Introduction to Statistical Inference

Marta Bofill Roig

March 17, 2025

GRBIO
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Outline

Warming up

Confidence interval

Estimators

- Estimators like $\bar{Y}$ and s^2 provide "likely" values for population parameters μ and σ^2.
- Due to chance, they are rarely exactly equal to the true parameters.
- However, "good" estimators tend to be "close" to the true values.

Confidence interval (cont'd)

Interval Estimates

- Instead of reporting a single value, we report an **interval** based on the estimator.
- This interval is likely to contain the true parameter value.
- "Likely" implies that **probability** is involved.

Here, we discuss the notion of such an interval, known as a **confidence interval**.

Confidence interval for μ

Example: we wish to estimate μ by the sample mean $\bar{Y}$.

- Assume $Y \sim N(\mu, \sigma^2)$, where μ and σ^2 are unknown.
- We have a random sample $Y_1, \ldots, Y_n$.
- Our estimator for μ is the sample mean, $\bar{Y}$.

Confidence interval for μ (cont'd)

- To make probability statements about $\bar{Y}$, we need to account for the unknown σ^2.
- We must estimate σ^2 even if it's not our primary interest.
- We use the statistic:

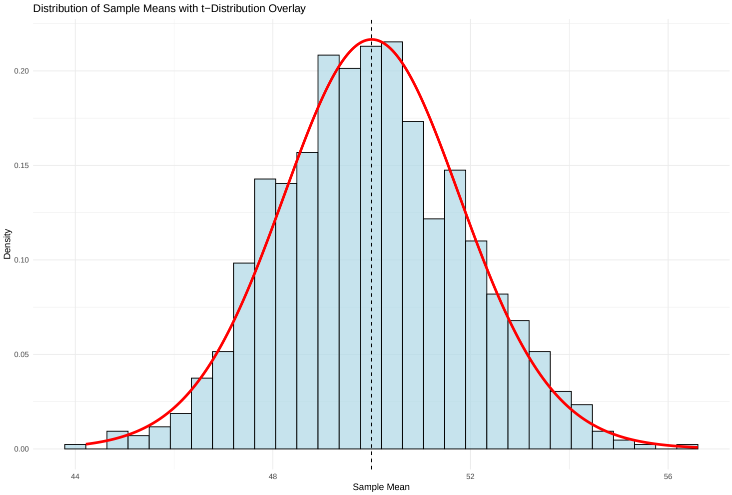
$$\frac{\bar{Y} - \mu}{s_{\bar{Y}}}$$

where $s_{\bar{Y}}$ is the estimated standard error of $\bar{Y}$.

Confidence interval for μ (cont'd)

- We want to quantify the uncertainty of estimating the fixed value μ using the random variable $\bar{Y}$.
- The randomness arises from the random sample.
- Probability statements reflect the uncertainty of trying to get an understanding of μ using $\bar{Y}$.

Distribution sample mean



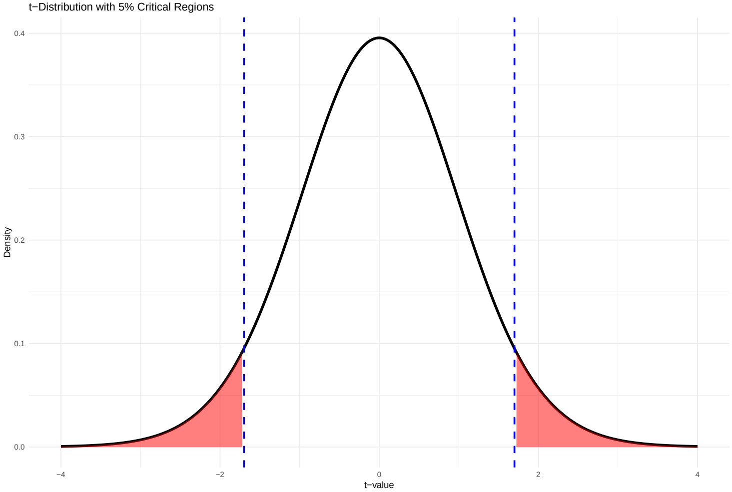
Confidence interval for μ (cont'd)

- We aim to create an interval that likely contains the true population mean μ.
- The formula is derived from the t-distribution:

$$P\left(\bar{Y} - t_{n-1,\alpha/2} \cdot s_{\bar{Y}} \leq \mu \leq \bar{Y} + t_{n-1,\alpha/2} \cdot s_{\bar{Y}}\right) = 1 - \alpha$$

- $t_{n-1,\alpha/2}$ is the critical t-value, $s_{\bar{Y}}$ is the standard error of the mean.

