56. ¿Qué significan los llamados *Grados de libertad*"[1]?

El concepto de grados de libertad es de mucha importancia en la estadística moderna, no obstante que los libros de texto, generalmente no dejan satisfechos a sus lectores con los intentos que hacen para explicar este concepto. Los matemáticos que leen los artículos originales y que manejan las ideas de espacios vectoriales n-dimensionales, van con ventaja en la comprensión de las ideas asociadas a los grados de libertad, pues a decir verdad, no es un concepto fácil de transmitir si la ayuda de la geometría vectorial.

Walter (1949), hizo un muy buen esfuerzo para dejar en claro el significado de los llamados "grados de libertad", lo hizo inicialmente de una manera muy intuitiva, para un publico que no tiene los conocimientos matemáticos mencionados, pero luego, avanza también en la abstracción para que aquellos con más bases geométricas, pudieran reforzar las preliminares ideas intuitivas. Estas ideas son las que hemos adaptado para nuestros lectores.

Este concepto de la estadística moderna, no se encuentra referenciado en la literatura científica antes del articulo de "Student" (1908), siendo Ronald Fisher (1915) el primero que hace referencia en forma explícita al mismo, en su artículo sobre la distribución del coeficiente de correlación, no obstante que Gauss y sus astrónomos asociados estaban familiarizados con la idea esencial de grados de libertad, pues en 1826, en su trabajo clásico "*Theory of the combination of Observations*" en la cual presenta la generalización del método de los mínimos cuadrados, el da fé de la necesidad de incluir la idea de "grados de libertad" (sin llamarlos de esta manera), cuando lo plantea no solo con palabras sino también en las fórmulas. Gauss expresa que el número de observaciones será disminuido por el número de parámetros desconocidos estimados con los datos, cuando sea usado como divisor en el estimado del error estándar de un conjunto de observaciones.

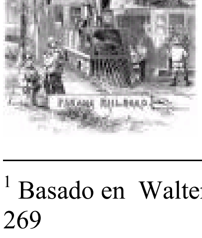
Pero, ¿que son los llamados "grados de libertad?

Consideremos, en principio la libertad de movimiento que poseen algunos objetos de la cotidianidad, los cuales serán tratados como si fueran solo puntos sin tamaño.

Consideremos por ejemplo, una partícula de aceite que se desliza por una tubería o una cuenta que se mueve ensartada en un alambre, tienen un solo grado de libertad pues su movimiento solo es posible en una sola dimensión que está definida por la ruta del tubo o del alambre respectivamente, sin importar que tan complicada sea la ruta. (es decir, no tiene que ser una recta).

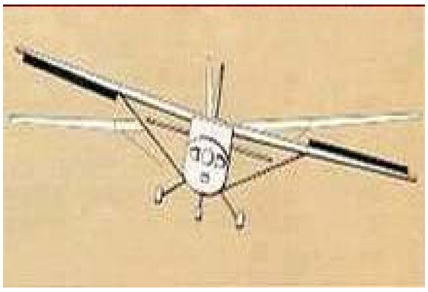
Un mosquito volando libremente en un espacio tridimensional, tiene tres grados de libertad.

Considerado como un punto en movimiento, un tren puede moverse solo hacia delante o hacia atrás sobre una ruta unidimensional (las vías del ferrocarril) que permanecen sobre un espacio bidimensional que es la superficie de la tierra, la cual se encuentra en el espacio tridimensional del universo. Una sola coordenada, distancia de algún origen, es suficiente para localizar el tren en cualquier momento dado, es decir que el tren tiene un solo grado de libertad.



[1] Basado en Walter M. Helen (1940). "Degree Freedom". Journal of Educational Psychology. 31(14), 253-269

Un automóvil se mueve sobre una superficie bidimensional, la cual es una porción de un espacio tridimensional. En un momento dado la posición del auto es determina por dos coordenadas, por lo tanto, el movimiento del automóvil tiene dos grados de libertad.



