**BIOST 518/515: Applied Biostatistics II/Biostatistics II**

Emerson, Winter 2015

**Homework 02**

January 26, 2015

**Total: 189/195 ( Excellent, it must be a biostatistic phd’s homework)**

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.(13 points given)

**Methods:**

Scatterplots are used as a descriptive statistical analysis for the association between CRP and FIB both overall and separately for groups having no prior history of diagnosed cardiovascular disease (CVD) or having prior diagnosed CVD. LOWESS lines were fitted to the scatterplots to visualize the trends in the entire data as well as subgroups of data. An additional table provided simple descriptive statistics, including mean, standard deviation, minimum and maximum of CRP and FIB in the sample as well as subgroups.

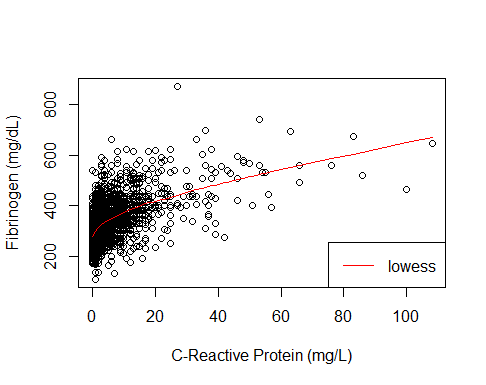
**Results:**

The total number of observations in the dataset is 5,000. 67 of those subjects (49 with no prior history of diagnosed CVD and 18 with prior diagnosed CVD) are missing data on C-reactive protein (CRP), while 85 subjects (60 with no prior history of diagnosed CVD and 25 with prior diagnosed CVD) are missing data on fibrinogen (FIB). In total, there were 101 subjects with missing data on either CRP or FIB. These subjects were omitted from all analyses. There is no additional information to determine how these omissions may impact our analysis.

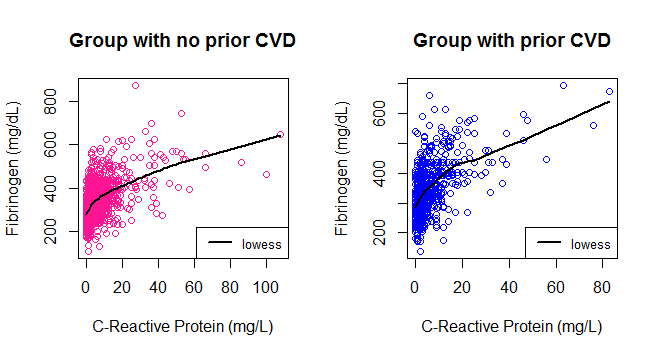
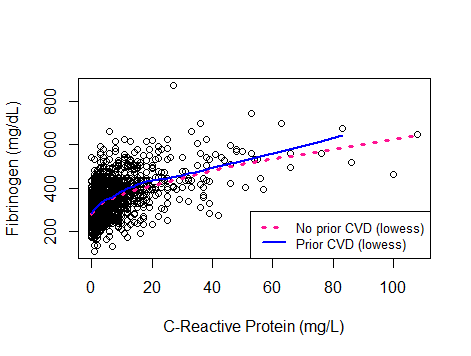
Figure 1 shows a scatterplot of CRP and FIB in the entire sample. A preliminary analysis using locally weighted scatterplot smoothing (LOWESS) shows a positive linear trend within the data: observations with higher mean CRP tend to have higher mean FIB and vice versa.

