

ESTADISTICA PARA LOS CLÍNICOS.

PRUEBAS ESTADÍSTICAS PARA EL CONTRASTE DE HIPÓTESIS(1,2)..

Introducción

A la hora de contrastar nuestra hipótesis nula de partida, varios factores van a determinar la prueba estadística que utilizaremos y se refieren al número, tipo y distribución de las variables y a la forma en que se eligió a los sujetos del estudio.

Cuando trabajamos con una sola variable, ésta podrá ser cualitativa o cuantitativa, por ejemplo, tal y como indicamos en los números anteriores, la prevalencia de demencia o la cifra media de tensión arterial (TA) de nuestra muestra, respectivamente. Si sólo tenemos **una muestra de sujetos**, nuestro interés será saber si dicha muestra difiere en los valores de la variable de los de una población u otra muestra de referencia. No obstante, es más habitual que dispongamos de **dos muestras o grupos de sujetos**, por ejemplo, hombres y mujeres, sanos y enfermos, pacientes con la enfermedad A y pacientes con la enfermedad B, etc. En este caso queremos saber también si existen diferencias entre los dos grupos con respecto a alguna de las variables que hemos estudiado, tal como la media de TA o la frecuencia de diabetes. Aunque el planteamiento estadístico en este caso es el de si dos muestras pertenecen a la misma población de referencia (por ejemplo, los hombres y las mujeres pertenecen a la misma población con respecto a las cifras de TA), los clínicos entendemos mejor el razonamiento anterior: si existen diferencias en la media de TA entre los hombre y las mujeres. De manera similar podemos comparar los valores de la variable de interés para el caso de que se trate de **más de dos muestras**.

Pruebas Paramétricas y Pruebas no Paramétricas

La distribución de la variable que nos interesa comparar también es importante a la hora de establecer la prueba estadística que nos permita evaluar la hipótesis nula de igualdad, de forma que si la distribución es normal y cumple algunas otras condiciones, utilizaremos los denominados **test paramétricos** y si no cumple dichos criterios tenemos la opción de usar los llamados **test no paramétricos** o de distribución libre, que se basan en los rangos de distribución de la variable. Las pruebas no paramétricas son menos potentes, es decir, son más exigentes al rechazar la hi-

pótesis nula de igualdad y por tanto tienen menos posibilidades de acertar cuando no la rechazan (mas posibilidades de cometer un error tipo beta). Por otro lado, la potencia de un test estadístico aumenta cuando lo hace el tamaño de la muestra; de esta forma, para conseguir la misma potencia con una prueba no paramétrica, necesitaremos aumentar el tamaño de la muestra en una cantidad determinada. Así, la potencia de una prueba no paramétrica suele calcularse en referencia a su homóloga paramétrica. Por ejemplo, la U de Mann-Whitney tiene una potencia relativa del 95% con respecto a la prueba paramétrica t de Student (significa que con una muestra de 100 sujetos, se consigue la misma potencia con la U de Mann-Whitney que con 95 sujetos para la t de Student).

Muestras apareadas o dependientes

Si los sujetos de las muestras han sido elegidos de forma que se parecen en bastantes de sus características (el prototipo serían los gemelos, pero también pueden ser compañeros de habitación en un colegio, etc) o se trata de los mismos individuos evaluados en dos momentos diferentes del tiempo, se habla de **muestras apareadas**. En este caso se utilizan pruebas estadísticas especiales para muestras apareadas. Lo que sucede es que, a la hora de valorar las diferencias que se presentan entre dos muestras, el investigador debe de ser muy prudente y, en consecuencia, exigente, cuando considera la importancia del error aleatorio, por si se diera el caso de que los sujetos de ambas muestras fueran muy diferentes con respecto a otras muchas variables presuntamente confundentes (error sistemático). Para el caso de muestras apareadas, es decir aquellas que suponemos muy parecidas con respecto a las referidas variables confundentes (se debe pensar que un sujeto al que se le hacen dos valoraciones, antes-después, de una intervención didáctica, sobre sus conocimientos respecto a una materia, no ha podido cambiar demasiado con respecto a su coeficiente intelectual, estado de ánimo, nivel de alerta, etc., y no digamos sus características demográficas, como edad, sexo, nivel cultural, etc.), esta situación no se presenta y el investigador puede ser menos estricto al considerar la importancia del error aleatorio.

