

USO DE ESTADÍSTICA BÁSICA EN EL ÁREA DE EPIDEMIOLOGÍA EN GRANJAS PORCINAS

Villar, G¹ * Camacho-Rea, C²

¹ Apligén S.A de C.V, ² Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán

La estadística es una herramienta básica, la cual puede ser empleada para evaluar y comparar los diferentes aspectos productivos de una granja al igual que el estatus de salud existente en ésta. Por lo que el estudio de la distribución de enfermedades y sus determinantes dentro de una granja, pueden llevarse a cabo si se tienen conocimientos básicos de estadística.

La distribución de enfermedad se refiere a la carga que existirá de ésta en una explotación dependiendo de varios factores como la zona geográfica, el grado de tecnificación de la granja, la raza, la etapa productiva o el sexo, entre otros. Mientras que los determinantes de una enfermedad son los factores que la causan o se relacionan a ella. Por lo que, estudiar la distribución de una enfermedad es una descripción de la misma, mientras que estudiar los determinantes se refiere a la etiología de ésta.

Por otro lado, la incidencia de una enfermedad se puede definir como el número de casos nuevos dentro de un periodo de tiempo específico, pudiendo ser este de un día, un mes o un año. A este término se le relaciona el de prevalencia que es el número de casos existentes de una enfermedad en un momento determinado. Las características a investigar dentro de una población pueden ser Cualitativas y Cuantitativas.

- A) Cualitativas, es toda variable que por su naturaleza no se puede medir como sexo, raza, color etc. Por esta característica se evalúan a través de Estadística No Paramétrica con tablas de contingencia y se calculan razones de Odds o las de Riesgos con sus respectivos intervalos de confianza.
- B) Si la variable es Cuantitativa se usa estadística Paramétrica calculando intervalos de confianza para la media y análisis de varianza.

A) Evaluación de Variables Cualitativas.

Por lo general recopilamos información de la exposición a un agente etiológico y la presentación de una enfermedad. Para hacer esto, se calculan las medidas de asociación entre ambas, estas nos indican la fuerza de relación entre ellas. Las dos medidas de asociación que se usan con más frecuencia son el riesgo relativo o la razón de riesgo (RR) y la oportunidad relativa u Odds (OR). La decisión de calcular una RR u OR depende del diseño del estudio.

Riesgo Relativo o Razón de Riesgo (RR)

Es un indicador que compara el riesgo entre dos poblaciones una expuesta y otra no expuesta a algún factor de riesgo. Dimensiona la probabilidad que tiene un animal con ciertos atributos para contraer una enfermedad, esa probabilidad se denomina Riesgo (R) o sea el riesgo de contraer la enfermedad y el atributo sería el Factor de Riesgo (FR). Para dimensionar el *Riesgo Relativo (RR)* hay que comparar el R en el grupo expuesto al FR y un grupo no expuesto al FR. El cálculo del R y del RR se puede hacer usando una Tabla 2x2:

Status Enf			
Status FR	Enfermo	No Enfermo	Total
Expuesto	a (Exp, Enf)	b (Exp, NEnf)	a + b (Exp)
No Expuesto	c (NoExp, Enf)	d (NoExp, NoEnf)	c + d (N Exp)
Total	a + c (Enf)	b + d (No Enf)	n (a+b+c+d)

El R de la enfermedad en los expuestos es $a/(a + b)$ es lo mismo que comparar a los Expuestos que se enfermaron entre el total de individuos expuestos; en los No Expuestos sería la misma comparación pero en su Status $c/(c+d)$.

El RR de enfermedad entre ambos grupos, los Expuestos contra los No Expuestos está dado por el cociente de su comparación:

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a \cdot (c+d)}{c \cdot (a+b)}$$

Razón de ODSS o de Oportunidad Relativa (OR)

Por otro lado, cuando no se conoce cuál es el total de la población en riesgo, no se puede determinar el riesgo de enfermedad entre los animales expuestos y no expuestos. Esto significa que no se podrá usar la RR. En lugar de eso, se usa una medida indirecta del riesgo llamada *Razón de ODSS o de Oportunidad Relativa (OR)*, este indicador mide el número de veces que la respuesta ocurre (con probabilidad r) por cada vez que no ocurre (con una probabilidad 1-r):

$$Odds = r/(1-r)$$

Este método también mide a los animales Expuestos como los No Expuestos. Utilizando la tabla anterior de 2x2 se tendría que el Odss para los animales Expuestos estaría definido como a/b o lo que es lo mismo en nuestro ejemplo Enfermos/ No Enfermos y de igual manera sucedería para el caso de los No Expuestos a los factores de riesgo causantes de la enfermedad. Por lo que la Razón de Oportunidad Relativa (RO) compara que la enfermedad entre animales Expuestos y los No Expuestos está dada por la siguiente formula:

$$RO = \frac{a/b}{c/d} = \frac{a \cdot d}{b \cdot c}$$

En resumen, al comparar el Riesgo (R) y el Odds de una determinada enfermedad entre animales Expuestos y No Expuestos para cierto factor de riesgo, se tendría la siguiente interpretación:

RR	OR	Interpretación
>1	>1	Asociación (Factor confiere Riesgo).
=1	=1	No asociación (Factor No confiere Riesgo Ni Protección).
<1	<1	Asociación (Factor confiere Protección).

