Lecture 4: Estimation in Statistics & 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

- Estimation in Statistics
    - Confidence interval: confidence level, sample statistic and margin of error.
- 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

Estimation in Statistics

Point Estimate vs. Interval Estimate

Statisticians use **sample** **statistics** to estimate **population** **parameters**. For example, sample means are used to estimate population means.

An estimate of a population parameter may be expressed in two ways:

- **Point estimate**. A point estimate of a population parameter is a single value of a sample statistic. For example, the sample mean $\bar{x}$ is a point estimate of the population mean μ.
- **Interval estimate**. An interval estimate is defined by two numbers, between which a population parameter is said to lie. For example, $a < \bar{x} < b$ is an interval estimate of the population mean μ. It indicates that the population mean is greater than a but less than b.

Source: https://stattrek.com/

Confidence Intervals

Statisticians use a **confidence interval** to express the precision/error and certainty/uncertainty associated with a sample estimate of a population parameter (e.g., the population mean). A confidence interval consists of three parts.

- A confidence level.
- A sample statistic.
- A margin of error.

The certainty/uncertainty associated with the **confidence interval** is specified by the **confidence level**. The **sample statistic** and the **margin of error** define an interval estimate that describes the precision of the sampling method. The interval estimate of a confidence interval is defined by the *sample statistic ± margin of error*.

For example, suppose we compute an interval estimate of a population parameter. We might choose a **95% confidence level** to describe the interval estimate as a **95% confidence interval**. This means that if we used the same sampling method to select different samples and compute different interval estimates, with **95% of the time** the true population parameter would fall within a range defined by the *sample statistic ± margin of error* .

Source: https://stattrek.com/

Confidence Level

- The probability part of a confidence interval is called a **confidence level**. The confidence level describes the likelihood that a particular sampling method will produce a confidence interval that includes the true population parameter.
- Here is how to interpret a confidence level. Suppose we collected all possible samples from a given population, and computed confidence intervals for each sample. Some confidence intervals would include the true population parameter; others would not. A 95% confidence level means that with 95% of the time/chance the intervals contain the true population parameter (e.g., the population mean); a 90% confidence level means that with 90% of the time/chance the intervals contain the population parameter; and so on.

Source: https://stattrek.com/

Margin of Error (1)

- In a confidence interval, the range of values above and below the sample statistic is called the **margin of error**.
- For example, suppose the local newspaper conducts an election survey and reports that the independent candidate will receive 30% of the vote. The newspaper states that the survey had a 5% margin of error and a confidence level of 95%. These findings result in the following confidence interval: We are 95% confident that the independent candidate will receive between 25% and 35% of the vote.
- Note: Many public opinion surveys report interval estimates, but not confidence intervals. They provide the margin of error, but without the confidence level. To clearly interpret survey results you need to know both! We are much more likely to accept survey findings if the confidence level is high (say, 95%) than if it is low (say, 50%).

Source: https://stattrek.com/

Test Your Understanding (Polling 14)

Problem 1

Which of the following statements is true.

I. When the margin of error is small, the confidence level is high.
II. When the margin of error is small, the confidence level is low.
III. A confidence interval is a type of point estimate.
IV. A population mean is an example of a point estimate.

(A) I only
(B) II only
(C) III only
(D) IV only.
(E) None of the above.

Source: https://stattrek.com/

Test Your Understanding (Polling 14)

Solution

The correct answer is (E). The confidence level is not affected by the margin of error. When the margin of error is small, the confidence level can low or high or anything in between. A confidence interval is a type of interval estimate, not a type of point estimate. A *population* mean is not an example of a point estimate; a *sample* mean is an example of a point estimate.

Source: https://stattrek.com/

Margin of Error (2)

- Recall that in a confidence interval, the range of values above and below the sample statistic is called the **margin of error**.
- For example, suppose we wanted to know the percentage of adults that exercise daily. We could devise a sample design to ensure that our sample estimate will not differ from the true population value by more than, say, 5 percent (the margin of error) 90 percent of the time (the confidence level).

Source: https://stattrek.com/

How to Compute the Margin of Error

- The margin of error can be defined by either of the following equations.
  - Margin of error = Critical value x Population standard deviation of the statistic / sqrt(n)
  - Margin of error = Critical value x Sample standard deviation of the statistic / sqrt(n)
- Note that Sample/Population standard deviation of the statistic / sqrt(n) is called **Standard error**.
- If you know the population standard deviation of the statistic, use the first equation to compute the margin of error. Otherwise, use the second equation.

Source: https://stattrek.com/

How to Find the Critical Value

The **critical value** is a factor used to compute the margin of error. Here, we describe how to find the critical value, when the sampling distribution of the statistic is normal or nearly normal.

