5.46

a. The sampling distribution of $\bar{X}$ is centered at $E(\bar{X}) = \mu = 12$ cm, and the standard deviation of the $\bar{X}$ distribution is $\sigma_{\bar{X}} = \dfrac{\sigma_{\bar{X}}}{\sqrt{n}} = \dfrac{.04}{\sqrt{16}} = .01$ cm.

b. With $n = 64$, the sampling distribution of $\bar{X}$ is still centered at $E(\bar{X}) = \mu = 12$ cm, but the standard deviation of the $\bar{X}$ distribution is $\sigma_{\bar{X}} = \dfrac{\sigma_{\bar{X}}}{\sqrt{n}} = \dfrac{.04}{\sqrt{64}} = .005$ cm.

c. $\bar{X}$ is more likely to be within .01 cm of the mean (12 cm) with the second, larger, sample. This is due to the decreased variability of $\bar{X}$ that comes with a larger sample size.

5.50

a. $P(9{,}900 \le \bar{X} \le 10{,}200) \approx P\left(\dfrac{9{,}900-10{,}000}{500/\sqrt{40}} \le Z \le \dfrac{10{,}200-10{,}000}{500/\sqrt{40}}\right)$
$= P(-1.26 \le Z \le 2.53) = \Phi(2.53) - \Phi(-1.26) = .9943 - .1038 = .8905.$

b. According to the guideline given in Section 5.4, n should be greater than 30 in order to apply the CLT, thus using the same procedure for $n = 15$ as was used for $n = 40$ would not be appropriate.

5.59

a. $E(X_1 + X_2 + X_3) = 180$, $V(X_1 + X_2 + X_3) = 45$, $SD(X_1 + X_2 + X_3) = \sqrt{45} = 6.708$.
$P(X_1 + X_2 + X_3 \le 200) = P\left(Z \le \dfrac{200-180}{6.708}\right) = P(Z \le 2.98) = .9986$.
$P(150 \le X_1 + X_2 + X_3 \le 200) = P(-4.47 \le Z \le 2.98) \approx .9986.$

b. $\mu_{\bar{X}} = \mu = 60$ and $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{\sqrt{15}}{\sqrt{3}} = 2.236$, so
$P(\bar{X} \ge 55) = P\left(Z \ge \dfrac{55-60}{2.236}\right) = P(Z \ge -2.236) = .9875$ and
$P(58 \le \bar{X} \le 62) = P(-.89 \le Z \le .89) = .6266.$

c. $E(X_1 - .5X_2 - .5X_3) = \mu - .5\,\mu - .5\,\mu = 0$, while
$V(X_1 - .5X_2 - .5X_3) = \sigma_1^2 + .25\sigma_2^2 + .25\sigma_3^2 = 22.5 \Rightarrow SD(X_1 - .5X_2 - .5X_3) = 4.7434$. Thus,
$P(-10 \le X_1 - .5X_2 - .5X_3 \le 5) = P\left(\dfrac{-10-0}{4.7434} \le Z \le \dfrac{5-0}{4.7434}\right) = P(-2.11 \le Z \le 1.05) = .8531 - .0174 = .8357.$

d. $E(X_1 + X_2 + X_3) = 150$, $V(X_1 + X_2 + X_3) = 36 \Rightarrow SD(X_1 + X_2 + X_3) = 6$, so
$P(X_1 + X_2 + X_3 \le 200) = P\left(Z \le \dfrac{160-150}{6}\right) = P(Z \le 1.67) = .9525.$
Next, we want $P(X_1 + X_2 \ge 2X_3)$, or, written another way, $P(X_1 + X_2 - 2X_3 \ge 0)$.
$E(X_1 + X_2 - 2X_3) = 40 + 50 - 2(60) = -30$ and $V(X_1 + X_2 - 2X_3) = \sigma_1^2 + \sigma_2^2 + 4\sigma_3^2 = 78 \Rightarrow$
$SD(X_1 + X_2 - 2X_3) = 8.832$, so
$P(X_1 + X_2 - 2X_3 \ge 0) = P\left(Z \ge \dfrac{0-(-30)}{8.832}\right) = P(Z \ge 3.40) = .0003.$

