Statistical Literacy — MINT

Lecture 9: A second step in inference - Hypothesis testing for means

Rémi Viné

The Graduate Institute | Geneva

November 20th, 2023

Outline

Recap' - Confidence intervals $\rightarrow$ Get the right critical value

- Proportions: take critical value from Standard Normal Distribution
- Means:
  - If population variance known (σ^2)
    - Take critical value from the Standard Normal Distribution
  - If population variance unknown
    - Take critical value from the Student (T) Distribution

Case (and information)	Sample size	α	Critical value
Proportion ($\hat{p} = 0.3$)	28	5%	$z_{0.025} = 1.96$
Means ($\bar{x} = 12$)	41	1%	$t^{40}_{0.005} = 2.704$
Means ($\bar{x} = -34$, $\sigma = 3$)	12	0.2%	X
Means ($\bar{x} = -6$, Normal population)	19	0.2%	$t^{18}_{0.001} = 3.611$

Housekeeping

- Problem set 9 is now available
- Second assignment is due by lecture 11. Send it to your *Teaching Assistant* by lecture 11
  - Follow the guidelines
  - Build 2 confidence intervals
  - Run one hypothesis test
  - Use classical tests for hypotheses for each output (less options leading to different confidence intervals)

What is hypothesis testing?

- A **significant** leap for statisticshood!
  - The term "significant" can be used only after conducting a statistical test
- Hypothesis testing consists in stating a system of hypotheses that are to be tested. One will be rejected and this will lead to a meaningful interpretation
  - For example the statement: "There is a significant difference between men and women incomes" can be tested using hypotheses testing
    - Shrestha & Sakellariou (1996) for Canada
    - Iwasaki & Satogami (2023) for the EU

6 Steps for hypothesis testing

1. **Validity conditions:** Verify you can use inference (if not, stop here)
2. State the hypotheses
3. Calculate the **test-statistic** (and/or the corresponding **p-value**)
4. Find the **critical value**, having decided on the **level of significance** (α)
5. Compare
   - Test-statistic *versus* critical value (in absolute terms)
   - $P - value$ *versus* level of significance
6. Conclude

I. Validity conditions

- As long as the sampling distribution can be considered as Normally distributed, then the validity conditions are met
  - Verified if the population distribution is Normal, or in any other cases where $n > 30$

II. State the hypotheses

- State a **null hypothesis**: H_0; and a counterpart, the **alternative hypothesis**: H_a, also written H_1 (they cannot overlap)

  $\Rightarrow$ You **reject** H_0, or you **fail to reject** H_0

One-tail test

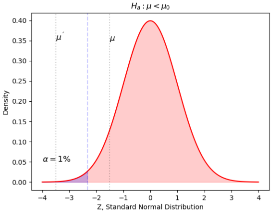
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu < \mu_0 \end{cases}$$

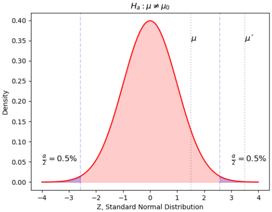
Two-tail test

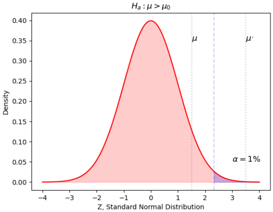
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{cases}$$

One-tail test

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu > \mu_0 \end{cases}$$







with μ_0 a constant to compare the population average with

II. Validity & Hypotheses

Example

- Assume a computer manufacturer states that the produced laptops have an battery life of 400 minutes ($= \mu_0$). You buy 100 laptops ($= n$) and observe that the average battery life in this sample is 397 minutes ($= \bar{x}$) with a variance in the sample of 121 minutes ($= s^2$).
- You want to test whether the manufacturer misestimated the average battery life.

1. Validity: $n > 30$, so the sampling distribution is approximately Normal
2. State the hypotheses (*two tail test, as this is about misestimation, neither just under- nor just over-estimation*):

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{cases}$$

III. Calculate the test-statistic

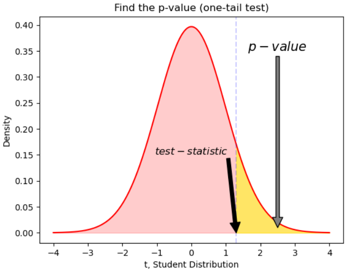
- The test-statistic is the **standardized** or the **studentized** value of the sample average
  - It tells where the sample is located compared to the tested value (μ_0) on the sampling distribution
  - $test - statistic = Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ when standardized (because the population variance is available)
  - $test - statistic = T = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ when studentized (because the population variance is **NOT** available)