When the sampling distribution is nearly normal, the critical value can be expressed as a t statistic (or t-score) or as a z-score. To find the critical value, follow these steps.

- Compute alpha (α): α = 1 - (confidence level / 100)
- Find the critical probability (p*): p* = 1 - α/2

Source: https://stattrek.com/

How to Find the Critical Value (continued)

- To express the critical value as a z-score, find the z-score having a cumulative probability equal to the critical probability (p*).
- To express the critical value as a t statistic, follow these steps.
  - Find the degrees of freedom (DF). When estimating a mean from a single sample, DF is equal to the sample size minus one.
  - The critical t statistic (t*) is the t statistic having degrees of freedom equal to DF (i.e., the sample size minus one) and a cumulative probability equal to the critical probability (p*).

Source: https://stattrek.com/

T-Score (t statistic) vs. Z-Score

Should you express the critical value as a t statistic or as a z-score? One way to answer this question focuses on the population standard deviation.

- If the **population standard deviation is known**, use the **z-score**.
- If the **population standard deviation is unknown**, use the **t statistic**.

You can use the Normal Distribution Calculator (https://stattrek.com/online-calculator/normal.aspx) to find the critical z-score, and the t Distribution Calculator (https://stattrek.com/online-calculator/t-distribution.aspx) to find the critical t statistic.

Source: https://stattrek.com/

Test Your Understanding (Polling 15)

Problem

Nine hundred (900) high school freshmen were randomly selected for a national survey. Among survey participants, the sample mean grade-point average (GPA) was 2.7, and the sample standard deviation was 0.4. What is the margin of error, assuming a 95% confidence level?

(A) 0.013
(B) 0.025
(C) 0.500
(D) 1.960
(E) None of the above.

Source: https://stattrek.com/

Test Your Understanding (Polling 15)

Solution

The correct answer is (B). To compute the margin of error, we need to find the critical value and the standard error of the mean. To find the critical value, we take the following steps.

- Compute alpha (α): α = 1 - (confidence level / 100) = 1 - 0.95 = 0.05
- Find the critical probability (p*): p* = 1 - α/2 = 1 - 0.05/2 = 0.975
- Find the degrees of freedom (df): df = n - 1 = 900 -1 = 899
- Find the critical value. Since we don't know the population standard deviation, we'll express the critical value as a t statistic. For this problem, it will be the t statistic having 899 degrees of freedom and a cumulative probability equal to 0.975. Using the t Distribution Calculator, we find that the critical value is 1.96.

Source: https://stattrek.com/

Test Your Understanding (Polling 15)

Solution

Next, we find the standard error of the mean ($SE\bar{x}$), using the following equation:

$SE\bar{x}$ = s / sqrt(n) = 0.4 / sqrt(900) = 0.4 / 30 = 0.013

And finally, we compute the margin of error (ME).

ME = Critical value x Standard error = 1.96 * 0.013 = 0.025

This means we can be 95% confident that the mean grade point average in the population is 2.7 plus or minus 0.025, since the margin of error is 0.025.

Note: The larger the sample size, the more closely the t distribution looks like the normal distribution. For this problem, since the sample size is very large, we would have found the same result with a z-score as we found with a t statistic. That is, the critical value would still have been 1.96. The choice of t statistic versus z-score does not make much practical difference when the sample size is very large.

Source: https://stattrek.com/

Procedure for a Statistical Test

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: $z = \left| \frac{\bar{x} - \mu}{\delta/\sqrt{n}} \right|$
  - Use Normal Distribution Calculator 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: $t = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right|$
  - Use T Distribution Calculator 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 $P < \alpha$.

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)

Apply Student's t-test with t statistic = 1.17 and degree of freedom = 5, we have the cumulative probability

$$\mathrm{P}(\mathrm{T} < 1.17) = 0.8526$$

This is a two-tailed test so the *P*-value is obtained from

$$P = 2[1 - \mathrm{P}(\mathrm{T} < 1.17)] = 0.2944$$

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

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

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: $\mu = 300$

    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

- **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 = $|(\bar{x} - \mu) / SE|$ = |(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) = 0.96. Thus, the *P*-value = 2[1- P(t < 1.77)] = 0.08.

- **Interpret results**. Since the *P*-value (0.08) is greater than the significance level (0.05), we cannot reject 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 = $|(\bar{x} - \mu) / SE|$ = |(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) = 0.81. Hence, the P-value = 1-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/

EE1004 Teaching and Learning Survey (30 March 2021)

https://cityuhk.questionpro.com/t/AR91DZltjA