Homework 2

1. **Methods:** We classified every individual into one of three groups based on CRP levels dictated by prior scientific knowledge; low CRP (below 1 mg/L), mid CRP (1 – 3 mg/L) and high CRP (above 3 mg/L). We then provided descriptive statistics for FIB count stratified on CRP score. We did this for all subjects combined, for only subjects who had prior CVD, and for only subjects who did not have prior CVD.

**Results:** We have provided tables below illustrating descriptive statistics (mean (std dev; min – max) for FIB stratified on CRP levels for all subjects, subjects with prior CVD, and subjects without prior CVD. As you can see, in all groups defined by prior CVD, higher levels of CRP are associated with higher levels of FIB. Also, for all groups whose CRP was high (above 3 mg/L), the variation of FIB levels is higher than in other CRP groups.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | All Subjects |  |  |
|  | Below 1 mg/L | 1 - 3 mg/L | Above 3 mg/L | All Subjects |
| N | 428 | 3330 | 1175 | 4933 |
| FIB | 280 (50.5; 172 - 540) | 311 (53.2; 109 - 592) | 373 (90.0; 132 - 872) | 323 (67.4; 109 - 872) |
|  |  |  |  |  |
|  |  | Subjects with Prior CVD |  |  |
|  | Below 1 mg/L | 1 - 3 mg/L | Above 3 mg/L | All Subjects |
| N | 78 | 715 | 338 | 1131 |
| FIB | 290 (57.9; 180 - 540) | 315 (55.6; 138 - 592) | 386 (84.5; 175 - 695) | 335 (74.1; 138 - 695) |
|  |  |  |  |  |
|  |  | Subjects W/out Prior CVD |  |  |
|  | Below 1 mg/L | 1 - 3 mg/L | Above 3 mg/L | All Subjects |
| N | 350 | 2615 | 837 | 3802 |
| FIB | 278 (48.5; 172 - 436) | 310 (52.5; 109 - 562) | 367 (78.9; 132 - 872) | 320 (64.8; 109 - 872) |

* Data presented is: mean (standard deviation; min – max)

1. A) **Methods:** We ran a t test analysis of association testing the equality of mean FIB between people with prior history of CVD and people with no prior history of CVD at a .05 level of significance. We assumed that the standard deviation of FIB is similar between the groups with prior CVD and without prior CVD.

**Results:** Based on a .05 level of significance, we reject the null hypothesis that the mean FIB levels are equal between those with prior CVD and those without prior CVD (p-value < .0001). Of the 3802 subjects without prior CVD, the mean Fib level was 320. Of the 1131 subjects with prior CVD, the mean FIB level was 334. This 14 point difference in FIB would not be unusual if the true difference in FIB levels was between 10 and 19 points. This data provides evidence that people with prior CVD have a higher FIB than those without prior CVD.

B) **Methods:** We performed a standard linear regression with FIB as the response and CVD as the predictor without robust standard error estimates which is the analog to assuming equal variances in the t test of mean FIB for groups stratified on prior CVD history. We then compared the sign of the slope of the regression model with our conclusion in part A, and compared the p-value to test that the slope is 0 to the p-value for the t test for equal means in part A.

**Results:** The linear regression analyses between our response (FIB level) and predictor (Prior CVD history) provided the same conclusions of association as in part A. The slope provided by the linear regression analysis was positive at 15 which is the regression models’ predicted increase in FIB due to having a prior CVD history. This is very similar to the 14 point difference estimated in part A. Also, the p-value for testing for a slope of 0 was < .0001, just as the p-value for testing equal means in part A. The linear regression analysis provided the same evidence that higher FIB levels are associated with having prior CVD.

C) **Methods:** We ran a t test analysis of association testing the equality of mean FIB between people with prior history of CVD and people with no prior history of CVD at a .05 level of significance. We did not assume that the standard deviation of FIB is similar between the groups with prior CVD and without prior CVD.

