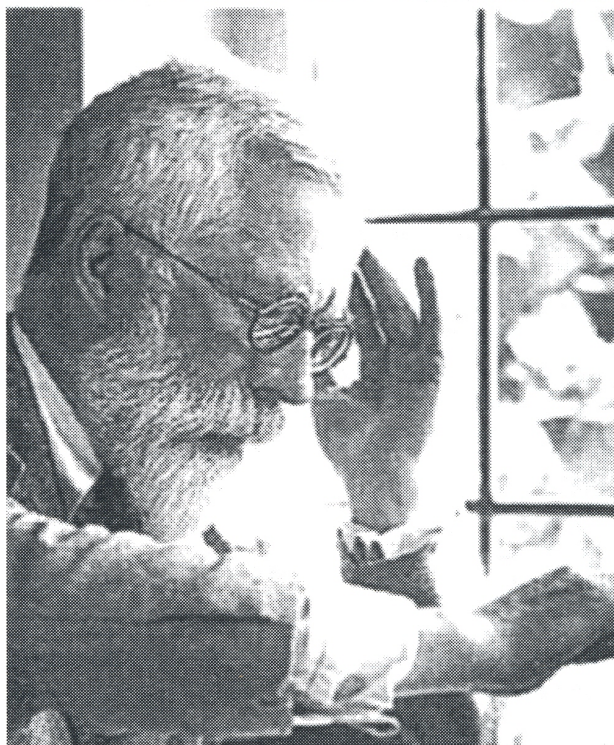


LOS GRADOS DE LIBERTAD EN ESTADÍSTICA: UNA APROXIMACIÓN PEDAGÓGICA



Sir Ronald Aylmer Fisher

Roberto Behar Gutiérrez, Javier Olaya Ochoa

*Profesores Titulares, Escuela de Ingeniería Industrial y Estadística,
Universidad del Valle, Cali, Colombia.*

Contactos: olaya@univalle.edu.co,

Pere Grima Cintas

*Profesor Titular, Departamento de Investigación de Operaciones y Estadística,
Universidad Politécnica de Cataluña, Barcelona, España*

Resumen

En ausencia de una definición formal del concepto de grados de libertad en los textos, este trabajo presenta tres aproximaciones al significado, que se denotan como Probabilista, Geométrica y Estadística. Las tres maneras de intentar una definición no son en lo absoluto excluyentes si no más bien complementarias.

Palabras Clave: Dimensión del análisis, Distribución Chi-cuadrado, Ecuaciones de regresión.

1. Introducción

Resulta un tanto difícil imaginarse los métodos estadísticos sin el concepto de grados de libertad. Sin embargo, los libros de texto no suelen entrar en él con mucho detalle —como se verá un poco más adelante— quizás porque el hacerlo podría implicar meterse en el terreno de las transformaciones lineales, lo cual podría complicar las cosas para muchos usuarios de los métodos estadísticos. Y por otra parte podría apartarse de los objetivos que se pretenden en un curso de estadística general. Existen varias aproximaciones posibles al intento de presentar esta noción tan importante. A continuación se presentan algunas, sin la pretensión de ser exhaustivos. En algunas, la intuición será la herramienta principal. En otras, no tanto. Pero el propósito es explorar este concepto y ofrecer varias ideas. El lector escogerá aquella que le ayude a comprenderlo mejor.

2. Acercamiento probabilista

Cuando se trata de definir los grados de libertad, algunos autores no ven la necesidad de entrar en explicaciones de significado. Para Dudewicz y Mishra (1988), por ejemplo, basta con indicar que se trata de una cantidad presente en la ecuación de la función de densidad de una variable que se distribuye Chi-cuadrado. En este contexto, podría pensarse que se trata de un parámetro de la distribución, más o menos como λ y α para la distribución Gamma. Se dice entonces (Ross, 2002) que una variable sigue una distribución Gamma con parámetros λ y α . O que una variable sigue una distribución Chi-cuadrado con V grados de libertad. De hecho una distribución Chi-cuadrado se presenta como un caso particular de una distribución Gamma en la cual $\lambda = V/2$ y $\alpha = 1/2$.

Otros autores ni siquiera consideran necesario mencionar el punto. Para Lehmann y Casella (1998), los grados de libertad sencillamente están allí. Para ellos la distribución Chi-cuadrado es una distribución de un solo parámetro perteneciente a la familia exponencial (Tabla 5.1, Pág. 25). Y la expresión "grados de libertad" aparece luego para mencionar simplemente que la variable χ_f^2 se distribuye Chi-cuadrado con f grados de libertad (Pág. 31).

