**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #2**

January 13, 2015

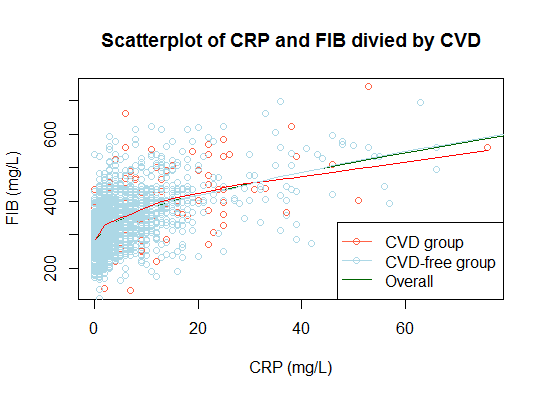
All questions relate to associations between the two biomarkers C-reactive protein (CRP) and fibrinogen (FIB), and how any such association might depend upon prevalence of prior cardiovascular disease (CVD). This homework again uses the subset of information that was collected to examine inflammatory 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.

**Method: Both CRP and CVD are continuous variables; hence I used scatter plots with lowess lines to analyze the association between CRP and FIB. After plotting, I compared the mean FIB across levels of CRP. If difference is detected, we can conclude association.**

**Results:**

**There are 5000 observations in the original data set. 101 of them contain missing values in CRP, FIB or history of CVD: 6 of the subjects with prior CVD are missing and 95 of the subjects without prior CVD are missing. These points are excluded in this analysis. The information from the missing data cannot be obtained otherwise by the information we have now.**



**In this picture, the points are color coded by sex (red for the CVD group and blue for the CVD-free group).The lowess smooth lines are also provided for both group (red for the CVD group and blue for the CVD-free group) and the overall group (green for the overall group).**

**The mean FIB is different across different levels of CRP, thus there exist an overall association between CRP and FIB. There is an overall upward trend in the data, but we have to take into consideration that samples are getting sparse with higher CRP. Also within (0 mg/L, 20 mg/L) CRP, the trend is slightly skewed, although that is most likely because of the marginal effect by the values close to 0. The variation does seem similar although the sample size is getting small as CRP gets larger. It is not totally unreasonable to assume homoscedasticity.**

**There are 1149 subjects that have prior history of CVD and 3851 subjects that have no prior history of CVD.**

**In both groups, the mean FIB varies across different levels of CRP, which indicates an association between FIB and CRP. The lowess curves are similar with an upwards trend. The skew near 0 is probably caused by the marginal effect of lowess curve. The range (can't tell variance directly) in each group appears to be similar as CRP varies. Thus it is not totally unreasonable to assume homoscedasticity in each group.**

**This upward tendency can also be observed from the following table. The mean CRP level goes up significantly as the level of FIB goes up, both in CVD group, CVD-free group and overall group.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **(Mean (SD; Minimum-Maximum;n))** | | |
| **FIB level** | **108-363 mg/L** | **364-618 mg/L** | **619-873 mg/L** |
| **CVD group** | **2.63 (2.89;0-32;805)** | **8.52 (9.68;0-76;314)** | **5.67(40;6-83;3)** |
| **CVD –free group** | **2.35 (3.03;0-42;3007)** | **7.07(9.80;0-100;761)** | **39.0(28.48;11-108;9)** |
| **Overall** | **2.41(3;0-42;3812)** | **7.50(9.79;0-100;1075)** | **41.9(30.1;6-108;12)** |

**(The reason I divided FIB in to these three levels is if divided into more levels, there will be empty group.)**

**Table: 3/5**

**Graph: 5/5**

**Discussion: 3/5**

**Total: 11/15**

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.

