Degrees of freedom (statistics)

From Wikipedia, the free encyclopedia

In statistics, the number of **degrees of freedom** is the number of values in the final calculation of a statistic that are free to vary.[1]

The number of independent ways by which a dynamic system can move, without violating any constraint imposed on it, is called *number of degrees of freedom*. In other words, the number of degree of freedom can be defined as the minimum number of independent coordinates that can specify the position of the system completely.

Estimates of statistical parameters can be based upon different amounts of information or data. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom. In general, the degrees of freedom of an estimate of a parameter is equal to the number of independent scores that go into the estimate minus the number of parameters used as intermediate steps in the estimation of the parameter itself (i.e. the sample variance has N-1 degrees of freedom, since it is computed from N random scores minus the only 1 parameter estimated as intermediate step, which is the sample mean).[2]

Mathematically, degrees of freedom is the number of dimensions of the domain of a random vector, or essentially the number of "free" components (how many components need to be known before the vector is fully determined).

The term is most often used in the context of linear models (linear regression, analysis of variance), where certain random vectors are constrained to lie in linear subspaces, and the number of degrees of freedom is the dimension of the subspace. The degrees of freedom are also commonly associated with the squared lengths (or "sum of squares" of the coordinates) of such vectors, and the parameters of chi-squared and other distributions that arise in associated statistical testing problems.

While introductory textbooks may introduce degrees of freedom as distribution parameters or through hypothesis testing, it is the underlying geometry that defines degrees of freedom, and is critical to a proper understanding of the concept. Walker (1940)[3] has stated this succinctly as "the number of observations minus the number of necessary relations among these observations."

Contents

Notation

In equations, the typical symbol for degrees of freedom is ν (lowercase Greek letter nu). In text and tables, the abbreviation "d.f." is commonly used. R.A. Fisher used n to symbolize degrees of freedom but modern usage typically reserves n for sample size.

Residuals

A common way to think of degrees of freedom is as the number of independent pieces of information available to estimate another piece of information. More concretely, the number of degrees of freedom is the number of independent observations in a sample of data that are available to estimate a parameter of the population from which that sample is drawn. For example, if we have two observations, when calculating the mean we have two independent observations; however, when calculating the variance, we have only one independent observation, since the two observations are equally distant from the mean.

In fitting statistical models to data, the vectors of residuals are constrained to lie in a space of smaller dimension than the number of components in the vector. That smaller dimension is the number of **degrees of freedom for error**.

Linear regression

Perhaps the simplest example is this. Suppose

$$X_1, \ldots, X_n$$

are random variables each with expected value μ, and let

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

be the "sample mean." Then the quantities

$$X_i - \overline{X}_n$$

are residuals that may be considered estimates of the errors $X_i - \mu$. The sum of the residuals (unlike the sum of the errors) is necessarily 0. If one knows the values of any $n - 1$ of the residuals, one can thus find the last one. That means they are constrained to lie in a space of dimension $n - 1$. One says that **"there are $n - 1$ degrees of freedom for errors."**

An only slightly less simple example is that of least squares estimation of a and b in the model

$$Y_i = a + bx_i + e_i \text{ for } i = 1, \ldots, n$$

where x_i is given, but e_i and hence Y_i are random. Let $\widehat{a}$ and $\widehat{b}$ be the least-squares estimates of a and b. Then the residuals

$$e_i = y_i - (\widehat{a} + \widehat{b}x_i)$$

are constrained to lie within the space defined by the two equations

$$e_1 + \cdots + e_n = 0,$$
$$x_1 e_1 + \cdots + x_n e_n = 0.$$

One says that **there are $n - 2$ degrees of freedom for error.**

Note about notation: the capital letter Y is used in specifying the model, while lower-case y in the definition of the residuals; that is because the former are hypothesized random variables and the latter are actual data.

We can generalise this to multiple regression involving p parameters and covariates (e.g. $p - 1$ predictors and one mean), in which case the cost in *degrees of freedom of the fit* is p.

Degrees of freedom of a random vector

Geometrically, the degrees of freedom can be interpreted as the dimension of certain vector subspaces. As a starting point, suppose that we have a sample of n independent normally distributed observations,

$$X_1, \ldots, X_n.$$

This can be represented as an *n*-dimensional random vector:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Since this random vector can lie anywhere in *n*-dimensional space, it has n degrees of freedom.