Para probar la hipótesis nula de que el RR o la OR es igual a UNO vs diferente, se puede realizar de dos formas:

- 1) Por Intervalo de Confianza en donde H_0 se rechaza si no contiene el UNO.

- 2) Realizando una prueba de Ji-cuadrada o una prueba exacta de Fisher.

1. Intervalos de confianza

El intervalo de confianza es un rango de valores (calculado de una muestra) en el cual se espera encontrar el verdadero valor del parámetro, con una probabilidad determinada.

Cuando calculamos un estimador (como R u Odds), o una medida de asociación (como RR o RO, a ese cálculo se le llama, Estimador Puntual.

El intervalo de confianza de un estimador puntual describe la precisión del cálculo. Representa un rango de valores a cada lado del estimador. Mientras más angosto es el intervalo de confianza, más preciso es el estimador puntual, normalmente este valor se ubica en el centro del intervalo de confianza. El intervalo de confianza más usado es el de 95%. Cuando se usa un intervalo de confianza de 95%, se concluye que el rango estimado tiene un 95% de posibilidades de contener el valor de la población verdadera. En epidemiología, la posibilidad de error del 5% es bien aceptada. Sin embargo, si desea una menor probabilidad, se podrá calcular un intervalo de confianza de 99%, con solo 1% de probabilidad de error. Esto tiene su costo, ya que con un intervalo de confianza de 99%, el rango calculado será más amplio que con uno de 95%. Idealmente, se desea un intervalo de confianza muy angosto, que pueda indicar que el cálculo es muy preciso, para lo cual se requiere una muestra más grande.

Análisis de datos Cualitativos

Una vez calculada una medida de asociación (RR o RO), y un intervalo de confianza para un rango de valores alrededor del estimador puntual, se deberá realizar una prueba estadística formal para determinar si los resultados son significativos.

2. Prueba de Ji-cuadrada

En epidemiología de campo las pruebas más empleadas son la Ji-cuadrada, como se mencionó anteriormente, un análisis común en epidemiología incluye variables dicotómicas o también llamadas binarias y se usa una tabla de 2x2. Para calcular el estadístico ji-cuadrada, se usan los valores observados y los valores esperados, empleando la siguiente fórmula:

$$[(\text{valor observado}-\text{esperado})^2/\text{esperado}]$$

para cada celda de la tabla, después se suman estos números para obtener el estadístico Ji-cuadrada.

En general, mientras más alto el valor de Ji-cuadrada, mayor es la probabilidad de que exista una diferencia significativa entre los dos grupos que se están comparando, sin embargo, se tendrá que observar el valor de **p** en una tabla de Ji-cuadrada.

Existen diferentes pruebas de ji-cuadrada. Cada una de ellas se usa en bajo diferentes circunstancias la más comúnmente usada es la de **Pearson, o la Ji-cuadrada no corregida**. De ahí que cuando se observa un resultado en el que simplemente aparece como Ji-cuadrada, es probable que se trate de la Ji-cuadrada de Pearson. Una regla general para usar Ji-cuadrado de Pearson es tener una muestra grande, más de 100 observaciones.

Sin embargo, si la muestra es menor a 30 observaciones o si una de las celdas en la tabla de 2x2 es menor a 5, se necesitará usar la **Prueba Exacta de Fisher**.

Asimismo, si se tienen datos apareados se deberá usar la prueba de **Mc Nemar** en vez de la Ji-cuadrada.

B) Evaluación de Variables Cuantitativas

La estadística inferencial abarca dos grandes campos denominados la *Estimación* y el de *Pruebas de hipótesis*. La estimación está dividida a su vez en estimación puntual y estimación por intervalos mientras que las pruebas de hipótesis constituyen la base para el diseño de experimentos más complicados como el Análisis de Varianza.

La estadística descriptiva organiza y resume datos obtenidos de una población, pero no permite emitir conclusiones de una muestra. Éstas se dividen en medidas de tendencia central que incluyen a la media, moda y mediana, y a las medidas de dispersión: varianza, desviación estándar, coeficiente de variación y media ponderada.

La media sirve para conocer el punto medio entre los valores mínimos y máximos de una medición; es un valor único y es simple de calcular pero puede verse afectada por los valores extremos.

La mediana es el valor que divide en dos partes iguales a un conjunto finito de valores arreglados u ordenados en el orden de su magnitud. La mediana no se afecta tan drásticamente por valores extremos como ocurre con la media.

La moda es el valor que ocurre con más frecuencia en un conjunto de valores.

La varianza mide la dispersión de los valores de un conjunto de observaciones con respecto a su media en unidades cuadradas.

La desviación estándar es la raíz cuadrada de la varianza y representa la dispersión de esos valores con respecto a su media en las mismas unidades que la variable original.

El error estándar es un ajuste de la desviación por el tamaño de muestra.

