Statistical inference

Anna Helga Jónsdóttir
Sigrún Helga Lund

University of Iceland

July 2020

Main topics:

1. Sampling distribution
2. Expected value and variance of the sum of two random variables
3. The statistic mean
4. Central limit theorem
5. Estimators and test statistics
6. Estimators
7. Confidence intervals
8. Hypothesis tests

Statistics

Statistic

A **statistic** is a number that is calculated by some method from our data.

- We look at our measurements as random variables because the outcome can change each time the experiment is repeated.
- Statistics are calculated from our measurements.
- If the outcomes change, the statistics can also change!
- That means that statistics are in fact random variables!

Sampling distribution

Sampling distribution

Every statistic is a random variable and has therefore some probability distribution. That distribution is called the **sampling distribution** of the statistic.

The sampling distribution depends on

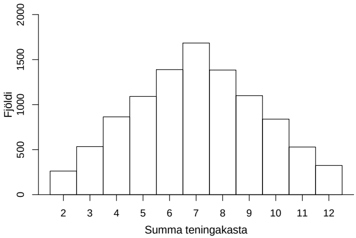
- The **probability distribution of the measurements** that the statistic is calculated for.
- The **number of measurements**.

When certain criteria are fulfilled the sampling distribution of some statistics follow certain known types. Statistical inference normally relies on that fact.

Example

Let X_1 and X_2 be random variables that describe the outcome when a dice is thrown.
Below are shown all possible outcomes of the statistic $X_1 + X_2$.

					(6,1)					
				(5,1)	(5,2)	(6,2)				
			(4,1)	(4,2)	(4,3)	(5,3)	(6,3)			
		(3,1)	(3,2)	(3,3)	(3,4)	(4,4)	(5,4)	(6,4)		
	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)	
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
2	3	4	5	6	7	8	9	10	11	12



Expected value of the sum of two random variables

Formulas for the expected value of random variables

If X and Y are two random variables, then

$$E[X+Y] = E[X] + E[Y]$$
$$E[X-Y] = E[X] - E[Y]$$

Variance of the sum of two random variables

Formulas for the variance of random variables

If X and Y are two independent random variables then

$$Var[X+Y] = Var[X] + Var[Y]$$
$$Var[X-Y] = Var[X] + Var[Y]$$

Expected value and variance of the mean

Expected value and variance of the mean

If $X_1, \ldots, X_n$ are independent and identically distributed random variables with expected value $E[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$, then the following holds for the mean of them, denoted $\bar{X}$:

$$E[\bar{X}] = \mu$$

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Standard error

Standard error

If $\bar{X}$ is the mean of $X_1, \ldots, X_n$, independent and identically distributed random variables with variance $Var[X_i] = \sigma^2$, then their **standard error** is

$$\sigma/\sqrt{n}$$

It is the standard deviation of the mean of the measurements.

The probability distribution of the mean of normally distributed random variables

The probability distribution of the mean of normally distributed random variables

If $X_1, \ldots, X_n$ are normally distributed random variables with mean μ and variance σ^2, then $\bar{X}$ follows also a normal distribution, with mean μ and variance σ^2/n.

$$\text{That is if } \; X_i \sim N(\mu, \sigma^2) \; \text{ then } \; \bar{X} \sim N(\mu, \sigma^2/n).$$

Central limit theorem

Central limit theorem

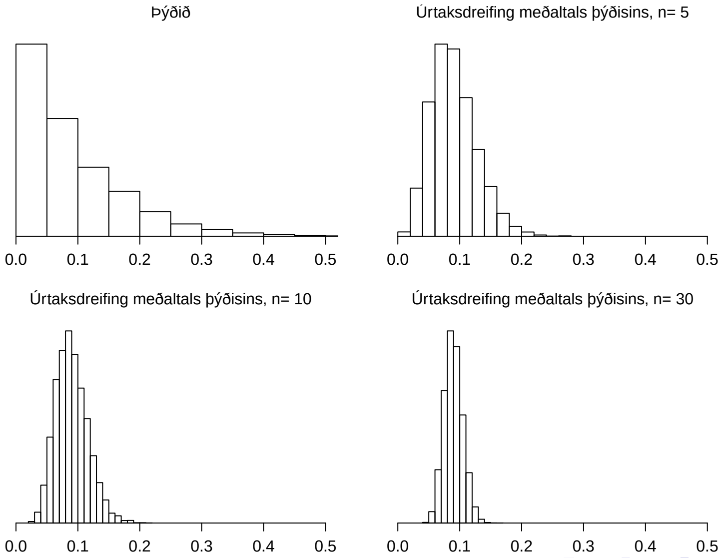
If $X_1, \ldots, X_n$ are independent and identically distributed variables then $\bar{X}$ follows normal distribution with mean μ and variance σ^2/n

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

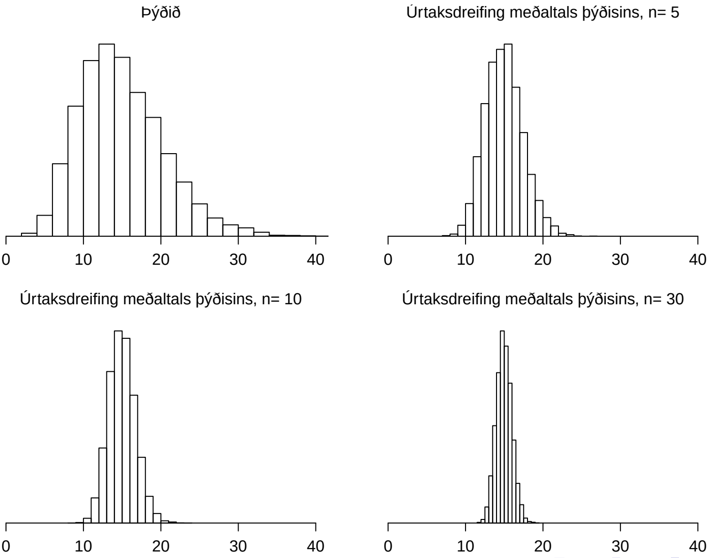
if n is large enaugh.