De igual manera, para Mood, Graybill y Boes (1976) los grados de libertad se definen como un número entero positivo V , el cual es un parámetro de la distribución Chi-cuadrado. Nótese que aquí el valor V se reconoce explícitamente como un parámetro de una distribución de probabilidad. La discusión posterior introduce, en el capítulo sobre distribuciones en el muestreo, tres variables de importancia en aplicaciones estadísticas:

1. Si $X_1, X_2, \dots, X_n$ son n variables aleatorias independientes que se distribuyen Normal con media μ y varianza σ^2 , entonces la variable

$$U = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} = \sum_{i=1}^n Z_i^2$$

se distribuye Chi-cuadrado con n grados de libertad.

2. Si U se distribuye Chi-cuadrado con n grados de libertad y V se distribuye Chi-cuadrado con m grados

de libertad, entonces la variable $L = \frac{U/n}{V/m}$ se distri-

buye F con n grados de libertad en el numerador y m grados de libertad en el denominador.

3. Si Z se distribuye Normal con media 0 y varianza 1 y U se distribuye Chi-cuadrado con n grados de liber-

tad, entonces la variable $T = \frac{Z}{\sqrt{U/n}}$ se distribuye t de Student con n grados de libertad.

Obsérvese la presencia decisiva de la distribución Normal. Y también que en todo caso los grados de libertad se derivan de una o más variables Chi-cuadrado. Posiblemente por esta razón tantos autores se limitan a proveer como única referencia a los grados de libertad su condición de parte integral de la densidad Chi-cuadrado.

En estos casos la presentación de los grados de libertad está muy asociada a la distribución de una variable aleatoria. Podría llamarse a esta presentación la aproximación probabilista. Desde esta aproximación,

todo queda dicho a partir de la definición de la densidad Chi-cuadrado y nada más pareciera necesario. Desde una óptica absolutamente pragmática podría pensarse que tienen razón, aún si la "utilidad práctica" de la definición queda en entredicho.

3. Una aproximación desde la modelación de ecuaciones de regresión

Entre los autores que escriben sobre este tema de la modelación de ecuaciones de regresión, hay quienes prefieren detenerse a buscar en los lectores algún grado de comprensión del sentido del término. Gujarati (1999) introduce un pie de página para explicar: "El término número de grados de libertad significa el número total de observaciones en la muestra ($= n$) menos el número de restricciones (lineales) independientes o de restricciones puestas en ellas." (Pág. 69, la parte en negrilla es del autor original). Aquí se ha producido un cambio radical de presentación, posiblemente dada por el tema que se está tratando: el texto de Gujarati se refiere a los modelos lineales en Economía. En la misma línea, pero en este caso en el cuerpo del libro, y no en un pie de página, Draper y Smith (1998) presentan los grados de libertad como un número asociado con toda suma de cuadrados. Según los autores, este número indica cuántas piezas independientes de información, que consideran los n números independientes $Y_1, Y_2, \dots, Y_n$, se necesitan para calcular la suma de cuadrados. Por ejemplo, la suma de cuadrados de las desviaciones alrededor de la media requiere $(n-1)$ piezas independientes de información [de los números $(Y_1 - \bar{Y}), (Y_2 - \bar{Y}), \dots, (Y_n - \bar{Y})$ solamente $(n-1)$ son independientes dado que la suma de estos números es cero por definición de la media]. En este caso el cambio de lenguaje es evidente. Podría llamarse a este tipo de presentación la aproximación heurística. No pareciera necesario mencionar que Draper y Smith escriben en el marco de los modelos lineales.

Ahora bien, no se puede extender la idea de que los autores de textos sobre modelos lineales comparten la aproximación heurística para contribuir a aclarar el concepto. Algunos textos muy avanzados sobre modelos lineales, como Graybill (1976) y Seber (1977), y otros poco avanzados, como Montgomery, Peck y

Vining (2002), no se detienen en aclaración alguna. De hecho Montgomery menciona los grados de libertad por primera vez en el cuerpo de una salida de computadora, sin miramientos ni explicaciones. Graybill es un poco más formal y, tal como Ross o Mood, Graybill y Boes prefieren definir la distribución Chi-cuadrado con V grados de libertad como una suma de V variables aleatorias independientes Normales estándar al cuadrado. Esta es una definición formal de una Chi-cuadrado. Pero ciertamente no lo es de los grados de libertad.