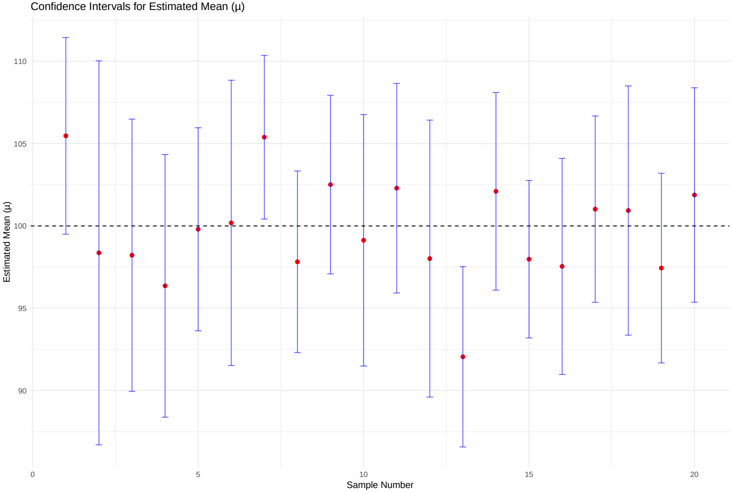
Confidence interval for μ (cont'd)



Confidence interval for μ (cont'd)

- The probability $(1-\alpha)$ is **not** about μ being in the interval.
- μ is a **fixed constant**.
- The probability refers to the **random interval** $(\bar{Y} \pm t_{n-1,\alpha/2} \cdot s_{\bar{Y}})$.
- It's the probability that this interval will contain the true μ.

Confidence interval for μ (cont'd)



Confidence interval for μ (cont'd)

The interval

$$(\bar{Y} - t_{n-1,\alpha/2} \cdot s_{\bar{Y}}, \bar{Y} + t_{n-1,\alpha/2} \cdot s_{\bar{Y}})$$

is called a $100 \cdot (1-\alpha)\%$ **confidence interval** for μ.

- For example, if $\alpha = 0.05$, then the interval would be called a 95% confidence interval.
- In general the value $1-\alpha$ is called the **confidence coefficient**.

Confidence interval for μ (cont'd)

- The interval is calculated from the sample mean ($\bar{Y}$) and its standard error ($s_{\bar{Y}}$).
- It is a **random interval** because it varies with different samples.
- Different samples yield different $\bar{Y}$ and $s_{\bar{Y}}$, and thus different intervals.

Confidence interval for μ (cont'd)

- The probability (e.g., 95%) refers to the likelihood that the **sample** produces an interval that contains the true μ.
- It's about the process of repeatedly taking samples and constructing intervals.
- In the long run, 95% of the intervals will contain μ.

Confidence interval for μ (cont'd)

The confidence coefficient (1 - α)...

- represents the probability a sample will produce an interval that covers the true population mean μ.
- measures our "confidence" in the sampling procedure and interval construction.

Confidence interval for μ (cont'd)

- The width of the confidence interval depends on the confidence coefficient and the size of the sample.
  - For a 90% interval would narrower than a 95% interval. For the 90% interval, we may be a little more stringent with the width as we aren't requiring such a high a level confidence.
  - If we wish greater confidence of 95%, we must widen the interval in order to be "more confident" that it will cover the fixed value μ.

Inference and hypothesis testing

Brief review: Estimation and Confidence Intervals

- We previously discussed:
  - Inference on a single population mean.
  - Inference on the difference of two means (under simplifying conditions).
- We explored how to:
  - Estimate a single population mean or difference of means.
  - Construct confidence intervals for these estimates.

Brief review: Estimation and Confidence Intervals (cont'd)

- Estimation and confidence intervals help us understand the unknown population parameters.
- Confidence intervals provide a range of plausible values for these parameters.
- They account for uncertainty due to sampling and biological variation.
- Probabilistic statements are crucial in statistical inference.

