02403 Introduction to Mathematical Statistics

Lecture 6: Confidence intervals and hypothesis testing

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Overview

Overview

Evidence from data

A very short summary

- In lecture 1 we discussed summary statistics (in particular the average $\bar{x}$ and the variance s^2)
- In lecture 2-5 we discussed probability theory, with special emphasis on function of summary statistics (e.g. T_{obs}, F)

In statistics we make quantitative statement about a population based on a sample hence

The sample should in a meaningful way represent the population we make statements about!

What does that mean?

Making sure that the sample represent the population might be difficult, but some points

- From a **finite sample** (e.g. height of males in DK): Make sure individuals are chosen at random (don't just ask your friends or users of a specific website)
- **Infinite samples:**
  - **Controlled experiments:** Make sure you only vary the variable of interest (don't contaminate your samles, use the same equitment, ect.)
  - **Real life measurements:** Make sure that measurements are taken in a standardlized way (e.g. don't expose the temperature censor with direct sunlight, make it clear what measurements represent, ect.)

If this is not taken care of you may end up making wrong, but very precise statements!

Overview

Interval Estimation: Confidence Intervals

Consider the $1-\alpha$ random interval $I(\boldsymbol{Y},\alpha)$ for the variable θ (e.g. μ or σ^2) such that

$$P(\theta \in I(\boldsymbol{Y},\alpha)) = 1-\alpha.$$

Notice that the interval (not θ) is random. The realization $I(\boldsymbol{y},\alpha)$ is referred to as thhe **confidence interval** for θ.

Repeated sampling interpretation

If the experiment is repeated K times $(K >> 1)$, then we expect that the true parameter is included in $(1-\alpha)K$ intervals.

Notice that in practice we have one interval and make statements based on that one interval!

What we know so far

Let $Y_1, \ldots, Y_n$ be iid. with $Y_i \sim N(\mu, \sigma^2)$ then

- If σ^2 is known then $P\left(\bar{Y} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$ implying that $I(\boldsymbol{Y}, \alpha) = \bar{Y} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$, and the $1-\alpha$ confidence interval is

$$I(\boldsymbol{y}, \alpha) = \bar{y} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = \bar{y} \pm ME$$

  where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal, and ME is "Margin of Error".

- If σ^2 is unknow then $P\left(\bar{Y} - t_{1-\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{Y} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha$ implying that $I(\boldsymbol{Y}, \alpha) = \bar{Y} \pm t_{1-\alpha/2}\frac{S}{\sqrt{n}}$, and the $1-\alpha$ confidence interval is

$$I(\boldsymbol{y}, \alpha) = \bar{y} \pm t_{1-\alpha/2}\frac{s}{\sqrt{n}} = \bar{y} \pm ME$$

  where $t_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a t-distribution with $n-1$ degees of freemdom.

Properties of confidence intervals

The Margin of Error is (for σ unknown)

$$ME = t_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

The margin of error (and hence the confidence interval width)

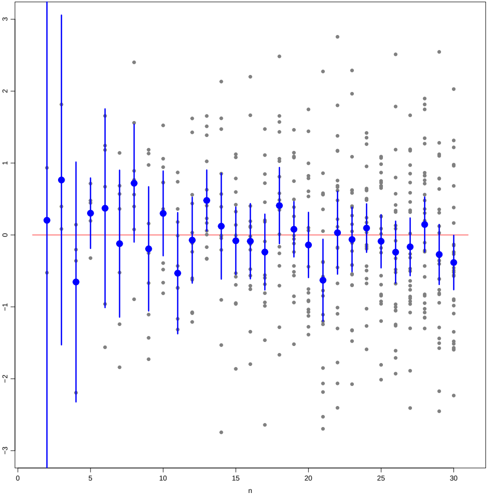
- increase when α decrease towards zero
- decrease with number of observations
- vary from sample to sample (through s)

Further the location varies (through $\bar{y}$) from sample to sample.

Interpretation

The onfidence interval is NOT about single observations. It is about the location of the true (unknown) mean value.

