STATISTICAL INFERENCES (2cr)
Chapter 9 One- and Two-Sample Estimation Problems

Zhong Guan

Math, IUSB

Outline

Outline

1. 9.5 Standard Error of a Point Estimate
   - Standard Error of $\bar{X}$
2. 9.6 Prediction Intervals
   - Use Prediction Interval to Detect Outlier
3. 9.7 Tolerance Interval(Limits)

Outline

Outline

When σ is Known.

If σ is known, the **standard error of point estimator $\bar{X}$** is the standard deviation of the sampling distribution of $\overline{X}$:

$$\text{s.e.}(\overline{X}) = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

In this case, the $100(1-\alpha)$**% confidence interval (limits) of μ is written as:**

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \bar{x} \pm z_{\alpha/2}\text{s.e.}(\bar{x})$$

When σ is Known.

If σ is known, the **standard error of point estimator $\bar{X}$** is the standard deviation of the sampling distribution of $\overline{X}$:

$$\text{s.e.}(\overline{X}) = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

In this case, the $100(1-\alpha)$**% confidence interval (limits) of μ is written as:**

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \bar{x} \pm z_{\alpha/2}\text{s.e.}(\bar{x})$$

When σ is Not Known.

If σ is not known, the **standard error of point estimator $\bar{X}$** is the standard deviation of the sampling distribution of $\overline{X}$:

$$\text{s.e.}(\overline{X}) = \sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

In this case, the **$100(1-\alpha)\%$ confidence interval (limits) of μ is written as:**

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2}(n-1)\text{s.e.}(\bar{x})$$

When σ is Not Known.

If σ is not known, the **standard error of point estimator $\bar{X}$** is the standard deviation of the sampling distribution of $\overline{X}$:

$$\text{s.e.}(\overline{X}) = \sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

In this case, the $100(1-\alpha)$**% confidence interval (limits) of μ is written as:**

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2}(n-1)\text{s.e.}(\bar{x})$$

An Example

Example 1: Predict a Future Value Due to the decrease in interest rates, the First Citizen Bank received a lot of mortgage applications. A recent sample of 50 mortgage loans resulted in an average of $257,300. Assume a population standard deviation of $25,000, If the next customer called in for mortgage loan application, what would be the next customer's loan amount?

Solution

Denote sample mean of of 50 mortgage loans by $\bar{X}$ and the loan amount of the next customer by X_0, then X_0 and $\bar{X}$ are independent normal random variables.

$$X_0 \sim N(\mu, \sigma^2), \quad \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

So

$$X_0 - \overline{X} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$$

$$Z = \frac{X_0 - \overline{X}}{\sqrt{\sigma^2 + \frac{\sigma^2}{n}}} \sim N(0, 1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Solution

Denote sample mean of of 50 mortgage loans by $\bar{X}$ and the loan amount of the next customer by X_0, then X_0 and $\bar{X}$ are independent normal random variables.

$$X_0 \sim N(\mu, \sigma^2), \quad \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

So

$$X_0 - \overline{X} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$$

$$Z = \frac{X_0 - \overline{X}}{\sqrt{\sigma^2 + \frac{\sigma^2}{n}}} \sim N(0, 1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Solution

Denote sample mean of of 50 mortgage loans by $\bar{X}$ and the loan amount of the next customer by X_0, then X_0 and $\bar{X}$ are independent normal random variables.

$$X_0 \sim N(\mu, \sigma^2), \quad \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

So

$$X_0 - \overline{X} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$$

$$Z = \frac{X_0 - \overline{X}}{\sqrt{\sigma^2 + \frac{\sigma^2}{n}}} \sim N(0, 1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Solution

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2}\right) = 1-\alpha$$

Solving inequalities for X_0, we have

$$P\left(\overline{X} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < X_0 < \overline{X} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}}\right) = 1-\alpha$$

$\bar{x} = \$257,300$, $n = 50$, and $\sigma = \$25,000$. Choose $\alpha = 0.05$, so that $1-\alpha = 0.95$ and $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} = (\$207,812, \$306,787)$$

$$P(\$207,812 < X_0 < \$306,787) = 95\%$$

Solution

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2}\right) = 1-\alpha$$

Solving inequalities for X_0, we have

$$P\left(\overline{X} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < X_0 < \overline{X} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}}\right) = 1-\alpha$$

