MAT 2377 3X (Spring 2011)
Interval Estimation - Sections 8-1 to 8-3 and 8-5

Recall: The observed value of an estimator for an unknown parameter is called a point estimate. For example, if the sample mean is $\overline{x} = 35$, then the latter is a point estimate for the population mean μ.

Remark: The point estimate tells us nothing about the accuracy of the estimate. That is, how close is $\overline{x}$ of μ? One way to answer the question of accuracy is to construct an interval of values such that we will be highly confident that μ will be in this interval.

General Idea for Confidence Intervals: Suppose that θ is some unknown parameter and that we are using the statistic $\widehat{\Theta} = h(X_1, \ldots, X_n)$ as an estimator for θ. Suppose furthermore, that we can find two statistics

$$L = L(X_1, \ldots, X_n) \quad \text{and} \quad U = U(X_1, \ldots, X_n)$$

(usually based upon $\widehat{\Theta}$) such that

$$1 - \alpha = P(L \leq \theta \leq U),$$

where α is a constant between 0 and 1. Usually $1 - \alpha$ will be a large value, e.g. 90%, 95% or 99%.

Let l and u be the observed values of L and U, respectively.

We say that $[l, u]$ is a $100\,(1 - \alpha)\%$ **confidence interval** for θ.

Terminology:

1. $1 - \alpha$ is called the confidence level or confidence coefficient.
2. α is called the error rate.
3. $u - l$=length of the interval is a measure of the precision of the estimate. A smaller length interval is interpreted as a more precise estimation.

Interpretation: Suppose that $1 - \alpha = .95$ and the from the n observations we get the following confidence interval $[10, 15]$ for μ. We say that we are 95% **confident** that $10 \leq \mu \leq 15$. What does that mean?

We do not know if μ is between 10 and 15, but the technique that we used to produce the interval yields an interval that contains μ 95% of the time.

Quantiles: When constructing confidence intervals the following quantile notation will be useful.

Definition: Let Z be a standard normal random variable. Its upper quantile of order A is a value z_A such that

$$P(Z > z_A) = A,$$

that is the area under the prob. density function of Z to the right of z_A is A.

Some common quantiles for the standard normal:

A	z_A
.25	.674
.10	1.282
.05	1.645
.025	1.96
.01	2.326
.005	2.576

Estimating μ (σ known)

Assumptions:

- Population is normal or the sample size is large ($n \geq 30$).
- Population variance σ^2 is known.

Under these assumptions,

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

follows a standard normal distribution. It is approximate in the case of a non-normal population with a large sample size. It follows that

$$\begin{aligned} 1 - \alpha &= P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha}\right) \\ &= P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

Therefore, if the population is normal or $n \geq 30$, a $100\,(1-\alpha)\%$ confidence interval for μ is

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Example 1: Consider a random sample of $n = 10$ observations that yield a mean of 64.46. Assume that the population is normal with standard deviation $\sigma = 4.5$.

(a) Construct a 95% C.I. for μ.

(b) Construct a 98% C.I. for μ.

(c) Compare the precision of the two intervals.

Precision:

The precision of the interval is defined as the length of the interval, that is

$$\overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - (\overline{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}) = 2\,\frac{z_{\alpha/2}\sigma}{\sqrt{n}}.$$

Remarks:

- The precision is a function of the confidence level and of the sample size.
- As we increase the confidence level, the estimation is less precise.
- As we increase the sample size, we increase precision.
- In practice, we would like high precision and high confidence. To do so, we can fix the confidence level, and then choose the appropriate sample size to control the precision of the estimate.

Sample Size:

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

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2.$$

Remarks:

- If n that you compute is not an integer, then it must be rounded-up to the nearest integer.
- If σ is unknown, then it is common practice to collect a preliminary sample and use the sample standard deviation s instead of σ in the formula to compute n.

Example 2:
Consider a population with a standard deviation of $\sigma = 10$. Suppose that we wanted to be 95% confident that the error in estimating the true mean μ with the sample mean $\overline{x}$ is less than 0.3. What sample size should be used?

One-Sided Confidence Intervals

We sometimes want a lower-bound (or upper bound) for μ. We can do this with a one-sided confidence interval.

If the population is normal or the sample size is large ($n \geq 30$), then
(i) a $100\,(1-\alpha)\%$ upper-confidence bound for μ is

$$\mu \leq \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}};$$

(ii) a $100\,(1-\alpha)\%$ lower-confidence bound for μ is

$$\mu \geq \overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Example 3: Consider the scenario from Example 1.
(a) Find a 95% upper-confidence bound for μ.
(b) Give an interpretation for this bound.

Large-Sample Confidence Interval for μ

Recall that by the Central Limit Theorem, if n is large (i.e. $n \geq 30$), then

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

follows approximately a $N(0, 1)$ distribution. In practice, we do not know σ, so we use the sample standard deviation S instead. It can be shown that the limiting distribution, i.e. as $n \to \infty$ is still $N(0, 1)$ even if we use S instead of σ.

