DISTRIBUTION OF SAMPLE MEANS

A. Concepts

1. Special relationships obtain among z, PR, and area under the standard normal curve **for raw scores**

   a. **IFF** (if and only if) the original distribution is normal

   b. but we can use the Central Limit Theorem and a sampling distribution to use the standard normal curve on samples from populations that are **not** distributed normally

Concepts (cont'd)

2. we'll use the sampling distribution of the mean, $\bar{x}$

3. also called the distribution of sample means

4. this is a theoretical distribution

Concepts (cont'd)

5. there are three steps in generating a theoretical sampling distribution

a. every sample generating the sample statistic is drawn randomly from the same population

b. the sample size, n, is the same for all samples

c. the number of samples is very large

Concepts (cont'd)

6. for example, the $\bar{x}$ distribution tends to be normal for any distribution of x, no matter how skewed the original distribution is

Refer to Busha & Harter (1980, p. 256, Figure 10.6), means of 10 hypothetical random samples

B. Central Limit Theorem

1. these concepts apply because of the Central Limit Theorem

2. Roscoe (1975, p. 163)

As the number of samples of the same size is taken from the same population and their means computed, "the sampling distribution of the mean approaches a normal distribution with a mean μ and variance σ^2/N," where N is the number of observations in each sample – we would use n instead

Central Limit Theorem (cont'd)

3. the sampling distribution of the mean, also called the distribution of sample means, approaches the normal as the number of $\bar{x}$'s increases

C. Values related to the distribution of sample means, also called the sampling distribution of the mean

1. the mean of the sampling distribution of the mean is μ, the population mean

a. $E(\overline{x}) = \mu$

b. $E(\overline{x})$ is the expected value of x-bar

c. Roscoe (1975, p. 161): "The expected value of any statistic – the predicted value which would give the least error . . . for many samples – is the mean of the sampling distribution."

Values related to the distribution of sample means (cont'd)

2. The standard deviation of the distribution of sample means is defined as the standard error of μ, SE_μ

a. sometimes called the Standard Error or Standard Error of Estimate

b. $SE_\mu = \sigma_x = \frac{\sigma}{\sqrt{n}}$

c. the variance of the sampling distribution of the mean is defined as $\frac{\sigma^2}{n}$

Values related to the distribution of sample means (cont'd)

d. if we do **not** know σ , we estimate the standard error of μ, using a calculation different from Spatz:

$$SE_{\mu} = \frac{s}{\sqrt{n-1}}$$

e. as n increases, $SE_{\mu} \rightarrow 0$, that is, the standard error of μ approaches 0

(i) i.e., the mean of the sampling distribution of the mean approaches μ

(ii) as n increases, we get a better estimate of the population parameter

(iii) what is the standard deviation of this distribution?

D. Examples

1. Busha & Harter (1980, pp. 257-258) discuss a population of books, measuring their length in pages

From previous research, a highly **unusual** situation, we know that

μ = 260 pp.

σ = 180 pp.

Examples (cont'd)

a. the standard deviation is relatively high, i.e.,

$$CV - \frac{\sigma}{\mu} = 0.69$$

b. let's compute the mean and SE_μ when n = 16, n = 49, and n = 100; what that means is that we will consider three conditions, when we take sample sizes of 16, 49, and 100

the number of samples is presumed to be infinitely large

(i) $E(\bar{x}) = \mu = 260 \quad pp$

Examples (cont'd)

(ii) $SE_{\mu} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{180}{\sqrt{16}}, = \frac{180}{\sqrt{49}}, \frac{180}{\sqrt{100}} = 45, 25.7, 18\, pp.$

c. SE_{μ} gets smaller (approaches 0) as the curve of the $\overline{x}$ distribution approaches the normal, $E(\overline{x}) = \mu$

Examples (cont'd)

2. let's start with a population of measurements of individuals' use of a library measured by the number of circulations per month

from previous research, we know that

μ = 9 circulations/month

σ = 1.4 circulations/month

Examples (cont'd)

a. a. compute the mean of the sampling distribution of the mean (the expected value of $\bar{x}[E(\bar{x})]$) and the standard deviation of the sampling distribution of the mean (the standard error of μ [SE_{μ}]) as n = 40, 120

(i) $E(\bar{x})$ = μ = 9 circs/month

(ii) $SE_{\mu} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.4}{\sqrt{40}}, = \frac{1.4}{\sqrt{120}} = 0.22, 0.13$ circs/month

c. $SE_{\mu} = \sigma_{\bar{x}} \rightarrow 0$ as $n \rightarrow \infty$

E. Sampling error

1. SE_{μ} is a measure of sampling error

2. sampling error results from "random fluctuations in samples" (Vogt, 1999, p. 214)

3. Definitions of sampling error

a. Williams (1992, 4th ed., p. 61): "an estimate of how statistics may be expected to deviate from parameters when sampling randomly from a given population"

Sampling error (cont'd)

b. Katzer et al. (1998, p. 282): "The degree to which a sample statistic (e.g., the mean) fails to equal the average of an infinite number of determinations of the statistic from the same population. Sampling error is the result of noise from random sampling."

c. Vogt (1999, p. 253): "The inaccuracies in inferences about a population that come about because researchers have taken a sample rather than studying the entire population."

Sampling error (cont'd)

d. Generally Vogt (1999, p. 274) says, "The smaller the standard error [a result and measure of sampling error], the better the sample statistic is as an estimate of the population parameter. . . . The standard error is the standard deviation of the sampling distribution of a statistic."

F. **A final word on sampling distributions**

The sampling distribution answers the question: "how would the summary statistic behave if we were to repeat the whole process (produce the data and compute the summary) a large number of times?" (Cobb et al., 1997, p. 204)