2809

Biostats HW02

20 Jan 15

Total grade: 149/195

1. Methods: Descriptive statistics included a scatter plot (and Lowess curve) between blood fibrinogen (FIB) and C reactive protein (CRP) levels by CVD history. Additionally, descriptive statistics of FIB levels (including mean, standard deviation, minimum and maximum) are presented by categorized CRP levels and CVD history. CRP was defined as less than 1 mg/L (low risk), 1-3 mg/L inclusive (average risk), and greater than 3 mg/L (high risk).

Results: The original dataset included 5000 study subjects. Measurements of blood FIB and CRP were missing for 67 and 85 subjects respectively. Fifty one subjects were missing both FIB and CRP measurements. Observations with missing FIB or CRP (101 participants) were omitted from study analysis, resulting in a total 4899 subjects for the analysis. No patients were missing information of CVD history.

The figure below demonstrates a general positive association between FIB and CRP blood levels. However much of the data is group in the lower measurements of both FIB and CRP. This makes it difficult to tease out a difference between those with CVD. It appears that more subjects with higher levels of CRP and FIB have a history of CVD. This observation is confirmed when looking at the data in the table. As CRP increases in risk level the mean FIB level also increases. Additionally the mean FIB level is consistently higher among subjects with prior history of CVD than those without prior history.





2. a. Methods: This analysis compares the mean levels of FIB between subjects with a history of CVD and those without CVD history. The difference of means were tested using a t test that did assumed equal variances in the two groups. The 95% confidence interval was calculated using methods with the same assumption about variances.

Results: The mean blood FIB level for 3777 subjects without previous history of CVD was 319.6 mg/dL and 334.4 for 1122 subjects with previous CVD. Based on a 95% confidence level we observed subjects without CVD history to have a mean FIB 14.8 mg/dL lower than those with CVD history. Our data would not be unusual if the true mean FIB was between 10.4 and 19.3 mg/dL lower among those without CVD history compared to those with CVD history. These results of a t test assuming equal variance (two sided pvalue< 0.001) are statistically significant at alpha level of 0.05. We can therefor reject the null hypothesis that the mean blood FIB levels are no different between those with CVD history and those without CVD history.

B. Linear regression assuming equal variances (without the robust option) performs similarly to a ttest when you have one binary variable. Results from a simple linear regression model using CVD history as a binary predictor variable and FIB as the outcome will give you the same (or very similar) results as the ttest. The intercept from the regression will be the same as the mean FIB level among subjects without CVD history (319.6 mg/dL). The confidence interval for the intercept then also corresponds to the confidence interval for the estimate of mean FIB among those without CVD history (317.5, 321.7 mg/dL). The regression measure for beta for the X(CVD history), or the slope for CVD history corresponds to the difference between means of the two groups from the ttest (14.8 mg/L lower in the group with no CVD history). (Although it will likely be -1 times the difference if you don’t change the coding because the ttest compares group 0 to 1 and regression works the other way around comparing 1 to 0.) The interpretation of the beta is the change in the outcome (FIB) resulting in a one unit increase in the variable X (CVD). When X is binary as the case here there is only one possible 1 unit increase and corresponds to the difference between those with CVD history to those without.

The confidence interval and pvalue for the beta also corresponds to the CI and pvalue for the difference in mean FIB between those with CVD history and those without CVD history. This is demonstrated below where boxes are placed around the mentioned numbers. Color and line pattern match the corresponding output for the two different statistical tests, t test and linear regression.

Lastly you could get the mean FIB among those with CVD history by taking the intercept and adding the estimate for beta (319.6+14.8=334.4 mg/dL).





c. Methods: This analysis compares the mean levels of FIB between subjects with a history of CVD and those without CVD history. The difference of means were tested using a t test that allowed for unequal variances in the two groups using Satterwaite approximation. The 95% confidence interval was calculated using methods with the same assumption about unequal variances.

Results: The mean blood FIB level for 3777 subjects without previous history of CVD was 319.6 mg/dL and 334.5 for 1122 subjects with previous CVD. Based on a 95% confidence level we observed subjects without CVD history to have a mean FIB 14.8 mg/dL lower than those with CVD history. Our data would not be unusual if the true mean FIB was between 10.4 and 19.3 mg/dL lower among those without CVD history compared to those with CVD history. These results of a t test allowing for unequal variance (two sided pvalue< 0.001) are statistically significant at alpha level of 0.05. We can therefor reject the null hypothesis that the mean blood FIB levels are no different between those with CVD history and those without CVD history. We can conclude prior CVD history is associated with higher mean FIB level.

