**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2014

**Homework #7**

February 17, 2014

**Written problems:** To be submitted as a MS-Word compatible file to the class Catalyst dropbox by 9:30 am on Monday, February 24, 2014. See the instructions for peer grading of the homework that are posted on the web pages.

***Note: You may find the keys to homeworks 1 and 3 from the Winter 2006 offering of Biost 518 of use in solving the questions on this homework.***

*On this (as all homeworks) Stata / R code and unedited Stata / R output is* ***TOTALLY*** *unacceptable. Instead, prepare a table of statistics gleaned from the Stata output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*

***Unless explicitly told otherwise in the statement of the problem, in all problems requesting “statistical analyses” (either descriptive or inferential), you should present both***

* ***Methods: A brief sentence or paragraph describing the statistical methods you used. This should be using wording suitable for a scientific journal, though it might be a little more detailed. A reader should be able to reproduce your analysis. DO NOT PROVIDE Stata OR R CODE.***
* ***Inference: A paragraph providing full statistical inference in answer to the question. Please see the supplementary document relating to “Reporting Associations” for details.***

**Questions 1 and 2** suppose that you are reading a scientific article in a journal with inadequate statistical review. The scientific question addressed by the article is the association between blood lipid profiles (especially total cholesterol), biomarkers of inflammation (fibrinogen), and mortality from cardiovascular disease. The authors were also interested in the role of race (as categorized by Caucasian and Noncaucasian) in the relationship between sex and the serum measurements of total cholesterol and fibrinogen.

The authors reported gathering data on 3,015 subjects, of whom 1,258 were male and 1,757 were female. The subjects were further characterized as 2,534 Caucasians, 481 Noncaucasians. The data analysis presented in the manuscript is limited to the means and standard errors of the serum measures within subgroups as given in the following table.

**Table 1. Means (standard errors) of serum cholesterol and fibrinogen according to patient sex and race.**

|  |  |  |
| --- | --- | --- |
|  | **Males** | **Females** |
| **Caucasians** | **Noncaucasians** | **Caucasians** | **Noncaucasians** |
| **Cholesterol (mg/dl)** | 197.5 (1.092) | 197.9 (2.557) | 222.8 (1.103) | 213.6 (2.321) |
| **Fibrinogen (mg/dl)** | 317.8 (2.126) | 333.7 (5.628) | 320.7 (1.627) | 349.4 (4.643) |

1. You desire to do a more careful evaluation of the evidence at hand for associations between sex and cholesterol. You therefore desire to compute estimates, 95% confidence intervals, and P values to address questions of associations within subgroups, associations adjusted for race, and effect modification. In addressing the following questions, provide a sentence that interprets your inferential statistics in a manner suitable for inclusion in a scientific journal article. Avoid statistical jargon. (You note that without the sample sizes by subgroup, you will not be able to use the exact statistical methods (i.e., t tests) that you might otherwise have, but you will be able to perform analyses based on large sample approximations and the fact that sample means are approximately normally distributed. The Stata function normal() will return the cumulative distribution function for the standard normal. Hence,

di normal(1.96)

 will display 0.9750021. In R, the equivalent function is pnorm().)

For problems 1 and 2, the following equations were used:

95% Confidence Intervals: ϴlower = ϴ - z(1-alpha/2)\*se(ϴ); ϴupper=ϴ+z(1-alpha/2)\*se(ϴ)

ϴ= difference in means (male-female)

se(ϴ) = sqrt(se1^2-se2^2) where se1 is the standard error for males, se2 is the standard error for females

p values: p=2\*pnorm(-abs((ϴ-0)/se(ϴ))) [pnorm function used in R]

Δadj = (wc\*Δc+wn\*Δn)/(wc+wn)

se(Δadj) = sqrt((Δc\*sec^2 + Δn\*sen^2)/( Δc+Δn)^2)

* 1. Are mean cholesterol levels associated with sex in Caucasians? (Recall that the standard error of two independent statistics is the square root of the sum of the squares of the individual standard errors. Thus calculate the standard error for the difference in mean cholesterol using the standard errors for the males and females.)

