Homework 01

ID: 2294

1. Methods: Per each level of the prior cardiovascular disease descriptives (mean, standard deviation, minimum and maximum) are computed for CRP (blood C reactive protein in mg/l) and FIB (blood fibrinogen in mg/dl). Pearson’s correlation coefficient is used to measure the association between CRP and FIB. It is computed for over subjects and separately for groups having no prior history of diagnosed cardiovascular disease (CVD) or having prior diagnosed CVD. Also scatter plots and LOWESS plots are used to explore this association.

Over the 5000 subjects on the dataset 51 haven’t both CRP and FIB values, 34 miss FIB and another 16 miss measurement of CRP. Only cases with both measurements are kept for the correlation computations.

Results: The CRP and FIB descriptive are depicted on the next table. Both tend to have higher means on subjects with prior CVD. Among the 4899 with both measurements on CVD and CRP the correlation coefficient is 0.49 whereas among the 1122 subjects with prior CVD the correlation is 0.52 a little bit higher than the correlation of 0.46 on 3777 subjects who haven’t prior CVD.

|  |  |  |
| --- | --- | --- |
|  | **Previous Cardiovascular Disease** |  |
| **Characteristic** | **Yes** | **No** | **Total** |
| N = 1149 | N = 3851 | N = 5000 |
| Blood C reactive protein levels (mg/l) | 4.4 (6.88; 0.0 - 83.0) [18] | 3.4 (5.90; 0.0 - 108.0) [49] | 3.6 (6.15; 0.0 - 108.0) [67] |
| Blood fibrinogen in mg/dl | 334.5 (74.06; 138.0 - 695.0) [25] | 319.6 (64.76; 109.0 - 872.0) [60] | 323.0 (67.3; 109.0 - 872.0) [85] |
|   |   |   |   |
| \* Descriptive statistics: mean (standard deviation; minimum - maximum)[Total missing values] |



Figure – Relation between CRP and FIB through different levels of prior CVD.

2.a) Methods: Mean values of serum fibrinogen (FIB) are compared between subjects who had prior CVD and those who hadn’t prior CVD. T-test with equal variances assumption is used to test the mean difference to be different from zero (bilateral test with 5% level). Also 95% confidence intervals (95% CI) are presented. FIB values were missing on 85 subjects; they were dropped.

Results: The mean serum level of FIB was 334.5 mg/dl (95% CI: 330.1 – 338.8) among the 1124 subjects with prior CVD compared to 319.6 mg/dl (95% CI: 317.5 – 321.6) among 3791 subjects without prior CVD. There is a tendency of 14.9 mg/dl higher FIB levels among the subjects with prior CVD and it wouldn’t be unusual to find this difference between 10.4 and 19.3 mg/dl. The p-value is below 0.0001 so we reject the null hypothesis on no mean differences on both groups.

2.b) Methods: A linear regression with FIB as the outcome variable and prior CVD status as the predictor is computed using least squares method. . Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%.

Results: The estimated regression coefficients are shown on table 1. The mean serum FIB level among those without prior CVD is 319.6 mg/dl which wouldn’t be unusual to find it between 317.4 and 321.7 mg/dl (p < 0.001). Subjects with a prior CVD have 14.9 mg/dl higher FIB levels than those without prior CVD, and it wouldn’t be unusual to find this difference between 10.4 and 19.3 mg/dl. The p-value is below 0.0001 so we reject the null hypothesis on no mean differences on both groups.

 Table 1 – Linear regression coefficients

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **95% CI** | **p** |
| Total cases in the model | 4915 |  |  |
|  |  |  |  |
| Had a previous CVD | 14.9 | 10.4 - 19.3 | < 0.001 |
| Intercept | 319.6 | 317.4 - 321.7 | < 0.001 |
|   |   |   |   |

Therefore the mean FIB level among subjects with prior CVD is 334.5 mg/dl and wouldn’t be unusual to lie between 330.5 and 338.4 mg/dl.

The linear regression intercept corresponds to the mean on the group without prior CVD. The regression coefficient of the dummy variable representing prior CVD is the mean difference between both groups on the t-test with equal variances assumption. Also the p-value and standard error of the regression coefficient of the dummy variable representing prior CVD is the same on the difference between both groups on the t-test with equal variances assumption.

