Lesson 4. Hypothesis Testing – Part 1

1 The hypothesis testing framework

- The goal of **hypothesis testing** is to test competing claims about a parameter
- The general framework:
  1. State the hypotheses:
     - **Null hypothesis** H_0: nothing unusual is happening, no relationship exists, etc.
     - **Alternative hypothesis** H_A: something unusual is happening, some relationship exists, etc.
  2. Calculate the **test statistic**
  3. Calculate the p-**value**
  4. State your **conclusion**

2 t-test for one population mean

- Question: **Is an unknown population mean μ different from a specific value μ_0?**
- Three versions:

Two-tailed test:	Is μ different from μ_0?
Left-tailed test:	Is μ less than μ_0?
Right-tailed test:	Is μ greater than μ_0?

- Formal steps:
  1. State the hypotheses:

Two-tailed test:	$H_0 : \mu = \mu_0$	$H_A : \mu \neq \mu_0$
Left-tailed test:	$H_0 : \mu = \mu_0$	$H_A : \mu < \mu_0$
Right-tailed test:	$H_0 : \mu = \mu_0$	$H_A : \mu > \mu_0$

  2. Calculate the test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where

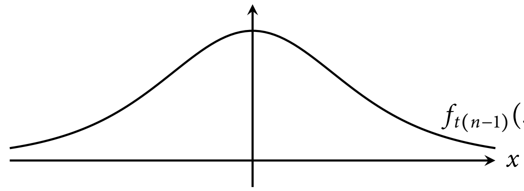
$$\bar{x} = \text{sample mean (estimate)}$$
$$s = \text{sample standard deviation (estimate)}$$
$$n = \text{sample size}$$

The test statistic measures how far the data is from the null hypothesis.

3. Calculate the p-value:

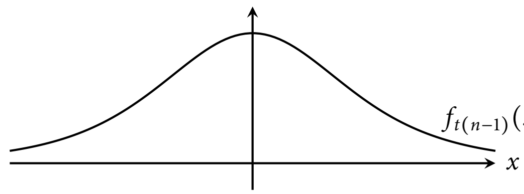
   If H_0 is true, the test statistic t follows a t-distribution with $n-1$ degrees of freedom.

   Two-tailed test:

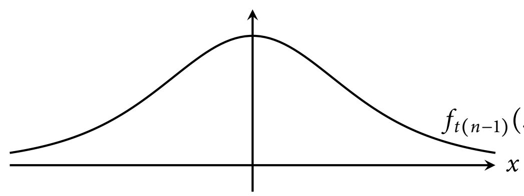
   $$p\text{-value} = P(t(n-1) \leq -|t|) + P(t(n-1) \geq |t|) = 2P(t(n-1) \geq |t|)$$

   Left-tailed test:

   $$p\text{-value} = P(t(n-1) \leq t)$$

   Right-tailed test:

   $$p\text{-value} = P(t(n-1) \geq t)$$

   The p-value is how likely our data could have occured, given that the null hypothesis is true.
   $\Rightarrow$ A small p-value is evidence against H_0.

4. State your conclusion, based on the given significance level α:

   If p-value $\leq \alpha$, we **reject** H_0:

   > At a significance level of α, we reject the null hypothesis. We see evidence that the population mean is different from/less than/greater than μ.

   If p-value $> \alpha$, we **fail to reject** H_0:

   > At a significance level of α, we fail to reject the null hypothesis. We do not see evidence that the population mean is different from/less than/greater than μ.

   The underlined parts above should be rephrased to correspond to the context of the problem.

- Technical conditions to check – this test may be used if the following conditions hold:

  1. Data must be from a simple random sample
  2. Either the population distribution is normal or the sample size (n) is large

Example 1. A Keurig machine is supposed to output 6 ounces of coffee when the smallest size is selected. For quality control, one machine is selected to be tested extensively to determine whether its average output is actually 6 ounces. The mean output of 20 cups of coffee is 6.1 ounces, and the standard deviation is 0.3 ounces. Use a significance level of 0.10 to test whether this machine's average output differs from 6 ounces.

Example 2. Prof. Moriarty reads a study saying a typical commute to work takes 30 minutes. She wants to test if her average commute time is greater than this value, at a significance level of 0.05. She records her commute time on 20 random days and calculates a mean of 31.5 minutes and a standard deviation of 2.9 minutes.

3 Statistical significance vs. practical importance

- "Significance" in the statistical sense ⟷ size of the p-value
- Suppose in Example 1, we concluded that the mean output was (statistically) significantly different from 6 ounces
- This means that the data is very unlikely to have occured by chance if $\mu = 6$
- It says nothing about the size of the difference

4 Type I and Type II errors

	Reject H_0	Fail to reject H_0
H_0 true		
H_0 false		

- We control the probability of a Type I error by setting the significance level α