BIOST 518
Homework 2

1. **Methods**: Of the 5000 participants of the study, only 4899 had data for serum CRP levels and serum fibrinogen levels. Subjects with missing data were excluded from the analysis. Descriptive statistics are presented here by groups defined by serum CRP levels as measured by the study (less than 1 mg/L, 1-3 mg/L, and greater than 3 mg/L). Further, the population was divided into those who had prior history of cardiovascular history and those who did not. The values reported in the table below are the mean, standard deviation, min and max.

**Results**: The results of the analysis indicates that of the 4899 subject with available data, 1956 had low CRP levels (below 1 mg/L), 1081 had medium CRP levels (between 2 and 3 mg/L) and 1862 had high serum CRP levels (greater than 3 mg/L). Further, 3777 of the participants had no previous atherosclerotic event prior to the study, while 1122 had prevalent atherosclerotic events at the time of enrollment. Within the total population, the mean fibrinogen level was 294.60 mg/dL among those with low serum CRP levels, 314.34 mg/dL among those with medium CRP levels, and 357.91 mg/dL among those with high serum CRP levels. This is indicative of a positive association between serum CRP and fibrinogen levels. In the population with no prior cardiovascular disease, the mean fibrinogen levels were 293.17 mg/dL among those with low CRP levels, 313.63 mg/dL among those with medium CRP levels and 354.24 mg/dL among those with high CRP levels. Looking at the population who had prior cardiovascular disease, the mean fibrinogen levels were 300.75 mg/dL among those with low serum CRP, 316.75 mg/dL among those with medium CRP and 367.73 mg/dL among those with high CRP. Within these two populations, the trend of increasing serum CRP levels as fibrinogen levels increase appears to hold, providing more evidence that serum CRP and fibrinogen levels are positively associated.

**Table 1**: Fibrinogen levels among the study population by serum CRP levels , stratified by CVD history.

|  |  |
| --- | --- |
|  | **Serum C Reactive Protein (CRP)** |
|  | **< 1mg/L** | **1-3 mg/L** | **< 3 mg/L** |
| **Total Population (N=4899)** |
| **N** | 1956 | 1081 | 1862 |
| **Fibrinogen (mg/dL)** | 294.60 (50.17; 109-540) | 314.34 (51.45; 138-482) | 357.91 (75.16; 132-872) |
| **Prior CVD (N=1122)** |
| **N** | 370 | 246 | 506 |
| **Fibrinogen (mg/dL)** | 300.75 (53.90; 171-540) | 316.75 (52.52; 138-470) | 367.73 (81.47; 175-695) |
| **No Prior CVD (N=3777)** |
| **N** | 1586 | 835 | 1356 |
| **Fibrinogen (mg/dL)** | 293.17 (49.17; 109-482) | 313.63 (51.15; 183-482) | 354.24 (72.35; 132-872) |

Fibrinogen values are reported as mean (standard deviation; min-max)

* 1. **Methods**: The differences in the mean serum fibrinogen levels between those prior CVD and those with no prior CVD were tested using a t test which assume equal variances. 95 % confidence intervals were also generated with the same assumption of equal variances.

	**Inference**: Serum fibrinogen was 319.57 mg/dL among those with no prior CVD and 334.46 mg/dL among those with a history of CVD. The observed difference of those with no history of CVD having 14.89 mg/dL lower fibrinogen levels than those with prior history of CVD would not be unusual if the true difference in mean CVD levels were between 10.42 mg/dL and 19.35 mg/dL lower in the population with a history if CVD, based on calculated 95% confidence intervals. Furthermore, a t test which assumes equal variance indicates that this observation is statistically significant at a level of 0.05 (two-sided P=0.0000). We can therefore conclude that the mean serum fibrinogen levels are different between those who have CVD history and those who do not.
	2. Since linear regression analysis is essentially the t test, especially when assuming equal variance among populations, then running linear regression analysis on the same data set should yield the same values. In particular, when using the Stata software, the difference in means reported using the t test, standard deviation, P value, and 95% confidence intervals are exactly the same as those reported in the linear regression analysis. When reading the output of the linear regression analysis, these values can be found on the second table provided by the software, though the difference in the mean is reported as the coefficient and, in this particular case, is positive instead of negative. Linear regression analysis indicates this difference in means, but the t test will indicate how it is different based on the sign (i.e. a negative difference means that the mean is lower in the population without history of CVD).
	3. **Methods**: The differences in the mean serum fibrinogen levels between those prior CVD and those with no prior CVD were tested using a t test which allows for unequal variances. 95 % confidence intervals were also generated allowing for unequal variances.