Variable dependiente y Variable Independiente

El esquema que se sigue en la situación en que comparamos dos o más muestras o grupos es que la variable de clasificación de los sujetos en cada una de las muestras se denomina **variable independiente o predictora**, de forma que es la pertenencia a uno de dichos grupos lo que biológica o clínicamente determina los valores que toma la variable que estamos analizando, a la que denominamos **variable dependiente o resultado**. Por ejemplo, en un estudio de seguimiento de una cohorte de sujetos atendidos en el Servicio de Neurología de un Hospital, el hecho de padecer Diabetes Mellitus puede conllevar una mayor incidencia de ictus trombótico o el hecho de ser hipertenso significar mayor incidencia de demencia. En estos ejemplos, la Diabetes Mellitus o la HTA (variables categóricas dicotómicas: los sujetos son o no diabéticos, son o no hipertensos) actúan como variables independientes o predictoras, responsables de la frecuencia de ictus trombótico o demencia, respectivamente. Estas últimas, siendo también variables dicotómicas, resultan ser las variables dependientes o resultado del efecto de las anteriores.

Pruebas de dos colas y pruebas de una cola

En las pruebas de hipótesis se puede partir de la idea de que uno de los grupos de estudio va a tomar valores mayores o menores de la variable resultado que el otro grupo. Por ejemplo, los pacientes hipertensos tendrían mayor frecuencia de demencia que los pacientes normotensos. En este caso utilizaríamos **pruebas de hipótesis de una cola**. Esta no es una actitud prudente en investigación y es más razonable plantear lo que se denominan **pruebas de dos colas**, es decir, considerar la hipótesis "a priori" de que la hipertensión puede conllevar tanto una mayor frecuencia como una menor frecuencia de demencia cuando comparamos con la situación que se presentaría en los pacientes normotensos. La consecuencia será que las diferencias en la frecuencia de demencia entre los grupos deberán ser mayores para que podamos atrevernos a rechazar la hipótesis nula de igualdad, es decir, el valor del estadístico teórico es mayor para el mismo riesgo alfa (no obstante, debemos recordar siempre que esto conlleva mayor posibilidad de cometer un error de tipo beta).

Los test estadísticos de análisis univariante más importantes se detallan en la tabla I. Los paquetes estadísticos disponibles en la actualidad permiten aplicar las diferentes pruebas sin mayor dificultad para el investigador. No obstante, debe recordarse que estos paquetes no piensan por nosotros y suelen dar resultados que debemos interpretar con cautela y siempre con un adecuado sentido clínico.

Tabla I. Tipo de test estadístico para hacer inferencias (comparaciones entre muestras).				
DISTRIBUCION	VARIABLE INDEPENDIENTE (PREDICTORA)	VARIABLE DEPENDIENTE (RESULTADO)	RELACIÓN ENTRE LAS MUESTRAS	PRUEBA ESTADÍSTICA
Normal (Paramétricos)	Una sola muestra (se compara con valor teórico)	Cuantitativa		t-student para una muestra
	Dicotómica	Categoría	No relacionadas	No existe (usar Chi-cuadrado de Pearson)
			Relacionadas	No existe (usar no paramétricos)
	Policotómica	Categoría	Cuantitativa	No relacionadas Relacionadas
Cuantitativa			No relacionadas Relacionadas	No existe (usar Chi-cuadrado de Pearson) ANOVA de una vía ANOVA de medidas repetidas
No normal (No paramétricos)	Una sola muestra (se compara con valor teórico)	Categoría	Relacionadas	Binomial Chi-cuadrado de Pearson Chi-cuadrado de Mantel-Haenzsel Prueba de Kolmogorow-Smirnov Prueba de las Rachas
			No relacionadas	Test exacto de McNemar Prueba de los Signos Chi-cuadrado de Pearson Test exacto de Fisher Test de Wilcoxon Prueba de los signos
	Cuantitativa	Relacionadas	Mann-Whitney Mediana	
		No relacionadas	Z Kolmogorov-Smirnov Rachas de Wald-Wolfowitz Valores extremos de Moses	
	Policotómica	Categoría	No relacionadas	Prueba Q de Cochran
		Cuantitativa	Relacionadas No relacionadas	Prueba de Friedman W de Kendall (concordancia) Prueba de Kruskal-Wallis Mediana K variables ANOVA de dos vías por rangos
COVARIACION (medidas de dos variables en los mismos sujetos o unidades de análisis del estudio)				
Paramétrico	Cuantitativa	Cuantitativa		Correlación de Pearson
No paramétrico	Cuantitativa	Cuantitativa		Correlación de Spearman

COMPARACIÓN DE DATOS CUANTITATIVOS ENTRE DOS O MAS GRUPOS

Como se puede ver en la tabla I, la prueba estadística que se utiliza para contrastar la hipótesis nula de igualdad de medias para una muestra o entre dos muestras

o grupos es la **t de Student**. Para la aplicación de la t de Student se requiere que la distribución de la variable cuantitativa sea normal en ambos grupos de comparación. Existe la posibilidad de que las varianzas de ambos grupos sea iguales o desiguales o que las muestras sean apareadas o no apareadas.