**Results:** Based on a .05 level of significance, we reject the null hypothesis that the mean FIB levels are equal between those with prior CVD and those without prior CVD (p-value < .0001). Of the 3802 subjects without prior CVD, the mean Fib level was 320. Of the 1131 subjects with prior CVD, the mean FIB level was 334. This 14 point difference in FIB would not be unusual if the true difference in FIB levels was between 10 and 20 points. This data provides evidence that people with prior CVD have a higher FIB than those without prior CVD.

D) **Methods:** We performed a simple linear regression with FIB as the response and CVD as the predictor using robust standard error estimates which is the analog to not assuming equal variances in the t test of mean FIB for groups stratified on prior CVD history. We then compared the sign of the slope of the regression model with our conclusion in part C, and compared the p-value to test that the slope is 0 to the p-value for the t test for equal means in part C.

**Results:** The linear regression analyses between our response (FIB level) and predictor (Prior CVD history) provided the same conclusions of association as in part C. The slope provided by the linear regression analysis was positive at 13 which is the regression models’ predicted increase in FIB due to having a prior CVD history. This is very similar to the 14 point difference estimated in part C. Also, the p-value for testing for a slope of 0 was < .0001, just as the p-value for testing equal means in part C. The linear regression analysis provided the same evidence that higher FIB levels are associated with having prior CVD.

E) In part A, we are assuming equal variances of FIB between groups with prior CVD and groups without prior CVD. Therefore, any detected differences in the distributions of FIB between prior CVD groups is attributed to differences in the mean FIB. In part C, we are not assuming equal variances, so differences in the distributions of FIB can be attributed to differences in the mean or variance or other differences. Because of this, we expect the method in part A to provide stronger evidence (lower p-value, higher absolute value of the t test statistic) for a difference in mean FIB between prior CVD groups. This idea is supported by the absolute value of our t test statistic of 6.5 using method A and 6.1 using method C.

3. **Methods:** We analyzed the association between mean FIB across groups defined by CRP. We analyzed this association using linear regression with robust standard error estimates treating CRP and FIB as continuous data types with FIB as the response and CRP as the predictor. Both CRP and FIB are untransformed. We then offer an interpretation of the fitted intercept, the fitted slope, and a full analysis of association using this regression. Finally, after the results from problem 6, we provide a full table of FIB and CRP levels pertaining to problems 3 – 6.

**A) Results:** Our fitted intercept of 300.076 is the estimated mean FIB level among those individuals with a CRP of 0.

**B) Results:** Our fitted slope of 5.864 is the average of the estimated change in FIB for a 1 unit increase in CRP.

**C) Results:** Based on our linear regression analysis (using robust standard error estimates) between mean CRP (predictor) and mean FIB (response) levels, we conclude that there is an association between CRP and FIB levels. As CRP level increases, so too does FIB levels. For people with a CRP of 0, we estimate a mean FIB level of 300.076. Then, for every 1 unit increase in CRP, we estimate that FIB will increase by 5.864 units on average.

**D) Results:** Full table provided after all results from problem 6.

4. **Methods:** We analyzed the association between mean FIB across groups defined by logged (natural log) CRP. We analyzed this association using linear regression with robust standard error estimates treating log CRP and FIB as continuous data types with FIB as the response and log CRP as the predictor. FIB is untransformed. CRP values of 0 are replaced by 0.5. We then offer an interpretation of the fitted intercept, the fitted slope, and a full analysis of association using this regression. Finally, after the results from problem 6, we provide a full table of FIB and CRP levels pertaining to problems 3 – 6.

**A) Results:** Our fitted intercept of 294.997 is the estimated mean FIB level among those individuals with a CRP of 1.

**B) Results:** Our fitted slope of 34.900 is the average of the estimated change in FIB for a 1 unit increase in log(CRP).