	**Inference**: Serum fibrinogen was 319.57 mg/dL among those with no prior CVD and 334.46 mg/dL among those with a history of CVD. The observed difference of those with no history of CVD having 14.89 mg/dL lower fibrinogen levels than those with prior history of CVD would not be unusual if the true difference in mean CVD levels were between 10.09 mg/dL and 19.68 mg/dL lower in the population with a history if CVD, based on calculated 95% confidence intervals, and allowing for unequal variances between the two populations. The t test, allowing for unequal variance, indicates that this observation is statistically significant at a level of 0.05 (two-sided P=0.0000). We can therefore conclude, with high confidence, that the mean serum fibrinogen levels are different between those who have CVD history and those who do not and reject the null hypothesis that there is no difference in serum fibrinogen levels between the two populations.
	4. When assuming heteroscedastic data, or unequal variance, performing linear regression in Stata requires that a “robust” analysis is specified in the command. While the values obtained from the output are not exactly the same (as with regular linear regression and t tests assuming equal variances), they are very similar. The values of interest with the robust linear regression analysis are the values reported as “coef.”, “robust std. err.” , “t”, “P>|t|”, and the 95% confidence intervals. The “coef.” Value is the difference in means reported in the t test.
	5. Given the identical outputs between the two t tests (assuming equal and unequal variances), I would say that it is not likely that running a t test assuming equal variances would allow you to predict whether a second t test allowing for unequal variances would provide a stronger/weaker association. With regards to these specific analyses using the inflammation data, the reported P values are identical and so that would not matter so much. Additionally, the 95% confidence intervals are, while different between the two analyses, are only **slightly** different. The case is the same for the t statistics. In general, with a population difference between the two groups as large as in this population (i.e. 1124 vs 3791, or roughly 3.5-fold) then an analysis which assumes unequal variance would be recommended.
1. **Methods**: The association between mean fibrinogen levels across groups defined by CRP was evaluated by performing a robust linear regression analysis to evaluate linear trends. The Huber-White sandwich estimator was used to determine standard errors.

	1. The estimated intercept was found to be 304.015. With regards to the fibrinogen levels, this intercept can be regarded as the estimated level of serum fibrinogen if serum CRP levels are exactly 0.
	2. The slope was found to be 5.251. With regards to fibrinogen levels, this slope indicates that for every mg/L rise in serum CRP concentrations, the serum fibrinogen level is estimated to increase by 5.251 mg/dL.
	3. **Inference**: Based on the linear regression analysis conducted, we find that for every 1 mg/L increase in mean serum CRP levels, the mean fibrinogen level is estimated to increase by 5.251 mg/dL. Those with higher serum CRP levels were observed have higher serum fibrinogen levels. This observed association between serum CRP and fibrinogen levels would not be judged to be unusual if the true increase in fibrinogen levels were between 4.60 and 5.898 mg/d, based on calculated 95% confidence intervals. At a statistical significance level of 0.05, these results are indicative of an association between the serum levels of CRP and fibrinogen (two-sided P=0.0000).
	4. **See table below**.
2. **Methods**: The CRP values were first log-transformed (base e). Subjects whose serum CRP levels were 0 were assigned CRP levels of 0.5 mg/L in order to allow for log-transformation. The association between the mean fibrinogen levels as defined by the continuous log-transformed CRP levels was evaluated using linear regression analysis. The Huber-White sandwich approximation was used to determine standard errors.

