

QUIZ ESTADÍSTICA 2019-II, SEMANAS 15 Y 16

Maria Alejandra Navarro Corredor – 2161578

Pablo Emilio Mojica Pradilla – 2160008

MANOVA: Paramétrico, no tiene análogo no paramétrico. Análisis multifactoriales de varianza o análisis de varianza con más de un factor o variable independiente (1). No solo puede tener en cuenta muchas variables independientes, también usa varias variables dependientes que, de algún modo miden la misma respuesta desde distintos puntos de vista. El ANOVA de dos vías es el modelo más simple, ya que solo hay dos factores o variables independientes (de agrupación) y una variable dependiente (la que se compara) (4).

Explicación del ejemplo: El ejemplo del crecimiento de dientes nos muestra que sí hay diferencias según tanto la dosis como el suplemento, y también relacionando ambos factores. Según la comparación entre pares, las diferencias se presentan entre todas las dosis y entre ambos suplementos. Por último, nos indica que existe diferencia entre todas las combinaciones de dosis y suplementos menos en cuatro (VC:D1-OJ:D.05, OJ:D2-OJ:D1, VC:D2-OJ:D1 y VC:D2-OJ:D2).

Regresión logística: Paramétrico. Se usa la regresión logística binaria cuando se desea conocer el modo en que diversos factores (variables cuantitativas o categóricas) se asocian simultáneamente a una variable cualitativa o categórica dicotómica. Hay otros modelos, como la regresión logística nominal para variables categóricas con más de dos categorías o la regresión logística ordinal (4).

Regresión múltiple: Paramétrico. Puede incluir dos o más variables independientes (predictor, variable X) y una sola variable dependiente Y. en este caso la variable Y es cuantitativa (4). El objetivo del análisis de regresión es predecir los valores de la variable respuesta, en función de los valores de las variables independientes (1).

Análisis factorial: Paramétrico. En éste análisis no hay una variable dependiente o muchas independientes que se usen para explicarla o predecirla, sino que pretende extraer de una base de datos con muchas variables un pequeño grupo de factores que consigan proporcionar de manera resumida gran parte de la información contenida en todas las variables iniciales. Por lo tanto, es una técnica de reducción de variables (4).

Análisis de discriminantes: Paramétrico. Sirve para determinar si existen diferencias estadísticamente significativas entre los perfiles de las puntuaciones medias sobre el conjunto de variables de dos o más grupos definidos a priori (5). Parte de una tabla de datos de n individuos a los que se les ha medido p variables cuantitativas independientes, que actúan como perfil de características de cada uno de ellos. Una variable cualitativa adicional, dependiente o clasificatoria, con dos o más categorías, define el grupo al que cada individuo pertenece (6).

Análisis de clúster, conglomerado o grupos: Paramétrico y no paramétrico. Tiene por objetivo agrupar elementos en grupos homogéneos en función de las similitudes o similaridades entre ellos. El análisis de conglomerados estudia tres tipos de problemas: Partición de los datos, se dispone de datos que son heterogéneos y se dividen en un número de grupos prefijado; Construcción de jerarquías, se estructuran los elementos de un conjunto de forma jerárquica por su similitud; Clasificación de variables, en problemas con muchas variables es preferible dividir las en grupos (2).

K-medias: Paramétrico y no paramétrico. Análisis de conglomerado k-medias. Hay una muestra de n elementos y p variables, la idea es dividir la muestra en K grupos (2). Cada observación es asignada al clúster o grupo que posee un centroide (media que se usa para describir el clúster y cuyo valor es comparado con el valor que toma cada dato) con un valor más cercano o similar. Al

introducirse un nuevo dato en el clúster, el valor del centroide cambia automáticamente, con ello algunos individuos podrían cambiar de centroide. Se debe recalcular el centroide y estos cálculos se repiten hasta que ningún dato cambie de grupo (4).

Escalamiento multidimensional: No paramétrico. Las técnicas de escalado multidimensional son una generalización de la idea de componentes principales cuando en lugar de disponer de una matriz de observaciones por variables, se dispone de una matriz cuadrada $n \times n$ de distancias o disimilitudes entre n elementos de un conjunto (2).

Matriz de correlación: Paramétrico y no paramétrico. Es una matriz cuadrada y simétrica que tiene unos en la diagonal y es completada con los coeficientes de correlación entre las variables (cuando en lugar de unos tiene ceros significa que la relación entre variables al menos no es lineal, sino que puede ser de otro tipo). Nos explica cómo se relacionan las variables que estamos estudiando (9).

Explicación del ejemplo: El ejemplo de la relación entre la temperatura del aire y agua nos permite hallar una matriz de correlación que indica que existe una correlación del 92.4% entre las dos variables.

Análisis de mantel: No paramétrico. El test de Mantel estima el grado de correlación existente entre dos matrices X y Y. Los análisis se centran en matrices de similitud/distancias derivadas de los datos originales (8).

