**Biostats 518 Homework 7**

1. Question 1
   1. Are mean cholesterol levels associated with sex in Caucasians?
      1. **Methods:** Run a t-test with unequal variances using the mean of each sex as the inputs for the test. We use the point estimate of the mean +/- 1.96\*standard error to find the 95% CI for the data. We then view the p-values generated by the t-test to understand the significance of the results.
      2. **Inference:** The mean cholesterol for Caucasian men is 197.5 with a 95% CI between 195.35 and 199.64. The mean cholesterol for Caucasian women is 222.8 with a 95% CI between 220.63 and 224.96. Assuming both groups have proportional sample sizes (58% women and 41% men) but unequal variances, the t-test seems to suggest a significant difference between males and females with a very low p-value.
   2. Are mean cholesterol levels associated with sex in noncaucasians?
      1. **Methods:** Run a t-test with unequal variances using the mean of each sex as the inputs for the test. We use the point estimate of the mean +/- 1.96\*standard error to find the 95% CI for the data. We then view the p-values generated by the t-test to understand the significance of the results.
      2. **Inference:** The mean cholesterol for nonaucasian men is 197.9 with a 95% CI between 192.88 and 202.91. The mean cholesterol for noncaucasian women is 213.6 with a 95% CI between 209.05 and 218.14. Assuming both groups have proportional sample sizes (58% women and 41% men) but unequal variances, the t-test seems to suggest a significant difference between males and females with a very low p-value.
   3. Are mean cholesterol levels associated with sex after adjustment for race?
      1. **Method:** Combine the data using importance weights and generate a new standard error for the groups. Use the new standard errors in a t-test calculation and test for significance. Generate 95% CI for the results.
      2. **Inference:** After generating a new SE estimate for the groups and running a t-test, the results still seem to be significant between groups. The 95% CI for Caucasians seemed to get a little smaller and the 95% CI for noncaucasians seemed to get a little wider (perhaps because they “averaged” each other out).
   4. Race does seem to modify the association between mean cholesterol level and sex
2. Fibrogen
   1. Are mean fibrogen levels associated with sex in Caucasians?
      1. **Methods:** Run a t-test with unequal variances using the mean of each sex as the inputs for the test. We use the point estimate of the mean +/- 1.96\*standard error to find the 95% CI for the data. We then view the p-values generated by the t-test to understand the significance of the results.
      2. **Inference:** The mean fibrogen for Caucasian men is 317.8 with a 95% CI between 313.63 and 321.96. The mean fibrogen for Caucasian women is 320.7 with a 95% CI between 317.51 and 323.88. Assuming both groups have proportional sample sizes (58% women and 41% men) but unequal variances, the t-test seems to suggest a significant difference between males and females with a very low p-value.
   2. Are mean fibrogen levels associated with sex in noncaucasians?
      1. **Methods:** Run a t-test with unequal variances using the mean of each sex as the inputs for the test. We use the point estimate of the mean +/- 1.96\*standard error to find the 95% CI for the data. We then view the p-values generated by the t-test to understand the significance of the results.
      2. **Inference:** The mean fibrogen for nonaucasian men is 333.7 with a 95% CI between 322.66 and 344.73. The mean fibrogen for noncaucasian women is 349.4 with a 95% CI between 340.29 and 358.50. Assuming both groups have proportional sample sizes (58% women and 41% men) but unequal variances, the t-test seems to suggest a significant difference between males and females with a very low p-value.
   3. Are mean cholesterol levels associated with sex after adjustment for race?
      1. **Method**: Combine the data using importance weights and generate a new standard error for the groups. Use the new standard errors in a t-test calculation and test for significance. Generate 95% CI for the results.
      2. **Inference**: After generating a new SE estimate for the groups and running a t-test, the results still seem to be significant between groups. The 95% CI for Caucasians seemed to get a little smaller and the 95% CI for noncaucasians seemed to get a little wider (perhaps because they “averaged” each other out).
   4. Race does seem to modify the association between mean fibrogen level and sex
3. Obtaining estimates for sample size calculations
   1. 39.288 mg/dl
   2. 58.932 mg/dl
   3. 37.492 mg/dl (based off of RMSE from regression model)
4. Two-arm study
   1. 1780 (using a correlation value of 0.4, u1 of 211, u2 of 208)
   2. 2383 (using a correlation value of 0.4, u1 of 211, u2 of 208)
   3. It would be lower if we didn’t adjust for sex or age
   4. It would be lower since we don’t care about the delta between the final state and the initial state
   5. It would be higher since we are accounting for the variance of sex and age. These factors would make the model more exact, but possibly overfitted.
5. Two-arm study
   1. 1960/2993 = 0.65
   2. 2448/2505 = 0.97
   3. We need at least 77 people in our sample
   4. We need a smaller sample size in this study because we are looking for the actual end result as opposed to the delta. The results of this experiment will just tell us that there is a difference between the two groups, but the results of the 4b experiment will tell us that there is a difference between the deltas in the two groups. 4b is stronger evidence than this experiment.