Véase ahora que ocurre en el caso de querer ajustar una línea recta a un conjunto de puntos usando el método de los mínimos cuadrados, que consiste en elegir de todas las posibles rectas aquella que haga menor la suma de cuadrados de los residuos e_i (distancia de los puntos a la recta ajustada).

La familia de modelos a considerar es de la forma: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ y de entre todas las posibles rectas, se busca aquella con coeficientes b_0, b_1 que haga mínima la suma de los cuadrados de los residuos. Para esto se realiza un proceso de optimización matemática, obteniéndose que dichos valores b_0, b_1 deben cumplir con las siguientes restricciones para los residuos e_j

$$e_1 + e_2 + e_3 + \dots + e_{n-1} + e_n = 0$$

$$y \quad e_1 x_1 + e_2 x_2 + e_3 x_3 + \dots + e_{n-1} x_{n-1} + e_n x_n = 0$$

La primera restricción quita un grado de libertad a los residuos, ya que conocidos cualesquiera $(n-1)$ queda unívocamente definido el enésimo. Además, esto implica que la recta debe pasar por el punto $(\bar{x}, \bar{y})$

Así que las posibilidades se restringen a un haz de rectas que pasan por $(\bar{x}, \bar{y})$, pero de todas ellas se escoge aquella que cumpla con la segunda restricción, con lo cual el error pierde otro grado de libertad, quedando con $(n-2)$.

Es conveniente saber que $SCR/(n-2)$ es un estimador insesgado de la varianza del error σ^2 y que, en general, para un modelo de regresión múltiple con p

parámetros ((p-1) variables predictoras), el número de grados de libertad de los residuos es (n-p) y por lo tanto $SCR/(n-p)$ es también un estimador insesgado para la varianza del error.

4. Visión geométrica

La importancia de la distribución Normal se realiza particularmente en Cramer (1953)¹, quien indica que en el estudio de las distribuciones en el muestreo referentes a variables cuya distribución sea Normal, resulta útil una transformación lineal sugerida por Fisher (1925, citado por Cramer). Lo que sigue es una versión libre de lo expuesto por Cramer (Pág. 436-438). Asíumase que se dispone de n variables aleatorias independientes $X_1, X_2, \dots, X_n$, cada una de las cuales es Normal con media 0 y varianza σ^2 , es decir, lo que se denomina comúnmente una muestra aleatoria tomada de una población Normal $(0, \sigma^2)$. Considérese la transformación ortogonal

$$Y_i = c_{i1}X_1 + c_{i2}X_2 + \dots + c_{in}X_n, \text{ con } i=1, 2, \dots, n, (1)$$

que substituye las variables X_i originales por las nuevas variables Y_i . Se sigue que $E[Y_i]=0$ y que

$$E[Y_i Y_k] = \sigma^2 \sum_{j=1}^n c_{ij} c_{jk} = \begin{cases} \sigma^2 & \text{para } i=k \\ 0 & \text{para } i \neq k \end{cases}$$

Así que las nuevas variables Y_i , cada una de las cuales se distribuye Normal, tienen la misma distribución Normal $(0, \sigma^2)$. Y además son no-correlacionadas, por lo que además son independientes.

El significado geométrico de este resultado es que la transformación ortogonal que se ha introducido produce una rotación del sistema de coordenadas alrededor del origen, de tal manera que la distribución Normal en R_n que aquí se considera (cuya media es 0 y cuya varianza es $\sigma^2 I$), es invariante con respecto a esta transformación. Podría llamarse a este punto de vista la *visión geométrica*.

Nótese que, sin pérdida de generalidad, es posible asumir que $\sigma^2 = 1$.

Si se contara únicamente con un cierto número $p < n$ de funciones lineales $Y_1, Y_2, \dots, Y_p$, tales que los coeficientes c_{ik} utilizados en la definición de las Y_i satisfacen las condiciones de ortogonalidad para $i=1, 2, \dots, p$ y $k=1, 2, \dots, p$, siempre será posible encontrar otras (n-p) filas de coeficientes c_{ik} con $i=p+1, p+2, \dots, n$, de tal manera que la matriz completa de coeficientes sea ortogonal.

