**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**TOTAL POINTS: 98/195**

**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.

QUESTION 1: 6/15 points (see below for detailed comments)

**Methods: Descriptive statistics were produced for CRP, fibrinogen in all subjects and separately for participants with and without prior diagnosis of CVD using t-test by “prevdis,” assuming unequal variances. The test for association of CRP and fibrinogen was evaluated with logistic regression using robust standard errors. Then, the evaluation for an association between CRP and fibrinogen was performed separately for participants with and without prior diagnosis of CVD.**

**Results:** TABLE: 3 points (minus 2 points for not stratifying Fibrinogen by levels of CRP)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE 1: CRP and fibrinogen summary measures, stratified by prior CVD diagnosis | | | | | | | |
|  | Overall  (N = 5000) | | With prior diagnosis of CVD (N = 1149) | | Without prior diagnosis of CVD (N = 3851) | | Difference between groups (unequal var.) |
|  | Observation | N  (% missing) | Observation | N  (% missing) | Observation | N  (% missing) | Est (95% CI, P-value) |
| CRP – mean (SD, min-max) | **3.61**  (6.15, 0-108) | 4915  (1.70%) | **4.40**  (6.88, 0-83) | 1131  (1.57%) | **3.38**  (5.90, 0-108) | 3802  (1.27%) | **-1.02** (-1.46, -0.57, <0.0001) |
| Fibrinogen – mean (SD, min-max) | **322.98** (67.29, 109-872) | 4933  (1.34%) | **334.46** (74.06, 138-695) | 1124  (2.18%) | **319.57** (64.76, 109-872) | 3791  (1.56%) | **-14.89** (-19.68, -10.09, <0.0001) |

SCATTER PLOT: 2 points (minus 1 point for not labeling axes with units, minus 2 points for not stratifying by history of prior CVD or providing a verbal explanation of the scatter plot)



DISCUSSION OF FINDINGS: 1 point (Minus 4 points for no discussion of the descriptive data, effect modification, or confounding. The discussion of the t-test results was accurate but not the appropriate answer according to the homework key. Logistic regression does not seem to be the correct model to use in this case because the response variable (presumably CRP – though it was not clearly defined in the methods which was the predictor and which was the response) is a continuous variable and not a binary one (unless CRP was dichotomized, but it didn’t state anywhere that was done).

**Logistic regression of CRP, fibrinogen, revealed an OR of 1.0145 with SE 0.0011 (95% CI 1.012-1.017, p<0.001). In patients with no prior diagnosis of CVD, the OR was 1.0152 with SE 0.0012 (95% CI 1.013-1.018, p<0.001). In patients with a prior diagnosis of CVD, the OR was 1.0117 with SE 0.0022 (95% CI 1.007-1.016, p<0.001).**

**Inference:**

**There is a statistically significant difference in the difference in means for both CRP and fibrinogen in participants with and without a prior diagnosis of CVD. The 95% CI suggests that the observations in our data would not be unusual if the true difference in mean CRP was anywhere from 0.57 lower to 1.46 lower in participants without a prior diagnosis of CVD. Additionally, the 95% CI suggests that the observations in our data would not be unusual if the true difference in mean fibrinogen was anywhere from 10.09 lower to 19.68 lower in participants without a prior diagnosis of CVD. By logistic regression, for each 1 unit change in fibrinogen, CRP increases by 1.45% in all patients, by 1.52% in patients with no prior diagnosis of CVD, and by 1.17% in patients with a prior diagnosis of CVD.**

1. Perform t test analyses exploring an association between mean fibrinogen and prior history of CVD.

QUESTION 2: 30/45 points (see below for detailed comments)

* 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.

8 points (Minus 2 points for not explicitly stating the point estimates and sample sizes in the discussion and also for not specifying that the results can only be considered a test of the means between the two groups if the variances are known to be equal.)

