**Biostats 518 - HW7**

**Question 1:**

**a) Are mean cholesterol levels associated with sex in Caucasians?**

The mean cholesterol level among Caucasians was found to be 25.3 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean cholesterol were between 22.26 mg/dl and 28.34 mg/dl, in favor of higher mean cholesterol among females. Based on the 2-sided p-value of <0.0001, we can reject the null hypothesis of no difference in mean cholesterol level between Caucasian males and females.

Calculations:

Difference in mean cholesterol between Caucasian males and females

= 197.5-222.8 = -25.3 mg/dl.

Standard error for comparison:

 = 1.552

Z statistic = -25.3 mg/dl/1.552 = -16.3

Two sided p-value using Stata commands [di 2\*normal(-16.3)] = 9.869x10^-60 = <0.0001

95% CI = -25.3 mg/dl ± 1.96(1.552) = -25.3 mg/dl ± 3.042 = (-28.34, -22.26)

**b) Are mean cholesterol levels associated with sex in Noncaucasians?**

The mean cholesterol level among Noncaucasians was found to be 15.7 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean cholesterol were between 8.932 mg/dl and 22.47 mg/dl, in favor of higher mean cholesterol among females. Based on the 2-sided p-value of <0.0001, we can reject the null hypothesis of no difference in mean cholesterol level between Noncaucasian males and females.

Calculations:

Difference in mean cholesterol between Noncaucasian males and females

=197.9-213.6= -15.7 mg/dl.

Standard error for comparison:

= 3.453

Z statistic = -15.7/3.453 = -4.547

Two sided p-value using Stata commands [di 2\*normal(-4.547)] = 5.442x10^-06 = < 0.0001

95% CI = -15.7 mg/dl ± 1.96(3.453) = -15.7 mg/dl ± 6.76788 = (-8.932, -22.47)

**c) Are mean cholesterol levels associated with sex after adjustment for race?**

Using Importance Weighting:

After adjusting for race (Caucasian and Noncaucasian) using importance weighting (weights based on the proportion of Caucasians and Noncaucasians in our sample), the mean cholesterol level among found to be -23.76 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean cholesterol were between 21.00 mg/dl and 26.54 mg/dl, in favor of higher mean cholesterol among females. Based on the 2-sided p-value of <0.0001, we can reject the null hypothesis of no difference in mean cholesterol level between males and females.

Using Efficiency Weighting:

After adjusting for race (Caucasian and Noncaucasian) using efficiency weighting, under the assumption of no effect modification (weights are the inverse of the square of the standard errors of the race specific estimate), the mean cholesterol level among found to be 23.68 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean cholesterol were between 20.91 mg/dl and 26.46 mg/dl, in favor of higher mean cholesterol among females. Based on the 2-sided p-value of <0.0001, we can reject the null hypothesis of no difference in mean cholesterol level between males and females.

Using efficiency weighting we get slightly different weighted point estimates for the difference in cholesterol, but the same standard error for this weighted difference, as using importance weighting.

Calculations:

Importance Weights:

3,015 subjects: 1,258 male (41.7%); 1,757 female (58.3%)

2,534 Caucasians (84.05%); 481 (15.95%) Noncaucasians

Adjusted Estimate: Δadj = (wC × ΔC + wN × ΔN) / (wC + wN)

= 0.8405 x -25.3+0.1595x-15.7 = -21.26 + -2.504 = -23.76 mg/dl

Standard error:



=squareroot[0.7064(2.4087)+0.02544(11.923)]

=squareroot[1.7015+0.3033] = squareroot(2) = 1.416

95% CI: -23.76 ± 1.96 (1.416) = -23.76 ± 2.775 = -26.535, -20.985

Z statistic = -23.76/1.416 = 16.78

Two sided p-value using Stata commands [di 2\*normal(-16.78)] = 3.418x10^63 = <0.0001

Efficiency Weights

Adjusted Estimate: Δadj = (wC × ΔC + wN × ΔN) / (wC + wN)

Numerator:

