**Question 1**

*Methods*:

The data set was broken up into groups that had a history of cardiovascular disease and those who had not. Then the summary statistics for CRP and FIB were calculated based on these groupings.

*Inference*:

The following table shows the association between the two biomarkers C-reactive protein (CRP) and fibrinogen (FIB) based on the prevalence of prior cardiovascular disease (CVD).

|  |  |  |  |
| --- | --- | --- | --- |
| **Biomarkers** | **No Previous CVD**  *(N = 3851)* | **Previous CVD**  *(N = 1149)* | **All**  *(N = 5000)* |
| CRP 1  *(mg/l)* | 3.38 (5.899)  NA = 49 | 4.39 (6.881)  NA = 18 | 3.61 (6.152)  NA = 67 |
| FIB 1  *(mg/dl)* | 319.6 (64.764)  NA = 60 | 334.5 (74.062)  NA = 25 | 323 (67.287)  NA = 85 |

Note: 1 mean (SD)

**Question 2a**

*Methods*:

A t test was performed comparing the mean FIB level between those with and without a history of CVD. The t test assumed equal variances.

*Inference*:

The mean FIB for those with a history of CVD was 334.46 mg/dl, with 1124 observations; while the mean FIB for those without a history was 319.57 mg/dl with 3791 observations. Based on the t test, we can reject the null hypothesis and conclude that with *p* < 0.001 that there is a difference between the mean FIB levels based on history of CVD (*t* = 6.5412, *df* = 4913).

**Question 2b**

The same analysis as a t test with equal variances could be performed with a linear regression. However, the root mean square error (RMSE) would be used to calculate the confidence intervals. The result would lead to estimates that were slightly different.

**Question 2c**

*Methods*:

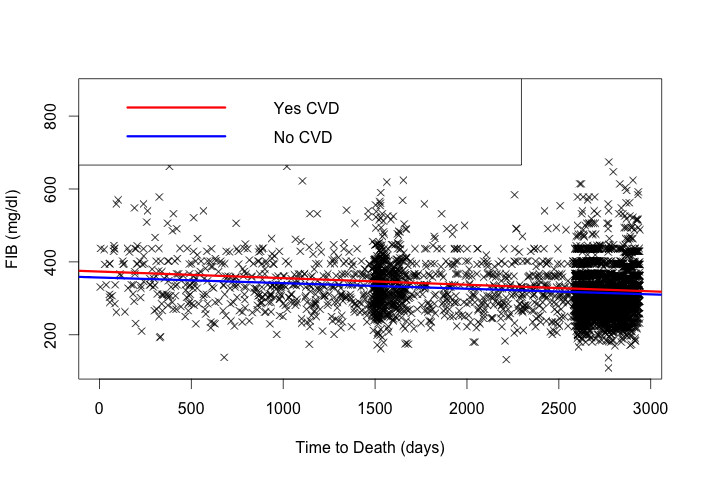
Similar to Question 2a, a t test was performed between mean FIB levels based on history of CVD. However this time, the t test assumed unequal variances.

*Inference*:

The mean FIB for those with a history of CVD was 334.46 mg/dl, with 1124 observations; while the mean FIB for those without a history was 319.57 mg/dl with 3791 observations. Based on the t test, we can reject the null hypothesis and conclude that with *p* < 0.001 that there is a difference between the mean FIB levels based on history of CVD (*t* = 6.0836, *df* = 1664).

**Question 2d**

The same analysis could be performed using linear regression, as the standard errors will lead to the same point estimates as a t test with unequal variances. This could be done by fitting a simple linear regression with time to death as the dependent variable and FIB as the independent variable. A separate regression would be used for the subjects with CVD and without CVD. Time to death should be used because this analysis is interested in mortality and FIB levels. For this analysis, the intercepts and slopes of the two regressions (history of CVD and no history of CVD) could be compared to explore the association between mean FIB and CVD history. A plot has been provided below to show the two fitted regressions between FIB and mortality based on CVD history.