En el mismo sentido un avión, tiene tres grados de libertad en el universo del espacio y puede ser localizado, con tres coordenadas, que pueden ser longitud, latitud y altitud. O puede ser altitud, distancia horizontal desde algún origen y un ángulo. O puede ser por la distancia directa desde algún origen y dos ángulos.

Ahora, hagamos un poco de abstracción para intentar establecer un paralelismo entre las situaciones cotidianas sobre movimiento de objetos y los problemas del muestreo.

Si a usted le pide que elija un par de números (x,y) al azar, usted tiene libertad completa de elección de los dos números, tiene dos grados de libertad. Las dos coordenadas, pueden ser representadas por un punto localizado en el plano XY, el cual es un espacio bidimensional. El punto es libre de moverse horizontal y verticalmente, hay dos variables y el punto tiene dos grados de libertad.

Ahora supongamos que nos ponen a elegir un par de números cuya suma es 7. Es claro que solo un número puede elegirse libremente, pues el segundo queda fijado, una vez se conozca el primero. Aunque aquí hay dos variables, en esta situación solo una es independiente, por lo que el número de grados de libertad se reduce de dos a solo uno, por la imposición de la restricción $x+y=7$, el punto ahora es libre de moverse en el plano XY pero restringido a permanecer sobre la recta $x+y=7$. Esta línea es un espacio unidimensional que está contenido en el espacio bidimensional original.

Supongamos ahora que nos piden escoger un par de números, tal que la suma de sus cuadrados sea 25. De nuevo, es claro que solo somos libres de escoger solo uno de los números, pues una vez seleccionemos el primero el otro queda fijado. El punto en cuestión, representado por un par de números (x,y), permanece en una circunferencia de radio 5 y con centro en el origen. La circunferencia es un espacio unidimensional, contenido en un plano bidimensional. El punto solo puede moverse hacia delante o hacia atrás a lo largo de la circunferencia y por eso solo tiene un grado de libertad. Hay dos números escogidos (N=2) sujetos a una restricción (r=1) y el número resultante de grados de libertad es $N-r=2-1=1$.

Supongamos ahora que imponemos simultáneamente las dos condiciones $x+y=7$ y también $x^2+y^2=25$. Si nosotros resolvemos algebraicamente estas ecuaciones, obtenemos que solo son posibles dos soluciones $x=3$, $y=4$ o $x=4$, $y=3$. Ninguna variable puede escogerse a voluntad. El punto está restringido por la ecuación $x+y=7$ a moverse a lo largo de una recta y además está restringido por la ecuación $x^2+y^2=25$, a moverse a lo largo de una circunferencia. Las dos restricciones simultaneas lo confinan a la intersección entre la recta y la circunferencia, dejándolo sin libertad de movimiento, es decir sin grados de libertad. Aquí, son dos los números a elegir (N=2) y r=2. El número de grados de libertad es $N-r=2-2=0$.

Consideremos ahora un punto (x,y,z) en un espacio tridimensional (N=3). Si no hay restricciones sobre sus coordenadas, el punto puede moverse libremente por todo el espacio, en todas las direcciones, tiene tres grados de libertad. Todas las variables son independientes. Si ahora le colocamos la restricción $x+y+z=c$, donde c es una constante, dos de los tres números pueden ser elegidos libremente, porque una vez seleccionados el tercero queda fijado, es decir, solo dos observaciones son independientes. Por ejemplo $x-y-z=10$, si nosotros escogemos x=7 y y=9, forzamos a que z tome el valor -12. La ecuación $x-y-z=c$ es la ecuación de un plano, es decir un espacio bidimensional que corta el espacio tridimensional original y un punto que pertenece a dicho espacio tiene solo dos grados de libertad, $(N-r=3-1=2)$. Si a las coordenadas del punto (x,y,z), se les exige el cumplimiento de la condición $x^2+y^2+z^2=k$, el punto será forzado a moverse en la superficie de una esfera que tiene centro en el origen y radio $\sqrt{k}$. La superficie de una esfera es un espacio bidimensional. $(N=3, r=2, N-r=3-1=2)$.