**Methods: Mean fibrinogen in all subjects and separately for participants with and without prior diagnosis of CVD will be analyzed using t-test by “prevdis,” assuming equal variances.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE 2: Fibrinogen summary measures, stratified by prior CVD diagnosis | | | | | | | |
|  | Overall  (N = 5000) | | With prior diagnosis of CVD (N = 1149) | | Without prior diagnosis of CVD (N = 3851) | | Difference between groups (equal var.) |
|  | Observation | N  (% missing) | Observation | N  (% missing) | Observation | N  (% missing) | Est (95% CI, P-value) |
| Fibrinogen – mean (SD, min-max) | **322.98** (67.29, 109-872) | 4933  (1.34%) | **334.46** (74.06, 138-695) | 1124  (2.18%) | **319.57** (64.76, 109-872) | 3791  (1.56%) | **-14.89** (-19.35, -10.42, <0.0001) |

**Inference: Mean fibrinogen is 14.89 lower in patients without a prior diagnosis of CVD. The 95% CI suggests that the observations in our data would not be unusual if the true difference in mean fibrinogen was anywhere from 10.42 lower to 19.35 lower in participants without a prior diagnosis of CVD.**

* 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.

6 points (Minus 4 points for not specifying classical linear regression, not stating that the standard errors will be different between the two tests because regression uses a pooled variance estimate, not specifying that the difference in the means in the t-test is the **slope** in the regression equation, and not stating that the t-statistics and p-values are the same for both tests.)

**Linear regression of fibrinogen (fib) and prior diagnosis of CVD (prevdis) gives information similar to the t-test assuming equal variance. As shown in the tables above (2) and below (3), linear regression gives the same estimate of the difference in means as the t-test. The 95% CI is exactly the same for the difference in means from the t-test and from linear regression. The constant from the linear regression (319.57) is equal to the observed mean fibrinogen in the participants without a prior diagnosis of CVD.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TABLE 3: Linear regression of fibrinogen, prior diagnosis of CVD | | | | |
| Fibrinogen | Coefficient | SE | p-value | 95% CI |
| Prior CVD diagnosis | 14.88508 | 2.275603 | <0.001 | (10.42389, 19.34628) |
| \_cons | 319.574 | 1.088223 | <0.001 |  |

* 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.

5 points (Minus 5 points for no discussion of the results in the table.)

**Methods: Mean fibrinogen in all subjects and separately for participants with and without prior diagnosis of CVD will be analyzed using t-test by “prevdis,” assuming unequal variances.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE 4: Fibrinogen summary measures, stratified by prior CVD diagnosis | | | | | | | |
|  | Overall  (N = 5000) | | With prior diagnosis of CVD (N = 1149) | | Without prior diagnosis of CVD (N = 3851) | | Difference between groups (unequal var.) |
|  | Observation | N  (% missing) | Observation | N  (% missing) | Observation | N  (% missing) | Est (95% CI, P-value) |
| Fibrinogen – mean (SD, min-max) | **322.98** (67.29, 109-872) | 4933  (1.34%) | **334.46** (74.06, 138-695) | 1124  (2.18%) | **319.57** (64.76, 109-872) | 3791  (1.56%) | **-14.89** (-19.68, -10.09, <0.0001) |

* 1. How could a smilar 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.

6 points (Minus 4 points for not stating that the standard errors for both the slope and Y-intercept will be very similar between the two tests, not specifying that the difference in the means in the t-test is the **slope** in the regression equation, and not stating that the p-values are approximately the same for both tests.)

**Linear regression of fibrinogen (fib) and prior diagnosis of CVD (prevdis) using Robust SE gives information similar to the t-test assuming unequal variance. As shown in the tables above (4) and below (5), linear regression gives the same estimate of the difference in means as the t-test. The 95% CI is exactly the same for the difference in means from the t-test and from linear regression. The constant from the linear regression (319.57) is equal to the observed mean fibrinogen in the participants without a prior diagnosis of CVD.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TABLE 5: Linear regression of fibrinogen, prior diagnosis of CVD | | | | |
| Fibrinogen | Coefficient | SE | p-value | 95% CI |
| Prior CVD diagnosis | 14.88508 | 2.4462 | <0.001 | (10.08926, 19.68091) |
| \_cons | 319.574 | 1.0519 | <0.001 |  |

* 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)?

5 points

**The participants with a prior diagnosis of CVD make up the group with the smaller sample size. In that group, the SD and SE are higher indicating a higher variance in the smaller group. We have been taught that in the t-test that presumes equal variance, if the group with smaller sample size has higher variance then the t-test will be anti-conservative (reporting a CI that is too narrow and p-values that are too small). When we did the same t-test analysis but presumed unequal variances, our CI was slightly wider but the p-value remained <0.0001.**

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.