Confidence intervals for increasing sample stikprøvestørrelse



Example[1]

The quality assurance department at a candy factory has taken a random sample of 26 chocolate bars of a certain brand. Each chocolate bar in the sample is weighted, and it is found that the average weight is 200.3 grams and the observed standard deviation is 0.75 grams.

- What is the 95% confidence interval for the mean?

[1] Adapted from June 2023

Overview

Theorem 3.14: The Central Limit Theorem (CLT)

In large samples the distribution of the average follows a normal distribution

Let $\bar{Y}$ be the average of a randomly drawn sample of size n taken from a population with mean μ and variance σ^2. Then, the distribution of

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

approaches the standard normal distribution, $N(0, 1^2)$, as $n \to \infty$.

That is, if n is large enough, we can (approximately) assume:

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1^2).$$

Consequence of the Central Limit Theorem:

The confidence interval for μ also applies to non-normal data:

Confidence intervals for the mean can be calculated based on the t-distribution in almost all situations, as long as n is "large enough."

When is n "large enough"?

Difficult to give a precise answer, BUT:

- Rule of thumb: $n \geq 30$
- Even for smaller n, the formula may be (almost) valid for non-normal data.

I.e. if $Y_1, ..., Y_n$ are iid. with $E[Y_i] = \mu$ and $V[Y] = \sigma^2$, and n large enough the a confidence interval for μ can be calculated as

$$\bar{y} \pm t_{1-\alpha/2} \frac{s}{\sqrt{n}} \approx \bar{y} \pm z_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

where $t_{1-\alpha/2}$ is based on a t-distribution with $n-1$ degrees of freedom.

Overview

Distribution of the sample variance, Theorem 2.81

Assume that $Y_1, \ldots, Y_n$ are independent and identically distributed (iid.) random variables, $Y_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$.
We know that (lecture 3)

1. The central estimator for the variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

2. The esimator is related to the χ^2-distribution by

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

The aim is to find a confidence interval for σ^2, i.e. an interval such that

$$P(\sigma^2 \in I(\mathbf{Y}, \alpha)) = 1 - \alpha.$$

The confidence interval

Using the distribution of $\frac{(n-1)S^2}{\sigma^2}$ we get

$$P\left(\chi^2_{\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2}\right) = 1-\alpha$$

where χ^2_τ is the τ-quantile of a χ^2-distribution with $(n-1)$ degrees of freedom. The probability can be rewriten as

$$\begin{aligned} P\left(\chi^2_{\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2}\right) &= P\left(\frac{\chi^2_{\alpha/2}}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{1-\alpha/2}}{(n-1)S^2}\right) \\ &= P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right) \\ &= 1-\alpha \end{aligned}$$

Method 3.19: Confidence intervals for variance and standard deviation

Let $Y_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$ be iid.

Variance:

A $100(1-\alpha)\%$ confidence interval for the variance σ^2 is given by:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}};\ \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right],$$

where the quantiles come from a χ^2 distribution with $n-1$ degrees of freedom.

Standard deviation:

A $100(1-\alpha)\%$ confidence interval for the standard deviation σ is:

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}};\ \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right].$$

Example: Tablet Production

Tablet Production:

In the production of tablets, an active ingredient is mixed with a powder, after which the mixture is formed into tablets. The goal is to produce a homogeneous mixture so that the strength of the tablets is consistent.

From the sample of 20 measurements (representing 20 tablets), we find that the average concentration is 1.01 mg/g, with a sample standard deviation of 0.07 mg/g.

- In order to assess consistency between tablets find a 95% CI of the standard deviation.

Overview

Statistical Significance and Terminology

We say that we are *performing a hypothesis test* when we decide to reject or accept a null hypothesis based on data.

A null hypothesis is *rejected* if p-value $< \alpha$ (α is chosen in advance).

Otherwise, the null hypothesis is said to be *'accepted'*. It is more accurate (and preferable) to say that the null hypothesis cannot be rejected.