About today's lecture

- We will now explore statistical inference in more depth.
- Formal statistical inference involves estimating population parameters and quantifying uncertainty.
- Probabilistic reasoning is essential for making valid inferences – Probabilistic statements will continue to play a key role.

Hypothesis testing

Rat Example:

- We want to study the effect of Vitamin A on rat weight gain.
- Known: Untreated rats gain 27.8 mg on average in 3 weeks.
- Question: Does Vitamin A treatment change this mean weight gain?

Hypothesis testing (cont'd)

Experimental approach:

- We cannot test all rats; we use a sample.
- Treat the sample rats with Vitamin A.
- Sample represents the population of Vitamin A treated rats.
- Population mean weight gain is unknown: μ.

Hypothesis testing (cont'd)

What are we testing?

- We formulate two hypotheses about μ:

  1. $\mu = 27.8$ mg (Vitamin A has no effect).

  2. $\mu \neq 27.8$ mg (Vitamin A has an effect).

Hypothesis testing (cont'd)

Null Hypothesis (H_0):

- $H_0 : \mu = 27.8$
- Vitamin A has no effect.

Alternative Hypothesis (H_1):

- $H_1 : \mu \neq 27.8$
- Vitamin A has an effect.

Hypothesis testing (cont'd)

Rationale:

- We want to investigate (**research question**) if Vitamin A affects rat weight gain.
- We set up **two hypotheses**: a null hypothesis (no effect) and an alternative hypothesis (effect).
- The goal is to use **sample data** to decide between these hypotheses.

Hypothesis testing (cont'd)

A formal statistical procedure for "deciding" between H_0 and H_1 is called a hypothesis test or test of significance.

- Decisions are based on observations from a **sample**.
- Decisions are influenced by sampling procedure and biological variation.
- **Probability** plays a crucial role.

Rat Example

- Assume H_0 is true: $\mu = 27.8$ mg (Vitamin A has no effect).
- We observed $\overline{Y} = 41.0$ mg from a sample of $n = 5$.
- Key question: How likely is it to see $\overline{Y} = 41.0$ mg if $\mu = 27.8$ mg?

Rat Example (cont'd)

- If "likely":
  - 41.0 is not unusual.
  - Do not reject H_0.
- If "not likely":
  - 41.0 is unusual and unexpected.
  - Reject H_0.
- "Likely" is defined in terms of **probability**.

Hypothesis testing (cont'd)

- Hypothesis testing helps decide between null and alternative hypotheses.
- Decisions are based on sample data and probabilities.
- We assess the **likelihood** of observed data under the null hypothesis.

Hypothesis testing (cont'd)

How do we decide between H_0 and H_1?

Rationale: "Pretend" H_0 is true and assess the probability of seeing the $\overline{Y}$ value we saw for our particular sample.

- If this probability is **small**, reject H_0.
- If this probability is **not small**, do not reject H_0.

How "small" is small?

Generic hypothesis test

- Consider the generic situation:
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu \neq \mu_0$
- μ_0 is the value of interest (e.g., 27.8 mg in the rat example).
- Assume H_0 is true: $\mu = \mu_0$.
- We want to determine the probability of seeing $\overline{Y}$ (our sample mean).

Generic hypothesis test (cont'd)

Using the t-distribution

- If Y is normally distributed, under H_0:

$$\frac{\overline{Y} - \mu_0}{s_{\overline{Y}}} \sim t_{n-1}$$

  where $s_{\overline{Y}}$ is the standard error of the mean.

- This statistic follows a t-distribution with $n-1$ degrees of freedom.

Generic hypothesis test (cont'd)

- "Likely" $\overline{Y}$ values are close to μ_0.
- This means the statistic $\left|\frac{\overline{Y}-\mu_0}{s_{\overline{Y}}}\right|$ is close to 0.
- "Unlikely" $\overline{Y}$ values result in a large magnitude of the statistic.

Generic hypothesis test (cont'd)

How to formalise "Unlikely"

- Define a small probability α (e.g., 0.05).
- If the probability of seeing our $\overline{Y}$ is less than α, we reject H_0.
- This means the evidence is strong enough to refute H_0.

