BIOST 515/518 Winter 2014

Homework 7 ID 7868

1. (a) Association of mean cholesterol levels with sex in Caucasians:

Point estimate = 222.8 – 197.5 = 25.3

Combined standard error = √(1.1032 + 1.0922) = 1.552

95% confidence intervals = 25.3± 1.96\*1.552 = (22.3, 28.3)

Z-statistic = PE/SE = 16.30

p-value = <0.0005

Mean cholesterol levels were found to be 25.3 mg/dL lower in Caucasian males than in Caucasian females. Such a difference was found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean cholesterol across groups defined by sex (two-sided p < 0.0005). Based on a 95% confidence interval, we find that the observed difference in mean cholesterol is consistent with a true difference of mean cholesterol 22.3 mg/dL lower to 28.3 mg/dL lower comparing Caucasian males to females.

(b) Association of mean cholesterol levels with sex in Noncaucasians:

Point estimate = 213.6 -197.9 = 15.7

Combined standard error = sqrt(2.321^2 + 2.557^2) = 3.453

95% confidence intervals = 15.7 ± 1.96\*3.453 = (8.932, 22.5)

Z-statistic = PE/SE = 4.546

p-value = 5.468e-6 (<0.0005)

Mean cholesterol levels were found to be with 15.7 mg/dL lower in Caucasian males than in Caucasian females. Such a difference was found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean cholesterol across groups defined by sex (two-sided p < 0.0005). Based on a 95% confidence interval, we find that the observed difference in mean cholesterol is consistent with a true difference of mean cholesterol 8.93 mg/dL lower to 22.5 mg/dL lower comparing Noncaucasian males to females.

(c) Importance weights: wC = 0.8405, wN = 0.1595 from sample population

So point estimate after adjustment for race

= ((0.8405\*25.3)+(0.1595\*15.7))/(0.8405+0.1595)

= 23.77

Standard error after adjustment for race

= sqrt((0.8405^2\*1.552^2 + 0.1595^2\*3.453^2)/(0.8405+0.1595)^2)

= 1.416

95% confidence intervals = 23.77 ± 1.96\*1.416 = (20.99, 26.54)

Z-statistic = 16.79

p-value < 0.0005

After adjustment for race using importance weights, mean cholesterol levels were found to be with 23.8 mg/dL lower in males compared to females of the same race. Such a difference was found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean cholesterol across groups defined by sex (two-sided p < 0.0005). Based on a 95% confidence interval, we find that the observed difference in mean cholesterol is consistent with a true difference in mean cholesterol of 21.0 mg/dL lower to 26.5 mg/dL lower comparing males to females of the same race.

Efficiency weights: wC = 1/1.552^2 = 0.4152, wN = 1/3.453^2 = 0.08387 from inverse of square of the standard error of the stratum specific estimate.

So point estimate after adjustment for race

= ((0.4152\*25.3)+(0.08387\*15.7))/(0.4152+0.08387)

= 23.69

Standard error after adjustment for race

= sqrt((0.4152^2\*1.552^2 + 0.08387^2\*3.453^2)/(0.4152+0.08387)^2)

= 1.416

95% confidence intervals = 23.69 ± 1.96\*1.416= (20.91, 26.46)

Z-statistic = 16.73

p-value <0.0005

After adjustment for race using efficiency weights, mean cholesterol levels were found to be with 23.7 mg/dL lower in males compared to females of the same race. Such a difference was found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean cholesterol across groups defined by sex (two-sided p < 0.0005). Based on a 95% confidence interval, we find that the observed difference in mean cholesterol is consistent with a true difference in mean cholesterol of 20.9 mg/dL lower to 26.5 mg/dL lower comparing males to females of the same race.

(d) Estimate of effect = Difference of difference in sample means = 15.7 – 25.3 = -9.6

Estimate std error for estimated interaction contrast

= sqrt(1.552^2+3.453^2) = 3.786

95% confidence intervals = -9.6 ± 1.96\*3.786= (-17.02, -2.178)

Z-statistic = -9.6/3.786 = -2.536

p-value = 0.0112

The difference in mean cholesterol across groups defined by sex was found to be 9.6 mg/dL lower in Noncaucasians than in Caucasians. Such a difference was found to be sufficiently extreme to be able to rule out a null hypothesis of no effect modification by race in the association between cholesterol levels and sex (two-sided p = 0.0112). Based on a 95% confidence interval, we find that the observed difference in the association between cholesterol and sex across race is consistent with a true difference of difference in cholesterol of 17.0 mg/dl lower to 2.18 mg/dl lower comparing Noncaucasians to Caucasians.

