

ESTADÍSTICA MULTIVARIADA

Dr. José Dionicio Zacarias Flores

¿PORQUÉ ESTADÍSTICA MULTIVARIADA?

- Como su nombre lo indica, la estadística multivariada se refiere a una variedad de técnicas descriptivas e inferenciales que se han desarrollado para manejar situaciones en las que intervienen conjuntos de variables como predictores o como medidas de rendimiento. Si los investigadores tuvieran una mentalidad lo suficientemente estrecha o las teorías y técnicas de medición tan bien desarrolladas o la naturaleza fuera tan simple como para dictar una sola variable independiente y una única medida de resultado según sea apropiado en cada estudio, no habría necesidad de técnicas estadísticas multivariadas. En el experimento científico clásico que involucra una única medida de resultado y una sola variable manipulada (todas las demás variables se eliminan como posibles factores causales a través del control experimental explícito o el control estadístico proporcionado por la aleatorización), apenas surgen preguntas de patrones o combinaciones óptimas de variables.

- Del mismo modo, los problemas de las comparaciones múltiples no confunden las interpretaciones de ninguna prueba t o coeficiente de correlación utilizado para evaluar la relación entre la variable independiente (o predictor) y la variable dependiente (o resultado).
- Sin embargo, por razones muy excelentes, los investigadores en todas las ciencias conductuales, biológicas o físicas desde hace mucho tiempo han abandonado la dependencia exclusiva del diseño clásico univariante. Ha quedado muy claro que una manipulación experimental dada (por ejemplo, reforzar positivamente una clase de respuestas en cada uno de los N ensayos) afectará muchos aspectos algo diferentes pero parcialmente correlacionados (por ejemplo, velocidad, fuerza, consistencia y "corrección") de comportamiento del organismo.

- Del mismo modo, muchos datos diferentes sobre un solicitante (por ejemplo, calificaciones de la escuela secundaria en matemáticas, inglés y periodismo; actitud hacia la autoridad y el estado socioeconómico de sus padres) **pueden ser valiosos** para predecir su calificación, promedio en la universidad, y es necesario considerar cómo combinar todos estos datos en una sola "mejor" predicción del rendimiento universitario. (Es ampliamente conocido en nuestra discusión sobre la regresión múltiple que **los predictores que tienen las correlaciones más altas con la variable de criterio cuando se consideran individualmente pueden contribuir muy poco a esa combinación de variables de predicción que se correlaciona más altamente con el criterio**).

- Como está implícito en la discusión del párrafo anterior, las técnicas estadísticas multivariadas logran dos tipos generales de cosas para nosotros, y estas dos funciones corresponden aproximadamente a la distinción entre estadística descriptiva e inferencial.
- **En el lado descriptivo**, proporcionan reglas para combinar las variables de manera óptima. Lo que se entiende por "óptimo" varía de una técnica a otra, como se verá.
- **En el lado inferencial**, proporcionan una solución al problema de comparación múltiple. Casi cualquier situación en la que se apliquen técnicas multivariadas podría analizarse a través de una serie de pruebas de significación univariadas (por ejemplo, pruebas t), utilizando una de esas pruebas univariadas para cada combinación posible de una de las variables predictoras con una de las variables de resultado.

- Sin embargo, debido a que cada una de las pruebas univariadas está diseñada para producir un resultado significativo $\alpha \times 100\%$ del tiempo (donde α es el "nivel de significancia" de la prueba) cuando la hipótesis nula es correcta, la probabilidad de tener al menos una de las pruebas que producen un resultado significativo cuando, de hecho, no hay nada más que una variación casual que aumenta rápidamente a medida que aumenta el número de pruebas. Por lo tanto, **es altamente deseable tener un medio para controlar explícitamente la tasa de error experimental.** Las técnicas estadísticas multivariadas proporcionan este control