Si imponemos ambas condiciones al punto (x,y,z), es decir que pertenezca al plano y a la esfera, entonces el punto solo podrá moverse en la intersección del plano y la esfera que es una circunferencia, la cual es una figura unidimensional, en un espacio original de tres dimensiones. $(N-r=3-2=1)$. Si consideramos algebraicamente la solución de las dos ecuaciones en tres variables, al despejar una de ellas y reemplazarla en la otra nos resulta una sola ecuación con dos variables, quedando de nuevo en el caso de poder elegir el valor para una de ellas y quedando fijado así el valor para la otra, es decir, tenemos un solo grado de libertad.

Estas ideas pueden ser generalizadas para N más grande que 3 y esta generalización necesariamente corresponderá a una abstracción. Cualquier conjunto de N números, determinan un solo punto en espacio N-dimensional, cada número proporcionando una coordenada para el punto. Si no se impone ninguna restricción, cada número es libre de variar independientemente de los otros y por lo tanto el número de grados de libertad es N. Cada relación necesaria impuesta sobre ellos, reduce el número de grados de libertad en 1. Cualquier ecuación de primer grado, que conecte las N variables mediante la ecuación, es un espacio de (N-1) dimensiones. Si por ejemplo, consideramos solamente los puntos cuya suma de coordenadas es una constante, $\sum x_i = c$, hemos limitado el punto a un espacio de N-1 dimensiones. Si nosotros consideramos solo los puntos que cumplan con $\sum (x_i - M)^2 = k$, que corresponde a la superficie de una "hiperesfera" con centro en el origen y radio $\sqrt{k}$. Esta superficie es un espacio de dimensión (N-r) dentro de un espacio original de dimensión N y por lo tanto el número de grados de libertad debería ser N-r.

Abordemos ahora la situación que corresponde a la problemática de la estadística.

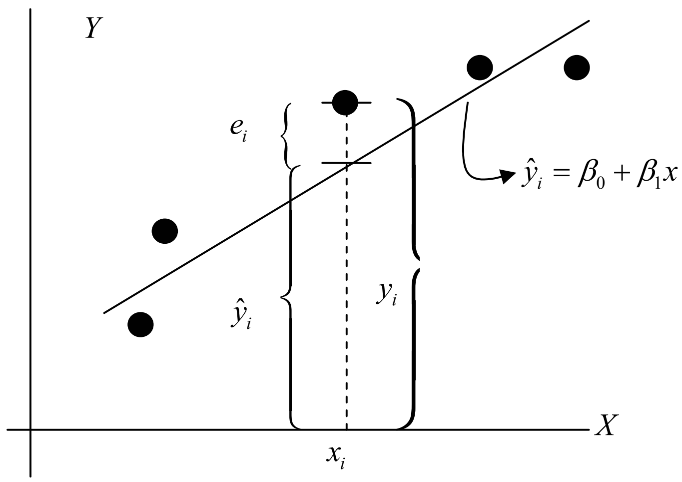
Una muestra de tamaño N, puede ser representada por un punto $(X_1, X_2, X_3, \ldots X_N)$ en un espacio N-dimensional, cuyo origen tomado como la verdadera media poblacional μ. (vector)

Definamos $x_1 = X_1 - \mu$, $x_2 = X_2 - \mu$, $x_3 = X_3 - \mu$, etc. Donde $X_1, X_2, X_3, \ldots X_N$ son las medidas crudas (originales) que se realizaron a los N individuos de la muestra. Sean $\bar{X}$ y S, la media y la desviación estándar para los N individuos de la muestra. Cualquier conjunto de observaciones determina un solo punto. Este punto tiene N grados de libertad, siempre y cuando no sea impuesta ninguna restricción a sus coordenadas.