Figure 2a shows the same scatterplot with different LOWESS lines for observations with no prior history of diagnosed CVD and observations with prior diagnosed CVD. Figure 2b and 2c depict the same information, with observations of those two groups in two separate graphs. The group LOWESS lines also show similar positive linear relationship between CRP and FIB. Group with prior CVD has a slightly greater slope of the LOWESS line.(plots better be displayed on a log-log scale, and incorporated in one figure for conparison, -1 point)

Table 1 provides additional information, such as the number of observations, mean, standard deviation, minimum and maximum of FIB in the entire sample of 4,899 subjects with complete data on CRP and FIB, as well as in each group determined by prior history of diagnosed CVD and by CRP levels. Across subgroups defined by the presence and absence of diagnosed CVD, higher CRP level corresponded with higher mean and standard deviation of FIB.（the number of subjects included in the analysis should be 4933, -1 point; )



**Figure 1. Relationship between C-Reactive Protein and Fibrinogen**



**(a) Entire Sample**

**Figure 2. Relationship between C-Reactive Protein and Fibrinogen**

**(b) No Prior CVD**

**(c) With Prior CVD**

**Table 1. Descriptive Statistics for Fibrinogen Levels within Groups Defined by C-Reactive Protein (CRP) Level and Status of Prior Diagnosed Cardiovascular Diseases (CVD)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Fibrinogen (mg/dL)1** | | |
| **Serum CRP Level** | **No prior diagnosed CVD** | **Prior diagnosed CVD** | **Entire Sample** |
| Below 1 mg/L | n = 348  277.48 (48.52; 172 – 436) | n = 78  290.23 (57.93; 180 – 540) | n = 426  279.81 (50.55; 172 – 540) |
| Between 1 and 3 mg/L | n = 2,597  310.02 (52.46; 109 – 562) | n = 709  314.84 (55.60, 138 – 592) | n = 3,306  311.05 (53.18; 109 – 592) |
| Above 3 mg/L | n = 832  367.20 (78.88; 132 – 872) | n = 335  386.29 (84.50, 175 – 695) | n = 1,167  372.68 (80.96; 132 – 872) |
| **All Levels** | **n = 3,777**  **319.62 (64.83; 109 – 872)** | **n = 1,122**  **334.46 (74.11; 138 – 695)** | **n = 4,899**  **323.02 (67.35; 109 – 872)** |

*1 Descriptive statistics presented: number of observations, followed by mean (standard deviation; minimum – maximum)*

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

a. 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. (9 points given)

**Methods:**

Mean FIB values were compared across the two groups defined by presence or absence of prior history of CVD. Difference in the means was tested using a two-sample t-test that assumes equal variance within each group. A 95% confidence interval was constructed accordingly, assuming equal variance.( two sided p value were used here, and how missing data were handled)

**Results**

Defined by presence or absence of prior history of CVD, there were 3,777 patients with no prior history of diagnosed CVD and 1,122 patients with prior diagnosed CVD when omitting 101 patients with missing data in either FIB or CRP. The mean FIB values were 334.46 mg/dL among those with prior diagnosed CVD, and 319.62 mg/dL among those with no prior history of diagnosed CVD. The mean FIB among subjects with prior CVD tended to be 14.85 mg/dL higher than the mean FIB among those without CVD. With 95% confidence, this result is consistent with the true mean FIB among those with prior CVD being anywhere between 10.38 mg/dL to 19.32 mg/dL higher than the true mean FIB among those with no prior CVD. Assuming equal variance across two groups, this result is statistically significant at the 0.05 critical threshold with a two-sided p-value of less than 0.0001. As a result, we reject the null hypothesis that the mean FIB values were not different by presence or absence of prior history of CVD, in favor of an alternative hypothesis that mean FIB level were associated with prior history of CVD.

If only 85 patients with missing data in FIB were omitted, there were 3,791 patients with no prior history of diagnosed CVD and 1,124 patients with prior diagnosed CVD. The mean FIB values were 334.46 mg/dL among those with prior diagnosed CVD, and 319.57 mg/dL among those with no prior history of diagnosed CVD. The mean FIB among subjects with prior CVD tended to be 14.89 mg/dL higher than the mean FIB among those without CVD. With 95% confidence, this result is consistent with the true mean FIB among those with prior CVD being anywhere between 10.42 mg/dL to 19.35 mg/dL higher than the true mean FIB among those with no prior CVD. Assuming equal variance across two groups, this result is statistically significant at the 0.05 critical threshold with a two-sided p-value of less than 0.0001. As a result, we reject the null hypothesis that the mean FIB values were not different by presence or absence of prior history of CVD, in favor of an alternative hypothesis that mean FIB level were associated with prior history of CVD.

b. 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. (8 points given)

**Methods:**

An ordinary least squares regression that presumes homoscedasticity was performed to examine the relationship between the response variable, FIB, and the predictor of interest, prior history of CVD, that defines two groups.

