Module 4

General Sir John Kotelawala Defense University

Department of Mathematics

MA 3102 APPLIED STATISTICS

One-side confidence interval for population mean with one sample

There is one other aspect of confidence intervals that should be mentioned. So far, we have created only what are called **two-sided confidence intervals** for the mean μ. Sometimes, however, we might want only a lower (or upper) bound on μ. We proceed as follows.

Say $\overline{X}$ is the mean of a random sample of size n from the normal distribution $N(\mu, \sigma^2)$, where, for the moment, assume that σ^2 is known. Then

$$P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_\alpha\right) = 1-\alpha,$$

or equivalently,

$$P\left[\overline{X} - z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu\right] = 1-\alpha.$$

Once $\overline{X}$ is observed to be equal to $\overline{x}$, it follows that $[\overline{x} - z_\alpha(\sigma/\sqrt{n}), \infty)$ is a $100(1-\alpha)\%$ **one-sided confidence interval** for μ. That is, with the confidence coefficient $1-\alpha$, $\overline{x} - z_\alpha(\sigma/\sqrt{n})$ is a lower bound for μ. Similarly, $(-\infty, \overline{x} + z_\alpha(\sigma/\sqrt{n})]$ is a one-sided confidence interval for μ and $\overline{x} + z_\alpha(\sigma/\sqrt{n})$ provides an upper bound for μ with confidence coefficient $1-\alpha$.

When σ is unknown, we would use $T = (\overline{X}-\mu)/(S/\sqrt{n})$ to find the corresponding lower or upper bounds for μ, namely,

$$\overline{x} - t_\alpha(n-1)(s/\sqrt{n}) \quad \text{and} \quad \overline{x} + t_\alpha(n-1)(s/\sqrt{n}).$$

7.1-10. A leakage test was conducted to determine the effectiveness of a seal designed to keep the inside of a plug airtight. An air needle was inserted into the plug, and the plug and needle were placed under water. The pressure was then increased until leakage was observed. Let X equal the pressure in pounds per square inch. Assume that the distribution of X is $N(\mu, \sigma^2)$. The following $n = 10$ observations of X were obtained:

3.1 3.3 4.5 2.8 3.5 3.5 3.7 4.2 3.9 3.3

Use the observations to

(a) Find a point estimate of μ.

(b) Find a point estimate of σ.

(c) Find a 95% one-sided confidence interval for μ that provides an upper bound for μ.

CONFIDENCE INTERVALS FOR PROPORTIONS

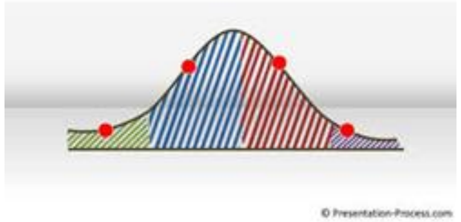
$$Y \sim Binomial(n, p) \;\; ,$$

$$E(Y) = np, \qquad Var(Y) = np(1-p)$$

$$E\left(\frac{Y}{n}\right) = p, \qquad Var\left({}^{Y}\!/_{n}\right) = \frac{p(1-p)}{n}$$

$$\frac{Y - np}{\sqrt{np(1-p)}} = \frac{{}^{Y}\!/_{n} - p}{\sqrt{p(1-p)/n}}$$

has an approximate normal distribution N(0,1) provided n is large enough.



$$P\left[-Z_{\alpha/2} \leq \frac{Y/n - p}{\sqrt{p(1-p)/n}} \leq Z_{\alpha/2}\right] \approx 1 - \alpha$$

$$P\left[\frac{Y}{n} - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{Y}{n} + Z_{\alpha/2}\sqrt{p(1-p)/n}\right] \approx 1 - \alpha$$

Replacing *p = y/n*

$$\left[\frac{Y}{n} + Z_{\alpha/2}\sqrt{\frac{Y/n\,(1 - Y/n)}{n}} \leq p \leq \frac{Y}{n} - Z_{\alpha/2}\sqrt{\frac{Y/n\,(1 - Y/n)}{n}}\right] \approx 1 - \alpha$$

Thus, for large *n*, then the interval

$$\mathrm{P}\left[\frac{Y}{n} - Z_{\alpha/2}\sqrt{\frac{Y/n(1-Y/n)}{n}} \le p \le \frac{Y}{n} - Z_{\alpha/2}\sqrt{\frac{Y/n(1-Y/n)}{n}}\right] \approx 1-\alpha$$

Serve as an approximate 100(1-α) % CI for p.

$$\frac{Y}{n} \pm Z_{\alpha/2}\sqrt{\frac{Y/n\,(1-Y/n)}{n}}$$

EXAMPLE

$$n = 40$$
$$Y = 8$$
Find 90 % CI

Let us return to the example of the histogram of the candy bar weights, Example 6.1-1, with $n = 40$ and $y/n = 8/40 = 0.20$. If $1 - \alpha = 0.90$, so that $z_{\alpha/2} = 1.645$, then, using Equation 7.3-2, we find that the endpoints

$$0.20 \pm 1.645\sqrt{\frac{(0.20)(0.80)}{40}}$$

[0.096, .304] or [9.6%, 30.4%]

If *n* increases, the width of the interval decreases.

In a certain political campaign, one candidate has a poll taken at random among the voting population. The results are that $y = 185$ out of $n = 351$ voters favor this candidate. Even though $y/n = 185/351 = 0.527$, should the candidate feel very confident of winning? From Equation 7.3-2, an approximate 95% confidence interval for the fraction p of the voting population who favor the candidate is

$$0.527 \pm 1.96\sqrt{\frac{(0.527)(0.473)}{351}}$$

or, equivalently, $[0.475, 0.579]$. Thus, there is a good possibility that p is less than 50%, and the candidate should certainly take this possibility into account in campaigning.

