Biost 518: Applied Biostatistics II
Biost 515: Biostatistics II
Emerson, Winter 2014

Homework #7 Key
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()`.)

a. 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.)

Ans: (10 pts,total) ***Statistical Methods:*** **An analysis of the difference between the sexes in mean serum cholesterol was computed within each stratum by subtracting the sample mean for males from the sample mean for females. A two-sided p value and 95% confidence interval were computed assuming an approximate normal distribution for the estimated difference in means, with the estimated standard errors for the difference based on the square root of the sum of the squared standard errors for each sex within the strata.**

Results: **Among the 2,534 Caucasians, females average a serum cholesterol of 222.8 mg/dL, while male Caucasians average 197.5 mg/dL. The observed difference of 25.3 mg/dL is larger than that that might be expected by random chance when the sexes truly have equal average cholesterol levels (two-sided P < 0.0001). Such an observed difference would not be unexpected if the true difference were such that the females' average cholesterol were anywhere between 22.3 mg/dL and 28.3 mg/dL higher than the males'average.**

Calculations: The estimated difference and standard error for the comparison in Caucasians are

$$\hat{\Delta}_C = 222.8 - 197.5 = 25.3 \qquad \hat{SE}\left(\hat{\Delta}_C\right) = \sqrt{1.092^2 + 1.103^2} = 1.552$$

Hence, we can compute a Z statistic as

$$Z = \frac{222.8 - 197.5}{\sqrt{1.092^2 + 1.103^2}} = \frac{25.3}{1.552} = 16.3$$ *with two-sided P value computed using Stata code*

```
di 2 * normal( - abs( (222.8 - 197.5) / sqrt(1.092^2 + 1.103^2) ) )
```

and 95% CI computed as

$$\hat{\Delta}_C \pm z_{.025} \times \hat{SE}\left(\hat{\Delta}_C\right) = 25.3 \pm 1.96 \times 1.552 = (22.3,\ 28.3)$$ **(***or computed using Stata code)*

```
di (222.8 – 197.5) - invnorm(.975) * sqrt(1.092^2 + 1.103^2)
di (222.8 – 197.5) + invnorm(.975) * sqrt(1.092^2 + 1.103^2)
```

b. Are mean cholesterol levels associated with sex in Noncaucasians?

Ans: (10 pts,total) ***Statistical Methods:*** **(same as part a)**

Results: **Among the 481 Noncaucasians, females average a serum cholesterol of 213.6 mg/dL, while male Noncaucasians average 197.9 mg/dL. The observed difference of 15.7 mg/dL is larger than that that might be expected by random chance when the sexes truly have equal average cholesterol levels (two-sided P < 0.0001). Such an observed difference would not be unexpected if the true difference were such that the females' average cholesterol were anywhere between 8.93 mg/dL and 22.5 mg/dL higher than the males'average.**

Calculations: The estimated difference and standard error for the comparison in Caucasians are

$$\hat{\Delta}_C = 213.6 - 197.9 = 15.7 \qquad \hat{SE}\left(\hat{\Delta}_C\right) = \sqrt{2.321^2 + 2.557^2} = 3.453$$

Hence, we can compute a Z statistic as

$$Z = \frac{15.7}{3.453} = 4.546$$ *with two-sided P value computed using Stata code*

```
di 2 * normal( - abs( (213.6 - 197.9) / sqrt(2.321^2 + 2.557^2) ) )
```

and 95% CI computed as

$$\hat{\Delta}_C \pm z_{.025} \times S\hat{E}\left(\hat{\Delta}_C\right) = 15.7 \pm 1.96 \times 3.453 = (8.93, 22.5)$$ *(or computed using Stata code)*

```
di (213.6 - 197.9) - invnorm(.975) * sqrt(2.321^2 + 2.557^2)
di (213.6 - 197.9) + invnorm(.975) * sqrt(2.321^2 + 2.557^2)
```

c. 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 Δ_C and Δ_N you derived in parts a and b, respectively: Let the adjusted estimated be defined according to

$$\Delta_{adj} = (w_C \times \Delta_C + w_N \times \Delta_N) / (w_C + w_N)$$

where w_C and w_N 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) = c^2 Var(X). Hence, you can find the standard error of the adjusted estimate can be found by

$$se\left(\Delta_{adj}\right) = \sqrt{\frac{w_C^{\,2} \times se^2\left(\Delta_C\right) + w_N^{\,2} \times se^2\left(\Delta_N\right)}{\left(w_C + w_N\right)^2}}$$