**Results:**

The same analysis as presented in part a could have been performed with an ordinary least squares regression that presumes homoscedasticity to evaluate the association between the mean FIB (response variable) and the presence of prior diagnosed CVD (predictor of interest).

In the case of omitting 101 patients, the estimated coefficient for the presence of prior diagnosed CVD is 14.85, which indicates that the presence of prior diagnosed corresponds to 14.85 mg/dL increase in mean FIB. This coefficient is exactly the same as the point estimate for difference of mean FIB in two groups in part a. The 95% confidence interval from the classical linear regression is between 10.38 and 19.32, which is exactly the same as the 95% confidence interval found in part a. The test statistics are slightly different, but the conclusions are similar, as the p-value is less than 0.0001 from the linear regression model.

In the case of omitting 85 patients, the estimated coefficient for the presence of prior diagnosed CVD is 14.89, which indicates that the presence of prior diagnosed corresponds to 14.89 mg/dL increase in mean FIB. This coefficient is exactly the same as the point estimate for difference of mean FIB in two groups in part a. The 95% confidence interval from the classical linear regression is between 10.42 and 19.35, which is exactly the same as the 95% confidence interval found in part a. The test statistics are slightly different, but the conclusions are similar, as the p-value is less than 0.0001 from the linear regression model. ( t statistics will be equal in absolute value and no mentioning about SE, -2 points)

c. 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. (10 points given)

**Methods:**

Mean FIB values were compared across the two groups defined by presence or absence of prior history of CVD. Difference in the means was tested using a two-sample t-test that does not assume equal variance within each group. A 95% confidence interval was constructed accordingly, without assuming equal variance.

**Results**

Defined by presence or absence of prior history of CVD, there were 3,777 patients with no prior history of diagnosed CVD and 1,122 patients with prior diagnosed CVD. The mean FIB values were 334.46 mg/dL among those with prior diagnosed CVD, and 319.62 mg/dL among those with no prior history of diagnosed CVD. The mean FIB among subjects with prior CVD tended to be 14.85 mg/dL higher than the mean FIB among those without CVD. With 95% confidence, this result is consistent with the true mean FIB among those with prior CVD being anywhere between 10.04 mg/dL to 19.65 mg/dL higher than the true mean FIB among those with no prior CVD. Without assuming equal variance across two groups, this result is statistically significant at the 0.05 critical threshold with a two-sided p-value of less than 0.0001. As a result, we reject the null hypothesis that the mean FIB values were not different by presence or absence of prior history of CVD, in favor of an alternative hypothesis that mean FIB level were associated with prior history of CVD.

If only 85 patients with missing data in FIB were omitted, there were 3,791 patients with no prior history of diagnosed CVD and 1,124 patients with prior diagnosed CVD. The mean FIB values were 334.46 mg/dL among those with prior diagnosed CVD, and 319.57 mg/dL among those with no prior history of diagnosed CVD. The mean FIB among subjects with prior CVD tended to be 14.89 mg/dL higher than the mean FIB among those without CVD. With 95% confidence, this result is consistent with the true mean FIB among those with prior CVD being anywhere between 10.09 mg/dL to 19.68 mg/dL higher than the true mean FIB among those with no prior CVD. Without assuming equal variance across two groups, this result is statistically significant at the 0.05 critical threshold with a two-sided p-value of less than 0.0001. As a result, we reject the null hypothesis that the mean FIB values were not different by presence or absence of prior history of CVD, in favor of an alternative hypothesis that mean FIB level were associated with prior history of CVD.

d. 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. (9 points given)

**Methods:**

An ordinary least squares regression that does not presume homoscedasticity, i.e. using robust standard error estimate, was performed to examine the relationship between the response variable, FIB, and the predictor of interest, prior history of CVD, that defines two groups.

