Biost 518 HW7

2/17/14

1a. **Methods**: The association between sex and mean cholesterol levels was assessed in 2534 Caucasians. Cholesterol was measured as a continuous variable (m/dL). A Z score was calculated using sex-specific means and standard errors of cholesterol levels. Sample means are assumed to be approximately normally distributed.

**Results**: The mean (SE) cholesterol level was 197.5 (1.092) mg/dL for Caucasian males and 222.8 (1.103) mg/dL for Caucasian females. The standard error of the difference in mean cholesterol between sexes is √(1.092^2+1.103^2)=1.552. Thus the Z score is (222.8-197.5)/1.552=16.30, which is statistically significant (p<0.0001). The 95% confidence interval of the difference (difference±1.96\*SE ) is 25.3±1.96\*1.552 (22.26, 28.34).

**Interpretation**: We found an average difference in mean cholesterol of 25.3 mg/dL (95% CI: 22.26, 28.34) between Caucasian females and males (higher in females), an observation that would be highly unusual in the absence of an association between sex and cholesterol levels in Caucasians (p<0.0001).

1b. **Methods**: The association between sex and mean cholesterol levels was assessed in 481 Non-Caucasians. Cholesterol was measured as a continuous variable (m/dL). A Z score was calculated using sex-specific means and standard errors of cholesterol levels. Sample means are assumed to be approximately normally distributed.

**Results**: The mean (SE) cholesterol level was 197.9 (2.557) mg/dL for Non-Caucasian males and 213.6 (2.321) mg/dL for Non-Caucasian females. The standard error of the difference in mean cholesterol between sexes is √(2.557^2+2.321^2)=3.453. Thus the Z score is (213.6-197.9)/3.453=4.547, which is statistically significant (p<0.0001). The 95% confidence interval of the difference is 15.7±1.96\*3.453 (8.932, 22.47).

**Interpretation**: We found an average difference in mean cholesterol of 15.7 mg/dL (95% CI: 8.932, 22.47) between Non-Caucasian females and males (higher in females), an observation that would be highly unusual in the absence of an association between sex and cholesterol levels in Non-Caucasians (p<0.0001).

1c. **Methods**: The association between sex and mean cholesterol levels after adjustment for race was assessed in 3015 subjects. Cholesterol was measured as a continuous variable (m/dL). Sample means are assumed to be approximately normally distributed. Importance weights for Z score calculation based on sample estimates of race were used (i.e. 84.05% for Caucasians, 15.95% for Non-Caucasians).

**Results**: The adjusted estimated difference is (0.8405\*25.3+0.1595\*15.7)/1= 23.77. The standard error of the adjusted estimate can be calculated by:

**

Thus, our SE of the adjusted estimate is √(0.8405^2\*1.552^2+0.1595^2\*3.453^2)/(1^2)=1.416, giving us a Z score of 23.77/1.416=16.79, which is statistically significant (p<0.0001). The 95% confidence interval is 23.77±1.96\*1.416 (20.99, 26.54).

**Interpretation**: We found an average difference in mean cholesterol of 23.77 mg/dL (95% CI: 20.99, 26.54) between females and males (higher in females) after adjustment for race, an observation that would be highly unusual in the absence of an association between sex and cholesterol levels after adjustment for race (p<0.0001).

1d. The mean difference in cholesterol between females and males among Caucasians was 25.3 mg/dL (SE=1.552) while the mean difference among Non-Caucasians was 15.7 mg/dL (SE=3.453). The different mean differences of cholesterol between females and males for Caucasians and Non-Caucasians suggest that race does modify the association between mean cholesterol level and sex.

2a. **Methods**: The association between sex and mean fibrinogen levels was assessed in 2534 Caucasians. Fibrinogen was measured as a continuous variable (m/dL). A Z score was calculated using sex-specific means and standard errors of cholesterol levels. Sample means are assumed to be approximately normally distributed.

**Results**: The mean (SE) fibrinogen level was 317.8 (2.216) for Caucasian males and 320.7 (1.627) for Caucasians females. The standard error of the difference in mean fibrinogen between sexes is √(2.216^2+1.627^2)=2.749. Thus the Z score is (320.7-317.8)/2.749=1.055, which is not statistically significant (p=0.2914). The 95% confidence interval of the difference is 2.9±1.96\*2.749 (-2.488, 8.288).

**Interpretation**: We found an average difference in mean fibrinogen of 2.9 mg/dL (95% CI: -2.488, 8.288) between Caucasian females and males (higher in females), an observation that was not statistically significant (p=0.2914).

2b. **Methods**: The association between sex and mean fibrinogen levels was assessed in 481 Non-Caucasians. Fibrinogen was measured as a continuous variable (m/dL). A Z score was calculated using sex-specific means and standard errors of cholesterol levels. Sample means are assumed to be approximately normally distributed.