2.c) Methods: Mean values of serum fibrinogen (FIB) are compared between subjects who had prior CVD and those who hadn’t prior CVD. T-test with unequal variances assumption is used to test the mean difference to be different from zero (bilateral test with 5% level). Also 95% confidence intervals (95% CI) are presented. FIB values were missing on 85 subjects; they were dropped.

Results: The mean serum level of FIB was 334.5 mg/dl (95% CI: 330.1 – 338.8) among the 1124 subjects with prior CVD compared to 319.6 mg/dl (95% CI: 317.5 – 321.6) among 3791 subjects without prior CVD. There is a tendency of 14.9 mg/dl higher FIB levels among the subjects with prior CVD and it wouldn’t be unusual to find this difference between 10.1 and 19.7 mg/dl. The p-value is below 0.0001 so we reject the null hypothesis on no mean differences on both groups.

2.d) Methods: A linear regression with FIB as the outcome variable and prior CVD status as the predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%.

Results: The estimated regression coefficients are shown on table 2. The mean serum FIB level among those without prior CVD is 319.6 mg/dl which wouldn’t be unusual to find it between 317.5 and 321.6 mg/dl (p < 0.001 i.e we reject the null hypothesis that intercept is zero). Subjects with a prior CVD have 14.9 mg/dl higher FIB levels than those without prior CVD, and it wouldn’t be unusual to find this difference between 10.1 and 19.7 mg/dl. The p-value is below 0.0001 so we reject the null hypothesis on no mean differences on both groups.

Table 2 – Linear regression coefficients

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **95% CI** | **p** |
| Total cases in the model | 4915 |  |  |
|  |  |  |  |
| Had a previous CVD | 14.9 | 10.1 - 19.7 | < 0.001 |
| Intercept | 319.6 | 317.5 - 321.6 | < 0.001 |
|   |   |   |   |

The linear regression intercept corresponds to the mean on the group without prior CVD. The regression coefficient of the dummy variable representing prior CVD is the mean difference between both groups on the t-test with unequal variances assumption. Also the p-value and standard error of the regression coefficient of the dummy variable representing prior CVD is the same on the difference between both groups on the t-test with unequal variances assumption. The use of the robust SE in linear regression leads to inference that is approximately (but not exactly) the same as a t test that allows for the possibility of unequal variances. The point estimates will agree exactly.

2.e) On a) we performed a Sudent’s t-test with equal variances assumption. According to the software output the there is an imbalance in group sample sizes and variances are not equal on the sample. Under these circumstances even though the bilateral p-value is below 0.0001. That suggest a statistical significant p-value for the t-test with unequal variances assumption.

3.a) Methods: A linear regression with FIB as the outcome variable and CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%.

Results: The mean serum FIB levels among subjects with serum CRP held at zero mg/l is 304.015 mg/dl and it wouldn’t be unusual to lie between 301.514 – 306.517mg/dl. We reject the null hypothesis that this mean is equal to zero with p-value < 0.001 (table 3).

Table 3 – Linear regression coefficients

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **95% CI** | **p** |
| Total cases in the model | 4899 |  |  |
|  |  |  |  |
| Serum CRP levels (mg/l) | 5.251 | 4.604 - 5.898 | < 0.001 |
| Intercept | 304.015 | 301.514 - 306.517 | < 0.001 |
|   |   |   |   |

3.b) Methods: A linear regression with FIB as the outcome variable and CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%.

Results: Refer to table 3. Subjects separated per each higher 1 mg/dl in serum CRP levels tend to have a 5.251 mg/l higher mean in serum FIB levels.

3.c) Methods: A linear regression with FIB as the outcome variable and CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%.

Results: Subjects separated per each higher 1 mg/dl in serum CRP levels tend to have a 5.251 mg/l higher mean in serum FIB levels. This mean difference would not be unusual to be found between 4.604 and 5.898 mg/dl and it is statistically different from zero (p-value < 0.001). Refer to table 3. So there is an association between fibrinogen and CRP, at least a first trend one.

3.d) Methods : Is the same as on the 3.c.

Results: Refer to the next table. Fitted values are computed using the coefficients computed 3a. They represent mean values of FIB serum levels in mg/dl among subjects group defined by levels CRP.

|  |  |
| --- | --- |
|   | **Fitted Values for Fibrinogen (mg/dL)** |
| **CRP level** | **Mean** |
| **1 mg/L** | 309.266 |
| **2 mg/L** | 314.517 |
| **3 mg/L** | 319.768 |
| **4 mg/L** | 325.019 |
| **6 mg/L** | 335.520 |
| **8 mg/L** | 346.022 |
| **9 mg/L** | 351.273 |
| **12 mg/L** | 367.026 |

4.a) Methods: A linear regression with FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. Log transformation uses natural logarithm.