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

- *<u>Importance weights:</u> 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).*
- *<u>Efficiency weights:</u> 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.*

<u>Ans:</u> (10 pts,total) *<u>Statistical Methods:</u>* An analysis of the difference between the sexes in mean serum cholesterol adjusted by race was computed by using a weighted average of stratum specific estimates. In each race, the sample mean for males was subtracted from the sample mean for females. The standard errors for the stratum specific estimated differences were based on the square root of the sum of the squared standard errors for each sex within the strata. Race adjusted estimates across the strata were then computed as a weighted average using the inverse of the squared standard errors from each stratum. As a sensitivity analysis, a weighted average was also computed using weights derived from 2010 US census data, in which 86.37% of US residents aged 65 years or older are Caucasian. Two-sided p values and 95% confidence intervals were computed assuming an approximate normal distribution for the estimated difference in means.

***<u>Results:</u>* The sample consists of 1,757 females and 1,258 males, and of 2,534 Caucasians and 481 Noncaucasians. The distribution of sex within each racial group was not provided. Female Caucasians average a serum cholesterol of 222.8 mg/dL, while male Caucasians average 197.5 mg/dL, yielding an**

observed difference of 25.3 mg/dL (SE 1.552 mg/dL). In Noncaucasians, females average a serum cholesterol of 213.6 mg/dL, while males average 197.9 mg/dL, yielding an observed difference of 15.7 mg/dL (SE 3.453 mg/dL). Using efficient weights based on the inverse squared standard errors, a race adjusted difference in mean serum cholesterol across the sexes estimates that females average serum cholesterol 23.69 mg/dL higher than males of the same racial group, and that observed difference is greater than that might be expected by random chance when the sexes truly have equal average cholesterol levels (two-sided P < 0.0001). Such an observed difference would not be unexpected if the true difference were such that the females' average cholesterol were anywhere between 20.9 mg/dL and 26.5 mg/dL higher than the average for males in the same racial group. As a sensitivity analysis, race adjusted estimates were also computed using weights based on US census data. The estimated difference of 23.99 mg/dL (95% CI 21.2 to 26.8 mg/dL, two-sided P < 0.0001) is little different than that estimated using efficiency weights. *(I also provide the inference below that would have been obtained using the sample sizes instead of the US census data for the importance weights.)*

Calculations: The estimated difference and standard error for the race adjusted comparison ***using efficiency weights*** *are*

$$\hat{\Delta} = \frac{\left(25.3/1.552^2\right)+\left(15.7/3.453^2\right)}{\left(1/1.552^2\right)+\left(1/3.453^2\right)} = 23.69 \qquad \hat{SE}\left(\hat{\Delta}\right) = \sqrt{\frac{\left(1.552^2/1.552^4\right)+\left(3.453^2/3.453^4\right)}{\left\{\left(1/1.552^2\right)+\left(1/3.453^2\right)\right\}^2}} = 1.416$$

Hence, we can compute a Z statistic as

$$Z = \frac{23.69}{1.416} = 16.73$$ *with two-sided P value computed using Stata code*

```
di 2 * normal( - abs( (16.73 ) )
```

and 95% CI computed as

$$\hat{\Delta} \pm z_{.025} \times \hat{SE}\left(\hat{\Delta}\right) = 23.69 \pm 1.96 \times 1.416 = (20.9,\ 26.5)$$ **(***or computed using Stata code)*

```
di (23.69) - invnorm(.975) * 1.416
di (23.69) + invnorm(.975) * 1.416
```