**Question 2e**

In part a under the assumption of equal variances, the standard error will be smaller than the standard error in part c with unequal variances. The degrees of freedom were larger in part a, which led to a smaller t statistic and p value. Therefore the results from part a could be used to predict that the analysis in part c would have a weaker association – with wider confidence intervals and a larger p value.

**Question 3**

*Methods*:

Two simple linear regression models were fitted: one for those with a history of CVD and the other for those without CVD. The models used CRP as the independent variable and FIB as the dependent variable. These two models were used the following parts a and b for Question 3. For part c, an ANOVA was performed between two regression models: one that accounted for FIB by CRP and another that accounted for FIB by CRP and CVD.

**Question 3a**

The intercept for those with a history of CVD was 309.665 mg/dl, while the intercept for those without CVD was lower at 302.574 mg/dl.

**Question 3b**

The slope for those with a history of CVD was 5.629, while the slope for those without CVD was less, at 5.034.

**Question 3c**

Based on an ANOVA with 4896 degrees of freedom, the prevalence of previous cardiovascular disease is relevant on FIB levels when also including CRP (*p* < 0.001). For those with a previous history of CVD, it is estimated that they have a difference in mean FIB of 5.629 mg/dl for each increase of mg/l in CRP. For those without a previous history of CVD, it is estimated that they have a difference in mean FIB of 5.034 mg/dl for each increase of mg/l in CRP.

**Question 3d**

The following table provides estimates for mean fitted FIB levels based on CRP level for each those with and without CVD history as well as for all combined.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Fitted Mean Values for Fibrinogen (mg/dL)* | | |
| *CRP Level* | **No CVD History** | **Yes CVD History** | **Both** |
| **1 mg/L** | 307.6089 | 315.2942 | 309.2661 |
| **2 mg/L** | 312.6434 | 320.9233 | 314.5170 |
| **3 mg/L** | 317.6779 | 326.5524 | 319.7679 |
| **4 mg/L** | 322.7124 | 332.1815 | 325.0188 |
| **6 mg/L** | 332.7814 | 343.4397 | 335.5206 |
| **8 mg/L** | 342.8504 | 354.6979 | 346.0224 |
| **9 mg/L** | 347.8849 | 360.3270 | 351.2733 |
| **12 mg/L** | 362.9884 | 377.2143 | 367.0260 |

**Question 4**

*Methods*:

All of the 0 values for CRP were replaced with 0.5 and then CRP was log transformed. Then the process was similar to question 3, where two simple linear regression models were fitted broken down by CVD history. The models used CRP as the independent variable and FIB as the dependent variable. These two models were used the following parts a and b. For part c, an ANOVA was performed between two regression models: one that accounted for FIB by CRP and another that accounted for FIB by CRP and CVD.

**Question 4a**

The intercept for those with a history of CVD was 298.039 mg/dl, while the intercept for those without CVD was lower at 295.013 mg/dl.

**Question 4b**

The slope for those with a history of CVD was 39.825, while the slope for those without CVD was less, at 35.402.

**Question 4c**

Based on an ANOVA with 4896 degrees of freedom, the prevalence of previous cardiovascular disease is relevant on FIB levels when also including CRP (*p* = 0.0005). For those with a previous history of CVD, it is estimated that they have a difference in mean FIB of 39.825 mg/dl for each increase of log(mg/l) in CRP. For those without a previous history of CVD, it is estimated that they have a difference in mean FIB of 35.42 mg/dl for each increase of log(mg/l) in CRP.

**Question 4d**

The following table provides estimates for mean fitted FIB levels based on CRP level for each those with and without CVD history as well as for all combined, using the log transformed values of CRP.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Fitted Mean Values for Fibrinogen (mg/dL)* | | |
| *CRP Level* | **No CVD History** | **Yes CVD History** | **Both** |
| **1 mg/L** | 295.0130 | 298.0390 | 295.5663 |
| **2 mg/L** | 305.6701 | 310.0275 | 306.6542 |
| **3 mg/L** | 311.9040 | 317.0404 | 313.1402 |
| **4 mg/L** | 316.3271 | 322.0160 | 317.7421 |
| **6 mg/L** | 322.5611 | 329.0289 | 324.2281 |
| **8 mg/L** | 326.9842 | 334.0046 | 328.8300 |
| **9 mg/L** | 328.7951 | 336.0417 | 330.7141 |
| **12 mg/L** | 333.2182 | 341.0174 | 335.3160 |