Results: The mean serum FIB levels among subjects with serum CRP held at 1 mg/l is 295.566 mg/dl and it wouldn’t be unusual to lie between 293.645 – 297.488 mg/dl. We reject the null hypothesis that this mean is equal to zero with p-value < 0.001 (table 4).

Table 4 – Linear regression coefficients with log serum CRP

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **95% CI** | **p** |
| Total cases in the model | 4899 |  |  |
|  |  |  |  |
| Log serum CRP levels (mg/l) | 36.833 | 34.577 - 39.089 | < 0.001 |
| Intercept | 295.566 | 293.645 - 297.488 | < 0.001 |
|   |   |   |   |

4.b) Methods: A linear regression with FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. All log transformation uses natural logarithm. For interpretation the log CRP linear regression coefficient is multiplied by log(2).

Results: Refer to table 4. A twice increase in serum CRP levels is associated with 25.531 mg/l higher in mean serum FIB levels.

4.c) Methods: A linear regression with FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. All log transformation uses natural logarithm. For interpretation the log CRP linear regression coefficient is multiplied by log(2).

Results: A twice increase in serum CRP levels is associated with 25.531 mg/l higher in mean serum FIB levels. This mean difference would not be unusual to be found between 23.987 and 27.094 mg/dl and we reject the null hypothesis of no linear association between FIB levels and log transformed CRP (p-value < 0.001). Refer to table 4.

4.d) Methods : Is the same as on the 4.c.

Results: Refer to the next table. Fitted values are computed using the coefficients computed 4a. They represent mean values of FIB serum levels in mg/dl among subjects group defined by levels CRP.

|  |  |
| --- | --- |
|   | **Fitted Values for Fibrinogen (mg/dL)** |
| **CRP level** | **Mean** |
| **1 mg/L** | 295.566 |
| **2 mg/L** | 321.097 |
| **3 mg/L** | 336.032 |
| **4 mg/L** | 346.628 |
| **6 mg/L** | 361.562 |
| **8 mg/L** | 372.159 |
| **9 mg/L** | 376.497 |
| **12 mg/L** | 387.093 |

5.a) Methods: A linear regression with log transformed FIB as the outcome variable and CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. All log transformation uses natural logarithm. To interpret the intercept natural exponentiation is used.

Results: The geometric mean of serum FIB levels among subjects with serum CRP held at 0 mg/l is 300.896 mg/dl and it wouldn’t be unusual to lie between 298.647 – 303.162 mg/dl. We reject the null hypothesis that this mean is equal to zero with p-value < 0.001 (table 5).

Table 5 – Linear regression coefficients with serum CRP, FIB is on log scale

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **95% CI** | **p** |
| Total cases in the model | 4899 |  |  |
|  |  |  |  |
| Serum CRP levels (mg/l) | 0.014 | 0.012 - 0.016 | < 0.001 |
| Intercept | 5.707 | 5.699 - 5.714 | < 0.001 |
|   |   |   |   |

5.b) Methods: A linear regression with FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. All log transformation uses natural logarithm. For interpretation of the predictor coefficient the following formula is used $100(e^{β}-1)$which represents the percentage change per one unit higher of β.

Results: Refer to table 5. The geometric mean of FIB tends to be 1.402% higher per each 1 mg/l higher of serum CRP levels.

5.c) Methods: A linear regression with FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. All log transformation uses natural logarithm. For interpretation the following formula is used $100(e^{β}-1)$which represents the percentage change per one unit higher of β.

Results: The geometric mean of FIB tends to be 1.402% higher per each 1 mg/l higher of serum CRP levels. This mean ratio would not be unusual to be found between 1.223 and 1.558% and we reject the null hypothesis of no linear association between log FIB levels and CRP (p-value < 0.001). Refer to table 5.

5.d) Methods : Is the same as on the 5.c.