Considérese la forma cuadrática

$$Q(X_1, X_2, \dots, X_n) = \sum_{i=1}^n X_i^2 - Y_1^2 - Y_2^2 - \dots - Y_p^2$$

Si se utiliza la transformación ortogonal (1), se encuentra que $\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2$, por lo que Q es en realidad $Q = \sum_{i=1}^n Y_i^2$. Y esta es una suma de (n-p) variables aleatorias que se distribuyen Normal $(0,1)$ elevadas al cuadrado, la cual, siguiendo a Mood, se distribuye Chi-cuadrado con (n-p) grados de libertad.

El número (n - p) es el rango de la forma cuadrática Q, que es el menor número de variables independientes a que puede reducirse la forma cuadrática Q mediante una transformación lineal no-singular. Cramer señala entonces que es a este número al que Fisher bautizó como grados de libertad (del problema, o de la distribución de la variable aleatoria asociada al problema).

El siguiente ejemplo podría ayudar a ver la importancia de esta argumentación y la de su importancia en Estadística: seguramente las primeras reflexiones sobre la idea de grados de libertad en el contexto de la estadística surgen cuando se intenta comprender por qué para calcular la varianza muestral se recomienda dividir la suma de las desviaciones con respecto a la media entre (n-1), en lugar de dividir entre n, siendo este es el número total de datos que compone la muestra, como se ve en la ecuación siguiente.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Sin embargo, se sabe que la suma de las desviaciones de los datos con respecto a su media es siempre nula, es decir que: $\sum_{i=1}^n (x_i - \bar{x}) = 0$

Esto significa que de los n sumandos que tiene el numerador de la varianza solo $(n-1)$ son independientes, es decir, sólo podemos "inventarnos" $(n-1)$ si queremos que los n tengan la media definida $\bar{x}$.

Por esta razón, aunque la varianza se ha calculado a partir de n datos, es importante el hecho de que estos tengan sólo $(n-1)$ grados de libertad, ya que para obtener un estimador insesgado de la varianza poblacional, debemos dividir la suma de cuadrados por el número de sus grados de libertad, y no por el número de datos. La asociación de este ejemplo con la exposición de Cramer y con su relación con la Estadística ya mencionadas en este texto es evidente.

La siguiente presentación de la visión geométrica aparece desarrollada en Behar y Grima (2004). Supóngase que se pide elegir un par de números (x, y) al azar. En este caso se tiene completa libertad para la elección de estos dos números, los cuales pueden ser representados como un punto localizado en el plano XY , que es un espacio bidimensional. Como el punto es libre de moverse en ambas direcciones, entonces se dice que tiene dos grados de libertad.

Pero si se restringe la elección de estos dos números a un par de números cuya suma es 7, se sigue que sólo puede elegirse libremente un número, pues el segundo queda fijado una vez se conozca el primero. Aunque aquí también hay dos variables, sólo una es independiente, por lo que el número de grados de libertad se reduce de dos a uno. 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.

Supóngase ahora se debe escoger un par de números tales que la suma de sus cuadrados sea 25. En este caso, como en el anterior, sólo se está en libertad de elegir uno de los números, puesto que una vez seleccionado el primero el otro queda inmediatamente fijado. El par de números a elegir 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 sobre la circunferencia y por eso tiene un solo grado de libertad. Hay dos números escogidos

$(n = 2)$ sujetos a una restricción $(r = 1)$ y el número resultante de grados de libertad es $(n-r)=2-1=1$.

Ahora, si se imponen simultáneamente las dos condiciones $x+y=7$ y $x^2+y^2=25$ y se resuelven simultáneamente estas ecuaciones, se encuentra que sólo son posibles dos soluciones: o bien $x = 3, y = 4$, o bien $x = 4, y = 3$. 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 sobre una circunferencia. Las dos restricciones simultáneas 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 dos las restricciones impuestas $(r=2)$. El número de grados de libertad es $(n-r)=2-2=0$.

Estas ideas pueden ser generalizadas para n mayor que 2, de forma que cualquier conjunto de n números determina un punto en el espacio n -dimensional. 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 es un espacio de $(n-1)$ dimensiones. Si por ejemplo se consideran únicamente aquellos puntos cuya suma de coordenadas es una constante, por ejemplo $\sum x_i = c$, se ha limitado el punto a un espacio de $(n-1)$ dimensiones. Si se consideran solamente los puntos que cumplan con $\sum (x_i - m)^2 = k^2$, que corresponde a la superficie de una "hiperesfera" con centro en el origen y radio k , esta superficie es un espacio de dimensión $(n-1)$ dentro de un espacio original de dimensión n y por lo tanto el número de grados de libertad debería ser $(n-1)$.

