Grado de libertad (estadística)

https://es.wikipedia.org/wiki/Grado_de_libertad_(estadistica)

En estadística, **grados de libertad,** expresión introducida por Ronald Fisher, dice que, de un conjunto de observaciones, los grados de libertad están dados por el número de valores que pueden ser asignados de forma arbitraria, antes de que el resto de las variables tomen un valor automáticamente, producto de establecerse las que son libres, esto, con el fin de compensar e igualar un resultado el cual se ha conocido previamente.

Se encuentran mediante la fórmula $n - r$, donde n = número de sujetos en la muestra que pueden tomar un valor y r es el número de sujetos cuyo valor dependerá del que tomen los miembros de la muestra que son libres.

También pueden ser representados por $k - r$, donde k = número de grupos, cuando se realizan operaciones con grupos y no con sujetos individuales.

Cuando se trata de eliminar los estadísticos con 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 ejemplo puede aclarar el concepto.

Supóngase $X_1, \ldots, X_n$ que son variables aleatorias, cada una de ellas con media μ.

Así, $\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$ será es la "media muestral".

Entonces las cantidades $X_i - \overline{X}_n$ son los residuos.

Estos pueden ser considerados estimaciones de los errores $X_i - \mu$.

La suma de los residuos (a diferencia de la suma de los errores, que no es conocida) es necesariamente 0, Así,

$$\sum_{i=1}^{n}(X_i - \overline{X}_n) = \sum_{i=1}^{n} X_i - n\overline{X}_n = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0$$

ya que existen variables con valores superiores e inferiores a la media muestral. Esto también significa que los residuos están restringidos a encontrarse en un espacio de dimensión $n - 1$ (en este ejemplo, en el caso general a $n - r$) 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" (el error tiene $n - r$ grados de libertad general).