Methods:

In order to compare the mean cholesterol levels of Caucasian males and females, an estimate of the difference in means was computed along with a standard error for the estimate. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

The mean serum Cholesterol for Caucasian males is estimated to be 25.3 mg/dl lower than that for Caucasian females, with a standard error of 1.55 mg/dl. According to a 95% confidence interval, finding a mean cholesterol level for Caucasian males anywhere between 22.3 and 28.3 mg/dl lower than that of Caucasian females would not be unusual. A 2-sided p-value testing the null hypothesis that the difference in means is equal to zero is less than 0.005, that indicates statistical significance allowing for the rejection of the null hypothesis of no association.

* 1. Are mean cholesterol levels associated with sex in Noncaucasians?

Methods:

In order to compare the mean cholesterol levels of non-Caucasian males and females, an estimate of the difference in means was computed along with a standard error for the estimate. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

The mean serum Cholesterol for non-Caucasian males is estimated to be 15.7 mg/dl lower than that for non-Caucasian females, with a standard error of 3.45 mg/dl. According to a 95% confidence interval, finding a mean cholesterol level for non-Caucasian males anywhere between 8.93 and 22.5 mg/dl lower than that of non-Caucasian females would not be unusual. A 2-sided p-value testing the null hypothesis that the difference in means is equal to zero is less than 0.005, indicating statistical significance allowing for the rejection of the null hypothesis of no association.

* 1. Are mean cholesterol levels associated with sex after adjustment for race? Provide adjusted estimates using both importance and efficiency weights.

*An approach that can be used here is to find a weighted average of the measures of effect in each race group. Hence, you might use a weighted average of the estimates ΔC and ΔN you derived in parts a and b, respectively: Let the adjusted estimated be defined according to*

*Δadj = (wC × ΔC + wN × ΔN) / (wC + wN)*

*where wC and wN are relative weights to be applied to the two strata. (Note that the equation becomes simpler if we ensure that the relative weights sum to 1.) The SE of the adjusted estimate of effect is then found by using the properties of variances. Recall that when multiplying a random variable by a constant, Var(cX) = c2 Var(X). Hence, you can find the standard error of the adjusted estimate can be found by*

**

*Many options could be considered for choosing the weights. Two that might be considered include:*

* + - *Importance weights: We weight each stratum according to its relative importance in the population of interest. This could be estimated from our sample (84.05% of our sample was Caucasian, so we could assume that that was also the frequency in the general population of elderly adults) or taken from, say, US census data (86.37% of US residents aged 65 years or older are Caucasian).*
		- *Efficiency weights: Under the assumption of no effect modification, the most efficient analysis would be to weight each stratum in proportion to the inverse of the square of the standard error of the stratum specific estimate.*

Methods:

In order to compare the mean cholesterol levels of non-Caucasian males and females after adjustment for race, an estimate of the difference in means was computed along with a standard error for the estimate using both importance weights and efficiency weights. Importance weights were based on proportions of the sample population that were Caucasian (84.05%) and those that were non-Caucasian (15.95%). Efficiency weights were based on the inverse of the square of the standard error of the difference in mean Cholesterol between the sexes for both Caucasian and non-Caucasian subjects, respectively. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

 *Importance Weighted:*

After adjustment for race (Caucasian or non-Caucasian), the difference in mean serum Cholesterol level was estimated to be 23.77 mg/dl lower for males than for females, with a standard error of 1.416 mg/dl. A 95% confidence interval suggests that finding mean serum cholesterol levels for males anywhere between 20.99 and 26.54 mg/dl lower than that for females would not be unusual. A two-sided p-value testing the null hypothesis that the difference in means is zero is less than 0.005, suggesting statistical significance at the 0.05 alpha level allowing for the rejection of the null hypothesis of no association between serum cholesterol and sex after adjustment for race.

*Efficiency Weighted:*

After adjustment for race (Caucasian or non-Caucasian), the difference in mean serum Cholesterol level was estimated to be 11.82 mg/dl lower for males than for females, with a standard error of 0.7074 mg/dl. A 95% confidence interval suggests that finding mean serum cholesterol levels for males anywhere between 10.43 and 13.20 mg/dl lower than that for females would not be unusual. A two-sided p-value testing the null hypothesis that the difference in means is zero is less than 0.005, suggesting statistical significance at the 0.05 alpha level allowing for the rejection of the null hypothesis of no association between serum cholesterol and sex after adjustment for race.