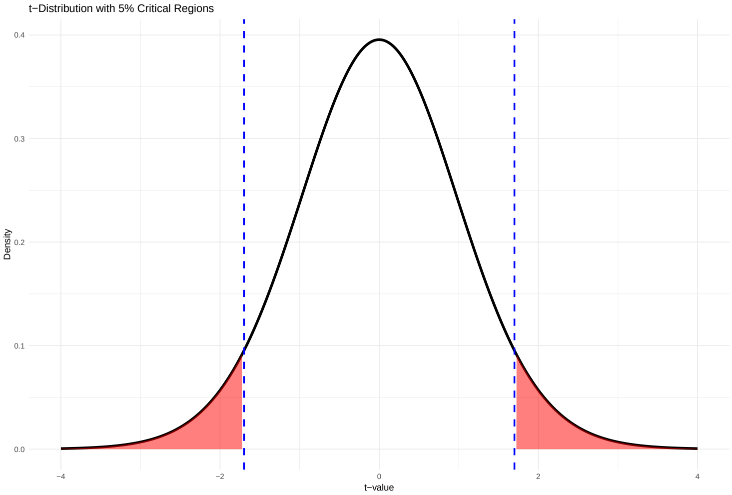
Critical values and the t-distribution

- The statistic follows a t_{n-1} **distribution**.
- There exists a **critical value** $t_{n-1,\alpha/2}$ such that:

$$P\left(\left|\frac{\overline{Y}-\mu_0}{s_{\overline{Y}}}\right| > t_{n-1,\alpha/2}\right) = \alpha$$

- $t_{n-1,\alpha/2}$ defines the **rejection region**.

Critical values and the t-distribution



- Each shaded region has area $\alpha/2$.
- Total shaded area is α.

Wrapping up – let's put all together

- We use the t-**distribution** to assess the **likelihood** of our **sample mean**.
- We define a **critical value** based on α.
- If the test statistic falls in the **rejection region**, we reject H_0.

Wrapping up – let's put all together

- Values of the statistic greater than $t_{n-1,\alpha/2}$ are "unlikely" (probability $< \alpha$).
- If our sample statistic is greater than $t_{n-1,\alpha/2}$, we reject H_0.
- We favour the alternative hypothesis H_1.

Rat Example

Continuing with the Rat experiment:

- $H_0 : \mu = \mu_0 = 27.8$ mg (Vitamin A has no effect).
- We have $n = 5$, $s_{\overline{Y}} = 4.472$.
- Calculate the test statistic:

$$\left| \frac{\overline{Y} - \mu_0}{s_{\overline{Y}}} \right| = \left| \frac{41.0 - 27.8}{4.472} \right| = 2.952$$

Rat Example (cont'd)

Comparing to the critical value:

- For $\alpha = 0.05$ and $n - 1 = 4$ degrees of freedom, $t_{4,0.025} = 2.776$.
- Compare the calculated statistic to the critical value:

$$2.952 > 2.776$$

- Since the statistic is greater than the critical value, we reject H_0.

Rat Example (cont'd)

Conclusion:

- We reject H_0: The evidence supports that mean weight gain is different from 27.8 mg.
- Vitamin A does have an effect on weight gain.

Rat Example (cont'd)

Summary:

- We calculated the test statistic and compared it to the critical value.
- Since the statistic fell in the rejection region, we rejected the null hypothesis.
- We concluded that Vitamin A has a significant effect on rat weight gain.

Decision tools for hypothesis testing

Decision tools for hypothesis testing

- The statistic

$$\frac{\overline{Y} - \mu_0}{s_{\overline{Y}}}$$

  is called a **test statistic**.

- It's a function of sample information used to decide between H_0 and H_1.

Decision tools for hypothesis testing (cont'd)

- Instead of comparing the test statistic to a critical value, we can use probabilities.
- Find the probability of observing a test statistic as extreme as the one we calculated.
- This probability is called the p**-value**.

Decision tools for hypothesis testing (cont'd)

Calculating the p-value:

- If t_{n-1} is a t-distributed random variable with $n-1$ degrees of freedom, find:

$$P(|t_{n-1}| > \text{value of test statistic we saw})$$

- Compare this probability (p-value) to α.
- If p-value $< \alpha$, reject H_0.

