Solutions Project 1
17 points

1. (3 pts, Exercise 7.10)

For any n, $P(|\bar{Y} - \mu| \le .3) = P\left(\left|\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \le \frac{0.3}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(|Z| \le \frac{0.3\sqrt{n}}{\sigma}\right)$.

b. With $\sigma=2$,

For $n = 25$: $P(|\bar{Y} - \mu| \le .3) = P(|Z| \le 0.75) = 1 - 2P(Z > .75) = .5468$

For $n = 36$: $P(|\bar{Y} - \mu| \le .3) = P(|Z| \le 0.9) = 1 - 2P(Z > .9) = .6318$

For $n = 49$: $P(|\bar{Y} - \mu| \le .3) = P(|Z| \le 1.05) = 1 - 2P(Z > 1.05) = .7062$

For $n = 64$: $P(|\bar{Y} - \mu| \le .3) = P(|Z| \le 1.2) = 1 - 2P(Z > 1.2) = .7698$

c. For larger sample sizes, the sampling distribution of the sample mean has smaller variance. Thus, the probably of the sample mean being within 0.3 of the population mean increases with the sample size n.

d. With a larger population standard deviation σ, the pdf has thicker tails. Thus, the probabilities $P(|\bar{Y} - \mu| \le .3)$ are smaller with a larger population standard deviation.

2. (1 pt, Exercise 7.12)

We require $P(|\bar{Y} - \mu| \le 1) = P(-1 \le \bar{Y} - \mu \le 1) = P\left(\frac{-1}{\frac{\sigma}{\sqrt{n}}} \le \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{1}{\frac{\sigma}{\sqrt{n}}}\right) = P(-.25\sqrt{n} \le Z$ $\le .25\sqrt{n}) = .90$. This probability statement holds as long as $0.25\sqrt{n} = 1.645$, so $n = 43.296$. Hence, sample 44 trees.

3. (3 pts, Exercise 7.20)

a. Using the fact that $\chi^2(\nu) = \text{GAM}(\alpha = \nu/2, \beta = 2)$, then $E(U) = \alpha\beta = \nu$, and $V(U) = \alpha\beta^2 = 2\nu$.

b. By Theorem 7.3, $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$. Thus, the result from part **a** shows that $E(\frac{n-1}{\sigma^2}S^2) = n-1$. Since expectation is a linear operator, $\frac{n-1}{\sigma^2}E(S^2) = n-1$ and so

$E(S^2) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$.

Part **a** also shows that $V(\frac{n-1}{\sigma^2}S^2) = 2(n-1)$ and so $\left(\frac{n-1}{\sigma^2}\right)^2 V(S^2) = 2(n-1)$ from which it follows that $V(S^2) = \left(\frac{\sigma^2}{n-1}\right)^2\left[2(n-1)\right] = 2\sigma^4/(n-1)$.

4. (2 pt) (Exercise 7.42)

Let $\bar{Y}$ denote the sample mean strength of 100 random selected pieces of glass. Thus, the quantity $\frac{\bar{Y}-14}{2/\sqrt{100}} = (\bar{Y} - 14)/.2$ has an approximate standard normal distribution by CLT.

a. $P(\bar{Y} > 14.5) = P\left(\frac{\bar{Y}-14}{2/\sqrt{100}} > \frac{14.5-14}{2/\sqrt{100}}\right) \approx P(Z > 2.5) = .0062$.

b. We have that $P\left(-1.96 < \frac{\bar{Y}-14}{2/\sqrt{100}} < 1.96\right) = P(-1.96 < Z < 1.96) = .95$. Thus
$P\left(14 - 1.96 \cdot 2/\sqrt{100} < \bar{Y} < 14 + 1.96 \cdot 2/\sqrt{100}\right) = P\left(13.608 < \bar{Y} < 14.392\right) = 0.95$.
Thus, the interval which captures 95% of all sample mean fracture strengths from samples of size 100 is [13.608, 14.392].

5. (2 pts, Exercise 7.58)

For $W_i = X_i - Y_i$, we have that $E(W_i) = E(X_i) - E(Y_i) = \mu_1 - \mu_2$ and $V(W_i) = V(X_i) - V(Y_i) = \sigma_1^2 + \sigma_2^2$ since X_i and Y_i are independent. Thus, $\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \bar{X} - \bar{Y}$ so $E(\bar{W}) = \mu_1 - \mu_2$, and $V(\bar{W}) = (\sigma_1^2 + \sigma_2^2)/n$. Thus, since the W_i are independent,

$$U_n = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} = \frac{\bar{W} - E(\bar{W})}{\sqrt{V(\bar{W})}}$$

satisfies the conditions of the CLT and has a limiting standard normal distribution.

6. (2 pts)

a. $P(Y < 1) = F(0)$, where F is the cdf for Bin($n = 5$, $p = 0.1$). In R, you can use either dbinom(0,5,.1) or pbinom(0,5,.1), both giving 0.59049.
b. $N(.5, .45)$
c. $P(Y < 1)$ is approximated in R by pnorm(1,0.5,sqrt(0.45)) = 0.7719717.
d. the binomial pmf does not resemble a mound–shaped distribution (n is not large here).

7. (1 pt) I used the following R code to determine that the sample size required is $n = 66$:

```
> n=65; abs(pnorm(1,n*.1,sqrt(.1*.9*n))-pbinom(0,n,.1))
[1] 0.01042277
> n=66; abs(pnorm(1,n*.1,sqrt(.1*.9*n))-pbinom(0,n,.1))
[1] 0.009834207
```

8. (1 pt) (Exercise 7.80)

Let Y = # in the sample that are younger than 31 years of age. Since 31 is the median age, $Y \sim$ Bin($n = 100$ and $p = 1/2$). Because n is large, this Binomial distribution is well approximated by N($np = 50$, $np(1 - p) = 25$). Therefore,

$$P(Y \geq 60) \approx P(Z \geq \tfrac{60-50}{\sqrt{25}}) = P(Z \geq 2) = 0.0228.$$

9. (2 pts) (Exercise 7.84)

a. $E\left(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right) = \frac{E(Y_1)}{n_1} - \frac{E(Y_2)}{n_2} = \frac{n_1 p_1}{n_1} - \frac{n_2 p_2}{n_2} = p_1 - p_2$. Note that distributing the expectation across the difference does NOT require independence.
b. Since the two samples are independent of each other, then

$$V\left(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right) = \frac{V(Y_1)}{n_1^2} + \frac{V(Y_2)}{n_2^2} = \frac{n_1 p_1(1-p_1)}{n_1^2} + \frac{n_2 p_2(1-p_2)}{n_2^2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$