**Method: I used two-sample t-test that assumes equal variance to test the mean difference between the group of 1124 subjects with prior history of CVD and the group of 3791 subjects who do not. (85 of the 5000 subjects have missing value on CVD history or FIB) Also, the 95% of the confidence interval for the difference was provided based on the same variance processing strategy.**

**Result:**

|  |  |  |
| --- | --- | --- |
|  | **Sample size** | **Mean** |
| **Prior history of CVD** | 1124 | 334.5 mg/L |
| **No prior history of CVD** | 3791 | 319.6 mg/L |

**As presented in the form above, the mean in the two CVD groups were 334.5 mg/L (330 mg/L, 339 mg/L) and 319.6 mg/L (318 mg/L, 322 mg/L). According to the 95% confidence interval, the 14.9 mg/L higher in mean FIB in the group of subjects that have prior history of CVD would not be considered unusual if the true difference between means was between 10.4 mg/L and 19.3 mg/L, higher in the group with prior history of CVD. The t-test adopted has the assumption of equal variance and the result is significant under the significance level of 0.05 (two-sided P-value<0.001), therefore we can reject the null hypothesis that the means are the same in the group with prior history of CVD and the group without.**

**Total: 10/10**

* 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.

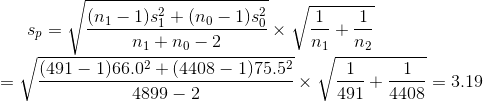
**Methods: The corresponding linear regression is the one with a binary predictor of interest under assumptions of homoscedasticity. The predictor of interest is the CVD history, the outcome is the levels of FIB. They are modeled by a simple linear regression model with 2 parameters: slope and intercept. The standard error used in this analysis is the classic pooled standard error.**

**Results:**

**The intercept of the regression on CVD and FIB is 319.5 mg/L, with a 95% confidence interval (317 mg/L, 322 mg/L). The intercept is correspondent to the mean FIB in the group without prior history of CVD. The confidence interval given here is different from the confidence interval given by t test, for the standard error here is calculated by: and the standard error in the t test is the standard error of the group without prior history of CVD. Also the critical values in the confidence intervals have different degree of freedom: in the regression the degree of f914reedom is 3790, and in t-test the degree of freedom is 4914.**

**The slope of the simple linear regression model is 14.9 mg/L, with a 95% confidence interval (10.4 mg/L, 19.3 mg/L). The slope is correspondent to the difference in the two CVD groups. The confidence interval given by t test and simple linear regression are the same. They handle the standard error in the same way by using the pooled standard error.**

**The pooled standard error is given by:**

****

**The p-value is also the same in the t test and the regression model (for slope). They are both getting the p-value from the sampling distribution of the mean difference between the group with CVD history and the group without. Also, they handle the standard error (the standard deviation of the sampling distribution) in the same way.**

**Total: 9/10**

* 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:**

**I used two sample t-test without assuming same variance to test the mean difference in FIB level of subjects with prior history of CVD and those without. Also, the 95% of the confidence interval for the difference was provided based on the same variance processing strategy.**

**Result:**

|  |  |  |
| --- | --- | --- |
|  | **Sample size** | **Mean** |
| **Prior history of CVD** | 1124 | 334.5 mg/L |
| **No prior history of CVD** | 3791 | 319.6 mg/L |

**As presented in the form above, the mean in the two CVD groups were 334.5 mg/L (330 mg/L, 339 mg/L) and 319.6 mg/L (318 mg/L, 322 mg/L). According to the 95% confidence interval, the 14.9 mg/L higher in mean FIB in the group of subjects that have prior history of CVD would not be considered unusual if the true difference between means was between 10.1 mg/L and 19.7 mg/L, higher in the group with prior history of CVD. The t-test does not assume equal variance and the result is significant under the significance level of 0.05 (two-sided P-value<0.001), therefore we can reject the null hypothesis that the means are the same in the group with prior history of CVD and the group without.**

**Total:8/10**

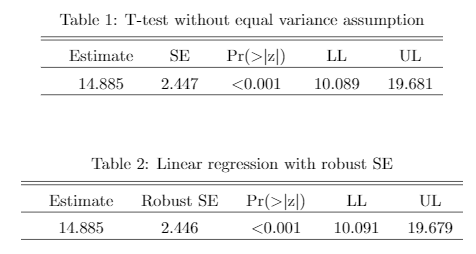
* 1. How could a smilar analysis as presented in part c have been performed with linear regression? Explicitly provide the correspondences between the various statistical output from each of the analyses.