=(1/1.5522)(-25.3)+(1/3.4532)(-15.7) = =(1/2.4087)(-25.3)+(1/11.923)(-15.7) =

=(0.4152)(-25.3)+(0.08387)(-15.7)= -11.820

Denominator: (1/1.5522)+(1/3.4532) = 0.4152+0.08387 = 0.4991

Weighted Estimate: 11.820/0.4991 = -23.68

SE = 1.416 (same as importance sampling)

95% CI: -23.68 ± 1.96 (1.416) = -23.68 ± 2.775 = -26.455, -20.905

Z statistic = -23.68/1.416 = 16.72

Two sided p-value using Stata commands [di 2\*normal(-16.72)] = 9.372x10^-63 = <0.0001

**d) Does race modify the effect of sex on cholesterol?**

The difference in mean cholestesterol across groups defined by sex was found to be 9.6 higher in Caucasians than Noncaucasions. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean cholesterol difference were between 2.179 mg/dl and 17.02 mg/dl, in favor of a greater difference in mean cholesterol for Caucasians. Based on the 2-sided p-value of 0.0112, we have sufficient evidence to reject the null hypothesis of no effect modification by race in the association of mean cholesterol level between males and females.

Calculations:

Difference in differences: (-25.3) - (-15.7) = -9.6

SE = squareroot(1.552^2 + 3.453^2) = squareroot(2.4087 + 11.923) = squareroot(14.33)

SE= 3.786

95% CI = -9.6 ± 1.96 (3.786) = -9.6 ± 7.42056 = -17.02, -2.179

Z statistic = -9.6/3.786 = -2.5356

Two sided p-value using Stata commands [di 2\*normal(-2.5356)] =0.0112

**Question 2**

**a) Are mean fibrinogen levels associated with sex in Caucasians?**

The mean fibrinogensterol level among Caucasians was found to be 2.9 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean fibrinogencomparing womento men were between 2.347 mg/dl lower and 8.147 mg/dl higher. Based on the 2-sided p-value of 0.2787, there is not sufficient evidence to reject the null hypothesis of no difference in mean fibrinogenlevel between Caucasian males and females.

Calculations:

Difference in mean fibrinogenbetween Caucasian males and females

= 317.8 mg/dl - 320.7 mg/dl = -2.9 mg/dl.

Standard error for comparison:

 = 2.677

Z statistic = -2.9/2.677 = -1.0833

Two sided p-value using Stata commands [di 2\*normal(-1.0833)] = 0.2787

95% CI = 2.9 mg/dl ± 1.96(2.677) = -2.9 mg/dl ± 5.247 = (-8.147, 2.347)

**b) Are mean fibrinogen levels associated with sex in Noncaucasians?**

The mean fibrinogenlevel among Noncaucasians was found to be 15.7 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean fibrinogenwere between 1.3998 mg/dl and 30.00 mg/dl higher in females compared to male Noncaucasians. Based on the 2-sided p-value of 0.0314, we can reject the null hypothesis of no difference in mean fibrinogenlevel between Noncaucasian males and females.

Calculations:

Difference in mean fibrinogenbetween Noncaucasian males and females

=333.7-349.4= -15.7 mg/dl.

Standard error for comparison:

= 7.296

Z statistic = -15.7/7.296= -2.152

Two sided p-value using Stata commands [di 2\*normal(-2.152)] = 0.03139735

95% CI = -15.7 mg/dl ± 1.96(7.296) = -15.7 mg/dl ± 14.3 = (-30, -1.3998)

**c) Are mean fibrinogen levels associated with sex after adjustment for race?**

Using Importance Weighting:

After adjusting for race using importance weighting (weights based on the proportion of Caucasians and Noncaucasians in our sample), the mean fibrinogenlevel among found to be 4.941 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean fibrinogenwere between 9.906 mg/dl lower to 0.024 mg/dl higher for males compared to females. Based on the 2-sided p-value of 0.05106, there is not sufficient evidence to reject the null hypothesis of no difference in mean fibrinogenlevel between males and females.