2. (a) Association of mean fibrinogen levels with sex in Caucasians:

Point estimate = 320.7-317.8 = 2.9

Combined standard error = sqrt(1.627^2 + 2.126^2) = 2.677

95% confidence intervals = 2.9± 1.96\*2.677 = (-2.347, 8.147)

Z-statistic = PE/SE = 1.083

P-value = 0.2788

Mean fibrinogen levels were found to be with 2.9 mg/dL lower in Caucasian males than in Caucasian females. Such a difference was not found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean fibrinogen across groups defined by sex (two-sided p = 0.2788). Based on a 95% confidence interval, we find that the observed difference in mean fibrinogen is consistent with a true difference of mean fibrinogen 2.35 mg/dL lower to 8.15 mg/dL lower comparing Caucasian males to females.

(b) Association of mean fibrinogen levels with sex in Noncaucasians:

Point estimate = 349.4-333.7 = 15.7

Combined standard error = sqrt(4.643^2 + 5.628^2) = 7.296

95% confidence intervals = 15.7± 1.96\*7.296 = (1.3998, 30.0002)

Z-statistic = PE/SE = 2.152

P-value = 0.0314

Mean fibrinogen levels were found to be with 15.7 mg/dL lower in Noncaucasian males than in Noncaucasian females. Such a difference was found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean fibrinogen across groups defined by sex (two-sided p = 0.0314). Based on a 95% confidence interval, we find that the observed difference in mean fibrinogen is consistent with a true difference of mean fibrinogen 1.40 mg/dL lower to 30.0 mg/dL lower comparing Noncaucasian males to females.

(c) Importance weights: wC = 0.8405, wN = 0.1595 from sample population

So point estimate after adjustment for race

= (0.8405\*2.9)+(0.1595\*15.7)/(0.8405+0.1595)

= 4.942

Standard error after adjustment for race

= sqrt((0.8405^2\*2.677^2 + 0.1595^2\*7.296^2)/(0.8405+0.1595)^2)

= 2.533

95% confidence intervals = 4.942 ± 1.96\*2.533 = (-0.02268, 9.907)

Z-statistic = 1.951

p-value = 0.05108

After adjustment for race using importance weights, mean fibrinogen levels were found to be with 4.94 mg/dL lower in males compared to females of the same race. Such a difference was not found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean fibrinogen across groups defined by sex (two-sided p = 0.05108). Based on a 95% confidence interval, we find that the observed difference in mean fibrinogen is consistent with a true difference in mean fibrinogen of 0.0227 mg/dL higher to 9.91 mg/dL lower comparing males to females of the same race.

Efficiency weights: wC = 1/2.677^2 = 0.1395, wN = 1/7.296^2 = 0.01879 from inverse of square of the standard error of the stratum specific estimate.

So point estimate after adjustment for race

= ((0.1395\*2.9)+( 0.01879 \*15.7))/( 0.1395+0.01879)

= 4.419

Standard error after adjustment for race

= sqrt((0.1395^2\*2.677^2 + 0.01879^2\*7.296^2)/(0.1395+ 0.01879)^2)

= 2.513

95% confidence intervals = 4.419 ± 1.96\*2.513 = (-0.5064, 9.345)

Z-statistic = 1.759

p-value = 0.0787

After adjustment for race using efficiency weights, mean fibrinogen levels were found to be with 4.42 mg/dL lower in males compared to females of the same race. Such a difference was not found sufficiently extreme to be able to rule out a null hypothesis of no difference in mean fibrinogen across groups defined by sex (two-sided p = 0.0787). Based on a 95% confidence interval, we find that the observed difference in mean fibrinogen is consistent with a true difference in mean fibrinogen of 0.506 mg/dL higher to 9.35 mg/dL lower comparing males to females of the same race.

(d) Estimate of effect = Difference of difference in sample means = 15.7-2.9 = 12.8

Estimate std error for estimated interaction contrast

= sqrt(2.677^2+7.296^2) = 7.772

95% confidence intervals = 12.8 ± 1.96\*7.772= (-2.432, 28.03)

Z-statistic = 12.8/7.772= 1.647

p-value = 0.09955

The difference in mean fibrinogen across groups defined by sex was found to be 12.8 mg/dL higher in Noncaucasians than in Caucasians. Such a difference was not found to be sufficiently extreme to be able to rule out a null hypothesis of no effect modification by race in the association between fibrinogen levels and sex (two-sided p = 0.09955). Based on a 95% confidence interval, we find that the observed difference in the association between fibrinogen and sex across race is consistent with a true difference of difference in fibrinogen of 2.432 mg/dl lower to 28.03 mg/dl lower comparing Noncaucasians to Caucasians.