Now, let $\bar{X}$ be the sample mean. The random vector can be decomposed as the sum of the sample mean plus a vector of residuals:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \bar{X} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix}.$$

The first vector on the right-hand side is constrained to be a multiple of the vector of 1's, and the only free quantity is $\bar{X}$. It therefore has 1 degree of freedom.

The second vector is constrained by the relation $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$. The first $n - 1$ components of this vector can be anything. However, once you know the first $n - 1$ components, the constraint tells you the value of the nth component. Therefore, this vector has $n - 1$ degrees of freedom.

Mathematically, the first vector is the orthogonal, or least-squares, projection of the data vector onto the subspace spanned by the vector of 1's. The 1 degree of freedom is the dimension of this subspace. The second residual vector is the least-squares projection onto the $(n - 1)$-dimensional orthogonal complement of this subspace, and has $n - 1$ degrees of freedom.

In statistical testing applications, often one isn't directly interested in the component vectors, but rather in their squared lengths. In the example above, the residual sum-of-squares is

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left\| \begin{matrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{matrix} \right\|^2.$$

If the data points X_i are normally distributed with mean 0 and variance σ^2, then the residual sum of squares has a scaled chi-squared distribution (scaled by the factor σ^2), with $n - 1$ degrees of freedom. The degrees-of-freedom, here a parameter of the distribution, can still be interpreted as the dimension of an underlying vector subspace.

Likewise, the one-sample t-test statistic,

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n - 1)}}$$

follows a Student's t distribution with $n - 1$ degrees of freedom when the hypothesized mean μ_0 is correct. Again, the degrees-of-freedom arises from the residual vector in the denominator.

Degrees of freedom in linear models

The demonstration of the t and chi-squared distributions for one-sample problems above is the simplest example where degrees-of-freedom arise. However, similar geometry and vector decompositions underlie much of the theory of linear models, including linear regression and analysis of variance. An explicit example based on comparison of three means is presented here; the geometry of linear models is discussed in more complete detail by Christensen (2002).[4]

Suppose independent observations are made for three populations, $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$. The restriction to three groups and equal sample sizes simplifies notation, but the ideas are easily generalized.

The observations can be decomposed as

$$X_i = \bar{M} + (\bar{X} - \bar{M}) + (X_i - \bar{X})$$
$$Y_i = \bar{M} + (\bar{Y} - \bar{M}) + (Y_i - \bar{Y})$$
$$Z_i = \bar{M} + (\bar{Z} - \bar{M}) + (Z_i - \bar{Z})$$

where $\bar{X}, \bar{Y}, \bar{Z}$ are the means of the individual samples, and $\bar{M} = (\bar{X} + \bar{Y} + \bar{Z})/3$ is the mean of all $3n$ observations. In vector notation this decomposition can be written as

$$
\begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_n \\ Z_1 \\ \vdots \\ Z_n \end{pmatrix}
= \bar{M}
\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
+
\begin{pmatrix} \bar{X} - \bar{M} \\ \vdots \\ \bar{X} - \bar{M} \\ \bar{Y} - \bar{M} \\ \vdots \\ \bar{Y} - \bar{M} \\ \bar{Z} - \bar{M} \\ \vdots \\ \bar{Z} - \bar{M} \end{pmatrix}
+
\begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \\ Y_1 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \\ Z_1 - \bar{Z} \\ \vdots \\ Z_n - \bar{Z} \end{pmatrix}.
$$

The observation vector, on the left-hand side, has $3n$ degrees of freedom. On the right-hand side, the first vector has one degree of freedom (or dimension) for the overall mean. The second vector depends on three random variables, $\bar{X} - \bar{M}, \bar{Y} - \bar{M}$ and $\bar{Z} - \bar{M}$. However, these must sum to 0 and so are constrained; the vector therefore must lie in a 2-dimensional subspace, and has 2 degrees of freedom. The remaining $3n - 3$ degrees of freedom are in the residual vector (made up of $n - 1$ degrees of freedom within each of the populations).

Sum of squares and degrees of freedom

In statistical testing problems, one usually isn't interested in the component vectors themselves, but rather in their squared lengths, or Sum of Squares. The degrees of freedom associated with a sum-of-squares is the degrees-of-freedom of the corresponding component vectors.