d. Linear regression allowing for unequal variances using a Huber-White sandwich estimator for the standard errors performs similarly to a ttest that allows for unequal variances. Results from a simple linear regression model using CVD history as a binary predictor variable and FIB as the outcome will give you the same (or very similar) results as the ttest. The intercept from the regression will be the same as the mean FIB level among subjects without CVD history (319.6 mg/dL). The confidence interval for the intercept then also roughly corresponds to the confidence interval for the estimate of mean FIB among those without CVD history (317.5, 321.7 mg/dL). The regression measure for beta for the X(CVD history), or the slope for CVD history corresponds to the difference between means of the two groups from the ttest (14.8 mg/L lower in the group with no CVD history). (Although it will likely be -1 times the difference if you don’t change the coding because the ttest compares group 0 to 1 and regression works the other way around comparing 1 to 0.) The interpretation of the beta is the change in the outcome (FIB) resulting in a one unit increase in the variable X (CVD). When X is binary as the case here there is only one possible 1 unit increase and corresponds to the difference between those with CVD history to those without. The confidence interval and pvalue for the beta also corresponds to the CI and pvalue for the difference in mean FIB between those with CVD history and those without CVD history. This is demonstrated below where boxes are placed around the mentioned numbers. Color and line pattern match the corresponding output for the two different statistical tests, t test and linear regression.

Lastly you could get the mean FIB among those with CVD history by taking the intercept and adding the estimate for beta (319.6+14.8=334.6 mg/dL).





e. You could look at the standard of deviation output from part a to predict if the analysis in part c would find a stronger or weaker association. You would see that the standard of deviation is close to the same in the two groups (64.9 without prior CVD; and 74.2 with prior CVD). You would expect the association to be close to the same, but maybe expect a weaker association. I conclude this because the group with the slightly higher SD (group with prior CVD) also is the group with the smaller sample size. This would lead us to hypothesize that using the test that allows for unequal variances could lead to a slightly larger p value and wider CI. However, again the results from part A show very similar SD between the two groups and so you wouldn’t expect much of a difference between the two tests that assume and do not assume equal variances.

3. Method: A linear regression assesses an association between mean FIB across groups defined by CRP. CRP is a continuous and untransformed random variable.

a. The intercept from linear regression analysis using Huber-white estimates for standard error indicates the estimated mean value (304.0 mg/dL) of blood FIB when CRP levels equal to 0.

b. Linear regression analysis using Huber-White estimates for standard error estimates the difference in mean FIB level across groups differing by 1 mg/L of CRP. In other words 1 mg/L increase in CRP is associated with 5.25 mg/dL increase in blood FIB.

c. From linear regression analysis using Huber-White estimates for standard error, we estimate the difference in mean FIB levels (mg/dL) for two groups differing by 1 mg/L of CRP to be 5.35 mg/dL higher in the group with higher CRP levels. Using 95% confidence intervals, our data would not be unusual if the true difference in mean FIB were between 4.60 and 5.90 mg/dL higher per 1 mg/L higher level of blood CRP. Using alpha 0.05 our results from a linear regression allowing for unequal variances (two sided p value < 0.001), we reject the null hypothesis that there is no linear trend on the mean FIB across groups of blood CRP. This supports the conclusion that an increase in mean CRP is associated with an increase in mean FIB levels.

**4.** Method: A linear regression assesses an association between mean FIB across groups defined by CRP. CRP is a continuous and log transformed random variable.

a. The intercept from linear regression analysis using Huber-white estimates for standard error indicates the estimated mean value (296.0 mg/dL) of blood FIB when logCRP levels equal to 0, thus when CRP equals 1.

b. Linear regression analysis using Huber-White estimates for standard error estimates the difference in mean FIB level (mg/dL) across groups differing by 10% increase in CRP. In other words a group with 10% higher CRP levels is estimated to have mean blood FIB by 3.70 mg/dL higher than the lower CRP level group. 1 unit increase in log CRP

c. From a linear regression analysis using Huber-White estimates for standard error, we estimate the difference in mean FIB levels (mg/dL) associated with a 10% increase in CRP to be 3.70 mg/dL higher in the group with higher CRP levels. Using 95% confidence intervals, our data would not be unusual if the true difference in mean FIB were between 3.45 and 3.95 mg/dL higher per 10% increase in of blood CRP. Using alpha 0.05 our results from a linear regression allowing for unequal variances (two sided p value < 0.001), we reject the null hypothesis that there is no linear trend on the mean FIB across groups of blood CRP. This supports the conclusion that an increase in CRP (increase in logCRP) is associated with increase in mean FIB on a multiplicative scale.

