Lecture 5:
Procedure for a Statistical Test

In statistics, **estimation** refers to the process by which one makes inferences about a population, based on information obtained from a sample.

Outline

- Procedure for a Statistical Test
  - Statistical Hypotheses: Null hypothesis, Alternative hypothesis
  - Decision Errors: Type I error, Type II error
  - Decision Rules: P-value, significance level (α)
  - One-Tailed and Two-Tailed Tests
  - Procedure of Hypothesis Testing

Statistical Hypotheses

There are two types of statistical hypotheses.

- **Null hypothesis**. The null hypothesis, denoted by H_o, is usually the hypothesis that sample observations result purely from chance.
- **Alternative hypothesis**. The alternative hypothesis, denoted by H_1, is the hypothesis that sample observations are influenced by some non-random cause.

Source: https://stattrek.com/

Statistical Hypotheses (continued)

- For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Symbolically, these hypotheses would be expressed as

$$H_o: P = 0.5$$
$$H_1: P \neq 0.5$$

- Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we would be inclined to reject the null hypothesis. We would conclude, based on the evidence, that the coin was probably unfair and unbalanced.

Source: https://stattrek.com/

Decision Errors

Two types of errors can result from a hypothesis test.

- **Type I error**. A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the **significance level**. This probability is also called **alpha**, and is often denoted by α.
- **Type II error**. A Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called **Beta**, and is often denoted by β. The probability of *not* committing a Type II error is called the **Power** of the test.

Source: https://stattrek.com/

Decision Rules

- If many experiments are conducted, then one can expect some of them to give an “abnormal” outcome.
- For a result to be worth reporting, the probability of any rejection error in the test (called **P-value**) should be less than the significance level $\boldsymbol{\alpha}$ (i.e., $P < \alpha$).
- **P-value**. The strength of evidence in support of a null hypothesis is measured by the **P-value**. Suppose the test statistic is equal to *S*. The P-value is the probability of observing a test statistic as extreme as *S*, assuming the null hypothesis is true. Then, the null hypothesis is:
  - Accepted if $P \geq \alpha$.
  - Rejected if $P < \alpha$.

Source: https://stattrek.com/

One-Tailed and Two-Tailed Tests

- A test of a statistical hypothesis, where the region of rejection is on only **one side** of the sampling distribution, is called a **one-tailed test**. For example, suppose the null hypothesis states that the mean is less than or equal to 10. The alternative hypothesis would be that the mean is greater than 10. The region of rejection would consist of a range of numbers located on the right side of sampling distribution; that is, a set of numbers greater than 10.
- A test of a statistical hypothesis, where the region of rejection is on **both sides** of the sampling distribution, is called a **two-tailed test**. For example, suppose the null hypothesis states that the mean is equal to 10. The alternative hypothesis would be that the mean is less than 10 or greater than 10. The region of rejection would consist of a range of numbers located on both sides of sampling distribution; that is, the region of rejection would consist partly of numbers that were less than 10 and partly of numbers that were greater than 10.

Source: https://stattrek.com/

Procedure of Hypothesis Testing

Define Hypothesis

- Null hypothesis H_0: e.g., the mean value = μ
- Alternative hypothesis H_1: e.g., the mean value $\neq \mu$

Perform hypothesis testing

- If the **population standard deviation is known**, then
  - Calculate the z-score from the data set: $z = \frac{\bar{x} - \mu}{\delta/\sqrt{n}}$
  - Use Normal Distribution calculator/table* to obtain the cumulative probability, i.e., P(Z < z).
  - Compute the P-value from $P(Z < z)$.
- If the **population standard deviation is unknown**, then
  - Calculate the t statistic from the data set: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
  - Use T Distribution calculator/table (http://www.ttable.org/) with degree of freedom = $n - 1$ to obtain the cumulative probability, i.e., $P(T < t)$.
  - Compute the P-value from $P(T < t)$.

Conclusion

- Accept the null hypothesis H_0 if $P \geq \alpha$.
- Reject the null hypothesis H_0 if $P < \alpha$.

* https://www.dummies.com/education/math/statistics/how-to-find-probabilities-for-z-with-the-z-table/

Example: Soft-drink vending machine

When you buy a large size Coke in McDonalds, the soft-drink vending machine is supposed to produce 30 ounces of Coke on average, Now, the machine was tested by taking a sample of 6 cups and we measure the amount of Coke in each cup. The measurement gave the following values (in ounce):

32.2, 30.3, 29.3, 34.0, 28.0, 32.2

The question: Is the machine correctly calibrated?

