Basic concepts and introduction to statistical inference

Anna Helga Jonsdottir
Gunnar Stefansson
Sigrun Helga Lund

University of Iceland

A review of concepts

- Basic concepts
- Confidence intervals
- Hypotheses test
- p-values

Population and sample

Inferential statistics: To generalize from a sample to a larger group of people or items

We take a **sample** from the population and use that to generalize about an underlying **population**.

If the sampling scheme is appropriate then the generalization works.

Estimators and estimates

Estimator

An **estimator** is a statistic which estimates parameters of a population/probability distributions.

- Estimators for parameters of normal distribution, Poisson distribution and binomial distribution e.g. μ, σ, λ and π.
- Estimators are denoted e.g. by $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ and $\hat{\pi}$.
- The outcome of an estimator is an estimate, typically $\bar{x},\ s,\ldots$.

Confidence intervals

Usually the probability is zero of the estimate becoming exactly the true value of the parameter.

Confidence intervals

An interval which contains the true value with a confidence level 1 - α.

Confidence level

The proportion of cases where the confidence interval contains the true parameter, in repeated experiments.

Confidence limits

The endpoints of the confidence interval, called the **lower and upper confidence limit** (or bounds)

The ideology behind hypothesis tests

The ideology behind hypothesis tests

A hypothesis is found which describes what we want to demonstrate and another that describes a neutral (null) case.

A statistic is found which has a known probability distribution in the neutral case. This statistic is our test statistic.

It is defined what values of the test statistic are "improbable" according to the probability distribution in the neutral case.

If the retrieved estimate classifies as "improbable" the hypothesis for the neutral stage is rejected and the hypothesis we want to demonstrate is claimed.

If the estimate is not "improbable" no claims are made.

Hypotheses

Null hypothesis

A **null hypothesis** is a hypothesis that can be rejected with observed data. It can never we be claimed. It is usually denoted with H_0.

Alternative hypothesis

An **alternative hypothesis** is the hypothesis we wish to confirm with the experiment. It can only be claimed but not rejected. It is either denoted with H_1 or H_a.

Test statistics

Test statistic

A **test statistic** is a statistic that can be used to reject a null hypothesis if the measurements allow.

Null hypothesis rejected

A null hypothesis is **rejected** if the test statistic receives an improbable value compared to the probability distribution it should have if the null hypothesis would be true.

Rejection areas and α-levels

α-level

The α **level** of a hypothesis test is the highest acceptable probability that we receive an improbable value when the null hypothesis is true.

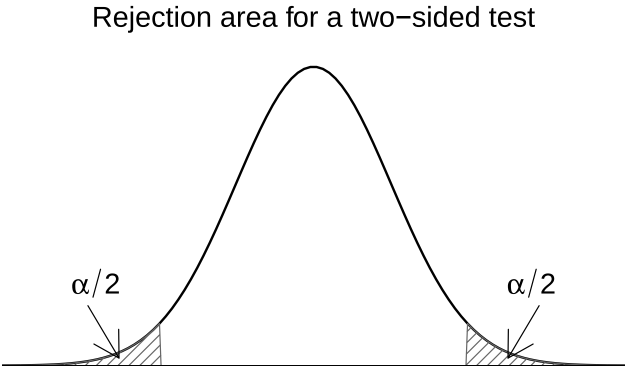
Rejection areas of hypothesis tests

Rejection areas of hypothesis tests are the intervals that contain **all** of the improbable values and **only** those values.

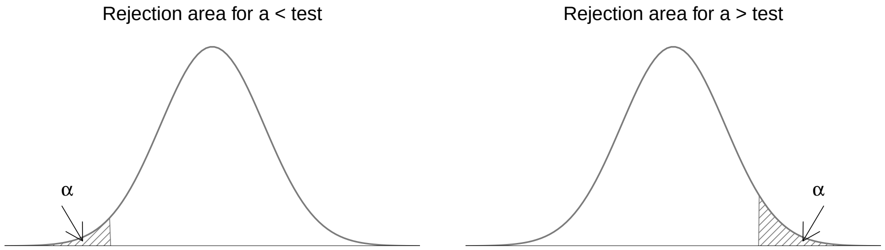
If the test statistics falls within the rejection interval of the hypothesis test, we reject the null hypothesis.

