MAT2377

Ali Karimnezhad

Version December 15, 2016

Comments

- These slides cover material from Chapter 5.
- In class, I may use a blackboard. I recommend reading these slides before you come to the class.
- I am planning to spend 2-3 lectures on this chapter.
- I am not re-writing the textbook. The reference book contains many interesting and practical examples.
- There may be some typos. The final version of the slides will be posted $after$ the chapter is finished.

Point Estimation

A point estimate of some population parameter θ is a single value $\hat{\theta}$ of a statistic Θ.

- if $X \sim B(n, p)$, what is $\hat{p}$?
- if $X \sim N(\mu, 1)$, what is $\hat{\mu}$?
- An estimator is not expected to estimate the population parameter without error.
- We are making decisions.

Unbiased Estimator

What are the desirable properties of a good decision that would influence us to choose one estimator rather than another?

- Let $\hat{\Theta}$ be an estimator whose value $\hat{\theta}$ is a point estimate of some unknown population parameter θ. Certainly, we would like the sampling distribution of $\hat{\Theta}$ to have a mean equal to the parameter estimated.

- A statistic $\hat{\Theta}$ is said to be an unbiased estimator of the parameter θ if

$$E(\hat{\Theta}) = \theta.$$

Examples:

If a sample $X_1, \ldots, X_n$ has unknown population mean and variance μ and σ^2, respectively, show that

- the sample mean $\bar{X}$ is an unbiased estimator for μ;
- the sample variance S^2 is an unbiased estimator for σ^2.

Efficient Estimators

- Suppose $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two unbiased estimators of the same population parameter θ. If

$$Var(\hat{\Theta}_1) < Var(\hat{\Theta}_2),$$

  we say that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$.

- If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the most efficient estimator of θ.

Examples:

If X_1, X_2 are two independent random variables having unknown population mean and variance μ and σ^2, respectively, which one of the following estimators of μ is more efficient?

$$T_1 = \frac{1}{2}(X_1 + X_2) \qquad T_2 = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

Interval Estimation

There are many situations in which it is preferable to determine an interval within which we would expect to find the value of the parameter.

An interval estimate of a population parameter θ is an interval of the form

$$\hat{\theta}_L < \theta < \hat{\theta}_U,$$

where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\hat{\Theta}$ for a particular sample and also on the sampling distribution of $\hat{\Theta}$.

Ideally, we prefer a short interval with a high degree of confidence!

Interval Estimation

- We find $\hat{\theta}_L$ and $\hat{\theta}_U$ such that

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha, \quad \text{for } 0 < \alpha < 1.$$

  Then, we have a probability of $1 - \alpha$ of selecting a random sample that will produce an interval containing θ.

- The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ is called a $100(1 - \alpha)\%$ confidence interval,

- the fraction $1 - \alpha$ is called the confidence coefficient or the degree of confidence,

- the endpoints, $\hat{\theta}_L$ and $\hat{\theta}_U$ are called the lower and upper confidence limits.

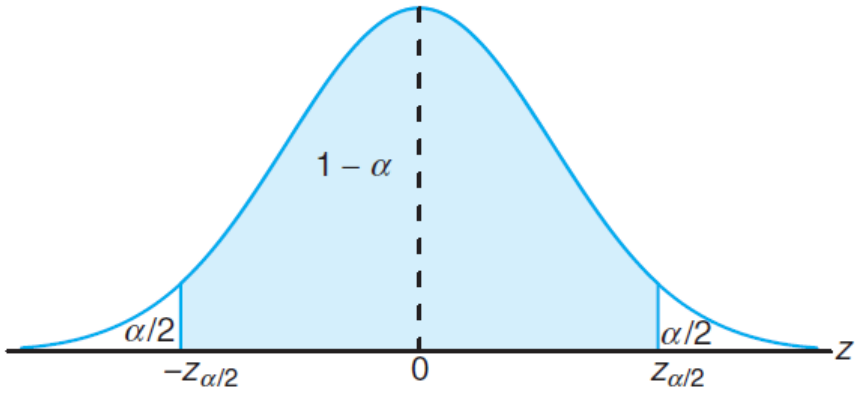
Constructing Two-Sided Confidence Interval for Mean