**Question 5**

*Methods*:

The FIB values were first log transformed. Then the process was similar to question 3 and 4, where two simple linear regression models were fitted broken down by CVD history. The models used CRP as the independent variable and FIB as the dependent variable. These two models were used the following parts a and b. For part c, an ANOVA was performed between two regression models: one that accounted for FIB by CRP and another that accounted for FIB by CRP and CVD.

**Question 5a**

The intercept for those with a history of CVD was 5.725, while the intercept for those without CVD was lower at 5.702.

**Question 5b**

The slope for those with a history of CVD was 0.0146, while the slope for those without CVD was less, at 0.013.

**Question 5c**

Based on an ANOVA with 4896 degrees of freedom, the prevalence of previous cardiovascular disease is relevant on FIB levels when also including CRP (*p* < 0.001). For those with a previous history of CVD, it is estimated that they have a difference in mean FIB of exp(5.725 mg/dl) for each increase of mg/l in CRP. For those without a previous history of CVD, it is estimated that they have a difference in mean FIB of exp(5.702 mg/dl) for each increase of mg/l in CRP.

**Question 5d**

The following table provides estimates for mean fitted FIB levels based on CRP level for each those with and without CVD history as well as for all combined, using the log transformed values of CRP.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Fitted Mean Values for Fibrinogen (mg/dL)* | | |
| *CRP Level* | **No CVD History** | **Yes CVD History** | **Both** |
| **1 mg/L** | 303.3842 | 310.9400 | 305.0936 |
| **2 mg/L** | 307.3539 | 315.5130 | 309.3698 |
| **3 mg/L** | 311.3756 | 320.1533 | 313.7059 |
| **4 mg/L** | 315.4499 | 324.8618 | 318.1027 |
| **6 mg/L** | 323.7592 | 334.4877 | 327.0822 |
| **8 mg/L** | 332.2873 | 344.3987 | 336.3152 |
| **9 mg/L** | 336.6353 | 349.4638 | 341.0290 |
| **12 mg/L** | 350.0234 | 365.1105 | 355.5704 |

**Question 6**

*Methods*:

First all CRP values of 0 were replaced with 0.5 and then all FIB and CRP values were log transformed. Then all CRP values of 0 were replaced with 0.5 and also log transformed. Then the process was similar to question 3, 4, and 5, where two simple linear regression models were fitted broken down by CVD history. The models used CRP as the independent variable and FIB as the dependent variable. These two models were used the following parts a and b. For part c, an ANOVA was performed between two regression models: one that accounted for FIB by CRP and another that accounted for FIB by CRP and CVD.

**Question 6a**

The intercept for those with a history of CVD was 5.688, while the intercept for those without CVD was lower at 5.676.

**Question 6b**

The slope for those with a history of CVD was 0.110, while the slope for those without CVD was less, at 0.103.

**Question 6c**

Based on an ANOVA with 4896 degrees of freedom, the prevalence of previous cardiovascular disease is relevant on FIB levels when also including CRP (*p* = 0.0018). For those with a previous history of CVD, it is estimated that they have a difference in mean FIB of exp(0.110 mg/dl) for each increase of log(mg/l) in CRP. For those without a previous history of CVD, it is estimated that they have a difference in mean FIB of exp(0.103 mg/dl) for each increase of log(mg/l) in CRP.