The three-population example above is an example of one-way Analysis of Variance. The model, or treatment, sum-of-squares is the squared length of the second vector,

$$SSTr = n(\bar{X} - \bar{M})^2 + n(\bar{Y} - \bar{M})^2 + n(\bar{Z} - \bar{M})^2$$

with 2 degrees of freedom. The residual, or error, sum-of-squares is

$$SSE = \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2 + \sum_{i=1}^{n}(Z_i - \bar{Z})^2$$

with $3(n−1)$ degrees of freedom. Of course, introductory books on ANOVA usually state formulae without showing the vectors, but it is this underlying geometry that gives rise to SS formulae, and shows how to unambiguously determine the degrees of freedom in any given situation.

Under the null hypothesis of no difference between population means (and assuming that standard ANOVA regularity assumptions are satisfied) the sums of squares have scaled chi-squared distributions, with the corresponding degrees of freedom. The F-test statistic is the ratio, after scaling by the degrees of freedom. If there is no difference between population means this ratio follows an F distribution with 2 and $3n − 3$ degrees of freedom.

In some complicated settings, such as unbalanced split-plot designs, the sums-of-squares no longer have scaled chi-squared distributions. Comparison of sum-of-squares with degrees-of-freedom is no longer meaningful, and software may report certain fractional 'degrees of freedom' in these cases. Such numbers have no genuine degrees-of-freedom interpretation, but are simply providing an *approximate* chi-squared distribution for the corresponding sum-of-squares. The details of such approximations are beyond the scope of this page.

Degrees of freedom parameters in probability distributions

Several commonly encountered statistical distributions (Student's t, Chi-Squared, F) have parameters that are commonly referred to as *degrees of freedom*. This terminology simply reflects that in many applications where these distributions occur, the parameter corresponds to the degrees of freedom of an underlying random vector, as in the preceding ANOVA example. Another simple example is: if $X_i; i = 1, \ldots, n$ are independent normal (μ, σ^2) random variables, the statistic

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

follows a chi-squared distribution with $n−1$ degrees of freedom. Here, the degrees of freedom arises from the residual sum-of-squares in the numerator, and in turn the $n−1$ degrees of freedom of the underlying residual vector $\{X_i - \bar{X}\}$.

In the application of these distributions to linear models, the degrees of freedom parameters can take only integer values. The underlying families of distributions allow fractional values for the degrees-of-freedom parameters, which can arise in more sophisticated uses. One set of examples is problems where chi-squared approximations based on effective degrees of freedom are used. In other applications, such as modelling heavy-tailed data, a t or F distribution may be used as an empirical model. In these cases, there is no particular *degrees of freedom* interpretation to the distribution parameters, even though the terminology may continue to be used.

Effective degrees of freedom

Many regression methods, including ridge regression, linear smoothers and smoothing splines are not based on ordinary least squares projections, but rather on regularized (generalized and/or penalized) least-squares, and so degrees of freedom defined in terms of dimensionality is generally not useful for these procedures. However, these procedures are still linear in the observations, and the fitted values of the regression can be expressed in the form

$$\hat{y} = Hy,$$

where $\hat{y}$ is the vector of fitted values at each of the original covariate values from the fitted model, y is the original vector of responses, and H is the hat matrix or, more generally, smoother matrix.

For statistical inference, sums-of-squares can still be formed: the model sum-of-squares is $||Hy||^2$; the residual sum-of-squares is $||y - Hy||^2$. However, because H does not correspond to an ordinary least-squares fit (i.e. is not an orthogonal projection), these sums-of-squares no longer have (scaled, non-central) chi-squared distributions, and dimensionally defined degrees-of-freedom are not useful.

The *effective degrees of freedom* of the fit can be defined in various ways to implement goodness-of-fit tests, cross-validation and other inferential procedures. Here one can distinguish between *regression effective degrees of freedom* and *residual effective degrees of freedom*.

Regression effective degrees of freedom

Regarding the former, appropriate definitions can include the trace of the hat matrix,[5] tr(H), the trace of the quadratic form of the hat matrix, tr($H'H$), the form tr($2H - H H'$), or the Satterthwaite approximation, tr($H'H$)^2/tr($H'HH'H$). In the case of linear regression, the hat matrix H is $X(X'X)^{-1}X'$, and all these definitions reduce to the usual degrees of freedom. Notice that

$$\operatorname{tr}(H) = \sum_i h_{ii} = \sum_i \frac{\partial \hat{y}_i}{\partial y_i},$$

the regression (not residual) degrees of freedom in linear models are "the sum of the sensitivities of the fitted values with respect to the observed response values",[6] i.e. the sum of leverage scores.