- Remember sampling distribution of $\bar{X}$, i.e., $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
- We find $\hat{\mu}_L$ and $\hat{\mu}_U$ such that

$$P(\hat{\mu}_L < \mu < \hat{\mu}_U) = 1 - \alpha,$$

$$\vdots$$

$$P\left(\frac{\bar{X} - \hat{\mu}_U}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{\bar{X} - \hat{\mu}_L}{\sigma/\sqrt{n}}\right) = 1 - \alpha.$$



- Take $\frac{\bar{X}-\hat{\mu}_L}{\sigma/\sqrt{n}} = z_{\frac{\alpha}{2}}$ and $\frac{\bar{X}-\hat{\mu}_U}{\sigma/\sqrt{n}} = -z_{\frac{\alpha}{2}}$.
- Thus, $\hat{\mu}_L = \bar{X} - z_{\frac{\alpha}{2}}\,\sigma/\sqrt{n}$ and $\hat{\mu}_U = \bar{X} + z_{\frac{\alpha}{2}}\,\sigma/\sqrt{n}$.

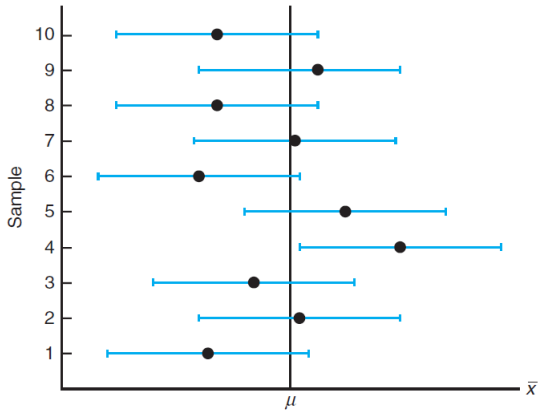
$$P(\bar{X} - z_{\frac{\alpha}{2}}\sigma/\sqrt{n} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\sigma/\sqrt{n}) = 1 - \alpha.$$

Two-Sided Confidence Interval on μ, σ^2 is Known

If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{x} - z_{\frac{\alpha}{2}}\sigma/\sqrt{n} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\sigma/\sqrt{n},$$

where $z_{\frac{\alpha}{2}}$ is the z-value leaving an area of $\frac{\alpha}{2}$ to the right.



Example:

The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint:

3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8

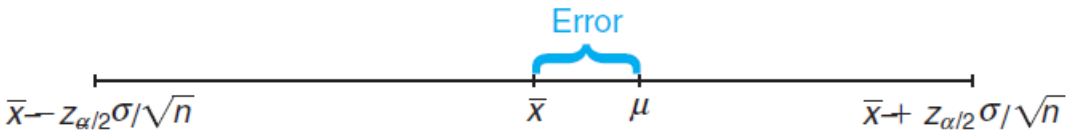
- Assuming that the measurements represent a random sample from a normal population with unity standard deviation, find a 95% interval for the average drying time in the population.

- What is maximum of the estimation error?

Sample Size Determination

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

$$n = \left(\frac{z_{\frac{\alpha}{2}}\,\sigma}{e}\right)^2$$



When solving for the sample size, n, we round all fractional values up to the next whole number. So, if we get 24, 24.1, 24.9 or 25, we report 25 as the sample size.

Example: How large a sample is required if we want to be 95% confident that our estimate of μ in our last Example is off by less than 0.05?