**C) Results:** Based on our linear regression analysis (using robust standard error estimates) between log(CRP) (predictor) and mean FIB (response) levels, we conclude that there is an association between log(CRP) and FIB levels. As log(CRP) level increases, so too does FIB levels. For people with a CRP of 1, we estimate a mean FIB level of 294.997. Then, for every 1 unit increase in log(CRP), we estimate that FIB will increase by 34.900 units on average.

**D) Results:** Full table provided after all results from problem 6.

5. **Methods:** We analyzed the association between geometric mean FIB across groups defined by CRP. We analyzed this association using linear regression with robust standard error estimates treating CRP and log(FIB) as continuous data types with log(FIB) as the response and CRP as the predictor. In order to analyze the association between the geometric mean FIB levels and CRP, we log transformed FIB, but did not transform CRP. We then offer an interpretation of the fitted intercept, the fitted slope, and a full analysis of association using this regression. Finally, after the results from problem 6, we provide a full table of FIB and CRP levels pertaining to problems 3 – 6.

**A) Results:** Our fitted intercept of 5.704 is the estimated log(FIB) level among those individuals with a CRP of 0.

**B) Results:** Our fitted slope of 0.016 is the average of the estimated change in log(FIB) for a 1 unit increase in CRP.

**C) Results:** Based on our linear regression analysis (using robust standard error estimates) between CRP (predictor) and log(FIB) (response) levels, we conclude that there is an association between CRP and log(FIB) levels. As CRP level increases, so too does FIB levels. For people with a CRP of 0, we estimate an average log(FIB) level of 5.704. Then, for every 1 unit increase in CRP, we estimate that log(FIB) will increase by .016 units on average.

**D) Results:** Full table provided after all results from problem 6.

6. **Methods:** We analyzed the association between geometric mean FIB across groups defined by log(CRP). We analyzed this association using linear regression with robust standard error estimates treating log(CRP) and log(FIB) as continuous data types with log(FIB) as the response and log(CRP) as the predictor. In order to analyze the association between the geometric mean FIB levels and log(CRP), we log transformed FIB and CRP. We then offer an interpretation of the fitted intercept, the fitted slope, and a full analysis of association using this regression. Finally, after the results from problem 6, we provide a full table of FIB and CRP levels pertaining to problems 3 – 6.

**A) Results:** Our fitted intercept of 5.681 is the estimated log(FIB) level among those individuals with a CRP of 1 (log(CRP) of 0).

**B) Results:** Our fitted slope of 0.106 is the average of the estimated change in log(FIB) for a 1 unit increase in log(CRP).

**C) Results:** Based on our linear regression analysis (using robust standard error estimates) between log(CRP) (predictor) and log(FIB) (response) levels, we conclude that there is an association between log(CRP) and log(FIB) levels. As log(CRP) level increases, so too does FIB levels. For people with a CRP of 1, we estimate an average log(FIB) level of 5.681. Then, for every 1 unit increase in log(CRP), we estimate that log(FIB) will increase by .106 units on average.

**D) Results:** Full table provided after all results from problem 6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fitted Values for Fibrenogen (mg/dL) | | | | |
| CRP Level (mg/L) | Mean vs CRP | Mean vs Log(CRP) | Geo Mean vs CRP | Geo Mean vs log(CRP) |
| 1 | 305.9400 | 294.9974 | 5.7197 | 5.6812 |
| 2 | 311.8036 | 319.1866 | 5.7355 | 5.7547 |
| 3 | 317.6673 | 333.3364 | 5.7513 | 5.7977 |
| 4 | 323.5310 | 343.3759 | 5.7671 | 5.8281 |
| 6 | 335.2584 | 357.5257 | 5.7987 | 5.8711 |
| 8 | 346.9858 | 367.5651 | 5.8303 | 5.9016 |
| 9 | 352.8495 | 371.6755 | 5.8461 | 5.9141 |
| 12 | 370.4406 | 381.7149 | 5.8935 | 5.9446 |