* 1. Does race modify the association between mean cholesterol level and sex?

Methods:

In order to assess modification of the association between mean cholesterol level and sex by race, the difference between the difference in means between sexes calculated for each race group (Caucasian and non-Caucasian) was estimated along with a standard error for the estimate. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

The difference in mean serum Cholesterol between males and females that are Caucasian is estimated to be 9.600 mg/dl lower than that for non-Caucasians, with a standard error of 3.786 mg/dl. According to a 95% confidence interval, finding a difference in mean serum cholesterol between males and females for Caucasians anywhere between 2.179 and 17.02 mg/dl lower than that for non-Caucasians would not be unusual. A 2-sided p-value testing the null hypothesis that the difference in means is equal to zero is 0.0112 (<0.05), indicating statistical significance allowing for the rejection of the null hypothesis of no association between the two differences in means (indicators of sex-cholesterol association) in the two different race groups. In other words, the null hypothesis that race does not modify the association between sex and serum cholesterol levels is able to be rejected.

1. You also desire to do a more careful evaluation of the evidence at hand for fibrinogen. You therefore answer the questions of problem 1 using the statistics for fibrinogen.
	1. Are mean fibrinogen levels associated with sex in Caucasians

Methods:

In order to compare the mean fibrinogen levels of Caucasian males and females, an estimate of the difference in means was computed along with a standard error for the estimate. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

The mean serum fibrinogen for Caucasian males is estimated to be 2.90 mg/dl lower than that for Caucasian females, with a standard error of 2.68 mg/dl. According to a 95% confidence interval, finding a mean fibrinogen level for Caucasian males anywhere between 2.35 mg/dl higher and 8.15 mg/dl lower than that of Caucasian females would not be unusual. A 2-sided p-value testing the null hypothesis that the difference in means is equal to zero is 0.279 indicates lack of statistical significance at the 0.05 alpha level, not allowing for the rejection of the null hypothesis of no association.

* 1. Are mean fibrinogen levels associated with sex in Noncaucasians?

Methods:

In order to compare the mean fibrinogen levels of non-Caucasian males and females, an estimate of the difference in means was computed along with a standard error for the estimate. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

The mean serum fibrinogen for non-Caucasian males is estimated to be 15.7 mg/dl lower than that for non-Caucasian females, with a standard error of 7.30 mg/dl. According to a 95% confidence interval, finding a mean fibrinogen level for non-Caucasian males anywhere between 1.40 and 30.0 mg/dl lower than that of non-Caucasian females would not be unusual. A 2-sided p-value testing the null hypothesis that the difference in means is equal to zero is 0.0314 (<0.05), indicating statistical significance allowing for the rejection of the null hypothesis of no association.

* 1. Are mean fibrinogen levels associated with sex after adjustment for race?

Methods:

In order to compare the mean fibrinogen levels of non-Caucasian males and females after adjustment for race, an estimate of the difference in means was computed along with a standard error for the estimate using both importance weights and efficiency weights. Importance weights were based on proportions of the sample population that were Caucasian (84.05%) and those that were non-Caucasian (15.95%). Efficiency weights were based on the inverse of the square of the standard error of the difference in mean fibrinogen between the sexes for both Caucasian and non-Caucasian subjects, respectively. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

 *Importance Weighted:*

After adjustment for race (Caucasian or non-Caucasian), the difference in mean serum fibrinogen level was estimated to be 4.942 mg/dl lower for males than for females, with a standard error of 2.533 mg/dl. A 95% confidence interval suggests that finding mean serum fibrinogen levels for males anywhere between 0.0235 mg/dl higher and 9.907 mg/dl lower than that for females would not be unusual. A two-sided p-value testing the null hypothesis that the difference in means is zero is 0.0511(>0.05), suggesting a lack of statistical significance at the 0.05 alpha level preventing the rejection of the null hypothesis of no association between serum fibrinogen and sex after adjustment for race.

*Efficiency Weighted:*

After adjustment for race (Caucasian or non-Caucasian), the difference in mean serum fibrinogen level was estimated to be 0.6996 mg/dl lower for males than for females, with a standard error of 0.3979 mg/dl. A 95% confidence interval suggests that finding mean serum fibrinogen levels for males anywhere between 0.0803 mg/dl higher and 1.479 mg/dl lower than that for females would not be unusual. A two-sided p-value testing the null hypothesis that the difference in means is zero is 0.0787 (>0.05), suggesting lack of statistical significance at the 0.05 alpha level preventing the rejection of the null hypothesis of no association between serum fibrinogen and sex after adjustment for race.

