

Name: Caleb McWhorter — Solutions

MATH 100

Fall 2023

HW 15: Due 12/11

*“If you torture the data long enough, it will confess.”*

*—Ronald Coase*

**Problem 1.** (10pt) Suppose that the average yearly cost of a private, four-year institution is normally distributed with mean \$26,489 and standard deviation \$3,204. Fahad is hoping to spend between \$18,000 and \$32,000 per year on his education. What percentage of private, four-year institutions meet Fahad’s criterion?

**Solution.** This would be the percentage of private, four-year institutions that cost between \$15,000 and \$32,000, i.e.  $P(\$18000 < X < \$32000)$ . We have...

$$z_{\$32000} = \frac{\$32000 - \$26489}{\$3204} = \frac{\$5511}{\$3204} = 1.72 \rightsquigarrow 0.9573 \quad z_{\$18000} = \frac{\$18000 - \$26489}{\$3204} = \frac{-\$8489}{\$3204} = -2.65 \rightsquigarrow 0.0040$$

Therefore,  $P(X < \$32000) = 0.9573$  and  $P(X < \$18000) = 0.0040$ . But then...

$$P(\$18000 < X < \$32000) = P(X < \$32000) - P(X < \$18000) = 0.9573 - 0.0040 = 0.9533$$

Therefore, 95.33% of private, four-year institutions meet Fahad’s criterion.

**Problem 2.** (10pt) State the Central Limit Theorem. Explain at least two ways in which it is used in Statistics.

**Solution.** The Central Limit Theorem states that if a simple random sample of size  $n$  (sufficiently 'large' or drawn from a normal distribution) with mean  $\mu$  and finite standard deviation  $\sigma$ , then the distribution of sample averages,  $\bar{X}$ , is normally distributed with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$ , ie.  $N(\mu, \sigma/\sqrt{n})$ .

There are *many* applications of the Central Limit Theorem. For instance, this allows one to compute probabilities not for individuals drawn from a population but rather groups of individuals drawn from a population. The Central Limit Theorem is also used to create confidence intervals and also to approximate binomial probabilities, i.e. the normal approximation to the binomial distribution. However, there are many, many other applications of the Central Limit Theorem.

**Problem 3.** (10pt) You have purchased a new 3D printer to create small board game pieces for your small business. The product description states that the variation (measured by the standard deviation) in production time for a project of your size should be no more than 1.2 hours. You create use the machine to create 15 sample pieces and find a mean production time of 241.2 minutes.

- (a) Create a 98% confidence interval for the mean production time for your product.
- (b) What does your computation in (a) assume? Explain.

**Solution.** First, observe that 1.2 hours is 72 minutes.

- (a) In a 98% confidence interval, 98% of the values will be captured by the interval. Therefore, 2% of values will not be captured. Because the normal distribution is symmetric, this leaves 1% of values in each ‘end.’ Therefore, the upper value in the confidence interval will be larger than  $98\% + 1\% = 99\%$  of values in the distribution. But then its  $z$ -score must correspond to 0.99. Therefore,  $z^* \approx 2.325$ . Then we have...

$$\begin{aligned} \bar{x} &\pm z^* \frac{\sigma}{\sqrt{n}} \\ 241.2 &\pm 2.325 \cdot \frac{72}{\sqrt{15}} \\ 241.2 &\pm 43.22 \end{aligned}$$

Therefore, there is a 99% chance that the true average time it will take to complete the pieces is between  $241.2 - 43.22 = 197.98$  minutes and  $241.2 + 43.22 = 284.42$  minutes, i.e. 3 hours and 17.98 minutes and 4 hours and 44.43 minutes.

- (b) For a confidence interval to produce a ‘valid’ estimate of  $\mu$ , the Central Limit Theorem need apply. Therefore, the underlying distribution need have a mean and a finite standard deviation. Clearly, both are the case here. One needs the sample to be a sample random sample—which we are not told. Therefore, we have assumed that the sample was a simple random sample. Furthermore, the Central Limit Theorem requires that the underlying distribution either be normal or that the sample be ‘sufficiently large,’ Clearly, the sample size of  $n = 15 < 30$  is not ‘sufficiently large.’ Therefore, we must also assume that the underlying distribution of construction time is normally distributed.

**Problem 4.** (10pt) You and your three friends would like to win a trivia competition to buy paper from your company to meet your quarterly sales targets. To make it to the final rounds, you need to be in the top 10% of teams. This requires team have at least some average number of questions answered correctly. Looking at past data, you see that the average *individual* gets only 64 questions correct with a standard deviation of 7 questions. Assume that the distribution of number of questions answered correctly by individuals is normally distributed.

- (a) What is the probability that you and your friends can average at least 70 correctly answered questions?
- (b) What is the probability that you and your friends will average less than 65 questions answered correctly?
- (c) What is the probability that you and your friends will average between 65 and 70 correctly answered questions?
- (d) What is the probability that *you* answer more than 70 questions correctly?

**Solution.** We need the distribution of sample means for a sample of size four. Because the underlying distribution of questions answered correctly is normally distributed with finite mean and standard deviation, the Central Limit Theorem applies.<sup>1</sup> Therefore, the distribution of the average questions a team of four can answer is normally distributed with mean  $\mu = 64$  and a standard deviation of  $\sigma/\sqrt{n} = 7/\sqrt{4} = 7/2 = 3.5$ , i.e.  $N(64, 3.5)$ .

(a)

$$z_{70} = \frac{70 - 64}{3.5} = \frac{6}{3.5} = 1.71 \rightsquigarrow 0.9564$$

Therefore,  $P(\bar{X} < 70) = 0.9564$ . Then  $P(\bar{X} \geq 70) = 1 - P(X < 70) = 1 - 0.9564 = 0.0436$ .

(b)

$$z_{65} = \frac{65 - 64}{3.5} = \frac{1}{3.5} = 0.29 \rightsquigarrow 0.6141$$

Therefore,  $P(\bar{X} < 65) = 0.6141$ .

(c) We have...

$$P(65 < \bar{X} < 70) = P(\bar{X} < 70) - P(\bar{X} < 65) = 0.9564 - 0.6141 = 0.3423$$

(d) This is a question about an individual—namely you. Therefore, we use the original, underlying distribution. We have...

$$z_{70} = \frac{70 - 64}{3.5} = \frac{6}{7} = 0.86 \rightsquigarrow 0.8051$$

Therefore,  $P(X < 70) = 0.8051$ . Then  $P(X \geq 70) = 1 - P(X < 70) = 1 - 0.8051 = 0.1949$ .

---

<sup>1</sup>This ignores that you and your friends do not constitute a random sample.