**Question 6d**

The following table provides estimates for mean fitted FIB levels based on CRP level for each those with and without CVD history as well as for all combined.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Fitted Mean Values for Fibrinogen (mg/dL)* | | |
| *CRP Level* | **No CVD History** | **Yes CVD History** | **Both** |
| **1 mg/L** | 291.7800 | 295.3024 | 292.5337 |
| **2 mg/L** | 300.9686 | 305.2446 | 301.9635 |
| **3 mg/L** | 306.4772 | 311.2148 | 307.6198 |
| **4 mg/L** | 310.4467 | 315.5214 | 311.6972 |
| **6 mg/L** | 316.1288 | 321.6927 | 317.5359 |
| **8 mg/L** | 320.2232 | 326.1443 | 321.7447 |
| **9 mg/L** | 321.9148 | 327.9846 | 323.4840 |
| **12 mg/L** | 326.0842 | 332.5233 | 327.7716 |

**Question 7**

*Methods*:

The results from 3 d, 4 d, 5 d, and 6 d were used to calculate the comparisons by CVD history group for each cell.

*Inference*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **Comparisons across CRP level** | **Problem 3: (mean No CVD)**  **(mean Yes CVD)**  **(mean All)** | **Problem 4: (mean No CVD)**  **(mean Yes CVD)**  **(mean All)** | **Problem 5: (mean No CVD)**  **(mean Yes CVD)**  **(mean All)** | **Problem 6: (mean No CVD)**  **(mean Yes CVD)**  **(mean All)** |
| ***Differences*** | | | | |
| **2 mg/L – 1 mg/L** | 5.0345 | 10.6571 | 3.9697 | 9.1886 |
| 5.6291 | 11.9885 | 4.5730 | 9.9422 |
| 5.2509 | 11.0879 | 4.2762 | 9.4298 |
| **3 mg/L – 2 mg/L** | 5.0345 | 6.2339 | 4.0217 | 5.5086 |
| 5.6291 | 7.0129 | 4.6403 | 5.9702 |
| 5.2509 | 6.4860 | 4.3361 | 5.6563 |
| **4 mg/L – 1 mg/L** | 15.1035 | 21.3141 | 12.0657 | 18.6667 |
| 16.8873 | 23.9770 | 13.9218 | 20.219 |
| 15.7527 | 22.1758 | 13.0091 | 19.1635 |
| **4 mg/L – 2 mg/L** | 10.0690 | 10.6570 | 8.0960 | 9.4781 |
| 11.2582 | 11.9885 | 9.3488 | 10.2768 |
| 10.5018 | 11.0879 | 8.7329 | 9.7337 |
| **6 mg/L – 3 mg/L** | 15.1035 | 10.6571 | 12.3836 | 9.6516 |
| 16.8873 | 11.9885 | 14.3344 | 10.4779 |
| 15.7527 | 11.0879 | 13.3763 | 9.9161 |
| **8 mg/L – 4 mg/L** | 20.1380 | 10.6571 | 16.8374 | 9.7765 |
| 22.5164 | 11.9886 | 19.5369 | 10.6229 |
| 21.0036 | 11.0879 | 18.2125 | 10.0475 |
| **9 mg/L – 6 mg/L** | 15.1035 | 6.2340 | 12.8761 | 5.7860 |
| 16.8873 | 7.0128 | 14.9761 | 6.2919 |
| 15.7527 | 6.4860 | 13.9468 | 5.9481 |
| **9 mg/L – 8 mg/L** | 5.0345 | 1.8109 | 4.3480 | 1.6916 |
| 5.6291 | 2.0371 | 5.0651 | 1.8403 |
| 5.2509 | 1.8841 | 4.7138 | 1.7393 |
| **12 mg/L – 6 mg/L** | 30.2070 | 10.6571 | 26.2642 | 9.9554 |
| 33.7746 | 11.9885 | 30.6228 | 10.8306 |
| 31.5054 | 11.0879 | 28.4882 | 10.2357 |
| ***Ratios*** | | | | |
| **2 mg/L / 1 mg/L** | 1.0164 | 1.0361 | 1.0131 | 1.0315 |
| 1.0179 | 1.0402 | 1.0147 | 1.0337 |
| 1.0170 | 1.0375 | 1.0140 | 1.0322 |
| **3 mg/L / 2 mg/L** | 1.0161 | 1.0204 | 1.0131 | 1.0183 |
| 1.0175 | 1.0226 | 1.0147 | 1.0196 |
| 1.0167 | 1.0212 | 1.0140 | 1.0187 |
| **4 mg/L / 1 mg/L** | 1.0491 | 1.0722 | 1.0398 | 1.0640 |
| 1.0536 | 1.0804 | 1.0448 | 1.0685 |
| 1.0509 | 1.0750 | 1.0426 | 1.0655 |
| **4 mg/L / 2 mg/L** | 1.0322 | 1.0349 | 1.0263 | 1.0315 |
| 1.0351 | 1.0387 | 1.0296 | 1.0337 |
| 1.0334 | 1.0362 | 1.0282 | 1.0322 |
| **6 mg/L / 3 mg/L** | 1.0475 | 1.0342 | 1.0398 | 1.0315 |
| 1.0517 | 1.0378 | 1.0448 | 1.0337 |
| 1.0493 | 1.0354 | 1.0426 | 1.0322 |
| **8 mg/L / 4 mg/L** | 1.0624 | 1.0337 | 1.0534 | 1.0315 |
| 1.0678 | 1.0372 | 1.0601 | 1.0337 |
| 1.0646 | 1.0349 | 1.0573 | 1.0322 |
| **9 mg/L / 6 mg/L** | 1.0454 | 1.0193 | 1.0398 | 1.0183 |
| 1.0492 | 1.0213 | 1.0448 | 1.0196 |
| 1.0470 | 1.0200 | 1.0426 | 1.0187 |
| **9 mg/L / 8 mg/L** | 1.0147 | 1.0055 | 1.0131 | 1.0053 |
| 1.0159 | 1.0061 | 1.0147 | 1.0056 |
| 1.0152 | 1.0057 | 1.0140 | 1.0054 |
| **12 mg/L / 6 mg/L** | 1.0908 | 1.0330 | 1.0811 | 1.0315 |
| 1.0983 | 1.0364 | 1.0916 | 1.0337 |
| 1.0939 | 1.0342 | 1.0871 | 1.0322 |