The estimated difference and standard error for the race adjusted comparison ***using importance weights based on US census data are***

$$\hat{\Delta} = \frac{\left(25.3 \times 0.8637\right)+\left(15.7 \times 0.1363\right)}{0.8637 + 0.1363} = 23.99 \qquad \hat{SE}\left(\hat{\Delta}\right) = \sqrt{\frac{\left(1.552^2 \times 0.8637^2\right)+\left(3.453^2 \times 0.1363^2\right)}{\left\{\ 0.8637 + 0.1363\ \right\}^2}} = 1.421$$

Hence, we can compute a Z statistic as

$$Z = \frac{23.99}{1.421} = 16.88$$ *with two-sided P value computed using Stata code*

```
di 2 * normal( - abs( (16.88 ) )
```

and 95% CI computed as

$$\hat{\Delta} \pm z_{.025} \times \hat{SE}\left(\hat{\Delta}\right) = 23.99 \pm 1.96 \times 1.421 = (21.2,\ 26.8)$$

The estimated difference and standard error for the race adjusted comparison ***using importance weights based on sample sizes in the data set are***

$$\hat{\Delta} = \frac{\left(25.3 \times 2534\right)+\left(15.7 \times 481\right)}{2534 + 481} = 23.77 \qquad \hat{SE}\left(\hat{\Delta}\right) = \sqrt{\frac{\left(1.552^2 \times 2534^2\right)+\left(3.453^2 \times 481^2\right)}{\left\{\ 2534 + 481\ \right\}^2}} = 1.416$$

(Note that the SE in this case is little different from the SE based on inverse of the stratum specific SEs. This might be expected if the female : male ratios were similar in each racial group and the standard deviations were fairly homoscedastic.) Hence, we can compute a Z statistic as

$$Z = \frac{23.77}{1.416} = 16.79 \text{ with two-sided P value computed using Stata code}$$

```
di 2 * normal( - abs( (16.79 ) )
```

and 95% CI computed as

$$\hat{\Delta} \pm z_{.025} \times \hat{SE}\left(\hat{\Delta}\right) = 23.77 \pm 1.96 \times 1.416 = (21.0, 26.5)$$

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

Ans: (10 pts,total) ***Statistical Methods:*** **An analysis of the difference between the racial groups with respect to the association between sex and mean serum cholesterol was computed by using a difference between the stratum specific estimates of the differences in mean cholesterol between the sexes. In each race, the sample mean for males was subtracted from the sample mean for females. The standard errors for the stratum specific estimated differences were based on the square root of the sum of the squared standard errors for each sex within the strata. Differences between the racial groups with respect to those associations were then computed as the difference of the association in Caucasians minus that for the Noncaucasians. Standard errors for the estimates of effect modification were computed as the square root of the summed squared standard errors from each stratum. Two-sided p values and 95% confidence intervals were computed assuming an approximate normal distribution for the estimated means.**

Results: **The sample consists of 1,757 females and 1,258 males, and of 2,534 Caucasians and 481 Noncaucasians. The distribution of sex within each racial group was not provided. Female Caucasians average a serum cholesterol of 222.8 mg/dL, while male Caucasians average 197.5 mg/dL, yielding an observed difference of 25.3 mg/dL (SE 1.552 mg/dL). In Noncaucasians, females average a serum cholesterol of 213.6 mg/dL, while males average 197.9 mg/dL, yielding an observed difference of 15.7 mg/dL (SE 3.453 mg/dL). The difference between the sex association with mean cholesterol across racial groups estimates that the association is 9.6 mg/dL higher in Caucausians than in Noncaucasians. That observed difference is greater than what might be expected by random chance when the sex-cholesterol association is truly equal in the two treatment groups (two-sided P = 0.0112). Such an observed difference would not be unexpected if the true difference between the sex-cholesterol associations across racial groups were such that the Caucasians' sex-cholesterol association (as measured by the difference of females' mean cholesterol minus males' mean cholesterol) were anywhere between 2.18 mg/dL and 17.0 mg/dL higher than the Noncaucasians' sex-cholesterol association.**

