

Statistic

A function $T(Y_1, \dots, Y_n)$ of observable random variables, which is itself an observable random variable and which does not contain any unknown parameters, is called a statistic.

In a general sense, a statistic just collects information of n different random variables. If it is used to make inferences about the parameters of a population, we call the statistic an estimator.

Estimators

- Any statistic whose values are used to estimate θ is defined to be an estimator of θ . If a parameter θ is estimated by an estimator, we usually write $\hat{\theta}$, where the hat indicates that we are dealing with an estimator of θ . Some text books use Greek letters for the unknown parameters and Roman letters for the estimators of the parameters.
- Estimators are random variables with a fixed (mean) and a random component (disturbance)
- Estimators are useful since we normally cannot observe the true underlying population and the characteristics of its distribution/density.
- The mean is an estimator / estimate of the expected population value and the sample variance is an estimator / estimate of the true population variance
- The formula/ rule to calculate the mean/ variance (characteristic) from a sample is called estimator, the value is called estimate. Thus, the realization of an estimator, i.e. the realization of the random variable $\hat{\theta}$ is called an estimate of θ .

Properties of Estimators

- **Unbiasedness:** the expected value of the estimator approaches the true value. Do we get the right answers on average across a large number of applications of a specific method of inference? The deviation from the true value is caused by random noise.
- **Efficiency:** efficiency of an estimator measures the sampling variation of all observations. Less variation is more efficient and produces better results. Confidence that the one estimate we get is close to the right one. Efficiency is a relative concept.

Testing

- Empirical research hypothesis testing is at least as important as estimation.
- An estimate of a parameter might have occurred by chance.
- Since an estimator is a random variable, the estimate is just ONE realization of the random variable. Thus, using an alternative random sample of the same size can possibly generate an estimate of the parameter with completely different implications.
- Therefore we need confidence in the precision of our estimate.
- In case we can derive the distribution of the estimator, we can make statements about the variability of the estimates using the variance of the estimator.
- Tests are tools that allow us to discriminate between the estimates for different models

Definition

A test is a decision problem with respect to an unknown parameter or a relationship between different parameters.

We condense the information of the sample by a function of the data to solve the decision problem. This function is called a test statistic.

Test Statistic

A test statistic is a known function of random variables, its purpose is to prove the null hypothesis. The test statistic has a known distribution under the null hypothesis, i.e. the distribution is known if the null is true. The test statistic follows another unknown distribution if the alternative hypothesis is true.

Example: two variables x and y .

H_0 : $\text{mean}(x) = \text{mean}(y)$

H_a : $\text{mean}(x) \neq \text{mean}(y)$

Types of Hypotheses

Let θ be a parameter and r a constant which is known under the null hypothesis:

1. Two-sided test of a simple hypothesis about parameter θ :
H₀: $\theta = r$
H_a: $\theta \neq r$
2. One-sided test of a simple hypothesis about parameter θ :
H₀: $\theta \geq r$
H_a: $\theta < r$
3. Two-sided test of a simple hypothesis about more than one parameters:
H₀: $R_1 \theta + R_2 \theta' = r$
H_a: $R_1 \theta + R_2 \theta' \neq r$
4. Joint hypothesis:
H₀: $\theta_1 = c_1, \theta_2 = c_2, \theta_3 = c_3, \theta_k = c_k$
H_a: at least one of the equalities under the null does not hold

Intuition

Tests can be designed as a kind of distance measure which measures the distance between the hypothesized parameter values and the evidence produced by the data.

If an estimate is close to the hypothesized value r we would tend not to reject the null. Large deviations of the estimate from r would lead us to reject the null.

But what is 'close' and 'far'? Classical inference divides the outcome space of the test statistic into disjoint regions, an acceptance region and a critical region. If the outcome of the test lies within the acceptance region, we do not reject the null.

Since by definition the distribution of the test statistic is known under the null, we define the critical regions by the probability for the occurrence of extreme outcomes of the test statistic, given the null is true. For example we define the 95 percent quantile which indicates that only 5 percent of all outcomes of the test are larger than this critical value.

Power of a test

The inference based on the outcome of the test statistic may be wrong.

The probability of rejecting the null for a given critical region as function of the parameter is called power function.

For a given test statistic and a critical region of a given significance level we define the probability of rejecting the null hypothesis as the power of a test

The power would be optimal if the probability of rejecting the null would be 0 if there is a relationship and 1 otherwise.