	1. The intercept was found to be 295.566. With regards to fibrinogen levels, this value is the fibrinogen value when the log-transformed CRP level is 0. This means that when CRP level is 1 (log 1=0), then the estimated fibrinogen level in the serum is 295.566 mg/dL.
	2. The slope of this analysis was found to be 36.833. With regards to fibrinogen levels, this means that for every unit logarithmic increase in CRP levels, fibrinogen levels are estimated to rise by 36.833 mg/dL.
	3. **Inference**: Based on the regression analysis conducted, we find that for every unit of logarithmic CRP level, the mean fibrinogen level is expected to increase by 36.833 mg/dL and those with higher log CRP (thus, higher CRP levels). The calculated 95% confidence intervals, this observed association between the natural log of CRP levels and fibrinogen levels would not be unusual if the true increase in fibrinogen level for every unit of log CRP was between 34.577 and 39.089 mg/dL. At a statistical significance level of 0.05, there is evidence that an association does exist between serum fibrinogen levels and the log-transformed serum CRP levels.
	4. **See table below.**
3. **Methods**: The serum fibrinogen values were first log-transformed (base e). The association between the geometric mean fibrinogen levels as defined by the continuous serum CRP levels was evaluated using robust regression analysis. In order to make inferences on the geometric mean of fibrinogen levels, the estimates and confidence intervals were exponentiated.

	1. The geometric mean fibrinogen intercept obtained from the analysis was found to be 300.896 mg/dL. This means that when serum CRP levels are 0, the estimated geometric mean is estimated to be 300.896 mg/dL.
	2. The slope of the analysis was found to be 1.0140. With regards to the fibrinogen levels, this means that for each incremental increase in serum CRP levels, serum fibrinogen levels are estimated to increase by 1.40 %.
	3. **Inference**: The regression analysis indicates that for every integral increase in serum CRP levels, the geometric mean fibrinogen level is estimated to increase by 1.40%. 95% confidence intervals indicate that this observation would not be unusual if the true increase in the geometric mean of serum fibrinogen is between 1.223 % and 1.581%. A two-tailed P value of 0.000 indicates that the null hypothesis that there is no association between geometric mean fibrinogen levels and serum CRP levels can be rejected with high confidence (statistical significance of 0.05).
	4. **See table below**.
4. **Methods**: The serum fibrinogen and serum CRP values were first log-transformed (base e). The association between the geometric mean fibrinogen levels as defined by the continuous log-transformed serum CRP levels was evaluated using robust regression analysis. The Huber-White sandwich estimator was used to determine standard errors. In order to make inferences on the geometric mean of fibrinogen levels, the estimates and confidence intervals were exponentiated.

	1. The intercept obtained was 292.536. This means that when serum CRP levels are 1 mg/L (giving a log CRP=0), then the geometric mean fibrinogen level is estimated to be 292.536 mg/dL.
	2. The slope obtained is 1.11115. This means that for each integral increase of log-transformed CRP, the geometric mean fibrinogen levels are expected to increase by 11.115%.
	3. **Inference**: The regression analysis indicates that for every integral increase in log-transformed serum CRP levels, the geometric mean fibrinogen level is estimated to increase by 11.115%. 95% confidence intervals indicate that this observation would not be judged unusual if the true increase in the geometric mean of serum fibrinogen is between 10.465 % and 11.767%. A two-tailed P value of 0.000 when using statistical significance of 0.05 indicates we are able to infer an association between geometric mean fibrinogen and log transformed CRP levels with high confidence.
	4. **See table below**

**Table 2**: Fitted fibrinogen values for questions 3-6.

|  |  |
| --- | --- |
|  | **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** | 309.266 | 295.566 | 305.108 | 292.536 |
| **2 mg/L** | 314.517 | 321.097 | 309.321 | 315.074 |
| **3 mg/L** | 319.768 | 336.031 | 313.534 | 328.258 |
| **4 mg/L** | 325.019 | 346.627 | 317.746 | 337.612 |
| **6 mg/L** | 335.521 | 361.562 | 326.171 | 350.796 |
| **8 mg/L** | 346.023 | 372.158 | 334.596 | 360.150 |
| **9 mg/L** | 351.274 | 376.496 | 338.809 | 363.980 |
| **12 mg/L** | 367.027 | 387.093 | 351.447 | 373.334 |