Rule of Thumb: If $n \geq 40$, then

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

follows approximately a $N(0, 1)$ distribution.

Therefore, when n is large ($n \geq 40$), a $100(1 - \alpha)\%$ confidence interval for μ is

$$\overline{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}.$$

Example 4: Suppose that a sample of 45 observations yielded a mean of 24.5 and a standard deviation of 2.2. Give a 95% confidence interval for the true mean μ.

Estimating the mean of a normal population (σ unknown)

In the special case of a normal population, it is possible to construct a C.I. for the mean even when σ is unknown.

Definition: Consider the random variable T with probability density function

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\,\nu}\Gamma(\nu/2)} \left[\frac{x^2}{\nu} + 1\right]^{-(\nu+1)/2}, \quad -\infty < t < \infty.$$

We say that T follows a t distribution with ν degrees of freedom.

Properties: Consider a t distribution with ν degrees of freedom.

- Its upper quantile of order α is denoted $t_{\alpha,\nu}$. We find them in Table V.
- The density $f(t)$ is symmetric about $t = 0$, hence

$$t_{1-\alpha,\nu} = -t_{\alpha,\nu}.$$

- When $\nu \to \infty$, then the t distribution is a standard normal, hence

$$z_\alpha = t_{\alpha,\infty}.$$

  So we can use table V to find quantiles for the $N(0,1)$ distribution. For example, $z_{.025} = t_{.025,\infty} = 1.96$.

Theorem: Let $X_1, \ldots, X_n$ is a random sample from a **normal** population with mean μ and variance σ^2. The random variable

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

has a t distribution with $\nu = n - 1$ degrees of freedom.

Therefore, if the population is normal, then a $100\,(1-\alpha)\%$ confidence interval for μ is

$$\overline{x} \pm t_{\alpha/2,n-1}\,\frac{s}{\sqrt{n}}.$$

Example 5: Consider the scenario from Example 1, but suppose that σ^2 is unknown. The sample of 10 observations yielded a mean of 64.46 and a standard deviation of $s = 4.7$. Construct a 95% confidence interval for μ.

Estimating a population proportion p:

Consider the sample proportion $\widehat{P} = X/n$, where X is the number of successes among the n trials.

X follows a binomial distribution with parameters n and p. If n is large then X follows approximately a normal distribution. Since $\widehat{P}$ is a linear function of X, then it is also approximately normal. Hence,

$$\widehat{P} \text{ follows a } N\left(\mu_{\widehat{P}}, \sigma^2_{\widehat{P}}\right) = N\left(p, \frac{p\,(1-p)}{n}\right) \text{ distribution approximately.}$$

Hence

$$\begin{aligned} 1-\alpha &\approx P\left(-z_{\alpha/2} \leq \frac{\widehat{P}-p}{\sqrt{p\,(1-p)/n}} \leq z_{\alpha/2}\right) \\ &= P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}} \leq p \leq \widehat{P} + z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}}\right). \end{aligned}$$

Therefore, for large n, a $100\,(1-\alpha)\%$ C.I. for the true proportion p is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}}.$$

Problem: The confidence interval involves p, which is an unknown quantity. In practice, we simply replace p with its point estimate $\widehat{p}$ to obtain an approximate confidence interval.

For large n, that is

$$n\,\widehat{p} > 5 \quad \text{and} \quad n\,(1-\widehat{p}) > 5,$$

a $100\,(1-\alpha)\%$ confidence interval for p is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{n}}.$$

Example 6: A random sample of 300 lenses contains 11 defective lenses. Obtain a 95% C.I. for the true proportion of defective lenses.

Sample Size: If $\widehat{p}$ is used as an estimate of p, we can be $100\,(1-\alpha)\%$ confident that the error $|\widehat{p}-p|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \frac{p\,(1-p)}{n}.$$

Problem: The latter formula involves the unknown parameter p.

Solution: Consider $p\,(1-p) = p - p^2$. The latter is a quadratic function of p. Its curve is a turned down parabola with zeros at $p=0$ and $p=1$. The vertex of the parabola is at $p = 1/2$, which gives the maximum value of $p\,(1-p)$ which is $1/2(1-1/2) = .25$. If we use the n at $p = 1/2$, then it will be at least as large as required by the true value of p.

Hence, If $\widehat{p}$ is used as an estimate of p, we can be at least $100\,(1-\alpha)\%$ confident that the error $|\widehat{p}-p|$ will not exceed a specified amount E when the sample size is

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25)$$

Example 7: Suppose that we wanted to be at least 95% confident that the error in estimating the true proportion p with the sample proportion $\widehat{p}$ is less than 0.02. What sample size should be used?