**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #2**

January 13, 2015

**Written problems:** To be submitted as a MS-Word compatible file to the class Catalyst dropbox by noon on Tuesday, January 20, 2015. See the instructions for peer grading of the homework that are posted on the web pages.

*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.)*

***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.***

All questions relate to associations between the two biomarkers C-reactive protein (CRP) and fibrinogen (FIB), and how any such association might depend upon prevalence of prior cardiovascular disease (CVD). This homework again uses the subset of information that was collected to examine inflammatory biomarkers and mortality. The data can be found on the class web page (follow the link to Datasets) in the file labeled inflamm.txt. Documentation is in the file inflamm.pdf. See homework #1 for information about reading the data into R and/or Stata.

1. Provide a suitable descriptive statistical analysis for the association between CRP and FIB both overall, and separately for groups having no prior history of diagnosed cardiovascular disease or having prior diagnosed CVD.

**Methods:** Descriptive analysis was performed for blood C reactive protein (CRP) level data and blood fibrinogen level data (FIB), with each group divided into groups having prior cardiovascular disease (CVD), and no prior cardiovascular disease. CRP was divided into three groups based on level (<1 mg/l, 1-3 mg/l, >3 mg/l).. Mean blood fibrinogen levels in mg/dl, standard deviation, and ranges werecalculated for the group with no prior CVD, prior CVD, and overall for each CRP group. Of the 5001 subjects on whom data was collected, 68 subjects were missing data on CRP levels and 86 subjects were missing data on fibrinogen levels. These subjects were omitted from analysis.

**Results:** Mean blood fibrinogen level overall among patients with no prior history of cardiovascular disease was 319.62 mg/dL. Among patients with CRP level <1 mg/l, CRP level between 1 and 3, and CRP level >3, mean blood fibrinogen level was 277.48 mg/dL, 310.02 mg/dL, and 367.20 mg/dL, respectively. Mean blood fibrinogen level overall among patients with prior history of cardiovascular disease was 334.46 mg/dL. Among patients with CRP level <1 mg/l, CRP level between 1 and 3, and CRP level >3, mean blood fibrinogen level was 290.23 mg/dL, 314.85 mg/dL, and 386.29 mg/dL, respectively. Overall and in each CRP level category, mean blood fibrinogen level was higher among patients who had prior history of cardiovascular disease. For both patients with no history and with prior history of cardiovascular disease, there was an increase in mean blood fibrinogen level as CRP level increased.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Blood fibrinogen level (mean, sd, range) | CRP Level <1 mg/l | CRP Level Between 1 and 3 mg/l | CRP Level >3 mg/l | Overall |
| No CVD | 277.48 mg/dL (48.52, 172-436), n = 348 | 310.02 mg/dL (52.46, 109-562), n = 2597 | 367.20 mg/dL (78.88, 132-872), n = 832 | 319.62 mg/dL (64.83, 109-872), n = 3777 |
| Prior CVD | 290.23 mg/dL (57.93, 180-540), n = 78 | 314.85 mg/dL (55.60, 138-592), n = 709 | 386.29 mg/dL (84.50, 175-695), n = 695 | 334.46 mg/dL (74.11, 138-695), n = 1122 |

1. Perform t test analyses exploring an association between mean fibrinogen and prior history of CVD.
   1. Perform an analysis presuming that the standard deviation of fibrinogen is similar within each group defined by presence of absence of prior history of CVD.

**Methods:** Mean blood fibrinogen levels were compared between groups whose subjects had either presence or absence of prior history of cardiovascular disease. Differences in mean blood fibrinogen levels were obtained using a t test with the assumption of equal variances. 95% confidence intervals for the difference in mean blood fibrinogen level between groups were calculated, also assuming equal variance between groups. There were 86 subjects missing data on fibrinogen levels, and those subjects were excluded from the analysis.

**Results:** Mean blood fibrinogen was 319.57 mg/dl among the group with no prior history of cardiovascular disease (n =3791), and was 334.46 among the group with prior history of cardiovascular disease (n = 1124). The difference in mean blood fibrinogen was 14.89 mg/dl, with those subjects having prior history of cardiovascular disease having higher mean blood fibrinogen levels. Based on a 95% confidence interval, the data observed would not be unusual if the true value of mean blood fibrinogen of subjects with prior history of cardiovascular disease was between 19.34 and 10.42 mg/dl higher than the mean blood fibrinogen level of subjects with no prior history of cardiovascular disease. Using an alpha level of 0.05, we can reject the null hypothesis that there is no difference in mean blood fibrinogen levels between the two groups (two sided p < 0.001). Thus, we can conclude that there is an association between mean blood fibrinogen level and history of prior cardiovascular disease.

* 1. How could the same analysis as presented in part a have been performed with linear regression? Explicitly provide the correspondences between the various statistical output from each of the analyses.