* 1. Does race modify the association between mean fibrinogen level and sex?

Methods:

In order to assess modification of the association between mean fibrinogen level and sex by race, the difference between the difference in means between sexes calculated for each race group (Caucasian and non-Caucasian) was estimated along with a standard error for the estimate. 95% confidence intervals and a 2-sided p-value were computed using the Wald statistic.

Results:

The difference in mean serum Fibrinogen between males and females that are Caucasian is estimated to be 12.80 mg/dl higher than that for non-Caucasians, with a standard error of 7.772 mg/dl. According to a 95% confidence interval, finding a difference in mean serum fibrinogen between males and females for Caucasians anywhere between 2.432 mg/dl lower and 28.03 mg/dl higher than that for non-Caucasians would not be unusual. A 2-sided p-value testing the null hypothesis that the difference in means is equal to zero is 0.0996 (>0.05), indicating lack of statistical significance preventing the rejection of the null hypothesis of no association between the two differences in means (indicators of sex-fibrinogen association) in the two different race groups. In other words, the null hypothesis that race does not modify the association between sex and serum fibrinogen levels cannot be rejected.

**Questions 3 – 5** relate to the planning of a phase III clinical trial of a dietary intervention intended to improve cardiovascular health in a population of elderly adults by lowering serum cholesterol. Because we anticipate using an elderly patient population similar to that used in the cardiovascular health study, we will use the data in inflamm.txt (on the class web pages) to obtain estimates of the variances and correlations necessary to obtain power and sample size.

We consider below several different approaches which differ in the definition of the “treatment effect” θ. I note here (and again below), that several of the options we consider would be considered highly inappropriate for a real study.

We desire to calculate the sample size required to detect a hypothesized effect of the new treatment on patient outcome.

* We choose some summary measure of the treatment effect. We will call this θ.
	+ If we only have a single treatment group, common choices might be a mean, median, proportion above some threshold, etc.
	+ If we have both an experimental treatment group and a control group, then we might choose the difference in means, difference in medians, odds ratio, etc.
* We imagine that a treatment that does nothing beneficial would correspond to a “null treatment effect” of θ = θ0.
	+ In a one arm (i.e., single treatment group) study, the choice of null treatment effect will have to rely on some prior information. (And it is scientifically far less rigorous to have to rely on the “constancy” of estimates across studies.)
	+ In two arm studies (i.e., studies with a treatment group and a control group), the null treatment effect is most often a difference of 0 or a ratio of 1 for some summary measure across treatment groups.
* We want to a low probability of declaring statistical significance when the treatment has the null treatment effect of θ = θ0.
	+ The statistical “type 1 error” is the probability of declaring statistical significance for the value of θ = θ0.
	+ Common choices of type 1 error are 0.05 for a two-sided test and 0.025 for a one-sided test.
* We want to be relatively confident of declaring statistical significance when the treatment has a treatment effect of θ = θ1.
	+ The statistical “power” function is the probability of declaring statistical significance for each value of θ.
	+ Common choices of power are 80% - 97.5%.
* We will use frequentist hypothesis testing based on some test statistic *Z*.
	+ Typically *Z* will involve some estimated treatment effect, the null hypothesis, and an estimated standard error: Z = (estimate – hypothesis) / std.error
	+ For the problems we consider in this homework, *Z* will be approximately normally distributed, and under the null hypothesis, *Z* will have mean 0 and variance 1.
* Hence, if we observe *Z=z,* we can compute the one-sided upper P value as the probability that a standard normal random variable would be greater than *z,* This probability can be computed using a computer program.
	+ In Stata, the probability can be found by using normal( ) function. For instance, if we observed *Z* = 0.8410, the upper P value can be found from the Stata command disp 1 - normal(0.8410). (Stata would then display .20017397.)
	+ In Excel, we could use the function normdist( ). For instance, if *Z* = 0.8410, the lower P value can be found from by typing into an empty cell the Excel formula

=normdist(0.8410,0,1,TRUE).

where the 0 and 1 indicate that you want the normal distribution that has mean 0 and variance 1, and the TRUE indicates that you want the cumulative probability, rather than the density function. (Excel would then display .79982603.)