**Answer: A linear regression using robust SE is similar to the analysis in part c.**

|  |  |  |
| --- | --- | --- |
|  | **Sample size** | **Mean** |
| **Prior history of CVD** | 1124 | 334.5 mg/L |
| **No prior history of CVD** | 3791 | 319.6 mg/L |

**As presented in the form above, the point estimators are exactly the same. The mean difference is 14.9 mg/L higher in mean FIB in the group of subjects that have prior history of CVD.**

**The estimated difference in means in t-test and the slope in the regression have slightly different SE. Also the critical values in the confidence intervals have different degree of freedom: in t-test the degree of freedom is 1664.567, and in the regression the degree of freedom is 4913. The different SE, p-value and confidence interval are presented in the following table:**

****

**As presented, the confident interval for the mean difference in the t-test, because of the differences in SE and critical value.**

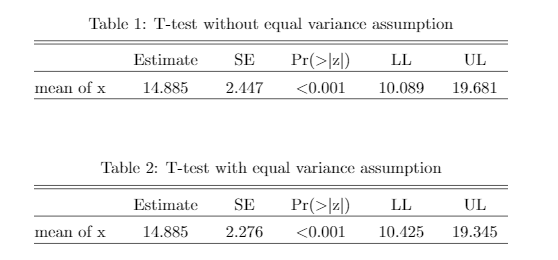
**Total = 5/10**

* 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)?

**Answer: With the sample size and the variance of each group calculated in part a (in the group with prior history of CVD, the sample size is 1124 and the variance is 5485; in the group without prior history of CVD, the sample size is 3791 and the variance is 4194), we know that the group with smaller sample size has higher variance, thus the t-test that presumes equal variance is anti-conservative with smaller p-value and narrower CI than the t-test without equal variance assumption.**

**(Weaker association)**

**The result is in concord with the analysis:**

****

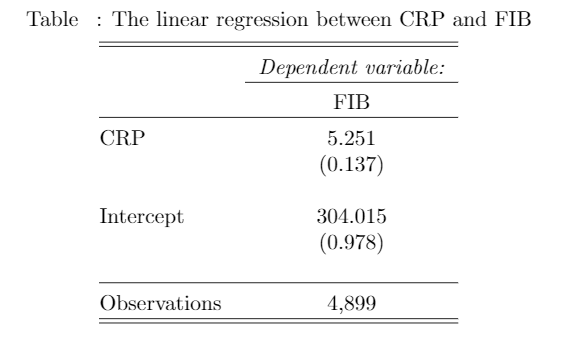
**Total:5/5**

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**.

**(101 of the 5000 subjects have missing value on CRP or FIB)**

1. Perform a statistical analysis evaluating an association between mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, untransformed random variable.

**The regression results are presented as follow:**

****

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

**Answer: The intercept from the fitted regression model is that the estimated mean FIB is 304 mg/L for population with CRP equals 0 mg/L.**

**Total: 5/5**

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

**Answer: The slope from the fitted regression model is that the estimated difference in mean FIB is 5.25 mg/L when the difference in CRP is 1 mg/L (higher in the greater CRP group) if the true relationship between FIB and CRP is linear. The slope is the average difference in mean FIB per 1 mg/L difference in CRP if the relationship between FIB and CRP is not linear.**

**Total:4/5**

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

**Answer: The slope of the linear regression suggests that there is a 5.25 mg/L difference in FIB when there is a 1 mg/L difference in CRP (higher in the greater CRP group). Also, from the 95% CI we can conclude that the observed mean difference is not unusual if the true difference in mean FIB per 1 mg/L CRP fell between 4.98 mg/L and 5.52 mg/L (higher in the greater CRP group). The result is significant with a two-sided p-value <0.001. Therefore, we can reject the null hypothesis that there is no linear trend in the average FIB across CRP groups.**

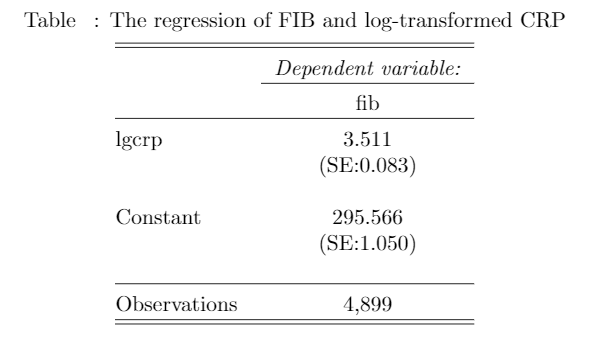
**(The regression has the assumption of homoscedasticity. When we use the robust standard error, the estimates are the same, but the 95% CI of mean FIB difference become 4.60 mg/L to 5.90 mg/L (higher in the greater CRP group).)**

**Total: 5/10**

* 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).

