Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT19

Föreläsning 8

Nancy Abdallah

Chalmers - Göteborgs Universitet

May 7, 2019

Interval Estimate - Confidence Intervals

If we draw a random sample of size n from a population, then $\overline{x}$ is a point estimate of μ. It is more meaningful to estimate μ by an interval that we call confidence interval.

How to construct a Confidence Interval?

Sampling from a normally distributed population

Suppose that the population is normally distributed with mean μ and variance σ^2, and let $X_1, \ldots, X_n$ be a random sample.

- The mean $\overline{X}$ is normally distributed with mean μ and variance σ^2/n.
- The interval $(\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}})$ contains 95% of the sample means. *(If we draw a sample of size n, the probability that its mean falls into the interval* $(\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}})$ *is 0.95).*
  We have seen this result with 2 instead of 1.96. 1.96 is more precise).
- μ is usually unknown. We use intervals involving the sample mean $\overline{x}$ instead.
- 95% of the intervals $(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\frac{\sigma}{\sqrt{n}})$ contain the mean μ. This interval is called a 95% Confidence Interval (C.I.).

Example

Suppose a researcher, interested in obtaining an estimate of the average level of some enzyme in a certain human population, takes a sample of 10 individuals, determines the level of the enzyme in each, and computes a sample mean of $\overline{x} = 22$. Suppose further it is known that the variable of interest is approximately normally distributed with a variance of 45. Find a 95% interval confidence of μ.

Solution We Need to find the interval $(\overline{x} - 1.96\sigma/\sqrt{n}, x + 1.96\sigma/\sqrt{n})$. $\sigma/\sqrt{n} = \sqrt{45/10} = 2.1213$, then the interval is

$$(22 - 1.96(2.1213), 22 + 1.96(2.1213))$$

which is

$$(17.84, 26.16)$$

Confidence interval of $\overline{x}$ from a normally distributed population with known σ

In general, a $100(1-\alpha)$ confidence interval is given by

$$\overline{x} \pm z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}$$

where $z_{(1-\alpha/2)}$ is the value of z for which $P(z \leq z_{(1-\alpha/2)}) = 1 - \frac{\alpha}{2}$ in the standard normal distribution and is called the **the reliability coefficient.**
In the book, $z_{\alpha/2}$ is used instead, and it means the value of z for which $P(z \geq z_{\alpha/2}) = 1 - \alpha$.
The C.I. can be written as

$$\text{estimator} \pm (\text{reliability coefficient}) \times (\text{standard error})$$

Interpretation

- **Probabilistic Interpretation:** In repeated sampling, from a normally distributed population with a known standard deviation, $100(1-\alpha)$ percent of all intervals of the form $\overline{x} \pm z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}$ will in the long run include the population mean μ.
- **Practical Interpretation:** When sampling is from a normally distributed population with known standard deviation, we are $100(1-\alpha)$ percent confident that the single computed interval, $\overline{x} \pm z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}$, contains the population mean μ.

Example

A physical therapist wished to estimate, with 99 percent confidence, the mean maximal strength of a particular muscle in a certain group of individuals. He is willing to assume that strength scores are approximately normally distributed with a variance of 144. A sample of 15 subjects who participated in the experiment yield a mean of 84.3.

Solution $\alpha = 0.01$. We need to find $z_{1-\alpha/2}$, i.e. z such that $P(z < z_{1-\alpha/2}) = 0.995$. Using the table, we get $z_{1-\alpha/2} = 2.58$. Therefore, a 99% confidence interval is

$$(84.3 - (2.58)\sqrt{144/15}, 84.3 + (2.58)\sqrt{144/155})$$

which is

$$(76.3, 92.3)$$

Sampling from non-normal distribution

Usually the distribution of the population is not normal or is not even known. How do we construct a confidence interval in this case?

Definition (The central limit theorem)

Let $X_1, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, $\overline{X}$ is approximately normal with mean μ and variance σ^2/n.

Remark n is considered large if $n \geq 25$.

Example

Punctuality of patients in keeping appointments is of interest to a research team. In a study of patient flow through the offices of general practitioners, it was found that a sample of 35 patients was 17.2 minutes late for appointments, on the average. Previous research had shown the standard deviation to be about 8 minutes. The population distribution was felt to be nonnormal.

1. Approximate the distribution of the sample mean $\overline{X}$. What are the parameters?
2. What is the 90 percent confidence interval for μ, the true mean amount of time late for appointments?

Example (Solution)

1. Since $n \geq 25$ then $\overline{X}$ is approximately normally distributed with mean equals to the population mean μ (unknown) and standard deviation $\sigma/\sqrt{n} = 8/\sqrt{35} = 1.3522$
2. We need to find first $z_{1-\alpha/2} = z_{0.95}$. Using the table of standard normal distribution we get $z_{0.95=1.645}$. Hence, a 90% confidence interval is given by

$$(17.2 - 1.645(1.3522), 17.2 + 1.645(1.3522))$$

which is

$$(15.0, 19.4)$$

Distribution for population variance

Let $X_1, \ldots, X_n$ be a random sample. Recall that $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ is an unbiased estimate for the variance σ^2.

Theorem

Suppose that $X_1, \ldots, X_n$ are drawn from a normal distribution with mean μ and variance σ^2. The random variable

$$(n-1)S^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/\sigma^2$$

has a chi-squared distribution with $n-1$ degrees of freedom.

Confidence Interval for population variance

- To find a $100(1-\alpha)\%$ C.I. for σ^2, we find first a $100(1-\alpha)\%$ C.I. for $(n-1)S^2/\sigma^2$.
  A C.I. for $(n-1)S^2/\sigma^2$ is given by
  $$(\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2})$$
  i.e. $P(\chi^2_{\alpha/2} \leq (n-1)S^2/\sigma^2 \leq \chi^2_{1-\alpha/2}) = 1-\alpha$
- By using operation on inequalities, we obtain the C.I on σ^2.
  A $100(1-\alpha)\%$ C.I. for σ^2 is then
  $$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right)$$

Example

Given the following data

$$9.7,\ 12.3,\ 11.2,\ 5.1,\ 24.8,\ 14.8,\ 17.7$$

Find a 95% confidence interval for the variance and for the standard deviation.

Solution The sample variance is $s^2 = 39.763$. The degrees of freedom are $n - 1 = 6$ and $\alpha = 0.05$. From the table, we get $\chi^2_{0.025} = 1.237$ and $\chi^2_{0.975} = 14.449$. Hence, our 95% C.I for σ^2 is

$$\left(\frac{6(39.763)}{14.449}, \frac{6(39.763)}{1.237}\right) = (16.512, 192.868)$$

and the 95% C.I. for σ is

$$(4.063, 13.888)$$

Student T-distribution

- Let Z and X be independent random variables such that $Z \in N(0,1)$ and $X \in \chi^2_\gamma$. Then

$$T = \frac{Z}{\sqrt{\chi^2/\gamma}}$$

is said to have a Student T distribution with γ degrees of freedom.

- Properties of T-distribution:
  - The graph of the density function is a bell-shape symmetrical about the mean which is equal to 0.
  - It approaches the normal distribution as n increases.
  - The variance of the t distribution is greater than 1. If df > 2, the variance is equal to $\frac{df}{df-2}$ where *df* is the degrees of freedom.

Confidence interval of $\overline{x}$ from a normally distributed population with unknown σ

Let $X_1, \ldots, X_n$ be a random sample drawn from a normally distributed population with unknown μ and σ.

- It is easy to show that $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ follows a T distribution with $n-1$ degrees of freedom.
- To find a $100(1-\alpha)\%$ C.I., we use the same method as for the normal distribution, i.e.

$$\overline{x} \pm t_{(1-\alpha/2)} s/\sqrt{n}$$

  where $\overline{x}$ is the mean of the sample, s its standard deviation, and $t_{(1-\alpha/2)}$ is the value of t for which $P(t < t_{(1-\alpha/2)}) = 1-\alpha/2$.
- The values of t can be found using table VI p.699-700.

Example

A set of data on the sulfur dioxide concentration (in micrograms per cubic meter) in a Bavarian forest has been collected. The sample contains 24 values, it mean is 53.92 $\mu g/m^3$, the sample variance is 101.480 and the standard deviation is $10.07\mu g/m^3$. To obtain a 95% C.I. we find $t_{(1-\alpha/2)}$ from the table. $t_{0.975} = 2.069$. Hence, a 95% C.I. is given by

$$(53.93 - 2.069(10.07)/\sqrt{24}, 53.93 + 2.069(10.07)/\sqrt{24})$$

which is equal to

$$(49.67, 58.17)$$

Summary - C.I. for the population mean