Chapter 7

Statistical Intervals Based on a Single Sample

Confidence Intervals

An alternative to reporting a single value for the parameter being estimated is to calculate and report an entire interval of plausible values – a *confidence interval* (CI). A *confidence level* is a measure of the degree of reliability of the interval.

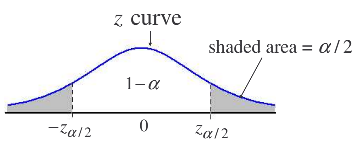
7.1 Basic Properties of Confidence Intervals

95% Confidence Interval

If after observing $X_1 = x_1, \ldots, X_n = x_n$, we compute the observed sample mean $\bar{x}$, then a *95% confidence interval* for μ can be expressed as

$$\left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Other Levels of Confidence



$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1 - \alpha$$

Other Levels of Confidence

A $100(1-\alpha)\%$ confidence interval for the mean μ of a normal population when the value of α is known is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Sample Size

The general formula for the sample size n necessary to ensure an interval width w is

$$n = \left(2z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$

Deriving a Confidence Interval

Let $X_1, \ldots, X_n$ denote the sample on which the CI for the parameter θ is to be based. Suppose a random variable satisfying the following properties can be found:

1. The variable depends functionally on both $X_1, \ldots, X_n$ and θ.
2. The probability distribution of the variable does not depend on θ or any other unknown parameters.

Deriving a Confidence Interval

Let $h(X_1, \ldots, X_n; \theta)$ denote this random variable. In general, the form of h is usually suggested by examining the distribution of an appropriate estimator $\hat{\theta}$. For any α between 0 and 1, constants a and b can be found to satisfy

$$P(a < h(X_1, \ldots, X_n; \theta) < b) = 1 - \alpha$$

$$P(l(X_1, \ldots, X_n)) < \theta < u(X_1, \ldots, X_n))$$

Deriving a Confidence Interval

Now suppose that the inequalities can be manipulated to isolate θ:

$$P(l(X_1, \ldots, X_n)) < \theta < u(X_1, \ldots, X_n))$$

lower confidence limit

upper confidence limit

For a $100(1-\alpha)\%$ CI.

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

Large-Sample Confidence Interval

If n is sufficiently large, the standardized variable

$$Z = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

has approximately a standard normal distribution. This implies that

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

is a large-sample confidence interval for μ with level $100(1-\alpha)\%$.

Confidence Interval for a Population Proportion p with level $100(1-\alpha)\%$

Lower(–) and upper(+) limits:

$$= \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \left(z_{\alpha/2}^2\right)/n}$$

Large-Sample Confidence Bounds for μ

Upper Confidence Bound:

$$\mu < \bar{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

Lower Confidence Bound:

$$\mu > \bar{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

7.3 Intervals Based on a Normal Population Distribution

Normal Distribution

The population of interest is normal, so that $X_1, \ldots, X_n$ constitutes a random sample from a normal distribution with both μ and σ unknown.

t Distribution

When $\overline{X}$ is the mean of a random sample of size n from a normal distribution with mean μ, the rv

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

has a probability distribution called a *t distribution* with n – 1 degrees of freedom (df).

Properties of t Distributions

Let t_v denote the density function curve for v df.

1. Each t_v curve is bell-shaped and centered at 0.
2. Each t_v curve is spread out more than the standard normal (z) curve.

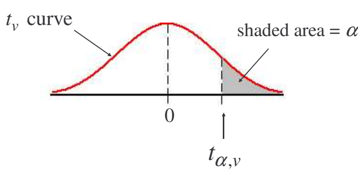
Properties of t Distributions

3. As v increases, the spread of the corresponding t_v curve decreases.
4. As $v \to \infty$, the sequence of t_v curves approaches the standard normal curve (the z curve is called a t curve with df = ∞.

t Critical Value

Let $t_{\alpha,v}$ = the number on the measurement axis for which the area under the t curve with v df to the right of $t_{\alpha,v}$ is α; $t_{\alpha,v}$ is called a *t critical value*.

Pictorial Definition of $t_{\alpha,v}$



Confidence Interval

Let $\overline{x}$ and s be the sample mean and standard deviation computed from the results of a random sample from a normal population with mean μ. The $100(1-\alpha)\%$ confidence interval is

$$\left(\overline{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2}, n-1 \cdot \frac{s}{\sqrt{n}}\right)$$