Rat Example

- From the t-table with $n - 1 = 4$ (or using R `pt()`), we found:

$$0.02 < P(|t_4| > 2.952) < 0.05$$

- The p-value is between 0.02 and 0.05 ("small").
- Since p-value $< \alpha = 0.05$, we reject H_0.

Rat Example (cont'd)

- The test statistic 2.952 is "unlikely".
- We reject H_0 based on the p-value.

Decision tools for hypothesis testing (cont'd)

- We defined the **test statistic** and introduced the concept of the p**-value**.
- The p-value is the probability of observing a test statistic as extreme as the one we calculated.
- If the p-value is less than α, we reject the null hypothesis.

Decision tools for hypothesis testing (cont'd)

- These two ways of conducting the hypothesis test are **equivalent**.

Decision tools for hypothesis testing (cont'd)

Decision based on the test statistic:

- Think about the size of the test statistic.
- If it is "large," it is "unlikely".
- "Large" depends on α, the chosen "unlikely" probability.

Decision tools for hypothesis testing (cont'd)

Decision based on the p-value:

- Think directly about the probability of seeing what we saw.
- If the probability is "small" (less than α), the test statistic was "unlikely".

Decision tools for hypothesis testing (cont'd)

- A "large" test statistic and a "small" probability are **equivalent**.
- Both methods lead to the same conclusion.
- The p-value method provides a measure of evidence strength.

Type I and II errors

Which sorts of error can occur in statistical tests?

		Truth	
		H_0 true	H_0 false
Decision	for H_0	✓	Type II error
	for H_1	Type I error	✓

- **Type I error**: Error of rejecting a null hypothesis when it is actually true. The significance level α must be predetermined in the study protocol, e.g., $\alpha = 0.025$.
- **Type II error**: Error of keeping a null hypothesis when it is actually false.
- **Power**: The power of a statistical test is the probability that the test will reject a false null hypothesis.

Errors in hypothesis testing

- α is chosen in advance to quantify "likely".
- Called the **significance level** or **error rate**.
- Probability of rejecting H_0 when it is actually true.

Errors in hypothesis testing (cont'd)

When do we reject H_0?

- Two scenarios:
    1. H_0 is false, leading to a large test statistic or small p-value.
    2. H_0 is true, but we got an "unusual" sample, leading to rejection.

Errors in hypothesis testing (cont'd)

- Scenario (ii) is a **mistake** - incorrect judgment.
- Called a **Type I Error**.
- Probability of Type I Error is at most α.
- α is the "error rate".

Errors in hypothesis testing (cont'd)

- We say "reject H_0 at level of significance α".
- States the criterion used to determine "likely".
- An observed test statistic leading to rejection is **statistically significant at level α**.
- Stating α is essential.

Errors in hypothesis testing (cont'd)

- α is the significance level and probability of Type I Error.
- Type I Error: Rejecting H_0 when it's true.
- Always state the significance level when reporting results.

Errors in hypothesis testing (cont'd)

- We've discussed Type I errors (rejecting H_0 when it's true).
- But what about failing to reject H_0 when it's false?

Errors in hypothesis testing (cont'd)

- **Type II Error**: Failing to reject H_0 when it's false.
- We denote the probability of a Type II error as β.

Errors in hypothesis testing (cont'd)

- We want β to be small, just like the probability of a Type I error.
- However, in many cases, a Type II error is less severe than a Type I error.

Example: Drug Testing

- **Type II Error**: Concluding a new drug is ineffective when it's actually effective.
- This means we miss out on a potentially better treatment.
- **Type I Error**: Concluding a new drug is effective when it's not.
- This means patients are exposed to unnecessary costs and risks.

References and other materials

1. D.R. Cox, E. J. Snell. Applied Statistics - Principles and Examples. Chapman & Hall, 1981.

2. Spiegelhalter, David. The art of statistics: Learning from data. Penguin UK, 2019.

3. James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning. Springer, 2013.

4. Teaching courses from `https://www4.stat.ncsu.edu/~davidian/`

5. Teaching courses from `https://www.drizopoulos.com/`

That's all folks!

GRBIO

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH