BIOSTATISTICS 515/518, Assignment #7

1a) Within Caucasian individuals, mean cholesterol is 25.30 mg/dL higher in females than in males. Based on an approximate 95% confidence interval, this observed difference is consistent with a true difference of between 22.26 mg/dL and 28.34 mg/dL higher in females. Thus we have statistically significant evidence of an association between gender and serum cholesterol in Caucasians.

b) Within non-Caucasian individuals, mean cholesterol is 15.7 mg/dL higher in females than in males. Based on an approximate 95% confidence interval, this observed difference is consistent with a true difference of between 8.932 mg/dL and 22.47 mg/dL higher in females. Thus we have statistically significant evidence of an association between gender and serum cholesterol in non-Caucasians.

c) We now adjust for race in our analysis by calculating the weighted difference in mean Cholesterol for Caucasian and non-Caucasians with weights determined by the distribution of race in the sample. The adjusted difference in mean cholesterol is 23.77 mg/dL higher in females. Based on an approximate 95% confidence interval, this observed difference is consistent with a true difference of between 20.52 mg/dL and 27.01 mg/dL higher in females. Thus we have statistically significant evidence of an association between gender and serum cholesterol after adjusting for race.

d) Within Caucasian individuals, we estimated mean cholesterol to be 25.30 mg/dL higher in females than in males. Within non-Caucasian individuals, we estimated mean cholesterol is 15.7 mg/dL higher in females than in males. Since the estimates of each race are not consistent with the other group’s 95% confidence interval, we have evidence of effect modification.

2a) Within Caucasian individuals, mean fibrinogen is 2.900 mg/dL higher in females than in males. Based on an approximate 95% confidence interval, this observed difference is consistent with a true difference of between 2.347 mg/dL lower and 8.147 mg/dL higher in females. Since the interval includes 0 we cannot conclude that we have statistically significant evidence of an association between gender and serum fibrinogen in Caucasians.

b) Within non-Caucasian individuals, mean fibrinogen is 15.7 mg/dL higher in females than in males. Based on an approximate 95% confidence interval, this observed difference is consistent with a true difference of between 1.400 mg/dL and 30.00 mg/dL higher in females. Thus we have statistically significant evidence of an association between gender and serum fibrinogen in non-Caucasians.

c) We now adjust for race in our analysis by calculating the weighted difference in mean fibrinogen for Caucasian and non-Caucasians with weights determined by the distribution of race in the sample (2,534 Caucasians, 481 non-Caucasians). The adjusted difference in mean cholesterol is 4.942 mg/dL higher in females. Based on an approximate 95% confidence interval, this observed difference is consistent with a true difference of between 0.8620 mg/dL lower and 10.75 mg/dL higher in females. Since the interval includes 0 we cannot conclude that we have statistically significant evidence of an association between gender and serum fibrinogen after adjusting for race.

d) In the Caucasian individuals, we estimated mean fibrinogen to be 2.900 mg/dL higher in females than in males. In non-Caucasian individuals, we estimated mean fibrinogen to be 15.7 mg/dL higher in females than in males. Since the non-Caucasian estimated mean difference of 15.7 mg/dL higher is not consistent with the 95% confidence interval of 2.347 mg/dL lower to 8.147 mg/dL, we have evidence of effect modification of the association between gender and fibrinogen by race.

3a) We note that there are 47 subjects missing cholesterol measurements. The sample standard deviation of cholesterol level among the remaining individuals was calculated and found to be 39.29 mg/dL.

b) Assuming that the correlation of cholesterol measurements made two years apart on the same individual is 0.40, the standard deviation of the change in cholesterol measurements made after two years within the population is estimated to be 60.86 mg/dL. (Calculated by squareroot(4\*(1-.4)\*39.29))

c) An estimate for the standard deviation of cholesterol within groups that are homogenous for age and sex was calculated as the mean square error from a multiple linear regression of cholesterol predicted by age and sex. This resulted in an estimated standard deviation of 37.48 mg/dL.

4a) To have 80% power to detect the design alternative will require 530 individuals in the sample.

b) To have 90% power to detect the design alternative will require 709 individuals in the sample.

c) If age and sex are not adjusted for in the study, to have 90% power to detect the design alternative will require 779 individuals in the sample.

d) If only the final cholesterol measurements are analyzed, adjusted for age and sex, the sample size to detect the design alternative will decrease because the variability contributed by each individual will go down.

e) If an Analysis of Covariance model adjusting for age, sex, and baseline cholesterol level is used, then the sample size required to detect the design alternative will decrease.

5a) The estimated number of subjects on the control arm with serum cholesterol below 200 mg/dL at the end of treatment is 0.4032. Of the 5000 observations in the inflammatory biomarkers data, 47 subjects were missing cholesterol measurements. The reported proportion does not include these missing values.

b) The estimated number of subjects on the treatment arm with serum cholesterol below 200 mg/dL at the end of treatment is 0.5045. Of the 5000 observations in the inflammatory biomarkers data, 47 subjects were missing cholesterol measurements. The reported proportion does not include these missing values.

c) The sample size required to provide 90% power to detect the design alternative is 1004 subjects.

d) This study design addresses the question of whether the treatment reduces cholesterol levels below 200 mg/dL, but if we were interested in estimating another effect size then we have lost information by using this dichotomized design rather than the design in 4b. This loss of information is demonstrated by the higher sample size required to obtain the same power and type I error rate.