QUESTION 3: 13/20 points (see below for detailed comments)

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

4 points (Minus 1 point for not specifying that the value of the Y-intercept is in the group where the value of the predictor (CRP) is zero and for not specifying the units of CRP or fibrinogen.)

**The estimated intercept from the linear regression of fibrinogen, CRP using Robust SE is a fibrinogen level of 304.015 (95% CI 301.514, 306.517).**

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

5 points

**The estimated slope from the fitted linear regression model of fibrinogen and CRP using Robust SE is a change in fibrinogen level of 5.251 for every change of 1 unit of CRP with a 95% CI of 4.604-5.898.**

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

4 points (Minus 6 points for no detailed discussion of the methods including the use of robust SE, no point estimates or sample sizes, no units for fibrinogen or CRP, and no p-value or statement about rejecting the null hypothesis)

**From linear regression, we have estimated that a change in CRP of 1 unit is associated with a change in fibrinogen of 5.251. Our data would not be unusual if the true association between CRP and fibrinogen had a slope between 4.604 and 5.898.**

* 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).

**See table below. Generated a categorical variable for CRP:**

**crp\_cat crp**

**0.5 0**

**1 [1,2)**

**2 [2,3)**

**3 [3,4)**

**4 [4,5)**

**6 [6,7)**

**8 [8,9)**

**9 [9,10)**

**12 [12,13)**

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.)

QUESTION 4: 6/20 points (see below for detailed comments)

* 1. **The estimated intercept: fibrinogen of 295.566 (95% CI 293.64, 297.49)** 3 points (Minus 2 points for not specifying mean fibrinogen level and not specifying that the value of the Y-intercept is in the group where the value of the predictor (CRP) is 1 mg/dL and for not specifying the units of CRP or fibrinogen.)
  2. **The estimated slope: change in fibrinogen level of 36.833 for every change of 1 in log(crp) when crp = 0 has been changed to 0.5.** 3 points (Minus 2 points for not specifying that the change in fibrinogen is seen for every 2.718-fold increase in CRP and for not specifying the units of CRP or fibrinogen.)
  3. **From linear regression, we have estimated that a change in log(crp) of 1 unit is associated with a change in fibrinogen of 295.566. Our data would not be unusual if the true association between CRP and fibrinogen had a slope between 293.64 and 297.49**. 0 points (Minus 10 points for no discussion of of the methods including the use of robust SE, no point estimates or sample sizes, no units for fibrinogen or CRP, and no p-value or statement about rejecting the null hypothesis. Also gives an incorrect answer where the Y-intercept is incorrectly interpreted as the slope and does not explain the use of log transformation.)
  4. **See table below. Generated a categorical variable for log(CRP):**

**Log\_crp\_cat log(crp)**

**0 <1**

**1 [1,1.5)**

**1.5 [1.5,2)**

**2 [2,2.5)**

**2.5 [2.5,3)**

**3 [3,3.5)**

**3.5 [3.5,4)**

**4 [4,4.5)**

**4.5 [4.5,5)**

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.

QUESTION 5: 11/20 points (see below for detailed comments)

* 1. **The estimated intercept: for the geometric mean of fibrinogen is 5.7068 (95% CI 5.699, 5.714). Reverse transforming these numbers gives an intercept of 300.896 (95% CI 298.65, 303.16).** 4 points (Minus 1 point for not specifying that the value of the Y-intercept is in the group where the value of the predictor (CRP) is zero and for not specifying the units of CRP or fibrinogen.)
  2. **The estimated slope: for the geometric mean of fibrinogen is 0.0139 (95% CI 0.0122, 0.0157). Reverse transforming these numbers gives a slope of 1.0140 (95% CI 1.0122, 1.0158).** 3 points (Minus 2 points for not specifying that the **percent change** in fibrinogen is seen for every 1 unit increase in CRP and for not specifying the units of CRP or fibrinogen.)
  3. **From linear regression, we have estimated that a change in CRP of 1 unit is associated with a change in fibrinogen of 1.4% (slope 1.0140). Our data would not be unusual if the true association between CRP and fibrinogen had a slope between 1.0122 and 1.0158.** 4 points (Minus 6 points for no detailed discussion of the methods including the use of robust SE, no point estimates or sample sizes, no units for fibrinogen or CRP, and no p-value or statement about rejecting the null hypothesis)
  4. **See table below. Used previously created categories for CRP level (crp\_cat).**

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.)