We could accomplish this same analysis by performing a linear regression with blood fibrinogen levels as the dependent variable (outcome) and prior history of cardiovascular disease as the independent variable (predictor of interest). This can be done using the STATA command “regress fib prevdis”. The statistical output gives the slope of 14.89, which corresponds to the difference in mean blood fibrinogen levels between groups (with groups being defined by either prevdis = 0 or prevdis = 1). The 95% confidence interval for the difference/slope is provided as well as (10.42, 19.34). The mean blood fibrinogen level for the group with no history of cardiovascular disease is given as well, as the intercept. Adding the slope and intercept gives the value of mean blood fibrinogen level for the group that does have a history of previous cardiovascular disease. The p-value on the top line gives the p-value for the t test comparing means between groups.

* 1. Perform an analysis allowing for the possibility that the standard deviation of fibrinogen might differ across groups defined by presence of absence of prior history of CVD.

**Methods:** Mean blood fibrinogen levels were compared between groups whose subjects had either presence or absence of prior history of cardiovascular disease. Differences in mean blood fibrinogen levels were obtained using a t test, this time with the assumption of unequal variances. 95% confidence intervals for the difference in mean blood fibrinogen level between groups were calculated, also assuming unequal variance between groups. There were 86 subjects missing data on fibrinogen levels, and those subjects were excluded from the analysis.

**Results:** Mean blood fibrinogen was 319.57 mg/dl among the group with no prior history of cardiovascular disease (n =3791), and was 334.46 among the group with prior history of cardiovascular disease (n = 1124). The difference in mean blood fibrinogen was 14.89 mg/dl, with those subjects having prior history of cardiovascular disease having higher mean blood fibrinogen levels. Based on a 95% confidence interval, the data observed would not be unusual if the true value of mean blood fibrinogen of subjects with prior history of cardiovascular disease was between 19.68 and 10.09 mg/dl higher than the mean blood fibrinogen level of subjects with no prior history of cardiovascular disease. Using an alpha level of 0.05, we can reject the null hypothesis that there is no difference in mean blood fibrinogen levels between the two groups (two sided p < 0.001). Thus, we can conclude that there is an association between mean blood fibrinogen level and history of prior cardiovascular disease.

* 1. How could a similar analysis as presented in part c have been performed with linear regression? Explicitly provide the correspondences between the various statistical output from each of the analyses.

Similarly to part b, we could accomplish this same analysis by performing a linear regression with blood fibrinogen levels as the dependent variable (outcome) and prior history of cardiovascular disease as the independent variable (predictor of interest). This can be done using the STATA command “regress fib prevdis, robust”. This command gives robust standard errors instead of standard errors, and takes into account the assumption of unequal variances. The statistical output still gives the slope of 14.89, which corresponds to the difference in mean blood fibrinogen levels between groups (with groups being defined by either prevdis = 0 or prevdis = 1). The 95% confidence interval for the difference/slope is provided as well as (10.08, 19.68). The mean blood fibrinogen level for the group with no history of cardiovascular disease is given as well, as the intercept. Adding the slope and intercept gives the value of mean blood fibrinogen level for the group that does have a history of previous cardiovascular disease. The p-value on the top line gives the p-value for the t test comparing means between groups.

* 1. How could you have used the results of the analysis performed in part a to predict whether the analysis in part c would have found a stronger or weaker association (as measured by the magnitude of the t statistic and p value)?

The only difference between the results of part a and part c was that with the assumption of unequal variances, the confidence interval for the differences between mean blood fibrinogen level. The assumption of equal variances is not conservative enough and thus, in part a, leads to a p-value that is too small and a confidence interval that is too narrow. This would falsely magnify the association between- such that when the assumption of equal variances was removed, as it was in part c, a weaker association would have been found.

For problems 3 – 6, we are interested in exploring alternative approaches to the use of simple linear regression to explore associations between CRP and FIB. In each of those problems, I ask you to report fitted values from the regression. **Please always use at least 4 significant figures when making calculations, and report the fitted values to three significant digits**.

1. Perform a statistical analysis evaluating an association between mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, untransformed random variable.
   1. Provide an interpretation of the estimated intercept from the fitted regression model as it pertains to fibrinogen levels.

The intercept is 304.015 mg/dl. This gives the value of blood fibrinogen when CRP level is 0 mg/L.

* 1. Provide an interpretation of the estimated slope from the fitted regression model as it pertains to fibrinogen levels.

The estimated slope from the fitted regression model is 5.251. This indicates that CRP and FIB are positively associated, and furthermore that a 5.251 unit increase in blood fibrinogen level for every unit increase in CRP level.

* 1. Provide full statistical inference about the presence of an association between fibrinogen and CRP using this regression analysis.