Constructing One-Sided Confidence Interval for Mean

- We find $\hat{\mu}_L$ and $\hat{\mu}_U$ such that

$$P(\mu > \hat{\mu}_L) = 1 - \alpha, \qquad P(\mu < \hat{\mu}_U) = 1 - \alpha,$$

$$\vdots$$

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{\bar{X} - \hat{\mu}_L}{\sigma/\sqrt{n}}\right) = 1 - \alpha, \qquad P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{\bar{X} - \hat{\mu}_U}{\sigma/\sqrt{n}}\right) = 1 - \alpha.$$

- Take $\frac{\bar{X} - \hat{\mu}_L}{\sigma/\sqrt{n}} = z_\alpha$ and $\frac{\bar{X} - \hat{\mu}_U}{\sigma/\sqrt{n}} = -z_\alpha$.

$$P(\mu > \bar{X} - z_\alpha\, \sigma/\sqrt{n}) = 1 - \alpha, \qquad P(\mu < \bar{X} + z_\alpha\, \sigma/\sqrt{n}) = 1 - \alpha.$$

Example:

The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint:

3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8

Assuming that the measurements represent a random sample from a normal population with unity standard deviation, find an upper 95% bound for the average drying time in the population.

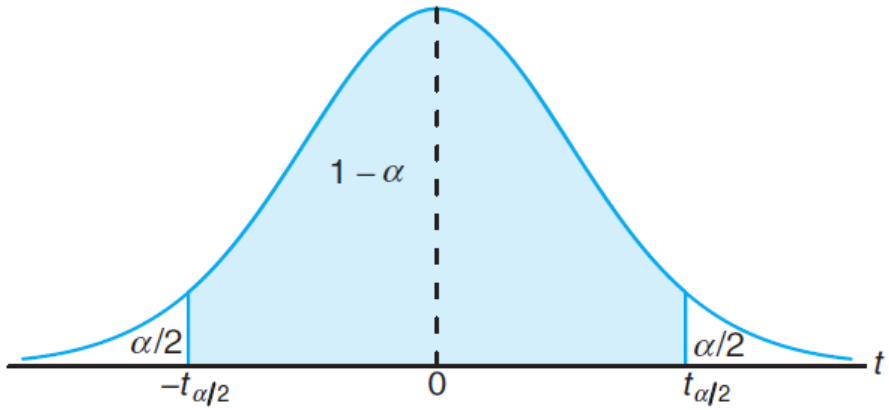
Two-Sided Confidence Interval on μ, σ^2 is Unknown

- Remember sampling distribution of $\bar{X}$, i.e., $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
- Remember sampling distribution of S^2, i.e., $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
- Remember $\bar{X}$ and S^2 are independent and $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$.
- We find $\hat{\mu}_L$ and $\hat{\mu}_U$ such that

$$P(\hat{\mu}_L < \mu < \hat{\mu}_U) = 1 - \alpha,$$

$$\vdots$$

$$P\left(\frac{\bar{X} - \hat{\mu}_U}{S/\sqrt{n}} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < \frac{\bar{X} - \hat{\mu}_L}{S/\sqrt{n}}\right) = 1 - \alpha.$$



- Take $\frac{\bar{X}-\hat{\mu}_L}{S/\sqrt{n}} = t_{\frac{\alpha}{2}}(n-1)$ and $\frac{\bar{X}-\hat{\mu}_U}{S/\sqrt{n}} = -t_{\frac{\alpha}{2}}(n-1)$.
- Thus, $\hat{\mu}_L = \bar{X} - t_{\frac{\alpha}{2}}(n-1)\, S/\sqrt{n}$ and $\hat{\mu}_U = \bar{X} + t_{\frac{\alpha}{2}}(n-1)\, S/\sqrt{n}$.

$$P(\bar{X} - t_{\frac{\alpha}{2}}(n-1)S/\sqrt{n} < \mu < \bar{X} + t_{\frac{\alpha}{2}}(n-1)S/\sqrt{n}) = 1-\alpha.$$

