Seminar 9 - 2025

Hypothesis Testing

A radio station *assumes* its average listener age is 30 years.

- **Null Hypothesis (H_0)**: $\mu = 30$.
- **Alternative Hypothesis (H_1)**: $\mu \neq 30$.

Process:

1. Collect a sample of listener ages and calculate the sample mean ($\bar{x}$).
2. Compare $\bar{x}$ to 30:
   - If $\bar{x}$ is "close" to 30, H_0 is supported.
   - If $\bar{x}$ is "not close" to 30, H_0 is rejected, and H_1 is supported.

Possible Outcomes:
Key Points:

- **Type I Error**: Rejecting H_0 when it is true (α).
- **Type II Error**: Accepting H_0 when H_1 is true (β).

Consider X to be some characteristic (relative to a population). Let $\bar{X}$ be the unbiased estimator for $\theta = \mu = \mathbb{E}(X)$ (where $\mathbb{E}(X)$ is theoretical mean of X), $\bar{X}$ has standard deviation $\frac{\sigma}{\sqrt{n}}$. As for confidence intervals, we start with the case where the test statistic has $N(0,1)$ distribution. If X has normal distribution or $n > 30$ we consider the test statistic

$$TS := Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

- For a given significance level $\alpha \in (0,1)$ and a given value $\theta_0 = \mu_0$, test:

$$H_0 : \mu = \mu_0$$

- Against one of the alternatives:

I. $H_1 : \mu \neq \mu_0$
II. $H_1 : \mu < \mu_0$
III. $H_1 : \mu > \mu_0$.

Given Data:

Decision	**Actual Situation**	
	H_0 true	H_1 true
Reject H_0	Type I error (prob. α)	Correct decision
Accept H_0	Correct decision	Type II error (prob. β)

- Parameters: $\alpha \in (0,1)$, μ_0, σ.
- Sample data: $x_1, \ldots, x_n$, where $X_1, \ldots, X_n$ are independent sample variables.
- Assumption: X is normally distributed or $n > 30$.

Steps:

- Under H_0:
$$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$
- Compute the test statistic using the data:
$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$
- Compare z to quantiles of the standard normal distribution $N(0,1)$.

A potential issue that can arise in hypothesis testing is as follows:

- We start by setting a significance level α, the probability of committing a type I error, and use this to define the rejection region (RR).
- When we compute the test statistic, its observed value does not fall within the rejection region, so we fail to reject the null hypothesis H_0. In other words, we accept H_0 as plausible.

However:

- When we calculate the probability of obtaining this value of the test statistic under the assumption that H_0 is true, we find this probability to be quite small—comparable to our preset α.
- This creates a paradox: we accept H_0, yet under the assumption that it is true, the observed value of the test statistic appears very unlikely.

Consider the following scenario:

- We test $H_0 : \mu = 50$ vs. $H_1 : \mu \neq 50$ at $\alpha = 0.05$ (two-tailed test).
- The rejection region (RR) is dictated by $|z| > 1.96$, based on the standard normal distribution, that is $RR = \{z \in \mathbb{R} : |z| \geq z_{1-\frac{0.05}{2}} = 1.96\} = (-\infty, -1.96] \cup [1.96, +\infty)$.

Observed Test Statistic: Assume $n = 36, \bar{x} = 48.7, \sigma = 10$. Then
$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = 1.9 \quad \text{(just outside the rejection region)}.$$

However:

- The probability of observing $z = 1.9$ under H_0 is (exercise!):
$$\mathbb{P}(|Z| > 1.9) = 0.058 \quad (\text{close to } \alpha = 0.05).$$
- Although $z \notin RR$, the probability is very small, making the observed value surprising under H_0.

Paradox:

- We fail to reject H_0, yet the observed test statistic seems unlikely if H_0 were true.

A **p-value** (probability value) is a measure used in hypothesis testing to assess the strength of the evidence against the null hypothesis (H_0) based on the sample data. It quantifies the probability of observing a test statistic at least as extreme as the one calculated from your sample, assuming that the null hypothesis is true.