- For a two tail test, the test-statistic is both the positive value and its mirror negative value (*e.g.*, $\{-2; 2\}$)

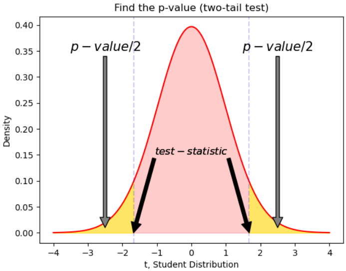
III. Or calculate the *p-value*

- The $p - value$ is the area further from the center of the distribution than the test-statistic
  - Samples in the area covered by the $p - value$ that would be even further from μ_0 than the current sample is

(a) One tail test



(b) Two tail test



III. Calculate the test-statistic

Example

- Assume a computer manufacturer states that the produced laptops have an battery life of 400 minutes ($= \mu_0$). You buy 100 laptops ($= n$) and observe that the average battery life in this sample is 397 minutes ($= \bar{x}$) with a variance in the sample of 121 minutes ($= s^2$).
- You want to test, at a level of significance of 1% ($= \alpha$), whether the manufacturer misestimated the average battery life.

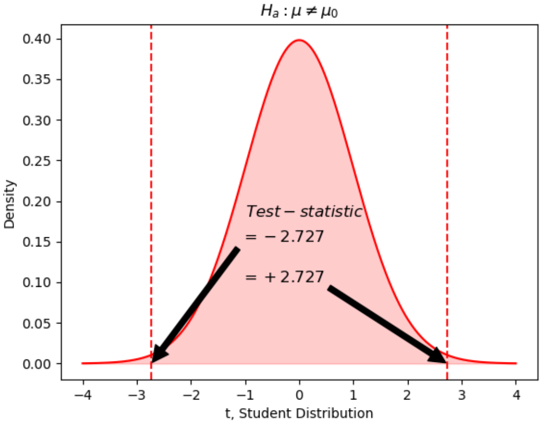
3. Test-statistic: the population variance is not available, so it is the *studentized* value (Student distribution):

$$test - statistic = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{397 - 400}{\frac{11}{10}} = -2.727$$

- Since two tail, the test-statistic is -2.727 and 2.727

III. Calculate the test-statistic

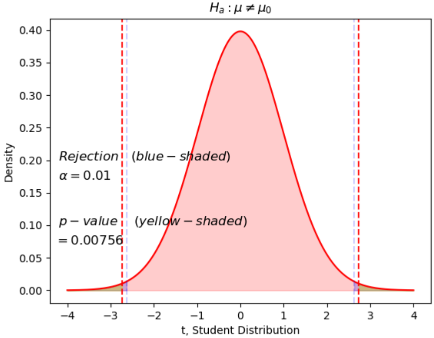
Example



III. Find the *p-value*

Example (this usually requires the use of a software, unless the test-statistic is a number that appears on the probability table → not asked directly in this class)

- Since the population variance is not available, the Student distribution must be used
  - Find on the Student distribution table the probability corresponding a value of 2.727 (two tail test with $\alpha = 0.01$)

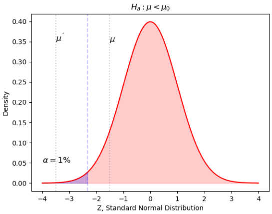


IV. Find the critical value

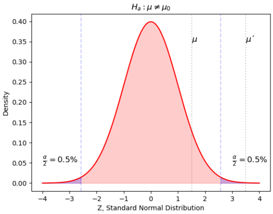
Define first the level of significance

- The critical value gives the thresholds beyond which there is the rejection area (blue areas in the plots)
  - The critical value is the blue dashed line threshold
- The critical value is derived based on the level of significance α (which does not depend on the sample itself)

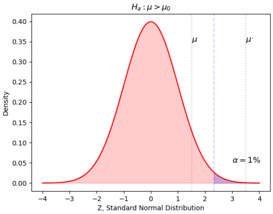
One-tail test



Two-tail test



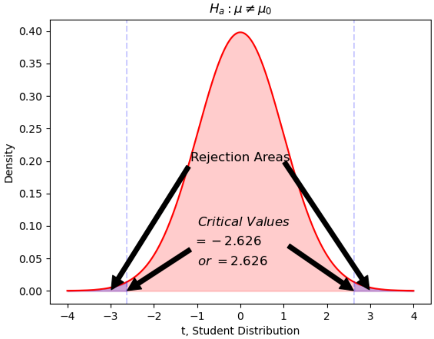
One-tail test



IV. Find the critical value

Example

- Find the right critical value
  - What is the correct table?
    - Since, in this case, the population variance is not available, the correct table is the Student table
    - Find $t_{\alpha/2}^{n-1} = t_{\alpha/2}^{n-1} = t_{0.005}^{99} = 2.626$ (and the negative opposite value because it is a two tail test)