One-Sided Confidence Interval on μ, σ^2 is Unknown

- We find $\hat{\mu}_L$ and $\hat{\mu}_U$ such that

$$P(\mu > \hat{\mu}_L) = 1 - \alpha, \qquad P(\mu < \hat{\mu}_U) = 1 - \alpha,$$

$$\vdots$$

$$P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} < \frac{\bar{X} - \hat{\mu}_L}{S/\sqrt{n}}\right) = 1 - \alpha, \qquad P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} > \frac{\bar{X} - \hat{\mu}_U}{S/\sqrt{n}}\right) = 1 - \alpha.$$

- Take $\frac{\bar{X} - \hat{\mu}_L}{S/\sqrt{n}} = t_\alpha(n-1)$ and $\frac{\bar{X} - \hat{\mu}_U}{S/\sqrt{n}} = -t_\alpha(n-1)$.

$$P(\mu > \bar{X} - t_\alpha(n-1)\, S/\sqrt{n}) = 1 - \alpha, \qquad P(\mu < \bar{X} - t_\alpha(n-1)\, S/\sqrt{n}) = 1 - \alpha.$$

Example:

The contents of seven similar containers of sulfuric acid in liter are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6.

Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

Confidence Interval for Proportion

If $X \sim B(n, p)$, then the point estimator, $\hat{p}$, for p is $\hat{p} = \frac{X}{n}$. Also,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately standard normal. Thus, for large n,

$$P\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

Similar to the confidence interval construction for Normal distribution, the *approximate* $(1-\alpha)$ confidence interval for p is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}.$$

But, this is a bit useless, so the *approximate* $(1-\alpha)$ confidence interval for p is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Example:

In a random sample of $n = 500$ families owning television sets in the city of Hamilton, Canada, it is found that $x = 340$ subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with television sets in this city that subscribe to HBO.

Prediction Intervals

- Sometimes, other than the population mean, the experimenter may also be interested in predicting the possible value of a future observation.

  Suppose $X_1, X_2, \ldots, X_n, X_{n+1}$ are $n+1$ independent random variables having the Normal distribution with unknown mean μ and known variance σ^2. We observe $x_1, x_2, \ldots, x_n$. Find a $100(1-\alpha)\%$ confidence interval for the next observation x_{n+1}.

$$X_{n+1} - \bar{X} \sim N(0, \sigma^2(1 + \frac{1}{n})),$$

  Thus

$$Z = \frac{X_{n+1} - \bar{X}}{\sigma^2(1 + \frac{1}{n})} \sim N(0, 1).$$

Two-Sided Prediction Interval of a Future Observation

For a Normal distribution of measurements with unknown mean μ and variance σ^2, a $100(1-\alpha)\%$ prediction interval of a future observation x_{n+1}

- if σ^2 is known, is given by

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

  where $z_{\frac{\alpha}{2}}$ is the z-value leaving an area of $\frac{\alpha}{2}$ to the right.

- if σ^2 is unknown, is given by

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

  where $t_{\frac{\alpha}{2}}(n-1)$ is the t-value with $\nu = n-1$ degrees of freedom, leaving an area of $\frac{\alpha}{2}$ to the right.

Example:

Due to the decrease in interest rates, the First Citizens Bank received a lot of mortgage applications. A recent sample of 50 mortgage loans resulted in an average loan amount of $257,300. Assume a population standard deviation of $25,000. For the next customer who fills out a mortgage application, find a 95% prediction interval for the loan amount.

Example:

A meat inspector has randomly selected 30 packs of 95% lean beef. The sample resulted in a mean of 96.2% with a sample standard deviation of 0.8%. Find a 99% prediction interval for the leanness of a new pack. Assume normality.

One-Sided Prediction Interval of a Future Observation

For a Normal distribution of measurements with unknown mean μ and variance σ^2, a $100(1-\alpha)\%$ prediction interval of a future observation x_{n+1}

- if σ^2 is known, is given by

$$?!?,$$

  where z_α is the z-value leaving an area of α to the right.

- if σ^2 is unknown, is given by

$$?!?,$$

  where $t_\alpha(n-1)$ is the t-value with $\nu = n-1$ degrees of freedom, leaving an area of α to the right.