The linear regression gave a p-value of <0.001. Assuming an alpha level of 0.05, this indicates statistical significance and thus a rejection of the null hypothesis that there is no association between CRP and FIB. Using the slope 5.2509, we can calculate that for every 10% increase in blood CRP level, there will be an increase of 5.78 mg/dL in blood fibrinogen level.

* 1. In a table similar to table 1 below, provide estimates of the central tendency for fibrinogen levels within groups having CRP of 1, 2, 3, 4, 6, 8, 9, and 12 mg/L. (Make clear what summary measure is being estimated).

Using the linear equation FIB = 304.0152 + 5.2509\*CRP, and plugging in value of CRP = 1, 2, 3, 4, 5, 6, 8, 9, and 12, the results when solving for FIB are presented in Table 1 below.

1. Repeat problem 3, except perform a statistical analysis evaluating an association between mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, log transformed random variable. (For the purpose of this problem in this homework, replace all observations of CRP=0 with CRP=0.5.)

The intercept is 295.57 mg/dL. This gives the value of blood fibrinogen when CRP level is 0 mg/L. The estimated slope form the fitted regression model is 36.83. Using the linear equation FIB = 295.5663 + 36.8332log(CRP), and plugging in values of CRP = 1, 2, 3, 4, 5, 6, 8, 9, and 12, the results when solving for FIB are presented in Table 1 below. Using the slope of 36.83 and multiplying by the log of 1.1, we can see that for every 10% increase in blood CRP level, there will be an increase of 3.51 mg/dL in blood fibrinogen level.

1. Repeat problem 3, except perform a statistical analysis evaluating an association between the geometric mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, untransformed random variable.

The intercept is 5.70 mg/dL. This gives the value of log transformed blood fibrinogen level when CRP level is 0 mg/L. The slope is 0.01392. Using the linear equation FIB = 5.7068 + 0.01392CRP, and plugging in values of CRP = 1, 2, 3, 4, 5, 6, 8, 9, and 12, the results when solving for FIB are presented in Table 1 below. For every 10% increase in blood CRP level, there will be a 14.93% increase in blood fibrinogen level.

1. Repeat problem 3, except perform a statistical analysis evaluating an association between the geometric mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, log transformed random variable. (For the purpose of this problem in this homework, replace all observations of CRP=0 with CRP=0.5.)

The intercept is 5.68 mg/dL. This gives the value of log transformed blood fibrinogen level when log transformed CRP level is 0 mg/dL. Using the linear equation FIB = 5.6786 + 0.1054CRP, and plugging in values of CRP = 1, 2, 3, 4, 5, 6, 8, 9, and 12, the results when solving for FIB are presented in Table 1 below. This information tells us that a 10% increase in blood CRP level is associated with a 1% increase in blood fibrinogen level.

**Table 1**: Example of possible display of fitted values. You should indicate the summary measure of the fibrinogen distribution that is being estimated in each column.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **CRP level** | **Problem 3:  mean** | **Problem 4:  mean** | **Problem 5:  geometric mean** | **Problem 6: geometric mean** |
| **1 mg/L** | 309.266 mg/dl | 295.566 mg/dl | 305.125 mg/dl | 292.540 mg/dl |
| **2 mg/L** | 314.517 mg/dl | 321.091 mg/dl | 309.402 mg/dl | 314.712 mg/dl |
| **3 mg/L** | 319.768 mg/dl | 336.032 mg/dl | 313.739 mg/dl | 328.453 mg/dl |
| **4 mg/L** | 325.019 mg/dl | 346.628 mg/dl | 318.136 mg/dl | 338.565 mg/dl |
| **6 mg/L** | 335.521 mg/dl | 361.563 mg/dl | 327.118 mg/dl | 353.347 mg/dl |
| **8 mg/L** | 346.022 mg/dl | 372.159 mg/dl | 336.353 mg/dl | 364.225 mg/dl |
| **9 mg/L** | 351.273 mg/dl | 376.497 mg/dl | 341.067 mg/dl | 368.775 mg/dl |
| **12 mg/L** | 367.026 mg/dl | 387.093 mg/dl | 355.612 mg/dl | 380.128 mg/dl |

1. Complete the following table that makes comparisons (differences or ratios) of the fitted values for each of the models.

