CONFIDENCE INTERVALS ON μ WHEN σ IS KNOWN

A. **Confidence intervals (CIs) generally**

1. **Katzer et al. (1998, p. 267): "A range of values within which there is a specified probability that the population parameter will exist, over all possible samples. The specific probability is called the confidence level. A confidence interval is a type of interval estimate, as opposed to $\bar{x}$, a simple point estimate."**

2. **Spatz (1997, p. 386): "an interval of scores which is expected, with specified confidence, to contain a parameter."**

Confidence intervals (CIs) generally (cont'd)

3. this kind of statistical inference depends upon the Central Limit Theorem and the sampling distribution of the mean

4. the width of an confidence interval is determined by

a. the level of confidence desired

b. the sample size used to generate the sampling distribution

B. Expanding the definition

1. we will determine confidence intervals on μ, and we will do so under two different circumstances
2. when σ , the standard deviation of the population, is known – this case is very unusual

   (i). we can use z-score tables, e.g., Table C in Spatz (1997), and what we know about the normal to generate confidence intervals on μ

   (ii) this is the first situation that we will examine; the next lesson will deal with the circumstances when we do not know σ

Expanding the definition (cont'd)

a. when σ , the s.d. of the population, is unknown – this case is very common

(i) we must use the Student's t distribution, e.g., Table D in Spatz (1997)

(ii) we will examine this circumstance in the next lesson

Expanding the definition (cont'd)

2. a common goal is to generate a 95% CI on μ

a. i.e., we want to generate an interval 95% certain to contain μ, the true population mean

(i) e.g., we take 100 random samples, with replacement, all samples of the same size, e.g. $n_1 = n_2 = n_3$, and so on

Expanding the definition (cont'd)

(ii) we then determine the $\bar{x}$ of each of the 100 samples of this same size

(iii) we determine 100 CI's on μ

(iv) 95% of them will **probably** contain μ

Expanding the definition (cont'd)

b. **not** that we are 95% certain that μ is in the interval

c. in doing calculations, we generally determine the mean of only one sample (one x-bar) and generate only one CI on μ

d. another common goal is to determine a 99% CI on μ -- we will do examples of both 95 and 99% CI's on μ in the next lesson

Calculating CI's on μ when sigma (σ) is known

1. 1. recall that the Central Limit Theorem tells us that the distribution of $\bar{x}$ approaches the normal as the number of random samples that generate $\bar{x}$'s increases without bound

2. in a sampling distribution of $\bar{x}$, 95% of the area under the standard normal curve or 95% of the values in the distribution are between $\mu + 1.96\ \sigma$ and $\mu - 1.96\ \sigma$

Calculating CI's on μ when sigma (σ) is known (cont'd)

3. Refer to Table C in Spatz, Area Under the Normal Curve (1997, pp. 369-370)

Calculating CI's on μ when sigma (σ) is known (cont'd)

4. we can say that $\mu = \bar{x} \pm zSE_{\mu}$

OR

when we generate a CI on μ: $\bar{x} - zSE_{\mu} \leq \mu \leq \bar{x} + zSE_{\mu}$

5. a 95% CI on μ: $\bar{x} - 1.96\,SE_{\mu} \leq \mu \leq \bar{x} + 1.96\,SE_{\mu}$

Calculating CI's on μ when sigma (σ) is known (cont'd)

6. suppose that we are doing a research study of how many retrieved records users get from a particular online bibliographic database on a particular set of questions

from one sample, we know that $\overline{x}$ = 24 retrieved records and that n = 100, as well as that σ = 12.1 records (from previous research)

Calculating CI's on μ when sigma (σ) is known (cont'd)

a. construct a 95% CI on μ, that is, generate CI's on μ that will probably contain the true population mean μ for 19 of 20, 95 of 100, 950 of 1000 samples of n=100

b. procedures

(i) divide the confidence level by 2 because the normal is symmetrical

(ii) convert area to appropriate z-score using Table C in Spatz (1997)

Calculating CI's on μ when sigma (σ) is known (cont'd)

(iii) determine $SE_{\mu} = \dfrac{\sigma}{\sqrt{n}} = \sigma_{\bar{x}}$

(iv) use the formula for generating a 95% CI on μ when we know sigma

Calculating CI's on μ when sigma (σ) is known (cont'd)

c. $\frac{0.95}{2} = 0.4750 \Rightarrow 1.96 = z$

(i) 95% CI on μ: $\bar{x} - zSE_{\mu} \leq \mu \leq \bar{x} + zSE_{\mu}$

(ii) $SE_{\mu} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12.1}{\sqrt{100}} = 1.21$ records

Calculating CI's on μ when sigma (σ) is known (cont'd)

(iii) 95% CI on μ: $\bar{x} - zSE_{\mu} \leq \mu \leq \bar{x} + zSE_{\mu}$

$$24 - (1.96)(1.21) \leq \mu \leq 24 + (1.96)(1.21)$$

$$24 - 2.37 \leq \mu \leq 24.2.37$$

$$21.6 \leq \mu \leq 26.4$$

Calculating CI's on μ when sigma (σ) is known (cont'd)

d. or we can also say that a 95% CI on μ: $\bar{x} \pm 1.96 SE_{\mu}$

$\bar{x} \pm 2.37$ records

e. width of any CI on μ when we know sigma = $2zSE_{\mu}$; in our example, the width of the 95% CI on μ is 2(2.37 records) = 4.8 records

D. Summary

1. Cobb et al. (1997, p. 218): confidence intervals answer the question: "what values of the unknown parameter are reasonable given the observed data?"

2. we start with a (sample) statistic $\bar{x}$, and we use that statistic to estimate the parameter, μ, the true population mean: $\bar{x} \pm$ z (some estimate of error [the standard error of μ, SE_{μ}])