QUESTION 6: 11/20 points (see below for detailed comments)

* 1. **The estimated intercept: for the geometric mean of fibrinogen is 5.678587 (95% CI 5.6724, 5.6848). Reverse transforming these numbers gives an intercept of 292.5358 (95% CI 290.7315, 294.3510) 4** points (Minus 1 point for not specifying that the value of the Y-intercept is in the group where the value of the predictor (CRP) is 1 mg/dL and for not specifying the units of CRP or fibrinogen.)
  2. **The estimated slope: for the geometric mean of fibrinogen is 0.1054 (95% CI 0.0995, 0.1113). Reverse transforming these numbers gives a slope of 1.1111 (95% CI 1.1047, 1.1177).** 3 points (Minus 2 points for not specifying that the change in fibrinogen is seen for every 2.718-fold increase in CRP and for not specifying the units of CRP or fibrinogen.)
  3. **From linear regression, we have estimated that a change in logCRP of 1 unit is associated with a change in fibrinogen of 11.11% (slope 1.1111). Our data would not be unusual if the true association between logCRP and fibrinogen had a slope between 1.1047, 1.1177.** 4 points (Minus 6 points for no detailed discussion of the methods including the use of robust SE, no point estimates or sample sizes, no units for fibrinogen or CRP, not specifying that percent change in fibrinogen is seen for a 2.718-fold increase in CRP, and no p-value or statement about rejecting the null hypothesis)
  4. **See table below**

TABLE: 10/20 points (Minus 10 points for incorrect fitted values in nearly all cells)

**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: fibrinogen – mean (SD)** | | **Problem 5: geometric mean fibrinogen – mean (SD)** | | **Log(crp) level** | **Problem 4: fibrinogen – mean (SD)** | **Problem 6: geometric mean fibrinogen**  **– mean (SD)** | |
| **Log data** | **Back-transformed** | **Log data** | **Back-transformed** |
| **1 mg/L** | 298.7 (49.30) | | 5.686 (0.169) | 294.7 (1.184) | **0** | 301.6 (51.50) | 5.694 (0.173) | 297.1 (1.189) |
| **2 mg/L** | 314.3 (51.46) | | 5.737 (0.166) | 310.1 (1.181) | **1** | 335.9 (58.15) | 5.802 (0.171) | 331.0 (1.186) |
| **3 mg/L** | 333.1 (56.14) | | 5.794 (0.167) | 328.3 (1.182) | **1.5** | 353.5 (63.28) | 5.851 (0.185) | 347.6 (1.203) |
| **4 mg/L** | 344.0 (63.08) | | 5.825 (0.178) | 338.7 (1.195) | **2** | 370.7 (70.75) | 5.897 (0.194) | 363.9 (1.214) |
| **6 mg/L** | 354.7 (65.91) | | 5.854 (0.185) | 348.6 (1.203) | **2.5** | 407.0 (82.73) | 5.987 (0.215) | 398.2 (1.240) |
| **8 mg/L** | 371.4 (71.09) | | 5.900 (0.183) | 365.0 (1.201) | **3** | 448.7 (105.25) | 6.082 (0.223) | 437.9 (1.250) |
| **9 mg/L** | 360.8 (71.64) | | 5.869 (0.201) | 353.9 (1.223) | **3.5** | 491.2 (114.26) | 6.168 (0.249) | 477.2 (1.283) |
| **12 mg/L** | 370.1 (78.17) | | 5.890 (0.227) | 361.4 (1.255) | **4** | 541.3 (97.32) | 6.280 (0.181) | 533.8 (1.198) |
|  |  | |  |  | **4.5** | 555.5 (129.40) | 6.306 (0.235) | 547.8 (1.265) |

1. Complete the following table that makes comparisons (differences or ratios) of the fitted values for each of the models. QUESTION 7: 2 points for appropriately labeling the table (Minus 8 points for all incorrect values of the comparisons across CRP levels)