6.11

a. $E\left(\frac{X_1}{n_1}-\frac{X_2}{n_2}\right)=\frac{1}{n_1}E(X_1)-\frac{1}{n_2}E(X_2)=\frac{1}{n_1}(n_1p_1)-\frac{1}{n_2}(n_2p_2)=p_1-p_2$.

b. $V\left(\frac{X_1}{n_1}-\frac{X_2}{n_2}\right)=V\left(\frac{X_1}{n_1}\right)+V\left(\frac{X_2}{n_2}\right)=\left(\frac{1}{n_1}\right)^2 V(X_1)+\left(\frac{1}{n_2}\right)^2 V(X_2)=$
$\frac{1}{n_1^2}(n_1p_1q_1)+\frac{1}{n_2^2}(n_2p_2q_2)=\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}$, and the standard error is the square root of this quantity.

c. With $\hat{p}_1=\frac{x_1}{n_1}$, $\hat{q}_1=1-\hat{p}_1$, $\hat{p}_2=\frac{x_2}{n_2}$, $\hat{q}_2=1-\hat{p}_2$, the estimated standard error is $\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$.

d. $(\hat{p}_1-\hat{p}_2)=\frac{127}{200}-\frac{176}{200}=.635-.880=-.245$

e. $\sqrt{\frac{(.635)(.365)}{200}+\frac{(.880)(.120)}{200}}=.041$

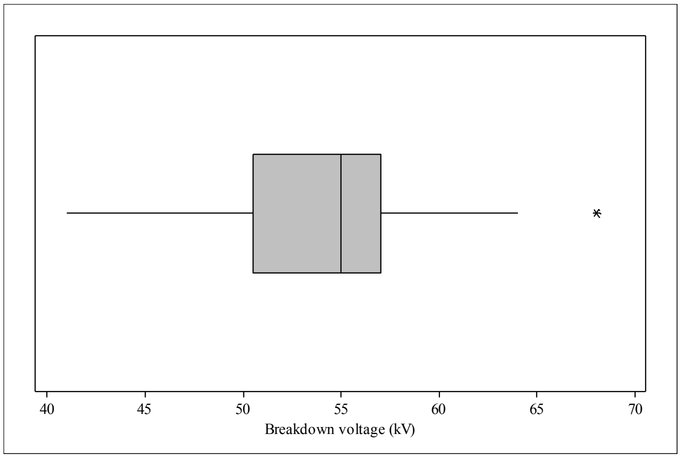
6.15

a. $E(X^2)=2\theta$ implies that $E\left(\frac{X^2}{2}\right)=\theta$. Consider $\hat{\theta}=\frac{\sum X_i^2}{2n}$. Then
$E\left(\hat{\theta}\right)=E\left(\frac{\sum X_i^2}{2n}\right)=\frac{\sum E\left(X_i^2\right)}{2n}=\frac{\sum 2\theta}{2n}=\frac{2n\theta}{2n}=\theta$, implying that $\hat{\theta}$ is an unbiased estimator for θ.

b. $\sum x_i^2=1490.1058$, so $\hat{\theta}=\frac{1490.1058}{20}=74.505$.

7.16

a. The boxplot shows a high concentration in the middle half of the data (narrow box width). There is a single outlier at the upper end, but this value is actually a bit closer to the median (55 kV) than is the smallest sample observation.



b. From software, $\bar{x}$ = 54.7 and s = 5.23. The 95% confidence interval is then

$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}} = 54.7 \pm 1.96\frac{5.23}{\sqrt{48}} = 54.7 \pm 1.5 = (53.2,\ 56.2)$$

We are 95% confident that the true mean breakdown voltage under these conditions is between 53.2 kV and 56.2 kV. The interval is reasonably narrow, indicating that we have estimated μ fairly precisely.

c. A conservative estimate of standard deviation would be (70 – 40)/4 = 7.5. To achieve a margin of error of at most 1 kV with 95% confidence, we desire

$$1.96\frac{s}{\sqrt{n}} \le 1 \Rightarrow n \ge \left[\frac{1.96s}{1}\right]^2 = \left[\frac{1.96(7.5)}{1}\right]^2 = 216.09.$$

Therefore, a sample of size at least n = 217 would be required.