Statistical Significance: An *effect* is said to be *(statistically) significant* if the p-value is less than the significance level α.
This terminology is most meaningful when the null hypothesis is: $H_0: \; \mu = \mu_0 = 0$, which translates to the null hypothesis of "no effect."

It can also be said that μ is *significantly different from* μ_0.

Sometimes, we also say that we 'accept' the *alternative hypothesis*: $\mathrm{H}_A : \mu \neq \mu_0$

Hypothesis test

A *null hypothesis* is rejected if the outcome (i.e. the data) is *unusual* under the null hypothesis (and the assumptions).

Definition

An observation is unusual if the probability of the observation or something more extreme is small (i.e. less than α). More extreme is understod in terms of distance to the null hypethesis.[a]

[a] This is the no-directional p-value and unidirectional tests also exist.

Example: Assume that $Y \sim N(0,1)$, we observe $y = 3$ in order to determine if y is unusual we calculate

$$P(Y > |y|) + P(Y < -|y|) = 2(1 - P(Y < |y|) = 2(1 - F(y)) = 0.0027$$

which means that it is unusual if α is chosen as 0.05.

The t-test

Assume that $Y_1, .., Y_n$ are idd. random variables with $Y_i \sim N(\mu, \sigma^2)$, Further we want to investigate the hypothesis

$$H_0: \quad \mu = \mu_0.$$

Under the null-hypothesis and the assumptions then

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

In the statistical analasis we have access to one realization

$$t_{obs} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

and we can calculate the p-value as

$$p - value = 2P(T > |t_{obs}|) = 2(1 - P(T < |t_{obs}|))$$

Critical values and hypothesis test

As an alternative to comparing the p-value to the significance level we may compare the obseved t-test statistics to the critical values

$$\text{Critical values: } t_{\alpha/2} \text{ and } t_{1-\alpha/2}$$

The null-hypothesis is rejected if

$$|t_{obs}| > t_{1-\alpha/2}$$

otherwise the null-hypothesis is accepted.

We have already seen the the $1-\alpha$ confidence interval is

$$\text{CI} = \bar{y} \pm t_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

and $\mu_0 \in$CI inply that $|t_{obs}| < t_{1-\alpha/2}$, hence an new interpretation of the $(1-\alpha)$ confidence interval is: **the collection of null-hypotheses that would be accepted on level α.**

Example[2]

The engineers at an international airport have conducted a survey, in which they have timed 40 randomly selected security checks. The average duration of the security checks included in the survey was 34.66 seconds, and the sample standard deviation was 10.12 seconds, it is assumed that the times are normally distributed.

- Based on the survey, what is the 99% confidence interval for the mean duration of the security checks?
- What is the p-value for the usual test of the null hypothesis $H_0 : \mu = 30$ against a two-sided alternative hypothesis?

[2] June 2024

t-test and Python

The t-test can be done using:

```
test = stats.ttest_1samp(y, popmean = mu_0)
```

and values can be extracted by e.g.

- `test.confidence_interval(0.95)`
- `test.pvalue`

Example: Water temperature in Skive fjord

Assume the the water temperature in Skive fjord in July are realization of an iid. normal random variables. What is

- The 95% confidence interval for the water temperature in Skive fjord in July?
- The p-value and conclusion (using $\alpha = 0.05$) for the test $\mu = 20$ against the alternative?

Overview

Type I and Type II Errors

There are two types of errors (only one at a time)

Type I: Rejecting H_0 when H_0 is true (false positive).
Type II: Accepting (not rejecting) H_0 when H_1 is true (false negative).

The risk of the two types of errors is usually referred to as:

$$P(\text{Type I error}) = \alpha$$
$$P(\text{Type II error}) = \beta$$

Courtroom Analogy

A person is brought before a court:

A person is being prosecuted under a specific charge.
The null and alternative hypotheses are:

H_0 : The person is innocent.

H_1 : The person is guilty.

Not being proven guilty is not the same as being proven innocent:

In other words:
Accepting a null hypothesis is not a statistical proof that the null hypothesis is true!