Key Points:

- **Definition:** The p-value is the probability of obtaining results at least as extreme as the results actually observed, given that the null hypothesis is true.
- **Interpretation:**
  - **Small p-value** ($\leq \alpha$): Suggests strong evidence against the null hypothesis. This leads to **rejecting the null hypothesis** (H_0).
  - **Large p-value** ($> \alpha$): Suggests weak evidence against the null hypothesis. This leads to **failing to reject** the null hypothesis.

Example:
Suppose we are testing the hypothesis that a coin is fair (i.e., the probability of heads is 0.5).

- **Null hypothesis (H_0):** The coin is fair, $p = 0.5$.
- **Alternative hypothesis (H_1):** The coin is biased, $p \neq 0.5$.

After flipping the coin 100 times, we observe 60 heads. We calculate a test statistic (like a z-score or t-score), and we obtain a p-value of 0.03.

- **If the significance level (α) is 0.05:** Since $0.03 < 0.05$, the p-value is less than the threshold. This suggests strong evidence against the null hypothesis, and we would **reject** H_0, concluding that the coin may not be fair.
- **If the p-value had been 0.08:** Since $0.08 > 0.05$, we would **fail to reject** the null hypothesis, suggesting that there is not enough evidence to conclude that the coin is biased.

Formula for p-value:
In hypothesis testing, the p-value is calculated based on the chosen test (such as a z-test or t-test) and the observed data. For example, in a **two-tailed z-test**, the p-value is:

$$\text{p-value} = 2 \times \mathbb{P}(Z > |z_{\text{observed}}|)$$

where z_{observed} is the test statistic (z-score) calculated from the sample data, and $\mathbb{P}(Z > |z_{\text{observed}}|)$ is the probability of getting a test statistic at least as extreme as the observed one.

Summary:

- A **p-value** helps determine whether to reject or fail to reject the null hypothesis.
- It provides evidence for or against the null hypothesis in the context of your data.
- The smaller the p-value, the stronger the evidence against H_0.

Problem 1

A company claims that the average weight of its sugar bags is 1 kg. A customer suspects that the bags weigh less than advertised. The customer takes a sample of 30 bags and finds the mean weight is $\bar{x} = 0.98\,\text{kg}$, with a known population standard deviation $\sigma = 0.05\,\text{kg}$. We test this claim at the 5% significance level.

State the hypotheses

- Null hypothesis (H_0): The average weight is 1 kg.

$$H_0 : \mu = 1$$

- Alternative hypothesis (H_1): The average weight is less than 1 kg.

$$H_1 : \mu < 1$$

Data

$$n = 30, \quad \sigma = 0.05, \quad \bar{x} = 0.98, \quad \alpha = 0.05.$$

Compute the test statistic

Using the formula for the z-test:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$$

where $\mu_0 = 1$.

Substitute the values:

$$z = \frac{0.98 - 1}{0.05/\sqrt{30}}.$$

Calculate the standard error:

$$\text{SE} = \frac{0.05}{\sqrt{30}} \approx 0.00913.$$

Now compute z:

$$z = \frac{-0.02}{0.00913} \approx -2.19.$$

Find the critical value

For a left-tailed test at $\alpha = 0.05$, the critical value is:

$$z_\alpha = -1.645.$$

Make the decision

Since $z = -2.19$ and $z_\alpha = -1.645$:

$$z < z_\alpha.$$

We reject the null hypothesis H_0.

Conclusion

At the 5% significance level, there is sufficient evidence to conclude that the average weight of the bags is less than 1 kg.

Problem 2

Imagine you have a coin, and you suspect it may not be fair (i.e., it does not have a 50% chance of landing heads). You decide to test this hypothesis by flipping the coin **50 times** and recording the number of heads.

- **Null Hypothesis** (H_0): The coin is fair. The probability of getting heads is 50%.

$$H_0 : p = 0.5$$

- **Alternative Hypothesis** (H_1): The coin is not fair. The probability of getting heads is different from 50%.

$$H_1 : p \neq 0.5$$

Flip the coin $n = 50$ times and record the number of heads X.

Example outcome: After flipping, you observe $X = 35$ heads.

Choose a significance level $\alpha = 0.05$ (5%).