Using Efficiency Weighting (achieve approximately same results):

After adjusting for race using efficiency weighting, under the assumption of no effect modification (weights are the inverse of the square of the standard errors of the race specific estimate), the mean fibrinogenlevel among found to be 4.419 mg/dl lower for males than females. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean fibrinogenwere between 9.384 mg/dl lower and 0.546 mg/dl higher for males than females. Based on the 2-sided p-value of 0.08105, there is not sufficient evidence to reject the null hypothesis of no difference in mean fibrinogenlevel between males and females.

Using efficiency weighting we get slightly different weighted point estimates for the difference in fibrinogen, but the same standard error for this weighted difference, as using importance weighting.

Calculations:

Importance Weights:

3,015 subjects: 1,258 male (41.7%); 1,757 female (58.3%)

2,534 Caucasians (84.05%); 481 (15.95%) Noncaucasians

Adjusted Estimate: Δadj = (wC × ΔC + wN × ΔN) / (wC + wN)

= 0.8405 x -2.9 mg/dl.+0.1595x-15.7 = -2.437 + -2.504 = -4.941 mg/dl

Standard error:



=squareroot[0.7064(7.166)+0.02544(53.23)]

=squareroot[5.06206+1.3542] = squareroot(6.416) = 2.533

95% CI: -4.941 ± 1.96 (2.533) = -4.941 ± 4.965 = (-9.906, 0.024)

Z statistic = -4.941 /2.533 = -1.951

Two sided p-value using Stata commands [di 2\*normal(-1.951)] = 0.05105705

Efficiency Weights

Adjusted Estimate: Δadj = (wC × ΔC + wN × ΔN) / (wC + wN)

Numerator:

=(1/2.6772)(-2.9)+(1/7.2962)(-15.7) = 0.13954(-2.9)+ 0.018786(-15.7)

= -0.40467+ - 0.29494 = -0.69961

Denominator: (1/2.6772)+(1/7.2962) = 0.13954+0.018786= 0.158326

Weighted Estimate: -0.69961/0.158326= -4.419

SE = 2.533 (same as importance sampling)

95% CI: -4.419 ± 1.96 (2.533) = -4.419 ± 4.965= -9.384, 0.546

Z statistic = -4.419/2.533= -1.7446

Two sided p-value using Stata commands [di 2\*normal(-1.7446)] = .08105452

**d) Does race modify the effect of sex on fibrinogen?**

The difference in mean fibrinogen across groups defined by sex was found to be 12.8 lower in Caucasians than Noncaucasions. Based on 95% confidences intervals, this observed difference would not be unusual if the true difference in mean fibrinogen difference were between 2.433 mg/dl lower and 28.033 mg/dl higher, comparing Noncaucasians to Caucasians. Based on the 2-sided p-value of 0.09956 we there is not sufficient evidence to reject the null hypothesis of no effect modification by race in the association of mean fibrinogen level between males and females.

Calculations:

Difference in differences: (-2.9) - (-15.7) = 12.8

SE = squareroot(2.677^2 + 7.296^2) = squareroot(7.1663+ 53.232) = squareroot(60.40)

SE= 7.772

95% CI = 12.8 ± 1.96 (7.772) = 12.8 ± 15.233 = (-2.433, 28.033)

Z statistic = 12.8/7.772= 1.647

Two sided p-value using Stata commands [di 2\*normal(1.647)] =0.09955805

**Question 3**

**a) Standard deviation at baseline and after 2 years of treatment (estimate using cross sectional measurement) =** 39.29 mg.dl

**b) Measurement made 2 years apart on same individual is 0.40.**



=

SD = 43.04

**c)** Root MSE = SD = 37.49

**Question 4**

**a) Sample size for 80% power, adjusted for age and sex**

δαβ= Z1- α = Z0.975 = 1.960

Zβ = Z0.80 =0.842

δαβ = 1.960 + 0.842 = 2.802

θ0 = 0; θ1 = -10;

∆= θ1-θ0= -10 - 0 = -10

V = 8σ 2(1-ρ)= 8(37.4922)(1-0.4)=6747

N = δαβ2 V / ∆2 = 2.8022× 6747 /(-10)2= 529.72

Sample size = 530

**b) Sample size for 90% power, adjusted for age and sex**

δαβ= Z1- α = Z0.975 = 1.960

Zβ = Z0.90 =1.282

δαβ = 1.960 + 1.282= 3.242

θ0 = 0; θ1 = -10;

