Solutions Project 3
(23 pts)

1. (8.64, 3 pts)

a. The point estimates are .35 (sample proportion of 18-34 year olds who consider themselves patriotic) and .77 (sample proportion of 60+ year olds who consider themselves patriotic. So, a 98% CI is, using $z_{0.01} = 2.326$,

$$(.77 - .35) \pm 2.326\sqrt{\tfrac{(.77)(.23)}{150} + \tfrac{(.35)(.65)}{340}}$$ or .42 ± .10 or (.32, .52).

b. Since the CI for the difference in proportions is BELOW .6, then the evidence suggests that there is a true difference in proportions, but this difference is less than .6.

c-d. We are 98% confident that, compared to 18-34 year olds, 32% to 52% more 60+ year olds consider themselves patriotic.

e. The assumptions are that: (1) The individuals in the 18-34 group were independently chosen from the individuals in the 60+ group; (2) The individuals within the 18-34 group were a SRS, and the individuals in the 60+ group were a SRS; (3) The sample sizes are large enough so that the normal approximation for the sampling distribution for $\hat{p}_1 - \hat{p}_2$ is appropriate, i.e., $n_i p_i \geq 10$ and $n_i(1 - p_i) \geq 10$.

2. (8.72, 2 pts)

a. Since the true proportions of men and women are unknown, substitute 0.5 in for both to maximize the margin of error (error bound) that corresponds to a 100(1- α)% CI:

$$z_{\alpha/2}\sqrt{\tfrac{.5(.5)}{1000} + \tfrac{.5(.5)}{1000}} = z_{\alpha/2}(0.022361).$$

When α = 0.05, then the margin of error for a 95% CI:

$$z_{0.025}\sqrt{\tfrac{.5(.5)}{1000} + \tfrac{.5(.5)}{1000}} = 1.96(0.022361) = .044.$$

b. Assuming that the two sample sizes are equal, and using $z_{0.05} = 1.645$ to generate a 90% CI for the true difference in proportions, solve the relation

$$1.645\sqrt{\tfrac{.5(.5)}{n} + \tfrac{.5(.5)}{n}} = .02,$$

so $n = .5\left(\dfrac{1.645}{0.02}\right)^2$, and rounding up gives n = 3383 in each group.

3. (8.74, 3 pts)

a. With α= 0.05, m = 0.1 and σ = 0.5, the sample size is $n = (1.96)^2\sigma^2/m^2$, so n = 97. It is not valid to select specimens from the same rainfall, since this would violate the assumption that the observations be independent.

b. A conservative range for pH on rain in North America is from 2 to 6. Thus, Using the range method, use $\sigma = (6 - 2)/4 = 1$ in the sample size calculation, $n = (1.96)^2 1^2/(.1)^2 = 384.16$. Thus, choose n = 385.

c. With $m = c\sigma$, then $n = z^2 \sigma^2/(c\sigma)^2 = (z/c)^2$. For 95% confidence, $z = 1.96$ which shows that for c = .5, 1 and 2, then the samples sizes are $n = 16$, $n = 4$, and $n = 1$ respectively.

4. (8.88, 3 pts)

a. The sample statistics are, with $n = 12$, $\bar{y} = 9$, $s = 6.4$. The 90% CI for μ = mean LC50 for DDT is, with 11 degrees of freedom and $t_{.05} = 1.796$,

$$9 \pm 1.796\left(6.4/\sqrt{12}\right) = 9 \pm 3.32 = (5.68, 12.32).$$

b. We are 90% confident that the true mean LC50 for DDT is between 5.68 and 12.32.

c. The two assumptions are that (1) the data are a SRS; and (2), since the sample size is so small, the data are from a normal distribution.

d. Since 6 is in the CI, the data do not support the conjecture that the mean LC50 measurement is larger then 6.

5. (8.90, 3 pts)

a. The pooled sample variance is $s_p^2 = \frac{14(42)^2 + 14(45)^2}{28} = 1894.5$ so $s_p = 43.5$. Because $n_1 = 15$ and $n_2 = 15$, there are 28 degrees of freedom giving $t_{0.025} = 2.048$, and so the 95% CI for the difference in mean verbal scores is

$$(446 - 534) \pm 2.048 \times 43.5\sqrt{\left(\tfrac{1}{15} + \tfrac{1}{15}\right)} = -88 \pm 32.55 = [-120.55, -55.45].$$

b-c. We are 95% confident that, on average, the literature majors outperform the engineering majors by between 55 and 121 point on the Verbal SAT.

d. We made four assumptions: (1) the sample measurements were SRSs from each group; (2) the math majors were chosen independently of the literature majors; (3) since the sample sizes are small, the data are from normal populations; and (4) the normal populations have equal variances, $\sigma_1 = \sigma_2$.

6. (8.100, 2 pts) A lower 1-sided CI for σ can be obtained from taking square root of the upper endpoint of the CI for σ^2 on page 435 of the textbook: $(0, \sqrt{\frac{(n-1)S^2}{\chi^2_{0.99}}}]$, where $\chi^2_{.99}$ is the 0.01 percentile with n-1 degrees of freedom. For this problem, the sample variance $s^2 = 34854.4$, $s = 186.7$ and n=20, so there are 19 degrees of freedom which shows that $\chi^2_{0.99} = 7.6327$. Thus, a 99% lower 1-sided CI for the standard deviation σ is $(0, \sqrt{\frac{19(34854.4)}{7.6327}}] = (0, 294.55]$.

In conclusion, we are 99% confident that the true standard deviation can be as large as 295 which is way larger than 150. In other words, because 150 is in this CI, then the evidence fails to suggest that the standard deviation is less than 150.

