Topic: Sampling distribution of sample mean; Point estimators;

Reading Assignment: 8.1 - 8.4, 9.1

Assigned Problems:

Chapter 8: **10, 11, 14, 16, 22, 30**

Chapter 9: **1 (a & b)**, and then these two additional sections:

(c) Based on your results for (a) and (b), which of the 4 estimators would you consider to be best for population parameter θ (and why)?

(d) Now also consider a 5th estimator: $\hat{\theta}_5 = \frac{X_1+X_2}{3}$

Determine if $\hat{\theta}_5$is better, equally good, or not as good as your choice of the best estimator in (c). Justify your answer.

8.10 **a** $P(199 < \bar{X} < 202) = P\left(\frac{\sqrt{25}(199-200)}{10} < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \frac{\sqrt{25}(202-200)}{10}\right)$

$\approx P(-0.5 < Z < 1) = 0.1915 + 0.3413 = 0.5328$

b $P\left(\sum_{i=1}^{25} X_i \leq 5100\right) = P\left(\frac{\sum_{i=1}^{25} X_i}{25} \leq \frac{5100}{25}\right) = P(\bar{X} \leq 204)$

$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq \frac{\sqrt{25}(204-200)}{10}\right) \approx P(Z \leq 2) = 0.5 + 0.4772 = 0.9772$

c The approximations in parts (a) and (b) assume that the resistances are independent, and random sampling.

8.11 $P(\bar{X} > 14) = P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} > \frac{\sqrt{100}(14-12)}{9}\right) \approx P(Z > 2.22) = 0.5 - 0.4868$

$= 0.0132$

8.14 **a** $P\left(\sum_{i=1}^{100} X_i > 45\right) = P(\bar{X} > 0.45) \approx P\left(Z > \frac{\sqrt{100}(0.45-0.5)}{0.2}\right)$

$= P(Z > -2.5) = 0.5 + 0.4938 = 0.9938$

b $P\left(\sum_{i=1}^{100} X_i > 50\right) = P(\bar{X} > \frac{50}{n}) \approx P\left(Z > \frac{\sqrt{n}((50/n)-0.5)}{0.2}\right) = 0.99$

This equation is true for $\frac{\sqrt{n}((50/n)-0.5)}{0.2} = -2.33$; i.e:

$0.5n - 0.466\sqrt{n} - 50 = 0$. Using the quadratic formula, we have:

$\sqrt{n} = \frac{0.466 \pm \sqrt{(0.466)^2 - 4(0.5)(-50)}}{2(0.5)} = -9.5449$ or 10.4769.

Using the positive root, we have $n = (10.4769)^2 = 109.77 \approx 110$

8.16 Let X_i= service time for i^{th} customer, $i = 1, 2, \ldots\ldots, 100$

$$P\left(\sum_{i=1}^{100} X_i < 120\right) = P\left(\bar{X} < \frac{120}{100}\right) = P(\bar{X} < 1.2)$$

$$= P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} < \frac{\sqrt{100}(1.2 - 1.5)}{1}\right)$$

$$\approx P(Z < -3) = 0.5 - 0.4987 = 0.0013$$

8.22 We can make a point estimate of the population standard deviation using the sample standard deviation of the given data $s = 0.3616$

The t-score for a sample mean 0.2 units above the true average pH would be:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{0.2}{0.3616/\sqrt{12}} = 1.92$$

The t-score for a sample mean 0.2 units below the population mean would then be $t = -1.92$, so, with $12 - 1 = 11$ degrees of freedom, our problem becomes:

$$P(|\bar{X} - \mu| < 0.2) = P(-1.92 < T < 1.92) = 0.9188$$

8.30 Let Y = number of disks that contain missing pulses out of the sample of 100. Then Y has a binomial distribution with parameters $n = 100$, $p = 0.2$. Let X be a normally distributed random variable with parameters:

$\mu = np = 100(0.2) = 20$ and $\sigma^2 = np(1 - p) = 100(0.2)(0.8) = 16$

$$P(Y \leq 15) \approx P(X \leq 15.5) = P\left(\frac{X - \mu}{\sigma} \geq \frac{15.5 - 20}{\sqrt{16}}\right)$$

$$= P(Z \leq -1.125) = 0.5 - 0.3708 = 0.1292$$

9.1 a $E(\hat{\theta}_1) = E(X_1) = \theta$

$$E(\hat{\theta}_2) = E\left(\frac{X_1 + X_2}{2}\right) = \left(\frac{E(X_1) + E(X_2)}{2}\right) = \frac{\theta + \theta}{2} = \theta$$

$$E(\hat{\theta}_3) = E\left(\frac{X_1 + 2X_2}{3}\right) = \left(\frac{E(X_1) + 2E(X_2)}{3}\right) = \frac{\theta + 2\theta}{3} = \theta$$

$$E(\hat{\theta}_4) = E(\bar{x}) = \theta$$

Hence, each of the four estimators is unbiased for θ.

b $V(\hat{\theta}_1) = V(X_1) = \theta^2$

$$V(\hat{\theta}_2) = V\left(\frac{X_1 + X_2}{2}\right) = \left(\frac{V(X_1) + V(X_2)}{4}\right) = \frac{\theta^2 + \theta^2}{4} = \frac{\theta^2}{2}$$

$$V(\hat{\theta}_3) = V\left(\frac{X_1 + X_2}{3}\right) = \left(\frac{V(X_1) + 4V(X_2)}{9}\right) = \frac{\theta^2 + 4\theta^2}{9} = \frac{5\theta^2}{9}$$

$$V(\hat{\theta}_4) = V(\bar{x}) = \frac{\sigma^2}{n} = \frac{\theta^2}{3}$$

Thus, $\hat{\theta}_4 = \bar{x}$ has the smallest variance among these four estimators.

c. Based on your results for (a) and (b), which of the 4 estimators would you consider to be best for population parameter θ (and why)?

$\hat{\theta}_4$ *is the best estimator* → All unbiased and this one has smallest variance

d. Now also consider a 5th estimator: $\hat{\theta}_5 = \frac{X_1+X_2}{3}$

Determine if $\hat{\theta}_5$ is better, equally good, or not as good as your choice of the best estimator in (c). Justify your answer.

$E(\hat{\theta}_5) = E\left(\frac{X_1+X_2}{3}\right) = \frac{E(X_1)+E(X_2)}{3} = \frac{\theta+\theta}{3} = \frac{2\theta}{3}$ → Biased estimator because $E(\hat{\theta}_5) \neq \theta$

Because one of the estimators is biased (Bias=$\frac{2\theta}{3} - \theta = \frac{-\theta}{3}$) we must compare mean squared errors.

$$MSE(\hat{\theta}_5) = V\left(\frac{X_1+X_2}{3}\right) + Bias^2 = \frac{V(X_1)+V(X_2)}{9} + \frac{\theta^2}{9} = \frac{2\theta^2}{9} + \frac{\theta^2}{9} = \frac{3\theta^2}{9} = \frac{\theta^2}{3}$$

$\hat{\theta}_5$ is as good of an estimator of θ as $\hat{\theta}_4$.