7.20

Because the sample size is so large, the simpler formula (7.11) for the confidence interval for p is sufficient:

$$.15 \pm 2.58\sqrt{\frac{(.15)(.85)}{4722}} = .15 \pm .013 = (.137, .163).$$

7.34

n = 14, $\bar{x} = 8.48$, $s = .79$; $t_{.05,13} = 1.771$

A 95% lower confidence bound: $8.48 - 1.771\left(\frac{.79}{\sqrt{14}}\right) = 8.48 - .37 = 8.11$. With 95% confidence, the value of the true mean proportional limit stress of all such joints is greater than 8.11 MPa. We must assume that the sample observations were taken from a normally distributed population.

8.26

Reject H_0 if $z \ge 1.645$; $\frac{s}{\sqrt{n}} = .7155$, so $z = \frac{52.7 - 50}{.7155} = 3.77$. Since 3.77 is ≥ 1.645, reject H_0 at level .05 and conclude that true average penetration exceeds 50 mils.

8.37

a. The parameter of interest is p = the proportion of the population of female workers that have BMIs of at least 30 (and, hence, are obese). The hypotheses are H_0: p = .20 versus H_a: p > .20.

With n = 541, np_0 = 541(.2) = 108.2 ≥ 10 and $n(1 - p_0)$ = 541(.8) = 432.8 ≥ 10, so the "large-sample" z procedure is applicable. Hence, we will reject H_0 if $z \geq z_{.05} = 1.645$.

From the data provided, $\hat{p} = \frac{120}{541} = .2218$, so $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{.2218 - .20}{\sqrt{.20(.80)/541}} = 1.27$.

Since 1.27 < 1.645, we fail to reject H_0 at the α = .05 level. We do not have sufficient evidence to conclude that more than 20% of the population of female workers is obese.

b. A Type I error would be to incorrectly conclude that more than 20% of the population of female workers is obese, when the true percentage is 20%. A Type II error would be to fail to recognize that more than 20% of the population of female workers is obese when that's actually true.

c. The question is asking for the chance of committing a Type II error when the true value of p is .25, i.e. $\beta(.25)$. Using the textbook formula,

$$\beta(.25) = \Phi\left[\frac{.20 - .25 + 1.645\sqrt{.20(.80)/541}}{\sqrt{.25(.75)/541}}\right] = \Phi(-1.166) \approx .121.$$

8.58

μ = the true average percentage of organic matter in this type of soil, and the hypotheses are H_0: μ = 3 v. H_a: $\mu \neq 3$. With n = 30, and assuming normality, we use the t test:

$t = \frac{\bar{x} - 3}{s/\sqrt{n}} = \frac{2.481 - 3}{.295} = \frac{-.519}{.295} = -1.759$. The P-value = 2[P(t > 1.759)] = 2(.041) = .082. At significance level .10, since .082 ≤ .10, we would reject H_0 and conclude that the true average percentage of organic matter in this type of soil is something other than 3. At significance level .05, we would not have rejected H_0.

8.75

a. With H_0: $p = 1/75$ v. H_a: $p \neq 1/75$, we reject H_0 if either $z \geq 1.96$ or $z \leq -1.96$. With $\hat{p} = \frac{16}{800} = .02$, $z \frac{.02 - .01333}{\sqrt{\frac{.01333(.98667)}{800}}} = 1.645$, which is not in either rejection region. Thus, we fail to reject the null hypothesis. There is not evidence that the incidence rate among prisoners differs from that of the adult population. The possible error we could have made is a type II.

b. P–value $= 2[1 - \Phi(1.645)] = 2[.05] = .10$. Yes, since .10 < .20, we could reject H_0.