7. (8.125, 2 pts)

a. From Definition 7.3, since $\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ and $\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$ are independent random variables, then

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2-1)} = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F\,(n_1-1,\, n_2-1).$$

b. By choosing percentiles from $F(n_1 - 1, n_2 - 1)$, we have

$$P(F_{n_1-1,n_2-1,1-\alpha/2} < F < F_{n_1-1,n_2-1,\alpha/2}) = 1-\alpha\,.$$

Using the F random variable from part (a) gives

$$P(F_{n_1-1,n_2-1,1-\alpha/2} < \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} < F_{n_1-1,n_2-1,\alpha/2}) = P(\frac{S_2^2}{S_1^2} F_{n_1-1,n_2-1,1-\alpha/2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2} F_{n_1-1,n_2-1,\alpha/2}) = 1-\alpha.$$

Thus, $\left(\frac{S_2^2}{S_1^2} F_{n_1-1,n_2-1,1-\alpha/2},\ \frac{S_2^2}{S_1^2} F_{n_1-1,n_2-1,\alpha/2}\right)$ is a 100(1 – α)% CI for σ_2^2/σ_1^2.

The F quantiles in this CI can easily be calculated in R using the `qf( )` function. Because your textbook only shows tables with upper percentiles from an F distribution, an alternative CI expression is found by noting that if $F_{1-\alpha/2} < F < F_{\alpha/2}$ then $1/F_{\alpha/2} < 1/F < 1/F_{1-\alpha/2}$. When $F \sim F(n_1-1, n_2-1)$ you showed previously that $1/F \sim F(n_2-1, n_1-1)$ which shows that $F_{n_1-1,n_2-1,1-\alpha/2} = \frac{1}{F_{n_2-1,n_1-1,\alpha/2}}$, and so the 100(1 – α)% CI for σ_2^2/σ_1^2 that only uses upper quantiles is $\left(\frac{1}{F_{n_2-1,n_1-1,\alpha/2}}\frac{S_2^2}{S_1^2},\ F_{n_1-1,n_2-1,\alpha/2}\frac{S_2^2}{S_1^2}\right)$.

8. (2 pts) Using #7 (Exercise 8.125) to build a CI for $\sigma^2_{summer}/\sigma^2_{spring}$, we'll consider spring alligators as population 1 and summer alligators as population 2. The quantiles necessary for the 90% CI are $F_{4,3,0.95} = 0.1517$ and $F_{4,3,0.05} = 9.1172$. Since $\frac{S^2_{summer}}{S^2_{spring}} = 5.9379$ then a 90% CI for $\sigma^2_{summer}/\sigma^2_{spring}$ is $(5.9379(0.1517),\ 5.9379(9.1172)) = (.90, 54.14)$. The R code and output:

```
> qf(c(.05,.95),4,3)          # Critical values
[1] 0.1517132 9.1171823
> qf(c(.05,.95),4,3)*5.9379   # 90% CI for variance ratio spring/summer
[1]  0.9008581 54.1369165
```

Thus, we are 90% confident that the true ratio $\sigma^2_{summer}/\sigma^2_{spring}$ is between 0.90 and 54.14. When assuming equal variances $\sigma^2_{summer}=\sigma^2_{spring}$ is equivalent to assuming that $\sigma^2_{summer}/\sigma^2_{spring}=1$. Since the 90% CI contains 1, then the evidence fails to refute the assumption that $\sigma^2_{summer}=\sigma^2_{spring}$.

Alternatively, we can consider $\dfrac{S^2_{summer}}{S^2_{spring}} \sim F(3,4)$ to get a CI for $\sigma^2_{spring}/\sigma^2_{summer}$. Since $\dfrac{S^2_{spring}}{S^2_{summer}} = 0.16841$ then a 90% CI for $\sigma^2_{spring}/\sigma^2_{summer}$ is $\left(0.16841(0.1097),\ 0.16841(6.5914)\right) = (0.0185, 1.11)$. The R code:

```
> qf(c(.05,.95),3,4)                # Critical values
[1] 0.109683 6.591382
> qf(c(.05,.95),3,4)*0.16841   # 90% CI for variance ratio spring/summer
[1] 0.01847172 1.11005466
```

Again, we see that 1 is in the CI. The endpoints of this CI are merely the reciprocals of the endpoints of $(0.90, 54.14)$.

9. (8.133, 3 pts) Because we assume $\sigma_1^2=\sigma_2^2=\sigma^2$ then from Chapter 7, we know that $E(S_i^2)=\sigma^2$ and $V(S_i^2)=\frac{2\sigma^2}{n_i-1}$ for i = 1, 2.

a. So $E(S_p^2)=\dfrac{(n_1-1)E(S_1^2)+(n_2-1)E(S_2^2)}{n_1+n_2-2}=\sigma^2$.

b. Because of the independence of the samples, then S_1^2 and S_2^2 are independent, and so $V(S_p^2)=\dfrac{(n_1-1)^2V(S_1^2)+(n_2-1)^2V(S_2^2)}{(n_1+n_2-2)^2}=\dfrac{2\sigma^4}{n_1+n_2-2}$.

c. All three estimators are unbiased. Since $Var(S_p^2)$ is less than either $Var(S_1^2)=\dfrac{2\sigma^4}{n_1-1}$ or $Var(S_2^2)=\dfrac{2\sigma^4}{n_2-1}$ (by Exercise 7.20) because it has the same numerator but a larger denominator, then S_p^2 is the preferred estimator of σ^2.