Sample mean of the data set is $\bar{x}$ = (32.2+30.3+29.3+34.0+28.0+32.2)/6 = 31, which is different from the target value μ = 30.

The next question: Is this difference statistically significant?

Example: Soft-drink machine (continued)

The sample standard deviation of the data set is

$$s = \sqrt{\frac{(32.2-31)^2+(30.3-31)^2+(29.3-31)^2+(34.0-31)^2+(28.0-31)^2+(32.2-31)^2}{5}}$$

$$= \sqrt{\frac{1.2^2+0.7^2+0.7^2+3^2+3^2+1.2^2}{5}} = \sqrt{\frac{21.86}{5}} = 2.09$$

When the null hypothesis (μ = 30) is correct, the t statistic is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{31-30}{2.09/\sqrt{6}} = 1.17$$

Example: Soft-drink machine (continued)

the *P*-value is the probability that the t statistic having 5 degrees of freedom is less than -1.17 or greater than 1.17.

We use the t Distribution Calculator to find P(t < -1.17) = 0.1474, and P(t > 1.17) = 0.1474. Thus, the P-value = 0.1474 + 0.1474 = 0.2948.

If we have α (significance level) set to 0.05 (5%), we accept the null hypothesis since $P \geq \alpha$. In other word, this means that the deviation of the measured data from the target mean is caused by random variation.

Conclusion: The soft-drink vending machine is calibrated correctly.

Test Your Understanding (Polling 16)

In this section, two sample problems illustrate how to conduct a hypothesis test of a mean score. The first problem involves a two-tailed test; the second problem, a one-tailed test.

Problem 1: Two-Tailed Test

An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for 5 hours (300 minutes) on a single gallon of regular gasoline. From his stock of 2000 engines, the inventor selects a simple random sample of 50 engines for testing. The engines run for an average of 295 minutes, with a standard deviation of 20 minutes. Test the null hypothesis that the mean run time is 300 minutes against the alternative hypothesis that the mean run time is not 300 minutes. Use a 0.05 level of significance. (Assume that run times for the population of engines are normally distributed.)

Source: https://stattrek.com/

Test Your Understanding (Polling 16)

Problem 1: Two-Tailed Test (continued)

Should we accept or reject the null hypothesis and why?

(A) Accept the null hypothesis since $P \geq \alpha$.
(B) Accept the null hypothesis since $P < \alpha$.
(C) Reject the null hypothesis since $P \geq \alpha$.
(D) Reject the null hypothesis since $P < \alpha$.
(E) None of the above.

Source: https://stattrek.com/

Test Your Understanding (Polling 16)

***Solution**:* **The answer is (A)**. The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

• **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

Null hypothesis: $\mu = 300$

Alternative hypothesis: $\mu \neq 300$

Note that these hypotheses constitute a two-tailed test. The null hypothesis will be rejected if the sample mean is too big or if it is too small.

• **Formulate an analysis plan**. For this analysis, the significance level is 0.05. The test method is a one-sample t-test.

Source: https://stattrek.com/

Test Your Understanding (Polling 16)

• **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the *t* statistic test statistic (*t*).

SE = *s* / sqrt(*n*) = 20 / sqrt(50) = 20/7.07 = 2.83

DF = *n* - 1 = 50 - 1 = 49

t = ($\bar{x}$ - μ) / SE = (295 - 300)/2.83 = -1.77

where *s* is the standard deviation of the sample, $\bar{x}$ is the sample mean, μ is the hypothesized population mean, and *n* is the sample size.

Since we have a two-tailed test, the *P*-value is the probability that the t statistic having 49 degrees of freedom is less than -1.77 or greater than 1.77.

We use the t Distribution Calculator to find P(t < -1.77) = 0.04, and P(t > 1.77) = 0.04. Thus, the P-value = 0.04 + 0.04 = 0.08.

• **Interpret results**. Since the *P*-value (0.08) is greater than the significance level (0.05), we accept the null hypothesis.