This is, however, not the case in reality. There is always a positive probability to draw the wrong conclusions from the results:

- One can reject the null-hypothesis even though it is true (type 1 error, alpha error):

$$\begin{aligned}\alpha &= \Pr[\text{Type I Error}] \\ &= \Pr[\text{rejecting the } H_0 \mid H_0 \text{ is true}]\end{aligned}$$

- Or not reject the null-hypothesis even though it is wrong (type 2 error, beta error)

$$\begin{aligned}\beta &= \Pr[\text{Type II Error}] \\ &= \Pr[\text{accepting the } H_0 \mid H_a \text{ is true}]\end{aligned}$$

Type I and Type II errors

Alpha and beta errors: an example:

A criminal trial: taking as the null hypothesis that the defendant is innocent, type I error occurs when the jury wrongly decides that the defendant is guilty. A type two error occurs when the jury wrongly acquits the defendant.

In significance test:

H_0 : beta is insignificant = 0:

Type I error: wrongly rejecting the null hypothesis

Type II error: wrongly accepting the null that the coefficient is zero.

Selection of significance levels increase or decrease the probability of type I and type II errors.

The smaller the significance level (5%, 1%) the lower the probability of type I and the higher the probability of type II errors.

Accepting Hypotheses?

From a research strategy point of view, we do not speak about acceptance of a hypothesis. We only infer that the data are not contradicting the null hypothesis. This terminology emphasizes that we do not know the real world after testing and the hypothesis about the world is just a working hypothesis which we use as long as we do not find a model that describes the world more appropriately.

One-Sample T-Test

Assumptions: Y_1, \dots, Y_n – iid (independent, identically distributed) sample, $Y \sim N(\mu, \sigma^2)$ – if Y is not normally distributed – central limit theorem.

$H_0: \mu = c$

$H_a: \mu \neq c$ (thus $\mu > c$ or $\mu < c$)

test statistic: student's t-statistic: follows a t-distribution under the null with $n-1$ degrees of freedom:

$$t = \frac{\bar{y} - c}{s / \sqrt{n}}$$

H_0 is rejected with respect to the level of alpha if: $|t| > t_{n-1, 1-\alpha/2}$

Since this is a two sided test, alpha defines the critical value to which to compare the test-statistic at the 2.5 and 97.5 percentile of the t-distribution.

Two-Sample T-Test

Assumptions: $Y_1, \dots, Y_n; X_1, \dots, X_n$ – iid (independent, identically distributed) sample, $Y, X \sim N(\mu, \sigma^2)$

$$H_0: \bar{X} = \bar{Y}$$

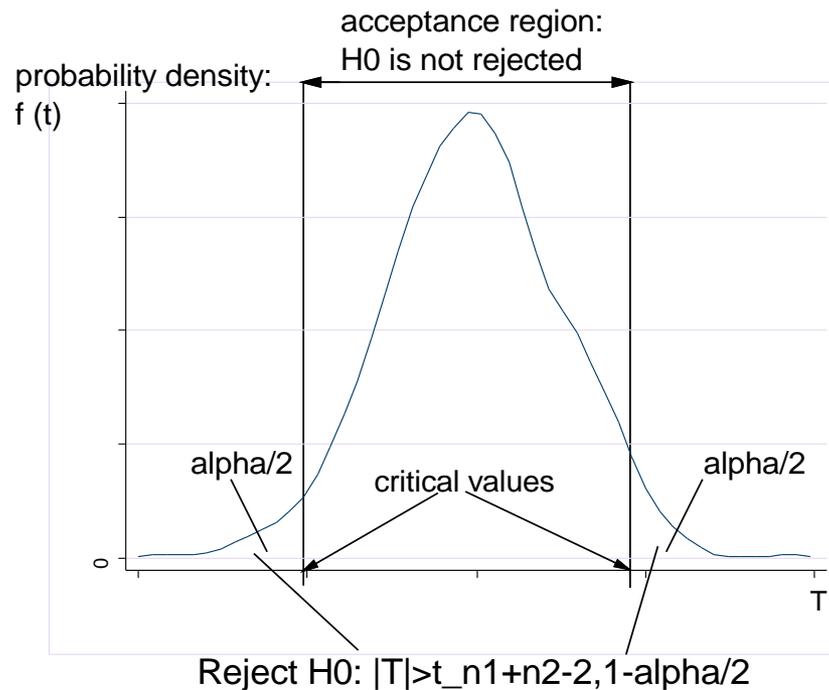
$$H_a: \bar{X} \neq \bar{Y}$$

test statistic: student's t-statistic: follows a t-distribution under the null:

$$T = \frac{\bar{y} - \bar{x}}{\sqrt{s_y^2 / n_1 + s_x^2 / n_2}}, \quad T \sim t_{(n_1+n_2-2), 1-\alpha/2}$$

H_0 is rejected with respect to the level of alpha if: $|T| > t_{(n_1+n_2-2), 1-\alpha/2}$

In a two sided test, alpha defines the critical value to which to compare the test-statistic at the 2.5 and 97.5 percentile of the t-distribution.



The area under the curve captures the probability for all outcomes of the test statistic T and has to amount to 1.