**Results**

The same analysis as presented in part c could have been performed with an ordinary least squares regression that does not presumes homoscedasticity, i.e. using robust standard error estimate, to evaluate the association between the mean FIB (response variable) and the presence of prior diagnosed CVD (predictor of interest).

In the case of omitting 101 patients, the estimated coefficient for the presence of prior diagnosed CVD is 14.85, which indicates that the presence of prior diagnosed corresponds to 14.85 mg/dL increase in mean FIB. This coefficient is exactly the same as the point estimate for difference of mean FIB in two groups in part c (as well as part a and b). The 95% confidence interval from the classical linear regression is between 10.04 and 19.65, which is similar to the 95% confidence interval found in part c. The test statistics are slightly different, but the conclusions are similar, as the p-value is less than 0.0001 from the linear regression model.

In the case of omitting 85 patients, the estimated coefficient for the presence of prior diagnosed CVD is 14.89, which indicates that the presence of prior diagnosed corresponds to 14.89 mg/dL increase in mean FIB. This coefficient is exactly the same as the point estimate for difference of mean FIB in two groups in part c (as well as part a and b). The 95% confidence interval from the classical linear regression is between 10.09 and 19.68, which is similar to the 95% confidence interval found in part c. The test statistics are slightly different, but the conclusions are similar, as the p-value is less than 0.0001 from the linear regression model. ( no mentioning about SEs, -1 point)

e. 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 given)

**Answer:**

As shown in Table 1 from question 1, the group with prior diagnosed CVD has a smaller sample size (1,122 observations, versus 3,777 observations without prior history of diagnosed CVD if omitting 101 patients; and 1,124 versus 3,791 observations when omitting 85 patients with missing data on FIB only). However, this group had a higher standard deviation compared to the group without CVD (74.11 versus 64.83 if omitting 101 patients; and 74.06 versus 64.76 if omitting 85 patients). Therefore, the t-test assuming equal variance will be anti-conservative, i.e. the reported p-value from such test would be too small and the reported confidence interval would be too narrow. As a result, when we used the t-test that does not assume equal variance as in part c, we could have predicted that the new t-statistic would be smaller than the t-statistic in part a, which leads to a larger reported p-value (still very much smaller than 0.0001). The reported confidence interval in part c, as a result, would be wider than the confidence interval found in part a.

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**.

3. Perform a statistical analysis evaluating an association between mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, untransformed random variable. (25 points given)

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

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

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

d. 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).

**Methods:**

An ordinary least squares regression that does not presume homoscedasticity, i.e. using robust standard error estimate, was used to examine the relationship between the mean FIB (response variable) and CRP levels (predictor of interest), modeling CRP as a continuous, untransformed random variable

**Results**

The estimated intercept from the fitted regression model is 304.0152. This value indicates that the mean FIB level for observation with CRP level of 0 mg/L is 304.0152 mg/dL.

The estimated slope from the fitted regression model is 5.2509. This value indicates that 1 mg/L increase in CRP level is corresponding to 5.2509 mg/dL increase in mean FIB level.

Presuming homoscedasticity, the 95% confidence interval for the estimated slope of the fitted regression model is from 4.983 to 5.519. Without such presumption, the 95% confidence interval is from 4.604 to 5.898. Both results are statistically significant at the 0.05 critical threshold with p-values less than 0.0001. As a result, we reject the null hypothesis that the mean FIB values were not different across groups defined by CRP levels, in favor of an alternative hypothesis that mean FIB level were associated with CRP levels.

Estimates of (arithmetic) mean of FIB levels are provided in Table 2 following question 6.