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 regression results are presented as follow:**



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

**Answer: The intercept from the fitted regression model is that the estimated mean FIB is 296 mg/L of population with CRP equals 1 mg/L. (log CRP=log 1=0)**

**Total: 5/5**

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

**Answer: The slope from the fitted regression model is that the estimated difference in mean FIB is 3.51 mg/L when the difference in CRP is by 10% (higher in the greater CRP group) if the true relationship between FIB and CRP is linear. The slope is the average difference in mean FIB per 10% difference in CRP if the relationship between FIB and CRP is not linear.**

**Total: 5/5**

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

**Answer: The slope of the linear regression suggests that there is a 3.51 mg/L difference in FIB when there is a 10% difference in CRP (higher in the greater CRP group). Also, from the 95% CI we can conclude that the observed mean difference is not unusual if the true difference in mean FIB per 10% CRP fell between 3.35 mg/L and 3.67 mg/L. The result is significant with a two-sided p-value <0.001. Therefore, we can reject the null hypothesis that there is no linear trend in the average FIB across log-transformed CRP groups.**

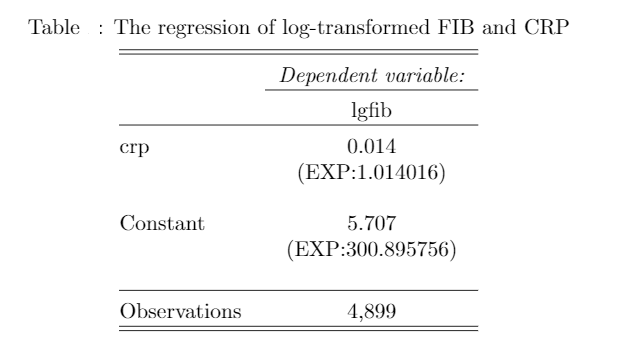
**(The regression has the assumption of homoscedasticity. When we use the robust standard error, the estimates are the same, but the 95% CI of the mean FIB become 3.30 mg/L to 3.73 mg/L (higher in the greater CRP group).)**

**Total: 5/10 (no methods)**

* 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).

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 regression results are presented as follow:**



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

**Answer: The intercept from the fitted regression model after being exponentiated is the estimated geometric mean FIB is 301 mg/L of population with CRP equals 0 mg/L.**

**Total: 5/5**

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

**Answer: The slope from the fitted regression model is the estimated ratio of geometric means in FIB when the difference in CRP is by 1 mg/L (higher in the greater CRP group) if the true relationship between log transformed FIB and CRP is linear. The slope is the average ratio in geometric mean FIB per 1 mg/L difference in CRP if the relationship between log transformed FIB and CRP is not linear.**

**Total: 5/5**

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

**Answer: The slope of the linear regression suggests that the geometric mean in FIB is 1.39% higher when there is a 1 mg/L difference in CRP (higher in the greater CRP group). Also, from the 95% CI we can conclude that the observed geometric mean ratio is not unusual if the true geometric mean of FIB in the higher CRP group were between 1.32% and 1.49% higher than that of the lower CRP group. The result is significant with a two-sided p-value <0.001. Therefore, we can reject the null hypothesis that there is no linear trend in the average FIB across log-transformed CRP groups.**

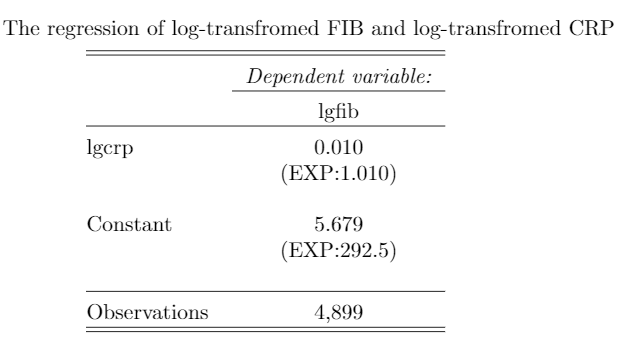
**(The regression has the assumption of homoscedasticity. When we use the robust standard error, the estimates are the same, but the 95% CI become 1.22% to 1.58% higher in geometric mean of FIB in the higher CRP group.)**

**Total: 5/10 (no methods)**

* 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).

