Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

In-Video Quiz Questions for Unit 3: Part 1 – (1) Sampling Variability and CLT

(02:27) – slide 3, after "Then, let's assume that we take random samples of 1,000 women from each state."

1. Which of the following would you assume to be more variable?

(a) Overall population distribution of all women in the US.

(b) Distribution of sample means of random samples of 1000 women from each state.

(12:55) – slide 5, after "We'll discuss these conditions in more detail in the next couple of slides."

2. Which of the below visualizations is not appropriate for checking the shape of the distribution of the sample, and hence the population?

(a) histogram

(b) boxplot

(c) normal probability plot

(d) barplot

(20:59) – slide 7, after "And also play around with the different parameters of the distributions, either picking how skewed they are, if it's a uniform distribution, what the minimum and the maximum are, of if it's a normal distribution, what the mean and the standard deviation are."

3. Suppose you have a slightly right skewed population distribution of annual incomes in a developed nation, with mean $30,000 and standard deviation $20,000. Suppose you take 10,000 random samples of size 625 from this population. Which of the following is most likely to be the distribution of the means of these samples?

(a) Right skewed, mean = \$30000, SD = \$20000

(b) Nearly normal, mean = \$30000, SD = \$20000

(c) Nearly normal, mean = \$30000, SD = \$20000 $/\sqrt{625} = 800$

(d) Nearly normal, mean = \$30000, SD = \$20000 $/\sqrt{10000} = 200$

(e) Left skewed, mean = \$30000, SD = \$20000 $/\sqrt{625} = 800$

Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

Answers:

1. a
2. d
3. c

Explanation: The CLT states $\overline{x} \sim N(mean = \mu, SE = \frac{\sigma}{\sqrt{n}})$. The population distribution is slightly right skewed but $n = 625$ is large, therefore we would expect CLT to hold and the sampling distribution to be nearly normal. Also, the n in the denominator of the standard error is the sample size (625), not the number of samples (10,000).