Todas las muestras con la misma media $\bar{X}$, serán representadas por los puntos que pertenecen al "hiperplano", $\sum X_i = N\bar{X}$, que será un espacio de (N-1) dimensiones.

Veamos ahora que ocurre en el caso de querer ajustar una línea recta al conjunto de datos, usando el método de los Mínimos Cuadrados, que consiste en elegir de todas las posibles rectas aquella que tenga la menor suma de cuadrados de los residuos e_i. Se dirá que el modelo hallado ha sido ajustado por Mínimos Cuadrados.

La familia de modelos a considerar es de la forma: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$



$e_i = y_i - \hat{y}_i$

Predicción en X=xi

Si llamamos SCE a la suma de los cuadrados de los residuos e_i

$$SCE = \sum e_i^2 = \sum_{i=1}^{5} [y_i - \hat{y}_i]^2 = \sum_{i=1}^{5} [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad (4)$$

Observemos que en la expresión (4) los datos (x_i, y_i) son conocidos. Por lo tanto lo único que puede hacer cambiar (disminuir) la SCE, son los valores de β_0, β_1. Es decir que SCE es una función de β_0, β_1.

$$SCE(\beta_0, \beta_1) = \sum_{i=1}^{5}\left[y_i - (\beta_0 + \beta_1 x_i) \right]^2 \quad (5)$$

Es función de la recta que escojamos

Entre todas las posibles rectas, queremos aquella con parámetros $\hat{\beta}_0, \hat{\beta}_1$ que haga que $SCE(\hat{\beta}_0, \hat{\beta}_1)$ sea la menos posible.

Para esto se realiza un proceso de optimización matemática, obteniéndose que dichos valores $\hat{\beta}_0, \hat{\beta}_1$ que optimizan SCE, deben cumplir con las siguientes restricciones:

- $e_1 + e_2 + e_3 + ...e_{n-1} + e_n = 0$ (5)
- $e_1 x_1 + e_2 x_2 + e_3 x_3 + ...e_{n-1} x_{n-1} + e_n x_n = 0$ (6)

Con la restricción (5), estamos obligando a pasar la recta por el punto $(\bar{x}, \bar{y})$, es decir que la mejor recta siempre debe pasar por ese punto, lo cual disminuye los grados de libertad del error, pues esto significa que la suma de los residuos e_i de los puntos que se encuentran por encima de la recta debe ser igual a la suma de los residuos e_i de los puntos que se encuentran por debajo. Conocidos (n-1) residuos, queda determinado el restante. Con esta restricción el error pierde un grado de libertad. Hasta aquí podemos imaginarnos que las posibilidades se restringen a un haz de rectas que pasan por $(\bar{x}, \bar{y})$, de todas ellas escogeremos aquella que cumpla con la restricción (6) y es cuando el error pierde otro grado de libertad, quedando con (n-2) grados de libertad.

En este caso el cuadrado medio del error CME, podrá calcularse como:

$$CME = \frac{SCE}{n-2}$$

Esta situación puede generalizarse. El proceso de minimizar la SCE, resultará en un número de restricciones (ecuaciones normales) igual al número de parámetros en el modelo.

El modelo de regresión ordinario

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ Supone generalmente que el error ε_i es una variable aleatoria con distribución normal, que tiene media cero y que además tiene una varianza constante σ^2 (que no depende de X).

Es conveniente saber que: $CME = \dfrac{SCE}{n-2}$ es un estimador insesgado para σ^2 y en general, para un modelo de regresión múltiple con "p" parámetros (p-1) variables predictoras, el número de grados de libertad del error es (n-p) y por lo tanto $CME = \frac{SCE}{n-p}$ es también un estimador insesgado para la varianza σ^2 del error.

Esta propiedad del insesgamiento, hace posible la construcción de importantes contrastes en estadística como el famoso F de Snedeccor, para Análisis de la Varianza