Explicación del ejemplo: El test de Mantel nos indica que existe una correlación del 75.1% entre las matrices de los dos lagos, lo que podría estar diciendo que la relación entre las variables y para los sitios es bastante fuerte y los datos son similares para ambos casos.

Correlación: Paramétrico y no paramétrico. Examina la dirección y la magnitud de la asociación entre variables cuantitativas. Así se conoce la intensidad de la relación entre ellas, es decir, se mide o cuantifica el grado de asociación que mantiene y se sabe si, al aumentar el valor de una variable, aumenta o disminuye el valor de la otra. Las variables son simétricas, no hay una variable dependiente y otra independiente. No hay variable causa ni variable efecto, por ende, es irrelevante en eje que ocupe cada variable. Se usan frecuentemente dos coeficientes de correlación: el de Pearson (paramétrico) y el de Spearman (no paramétrico). (4).

Explicación del ejemplo: El test de correlación nos indica que las variables de viento y temperatura presentan una relación negativa moderada (45.8%), es decir, que a medida que disminuye el viento, la temperatura tiende también a disminuir, pero no en gran proporción. Los resultados son significativos.

Componentes principales: Paramétrico y no paramétrico. Es útil para estudiar las relaciones que se presentan entre x variables correlacionadas que nos brindan una información determinada. Para lograr esto se transforma el conjunto original de variables en otro que presenta variables incorreladas entre sí (sin redundancias). No requiere de supuestos de normalidad, aunque si se cumplen es posible que se hagan interpretaciones más profundas de los componentes (10).

Explicación del ejemplo: El análisis de componentes principales explicó un 60.7% de la variabilidad de los datos en las 2 primeras tendencias de variación. El primer eje explicó un 33.5% del total de la varianza y está asociado positivamente con las variables tem. media y precipitación anual y negativamente con área y elevación. El segundo eje de ordenación explicó un 27.2% y se relacionó positivamente con la prec. seso y radiación del sol y negativamente con isoterminia. En general los páramos de la cord. occidental tienen precipitación anual mayor, los páramos de la cord. central y NP tienen una menor isoterminia en comparación con los demás paramos.

Análisis de correspondencia canónica: No paramétrico. Es un tipo de análisis estadístico lineal de múltiples variables. Se usa para analizar relaciones multidimensionales entre múltiples variables independientes y múltiples variables dependientes. Su objetivo es cuantificar la validez de la relación entre los dos conjuntos de variables (dependiente e independiente), determina si dos conjuntos de variables son independientes uno de otro o, inversamente, determinar la magnitud de las relaciones que pueden existir entre los dos conjuntos (7).

Análisis de covarianza (ANCOVA): Paramétrico. Es una técnica que combina una regresión y un ANOVA. Incluye una variable dependiente y una o más variables cuantitativas, llamadas covariables, que actúan como predictores. ANCOVA determina la correlación entre las covariables y la variable dependiente, y posteriormente retira la varianza asociada con las covariables a la variable dependiente antes de determinar si las diferencias en la variable dependiente son significativas. Su objetivo es eliminar cualquier error sistemático fuera del control del investigador que pudiera llegar a sesgar los resultados, además de tener en cuenta las diferencias debidas a las características propias de los sujetos incluidos en el estudio (4).

Bibliografía:

1. Cáceres R. 1995 Estadística multivariante y no paramétrica con SPSS.
2. Peña D. 2013 Análisis de datos multivariantes.
3. Cuadras C. 2007 Nuevos métodos de análisis multivariante.
4. Basterra-Gortari F, Bes-Rastrollo M, Estévez J, Fernández-Pachecho P, García-Lopez M, Gea-Sánchez A, Guillén-Grima F, Burgo C, López-Fidalgo J, Martín-Calvo N. 2014 Bioestadística amigable.
5. Mendoza-Méndez R, Dorantes-Coronado E, Cedillo-Monroy J, Jasso-Arriaga X. 2017 El método estadístico de análisis discriminante como herramienta de interpretación del estudio de adicción al móvil, realizado a los alumnos de la Licenciatura en Informática Administrativa del Centro Universitario UAEM.
6. Carvajal P, Trejos A, Soto J. 2004 Aplicación del análisis discriminante para explorar la relación entre el examen de ICFES y el rendimiento en álgebra lineal de los estudiantes de ingeniería de la UTP en el período 2001-2003.
7. Badii M, Castillo J, Cortez K, A. Wong, Villalpando P. 2017 Análisis de correlación canónica (ACC) e investigación científica.
8. Luzuriaga A, Olano J. 2006 Con los pies en el suelo: incluyendo la estructura espacial de los datos en los análisis multivariantes.
9. Hernández, Y. 2015. Matriz de correlación.
10. Fernández, S. 2011. Componentes principales.