Since this is a discrete counting experiment, we use the **binomial test**. Under the null hypothesis (H_0), the number of heads X follows a binomial distribution:

$$X \sim \text{Binomial}(n = 50, p = 0.5).$$

We are testing whether the observed result $X = 35$ (or more extreme results) is consistent with the null hypothesis.

The p-value is the probability of observing results as extreme as $X = 35$ heads under H_0. Specifically, we calculate:

$$\mathbb{P}(X \geq 35) + \mathbb{P}(X \leq 15),$$

where $15 = n - 35$ due to symmetry. Using statistical tools or tables, the p-value is approximately:

$$\text{p-value} = 0.044.$$

- If p-value $\leq \alpha$, reject the null hypothesis H_0.
- If p-value $> \alpha$, do not reject H_0.

Here:

$$\text{p-value} = 0.044 \quad \text{and} \quad \alpha = 0.05.$$

Since $0.044 < 0.05$, we reject the null hypothesis H_0.

In this example, we use the **binomial test** to calculate the p-value for the hypothesis that a coin is fair (i.e., $H_0 : p = 0.5$) based on $n = 50$ coin flips and the observed $X = 35$ heads.

1. Formula for the p-value

The p-value represents the probability of observing a result as extreme or more extreme than the observed $X = 35$, under the null hypothesis H_0 that the coin is fair ($p = 0.5$). Because this is a **two-tailed test**, "extreme" results mean:

- $X \geq 35$ (more heads than expected)
- $X \leq 15$ (more tails than expected, by symmetry).

Thus, the p-value is given by:

$$\text{p-value} = \mathbb{P}(X \geq 35) + \mathbb{P}(X \leq 15),$$

where $X \sim \text{Binomial}(n = 50, p = 0.5)$.

2. Calculating the p-value

To calculate the probabilities $\mathbb{P}(X \geq 35)$ and $\mathbb{P}(X \leq 15)$:

1. **The Binomial Formula** For a binomial distribution $X \sim \text{Binomial}(n, p)$, the probability of observing exactly k successes is:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $\binom{n}{k}$ is the binomial coefficient.

2. **Cumulative Probabilities** To get $\mathbb{P}(X \geq 35)$, we sum up the probabilities of all values $X = 35, 36, \ldots, 50$:

$$\mathbb{P}(X \geq 35) = \sum_{k=35}^{50} \binom{50}{k} (0.5)^k (0.5)^{50-k}.$$

Similarly, $\mathbb{P}(X \leq 15)$ is the sum of probabilities for $X = 0, 1, \ldots, 15$:

$$\mathbb{P}(X \leq 15) = \sum_{k=0}^{15} \binom{50}{k} (0.5)^k (0.5)^{50-k}.$$

3. **Symmetry Property** Because the binomial distribution with $p = 0.5$ is symmetric, we can simplify:

$$\mathbb{P}(X \leq 15) = \mathbb{P}(X \geq 35).$$

Therefore, the total p-value is:

$$\text{p-value} = 2 \cdot \mathbb{P}(X \geq 35).$$

4. **Using Statistical Tools** Computing the exact probabilities by hand is tedious, so we use statistical tools (e.g., tables, software like Python, R, or a calculator) to find:

$$\mathbb{P}(X \geq 35) \approx 0.022.$$

Thus:

$$\text{p-value} = 2 \cdot 0.022 = 0.044.$$

Interpretation:

- The p-value 0.044 is compared to the significance level $\alpha = 0.05$.
- Since $0.044 < 0.05$, we reject the null hypothesis H_0.

This suggests that the coin is likely not fair.

Based on the data, there is sufficient evidence to reject the null hypothesis at the 5% significance level. This suggests the coin may not be fair.

Step	**Description**
1. Hypotheses	$H_0 : p = 0.5$, $H_1 : p \neq 0.5$
2. Experiment	$n = 50$, observed $X = 35$ heads
3. Significance Level	$\alpha = 0.05$
4. Test	Binomial Test
5. p-value	0.044
6. Decision	Reject H_0 because $0.044 < 0.05$.
7. Conclusion	Coin may not be fair.