WIKIPEDIA

Test statistic

A **test statistic** is a statistic (a quantity derived from the sample) used in statistical hypothesis testing.[1] A hypothesis test is typically specified in terms of a test statistic, considered as a numerical summary of a data-set that reduces the data to one value that can be used to perform the hypothesis test. In general, a test statistic is selected or defined in such a way as to quantify, within observed data, behaviours that would distinguish the null from the alternative hypothesis, where such an alternative is prescribed, or that would characterize the null hypothesis if there is no explicitly stated alternative hypothesis.

An important property of a test statistic is that its sampling distribution under the null hypothesis must be calculable, either exactly or approximately, which allows *p*-values to be calculated. A *test statistic* shares some of the same qualities of a descriptive statistic, and many statistics can be used as both test statistics and descriptive statistics. However, a test statistic is specifically intended for use in statistical testing, whereas the main quality of a descriptive statistic is that it is easily interpretable. Some informative descriptive statistics, such as the sample range, do not make good test statistics since it is difficult to determine their sampling distribution.

Two widely used test statistics are the t-statistic and the F-test.

Contents

Example

For example, suppose the task is to test whether a coin is fair (i.e. has equal probabilities of producing a head or a tail). If the coin is flipped 100 times and the results are recorded, the raw data can be represented as a sequence of 100 heads and tails. If there is interest in the marginal probability of obtaining a head, only the number T out of the 100 flips that produced a head needs to be recorded. But T can also be used as a test statistic in one of two ways:

- the exact sampling distribution of T under the null hypothesis is the binomial distribution with parameters 0.5 and 100.
- the value of T can be compared with its expected value under the null hypothesis of 50, and since the sample size is large a normal distribution can be used as an approximation to the sampling distribution either for T or for the revised test statistic $T-50$.

Using one of these sampling distributions, it is possible to compute either a one-tailed or two-tailed p-value for the null hypothesis that the coin is fair. Note that the test statistic in this case reduces a set of 100 numbers to a single numerical summary that can be used for testing.

Common test statistics

One-sample tests are appropriate when a sample is being compared to the population from a hypothesis. The population characteristics are known from theory or are calculated from the population.

Two-sample tests are appropriate for comparing two samples, typically experimental and control samples from a scientifically controlled experiment.

Paired tests are appropriate for comparing two samples where it is impossible to control important variables. Rather than comparing two sets, members are paired between samples so the difference between the members becomes the sample. Typically the mean of the differences is then compared to zero. The common example scenario for when a paired difference test is appropriate is when a single set of test subjects has something applied to them and the test is intended to check for an effect.

Z-tests are appropriate for comparing means under stringent conditions regarding normality and a known standard deviation.

A *t*-test is appropriate for comparing means under relaxed conditions (less is assumed).

Tests of proportions are analogous to tests of means (the 50% proportion).

Chi-squared tests use the same calculations and the same probability distribution for different applications:

- Chi-squared tests for variance are used to determine whether a normal population has a specified variance. The null hypothesis is that it does.
- Chi-squared tests of independence are used for deciding whether two variables are associated or are independent. The variables are categorical rather than numeric. It can be used to decide whether left-handedness is correlated with libertarian politics (or not). The null hypothesis is that the variables are independent. The numbers used in the calculation are the observed and expected frequencies of occurrence (from contingency tables).
- Chi-squared goodness of fit tests are used to determine the adequacy of curves fit to data. The null hypothesis is that the curve fit is adequate. It is common to determine curve shapes to minimize the mean square error, so it is appropriate that the goodness-of-fit calculation sums the squared errors.

F-tests (analysis of variance, ANOVA) are commonly used when deciding whether groupings of data by category are meaningful. If the variance of test scores of the left-handed in a class is much smaller than the variance of the whole class, then it may be useful to study lefties as a group. The null hypothesis is that two variances are the same – so the proposed grouping is not meaningful.

In the table below, the symbols used are defined at the bottom of the table. Many other tests can be found in other articles. Proofs exist that the test statistics are appropriate.[2]

Name	Formula	Assumptions or notes
One-sample z-test	$z = \frac{\overline{x} - \mu_0}{(\sigma/\sqrt{n})}$	(Normal population **or** $n > 30$) **and** σ known. (z is the distance from the mean in relation to the standard deviation of the mean). For non-normal distributions it is possible to calculate a minimum proportion of a population that falls within k standard deviations for any k (see: *Chebyshev's inequality*).
Two-sample z-test	$z = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Normal population **and** independent observations **and** σ_1 and σ_2 are known
One-sample *t*-test	$t = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})},$ $df = n - 1$	(Normal population **or** $n < 30$) **and** σ unknown
Paired *t*-test	$t = \frac{\overline{d} - d_0}{(s_d/\sqrt{n})},$ $df = n - 1$	(Normal population of differences **or** $n < 30$) **and** σ unknown
Two-sample pooled *t*-test, equal variances	$t = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$ $df = n_1 + n_2 - 2$ [3]	(Normal populations **or** $n_1 + n_2 > 40$) **and** independent observations **and** $\sigma_1 = \sigma_2$ unknown
Two-sample unpooled *t*-test, unequal variances (Welch's *t*-test)	$t = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$ $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$ [3]	(Normal populations **or** $n_1 + n_2 > 40$) **and** independent observations **and** $\sigma_1 \neq \sigma_2$ both unknown
One-proportion z-test	$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)}}\sqrt{n}$	$n \cdot p_0 > 10$ **and** $n(1 - p_0) > 10$ **and** it is a SRS (Simple Random Sample), see notes.
Two-proportion z-test, pooled for $H_0\colon p_1 = p_2$	$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$	$n_1 p_1 > 5$ **and** $n_1(1 - p_1) > 5$ **and** $n_2 p_2 > 5$ **and** $n_2(1 - p_2) > 5$ **and** independent observations, see notes.