Notice that we do not need to know the probability distribution of the measurements!

Central limit theorem



Central limit theorem



Estimators and test statistics

There are two groups of important statistics.

- **Estimators** estimate the parameters of the probability distribution that the random variables follow.
  Example: Estimator that estimatesμ when the measurements are normally distribution.
  Example: Estimator that estimates p when the measurements are binomially distributed.
- **Test statistics** allow us to make statistical inference.
  Example: Test statistic that allows us to infer whether the variance of two population is the same.
  Example: Test statistic that allows us to infer whether the mean of a population differs from 20.

Estimators

Estimator

An **estimator** is a statistic that estimates parameters of probability distributions.

- Estimators for parameters of normal distribution, Poisson distribution and binomial distribution.
- μ, σ, λ and p.
- The outcome of the estimators are called estimates
- They are denoted with $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ and $\hat{p}$.

Estimator for the mean of a random variable

Metill á meðaltal slembistærðar

The estimator used for the mean of a random variable is

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

where n is the total number of measurements.

Estimator for the variance of a random variable

Estimator for the variance of a random variable

The estimator used for the variance of a random variable is

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

where $\bar{X}$ is the estimator for the mean of the measurements and n is the total number of measurements. mælinga.

Estimator for the ratio of a random variable

Estimator for the ratio of a random variable

The estimator used for the ratio of a random variable is

$$P = \frac{X}{n}$$

where X is the number of successful confidence intervals and n is the total number of confidence intervals.

Confidence level

Usually there is no probability that our estimate is exactly the true value of the parameter.

Confidence intervals

1 - α **confidence interval** is a numerical interval that contains the true value with the confidence level 1 - α.

Confidence level

Confidence level is the ratio of cases when the confidence interval contains the true value of the parameter, when the experiment is repeated very often.

Confidence limits

Confidence limits

Confidence limits are the endpoints of the confidence interval. The upper confidence limit is the upper endpoint of the interval (the highest value in the interval), but the lower confidence limit is the lower endpoint (the smallest value in the interval).

Type I error

Type I error denoted α, is the ratio of cases where the confidence interval contains the true value of the parameter, if the experiment is repeated very often.

The ideology behind hypothesis tests

The ideology behind hypothesis tests

A hypothesis is found that describes what we want to demonstrate and another that describes a neutral case.
A statistic is found that 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.

Hypothesis

Null hypothesis

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

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

Directions of hypothesis tests

Two-sided tests

If the data allows, a **two-sided test** claims that one or more parameters of the population or populations are **not equal** to each other or a certain value.

One-sided tests

There are two types of **one-sided tests**:
Those who claim that one parameter of the probability distribution is **larger** then another parameter or a certain value, if the measurements allow.
Those who claim that one parameter of the probability distribution is **smaler** then another parameter or a certain value, if the measurements allow.

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 a improbable value compared to the probability distribution it should have if the null hypothesis would be true.

Rejection areas and α-levels

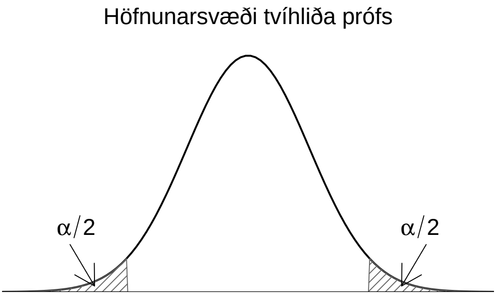
α-level

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

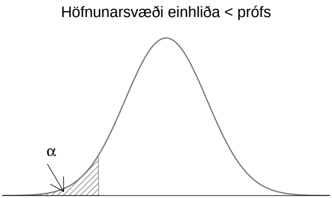
Rejection areas of hypothesis tests are the intervals that contain **all** of the improbable values and **only** those values.
If the test statistics falls within the rejection interval of the hypothesis test, we reject the null hypothesis.
If it does not fall within the rejection interval of the hypothesis test, we make no claims

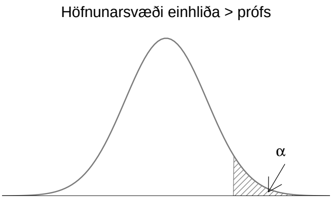
Rejection areas and α-levels



Mynd: Rejection ares of two-sided tests

Rejection areas and α-levels





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 ares 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 then α. If the p-value is greater then α the null hypothesis cannot be rejected.

Power

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

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

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 for our measurements.

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.

Example

If we conduct an hypothesis test with the α-level 5% and calculate a 95% confidence interval:

- We reject the null hypothesis that the statistic is equal to the number 1 if the number 1 is not within the confidence interval.
- The number 1 is not within the confidence interval if we reject the null hypothesis that the statistic is equal to the number 1.