∆= θ1-θ0= -10 - 0 = -10

V = 8σ 2(1-ρ)= 8(37.492)(1-0.4)=6747

N = δαβ2 V / ∆2 = 3.2422× 6747 /(-10)2= 709.15

Sample size = 710

**c) Sample size for 90% power, not adjusted for age and sex**

Use the SD calculated in 3a (39.29). This is a larger SD than used in 4b (37.49) where we did adjust).

δαβ= Z1- α = Z0.975 = 1.960

Zβ = Z0.90 =1.282

δαβ = 1.960 + 1.282= 3.242

θ0 = 0; θ1 = -10;

∆= θ1-θ0= -10 - 0 = -10

V = 8σ 2(1-ρ)= 8(39.292)(1-0.4)=7409.78

N = δαβ2 V / ∆2 = 3.2422× 7409.78 /(-10)2= 778.8

Sample size=779

With the unadjusted SD (greater variability), the sample size must be larger than when we used the adjusted SD, in order to detect an association with 90% power.

**d) Effect of sample size if analyze final measurement, not the change in cholesterol**

In a randomized control trial, if randomization is successful, the distribution of the baseline measurements will be the same in each arm. In this situation where the correlation is less than 0.5, it is more efficient to use only the final measurements. Our correlation between cholesterol measurements is 0.4, so it is more precise to use other final measurement compared to the change. Therefore, a smaller sample size is required if we used the final measurement only. Less variance, the smaller the sample size required.

Final only V=4(37.492) vs. Change V=8(37.492)(1-0.04)

Final V is smaller, so smaller sample size needed.

**e) Effect of sample size if analyze using ANCOVA adjusting for age and sex.**

Using ANCOVA and adjusting for baseline measurement as a covariate would reduce the overall variance, and therefore this method would reduce the sample size compared to using the change in measurements.

ANCOVA V=4(37.492)(1-0.04) vs. Change V=8(37.492)(1-0.04)

ANCOVA V is smaller, so smaller sample size needed.

**Question 5**

**a) Using the inflammatory biomarkers dataset, what is your estimate of the proportion *pC* of subjects on the control arm with serum cholesterol below 200 mg/dL at the end of treatment?**

pc= 0.3957

**b) Using the inflammatory biomarkers dataset, what is your estimate of the proportion *pT* of subjects on the treatment arm with serum cholesterol below 200 mg/dL at the end of treatment? (This is assumed to be equal to the number having cholesterol levels below 210 mg/dL in the CHS data.)**

pt=0.4942

**c) What sample size will provide 90% power to detect the design alternative?**

pc=0.3957

pt=0.4942

δαβ= Z1- α = Z0.975 = 1.960

Zβ = Z0.90 =1.282

δαβ = 1.960 + 1.282= 3.242

θ0 = 0.4942; θ1 = 0.3957;

∆= θ1-θ0= 0.4942 - 0.3957 = 0.0985

V = 2(pt(1-pt)+pc(1-pc)) = 2(0.4942(1-0.4942)+0.3957(1-0.3957)) = 0.978175

N = δαβ2 V / ∆2 = 3.2422× 0.978175/(0.0985)2= 1059.696

Sample size = 1,060

**d) What advantages or disadvantages does this study design have over the study design used in problem 4b?**

This method of dichotomizing cholesterol at 200mg/dl has an advantage of being easily interpretable (compare those with high or low cholesterol based on a cut off of 200mg/dl), and is useful especially if 200 mg/dl is important to the scientific question. However, by dichotomizing cholesterol we lose a lot of information, and a larger sample size is needed to be able to detect the difference between the trial arms. In 4b, the sample size needed is 710, whereas when cholesterol is dichotomized, the sample size needed is 1,060. Dichotomization is less efficient than when we look at a difference in cholesterol treating it continuously. The sample size depends on the scientific question of interest, as different questions will require different sample sizes to have the same statistical power.