7. Below is a table of differences and ratios of FIB levels compared across CRP levels using the various regression models/methods in problems 3 – 6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fitted Values for Fibrenogen (mg/dL) | | | | |
| CRP Level | Differences | | | |
| Comparison(mg/L) | Mean vs CRP | Mean vs Log(CRP) | Geo Mean vs CRP | Geo Mean vs log(CRP) |
| 2 mg/L - 1 mg/L | 5.8636 | 24.1892 | 0.0158 | 0.0735 |
| 3 mg/L - 2 mg/L | 5.8637 | 14.1498 | 0.0158 | 0.0430 |
| 4 mg/L - 1 mg/L | 17.5910 | 48.3785 | 0.0474 | 0.1469 |
| 4 mg/L - 2 mg/L | 11.7274 | 24.1893 | 0.0316 | 0.0734 |
| 6 mg/L - 3 mg/L | 17.5911 | 24.1893 | 0.0474 | 0.0734 |
| 8 mg/L - 4 mg/L | 23.4548 | 24.1892 | 0.0632 | 0.0735 |
| 9 mg/L - 6 mg/L | 17.5911 | 14.1498 | 0.0474 | 0.0430 |
| 9 mg/L - 8 mg/L | 5.8637 | 4.1104 | 0.0158 | 0.0125 |
| 12 mg/L - 6 mg/L | 35.1822 | 24.1892 | 0.0948 | 0.0735 |
|  | Ratios | | | |
| 2 mg/L / 1 mg/L | 1.0192 | 1.0820 | 1.0028 | 1.0129 |
| 3 mg/L / 2 mg/L | 1.0188 | 1.0443 | 1.0028 | 1.0075 |
| 4 mg/L / 1 mg/L | 1.0575 | 1.1640 | 1.0083 | 1.0259 |
| 4 mg/L / 2 mg/L | 1.0376 | 1.0758 | 1.0055 | 1.0128 |
| 6 mg/L / 3 mg/L | 1.0554 | 1.0726 | 1.0082 | 1.0127 |
| 8 mg/L / 4 mg/L | 1.0725 | 1.0704 | 1.0110 | 1.0126 |
| 9 mg/L / 6 mg/L | 1.0525 | 1.0396 | 1.0082 | 1.0073 |
| 9 mg/L / 8 mg/L | 1.0169 | 1.0112 | 1.0027 | 1.0021 |
| 12 mg/L / 6 mg/L | 1.1049 | 1.0677 | 1.0163 | 1.0125 |

8. A) The analysis that gave constant differences in the fitted values when comparing 2 groups that differed by an absolute increase in c units in CRP levels was the mean FIB vs CRP. Those differences are provided below. Notice that for every unit increase in CRP, mean FIB is estimated to increase by an average of 5.8636. That 5.8636 estimated increase remains constant, and therefore is not dependent on the current CRP level.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fitted Values for Fibrenogen (mg/dL) | | | | |
| CRP Level | Differences | | | |
| Comparison(mg/L) | Mean vs CRP |  |  |
| 2 mg/L - 1 mg/L | 5.8636 |  |  |
| 3 mg/L - 2 mg/L | 5.8637 |  |  |
| 4 mg/L - 1 mg/L | 17.5910 |  |  |
| 4 mg/L - 2 mg/L | 11.7274 |  |  |
| 6 mg/L - 3 mg/L | 17.5911 |  |  |
| 8 mg/L - 4 mg/L | 23.4548 |  |  |
| 9 mg/L - 6 mg/L | 17.5911 |  |  |
| 9 mg/L - 8 mg/L | 5.8637 |  |  |
| 12 mg/L - 6 mg/L | 35.1822 |  |  |