Calculations: The estimated difference and standard error for the interaction are

$$\hat{\Delta} = 25.3 - 15.7 = 9.6 \qquad \hat{SE}\left(\hat{\Delta}\right) = \sqrt{\left(1.552^2\right) + \left(3.453^2\right)} = 3.786$$

Hence, we can compute a Z statistic as

$$Z = \frac{9.6}{3.786} = 2.536 \text{ with two-sided P value computed using Stata code}$$

```
di 2 * normal( - abs( 2.536 ) )
```

and 95% CI computed as

$$\hat{\Delta} \pm z_{.025} \times \hat{SE}\left(\hat{\Delta}\right) = 9.6 \pm 1.96 \times 3.786 = (2.18, 17.0)$$

2. 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.

a. Are mean fibrinogen levels associated with sex in Caucasians

Ans: (10 pts,total) *Statistical Methods:* An analysis of the difference between the sexes in mean serum fibrinogen was computed within each stratum by subtracting the sample mean for males from the sample mean for females. A two-sided p value and 95% confidence interval were computed assuming an approximate normal distribution for the estimated difference in means, with the estimated standard errors for the difference based on the square root of the sum of the squared standard errors for each sex within the strata.

***Results:* Among the 2,534 Caucasians, females average a serum fibriinogen of 320.7 mg/dL, while male Caucasians average 317.8 mg/dL. The observed difference of 2.9 mg/dL is not larger than that that might be expected by random chance when the sexes truly have equal average cholesterol levels (two-sided P = 0.279). Such an observed difference would not be unexpected if the true difference were such that the females' average cholesterol were anywhere between 2.35 mg/dL lower and 8.15 mg/dL higher than the males'average.**

Calculations: The calculations proceed as in problem 1. They were performed using an Excel spreadsheet.

b. Are mean fibrinogen levels associated with sex in Noncaucasians?

Ans: (10 pts,total) *Statistical Methods:* (same as part a)

***Results:* Among the 481 Noncaucasians, females average a serum fibrinogen of 349.4 mg/dL, while male Noncaucasians average 333.7 mg/dL. The observed difference of 15.7 mg/dL is larger than that that might be expected by random chance when the sexes truly have equal average cholesterol levels (two-sided P = 0.0314). Such an observed difference would not be unexpected if the true difference were such that the females' average fibrinogen were anywhere between 1.40 mg/dL and 30.0 mg/dL higher than the males'average.**

Calculations: The calculations proceed as in problem 1. They were performed using an Excel spreadsheet.

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

Ans: (10 pts,total) *Statistical Methods:* An analysis of the difference between the sexes in mean serum fibrinogen adjusted by race was computed by using a weighted average of stratum specific estimates. In each race, the sample mean for males was subtracted from the sample mean for females. The standard errors for the stratum specific estimated differences were based on the square root of the sum of the squared standard errors for each sex within the strata. Race adjusted estimates across the strata were then computed as a weighted average using the inverse of the squared standard errors from each stratum. As a sensitivity analysis, a weighted average was also computed using weights derived from 2010 US census data, in which 86.37% of US residents aged 65 years or older are Caucasian. Two-sided p values and 95% confidence intervals were computed assuming an approximate normal distribution for the estimated difference in means.