**Table 2**: Example of possible display of comparisons of fitted values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **Comparisons across CRP level** | **Problem 3:  mean** | **Problem 4:  mean** | **Problem 5: geometric mean** | **Problem 6: geometric mean** |
| ***Differences*** | | | | |
| **2 mg/L – 1 mg/L** | 5.251 mg/dL | 25.525 mg/dL | 4.277 mg/dL | 22.172 mg/dL |
| **3 mg/L – 2 mg/L** | 5.251 mg/dL | 14.941 mg/dL | 4.337 mg/dL | 13.740 mg/dL |
| **4 mg/L – 1 mg/L** | 15.753 mg/dL | 51.062 mg/dL | 13.011 mg/dL | 46.025 mg/dL |
| **4 mg/L – 2 mg/L** | 10.502 mg/dL | 25.537 mg/dL | 8.734 mg/dL | 23.853 mg/dL |
| **6 mg/L – 3 mg/L** | 15.753 mg/dL | 25.531 mg/dL | 13.379 mg/dL | 24.894 mg/dL |
| **8 mg/L – 4 mg/L** | 21.003 mg/dL | 25.521 mg/dL | 18.217 mg/dL | 35.772 mg/dL |
| **9 mg/L – 6 mg/L** | 15.752 mg/dL | 14.934 mg/dL | 13.949 mg/dL | 15.428 mg/dL |
| **9 mg/L – 8 mg/L** | 5.752 mg/dL | 4.338 mg/dL | 4.714 mg/dL | 4.550 mg/dL |
| **12 mg/L – 6 mg/L** | 31.505 mg/dL | 25.531 mg/dL | 28.494 mg/dL | 26.781 mg/dL |
| ***Ratios*** | | | | |
| **2 mg/L / 1 mg/L** | 0.983 | 1.086 | 1.014 | 1.076 |
| **3 mg/L / 2 mg/L** | 1.017 | 1.047 | 1.014 | 1.044 |
| **4 mg/L / 1 mg/L** | 1.051 | 1.173 | 1.043 | 1.157 |
| **4 mg/L / 2 mg/L** | 1.033 | 1.080 | 1.028 | 1.076 |
| **6 mg/L / 3 mg/L** | 1.049 | 1.076 | 1.043 | 1.076 |
| **8 mg/L / 4 mg/L** | 1.065 | 1.074 | 1.057 | 1.076 |
| **9 mg/L / 6 mg/L** | 1.047 | 1.041 | 1.042 | 1.044 |
| **9 mg/L / 8 mg/L** | 1.015 | 1.012 | 1.014 | 1.012 |
| **12 mg/L / 6 mg/L** | 1.094 | 1.071 | 1.087 | 1.076 |

1. With respect to the results presented in Table 2, answer the following questions:
   1. Which analysis gave constant differences in the fitted values when comparing two groups that differed by an absolute increase in *c* units in CRP levels (i.e., comparing CRP=x to CRP = x+c)? Explicitly provide all those similar paired comparisons from the table.

The analysis in problem 3 gave constant differences in the fitted values when comparing two groups that differed by an absolute increase in c units in CRP levels.. All of those differences with a 1 unit increase in CRP levels (2 mg/L – 1 mg/L, 3 mg/L – 2 mg/L, 9 mg/L – 8 mg/L) were 5.251 mg/dL. All those with a 3 unit difference in CRP levels gave a difference of 15.752.

The analysis in problem 5 was similar but did not give exactly the same differences in the fitted values when comparing two groups that differed by an absolute increase in c units in CRP levels.

* 1. Which analysis gave constant ratios of the fitted values when comparing two groups that differed by an absolute increase in *c* units in CRP levels (i.e., comparing CRP=x to CRP = x+c)? Explicitly provide all those similar paired comparisons from the table..’’

The analysis in problem 5 gave constant ratios of the fitted values when comparing two groups that differed by an absolute increase in c units in CRP levels. A 1 unit increase in CRP levels gave a ratio of 1.014 in 3 cases, and a 3 unit increase in CRP levels gave a ratio of 1.043 in 3 cases.

* 1. Which analysis gave constant differences in the fitted values when comparing two groups that differed by a relative *c*-fold increase in CRP levels (i.e., comparing CRP=x to CRP = c \* x )? Explicitly provide all those similar paired comparisons from the table.

The analysis in problem 4 gave constant differences in the fitted values when comparing two groups that differed by a relative c-fold increase in CRP levels. A doubling of CRP levels gave a difference of 25.521 in 5 cases.

* 1. Which analysis gave constant ratios in the fitted values when comparing two groups that differed by a relative *c*-fold increase in CRP levels (i.e., comparing CRP=x to CRP = c \* x )? Explicitly provide all those similar paired comparisons from the table.

The analysis in problem 6 gave constant ratios in the fitted values when comparing two groups that differed by a relative c-fold increase in CRP levels. A doubling of CRP levels gave a ratio of 1.076 in 5 cases.

1. How would you decide which of the four potential analyses should be used to investigate associations between fibrinogen and CRP?

Both fibrinogen and CRP make more sense to look at on a logarithmic scale, because both of these values have a “normal” range with potential for extreme outliers. Having a ratio that is consistent with a doubling of CRP levels, as is the case with the analysis in problem 6, described in question 8d, is reliable. Thus, analysis six would be the best potential analysis to be used to investigate associations between fibrinogen and CRP.