* In R or S-Plus, we could use the function pnorm( ). For instance, if *zp* = 0.8410, the value of *p* can be found from the R or S-Plus command pnorm(0.8410). (The program would then display .79982603.)
* In the formulas for sample size, we more often want the value of the quantile *zp* such that the probability that a standard normal *Z* is less than *zp* is *p*.
	+ In Stata, the *p*-th quantile can be found by using invnorm( ) function. For instance, if we wanted *z0.80*, the 80th percentile can be found from the Stata command disp invnorm(0.80). (Stata would then display .8410.)
	+ In Excel, the value of *zp* can be found by using the function norminv( ). For instance, if α = 0.025, in our sample size formulas given below, we might want the 100(1 - .025)% percentile. The value of *z0.975* can be found by typing into an empty cell the Excel formula

=norminv(0.975,0,1)

where the 0 and 1 indicate that you want the normal distribution that has mean 0 and variance 1. (Excel would then display 1.959964.)

* + In R or S-Plus, we could use the function pqnorm( ). For instance, if we want *z0.975*, the value can be found from the R or S-Plus command qnorm(0.975). (The program would then display 1.959964.)

For our measure of treatment outcome, we could consider

* A surrogate clinical outcome of serum cholesterol after 2 years of treatment. We can summarize this clinical outcome according to (among others)
* mean cholesterol after 2 years of treatment,
* mean change in cholesterol after 2 years of treatment,
* geometric mean cholesterol after 2 years of treatment,
* median change in cholesterol after 2 years of treatment,
* probability of a cholesterol less than 200 mg/dL after 2 years of treatment
* The clinically relevant treatment outcome of myocardial infarction free survival (i.e., time to the earlier of myocardial infarction or death).

Recall from lecture that the most common formula used in sample size calculations is



where

* *N* is the total sample size to be accrued to the study,
* *V* is the average variability contributed by each subject to the estimate of the treatment effect θ (for each problem below, I provide the formula for *V*),
* *δαβ* is a “standardized alternative” which would allow a standardized one-sided level α hypothesis test to reject the null hypothesis with probability (power) β (note that many textbooks use notation in which the power is denoted 1-β), and
* *Δ* is some measure of the distance between the null and alternative hypotheses.

Often clinical trials are conducted with a stopping rule which allows early termination of the study on the basis of one or more interim analyses of the data. When such a “group sequential test” is to be used, the value of the standardized alternative *δαβ* must be found using special computer software. On the other hand, when a “fixed sample study” (i.e., one in which the data are analyzed only once) is to be conducted, the standardized alternative for a one-sided test is given by



where *zp* is the *p*th quantile of the standard normal distribution. For a two-sided level α test, the standardized alternative is given by



The value of *zp* can be found from Stata, Excel, or R as described above.

The formula for *Δ* depends on the statistical model used, but is usually either

* *Δ = θ1 - θ0* (used for inference in “additive models” for means and proportions, and sometimes medians), or
* *Δ = log(θ1 / θ0)* (used for inference in “multiplicative models” for geometric means, odds, and hazards, and sometimes means and medians),
1. **(Obtaining estimates for use in sample size calculations when using mean cholesterol)** When making inference about cholesterol using means (and differences of means), the formula for *V* will typically involve the standard deviation *σ* of measurements made within a treatment group. The following estimates should be used as needed to answer all other questions. Using the inflamm.txt dataset available on the class web pages.
	1. Ideally, we want the standard deviation of cholesterol at baseline and the standard deviation of cholesterol measured after two years of treatment. However, as we only have ready access to a single cross-sectional measurement, we will have to use that data to estimate both SDs. What is your best estimate of the standard deviation of cholesterol within the sample? Report using four significant digits.

The standard deviation of the data resulting from a single cross-sectional measurement is calculated to be 39.29 mg/dl.

* 1. Assuming that the correlation ρ of cholesterol measurements made two years apart on the same individual is ρ = 0.40, what is the standard deviation of the change in cholesterol measurements made after three years within the population? Report using four significant digits.