4. 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.) (25 points given)

**Methods:**

An ordinary least squares regression that does not presume homoscedasticity, i.e. using robust standard error estimate, was used to examine the relationship between the mean FIB (response variable) and CRP levels (predictor of interest), modeling CRP as a continuous, log transformed random variable.

**Results**

The estimated intercept from the fitted regression model is 295.5663. This value indicates that the mean FIB level for observation with the log transformed of CRP level at 0, i.e. the CRP level of 1 mg/L, is 295.5663 mg/dL.

The estimated slope from the fitted regression model is 36.8332. This value indicates that 1 unit increase in the log of CRP, i.e. about 2.7283-fold increase in CRP level, is corresponding to 36.8332 mg/dL increase in mean FIB level.

Presuming homoscedasticity, the 95% confidence interval for the estimated slope of the fitted regression model is from 35.122 to 38.544. Without such presumption, the 95% confidence interval is from 34.578 to 39.088. Both results are statistically significant at the 0.05 critical threshold with p-values less than 0.0001. As a result, we reject the null hypothesis that the mean FIB values were not different across groups defined by CRP levels, in favor of an alternative hypothesis that mean FIB level were associated with CRP levels.

Estimates of (arithmetic) mean of FIB levels are provided in Table 2 following question 6.

5. 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. (25 points given)

**Methods:**

An ordinary least squares regression that does not presume homoscedasticity, i.e. using robust standard error estimate, was used to examine the relationship between the geometric mean FIB (response variable) and CRP levels (predictor of interest), modeling CRP as a continuous, untransformed random variable. Instead of using the “identity link”, the regression used the “log link”, i.e. transforming the response variable, FIB, into a log scale.

**Results**

The estimated intercept from the fitted regression model is 5.7067. This value indicates that the geometric mean FIB level for observation with CRP level of 0 mg/L is 300.8958 mg/dL.

The estimated slope from the fitted regression model is 0.0139. This value indicates that 1mg/L increase in CRP level is corresponding to 1.0140-fold increase in the geometric mean of FIB level.

Presuming homoscedasticity, the 95% confidence interval for the estimated slope of the fitted regression model is from 0.0131 to 0.0147. Without such presumption, the 95% confidence interval is from 0.0122 to 0.0157. Both results are statistically significant at the 0.05 critical threshold with p-values less than 0.0001. As a result, we reject the null hypothesis that the geometric mean FIB values were not different across groups defined by CRP levels, in favor of an alternative hypothesis that geometric mean FIB level were associated with CRP levels.

Estimates of (geometric) mean of FIB levels are provided in Table 2 following question 6.

6. 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.) (25 points given)

**Methods:**

An ordinary least squares regression that does not presume homoscedasticity, i.e. using robust standard error estimate, was used to examine the relationship between the geometric mean FIB (response variable) and CRP levels (predictor of interest), modeling CRP as a continuous, log transformed random variable. Instead of using the “identity link”, the regression used the “log link”, i.e. transforming the response variable, FIB, into a log scale.

**Results:**

The estimated intercept from the fitted regression model is 5.6786. This value indicates that the geometric mean FIB level for observation with the log transformed of CRP level at 0, i.e. the CRP level of 1 mg/L, is 292.5357 mg/dL.

The estimated slope from the fitted regression model is 0.1054. This value indicates that 1 unit increase in the log of CRP, i.e. about 2.72-fold increase in CRP level, is corresponding to a 1.111-fold increase in the geometric mean of FIB level.

Presuming homoscedasticity, the 95% confidence interval for the estimated slope of the fitted regression model is from 0.1002 to 0.1106. Without such presumption, the 95% confidence interval is from 0.0995 to 0.1112. Both results are statistically significant at the 0.05 critical threshold with p-values less than 0.0001. As a result, we reject the null hypothesis that the geometric mean FIB values were not different across groups defined by CRP levels, in favor of an alternative hypothesis that the geometric mean FIB level were associated with CRP levels.

Estimates of (geometric) mean of FIB levels are provided in Table 2 following question 6.