5. Method: A linear regression assesses an association between the geometric mean FIB across groups defined by CRP. CRP is a continuous and untransformed random variable.

a. The intercept from linear regression analysis using Huber-white estimates for standard error indicates the estimated geometric mean value (301 mg/dL) of blood FIB when CRP levels equal to 0.

b. Linear regression analysis using Huber-White estimates for standard error estimates the ratio between geometric mean FIB levels across groups differing by 1 mg/L of CRP (1.014).

c. From linear regression analysis using Huber-White estimates for standard error, we estimate the ratio in geometric mean FIB levels for two groups differing by 1 mg/L of CRP to be 1.40% higher in the group with higher CRP levels. Using 95% confidence intervals, our data would not be unusual if the true ratio of geometric mean FIB were between 1.22% and 1.58% higher per 1 mg/L higher level of blood CRP. Using alpha 0.05 our results from a linear regression allowing for unequal variances (two sided p value < 0.001), we reject the null hypothesis that there is no linear trend on the geometric mean FIB across groups of blood CRP. This supports the conclusion that increase in CRP is associated with the increase in geometric mean FIB.

6. Methods: A linear regression assesses an association between the geometric mean FIB across groups defined by CRP. CRP is a continuous and log transformed random variable.

Results: a. The intercept from linear regression analysis using Huber-white estimates for standard error indicates the estimated geometric mean value (291.4 mg/dL) of blood FIB when logCRP levels equal to 0, so CRP is equal to 1.

b. Linear regression analysis using Huber-White estimates for standard error estimates the ratio between geometric mean FIB levels across groups differing by 10% increase in CRP. The group with 10% higher in CRP level is estimated to have a geometric mean blood FIB that is 1.04% higher.

c. From linear regression analysis using Huber-White estimates for standard error, we estimate the ratio in geometric mean FIB levels for two groups differing by 10% increase in CRP to be 1.04% higher in the group with higher CRP levels. Using 95% confidence intervals, our data would not be unusual if the true ratio of geometric mean FIB were between 0.972% and 1.10% in a group with 10% higher CRP level. Using alpha 0.05 our results from a linear regression allowing for unequal variances (two sided p value < 0.001), we reject the null hypothesis that there is no linear trend on the geometric mean FIB across groups of blood CRP. This supports the conclusion that the increase in CRP (log CRP) on the multiplicative scale is association with the increase in geometric mean FIB.

**Table 1**: Estimates of central tendency for Fibrinogen for four models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 mg/dL | 296 mg/dL | 305 mg/dL | 292.5 mg/dL |
| **2 mg/L** | 315 mg/dL | 307 mg/dL | 309 mg/dL | 302.0 mg/dL |
| **3 mg/L** | 320 mg/dL | 313 mg/dL | 314 mg/dL | 307.6 mg/dL |
| **4 mg/L** | 325 mg/dL | 318 mg/dL | 318 mg/dL | 311.7 mg/dL |
| **6 mg/L** | 336 mg/dL | 324 mg/dL | 327 mg/dL | 317.5 mg/dL |
| **8 mg/L** | 346 mg/dL | 329 mg/dL | 336 mg/dL | 321.7 mg/dL |
| **9 mg/L** | 351 mg/dL | 331 mg/dL | 341 mg/dL | 323.5 mg/dL |
| **12 mg/L** | 367 mg/dL | 335 mg/dL | 356 mg/dL | 327.8 mg/dL |

7. **Table 2**: Comparisons of fitted values across 4 different models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **Comparisons across CRP level** | **Problem 3: Difference in means** | **Problem 4: Mean** | **Problem 5: Geometric Mean** | **Problem 6: Geometric Mean** |
| **Differences** |  |  |  |  |
| **2 mg/L - 1 mg/L** | 5.25 | 11.09 | 4.28 | 9.43 |
| **3 mg/L - 2 mg/L** | 5.25 | 6.49 | 4.34 | 5.66 |
| **4 mg/L - 1 mg/L** | 15.80 | 22.18 | 13.01 | 19.16 |
| **4 mg/L - 2 mg/L** | 10.50 | 11.09 | 8.73 | 9.73 |
| **6 mg/L - 3 mg/L** | 15.80 | 11.09 | 13.38 | 9.92 |
| **8 mg/L - 4 mg/L** | 21.00 | 11.09 | 18.21 | 10.05 |
| **9 mg/L - 6 mg/L** | 15.80 | 6.49 | 13.95 | 5.95 |
| **9 mg/L - 8 mg/L** | 5.25 | 1.88 | 4.71 | 1.74 |
| **12 mg/L - 6 mg/L** | 31.50 | 11.09 | 28.49 | 10.24 |
| **Ratios** |  |  |  |  |
| **2 mg/L / 1 mg/L** | 1.017 | 1.038 | 1.014 | 1.032 |
| **3 mg/L / 2 mg/L** | 1.017 | 1.021 | 1.014 | 1.019 |
| **4 mg/L / 1 mg/L** | 1.051 | 1.075 | 1.043 | 1.066 |
| **4 mg/L / 2 mg/L** | 1.033 | 1.036 | 1.028 | 1.032 |
| **6 mg/L / 3 mg/L** | 1.049 | 1.035 | 1.043 | 1.032 |
| **8 mg/L / 4 mg/L** | 1.065 | 1.035 | 1.057 | 1.032 |
| **9 mg/L / 6 mg/L** | 1.047 | 1.020 | 1.043 | 1.018 |
| **9 mg/L / 8 mg/L** | 1.015 | 1.006 | 1.014 | 1.0054 |
| **12 mg/L / 6 mg/L** | 1.094 | 1.034 | 1.087 | 1.032 |