Cuando no se cumplen los criterios de normalidad, se utilizan test no paramétricos, tal como la prueba de **Mann-Whitney**, para el caso de muestras independientes y la prueba de **Wilcoxon** para muestras apareadas, entre otras.

Cuando hacemos comparaciones de datos cuantitativos entre más de dos grupos se utiliza el denominado **Análisis de la Varianza** (ANOVA). ANOVA evalúa la posible igualdad de medias de una variable cuantitativa entre varias muestras sin aumentar la tasa de error tipo I (resuelve el problema de las múltiples comparaciones). El funcionamiento de ANOVA se basa en la descomposición de la variabilidad total de una variable en sus diferentes partes, la debida a las diferencias entre los grupos (efectos principales) y la debida a las diferencias entre los individuos de cada grupo (residuales). Los estudios experimentales, aquellos en los que los sujetos son asignados aleatoriamente a un grupo de tratamiento, son los que utilizan con más frecuencia ANOVA. Este diseño está muy ligado al método experimental y es la base del análisis multivariante. Es posible analizar la información sobre medidas repetidas en los mismos sujetos con ANOVA.

Cuando no se cumplen las condiciones necesarias para aplicar ANOVA, la alternativa no paramétrica que se utiliza de manera más habitual es la prueba de **Kruskal-Wallis** para muestra independientes y la prueba de **Friedman** para muestras apareadas, entre otras.

COMPARACIÓN ENTRE DOS GRUPOS DE DATOS CUALITATIVOS

Si deseamos comparar la frecuencia de presentación de una variable categórica, ya sea dicotómica o policotómica, en dos o más grupos, es decir, en el caso en que la variable de comparación sea una variable cualitativa, los datos se resumen o agrupan en las denominadas **tablas cruzadas o tablas de contingencia**. En la situación más sencilla, cuando comparamos la frecuencia de presentación de una variable dicotómica (padecer o no padecer demencia) en dos grupos, por ejemplo, hipertensos

y normotensos, la información se distribuye en una tabla que tiene cuatro celdas, en cada una de las cuales se disponen los sujetos que reúnen o no reúnen las condiciones de ambas variables. La prueba estadística que se utiliza para contrastar la hipótesis nula de independencia de ambas variables es la prueba **chi-cuadrado de Pearson** u otras similares. Cuando los sujetos están apareados por otras variables presuntamente confundentes, se suele utilizar la prueba exacta de **McNemar**. La prueba de chi-cuadrado se basa en que las diferencias existentes entre lo observado en nuestra muestra y lo que sería de esperar bajo la hipótesis nula de independencia de ambas variables (en nuestro ejemplo, la frecuencia de demencia no cambiaría por el hecho de ser o no hipertenso) sean o no suficientemente grandes como para rechazar la hipótesis de independencia de ambas variables.

EVALUACIÓN DE LAS RELACIONES ENTRE DOS VARIABLES CUANTITATIVAS EN UN GRUPO DE SUJETOS: CORRELACIÓN Y REGRESIÓN.

En otras ocasiones se trata de establecer si existe relación entre los valores de dos variables cuantitativas en un grupo de sujetos. Por ejemplo, la edad de una muestra de pacientes y las cifras de su tensión arterial. Si la relación entre ambas variables conlleva que una de ellas es la responsable de los valores que toma la otra, hablamos de **regresión**. Si ambas variables están relacionadas pero no podemos establecer cual es la responsable, hablamos de **correlación**. El hecho de establecer la relación entre dos variables continuas puede tener uno de tres objetivos: evaluar el grado de asociación entre ambas variables (mediante el coeficiente de correlación de Pearson o el de Spearman para el caso de que no se cumplan las condiciones de distribución normal), predecir el valor de una variable conociendo el valor de la otra (mediante los coeficientes de regresión) y, por último, se puede intentar establecer el **grado de acuerdo** entre los valores de dos variables cuantitativas, por ejemplo cuando se lleva a cabo la misma determinación de un parámetro analítico sanguíneo con dos métodos diferentes con el objetivo de sustituir uno más costoso por otro más sencillo o barato. Existen métodos para establecer la asociación entre variables categóricas que derivan de las anteriormente referidas tablas de contingencia, tal como el coeficiente kappa.