Results: Refer to the next table. Fitted values are computed using the coefficients computed 5a. They represent mean values of FIB serum levels in mg/dl among subjects group defined by levels CRP.

|  |  |
| --- | --- |
|   | **Fitted Values for Fibrinogen (mg/dL)** |
| **CRP level** | **Geometric Mean** |
| **1 mg/L** | 305.113 |
| **2 mg/L** | 309.390 |
| **3 mg/L** | 313.726 |
| **4 mg/L** | 318.123 |
| **6 mg/L** | 327.103 |
| **8 mg/L** | 336.337 |
| **9 mg/L** | 341.051 |
| **12 mg/L** | 355.593 |

6.a) Methods: A linear regression with log transformed FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. All log transformation uses natural logarithm. To interpret the intercept natural exponentiation is used.

Results: The geometric mean of serum FIB levels among subjects with serum CRP held at 1 mg/l is 292.536 mg/dl and it wouldn’t be unusual to lie between 290.732 – 294.351 mg/dl. We reject the null hypothesis that this mean is equal to zero with p-value < 0.001 (table 6).

Table 6 – Linear regression coefficients with log serum CRP and log serum FIB

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **95% CI** | **p** |
| Total cases in the model | 4899 |  |  |
|  |  |  |  |
| Serum CRP levels (mg/l) | 0.105 | 0.100 - 0.111 | < 0.001 |
| Intercept | 5.679 | 5.672 - 5.685 | < 0.001 |
|   |   |   |   |

6.b) Methods: A linear regression with log transformed FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. All log transformation uses natural logarithm. For interpretation of the predictor coefficient the following formula is used $100(e^{βlog⁡(1.1)}-1)$which represents the associated percentage change per 10% higher of β.

Results: Refer to table 6. The geometric mean of FIB tends to be 12.292% higher per each 10% higher serum CRP level.

6.c) Methods: A linear regression with FIB as the outcome variable and log transformed CRP as predictor is computed using least squares method and with standard errors computed using the Huber-White sandwich estimator. Its coefficients, 95% confidence interval are reported and p-values. The significance level is set to 5%. Values of CRP that are zero replace by 0.5 prior the log transformation. All log transformation uses natural logarithm. For interpretation of the predictor coefficient the following formula is used $100(e^{βlog⁡(1.1)}-1)$which represents the associated percentage change per 10% higher of β.

Results: The geometric mean of FIB tends to be 12.292% higher per each 10% higher serum CRP level. This mean ratio would not be unusual to be found between 11.571 and 13.018% and we reject the null hypothesis of no linear association between log FIB levels and CRP (p-value < 0.001). Refer to table 6.

6.d) Methods : Is the same as on the 6.c.

Results: Refer to the next table. Fitted values are computed using the coefficients computed 5a. They represent mean values of FIB serum levels in mg/dl among subjects group defined by levels CRP.

|  |  |
| --- | --- |
|   | **Fitted Values for Fibrinogen (mg/dL)** |
| **CRP level** | **Geometric Mean** |
| **1 mg/L** | 292.536 |
| **2 mg/L** | 314.706 |
| **3 mg/L** | 328.446 |
| **4 mg/L** | 338.556 |
| **6 mg/L** | 353.337 |
| **8 mg/L** | 364.214 |
| **9 mg/L** | 368.764 |
| **12 mg/L** | 380.116 |

 Joining all tables with fitted values

Table 7 – Different fitted values for different levels of CRP.

|  |  |
| --- | --- |
|   | **Fitted Values for Fibrinogen (mg/dL)** |
| **CRP level** | **Problem 3 Mean** | **Problem 4 Mean** | **Problem 5 Geometric Mean** | **Problem6 Geometric Mean** |
| **1 mg/L** | 309.266 | 295.566 | 305.113 | 292.536 |
| **2 mg/L** | 314.517 | 321.097 | 309.390 | 314.706 |
| **3 mg/L** | 319.768 | 336.032 | 313.726 | 328.446 |
| **4 mg/L** | 325.019 | 346.628 | 318.123 | 338.556 |
| **6 mg/L** | 335.520 | 361.562 | 327.103 | 353.337 |
| **8 mg/L** | 346.022 | 372.159 | 336.337 | 364.214 |
| **9 mg/L** | 351.273 | 376.497 | 341.051 | 368.764 |
| **12 mg/L** | 367.026 | 387.093 | 355.593 | 380.116 |

7. Methods: Different fitted values on exercises 3, 4, 5 and 6 are compared through differences and ratios.

Results: Refer to table 8.

