.

Interpretación: la ecuación de regresión lineal que arroja el paquete estadístico es del tipo:

$$Y = a + bX$$

Donde:

Y: Es un valor de la variable dependiente que se desea predecir.

a: Es la ordenada en el origen o intercepto.

b: Es la pendiente o inclinación de la línea que se forma al graficar todos los datos.

X: Es el valor fijado para la variable independiente.

Análisis Paramétricos y No Paramétricos

Los datos a y b son proporcionados por el paquete estadístico. Para hacer una predicción se sustituyen los valores correspondientes en la ecuación.

Prueba t

Se utiliza para establecer si las medias de dos grupos difieren entre sí de manera significativa. Se utiliza el símbolo t .

Hipótesis a probar. diferencia entre dos grupos.

Variables: se realiza sobre una variable teóricamente dependiente. Si existen diferentes variables, hay que realizar una prueba para cada una de ellas.

Nivel de medición de las variables: intervalos o de razón.

Interpretación: la fórmula para esta prueba se encuentra en cualquier libro de estadística. Para saber si el resultado es significativo se requiere estimar los grados de libertad ($gl = (n_1 + n_2) - 2$, donde n_1 y n_2 corresponden al tamaño de los grupos que se comparan) y un nivel de significancia. El valor de t obtenido mediante la fórmula se compara con el valor de la tabla de t de Student (dado por el nivel de significancia y los grados de libertad). Si el valor calculado es igual o mayor al que aparece en la tabla, se rechaza la hipótesis nula. Si por el contrario, el valor calculado es menor al valor de tablas, se acepta.

Prueba de la Diferencia de las Proporciones

Sirve para determinar si dos proporciones varían significativamente o no.

Hipótesis a probar. de diferencia de proporciones en dos grupos.

Variables: una variable. Si existe más de una variable de interés, se realizará una prueba para cada una.

Análisis Paramétricos y No Paramétricos

Nivel de medición de las variables: de intervalo o de razón expresados en porcentajes o proporciones.

Interpretación: se obtienen las proporciones de los grupos y se aplica la siguiente fórmula.

$$z = \frac{P1 - P2}{\sqrt{\frac{P1q1}{n1} + \frac{P2q2}{n2}}}$$

El valor z resultante se compara con el valor de tablas z (normal) que corresponda al nivel de significancia elegido. Si el valor calculado es igual o mayor al valor de tablas se acepta la hipótesis de investigación (hipótesis alterna), de lo contrario, no se puede rechazar la Ho.

Análisis de Varianza

Sirve para evaluar si las medias y varianzas de dos grupos o más difieren significativamente entre sí. Para dos grupos se utiliza la prueba *t* y el *análisis de varianza unidireccional* se usa para tres o más grupos.

Hipótesis a probar: de diferencia entre dos o más grupos. La hipótesis de investigación propone que las medias y varianzas varían significativamente entre los grupos. La hipótesis nula que son similares.

Variables: una variable dependiente y una variable independiente.

Nivel de medición de las variables: la variable dependiente es por intervalo o de razón. La variable independiente debe ser categórica, es decir, que permita la formación de al menos dos grupos. La variable independiente puede ser nominal, ordinal, por intervalos o de razón. En los dos últimos casos la escala de la variable debe transformarse a categorías.

Análisis Paramétricos y No Paramétricos

Interpretación: se espera que los integrantes de un grupo sean similares entre sí pero diferentes a los integrantes de otros grupos. Con este fin el análisis de varianza unidireccional produce el valor de F , que nos indica si las diferencias entre los grupos son mayores a las diferencias intragrupos (dentro de los grupos). Estas diferencias se miden en términos de varianza ($S = (x - \bar{x})^2$), así la fórmula de F es la siguiente

$$F = \frac{\text{Media cuadrática entre los grupos}}{\text{Media cuadrática dentro de los grupos}}$$

Para aplicar esta fórmula se requieren las siguientes estimaciones de la media cuadrática:

$$\text{Media cuadrática entre los grupos} = \frac{\text{Suma de cuadrados entre grupos}}{\text{Grado de libertad entre grupos}}$$

Los grados de libertad entre grupos = $K - 1$ (donde K es el número de grupos).

$$\text{Media cuadrática dentro de los grupos} = \frac{\text{Suma de cuadrados intragrupos}}{\text{Grado de libertad intragrupos}}$$

Los grados de libertad intragrupos = $n - K$ (donde n es el tamaño de la muestra y K es el número de grupos).

Análisis Paramétricos y No Paramétricos

El resultado arrojado por un paquete estadístico incluye un valor para la significancia de F (α o probabilidad). Si este valor es menor al valor de significancia especificado por el investigador (e.g., 0.05 o 0.01) la prueba es significativa, es decir existe una diferencia entre las medias y varianzas entre grupos. Ver el siguiente ejemplo:

ANOVA

Tiempo

	Suma de cuadrados	gl	Media cuadrada	F	Sig.
Inter-grupos	8901,537	5	1780,307	6,450	,000
Intra-grupos	29810.297	108	276,021		
Total	38711.833	113			

Chi cuadrada

Se utiliza para determinar la relación entre dos variables categóricas. Se simboliza con χ^2

Hipótesis a probar. de tipo correlacional. No se consideran relaciones causales.

Variables: dos variables categóricas.

Análisis Paramétricos y No Paramétricos

Nivel de medición de las variables: de tipo nominal u ordinal. Las de intervalos o de razón se pueden utilizar si antes se convierten en categóricas.

Interpretación: se estima mediante una tabulación cruzada o tabla de contingencia. Al correr esta herramienta en un paquete estadístico tendremos las frecuencias *observadas* (las pertenecientes a nuestros datos de la muestra) y las *esperadas* que es una estimación de los valores si las variables fueran estadísticamente independientes o no estuvieran relacionadas. La frecuencia esperada se calcula de la siguiente forma:

$$f_e = \frac{(\text{Total de renglón})(\text{Total de columna})}{n}$$

Ya teniendo las frecuencias esperadas, se estima el valor de la Chi cuadrada mediante la siguiente fórmula:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Donde:

Σ = Es la sumatoria

O = Frecuencia observada en cada celda

E = Frecuencia esperada en cada celda

Análisis Paramétricos y No Paramétricos

El valor de la χ^2 está definido por los grados de libertad. La fórmula de los grados de libertad es:

$$gl = (r-1)(c-1)$$

Donde:

r = número de renglones de la tabla de contingencia

c = el número de columnas en la tabla.

En una tabla de Distribución de la χ^2 , que se puede consultar en cualquier libro de estadística o en Internet, con los grados de libertad estimados y el nivel de significancia especificado por el investigador, se determina el valor de la χ^2 .

Si el valor calculado de χ^2 es igual o mayor al valor de tablas, se dice que sí existe una relación entre las dos variables. La relación es significativa.

Coeficientes de Spearman y Kendall

Estos coeficientes de correlación sirven para estimar el nivel de relación entre dos variables. La decisión de cuál utilizar depende del tipo de escala que tengan las variables bajo análisis.

Análisis Paramétricos y No Paramétricos

Es conveniente recordar que la χ^2 sirve para evaluar la correlación entre dos variables expresadas en categorías.

El coeficiente *rho de Spearman* (r_s) y *tau de Kendall* (t), son medidas de correlación para variables con escalas ordinales, de tal manera que los miembros de la muestra se pueden ordenar por rangos. Son análisis no paramétricos.

Los resultados obtenidos de la aplicación de estas técnicas servirán para elaborar un reporte de resultados, como se verá en la siguiente lección.