V=4\*σ^2\*(1- ρ) for a general case of the change in mean difference in mean summary measure. By using σ= 39.29 (as calculated in part a) and ρ = 0.40, and the equation sd=sqrt(V/n), sd=0.8608.

* 1. We could also consider an analysis that would adjust for age and sex. In such a setting, we would want an estimate of the SD within groups that are homogenous for age and sex. What is your best estimate of the standard deviation of cholesterol within groups that had constant age and sex? Report using four significant digits. (Hint: Recall that the output from a regression model will provide an estimate of a common SD within groups as the “root mean squared error”. So you will need to perform a regression that allows each age-sex combination to have its own mean. A linear regression modeling age continuously along with sex would be one approach.)

A linear regression was performed utilizing the inflammation dataset, predicting serum cholesterol baseline levels using age and sex as predictor variables. The resulting root mean squared error, i.e., an estimate of the SD within groups homogenous for age and sex, is 37.49.

1. **(A two arm study of change in cholesterol after 2 years of treatment with adjustment for age and sex)** Suppose we randomly assign *N* subjects to receive either the new treatment or a control strategy. We use a randomization ratio of 1 subject on the new treatment to 1 subject on control. We use as our measure of treatment effect the mean change in cholesterol at the end of treatment for patients on the new treatment and mean change in cholesterol at the end of treatment for patients on control. The null hypothesis is that the difference in means is 0 mg/dL, and we want to detect whether the new treatment will result in an average change in cholesterol that is 10 mg/dL lower than might be expected on control.. We intend to perform a hypothesis test in which
* we adjust for age and sex,
* the one-sided level of significance is α = 0.025,
* the desired statistical power is β = 0.80 or 0.90,
* the measure of treatment effect is *θ = (μ T,2 - μ T,0 ) – (μ C,2 - μ C,0 )* (the mean change in cholesterol in the patients receiving the new treatment for 2 years of treatment minus the mean change in cholesterol in the patients treated with control for two years), and
* the average variability contributed by each subject to the estimated treatment effect (the difference in sample means) is *V= 8σ 2(1-ρ).* (Again, use a correlation of 0.4.)
* the comparison between alternative and null hypotheses is *Δ = θ1 - θ0*.
1. What sample size will provide 80% power to detect the design alternative?

For 80% power, β = 0.80.

N = (δ^2 \* V)/Δ^2

δ = $z\_{1-α}$+$z\_{β}$; $z\_{1-α}=z\_{1-0.025}=1.960; z\_{β}=z\_{0.80}=0.842; $δ=1.960+0.842=2.802

Δ = $θ\_{1}-θ\_{0}$; $θ\_{0}$=0; $θ\_{1}$=10; Δ= 10

V = 8\*σ^2\*(1- ρ); σ = standard deviation of sample adjusted for age and sex = 37.49; V=8\*(37.49^2)\*(1-0.4) = 6746.4

N = ((2.802^2)\*6746.4)/(10^2) = 529.7 -> **530=N**

1. What sample size will provide 90% power to detect the design alternative?

See calculations for part (a), except instead use β = 0.90; thus δ = 1.960 + $z\_{0.90}$= 1.960+1.282 = 3.242 and N = ((3.242^2)\*6746.4)/(10^2) = **709=N**

1. How would the sample size for 90% power change if you had not decided to adjust for age and sex?

The calculation for sample size adjusted for age and sex utilized an RMSE of 37.49 that was an ouput of a linear regression modeling cholesterol by both age and sex; this value was used for σ in the equation V = 8\*σ^2\*(1- ρ) used to calculate N = (δ^2 \* V)/Δ^2. If an uncorrected value was used for σ (39.30, as calculated in part (a) of problem 3), V (the average variability contributed by each subject) would be higher (7414) as would the sample size N (780).

1. What would be the effect on your sample size computation if you had decided to analyze only the final cholesterol measurement adjusted for age and sex (i.e., not the change)? (A qualitative answer is sufficient.)

If only the final cholesterol measurement was analyzed, the null and alternative hypotheses would be different; instead of the measure of treatment effect being θ = (μ T,2 - μ T,0 ) – (μ C,2 - μ C,0 ) it would be θ = (μ T,2 ) – (μ C,2 ) and this would affect the Δ term in the denomiator of the sample size equation. The value of V, the average variability contributed by each subject, would be different; instead of V = 8\*σ^2\*(1- ρ) (for a ρ value of 0.4, this would be V=4.8\* σ^2) , V would equal 4\* σ^2 and would be lower, resulting in a lower calculated sample size.