V. Compare the test-statistics and the critical value

- If the test-statistic belongs to the rejection region, then H_0 is rejected
  - $if\ |test - statistic| > |critical\ value|$, then $reject\ H_0$
  - $if\ |test - statistic| < |critical\ value|$, then $fail\ to\ reject\ H_0$

- One can also use the $p - value$:
  - $if\ p - value < \alpha$, then $reject\ H_0$
  - $if\ p - value > \alpha$, then $fail\ to\ reject\ H_0$

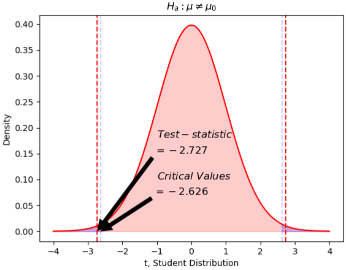
- Make sure to understand why these two comparisions boil down to the same thing

V. Compare the test-statistics and the critical value

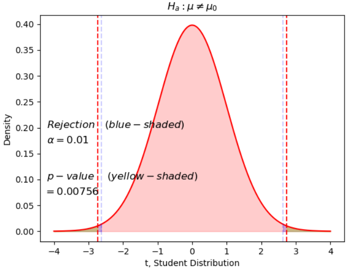
Example

- In the example
  - $|test - statistic| > |critical\ value|$
  - $p - value < \alpha$

(a) $test - statistic\ versus\ critical\ value$



(b) $p - value\ versus\ \alpha$



VI. Conclude

- H_0 is either **rejected** or **not rejected**
  - **Beware:** It is not about accepting H_0
- In the example, $|-2.727| > |-2.626|$ (and $0.00756 < 0.01$), thus H_0 is rejected
- Therefore, we conclude that, *at the significance level* 1%, we reject H_0, which implies that we conclude that the manufacturer misestimated the laptops' battery life.
  - Using a one-tail test, would you conclude that the manufacturer *overestimated* the laptops battery life?

Types of errors

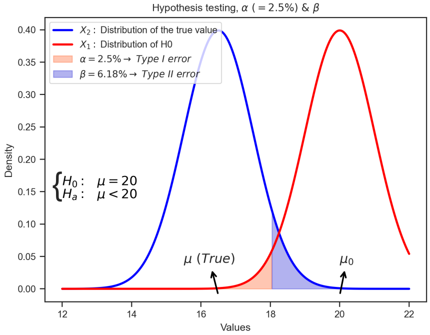
	H_0 is true	H_0 is false
Do not reject H_0 *(negative)*	Correct $(P(\cdot) = 1 - \alpha)$	Type II Error $(P(\cdot) = \beta)$
Reject H_0 *(positive)*	Type I Error $(P(\cdot) = \alpha)$	Correct $(P(\cdot) = 1 - \beta)$

- α is the level of significance, β is the power of the test

Types of errors

Illustration of the level of significance (α) and the power of the test (β)

- Assume you want to test whether a population average value (*e.g* age of first year student) is *lower* than 20 years old (μ_0).
  - Suppose that the true average (usually not observed) is $\mu = 16.5$, and set $\alpha = 5\%$



- Here, both distributions (call them X_1 and X_2) have $\sigma = 1$, hence

→ $\alpha = P(X_1 < 18.04) = P(Z < -1.96) = 2.5\%$

→ $\beta = P(X_2 > 18.04) = P(Z > 1.54) = 6.18\%$

Possible issues, be cautious!

- There can be issues due to the test itself, leading to type I or type II errors with some given probabilities
- There can also be issues because of the context of the test and the data gathering

Other points of caution

- Always bear in mind that the conclusion of your hypothesis might be due to other factors (**confounding factors**)
  - The clever Hans effect
    - Other actors might be involved in the change
    - If interested, see Samhita and Gross (2013)
  - The Hawthorne effect
    - Being observed/sampled changes your behavior
    - If interested, see Levitt and List (2011)
  - The placebo effect
    - It's not the drug but the information (the belief around its consumption)
    - If interested in an application in social sciences, see Jiménez-Buedo (2021)

Next session

- Next session is on univariate linear regression, basically adding up on everything we have seen so far.