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |  |
| **Comparisons across CRP level** | **Problem 3: fibrinogen – mean (SD)** | **Problem 5: geometric mean fibrinogen – mean (SD)**  **(Back transformed fibrinogen)** | **Log(crp) level** | **Problem 4: fibrinogen – mean (SD)** | **Problem 6: geometric mean fibrinogen**  **– mean (SD)**  **(Back transformed fibrinogen)** |
| ***Differences*** | | | | | |
| **2 mg/L – 1 mg/L** | 15.6 | 15.4 | **2 mg/L – 1 mg/L** | 34.8 | 32.9 |
| **3 mg/L – 2 mg/L** | 18.8 | 18.2 | **3 mg/L – 2 mg/L** | 78.0 | 74.0 |
| **4 mg/L – 1 mg/L** | 45.3 | 44.0 | **4 mg/L – 1 mg/L** | 205.4 | 202.8 |
| **4 mg/L – 2 mg/L** | 29.7 | 28.6 | **4 mg/L – 2 mg/L** | 170.6 | 169.9 |
| **6 mg/L – 3 mg/L** | 21.6 | 20.3 |  |  |  |
| **8 mg/L – 4 mg/L** | 27.4 | 26.3 |  |  |  |
| **9 mg/L – 6 mg/L** | 6.1 | 5.3 |  |  |  |
| **9 mg/L – 8 mg/L** | -10.6 | -11.1 |  |  |  |
| **12 mg/L – 6 mg/L** | 15.4 | 12.8 |  |  |  |
| ***Ratios*** | | | | | |
| **2 mg/L / 1 mg/L** | 1.0522 | 1.0523 | **2 mg/L/ 1 mg/L** | 1.1036 | 1.0994 |
| **3 mg/L / 2 mg/L** | 1.0598 | 1.0587 | **3 mg/L/ 2 mg/L** | 1.2104 | 1.2034 |
| **4 mg/L / 1 mg/L** | 1.1517 | 1.1493 | **4 mg/L/ 1 mg/L** | 1.6115 | 1.6127 |
| **4 mg/L / 2 mg/L** | 1.0945 | 1.0922 | **4 mg/L/ 2 mg/L** | 1.4602 | 1.4669 |
| **6 mg/L / 3 mg/L** | 1.0648 | 1.0618 |  |  |  |
| **8 mg/L / 4 mg/L** | 1.0797 | 1.0776 |  |  |  |
| **9 mg/L / 6 mg/L** | 1.0172 | 1.0152 |  |  |  |
| **9 mg/L / 8 mg/L** | 0.9715 | 0.9696 |  |  |  |
| **12 mg/L / 6 mg/L** | 1.0434 | 1.0367 |  |  |  |

1. With respect to the results presented in Table 2, answer the following questions:

QUESTION 8: 6/20 points (see below for detailed comments)

* 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.

**In problems 3 and 5, there was a nearly constant difference between 2 mg/L – 1 mg/L and 3 mg/L – 2 mg/L, when c=1. For problems 4 and 6, there was no constant difference.** (3 points for correctly selecting Problem 3, Minus 2 points for incorrectly selecting Problem 5)

* 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.

**Problems 3 and 5 had a nearly constant ratio for 2 mg/L/1 mg/L and 3 mg/L/2 mg/L.** (3 points for correctly selecting Problem 5, Minus 2 points for incorrectly selecting Problem 3)

* 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.

**In problems 3 and 5, there was a nearly constant difference for CRP values of 4-2 and 8-4.** 0 points for not selecting Problem 4. (See key for details)

* 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.

**In problems 3 and 5, the analysis gave constant ratios in groups that had a doubling, for example 2/1, 4/2, 6/3, 8/4, 12/6.** 0 points for not selecting Problem 6 (See key for details)

1. How would you decide which of the four potential analyses should be used to investigate associations between fibrinogen and CRP? 3/5 points (Minus 2 points for not discussing the biological basis for log transforming CRP since it likely increases multiplicatively in the setting of increasing inflammation)

The decision about deciding to log transform the data depends on if the data are likely to follow a linear trend relationship. With the spread of the data and the likelihood of outliers, it would likely be best to log transform at least fibrinogen and probably CRP as well.