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 regression results are presented as follow:**



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

**Answer: The intercept from the fitted regression model after being exponentiated is the estimated geometric mean FIB is 293 mg/L of population with CRP equals 1 mg/L. (log CRP=log 1=0)**

**Total: 5/5**

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

**Answer: The slope from the fitted regression model is the estimated ratio of geometric means in FIB (1.01) when the difference in CRP is by 10% (higher in the greater CRP group) if the true relationship between log transformed FIB and CRP is linear. The slope is the average ratio in geometric mean FIB per 10% difference in CRP (higher in the greater CRP group) if the relationship between log transformed FIB and CRP is not linear.**

**Total:4/5**

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

**Answer: The slope of the linear regression suggests that the geometric mean in FIB is 1% higher in the higher CRP group when there is a 10% difference in CRP (higher in the greater CRP group). Also, from the 95% CI we can conclude that the observed geometric mean ratio is not unusual if the true geometric mean of FIB in the higher CRP group were between 0.959% and 1.06% higher than that of the lower CRP group. The result is significant with a two-sided p-value <0.001. Therefore, we can reject the null hypothesis that there is no linear trend in the average log-transformed FIB across log-transformed CRP groups.**

**(The regression has the assumption of homoscedasticity. When we use the robust standard error, the estimates are the same, but the 95% CI become 0.953% to 1.07% higher in geometric mean of FIB in the higher CRP group.)**

**Total: 5/10 (no methods)**

* 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).

**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 FIB level)** | **Problem 4: (Mean FIB level)** | **Problem 5: (Geometric mean FIB level)** | **Problem 6: (Geometric mean FIB level)** |
| **1 mg/L** | 309 mg/L | 296 mg/L | 305 mg/L | 293 mg/L |
| **2 mg/L** | 315 mg/L | 321 mg/L | 309 mg/L | 315 mg/L |
| **3 mg/L** | 320 mg/L | 336 mg/L | 314 mg/L | 328 mg/L |
| **4 mg/L** | 325 mg/L | 347 mg/L | 318 mg/L | 339 mg/L |
| **6 mg/L** | 336 mg/L | 362 mg/L | 327 mg/L | 353 mg/L |
| **8 mg/L** | 346 mg/L | 372 mg/L | 336 mg/L | 364 mg/L |
| **9 mg/L** | 351 mg/L | 376 mg/L | 341 mg/L | 369 mg/L |
| **12 mg/L** | 367 mg/L | 387 mg/L | 356 mg/L | 380 mg/L |

Total:20/20 (for part d for questions 3,4,5,6)

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 FIB difference)** | **Problem 4:**  **(Mean FIB difference)** | **Problem 5: (Geometric mean FIB )** | **Problem 6: (Geometric mean FIB)** |
| ***Differences*** | | | | |
| **2 mg/L – 1 mg/L** | 5.25 mg/L | 25.5 mg/L | 4.28 mg/L | 22.2 mg/L |
| **3 mg/L – 2 mg/L** | 5.25 mg/L | 14.9 mg/L | 4.34 mg/L | 13.7 mg/L |
| **4 mg/L – 1 mg/L** | 15.75 mg/L | 51.1 mg/L | 13.0 mg/L | 46.0 mg/L |
| **4 mg/L – 2 mg/L** | 10.5 mg/L | 25.5 mg/L | 8.73 mg/L | 23.9 mg/L |
| **6 mg/L – 3 mg/L** | 15.75 mg/L | 25 .5 mg/L | 13.4 mg/L | 24.9 mg/L |
| **8 mg/L – 4 mg/L** | 21 mg/L | 25 .5 mg/L | 18.2 mg/L | 25.7 mg/L |
| **9 mg/L – 6 mg/L** | 15.75 mg/L | 14.9 mg/L | 13.9 mg/L | 15.4 mg/L |
| **9 mg/L – 8 mg/L** | 5.25 mg/L | 4.34 mg/L | 4.71 mg/L | 4.55 mg/L |
| **12 mg/L – 6 mg/L** | 31.5 mg/L | 25.5 mg/L | 28.5 mg/L | 26.8 mg/L |
| ***Ratios*** | | | | |
| **2 mg/L / 1 mg/L** | 1.02 | 1.09 | 1.01 | 1.08 |
| **3 mg/L / 2 mg/L** | 1.02 | 1.05 | 1.01 | 1.04 |
| **4 mg/L / 1 mg/L** | 1.05 | 1.17 | 1.04 | 1.16 |
| **4 mg/L / 2 mg/L** | 1.03 | 1.08 | 1.03 | 1.08 |
| **6 mg/L / 3 mg/L** | 1.05 | 1.08 | 1.04 | 1.08 |
| **8 mg/L / 4 mg/L** | 1.06 | 1.07 | 1.06 | 1.08 |
| **9 mg/L / 6 mg/L** | 1.05 | 1.04 | 1.04 | 1.04 |
| **9 mg/L / 8 mg/L** | 1.02 | 1.01 | 1.01 | 1.01 |
| **12 mg/L / 6 mg/L** | 1.09 | 1.07 | 1.09 | 1.08 |