**Results**: The mean (SE) fibrinogen level was 333.7 (5.628) mg/dL for male Non-Caucasians and 349.4 (4.643) mg/dL for female Non-Caucasians. The standard error of the difference in mean fibrinogen between sexes is √(5.628^2+4.643^2)=7.296. Thus the Z score is (349.4-333.7)/7.296=2.152, which is statistically significant (p=0.0314). The 95% confidence interval of the difference is 15.7±1.96\*7.296 (1.400, 30.00).

**Interpretation**: We found an average difference in mean fibrinogen of 15.7 mg/dL (95% CI: 1.400, 30.00) between Non-Caucasian females and males (higher in females), an observation that would be unusual in the absence of an association between sex and cholesterol levels in Non-Caucasians (p=0.0314).

2c. **Methods**: The association between sex and mean fibrinogen levels after adjustment for race was assessed in 3015 subjects. Fibrinogen was measured as a continuous variable (m/dL). Sample means are assumed to be approximately normally distributed. Importance weights for Z score calculation based on sample estimates of race were used (i.e. 84.05% for Caucasians, 15.95% for Non-Caucasians).

**Results**: The adjusted estimated difference is (0.8405\*2.9+0.1595\*15.7)/1=4.942. The standard error of the adjusted estimate can be calculated by:

**

Thus, our SE of the adjusted estimate is √(0.8405^2\*2.749^2+0.1595^2\*7.296^2)/(1^2)=2.587, giving us a Z score of 4.942/2.587=1.910, which is not statistically significant (p=0.0561). The 95% confidence interval is 4.942±1.96\*2.587 (-0.1285, 10.01).

Interpretation: We found an average difference in mean fibrinogen of 4.942 mg/dL (95% CI: -0.1285, 10.01) between females and males (higher in females) after adjustment for race, an observation that was not statistically significant (p=0.0561).

2d. The mean difference in fibrinogen between females and males among Caucasians was 2.9 mg/dL (SE=2.749) and 15.7 mg/dL (SE=7.296) for Non-Caucasians. The different mean differences of fibrinogen between females and males for Caucasians and Non-Caucasians suggest that race does modify the association between mean fibrinogen level and sex.

3a. The standard deviation of cholesterol within the sample is 39.29 mg/dL.

3b. We can use the following formula:



Thus, the standard deviation is √(39.29^2+39.29^2-2\*0.40\*39.29\*39.29)=43.04 mg/dL.

3c. The estimated standard deviation of cholesterol within groups with constant age and sex is 37.49 mg/dL.

4a. δαβ: z1-α=z0.975=1.960; zβ=z0.80=0.842; δαβ=1.960+0.842=2.802

Δ=10-0=10

V: 8\*37.49^2\*(1-0.4)=6746.4

N= (δαβ^2)V/Δ^2=2.802^2\*6746.4/10^2=529.7, so round up to 530.

4b. δαβ: z1-α=z0.975=1.960; zβ=z0.90=1.282; δαβ=1.960+1.282=3.242

Δ=10-0=10

V: 8\*37.49^2\*(1-0.4)=6746.4

N= (δαβ^2)V/Δ^2=3.242^2\*6746.4/10^2=709.1, so round up to 710.

4c. δαβ: z1-α=z0.975=1.960; zβ=z0.90=1.282; δαβ=1.960+1.282=3.242

Δ=10-0=10

V: 8\*39.29^2\*(1-0.4)= 7409.8

N= (δαβ^2)V/Δ^2=3.242^2\*7409.8/10^2=778.8, so round up to 779. Thus, not adjusting for age and sex would increase the calculated sample size.

4d. The V term in the sample size calculation would be smaller (2σ2 instead of 2.4σ2) while the other two terms would remain constant. Thus, analyzing only the final cholesterol measurement adjusted for age and sex would give a smaller sample size.

4e. The V term in the sample size calculation would be smaller (1.68σ2 instead of 2.4σ2) while the other two terms would remain constant, so using ANCOVA would give a smaller sample size.

5a. The estimated proportion of subjects on the control arm with serum cholesterol below 200 mg/dL is 0.3957.

5b. The estimated proportion of subjects on the treatment arm with serum cholesterol below 200 mg/dL at the end of treatment is 0.4942.

5c. δαβ: z1-α=z0.975=1.960; zβ=z0.90=1.282; δαβ=1.960+1.282=3.242

Δ=0.4942-0.3957=0.0985

V: 2\*(0.3957\*(1-0.3957)+0.4942\*(1-0.4942))=0.9782

N= (δαβ^2)V/Δ^2=3.242^2\*0.9782/0.0985^2=1059.7, so round up to 1060.

5d. If 200 mg/dL was a clinically important threshold, then dichotomizing the data would give the best scientific answer. However, dichotomizing also leads to loss of information as seen in the increased calculated sample size.