3. (a) The best estimate of the standard deviation of cholesterol within the sample for both baseline and measured after two years of treatment is 39.29 mg/dL.

(b) The standard deviation of the change in cholesterol measurements made after three years within the population is sqrt(39.29^2 + 39.29^2 – 2\*0.4\*39.29\*39.29) = 43.04 mg/dL.

(c) The RMSE from a classical linear regression modeling age and sex continuously is an estimate of the SD within groups that are homogeneous for age and sex, which is 37.49 mg/dL.

4. (a) To find δαβ, z1-α = z0.975 = 1.960; zβ = z0.80 = 0.842, so δαβ = 1.960 + 0.8416 = 2.802

Δ: θ0 = 0; θ1 = -10; Δ = -10

V: V = 8σ2(1-0.4) = 8\*37.49^2\*(1-0.4) = 6746.4

N = δαβ2 \* V/ Δ2 = 2.802^2\*8\*37.49^2\*(1-0.4)/(-10)^2 = 529.67

Rounding up gives N = 530

(b) δαβ, z1-α = z0.975 = 1.960; zβ = z0.90 = 1.282 , so δαβ = 1.960 + 1.282 = 3.242

Δ: θ0 = 0; θ1 = -10; Δ = -10

V: V = 8σ2(1-0.4) = 8\*37.49^2\*(1-0.4) = 6746.4

N = δαβ2 \* V/ Δ2 = 3.242^2\*8\*37.49^2\*(1-0.4)/(-10)^2 = 709.08

Rounding up gives N = 710

(c) Without adjusting for age and sex, SD = 39.29 mg/dL

So δαβ, z1-α = z0.975 = 1.960; zβ = z0.90 = 1.282 , so δαβ = 1.960 + 1.282 = 3.242

Δ: θ0 = 0; θ1 = -10; Δ = -10

V: V = 8σ2(1-0.4) = 8\*39.29^2\*(1-0.4) = 7409.78

N = δαβ2 \* V/ Δ2 = 3.242^2\*8\*39.29^2\*(1-0.4)/(-10)^2 = 778.91

Rounding up gives N = 779

For equal numbers in each group, we use N = 780.

(d) If only the final cholesterol measurement adjusted for age and sex instead of the change was analyzed, other variables remain the same while V = σ2(1/r+2+r) where r = 1, so V = 4σ2 < 4.8σ2, thus the sample size to be accrued will be smaller.

(e) If an Analysis of Covariance model adjusted for age, sex and baseline cholesterol level were used, other variables remain the same while V = 4σ2(1-0.4^2) = 3.36σ2 < 4.8σ2, thus the sample size to be accrued will be smaller, also even smaller than the situation in (d).

5. (a) Dichotomizing the data into groups defined by cholesterol levels, using 200 mg/dL as a cutoff point, the estimate of the proportion pC of subjects on the treatment arm with serum cholesterol below 200mg/dL at the end of treatment is 0.392.

(b) Dichotomizing the data into groups defined by cholesterol levels, using 210 mg/dL as a cutoff point, the estimate of the proportion pT of subjects on the treatment arm with serum cholesterol below 200mg/dL at the end of treatment is 0.4896.

(c) So δαβ, z1-α = z0.975 = 1.960; zβ = z0.90 = 1.282 , so δαβ = 1.960 + 1.282 = 3.242

Δ = θ1 = 0.4896-0.392 = 0.0976

V: V = 2\*(0.4896\*(1-0.4896) + 0.392 \*(1-0.392)) = 0.9766

N = δαβ2 \* V/ Δ2 = 3.242^2\*2\*(0.4996\*(1-0.4996) + 0.392\*(1-0.392))/(0.1004)^2

= 1077.56

Rounding up gives N = 1078

(d) The advantage of this study design is, if we care about the clinical relevance of lowering serum cholesterol levels by 10mg/dL, this study design answers the most relevant scientific question. However, if such a scientific threshold did not exist, then by dichotomizing the data, we lose information about how the treatment might tend to lower cholesterol levels across the population. This loss of information is reflected in the higher sample size requirements when dichotomizing the data: 1078 versus 779 (1077 versus 780 for equal sample sizes in each arm).