Table 8 – Comparison of different fitted values on problems 3, 4, 5 and 6.

|  |  |
| --- | --- |
|   | **Fitted Values for Fibrinogen (mg/dL)** |
| **CRP level** | **Problem 3 Mean** | **Problem 4 Mean** | **Problem 5 Geometric Mean** | **Problem6 Geometric Mean** |
| **Differences** |
| **2 mg/L – 1 mg/L** | 5.251 | 25.531 | 4.276 | 22.170 |
| **3 mg/L – 2 mg/L** | 5.251 | 14.935 | 4.336 | 13.740 |
| **4 mg/L – 1 mg/L** | 15.753 | 51.062 | 13.010 | 46.021 |
| **4 mg/L – 2 mg/L** | 10.502 | 25.531 | 8.734 | 23.850 |
| **6 mg/L – 3 mg/L** | 15.752 | 25.531 | 13.377 | 24.892 |
| **8 mg/L – 4 mg/L** | 21.003 | 25.531 | 18.214 | 25.658 |
| **9 mg/L – 6 mg/L** | 15.753 | 14.935 | 13.948 | 15.426 |
| **9 mg/L – 8 mg/L** | 5.251 | 4.338 | 4.714 | 4.549 |
| **12 mg/L – 6 mg/L** | 31.506 | 25.531 | 28.490 | 26.778 |
| **Ratios** |
| **2 mg/L – 1 mg/L** | 1.017 | 1.086 | 1.014 | 1.076 |
| **3 mg/L – 2 mg/L** | 1.017 | 1.047 | 1.014 | 1.044 |
| **4 mg/L – 1 mg/L** | 1.051 | 1.173 | 1.043 | 1.157 |
| **4 mg/L – 2 mg/L** | 1.033 | 1.080 | 1.028 | 1.076 |
| **6 mg/L – 3 mg/L** | 1.049 | 1.076 | 1.043 | 1.076 |
| **8 mg/L – 4 mg/L** | 1.065 | 1.074 | 1.057 | 1.076 |
| **9 mg/L – 6 mg/L** | 1.047 | 1.041 | 1.043 | 1.044 |
| **9 mg/L – 8 mg/L** | 1.015 | 1.012 | 1.014 | 1.012 |
| **12 mg/L – 6 mg/L** | 1.094 | 1.071 | 1.087 | 1.076 |

8a) Problem 3 (linear regression without any transformations). For example for an absolute 1mg/L difference in CRP levels (2mg/L – 1 mg/L, 3mg/L – 2mg/L, 9mg/l – 8 mg/L) fitted FIB maintained an absolute difference 5.251 mg/dl. For absolute 3mg/L (4 mg/L – 1 mg/L, 6 mg/L – 3 mg/L, 9 mg/L – 6 mg/L) in CRP the correspondent absolute increase in fitted FIB is 15.753.

8b) Problem 5 (linear regression with outcome log transformed). For example for an absolute 1mg/L difference in CRP levels (2mg/L – 1 mg/L, 3mg/L – 2mg/L, 9mg/l – 8 mg/L) fitted FIB maintained a relative increase 1.014. For absolute 3mg/L (4 mg/L – 1 mg/L, 6 mg/L – 3 mg/L, 9 mg/L – 6 mg/L) in CRP the correspondent relative increase in fitted FIB is 1.043.

8c) Problem 4 (linear regression with predictor log transformed). For example for a relative increase of 2 in CRP levels (2mg/L – 1 mg/L, 4mg/L – 2mg/L, 6mg/l – 3 mg/L, 8mg/l – 4 mg/L and 12mg/l – 6mg/L) fitted FIB maintained an absolute difference 25.331 mg/dl. For a relative increase of 1.5 in CRP levels (3 mg/L – 2 mg/L, 9 mg/L – 6 mg/L) fitted FIB maintained an absolute difference of 14.935.

8d) Problem 6 (linear regression with predictor log transformed and outcome log transformed). For example for a relative increase of 2 in CRP levels (2mg/L – 1 mg/L, 4mg/L – 2mg/L, 6mg/l – 3 mg/L, 8mg/l – 4 mg/L and 12mg/l – 6mg/L) fitted FIB maintained a relative increase of 1.076. For a relative increase of 1.5 in CRP levels (3 mg/L – 2 mg/L, 9 mg/L – 6 mg/L) fitted FIB maintained a relative increase of 1.044.

9. This should be specified prior to data analysis. According to the plots on problem 1 the association between FIB and CRP is not linear. FIB values grow fast in small CRP and grow almost linearly there after. I would consider to use the model on problem 4 or the model on problem 6.