**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().)

* 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.)
     + **Mean cholesterol levels were found to be 25.3 mg/dL greater in female Caucasians than in male Caucasians. 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 binary gender (P<0.0001). Based on a 95% confidence interval, we find that the observed difference in mean cholesterol is not atypical of settings in which the true difference in mean cholesterol were such that female Caucasians had mean cholesterol 22.26 mg/dL higher to 28.34 mg/dL higher than male Caucasians.**
  2. Are mean cholesterol levels associated with sex in Noncaucasians?
     + **Mean cholesterol levels were found to be 15.7 mg/dL greater in female Non-Caucasians than in male Non-Caucasians. Based on a 95% confidence interval, it would not be unusual if the true mean cholesterol for Non-Caucasian females were found to be 8.93 mg/dL to 22.5 mg/dL greater than mean cholesterol in Non-Caucasian males. At the 5% significance levels, we reject the null hypothesis of no difference in mean cholesterol across groups defined by binary gender (P<0.0001).**
  3. 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.*
    - **After adjustment for race, using importance weights based on sample proportions, mean cholesterol levels were found to be 23.8 mg/dL greater in female subjects than male subjects of the same race. Based on a 95% confidence interval, it would not be unusual if the true mean difference in mean cholesterol were 12.7 mg/dL to 34.9 mg/dL in female subjects than male subjects sharing the same race. At the 5% significance level, we reject the null hypothesis of no difference in mean cholesterol across groups defined by binary gender (P<0.0001).**
    - **Using efficiency weights as recommended in the prompt (inverse square of adjusted standard errors), after adjustment for race, mean cholesterol levels were found to be 23.7 mg/dL greater in female subjects than male subjects of the same race. Based on a 95% confidence interval, it would not be unusual if the true mean difference in mean cholesterol were 20.9 mg/dL to 26.5 mg/dL in female subjects than male subjects sharing the same race. At the 5% significance level, we reject the null hypothesis of no difference in mean cholesterol across groups defined by binary gender (P<0.0001).**
  1. Does race modify the association between mean cholesterol level and sex?
     + **The difference in mean cholesterol across groups defined by gender was found to be 7.6 mg/dL greater in Caucasians than in Non-Caucasians. Based on a 95% confidence interval, it would not be unusual if the true mean difference in effect were found anywhere between 0.179 mg/dL greater and 15.02 mg/dL greater in Caucasians than Non-Caucasians. At the 5% significance level, we reject the null hypothesis of no effect modification by race in the association between cholesterol and gender (P=0.0449).**

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
      * **Mean fibrinogen levels were found to be 2.9 mg/dl higher in female Caucasians than in male Caucasians. Based on a 95% confidence interval, it would not be unusual if the true mean difference in fibrinogen were found to be anywhere between 2.35 mg/dL lower or 8.15 mg/dL higher in female Caucasians than male Caucasians. At the 5% significance level, we fail to reject the null hypothesis of no difference in mean fibrinogen across groups defined by binary gender within Caucasians (P=0.2787).**
   2. Are mean fibrinogen levels associated with sex in Noncaucasians?
      * **Mean fibrinogen levels were found to be 15.7 mg/dL higher in female Non-Caucasians than in male Non-Caucasians. Based on a 95% confidence interval, it would not be unusual if the true mean difference in fibrinogen were found to be anywhere between 1.40 mg/dL and 30.0 mg/dL higher in female Non-Caucasians than male Non-Caucasians. At the 5% significance level, we reject the null hypothesis of no difference in mean fibrinogen across groups defined by binary gender within Non-Caucasians (P=0.0314).**
   3. Are mean fibrinogen levels associated with sex after adjustment for race?
      * **As it is not specified, importance weighting will be utilized for the adjustment. After adjustment for race, mean fibrinogen levels were found to be 4.94 mg/dL higher in female subjects than male subjects with shared race. Such a difference was not sufficiently extreme to rule out a null hypothesis of no difference in mean fibrinogen across groups defined by binary gender (P=0.05108). Based on a 95% confidence interval, we find that the observed difference in mean fibrinogen is not atypical of settings in which the true difference in mean fibrinogen were such that female patients had mean fibrinogen 0.0231 mg/dL lower to 9.906 mg/dL higher than male patients of the same race.**
   4. Does race modify the association between mean fibrinogen level and sex?
      * **The difference in mean fibrinogen across groups defined by gender was found to be 12.8 mg/dl higher in Non-Caucasians than in Caucasians. Based on a 95% confidence interval, it would not be unusual if the true mean difference in effect were found anywhere between 2.433 mg/dL lower and 28.033 mg/dL greater in Non-Caucasians than Caucasians. At the 5% significance level, we fail to reject the null hypothesis of no effect modification by race in the association between fibrinogen and gender (P=0.0996).**

**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.
      1. **S= 39.288 mg/dL**
   2. 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.
      1. **Since we are using the cross-sectional standard deviation from part (a), V = 2\*39.288^2\*(1-.4)=1852.3. Hence, standard deviation of the change in cholesterol measurements made after two years within the population is estimated as 43.04 mg/dL.**
   3. 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.)
      1. **Using the RMSE, best estimate of the standard deviation of cholesterol within groups that had constant age and sex is estimated as 37.49.**
2. **(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?
   * 1. **To find To find . To find N: .**
2. What sample size will provide 90% power to detect the design alternative?
   * 1. **To find To find . To find N: .**
3. How would the sample size for 90% power change if you had not decided to adjust for age and sex?
   * 1. **To find To find . To find N: . Hence, holding power and excluding adjustments decreases sample size.**
4. 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.)
   * 1. **Since the clinical trial is randomized, both treatment arms should tend to have the same mean cholesterol at baseline. Hence, the difference in final cholesterol should estimate the exact same quantity as the difference in change of cholesterol. The only aspect missing would be the correlation, which tends to decrease V in this design. Hence, V would increase without the correlation, and sample size would increase as well.**
5. 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.)
   * 1. **Since we are adjusting for the baseline cholesterol level, as well as age and sex, we will in fact have the same sample size as if we were using the analysis approach in this problem (excluding (d)).**
6. **(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?
   * 1. **From the data, the proportion is estimated as 0.4014.**
2. 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.)
   * 1. **From the data, the proportion is estimated as 0.499.**
3. What sample size will provide 90% power to detect the design alternative?
   * 1. **To find To find . To find N: .**
4. What advantages or disadvantages does this study design have over the study design used in problem 4b?
   * 1. **If an increase or decrease of 10mg/dl of cholesterol truly had clinical impact, then this study design addresses the question completely. Otherwise, we will have failed to understand the effect of interest across the population. The loss of information is reflected in the required greater sample size for 90% power after dichotomizing: from 779 to 1082 in this design.**

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

We begin discussion of the university salary dataset.