***Results:* The sample consists of 1,757 females and 1,258 males, and of 2,534 Caucasians and 481 Noncaucasians. The distribution of sex within each racial group was not provided. Female Caucasians average a serum fibrinogen of 320.7 mg/dL, while male Caucasians average 317.8 mg/dL, yielding an observed difference of 2.9 mg/dL (SE 2.677 mg/dL). In Noncaucasians, females average a serum fibrinogen of 349.4 mg/dL, while males average 333.7 mg/dL, yielding an observed difference of 15.7 mg/dL (SE 7.296 mg/dL). Using efficient weights based on the inverse squared standard errors, a race adjusted difference in mean serum fibrinogen across the sexes estimates that females average serum fibrinogen 4.42 mg/dL higher than males of the same racial group, and that observed difference is not**

greater than that might be expected by random chance when the sexes truly have equal average fibrinogen levels (two-sided P = 0.0787). Such an observed difference would not be unexpected if the true difference were such that the females' average fibrinogen were anywhere between 0.51 mg/dL lower and 9.34 mg/dL higher than the average for males in the same racial group. As a sensitivity analysis, race adjusted estimates were also computed using weights based on US census data. The estimated difference of 4.64 mg/dL (95% CI -0.29 to 9.58 mg/dL, two-sided P = 0.0650) is little different than that estimated using efficiency weights. *(Using the sample sizes instead of the US census data for the importance weights would have yielded a difference of 4.94 mg/dL (95% CI -0.02 to 9.91 mg/dL, two-sided P = 0.0511).)*

Calculations: The calculations proceed as in problem 1. They were performed using an Excel spreadsheet.

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

Ans: (10 pts,total) ***Statistical Methods:*** **An analysis of the difference between the racial groups with respect to the association between sex and mean serum fibrinogen was computed by using a difference between the stratum specific estimates of the differences in mean fibrinogen between the sexes. In each race, the sample mean for males was subtracted from the sample mean for females. The standard errors for the stratum specific estimated differences were based on the square root of the sum of the squared standard errors for each sex within the strata. Differences between the racial groups with respect to those associations were then computed as the difference of the association in Caucasians minus that for the Noncaucasians. Standard errors for the estimates of effect modification were computed as the square root of the summed squared standard errors from each stratum. Two-sided p values and 95% confidence intervals were computed assuming an approximate normal distribution for the estimated means.**

Results: **The sample consists of 1,757 females and 1,258 males, and of 2,534 Caucasians and 481 Noncaucasians. The distribution of sex within each racial group was not provided. Female Caucasians average a serum fibrinogen of 320.7 mg/dL, while male Caucasians average 317.8 mg/dL, yielding an observed difference of 2.9 mg/dL (SE 2.677 mg/dL). In Noncaucasians, females average a serum fibrinogen of 349.4 mg/dL, while males average 333.7 mg/dL, yielding an observed difference of 15.7 mg/dL (SE 7.296 mg/dL). The difference between the sex association with mean fibrinogen across racial groups estimates that the association is 12.8 mg/dL lower in Caucausians than in Noncaucasians. That observed difference is not greater than what might be expected by random chance when the sex-fibrinogen association is truly equal in the two treatment groups (two-sided P = 0.0996). Such an observed difference would not be unexpected if the true difference between the sex-fibrinogen associations across racial groups were such that the Caucasians' sex-fibrinogen association (as measured by the difference of females' mean fibrinogen minus males' mean fibrinogen) were anywhere between 28.0 mg/dL lower and 2.43 mg/dL higher than the Noncaucasians' sex-fibrinogen association.**

Calculations: The calculations proceed as in problem 1. They were performed using an Excel spreadsheet.

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 $\theta = \theta_0$.
  - 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 $\theta = \theta_0$.
  - The statistical "type 1 error" is the probability of declaring statistical significance for the value of $\theta = \theta_0$.
  - 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 $\theta = \theta_1$.
  - 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 $z_p = 0.8410$, 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 z_p such that the probability that a standard normal Z is less than z_p is p.
    - In Stata, the p-th quantile can be found by using `invnorm( )` function. For instance, if we wanted $z_{0.80}$, the 80th percentile can be found from the Stata command `disp invnorm(0.80).` (Stata would then display `.8410.`)
    - In Excel, the value of z_p can be found by using the function `norminv( ).` For instance, if $\alpha = 0.025$, in our sample size formulas given below, we might want the 100(1 - .025)% percentile. The value of $z_{0.975}$ 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 `qnorm( ).` For instance, if we want $z_{0.975}$, 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