B) The analysis that gave constant ratios in the fitted values when comparing 2 groups that differed by an absolute increase in c units in CRP levels was geometric mean FIB vs CRP. Those ratios are provided below. Notice that for a unit increase in CRP, the average estimated geometric mean FIB increased by .0158, independent of what the CRP level was.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fitted Values for Fibrenogen (mg/dL) | | | | |
| CRP Level | Differences | | | |
| Comparison(mg/L) |  |  | Geo Mean vs CRP |  |
| 2 mg/L - 1 mg/L |  |  | 0.0158 |  |
| 3 mg/L - 2 mg/L |  |  | 0.0158 |  |
| 4 mg/L - 1 mg/L |  |  | 0.0474 |  |
| 4 mg/L - 2 mg/L |  |  | 0.0316 |  |
| 6 mg/L - 3 mg/L |  |  | 0.0474 |  |
| 8 mg/L - 4 mg/L |  |  | 0.0632 |  |
| 9 mg/L - 6 mg/L |  |  | 0.0474 |  |
| 9 mg/L - 8 mg/L |  |  | 0.0158 |  |
| 12 mg/L - 6 mg/L |  |  | 0.0948 |  |

C) The analysis that gave constant differences in the fitted values when comparing 2 groups that differed by a relative c-fold increase in CRP levels was the mean FIB vs log(CRP). Those differences are provided below. Notice that for every 2-fold increase in CRP, mean FIB is estimated to increase by an average of 24.1892. That 24.1892 estimated average increase remains constant, and therefore is not dependent on the current CRP level.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fitted Values for Fibrenogen (mg/dL) | | | | |
| CRP Level | Differences | | | |
| Comparison(mg/L) |  | Mean vs Log(CRP) |  |  |
| 2 mg/L - 1 mg/L |  | 24.1892 |  |  |
| 3 mg/L - 2 mg/L |  | 14.1498 |  |  |
| 4 mg/L - 1 mg/L |  | 48.3785 |  |  |
| 4 mg/L - 2 mg/L |  | 24.1893 |  |  |
| 6 mg/L - 3 mg/L |  | 24.1893 |  |  |
| 8 mg/L - 4 mg/L |  | 24.1892 |  |  |
| 9 mg/L - 6 mg/L |  | 14.1498 |  |  |
| 9 mg/L - 8 mg/L |  | 4.1104 |  |  |
| 12 mg/L - 6 mg/L |  | 24.1892 |  |  |

D) The analysis that gave constant ratios in the fitted values when comparing 2 groups that differed by a relative c-fold increase in CRP levels was the geometric mean FIB vs log(CRP). Those ratios are provided below. Notice that for every 2-fold increase in CRP, the geometric mean FIB is estimated to increase by an average factor of 1.013. That 1.013 estimated average factor increase remains constant, and therefore is not dependent on the current CRP level.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fitted Values for Fibrenogen (mg/dL) | | | | |
| CRP Level | Differences | | | |
| Comparison(mg/L) |  |  |  | Geo Mean vs log(CRP) |
|  | Ratios | | | |
| 2 mg/L / 1 mg/L |  |  |  | 1.0129 |
| 3 mg/L / 2 mg/L |  |  |  | 1.0075 |
| 4 mg/L / 1 mg/L |  |  |  | 1.0259 |
| 4 mg/L / 2 mg/L |  |  |  | 1.0128 |
| 6 mg/L / 3 mg/L |  |  |  | 1.0127 |
| 8 mg/L / 4 mg/L |  |  |  | 1.0126 |
| 9 mg/L / 6 mg/L |  |  |  | 1.0073 |
| 9 mg/L / 8 mg/L |  |  |  | 1.0021 |
| 12 mg/L / 6 mg/L |  |  |  | 1.0125 |

9. The major determining factor in which analysis to use to investigate associations between fibrinogen and CRP is the science behind the relationships. When investigating associations we should first utilize our scientific knowledge about how our predictor and response behave. We generally believe that biological processes work on a multiplicative scale. Since both FIB and CRP are both affected and affect biological processes, I would analyze the association between the two using the geometric mean FIB vs log(CRP). When we do not know the science behind the question we may ask ourselves which association is the most likely to exist. Failing in that knowledge, we might consider which choice will give us the greatest precision, but only if we know nothing else. If we truly know nothing about the variables involved we will probably analyze association by comparing mean vs mean.