ANÁLISIS SIMULTÁNEO DE DIFERENTES VARIABLES PREDICTORAS: ANÁLISIS MULTIVARIANTE.

Resultados clínicos como los que acabamos de referir, ictus o demencia, normalmente no son debidos a una sólo causa. Es decir, en biología y en la clínica habitual estamos acostumbrados a ver que diferentes factores causales actúan juntos para producir una consecuencia. Frente al denominado **análisis univariante** (o bivariante, dado que en realidad se analizan dos variables, la variable predictora y la variable resultado), el **análisis multivariante** permite establecer la responsabilidad que tienen diferentes variables predictoras en el resultado clínico que nos interesa, ajustando o controlando por la influencia de cada una de ellas.

Como idea más general, el concepto de **modelización estadística** es el de un modelo matemático que relaciona dos o más variables y permite hacer una descripción aproximada de los datos observados. Los denominados **modelos lineales** son modelos estadísticos aditivos en los que la variable resultado puede ser explicada matemáticamente (el establecimiento de la causalidad va más allá del simple planteamiento estadístico) por diferentes variables predictoras, cada una de las cuales aportará cierta cantidad de “explicación” a la variable resultado. Lo que sucede es que aunque una variable resultado pueda presentar una asociación estadísticamente significativa con una o varias variables predictoras, la cantidad que estas variables explican de la variable resultado puede ser relativamente modesta. Esto nos ayuda a comprender el conocido hecho de que la existencia de una asociación estadística entre dos o más variables no significa que una de ellas quede explicada en su totalidad por la otra u otras variables, en la medida en que gran parte de la variabilidad de la primera podría justificarse por muchos otros factores cuyo conocimiento puede no estar a nuestro alcance.

El planteamiento que debe hacerse en una investigación clínica y en buena lógica también cuando leemos un artículo que nos la presente, es relacionar un efecto o resultado (la que hemos denominado variable dependiente o resultado) con una intervención o factor de exposición responsable (la que hemos denominado variable independiente o predictora). Pero en las demás circunstancias, es decir, en todas las demás características, ambos grupos, el que recibe la intervención o está expuesto y

el que no la recibe o no está libre de la exposición, deben ser iguales. Por ejemplo, si llevamos a cabo un estudio sobre tratamiento hipotensor para prevenir el desarrollo de demencia y los sujetos tratados son más jóvenes o las mujeres están más representadas, la menor tasa de déficit cognitivo puede estar en relación con los factores referidos más que con el efecto del tratamiento. Por ello, la edad y género de ambos grupos, tratados y no tratados, debe ser similar para evitar el efecto confundente de ambos factores. En los ensayos clínicos controlados, la primera tabla que se presenta en los resultados suele mostrar las características de ambos grupos de intervención, el tratado y el no tratado, al inicio del estudio, aclarando que no existen diferencias importantes entre ellos con respecto a los referidos factores presuntamente confundentes.

Para enfrentar el problema de las diferencias que puedan presentarse entre los grupos que se trata de comparar en un estudio se pueden adoptar dos estrategias. Por una parte, en el diseño del estudio, se puede **ser restrictivo** en los criterios de inclusión de los sujetos (seleccionar sólo aquellos con límites estrictos de edad o comorbilidades) o se puede proceder a **aparear** a los sujetos por las referidas características. Por otra parte, en el análisis de los resultados podemos "**estratificar**" según variables confundentes (por ejemplo analizar de forma separada el grupo de jóvenes y el de personas mayores) o podemos utilizar alguna de las técnicas de **análisis multivariantes** disponibles. Las técnicas multivariantes más utilizadas actualmente en la investigación clínica son las de regresión (lineal, logística y riesgo proporcional de Cox) (1,2,3,4,5), cuya pormenorización será motivo de otro artículo de esta serie.

BIBLIOGRAFÍA

- 1 .- Altman DG. Practical Statistics for Medical Research. London:Chapman and Hall, 1991.
- 2 .- Armitage P, Berry G. Statistical Methods in Medical Research. 3rd ed. Oxford:Blackwell Scientific Publications, 1994.
- 3 .- Carrasco JC, Hernán MA. Estadística multivariante en las ciencias de la vida. Madrid: Editorial Ciencia 3, S.L.1993.
- 4 .- Abraira Santos V, Pérez de Vargas Luque A. Métodos Multivariantes en Bioestadística. Madrid: Editorial Centro de Estudios Ramón Areces, S.A. 1996.

5 .- Field A. Discovering Statistics using SPSS for Windows. London: Sage Publications 2000.