**Table 2. Fitted Values for Fibrinogen from Problem 3 to Problem 6**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **CRP level** | **Problem 3: (arithmetic mean)** | **Problem 4: (arithmetic mean)** | **Problem 5: (geometric mean)** | **Problem 6: (geometric mean)** |
| **1 mg/L** | 309.2661 | 295.5663 | 305.1131 | 292.5357 |
| **2 mg/L** | 314.5169 | 321.0971 | 309.3895 | 314.7059 |
| **3 mg/L** | 319.7678 | 336.0316 | 313.7259 | 328.4456 |
| **4 mg/L** | 325.0186 | 346.6279 | 318.1230 | 338.5564 |
| **6 mg/L** | 335.5204 | 361.5624 | 327.1031 | 353.3374 |
| **8 mg/L** | 346.0221 | 372.1587 | 336.3366 | 364.2144 |
| **9 mg/L** | 351.2729 | 376.4970 | 341.0507 | 368.7637 |
| **12 mg/L** | 367.0255 | 387.0932 | 355.5930 | 380.1155 |

7. Complete the following table that makes comparisons (differences or ratios) of the fitted values for each of the models. (10 points given)

**Answer:**

Table 3 gives the comparisons (differences or ratios) of the fitted values for each of the models

**Table 3**: **Comparisons (Differences or Ratios) of the Fitted Values**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **Comparisons across CRP level** | **Problem 3: (arithmetic mean)** | **Problem 4: (arithmetic mean)** | **Problem 5: (geometric mean)** | **Problem 6: (geometric mean)** |
| ***Differences*** | | | | |
| **2 mg/L – 1 mg/L** | 5.2509 | 25.5308 | 4.2764 | 22.1702 |
| **3 mg/L – 2 mg/L** | 5.2509 | 14.9346 | 4.3364 | 13.7397 |
| **4 mg/L – 1 mg/L** | 15.7526 | 51.0616 | 13.0099 | 46.0207 |
| **4 mg/L – 2 mg/L** | 10.5017 | 25.5308 | 8.7335 | 23.8504 |
| **6 mg/L – 3 mg/L** | 15.7526 | 25.5308 | 13.3772 | 24.8917 |
| **8 mg/L – 4 mg/L** | 21.0034 | 25.5308 | 18.2136 | 25.6580 |
| **9 mg/L – 6 mg/L** | 15.7526 | 14.9346 | 13.9476 | 15.4263 |
| **9 mg/L – 8 mg/L** | 5.2509 | 4.3383 | 4.7141 | 4.5493 |
| **12 mg/L – 6 mg/L** | 31.5051 | 25.5308 | 28.4899 | 26.7782 |
| ***Ratios*** | | | | |
| **2 mg/L / 1 mg/L** | 1.0170 | 1.0864 | 1.0140 | 1.0758 |
| **3 mg/L / 2 mg/L** | 1.0167 | 1.0465 | 1.0140 | 1.0437 |
| **4 mg/L / 1 mg/L** | 1.0509 | 1.1728 | 1.0426 | 1.1573 |
| **4 mg/L / 2 mg/L** | 1.0334 | 1.0795 | 1.0282 | 1.0758 |
| **6 mg/L / 3 mg/L** | 1.0493 | 1.0760 | 1.0426 | 1.0758 |
| **8 mg/L / 4 mg/L** | 1.0646 | 1.0737 | 1.0573 | 1.0758 |
| **9 mg/L / 6 mg/L** | 1.0469 | 1.0413 | 1.0426 | 1.0437 |
| **9 mg/L / 8 mg/L** | 1.0152 | 1.0117 | 1.0140 | 1.0125 |
| **12 mg/L / 6 mg/L** | 1.0939 | 1.0706 | 1.0871 | 1.0758 |

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

a. 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. (5 points given)

**Answer:**

The analysis in problem 3 gave constant differences in the fitted values when comparing two groups that differed by an absolute increase of c units in CRP levels. In this model, an absolute increase of c units in CRP levels corresponded to 5.2509c absolute increase in mean FIB.