1. **Table 3**: Comparisons of Fitted Values

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Comparisons across CRP level** | **Problem 3: mean** | **Problem 4: mean** | **Problem 5: geometric mean** | **Problem 6: geometric mean** |
| ***Differences*** |
| **2 mg/L – 1 mg/L** | 5.251 | 25.531 | 4.213 | 22.538 |
| **3 mg/L – 2 mg/L** | 5.251 | 14.934 | 4.213 | 13.184 |
| **4 mg/L – 1 mg/L** | 15.753 | 51.061 | 12.638 | 45.076 |
| **4 mg/L – 2 mg/L** | 10.502 | 25.530 | 8.425 | 22.538 |
| **6 mg/L – 3 mg/L** | 15.753 | 25.531 | 12.637 | 22.538 |
| **8 mg/L – 4 mg/L** | 21.004 | 25.531 | 16.850 | 22.538 |
| **9 mg/L – 6 mg/L** | 15.753 | 14.934 | 12.638 | 13.184 |
| **9 mg/L – 8 mg/L** | 5.251 | 4.338 | 4.213 | 3.830 |
| **12 mg/L – 6 mg/L** | 31.506 | 25.531 | 25.276 | 22.538 |
| ***Ratios*** |
| **2 mg/L / 1 mg/L** | 1.017 | 1.086 | 1.014 | 1.077 |
| **3 mg/L / 2 mg/L** | 1.017 | 1.137 | 1.014 | 1.042 |
| **4 mg/L / 1 mg/L** | 1.051 | 1.173 | 1.040 | 1.154 |
| **4 mg/L / 2 mg/L** | 1.033 | 1.080 | 1.027 | 1.071 |
| **6 mg/L / 3 mg/L** | 1.049 | 1.076 | 1.040 | 1.069 |
| **8 mg/L / 4 mg/L** | 1.065 | 1.074 | 1.053 | 1.067 |
| **9 mg/L / 6 mg/L** | 1.049 | 1.041 | 1.040 | 1.038 |
| **9 mg/L / 8 mg/L** | 1.017 | 1.012 | 1.014 | 1.011 |
| **12 mg/L / 6 mg/L** | 1.094 | 1.071 | 1.077 | 1.064 |

* 1. The analyses of problems 3 and 5 gave constant differences in the fitted values. In particular, the fibrinogen levels were estimated to increase by the same increment as CRP values increase by 1. From the table, we can see that from CRP=1 mg/L to CRP=2 mg/L, fibrinogen increases by 5.251 mg/dL and 4.123 mg/dL in the analyses performed in problems 3 and 5, respectively, as CRP levels increase by 1 mg/L.
	2. Again, analyses in problems 3 and 5 gave constant ratios of fitted values. Specifically, in problem 3, the ratios were constant when comparing CRP=x to CRP=x+1 and gave a ratio of 1.017. In problem 5, this ratio was 1.014. For any value of x, the ratios were the same when comparing with a CRP value of x+c. For example, CRP=6 mg/L and CRP=3 mg/L (c=3), the ratio is 1.049 for problem 3 and 1.040 for problem 5. The ratio applies when CRP=9 and CRP=6, also when c=3.
	3. The analyses of problems 4 and 6 gave constant differences in the fitted values when comparing groups with a fold-increase in CRP levels. For example, when comparing CRP values that differed by 2-fold, the analysis in problem 4 gave a difference of 25.530 mg/dL while problem 6 yielded a difference of 22.538 mg/dL.
	4. None of the analyses appear to give constant ratios when comparing two groups that differed by fold-increases in serum-CRP levels.
1. Ideally analyses that are to be performed should be decided prior to actual data analysis, i.e. that the data should not be used to inform which analyses should be done (in other words, don’t choose an analytical method based on which will give the best association). However, if you expect that you are looking for ABSOLUTE differences in your data, then the analyses done in problems 3 and 5 would be more appropriate. These are the analyses that don not use log-transformed data (or are exponentiated back in the case of the geometric mean). If you are looking for relative differences in you data, then the analyses done, where the predictor of interest is log transformed as in problems 4 and 6 may be more appropriate.