Source: https://stattrek.com/

Test Your Understanding

Problem 2: One-Tailed Test

Bon Air Elementary School has 1000 students. The principal of the school thinks that the average IQ of students at Bon Air is at least 110. To prove her point, she administers an IQ test to 20 randomly selected students. Among the sampled students, the average IQ is 108 with a standard deviation of 10. Based on these results, should the principal accept or reject her original hypothesis? Assume a significance level of 0.01. (Assume that test scores in the population of engines are normally distributed.)

Source: https://stattrek.com/

Test Your Understanding

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

• **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

Null hypothesis: $\mu \geq 110$

Alternative hypothesis: $\mu < 110$

Note that these hypotheses constitute a one-tailed test. The null hypothesis will be rejected if the sample mean is too small.

• **Formulate an analysis plan**. For this analysis, the significance level is 0.01. The test method is a one-sample t-test.

Source: https://stattrek.com/

Test Your Understanding

- **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the *t* statistic test statistic (*t*).

  SE = *s* / sqrt(*n*) = 10 / sqrt(20) = 10/4.472 = 2.236

  DF = *n* - 1 = 20 - 1 = 19

  t = ($\bar{x}$ - μ) / SE = (108 - 110)/2.236 = - 0.894

where s is the standard deviation of the sample, $\bar{x}$ is the sample mean, μ is the hypothesized population mean, and *n* is the sample size.

Here is the logic of the analysis: Given the alternative hypothesis (μ < 110), we want to know whether the observed sample mean is small enough to cause us to reject the null hypothesis.

The observed sample mean produced a t statistic test statistic of -0.894. We use the t Distribution Calculator to find P(t < -0.894) = 0.19. This means we would expect to find a sample mean of 108 or smaller in 19 percent of our samples, if the true population IQ were 110. Thus the P-value in this analysis is 0.19.

- **Interpret results**. Since the P-value (0.19) is greater than the significance level (0.01), we cannot reject the null hypothesis.

Source: https://stattrek.com/

Test Your Understanding (Polling 17)

Problem 3: One-Tailed Test

An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously **at least** for 5 hours (300 minutes) on a single gallon of regular gasoline. From his stock of 2000 engines, the inventor selects a simple random sample of 50 engines for testing. The engines run for an average of 295 minutes, with a standard deviation of 20 minutes. Test the null hypothesis that the mean run time is 300 minutes against the alternative hypothesis that the mean run time is not 300 minutes. Use a 0.05 level of significance. (Assume that run times for the population of engines are normally distributed.)

Source: https://stattrek.com/

Test Your Understanding (Polling 17)

Problem 3: One-Tailed Test (continued)

Should we accept or reject the null hypothesis and why?

(A) Accept the null hypothesis since $P \geq \alpha$.
(B) Accept the null hypothesis since $P < \alpha$.
(C) Reject the null hypothesis since $P \geq \alpha$.
(D) Reject the null hypothesis since $P < \alpha$.
(E) None of the above.

Source: https://stattrek.com/

Test Your Understanding (Polling 17)

Solution: **The answer is (D).** The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

• **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

Null hypothesis: $\mu \geq 300$

Alternative hypothesis: $\mu < 300$

Note that these hypotheses constitute a one-tailed test. The null hypothesis will be rejected if the sample mean is too small.

• **Formulate an analysis plan**. For this analysis, the significance level is 0.05. The test method is a one-sample t-test.

Source: https://stattrek.com/

Test Your Understanding (Polling 17)

- **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the *t* statistic test statistic (*t*).

  SE = *s* / sqrt(*n*) = 20 / sqrt(50) = 20/7.07 = 2.83

  DF = *n* - 1 = 50 - 1 = 49

  t = ($\bar{x}$ - μ) / SE = (295 - 300)/2.83 = - 1.77

where *s* is the standard deviation of the sample, $\bar{x}$ is the sample mean, μ is the hypothesized population mean, and *n* is the sample size.

Since we have a one-tailed test, the *P*-value is the probability that the t statistic having 49 degrees of freedom is less than - 1.77.

We use the t Distribution Calculator to find P(t < -1.77) = 0.04. Thus, the *P*-value = 0.04.

- **Interpret results**. Since the *P*-value (0.04) is less than the significance level (0.05), we reject the null hypothesis.

Source: https://stattrek.com/

EE1004 Teaching and Learning Survey (11 Feb 2022)

https://cityuhk.questionpro.com/t/AR91DZrCsV