	$\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$	
Two-proportion z-test, unpooled for $\lvert d_0 \rvert > 0$	$z = \dfrac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$	$n_1 p_1 > 5$ **and** $n_1(1 - p_1) > 5$ **and** $n_2 p_2 > 5$ **and** $n_2(1 - p_2) > 5$ **and** independent observations, see notes.
Chi-squared test for variance	$\chi^2 = (n-1)\dfrac{s^2}{\sigma_0^2}$	Normal population
Chi-squared test for goodness of fit	$\chi^2 = \sum^k \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$	*df* = *k* − 1 − # *parameters estimated*, and one of these must hold. • All expected counts are at least 5.[4] • All expected counts are > 1 and no more than 20% of expected counts are less than 5[5]
Two-sample F test for equality of variances	$F = \dfrac{s_1^2}{s_2^2}$	Normal populations Arrange so $s_1^2 \geq s_2^2$ and reject H_0 for $F > F(\alpha/2, n_1 - 1, n_2 - 1)$[6]
Regression *t*-test of $H_0\colon R^2 = 0.$	$t = \sqrt{\dfrac{R^2(n-k-1^*)}{1-R^2}}$	Reject H_0 for $t > t(\alpha/2, n-k-1^*)$[7] *Subtract 1 for intercept; *k* terms contain independent variables.

In general, the subscript 0 indicates a value taken from the null hypothesis, H_0, which should be used as much as possible in constructing its test statistic. ... *Definitions of other symbols:*

- α, the probability of Type I error (rejecting a null hypothesis when it is in fact true)
- n = sample size
- n_1 = sample 1 size
- n_2 = sample 2 size
- $\overline{x}$ = sample mean
- μ_0 = hypothesized population mean
- μ_1 = population 1 mean
- μ_2 = population 2 mean
- σ = population standard deviation
- σ^2 = population variance
- s = sample standard deviation
- $\sum^k$ = sum (of *k* numbers)
- s^2 = sample variance
- s_1 = sample 1 standard deviation
- s_2 = sample 2 standard deviation
- t = t statistic
- df = degrees of freedom
- $\overline{d}$ = sample mean of differences
- d_0 = hypothesized population mean difference
- s_d = standard deviation of differences
- χ^2 = Chi-squared statistic
- $\hat{p}$ = *x/n* = sample proportion, unless specified otherwise
- p_0 = hypothesized population proportion
- p_1 = proportion 1
- p_2 = proportion 2
- d_p = hypothesized difference in proportion
- $\min\{n_1, n_2\}$ = minimum of n_1 and n_2
- $x_1 = n_1 p_1$
- $x_2 = n_2 p_2$
- F = F statistic

See also

- Likelihood-ratio test
- Neyman–Pearson lemma
- R^2 = coefficient of determination
- Sufficiency (statistics)

References

1. Berger, R. L.; Casella, G. (2001). *Statistical Inference*, Duxbury Press, Second Edition (p.374)
2. Loveland, Jennifer L. (2011). *Mathematical Justification of Introductory Hypothesis Tests and Development of Reference Materials* (https://digitalcommons.usu.edu/gradreports/14) (M.Sc. (Mathematics)). Utah State University. Retrieved April 30, 2013. Abstract: "The focus was on the Neyman–Pearson approach to hypothesis testing. A brief historical development of the Neyman–Pearson approach is followed by mathematical proofs of each of the hypothesis tests covered in the reference material." The proofs do not reference the concepts introduced by Neyman and Pearson, instead they show that traditional test statistics have the probability distributions ascribed to them, so that significance calculations assuming those distributions are correct. The thesis information is also posted at mathnstats.com as of April 2013.
3. NIST handbook: Two-Sample *t*-test for Equal Means (http://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm)
4. Steel, R. G. D., and Torrie, J. H., *Principles and Procedures of Statistics with Special Reference to the Biological Sciences.*, McGraw Hill, 1960, page 350.
5. Weiss, Neil A. (1999). *Introductory Statistics* (https://archive.org/details/introductorystat00neil/page/802) (5th ed.). pp. 802 (https://archive.org/details/introductorystat00neil/page/802). ISBN 0-201-59877-9.
6. NIST handbook: F-Test for Equality of Two Standard Deviations (http://www.itl.nist.gov/div898/handbook/eda/section3/eda359.htm) (Testing standard deviations the same as testing variances)
7. Steel, R. G. D., and Torrie, J. H., *Principles and Procedures of Statistics with Special Reference to the Biological Sciences.*, McGraw Hill, 1960, page 288.)