El coeficiente de variación expresa en porcentaje cuanto se dispersa la variable con respecto a su media, esto hace posible compararlo con otro conjunto de datos.

Estimación por Intervalo de Confianza

Los estimadores de una medida pueden ser por intervalo (construido por un rango limitado por dos valores) y sirve para reflejar con cierta confianza lo que ocurre realmente en una población, midiendo una muestra y no toda la población.

La ecuación estadística es:

$$\bar{x} \pm (s/\sqrt{n} \times Z_{\alpha/2})$$

Con esta ecuación el resultado obtenido es un rango que indica que con un alto porcentaje (95% comúnmente) de confianza se puede asegurar que si se dan las mismas condiciones se va a encontrar el resultado promedio dentro del rango estimado. Para que esta inferencia tenga valor es indispensable que las condiciones de los animales sobre los cuales se requiere predecir un futuro comportamiento sean similares a las condiciones de recolección de datos, sexo, tamaño, peso, granja, época año, estado sanitario, etc.

ANOVA o ANDEVA

Cuando se comparan variables continuas entre grupos de sujetos de estudio, se usa un análisis que otorga un valor **p**. Sin embargo, en lugar de usar pruebas de Ji-cuadrada como se hizo para datos categóricos, se usa una prueba t (para comparar 2 grupos) o una prueba f (para comparar 3 o más grupos).

El término Análisis de varianza se aplica al procedimiento estadístico desarrollado por Fisher (1918) para partir la varianza de una variable aleatoria en porciones independientes (ortogonales). La cual es resultado de la aplicación de los tratamientos y del efecto del error experimental.

Análisis de Varianza (ANOVA o ANDEVA)				
Fuente Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F calculada
Tratamientos	tx - 1	$\sum ni (\bar{Y}_i - \bar{Y}_{..})^2$	$\frac{\sum ni (\bar{Y}_i - \bar{Y}_{..})^2}{tx - 1}$	$\frac{CMTx}{CM Error}$
Error	n - tx	$\sum \sum (Y_{ij} - \bar{Y}_{i.})^2$	$\frac{\sum \sum (Y_{ij} - \bar{Y}_{i.})^2}{n - tx}$	
Total	n - 1	$\sum \sum (Y_{ij} - \bar{Y}_{..})^2$		

El modelo simple es el modelo más sencillo de ANDEVA, ya que solo considera el efecto de los tratamientos y del error experimental, sin embargo, aun cuando es sencillo puede ser de gran utilidad, si previamente se homogeneizaron los grupos en estudio.

El resultado final del ANDEVA es una F calculada, la cual se debe comparar con una F de tablas (P=0.05, gltx; gle), si el valor obtenido es superior al de tablas se rechaza la hipótesis nula (Ho), la cual indica: Tx1=Tx2=Tx3, por lo que se puede concluir que al menos uno de los tratamientos es diferente. Teniendo que utilizar alguna prueba de comparación de medias para determinar cuál es el mejor.

El método propuesto por Tukey es la herramientas más utilizadas por los investigadores para probar diferencias entre tratamientos. Son comparaciones múltiples a posteriori por lo que se realizan una vez obtenido el ANDEVA, el valor q obtenido se debe comparar con q valores críticos; q (tx, n-tx, α)

El Análisis de varianzas (ANDEVA) es la herramienta de uso para obtener resultados confiables y estadísticamente diferentes en caso de que algún tratamiento sea mejor que otro, es importante el diseño estadístico apropiado de los estudios, ya que el diseño de la “Prueba” controla todos los factores que pueden alterar los resultados, dejando actuar plenamente a los tratamientos. Con esto se disminuye la “Variación” de las medidas de respuesta procedente de factores externos al tratamiento. Existen varios diseños estadísticos, el Diseño Completamente al Azar, el de Bloques Aleatorios, Covariable,

Factorial, Cuadrado Latino, Parcelas Divididas, etc. Lo importante es conocer que variables hay que controlar y medir para aplicar el diseño apropiado antes de iniciar el experimento.

Conclusión

El análisis de datos epidemiológicos es vital para implicar exposiciones en la causalidad de enfermedades. Solo se han expuesto las técnicas más comunes de métodos estadísticos relacionados a epidemiología entre los que destacan los intervalos de confianza, pruebas de Ji-cuadrada, ANOVA y sus interpretaciones.

Bibliografía

Anderson M. and Nelson A. 2010, vol 3 num 6. Simple Interactive Statistical Analysis, Focus on Field Epidemiology, The North Carolina Institute for Public Health.

Dewey, C. y Marsh, W. 1996. Theory and Conduct of Experimental Field Trials. Pre-Conference Workshop. Allen D Leman Swine Conference, St Paul MN.

Herrero, H.J. y Barrera, S.A. 1999. Manual de Procedimientos Estadísticos de experimentos pecuarios, Instituto de Enseñanza e Investigación en Ciencias Agrícolas, Colegio Posgraduados Montecillos Texcoco.

Giesecke. 2002. J. Modern Infectious Disease Epidemiology. 2nd Ed. Londres. Arnold Publishing.