Errors in Hypothesis Testing

Two possible truths against two possible conclusions:

	Rejecting H_0	Not rejecting H_0
H_0 is true	Type I error (α)	Correct acceptance of H_0
H_0 is false	Correct rejection of H_0	Type II error (β)

Overview

One sample t-test and the LM

The general assumptions of the one-sample t-test ($Y_i \sim N(\mu, \sigma^2)$ and iid), can be written as

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

i.e. the design matrix is a vector of ones ($\boldsymbol{X} = \mathbf{1}$). The parameter estimator is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} = (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}$$

and

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1} = \frac{\sigma^2}{n}.$$

The fitted values are calculated by

$$\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y} = \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{Y} = \frac{1}{n} \boldsymbol{E}\boldsymbol{Y}$$

where $\boldsymbol{E}$ is an $n \times n$ matrix of ones.

One-sample t-test as a projection

If $\boldsymbol{Y} \sim N(\mathbf{1}\mu, \boldsymbol{\sigma}^2 \boldsymbol{I})$ then the partitioning of variation can be written as

$$\boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y} + \boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y},$$

and, regardless of the value of μ, then

$$\frac{1}{\sigma^2}\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y} \sim \chi^2(n-1).$$

further if $\mu = 0$ then

$$\frac{1}{\sigma^2}\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y} \sim \chi^2(1).$$

Implying that if $\mu = 0$ then

$$F = \frac{\frac{1}{\sigma^2}\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}/1}{\frac{1}{\sigma^2}\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}/(n-1)} = \frac{\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}}{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}/(n-1)} \sim F(1,n-1).$$

which can be used to test the null-hypothesis $\mu = 0$.

One sample t-test and the LM

We can calculate the observed F-statistics for the null-hypothesis $\mu = 0$ by

$$F_{obs} = \frac{\boldsymbol{y}^T \boldsymbol{H} \boldsymbol{y}}{\boldsymbol{y}^T (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y}/(n-1)},$$

and the corresponding p-value as

$$p - value = P(F > F_{obs}); \quad F \sim F(1, n-1).$$

The t-test and the F-test are equivalent since

1. If $T \sim t(n-1)$, then $T^2 \sim F(1, n-1)$
2. $F_{obs} = t_{obs}^2$

Finally a central estimator for the variance is

$$\hat{\sigma}^2 = \frac{1}{n-1} \boldsymbol{Y}^T (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{Y} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Overview

The formal framework for *statistical inference*

From Chapter 1 of the book:

- An *observational unit* is the single entity/level about which information is sought (e.g. a person) (**Observation Unit**)
- The *statistical population* consists of all possible "measurements" on each *observational unit* (**Population**)
- The *sample* from a statistical population is the actual set of data collected. (**Sample**)

Terminology and concepts:

- μ and σ are *parameters* that describe the population
- $\hat{\mu} = \bar{x}$ is the *estimate* for μ (specific outcome value)
- $\bar{X}$ is the *estimator* for μ (now seen as a random variable)
- The concept of *statistic* is a common term for both ($\bar{x}$ and $\bar{X}$)

Statistical inference: Learning from data

Learning from data:

We want to deduce the parameter values for the underlying population.

Important in this regard:

The sample must meaningfully be *representative* of a well-defined population.

How do you ensure this?

For example, by ensuring that the sample is fully *randomly selected*.

Random sampling

Definition 3.12:

- A random sample from an (infinite) population: The random variables $Y_1, Y_2, \ldots, Y_n$ constitute a random sample of size n from the infinite population if:

    1. All the random variables have the same distribution
    2. The n random variables are independent

What does this mean?

1. All observations must come from the same population
2. They must NOT share information with each other (e.g. if you had sampled entire families instead of individuals)

Agenda

1. Introduction and random sampling
2. Confidence interval
   - Confidence interval for the mean
3. Non-normal data, The Central Limit Theorem
4. Confidence interval for variance and standard deviation
5. Hypothesis test
6. Type I and Type II errors
7. One sample t-test and the LM
8. Summary of Statistical Inference