One sided CIs

One-sided confidence intervals are sometimes appropriate for p. For example, we may be interested in an upper bound on the proportion of defectives in manufacturing some item. Or we may be interested in a lower bound on the proportion of voters who favor a particular candidate. The one-sided confidence interval for p given by

$$\left[0, \frac{y}{n} + z_\alpha \sqrt{\frac{(y/n)[1-(y/n)]}{n}}\right]$$

provides an upper bound for p, while

$$\left[\frac{y}{n} - z_\alpha \sqrt{\frac{(y/n)[1-(y/n)]}{n}}, 1\right]$$

provides a lower bound for p.

Confidence Intervals Difference Proportions

$$E\left(\frac{Y_1}{n_1}\right) = p_1, \qquad Var\left({}^{Y_1}/_{n_1}\right) = \frac{p_1(1-p_1)}{n_1}$$

$$E\left(\frac{Y_2}{n_2}\right) = p_2, \qquad Var\left({}^{Y_2}/_{n_2}\right) = \frac{p_2(1-p_2)}{n_2}$$

$$= \frac{\left(\left({}^{Y_1}/_{n_1}\right) - \left({}^{Y_2}/_{n_2}\right)\right) - [p_1 - p_2]}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

has an approximate normal distribution N(0,1) provided n_1 and n_2 are large enough.

$$P\left[-Z_{\alpha/2} \leq \left[\frac{\left(\left(Y_1/n_1\right) - \left(Y_2/n_2\right)\right) - [p_1 - p_2]}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}\right] \leq Z_{\alpha/2}\right] \approx 1 - \alpha$$

Thus, for large n, if the observed Y equals y, then the interval

$$\frac{y_1}{n_1} - \frac{y_2}{n_2} \pm Z_{\alpha/2}\sqrt{\frac{\frac{y_1}{n_1}\left(1 - \frac{y_1}{n_1}\right)}{n_1} + \frac{\frac{y_2}{n_2}\left(1 - \frac{y_2}{n_2}\right)}{n_2}}$$

Serve as an approximate 100(1-α) % CI for p_1-p_2.

Two detergents were tested for their ability to remove stains of a certain type. An inspector judged the first one to be successful on 63 out of 91 independent trials and the second one to be successful on 42 out of 79 independent trials. The respective relative frequencies of success are $63/91 = 0.692$ and $42/79 = 0.532$. An approximate 90% confidence interval for the difference $p_1 - p_2$ of the two detergents is

$$\left(\frac{63}{91} - \frac{42}{79}\right) \pm 1.645\sqrt{\frac{(63/91)(28/91)}{91} + \frac{(42/79)(37/79)}{79}}$$

or, equivalently, $[0.039, 0.283]$. Accordingly, since this interval does not include zero, it seems that the first detergent is probably better than the second one for removing the type of stains in question.

SAMPLE SIZE

How large should be the sample size be to estimate a mean?

Example

A sample mean of a test score $\overline{X}$ is approximately $N(\mu, {}^{\sigma^2}/_{n})$. If the σ^2 is known to be 15 from the past experience, how much should be the sample size n, in order to estimate the mean μ = $\bar{x} \pm 1$ with 95 % confidence interval.

We estimate μ with 100(1-0.05) % CI as

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

$$1.96\frac{\sigma}{\sqrt{n}} = 1$$

$$1.96\frac{15}{\sqrt{n}} = 1$$

$$\sqrt{n} = 29.4 \; and \; n \approx 864.36$$

n = 865 because *n* must be an integer.

Suppose we need 80% CI with $\bar{x} \pm 2$

$$1.282\frac{15}{\sqrt{n}} = 2$$

$$\sqrt{n} = 9.615 \; and \; n = 93$$

Example: Average Number of Children in a family

- Say our population size is 5 and we don't want to ask all families.
- The population having **0, 2, 4, 6, 8 children**
- We ask only 2 families at random coming for a super market at 8.00 am and 2.00 pm.

The Maximum Error of the Estimate

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm \varepsilon$$

$$\varepsilon = \boldsymbol{maximum\ error\ of\ the\ estimate} = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$n = \frac{Z_{\alpha/2}^{2} \sigma^{2}}{\varepsilon^{2}}$$

Suppose we know that the unemployment rate has been about 8% (0.08). However, we wish to update our estimate in order to make an important decision about the national economic policy. Accordingly, let us say we wish to be 99% confident that the new estimate of p is within 0.001 of the true p. If we assume Bernoulli trials (an assumption that might be questioned), the relative frequency y/n, based upon a large sample size n, provides the approximate 99% confidence interval:

$$\frac{y}{n} \pm Z_{\alpha/2}\sqrt{\frac{{}^{y}/_{n}\,(1 - {}^{y}/_{n})}{n}}$$

$$Z_{\alpha/2}\sqrt{\frac{{}^{y}/_{n}\,(1 - {}^{y}/_{n})}{n}} = 0.001$$

$$2.576\sqrt{\frac{{}^{y}/_{n}\,(1 - {}^{y}/_{n})}{n}} = 0.001$$

Although we do not know *y/n* before sampling, since *y/n* will be near 0.08.

$$2.576\sqrt{\frac{0.08(0.92)}{n}} = 0.001$$

n = 48,394