Residual effective degrees of freedom

There are corresponding definitions of residual effective degrees-of-freedom (redf), with H replaced by $I - H$. For example, if the goal is to estimate error variance, the redf would be defined as tr($(I - H)'(I - H)$), and the unbiased estimate is (with $\hat{r} = y - Hy$),

$$\hat{\sigma}^2 = \frac{||\hat{r}||^2}{\operatorname{tr}\left((I - H)'(I - H)\right)},$$

or:[7][8][9]

$$\hat{\sigma}^2 = \frac{||\hat{r}||^2}{n - \operatorname{tr}(2H - HH')} = \frac{||\hat{r}||^2}{n - 2\operatorname{tr}(H) + \operatorname{tr}(HH')}.$$

$$\hat{\sigma}^2 \approx \frac{||\hat{r}||^2}{n - 1.25\operatorname{tr}(H) + 0.5}.$$

The last approximation above[8] reduces the computational cost from $O(n^2)$ to only $O(n)$. In general the numerator would be the objective function being minimized; e.g., if the hat matrix includes an observation covariance matrix, Σ, then $||\hat{r}||^2$ becomes $\hat{r}'\Sigma^{-1}\hat{r}$.

General

Note that unlike in the original case, non-integer degrees of freedom are allowed, though the value must usually still be constrained between 0 and n.

Consider, as an example, the k-nearest neighbour smoother, which is the average of the k nearest measured values to the given point. Then, at each of the n measured points, the weight of the original value on the linear combination that makes up the predicted value is just $1/k$. Thus, the trace of the hat matrix is n/k. Thus the smooth costs n/k effective degrees of freedom.

As another example, consider the existence of nearly duplicated observations. Naive application of classical formula, $n - p$, would lead to over-estimation of the residuals degree of freedom, as if each observation were independent. More realistically, though, the hat matrix $H = X(X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1}$ would involve an observation covariance matrix Σ indicating the non-zero correlation among observations. The more general formulation of effective degree of freedom would result in a more realistic estimate for, e.g., the error variance σ^2.

Other formulations

Similar concepts are the *equivalent degrees of freedom* in non-parametric regression,[10] the *degree of freedom of signal* in atmospheric studies,[11][12] and the *non-integer degree of freedom* in geodesy.[13][14]

Alternative

The residual sum-of-squares $||y - Hy||^2$ has a generalized chi-squared distribution, and the theory associated with this distribution[15] provides an alternative route to the answers provided above.

See also

- Pooled degrees of freedom
- Replication (statistics)
- Sample size
- Statistical model
- Variance

References

1. "Degrees of Freedom" (http://www.animatedsoftware.com/statglos/sgdegree.htm). *"Glossary of Statistical Terms"*. Animated Software. Retrieved 2008-08-21.
2. Lane, David M. "Degrees of Freedom" (http://davidmlane.com/hyperstat/A42408.html). *HyperStat Online*. Statistics Solutions. Retrieved 2008-08-21.
3. Walker, H. M. (April 1940). "Degrees of Freedom". *Journal of Educational Psychology* **31** (4): 253–269. doi:10.1037/h0054588 (https://dx.doi.org/10.1037%2Fh0054588).
4. Christensen, Ronald (2002). *Plane Answers to Complex Questions: The Theory of Linear Models* (Third ed.). New York: Springer. ISBN 0-387-95361-2.
5. Trevor Hastie, Robert Tibshirani, Jerome H. Friedman (2009), *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed., 746 p. ISBN 978-0-387-84857-0, doi:10.1007/978-0-387-84858-7 (https://dx.doi.org/10.1007%2F978-0-387-84858-7), [1] (http://books.google.com/books?

id=tVIjmNS3Ob8C&lpg=PA153&dq=degrees%20of%20freedom%20of%20a%20smoother&pg=PA154#v=on epage&q=degrees%20of%20freedom%20of%20a%20smoother&f=false) (eq.(5.16))

6. Ye, J. (1998), "On Measuring and Correcting the Effects of Data Mining and Model Selection", *Journal of the American Statistical Association*, 93 (441), 120–131. JSTOR 2669609 (http://www.jstor.org/stable/2669609) (eq.(7))