8. a. The analysis where the predictor variables CRP and FIB were not log transformed, model 3, is the analysis with constant differences in fitted values when comparing across CRP groups differing by an absolute increase in c units.

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Comparisons across CRP level** | **Problem 3: Difference in means** |
| **1 unit increase** |  |
| **2 mg/L - 1 mg/L** | 5.25 |
| **3 mg/L - 2 mg/L** | 5.25 |
| **9 mg/L - 8 mg/L** | 5.25 |
| **3 unit increase** |  |
| **4 mg/L - 1 mg/L** | 15.80 |
| **6 mg/L - 3 mg/L** | 15.80 |
| **9 mg/L - 6 mg/L** | 15.80 |

b. The analysis with untransformed CRP predictor estimating geometric mean FIB (problem 5) gave constant ratios of estimated values when comparing across levels of CRP that differed by an absolute increase in c units.

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Comparisons across CRP level**  **Ratios** | **Problem 5: Ratios of geometric means** |
| **1 unit increase** |  |
| **2 mg/L / 1 mg/L** | 1.014 |
| **3 mg/L / 2 mg/L** | 1.014 |
| **9 mg/L / 8 mg/L** | 1.014 |
| **3 unit increase** |  |
| **4 mg/L / 1 mg/L** | 1.043 |
| **6 mg/L / 3 mg/L** | 1.043 |
| **9 mg/L / 6 mg/L** | 1.043 |

c. The analyses with log transformed CRP predictor estimating the mean of the outcome FIB (problem 4) gave constant differences in estimated values when comparing across levels of CRP That differed by a relative c-fold increase.

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Comparisons across CRP level** | **Problem 4: Differences in Mean** | |
| **2 fold increase** |  | |
| **2 mg/L - 1 mg/L** | 11.09 | |
| **4 mg/L - 2 mg/L** | 11.09 | |
| **6 mg/L - 3 mg/L** | 11.09 | |
| **8 mg/L - 4 mg/L** | 11.09 | |
| **12 mg/L - 6 mg/L** | 11.09 | |
| **1.5 fold increase** |  | |
| **9 mg/L - 6 mg/L** | 6.49 | |
| **3 mg/L - 2 mg/L** | 6.49 | |

d. The analysis with log transformed CRP levels estimating the geometric mean FIB levels (problem 6) gave constant ratios when comparing levels of CRP differing by a relative c-fold increase.

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Comparisons across CRP level** | **Problem 6: Ratio of Geometric mean** |
| **2 fold increase** |  |
| **2 mg/L / 1 mg/L** | 1.032 |
| **4 mg/L / 2 mg/L** | 1.032 |
| **6 mg/L / 3 mg/L** | 1.032 |
| **8 mg/L / 4 mg/L** | 1.032 |
| **12 mg/L / 6 mg/L** | 1.032 |
| **1.5 fold increase** |  |
| **3 mg/L / 2 mg/L** | 1.019 |
| **9 mg/L / 6 mg/L** | 1.019 |

9. To decide which of the four potential analyses to investigate the association between fibrinogen and CRP I would think about the overall goal of the analysis and the specific aim of the analysis between FIB, CRP and cardiac death. Additionally you would want to think about the scale that makes since with these blood biomarker levels. After looking at the data you can see both CRP and FIB seem to make more since on the multiplicative scale, as do many blood biomarkers. Both measures have centered around low levels with several scattered higher levels (this is easily to see in the scatter plot). It might be that these high levels are not outliers but simply high levels. On the multiplicative scale these would not be quite as extreme. However you would want to make this judgment call with previous knowledge before seeing your data.