**Question 8a**

The analysis from Problem 3, with no transformations, gave constant differences in the fitted values when comparing two groups that differed by an absolute increase in c units in CRP. This is noticeable as the difference from 2 mg/L to 1 mg/L and 9mg/L to 8 mg/L was the same as 3 mg/L to 2 mg/L for each group (5.0345, 5.6291, 5.2509 same as 5.0345, 5.6291, 5.2509).

**Question 8b**

The analysis from problem 6, with log transformations of CRP and FIB gave constant ratios of the fitted values when comparing two groups that differed by an absolute increase in c units in CRP levels. This is transparent as the jump from 2 mg/L / 1 mg/L yielded the same results as from 4 / 2 and 6 / 3 and 8 / 4 and 12 / 6 (1.0315, 1.0337, 1.0322).

**Question 8c**

The analysis from Problem 3, with no transformations, gave constant differences in the fitted values when comparing two groups that differed by a relative c-fold increase in CRP levels. This is noticeable as the jump from 4 mg/L to 1 mg/L and 6 mg/L to 3 mg/L and 9 mg/L to 6 mg/L yielded values three times as large (each 15.1035, 16.8873, 15.7527) as the values for a 1 mg/L increase (e.g. 2mg/L to 1 mg/L). Similarly from 4 mg/L to 2 mg/L yielded values two times as large (10.0690, 11.2582, 10.5018). The jump from 8 mg/L to 4 mg/L was four times as large and the 12 mg/L to 6 mg/L jump was six times as large.

**Question 8d**

The analysis from Problem 5, with FIB log transformed yielded constant ratios in the fitted values when comparing two groups that differed by a relative c-fold increase in CRP levels. This is apparent when comparing the increase from a ratio of 2 to 4 (e.g. 2 mg/L / 1 mg/L to 4 mg/L / 1 mg/L.

**Question 9**

I would use literature to understand these two variables better. For instance, if the two variables were better associated through transformation (e.g. very small incremental changes) I would choose the log transformation. However, it is important to not use this dataset to sift through the data to determine which relationship worked best.