También una muestra de tamaño n , puede ser representada por un punto $(X_1, X_2, X_3, \dots, X_n)$ en un espacio n -dimensional con n grados de libertad, cuando no se impone ninguna restricción a sus coordenadas. Ahora bien, si forzamos que su media deba ser $\bar{x}$, ya tenemos una restricción y, por tanto, $(n-1)$ grados de libertad. Todas las muestras de tamaño n con la misma media $\bar{x}$, estarán representadas por los puntos que pertenecen al "hiperplano", $\sum x_i = n\bar{x}$, que será un espacio de $(n-1)$ dimensiones.

5. Una propuesta de generalización del concepto de grados de libertad

Una definición formal de grados de libertad, muy orientada a su cálculo podría ser: "Número de unidades independientes de información en una muestra, que son relevantes para la estimación de un parámetro o para el cálculo de un estadístico."

En el ejemplo del estadístico 'varianza' existe una sola restricción por lo que los grados de libertad se reducen a $(n-1)$, y el número de unidades de información que se consideran a efectos de la estimación de la varianza poblacional son $(n-1)$. De la misma forma, en un modelo de regresión con n puntos y p parámetros los residuos tienen $(n-p)$ grados de libertad, por lo que en la estimación de la varianza de los errores σ^2 tenemos $(n-p)$ unidades independientes de información y entonces se divide la suma de cuadrados entre $(n-p)$.

Y por último, la era de la información quedaría mal representada en un artículo contemporáneo si no se hiciera referencia alguna a documentos tomados de la Internet. Aquí está una, tomada de la página Web Wikipedia, la enciclopedia libre (Ver referencia. Tomado textualmente):

Cuando se trata de ajustar modelos estadísticos a un conjunto de datos, los residuos -expresados en forma de vector- se encuentran habitualmente en un espacio de menor dimensión que aquél en el que se encontraban los datos originales. Los grados de libertad del error los determina, precisamente, el valor de esta menor dimensión.

Un sencillo ejemplo aclara el concepto. Supongamos que $X_1, \dots, X_n$ son variables aleatorias cada una de ellas con media μ , y que $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ es la "media muestral".

Entonces las cantidades $X_i - \bar{X}_n$ son los residuos, que pueden ser considerados estimaciones de los errores $X_i - \mu$. La suma de los residuos (a diferencia de la suma de los errores,

que es desconocida) es necesariamente 0. Esto significa que los residuos están restringidos a encontrarse en un espacio de dimensión $n-1$ ya que si se conoce el valor de $n-1$ de estos residuos la determinación del valor del residuo restante es inmediata. Así, se dice que "el error tiene $n-1$ grados de libertad".

Y aunque lo ideal es que cada lector elabore sus propias conclusiones, esta argumentación se aproxima mucho más a las versiones estadística y geométrica que a la probabilista y vale decir que luce como un buen aporte a quienes usen el recurso de la Internet.

6. Referencias Bibliográficas

1. BEHAR, R. y Grima, P. (2004) 55 respuestas a dudas típicas de estadística, Díaz de Santos
2. CRAMER, H. (1953) Métodos matemáticos de estadística, Aguilar
3. DRAPER, N. R. y Smith, H. (1998) Applied regression analysis, 3Ed, Wiley
4. DUDEWICKS, E. J. y Mishra, S. M. (1988) Modern Mathematical Statistics, Wiley
5. FISHER, F. A. (1925) Applications of Student's distributions, Metron, 5:3, Pág. 90
6. GRAYBILL, F. A. (1976) Theory and application of the linear model, Wadsworth
7. GUJARATI, D. N. (1997) Econometría, 3Ed, McGraw Hill
8. LEHMANN, E. L. y Casella, G. (1998) Theory of point estimation, 2Ed, Springer
9. MONTGOMERY, D. C., Peck, E. A. y Vining, G. G. (2004) Introducción al análisis de regresión lineal, CECSA
10. MOOD, A. M., Graybill, F. A. y Boes, D. C. (1974) Introduction to the theory of statistics, 3Ed, McGraw Hill
11. ROSS, S. M. (2002) Probabilidad y Estadística para Ingenieros, 2Ed, McGraw Hill
12. SEBER, G. A. F. (1977) Linear regression analysis, Wiley
13. WIKIPEDIA, la enciclopedia libre http://es.wikipedia.org/wiki/Grados_de_libertad