$$N = \frac{\delta_{\alpha\beta}^2 V}{\Delta^2}$$

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),
- $\delta_{\alpha\beta}$ 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 $1-\beta$), 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 $\delta_{\alpha\beta}$ 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

$$\delta_{\alpha\beta} = z_{1-\alpha} + z_{\beta}$$

where z_p is the pth quantile of the standard normal distribution. For a two-sided level α test, the standardized alternative is given by

$$\delta_{\alpha\beta} = z_{1-\alpha/2} + z_{\beta}$$

The value of z_p 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

- $\Delta = \theta_1 - \theta_0$ (used for inference in "additive models" for means and proportions, and sometimes medians), or
- $\Delta = log(\theta_1 / \theta_0)$ (used for inference in "multiplicative models" for geometric means, odds, and hazards, and sometimes means and medians),

3. **(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.

    a. 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.

Ans: (5 pts) The standard deviation of cholesterol from the inflammatory biomarkers dataset is 39.29 mg/dl.

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

Ans: (5 pts) Assuming a correlation of $\rho = 0.40$ and presuming the standard deviation would be 39.29 mg/dL at baseline and at the final measurement at two years, we estimate the standard deviation of a difference of cholesterol measurements made two years apart as 43.04 mg/dl:

$$D_i = Chol_{i,final} - Chol_{i,bs\ln}$$

$$SD(D_i) = \sqrt{Var(Chol_{i,final}) + Var(Chol_{i,bs\ln}) - 2\,\rho\, SD(Chol_{i,final})\, SD(Chol_{i,bs\ln})}$$

$$= \sqrt{39.29^2 + 39.29^2 - 2 \times 0.40 \times 39.29 \times 39.29} = 39.29\sqrt{2 \times (1 - 0.4)} = 43.04$$

    c. 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.)

Ans: (5 pts) The root mean squared error (RMSE) from a regression of cholesterol on age and sex in the inflammatory biomarkers dataset is 37.49 mg/dl.

4. **(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 $\alpha = 0.025$,
    - the desired statistical power is $\beta = 0.80$ or 0.90,
    - the measure of treatment effect is $\theta = (\mu_{T,2} - \mu_{T,0}) - (\mu_{C,2} - \mu_{C,0})$ (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 $V = 8\sigma^2(1-\rho)$. (Again, use a correlation of 0.4.)
    - the comparison between alternative and null hypotheses is $\Delta = \theta_1 - \theta_0$.

    a. What sample size will provide 80% power to detect the design alternative?

Ans: (5 pts)

To find $\delta_{\alpha\beta}$: $z_{1-\alpha} = z_{0.975} = 1.960$; $z_\beta = z_{0.80} = 0.842$; $\delta_{\alpha\beta} = 1.960 + 0.8416 = 2.802$.

To find Δ: $\Delta = 10$.

To find V: $V = 8\sigma^2(1-\rho) = 8 \times 37.49^2 \times (1 - 0.40) = 6746$

To find N: $N = \delta_{\alpha\beta}^2 V / \Delta^2 = 2.802^2 \times 6746 / 10^2 = 530$.

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

Ans: (5 pts)

To find $\delta_{\alpha\beta}$: $z_{1-\alpha} = z_{0.975} = 1.960$; $z_\beta = z_{0.90} = 1.2816$; $\delta_{\alpha\beta} = 1.960 + 1.2816 = 3.2416$.

To find Δ: $\Delta = 10$.

To find V: $V = 8\sigma^2(1-\rho) = 8 \times 37.49^2 \times (1 - 0.40) = 6746$

To find N: $N = \delta_{\alpha\beta}^2 V / \Delta^2 = 3.2416^2 \times 6746 / 10^2 = 709$ (or 710 if we want an even number).