$\bar{x} = \$257,300$, $n = 50$, and $\sigma = \$25,000$. Choose $\alpha = 0.05$, so that $1-\alpha = 0.95$ and $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} = (\$207,812, \$306,787)$$

$$P(\$207,812 < X_0 < \$306,787) = 95\%$$

Solution

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2}\right) = 1-\alpha$$

Solving inequalities for X_0, we have

$$P\left(\overline{X} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < X_0 < \overline{X} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}}\right) = 1-\alpha$$

$\bar{x} = \$257,300$, $n = 50$, and $\sigma = \$25,000$. Choose $\alpha = 0.05$, so that $1-\alpha = 0.95$ and $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} = (\$207,812, \$306,787)$$

$$P(\$207,812 < X_0 < \$306,787) = 95\%$$

Solution

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2}\right) = 1-\alpha$$

Solving inequalities for X_0, we have

$$P\left(\overline{X} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < X_0 < \overline{X} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}}\right) = 1-\alpha$$

$\bar{x} = \$257,300$, $n = 50$, and $\sigma = \$25,000$. Choose $\alpha = 0.05$, so that $1-\alpha = 0.95$ and $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} = (\$207,812, \$306,787)$$

$$P(\$207,812 < X_0 < \$306,787) = 95\%$$

Solution

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2} \right) = 1-\alpha$$

Solving inequalities for X_0, we have

$$P\left(\overline{X} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < X_0 < \overline{X} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} \right) = 1-\alpha$$

$\bar{x} = \$257,300$, $n = 50$, and $\sigma = \$25,000$. Choose $\alpha = 0.05$, so that $1-\alpha = 0.95$ and $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} = (\$207,812, \$306,787)$$

$$P(\$207,812 < X_0 < \$306,787) = 95\%$$

Prediction Intervals

- If σ is known, given the observed mean $\bar{x}$ of a sample of size n, the $100(1-\alpha)\%$ **prediction interval** for the next value x_0 is

$$\bar{x} - z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}}$$

- If σ is unknown, given the observed mean $\bar{x}$ of a sample of size n, the $100(1-\alpha)\%$ **prediction interval** for the next value x_0 is

$$\bar{x} - t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}}$$

Prediction Intervals

- If σ is known, given the observed mean $\bar{x}$ of a sample of size n, the $100(1-\alpha)\%$ **prediction interval** for the next value x_0 is

$$\bar{x} - z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}}$$

- If σ is unknown, given the observed mean $\bar{x}$ of a sample of size n, the $100(1-\alpha)\%$ **prediction interval** for the next value x_0 is

$$\bar{x} - t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}}$$

Example with Unknown σ

Example 2. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed. If a sample of 30 bulbs has an average life of 780 hours and a standard deviation of 41 hours, find a 99% prediction interval for the next observed life of a bulb manufactured by this firm.

Solution $\bar{x} = 780$, $n = 30$, and $s = 41$. $\alpha = 0.01$, $1 - \alpha = 0.99$ and $t_{\frac{\alpha}{2}}(n-1) = 2.756$.

$$\bar{x} \pm t_{\frac{\alpha}{2}}(n-1)s\sqrt{1 + \frac{1}{n}} = (665.12, 894.88)$$

So the $100(1-\alpha)$**% prediction interval** for the next observed life is

$$(665.12, 894.88)$$

Example with Unknown σ

Example 2. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed. If a sample of 30 bulbs has an average life of 780 hours and a standard deviation of 41 hours, find a 99% prediction interval for the next observed life of a bulb manufactured by this firm.

Solution $\bar{x} = 780$, $n = 30$, and $s = 41$. $\alpha = 0.01$, $1 - \alpha = 0.99$ and $t_{\frac{\alpha}{2}}(n-1) = 2.756$.

$$\bar{x} \pm t_{\frac{\alpha}{2}}(n-1)s\sqrt{1 + \frac{1}{n}} = (665.12, 894.88)$$

So the $100(1-\alpha)$% **prediction interval** for the next observed life is

$$(665.12, 894.88)$$

Example with Unknown σ

Example 2. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed. If a sample of 30 bulbs has an average life of 780 hours and a standard deviation of 41 hours, find a 99% prediction interval for the next observed life of a bulb manufactured by this firm.