1. What would be the effect on your sample size computation if you had decided to use an Analysis of Covariance model that adjusted for age, sex, and the baseline cholesterol level? (A qualitative answer is sufficient.)

With an added adjustment for the baseline cholesterol level in the Analysis of Covariance, the value of V, average variability contributed by each subject, would be V=4\* σ^2\*(1- ρ^2). For ρ=0.4, this equals V=3.36\* σ^2, which would result in a lower value of V than by using a value for V that doesn’t correct for baseline cholesterol (for which V=4.8\* σ^2 for ρ=0.4) and a lower sample size.

1. **(A two arm study of cholesterol after 2 years of treatment and the effect of dichotomizing the data)** Suppose we choose to provide the new treatment to *N* subjects. We use as our measure of treatment effect the proportion of subjects having cholesterol below 200 mg/dL at the end of treatment. We are guessing that the new treatment will result instead in an average cholesterol of 135 mm Hg. We intend to perform a hypothesis test in which
* the one-sided level of significance is α = 0.025,
* the desired statistical power is β = 0.90,
* we presume that the proportion *pC* of subjects on the control arm with serum cholesterol below 200 mg/dL will be the same as was observed in the CHS inflamm.txt data set.
* we presume that the treatment will tend to lower serum cholesterol by 10 mg/dL on average, so the proportion *pT* of subjects on the treatment arm with serum cholesterol below 200 mg/dL will be the same as was observed in the CHS inflamm.txt data set for cholesterol levels below 210 mg/dL.
* the measure of treatment effect is *θ1 = pT, - pC* (the difference in the proportion of subjects receiving the new treatment who have cholesterol lower than 200 mg/dL minus the corresponding proportion on the control arm after 2 years of treatment). Under the null hypothesis, we assume there would be no difference between the treatment arms.,
* the average variability contributed by each subject to the estimated treatment effect (the sample proportion) is *V=2( pT,(1- pT, ) + pC (1 - pC ))*(most often, we would compute this under the alternative hypothesis in this setting),
* the comparison between alternative and null hypotheses is *Δ = θ1 - θ0 = θ1*.
1. 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?

Assuming the proportion of subjects in the control arm with serum cholesterol below 200 mg/dl at the end of treatment is equal to the proportion with serum cholesterol below 200 mg/dl in the dataset, Pc=0.4032. (This was calculated by creating a variable that =0 for cholest>=200 and =1 for cholest<200 and taking the average).

1. 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.)

Assuming the proportion of subjects in the treatment arm with serum cholesterol below 200 mg/dl at the end of treatment is equal to the proportion with serum cholesterol below 210 mg/dl in the dataset, Pc=0.5045. (This was calculated by creating a variable that =0 for cholest>=210 and =1 for cholest<210 and taking the average).

1. What sample size will provide 90% power to detect the design alternative?

N = (δ^2 \* V)/Δ^2

δ = $z\_{1-α}$+$z\_{β}$; $z\_{1-α}=z\_{1-0.025}=1.960; z\_{β}=z\_{0.90}=1.282; $δ=1.960+1.282=3.242

Δ = $θ\_{1}-θ\_{0}$; $θ\_{0}$=0; $θ\_{1}$=Pt-Pc=0.5045-0.4032; Δ= 0.1013

V = 2\*(Pt\*(1-Pt) + Pc\*(1-Pc)) = 2\*(0.5045\*(1-0.5045) + 0.4032\*(1-0.4032)) = 0.9812

N = ((3.242^2)\*0.9812)/(0.1013^2) = **1005=N**

1. What advantages or disadvantages does this study design have over the study design used in problem 4b?

For this particular application of using proportions as the summary measure, the statistical precision is low given the sample proportions close to 0.5 (0.4032 compared with 0.5045) as opposed to proportions close to 0 or 1 in the ideal case. However, given a correlation ρ=0.4 used in the study design in problem 4b, the use of change in mean effect over the study period is not terribly precise either.

**Discussion Sections: February 19 - 21, 2014**

We begin discussion of the university salary dataset.