Total:10/10

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. **Analysis in problem 3 Total: 5/5**

**Difference by 1: Difference by 3:**

|  |  |  |  |
| --- | --- | --- | --- |
| **2 mg/L – 1 mg/L** | 5.25 mg/L | **4 mg/L – 1 mg/L** | 15.75 mg/L |
| **3 mg/L – 2 mg/L** | 5.25 mg/L | **6 mg/L – 3 mg/L** | 15.75 mg/L |
| **9 mg/L – 8 mg/L** | 5.25 mg/L | **9 mg/L – 6 mg/L** | 15.75 mg/L |

* 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.

**Analysis in problem 5 Total: 5/5**

**Difference by 1: Difference by 3:**

|  |  |  |  |
| --- | --- | --- | --- |
| **2 mg/L / 1 mg/L** | 1.01 | **4 mg/L /1 mg/L** | 1.04 |
| **3 mg/L / 2 mg/L** | 1.01 | **6 mg/L / 3 mg/L** | 1.04 |
| **9 mg/L / 8 mg/L** | 1.01 | **9 mg/L / 6 mg/L** | 1.04 |

* 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.

**Analysis in problem 4 Total: 5/5**

**Ratio=2 : Ratio=1.5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **2 mg/L – 1 mg/L** | 25.5 mg/L | **3 mg/L – 2 mg/L** | | 1.05 |
| **4 mg/L – 2 mg/L** | 25 .5 mg/L | **9 mg/L -6 mg/L** | | 1.05 |
| **6 mg/L – 3 mg/L** | 25.5 mg/L | |  | |
| **8 mg/L – 4 mg/L** | 25 .5 mg/L | |  | |
| **12 mg/L – 6 mg/L** | 25 .5 mg/L | |  | |

* 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.

**Analysis in problem 6 Total: 5/5**

**Ratio=2 Ratio=1.5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **2 mg/L / 1 mg/L** | 1.08 | **3 mg/L / 2 mg/L** | | 1.04 |
| **4 mg/L / 2 mg/L** | 1.08 | **9 mg/L / 6 mg/L** | | 1.04 |
| **6 mg/L / 3 mg/L** | 1.08 | |  | |
| **8 mg/L / 4 mg/L** | 1.08 | |  | |
| **12 mg/L / 6 mg/L** | 1.08 | |  | |

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

**Answer: I will either use analysis in problem 3 or analysis in problem 6. FIB and CRP are both biomarkers of inflammation with same unit, hence I will choose to do the same operation on them due to their similar scientific nature. Analysis in problem 3 used difference, which is easier for people to compare and understand. Also difference works well when describing the scientific importance of the difference.**

**In our case, the number of FIB and CRP are not very small, so taking the difference is not unreasonable. And the difference is stable. It won’t be affected by the denominator like in the ratio. If there were scientific evident that the FIB and CRP are operating in multiplicative scale, the ratio will be a better choice, for we get back to additive scale by log-transform the data. Also, if we want to downweight outliers, the geometric mean would be a better choice.**

**Total: 5/5**