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

Ans: (5 pts) If we did not adjust for age and sex, we would have used 39.29 for the SD instead of 37.49 in the above formula: an increase by a factor of 39.29 / 37.49 = 1.0480. Sample size behaves like the square of that SD, so we would end up with a sample size estimate that was 1.0480^2 = 1.098 times as high, or 709 * 1.098 = 779 (or 780 for an even number).

d. 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.)

Ans: (5 pts) From problem 3a, the SD of a single measurement was 39.29 mg/dl, while the SD of the change was 43.04. Adjusting for sex and age would be expected to yield a similar ratio. Hence,we would expect lower sample size estimates if we only used the final measurements.

*We can again consider the ratio: 39.29 / 43.04 = 0.9129. Sample size behaves like the square of the SD, so we would end up with a sample size estimate that was 0.9129^2 = 0.8333 times as high, or 709 * 0.8333 = 591 (or 592 for an even number).*

So long as the correlation between the baseline and final measurements is less than 0.5, more precision is achieved when using only the final measurement compared to using the change over time. When the correlation is above 0.5, more precision is achieve when using the change over time compared to using only the final measurement.

e. 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.)

Ans: (5 pts) An ANCOVA model that adjusts for baseline as a predictor (in addition to treatment, and in this problem, age and sex) will always tend to precision at least as great as achieved with either an analysis of only the final measurements or an analysis of the change over time. Hence, the sample size estimates would be smaller.

In a two sample problem comparing only the final measurement, $V= 4\sigma^2$.

In a two sample problem comparing the change over time, $V= 8\sigma^2(1-\rho)$.

In a two sample problem that analyzes either the final measurement or the change over time using regression on the baseline value (along with an indicator of treatment), $V= 4\sigma^2(1-\rho^2)$. (Note that the regression models using the final measurement as response or using the change in measurements as response are just reparameterizations of each other.)

Note that because the correlation must be between -1 and 1, the ANCOVA model always has at least as low a value of V as the other analytic approaches.

5. **(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 intend to perform a hypothesis test in which
   - the one-sided level of significance is $\alpha = 0.025$,
   - the desired statistical power is $\beta = 0.90$,

- we presume that the proportion p_C 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 p_T 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 $\theta_1 = p_{T,} - p_C$ (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 $V=2(p_{T,}(1- p_{T,}) + p_C (1 - p_C))$(most often, we would compute this under the alternative hypothesis in this setting),
- the comparison between alternative and null hypotheses is $\Delta = \theta_1 - \theta_0 = \theta_1$.

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

Ans: (5 pts) Cholesterol measurements were available on 4,953 subjects, and 1,960 had cholesterol measurements below 200 mg/dl. We thus estimate p_c = 1960 / 4953 = 0.3957.

b. Using the inflammatory biomarkers dataset, what is your estimate of the proportion p_T 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.)

Ans: (5 pts) Cholesterol measurements were available on 4,953 subjects, and 2,448 had cholesterol measurements below 200 mg/dl. We thus estimate p_T = 2448 / 4953 = 0.4942.

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

Ans: (5 pts)

To find $\delta_{\alpha\beta}$: $z_{1-\alpha} = z_{0.975} = 1.960$; $z_\beta = z_{0.90} = 1.2816$; $\delta_{\alpha\beta} = 1.960 + 1.2816 = 3.2416$.

To find Δ: $\Delta = p_T - p_C = 0.4942 - 0.3957 = 0.0985$.

To find V: $V = 2(p_{T,}(1- p_{T,}) + p_C (1 - p_C) = 0.9782$

To find N: $N = \delta_{\alpha\beta}^2 V / \Delta^2 = 3.2416^2 \times 0.9782 / 0.0985^2 = 1059$ (or 1060 if we want an even number).

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

Ans: (5 pts) Clearly we lose efficiency (thus requiring a larger sample size) if we dichotomize our data. Hence the dichotomization is undesirable from a standpoint of statistical efficiency. However, if there

were a clinical reason (such as the importance of having a serum cholesterol below the threshold of 200 mg/dl), we might use the dichotomized analysis in order to ensure clinical importance.