Solution $\bar{x} = 780$, $n = 30$, and $s = 41$. $\alpha = 0.01$, $1 - \alpha = 0.99$ and $t_{\frac{\alpha}{2}}(n-1) = 2.756$.

$$\bar{x} \pm t_{\frac{\alpha}{2}}(n-1)s\sqrt{1 + \frac{1}{n}} = (665.12, 894.88)$$

So the $100(1-\alpha)$**% prediction interval** for the next observed life is

$$(665.12, 894.88)$$

Outline

Use Prediction Interval to Detect Outlier

Given a observed mean $\bar{x}$ of a sample of size n from a normal population, the next observed value x_0 is an **outlier**, if x_0 falls outside the $100(1-\alpha)\%$ **prediction interval**

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

Note: This method is a parametric or large sample method because we assume that either the population is normal or the sample size is large. The outlier detection method using Box-and-Whisker plot is a nonparametric method.

Use Prediction Interval to Detect Outlier

Given a observed mean $\bar{x}$ of a sample of size n from a normal population, the next observed value x_0 is an **outlier**, if x_0 falls outside the $100(1-\alpha)\%$ **prediction interval**

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

Note: This method is a parametric or large sample method because we assume that either the population is normal or the sample size is large. The outlier detection method using Box-and-Whisker plot is a nonparametric method.

Example

Example 1. A manufacturer is producing metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. Find an interval so that we are 99% confidence that this interval contains at least 95% of the metal pieces produced by this machine, assuming an approximate normal distribution.

Solution of Example 1.

Solution: If both μ and σ are known, then we are 100% confident that 95% of the metal pieces produced by this machine are contained in

$$(\mu - 1.96\sigma, \quad \mu + 1.96\sigma)$$

This is a tolerance interval.

Solution of Example 1.

Now, both μ and σ were unknown, we have to estimate them by $\bar{x}$ and s:

$$\bar{x} = \frac{1.01 + 0.97 + 1.03 + 1.04 + 0.99 + 0.98 + 0.99 + 1.01 + 1.03}{9}$$

$$= \frac{9.05}{9} = 1.005556,$$

$$\sum x_i^2 = 1.01^2 + 0.97^2 + 1.03^2 + 1.04^2 + 0.99^2 + 0.98^2 + 0.99^2 + 1.01^2 + 1.03^2$$

$$= 9.1051$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{8}[9.1051 - (9)(1.005556^2)] = 0.0006017722$$

$$s = \sqrt{0.0006017722} = 0.02453105$$

Solution of Example 1.

Now, both μ and σ were unknown, we have to estimate them by $\bar{x}$ and s:

$$\bar{x} = \frac{1.01 + 0.97 + 1.03 + 1.04 + 0.99 + 0.98 + 0.99 + 1.01 + 1.03}{9}$$

$$= \frac{9.05}{9} = 1.005556,$$

$$\sum x_i^2 = 1.01^2 + 0.97^2 + 1.03^2 + 1.04^2 + 0.99^2 + 0.98^2 + 0.99^2 + 1.01^2 + 1.03^2$$

$$= 9.1051$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{8}[9.1051 - (9)(1.005556^2)] = 0.0006017722$$

$$s = \sqrt{0.0006017722} = 0.02453105$$

Solution of Example 1.

Now, both μ and σ were unknown, we have to estimate them by $\bar{x}$ and s:

$$\bar{x} = \frac{1.01 + 0.97 + 1.03 + 1.04 + 0.99 + 0.98 + 0.99 + 1.01 + 1.03}{9}$$

$$= \frac{9.05}{9} = 1.005556,$$

$$\sum x_i^2 = 1.01^2 + 0.97^2 + 1.03^2 + 1.04^2 + 0.99^2 + 0.98^2 + 0.99^2 + 1.01^2 + 1.03^2$$

$$= 9.1051$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{8}[9.1051 - (9)(1.005556^2)] = 0.0006017722$$

$$s = \sqrt{0.0006017722} = 0.02453105$$

Solution of Example 1.