7. Clive Loader (1999), *Local regression and likelihood*, ISBN 978-0-387-98775-0, doi:10.1007/b98858 (https://dx.doi.org/10.1007%2Fb98858), [2] (http://books.google.com/books? id=D7GgBAfL4ngC&lpg=PP1&pg=PA28#v=onepage&q=degree%20of%20freedom&f=false) (eq.(2.18), p.30)

8. Trevor Hastie, Robert Tibshirani (1990), *Generalized additive models*, CRC Press, [3] (http://books.google.com/books? id=qa29r1Ze1coC&lpg=PR3&dq=Hastie%2C%20T.%20J.%2C%20and%20Tibshirani%2C%20R.%20J.%20 (1990)%2C%20Generalized%20Additive%20Models%2C%20London%3A%20Chapman%20and%20Hall.&p g=PA54#v=onepage&q=degrees%20of%20freedom&f=false) (p.54) and (eq.(B.1), p.305))

9. Simon N. Wood (2006), *Generalized additive models: an introduction with R*, CRC Press, [4] (http://books.google.com/books?id=hr17lZC- 3jQC&lpg=PA170&dq=Effective%20degrees%20of%20freedom&pg=PA172#v=onepage&q&f=false) (eq. (4,14), p.172)

10. Peter J. Green, B. W. Silverman (1994), *Nonparametric regression and generalized linear models: a roughness penalty approach*, CRC Press [5] (http://books.google.com/books?id=- AIVXozvpLUC&lpg=PA37&dq=generalized%20effective%20degrees%20of%20freedom&pg=PA37#v=onepa ge&q&f=false) (eq.(3.15), p.37)

11. Clive D. Rodgers (2000), *Inverse methods for atmospheric sounding: theory and practice*, World Scientific (eq.(2.56), p.31)

12. Adrian Doicu, Thomas Trautmann, Franz Schreier (2010), *Numerical Regularization for Atmospheric Inverse Problems*, Springer (eq.(4.26), p.114)

13. D. Dong, T. A. Herring and R. W. King (1997), Estimating regional deformation from a combination of space and terrestrial geodetic data, *J. Geodesy*, 72 (4), 200–214, doi:10.1007/s001900050161 (https://dx.doi.org/10.1007%2Fs001900050161) (eq.(27), p.205)

14. H. Theil (1963), "On the Use of Incomplete Prior Information in Regression Analysis", *Journal of the American Statistical Association*, 58 (302), 401–414 JSTOR 2283275 (http://www.jstor.org/stable/2283275) (eq.(5.19)-(5.20))

15. Jones, D.A. (1983) "Statistical analysis of empirical models fitted by optimisation", Biometrika, 70 (1), 67– 88

Further reading

- Bowers, David (1982). *Statistics for Economists*. London: Macmillan. pp. 175–178. ISBN 0-333-30110-2.
- Eisenhauer, J. G. (2008). "Degrees of Freedom". *Teaching Statistics* **30** (3): 75–78. doi:10.1111/j.1467-9639.2008.00324.x (https://dx.doi.org/10.1111%2Fj.1467-9639.2008.00324.x).
- Good, I. J. (1973). "What Are Degrees of Freedom?". *The American Statistician* **27** (5): 227–228.

doi:10.1080/00031305.1973.10479042 (https://dx.doi.org/10.1080%2F00031305.1973.10479042).
JSTOR 3087407 (https://www.jstor.org/stable/3087407).

- Walker, H. W. (1940). "Degrees of Freedom". *Journal of Educational Psychology* **31** (4): 253–
  269. doi:10.1037/h0054588 (https://dx.doi.org/10.1037%2Fh0054588). Transcription by C Olsen
  with errata (http://courses.ncssm.edu/math/Stat_Inst/PDFS/DFWalker.pdf)

External links

- Yu, Chong-ho (1997) Illustrating degrees of freedom in terms of sample size and dimensionality
  (http://www.creative-wisdom.com/computer/sas/df.html)
- Dallal, GE. (2003) Degrees of Freedom (http://www.tufts.edu/~gdallal/dof.htm)

Retrieved from "http://en.wikipedia.org/w/index.php?
title=Degrees_of_freedom_(statistics)&oldid=649822959"

Categories: Statistical terminology