Evidence is shown in Table 3:

- CRP levels of 2 mg/L vs. 1 mg/L, 3 mg/L vs. 2 mg/L, and 9 mg/L vs. 8 mg/L have a difference of 1 mg/L in CRP. These differences correspond to a difference of 5.2509 mg/dL in fitted FIB levels.

- CRP levels of 4 mg/L vs. 1 mg/L, 6 mg/L vs. 3 mg/L, and 9 mg/L vs. 6 mg/L have a difference of 3 mg/L in CRP. These differences correspond to the differences of 15.7526 mg/dL in fitted FIB levels

b. 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. (5 points given)

**Answer:**

The analysis in problem 5 gave constant ratios in fitted values when comparing two groups that differed by an absolute increase in c units in CRP levels. In this model, an absolute increase of c units in CRP levels corresponded to e0.0139c – fold increase in the geometric mean FIB.

Evidence is shown in Table 3:

- CRP levels of 2 mg/L vs. 1 mg/L, 3 mg/L vs. 2 mg/L, and 9 mg/L vs. 8 mg/L have a difference of 1 mg/L in CRP. These differences correspond to 1.0140-fold increase in fitted FIB levels.

- CRP levels of 4 mg/L vs. 1 mg/L, 6 mg/L vs. 3 mg/L, and 9 mg/L vs. 6 mg/L have a difference of 3 mg/L in CRP. These differences correspond to 1.0426-fold increase in fitted FIB levels.

c. 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. (5 points given)

**Answer:**

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. In this model, a c-fold increase in CRP levels corresponded to 36.8332log(c) of absolute increase in mean FIB.

Evidence is shown in Table 3:

- CRP levels of 2 mg/L vs. 1 mg/L, 4 mg/L vs. 2 mg/L, 6 mg/L vs. 3mg/L, 8 mg/L vs. 4mg/L, and 12 mg/L vs. 6 mg/L have a relative ratio of 2-fold. This ratio corresponded to a difference a 25.5308 mg/dL in fitted FIB values.

- CRP levels of 3 mg/L vs. 2 mg/L, and 9 mg/L vs. 6 mg/L have a relative ratio of 1.5-fold. This ratio corresponded to a difference of 14.9346 mg/dL in mean fitted FIB values.

d. 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. (5 points given)

**Answer:**

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. In this model, a c-fold increase in CRP levels corresponded to c0.1054-fold increase in geometric mean FIB.

Evidence is shown in Table 3:

- CRP levels of 2 mg/L vs. 1 mg/L, 4 mg/L vs. 2 mg/L, 6 mg/L vs. 3mg/L, 8 mg/L vs. 4mg/L, and 12 mg/L vs. 6 mg/L have a relative ratio of 2-fold. This ratio corresponded to a 1.0758-fold increase in fitted FIB values.

- CRP levels of 3 mg/L vs. 2 mg/L, and 9 mg/L vs. 6 mg/L have a relative ratio of 1.5-fold. This ratio corresponded to a difference of 1.0437-fold increase in fitted FIB values.

9. How would you decide which of the four potential analyses should be used to investigate associations between fibrinogen and CRP? (5 points given)

**Answer:**

It is typical to use log transformation in analyzing biological data, as we are usually interested in the multiplicative effects in biological process. It is also scientifically reasonable to suspect certain increase in CRP has some multiplicative impact, instead of absolute effect, on FIB level.

In addition, since the CRP measurements were taken from all older people, we expected a much larger range of CRP levels. Indeed, about 25% of the patients in the sample had measurements of CRP exceeding 3 mg/L, and the highest measurement of CRP was 108 mg/L. If one believes that larger mean tends to be associated with higher variance within the data, the use of log transformation is appropriate to reduce heteroscedasticity and improve the statistical precision.

As a result, I would choose the use the analysis in problem 6, i.e. log transformations on both the response variable and the predictor of interest.