Now, both μ and σ were unknown, we have to estimate them by $\bar{x}$ and s:

$$\bar{x} = \frac{1.01 + 0.97 + 1.03 + 1.04 + 0.99 + 0.98 + 0.99 + 1.01 + 1.03}{9}$$

$$= \frac{9.05}{9} = 1.005556,$$

$$\sum x_i^2 = 1.01^2 + 0.97^2 + 1.03^2 + 1.04^2 + 0.99^2 + 0.98^2 + 0.99^2 + 1.01^2 + 1.03^2$$

$$= 9.1051$$

$$s^2 = \tfrac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \tfrac{1}{8}[9.1051 - (9)(1.005556^2)] = 0.0006017722$$

$$s = \sqrt{0.0006017722} = 0.02453105$$

Solution of Example 1.

If μ and σ unknown, then the tolerance interval (limits) is

$$\bar{x} \pm ks,$$

where k is determined (See Table A.7) so that one can be $100(1-\gamma)\%$ confident that the given interval contains at least $100(1-\alpha)\%$ of all the measurements.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by Table A.7, $k = 4.55$.
So the 99% tolerance limits are

$$1.005556 \pm (4.55)(0.0246) = (0.978, 1.033)$$

A tolerance interval is an estimate of the "middle" 95% of the normal population.

Solution of Example 1.

If μ and σ unknown, then the tolerance interval (limits) is

$$\bar{x} \pm ks,$$

where k is determined (See Table A.7) so that one can be $100(1-\gamma)\%$ confident that the given interval contains at least $100(1-\alpha)\%$ of all the measurements.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by Table A.7, $k = 4.55$.
So the 99% tolerance limits are

$$1.005556 \pm (4.55)(0.0246) = (0.978, 1.033)$$

A tolerance interval is an estimate of the "middle" 95% of the normal population.

Solution of Example 1.

If μ and σ unknown, then the tolerance interval (limits) is

$$\bar{x} \pm ks,$$

where k is determined (See Table A.7) so that one can be $100(1-\gamma)\%$ confident that the given interval contains at least $100(1-\alpha)\%$ of all the measurements.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by Table A.7, $k = 4.55$.
So the 99% tolerance limits are

$$1.005556 \pm (4.55)(0.0246) = (0.978, 1.033)$$

A tolerance interval is an estimate of the "middle" 95% of the normal population.

Solution of Example 1.

If μ and σ unknown, then the tolerance interval (limits) is

$$\bar{x} \pm ks,$$

where k is determined (See Table A.7) so that one can be $100(1-\gamma)\%$ confident that the given interval contains at least $100(1-\alpha)\%$ of all the measurements.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by Table A.7, $k = 4.55$.
So the 99% tolerance limits are

$$1.005556 \pm (4.55)(0.0246) = (0.978, 1.033)$$

A tolerance interval is an estimate of the "middle" 95% of the normal population.

Approximation of Tolerance Factor

The tolerance factor k can also be approximated by

$$k = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right) z^2_{\alpha/2}}{\chi^2_{\gamma}(n-1)}}$$

If $n \to \infty$, then $k \to z_{\alpha/2}$.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by the above formula we have, $k = 4.554$.
So the 99% tolerance limits are

$$1.005556 \pm (4.554)(0.0246) = (0.978, 1.033)$$

Approximation of Tolerance Factor

The tolerance factor k can also be approximated by

$$k = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right) z^2_{\alpha/2}}{\chi^2_{\gamma}(n-1)}}$$

If $n \to \infty$, then $k \to z_{\alpha/2}$.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by the above formula we have, $k = 4.554$.
So the 99% tolerance limits are

$$1.005556 \pm (4.554)(0.0246) = (0.978, 1.033)$$

Approximation of Tolerance Factor

The tolerance factor k can also be approximated by

$$k = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right) z_{\alpha/2}^2}{\chi_{\gamma}^2(n-1)}}$$

If $n \to \infty$, then $k \to z_{\alpha/2}$.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by the above formula we have, $k = 4.554$.
So the 99% tolerance limits are

$$1.005556 \pm (4.554)(0.0246) = (0.978, 1.033)$$