If it does not fall within the rejection interval of the hypothesis test, we make no claims

Rejection area - one-sided



Rejection area - two-sided



Rejection areas and α-levels

The probability that a test statistic falls within the rejection are when the null hypothesis is true is exactly the α-level of the hypothesis test.

In order to define rejection areas one needs to decide:

- What is the direction of the test? (one- or two-sided test)
- What is an acceptable α-level for the test.

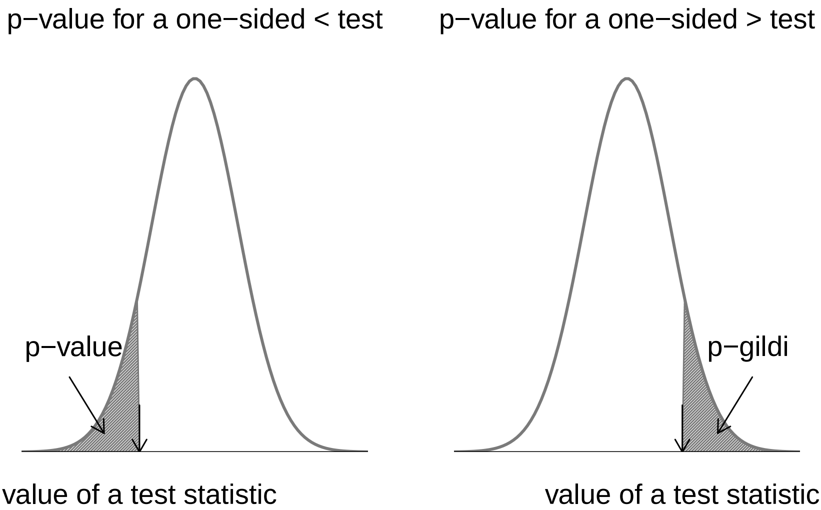
p-values

p-values

A **p-value** is the probability of receiving as improbable value or an value even more improbable as the one received with the measurements if the null hypothesis is true.

The H_0 shall be rejected if the p-value is less than α. If the p-value is greater then α the null hypothesis cannot be rejected.

p-values


p−value for a one−sided < test
p−value
value of a test statistic
p−value for a one−sided > test
p−gildi
value of a test statistic

Errors of type I and II

Type I error

Type I error is the error of rejecting a null hypothesis that was true. The probability of a type I error is the α-level of the hypothesis test.

Type II error

Type II error is the error of not rejecting a null hypothesis that was not true. The probability of a type II error is β, where $1 - \beta$ is the power of the hypothesis test.

	H_0 is true	H_0 is false
Reject H_0	Type I error Probability: α	Right decision Probability: $1 - \beta$
Not reject H_0	Right decision Probability: 1-α	Type II error Probability: β

Power

Power

The **power** of a hypothesis test is the probability of rejecting a null hypothesis that is not true. It is denoted with $1 - \beta$.

Not rejecting a null hypothesis

There can be various reasons behind one not rejecting a null hypothesis:

- The number of measurements was to small and therefore the hypothesis test had little power.
- The null hypothesis is true.
- Our model does not fit the measurements - the assumptions we made about the measurements do not hold.

We may never claim which one of the following cases was the reason! But we may make arguments for one reason being the most plausible.

Conducting hypothesis tests

Conducting hypothesis tests

1 Decide which hypothesis test is appropriate.

2 Decide the α-level.

3 Propose a null hypothesis and decide the direction of the test (one- or two-sided).

4 Calculate the test statistic for the hypothesis test.

5a See whether the test statistic falls within the rejection interval.

5b Look at the p-value of the test statistic.

6 Draw conclusions.

The relationship between confidence intervals and hypothesis tests

If the α-level is the same for both the confidence interval and the hypothesis test, the following are equivalent:

- We **reject** the null hypothesis that a particular statistic has a certain value.
- The confidence interval calculated does **not** contain that value.