Lecture 4: Estimation in Statistics

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
    - margin of error

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* $\pm$ *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* $\pm$ *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): $\alpha = 1 - (\text{confidence level} / 100)$
- Find the critical probability (p^*): $p^* = 1 - \alpha/2$

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/score. Alternatively, you can find the critical t statistic or z-score from T Table (http://www.ttable.org/) for calculating confidence interval (https://www.youtube.com/watch?v=DcWATePtE1s).

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/

EE1004 Teaching and Learning Survey (28 January 2022)

https://cityuhk.questionpro.com/t/AR91DZq3YY