**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #2**

January 13, 2015

ID: 2811

1.

Methods: Divided CRP into three groups, below 1mg/L (low risk), 1-2mg/L (average risk) and above 3 mg/L (high risk) and provided descriptive statistics by blood fibrinogen (fib).

Inference: Of the 426 people with low risk CRP, they had a sample mean fib of 279.8146 mg/dl and sample standard deviation of 50.5363 mg/dl. 3306 people had an average risk CRP, with a sample mean fib of 311.0532 mg/dl and sample standard deviation of 53.1751 mg/dl. And of the 1167 people with high risk CRP the sample mean fib 372.6821 mg/dl and sample standard deviation of 80.9594 mg/dl.

* 1. Methods: Performed a two sample t-test assuming equal variances of mean fib is equal between those with history of CVD and those without a history of CVD. Type 1 error 0.05.

Inference: Of the 3,791 people with no history of CVD the sample mean fib was 319.574 mg/dl (SD: 64.7637 mg/dl). And of the 1,124 people with a history of CVD the sample mean fib was 334.4591 mg/dl (SD: 74.0626 mg/dl). The mean fib for those without a history of CVD was 14.8851 mg/dl lower than those with a history of CVD and the SD was 14.36% higher for those with a history of CVD. With 95% confidence we would expect the true population mean difference to be between 10.424 mg/dl higher to19.346 mg/dl higher for those with a history of CVD. We reject the null hypothesis that the two means fib are the same in favor of the mean fib for those with CVD is higher than those without CVD (P<0.0001).

* 1. Methods: Performed a linear regression assuming common variance between the two groups.

Inference: The mean fib for each group CVD and non CVD is the same, with the same mean difference of 14.8851 mg/dl. The standard error and 95% CI is also the same. The t statistic is also the same.

* 1. Methods: Performed a two sample t-test not assuming equal variances of mean fib is equal between those with history of CVD and those without a history of CVD. Type 1 error 0.05.

Inference: The mean fib for each group CVD and non CVD is the same, with the same mean difference of 14.8851 mg/dl. The combined SE is slightly larger. SE of equal variances was 2.275 while the SE for unequal variances is 2.447. And the 95% confidence interval is wider for the group with unequal variances. Here, we would expect with 95% confidence that the population mean difference to be between 10.086 mg/dl higher to 19.684 mg/dl higher in people with a history of CVD. We would still reject the null hypothesis in favor of the alternative that the mean fib is unequal and that mean fib is higher for those with CVD. The t statistic differs though from 6.54 in the previous sections to 6.08 here.

* 1. Methods: Performed a linear regression assuming unequal variance between the two groups.

Inference: The mean fib for each group CVD and non CVD is the same, with the same mean difference of 14.8851 mg/dl. The standard error is the same as section c the t test run without assuming equal variances. The 95% CI is different. Here it is between 10.089 mg/dl higher to 19.681 mg/dl higher for those with a history of CVD. The SE also differs here. It is 2.44629 compared to 2.446739 in the t test. The t statistic is the same as part c.

* 1. Part a uses the assumption of equal variances and pools the variance (averages) to determine a test statistic and p value. The t test uses variance as the degrees of freedom and so this assumption will provide difference results from part c where the variances not the same. The degrees of freedom used for part a is 4913 while for part c is 1664.57. This will cause the crucial value to be larger for part c. This in turn is why the CI was wider for part c. This is also affected by the larger SE for non-pooled variances in section c.
1. 1. Intercept: 304.0152 mg/dl. At CRP of zero, the fib level is expected to be this estimate.
	2. Slope: 5.3509 /10dl. For every one point increase of CRP the fib level is expected to increase by this estimate.
	3. As CRP increases the fib is higher by 5.25 mg/dl. With 95% confidence we would expect the population to be between 4.603 and 5.897 mg/dl higher as CRP increases. We reject the null hypothesis that there is no linear trend (P<0.0001).
	4. Intercept: 295.5663 mg/dl. At CRP of zero, the fib level is expected to be this estimate.
	5. Slope: 36.8331. For every one point increase of log CRP the fib level is expected to increase by this estimate.
	6. As CRP increases, the fib is higher by 36.833. With 95% confidence we would expect the population to be between 34.577 and 39.088 higher as CRP increases. We reject the null hypothesis that there is no linear trend (P<0.0001).
2. Repeat problem 3, except perform a statistical analysis evaluating an association between the geometric mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, untransformed random variable.
	1. Intercept: 5.7067 e to the mg/dl. At CRP of zero, the geometric mean fib level is expected to be this estimate.
	2. Slope: 0.0139. For every one point increase of CRP the log fib level is expected to increase by this estimate.
	3. As CRP increases, the log fib is higher by 0.0139. With 95% confidence we would expect the log fib to be between 0.0121 and 0.0157 higher. We reject the null hypothesis that there is no linear trend (P<0.0001)
3. Repeat problem 3, except perform a statistical analysis evaluating an association between the geometric mean fibrinogen across groups defined by CRP, modeling CRP as a continuous, log transformed random variable. (For the purpose of this problem in this homework, replace all observations of CRP=0 with CRP=0.5.)
	1. Intercept: 5.6786 e to the mg/dl. At CRP of zero, the geometric mean fib level is expected to be this estimate.
	2. Slope: 0.1054. For every one point increase of log CRP the log fib level is expected to increase by this estimate.
	3. As log CRP increases the log fib is higher by 0.1054. With 95% confidence we would expect the log fib to be between 0.995 and .111 higher. We reject the null hypothesis that there is no linear trend (P<0.0001)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1 |   |   |   |   |
| CRP | 3 fib arithmetic mean | 4 fib arithmetic mean | 5 geometric mean | 6 geometric mean |
| 1 | 304.0152 | 295.5663 | 5.7067 | 5.6786 |
| 2 | 307.6548 | 321.0971 | 5.7163 | 5.7517 |
| 3 | 309.7839 | 336.0316 | 5.7220 | 5.7944 |
| 4 | 311.2945 | 346.6278 | 5.7260 | 5.8247 |
| 5 | 312.4662 | 354.8469 | 5.7291 | 5.8482 |
| 6 | 313.4235 | 361.5624 | 5.7316 | 5.8675 |
| 8 | 314.9341 | 372.1586 | 5.7356 | 5.8978 |
| 9 | 315.5526 | 376.4969 | 5.7372 | 5.9102 |
| 12 | 317.0632 | 387.0932 | 5.7412 | 5.9405 |

1.

**Table 2**: Example of possible display of comparisons of fitted values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2 | 3 | 4 | 5 | 6 |
| 2-1 | 3.6396 | 25.5308 | 0.0096 | 0.0731 |
| 3-2 | 2.1291 | 14.9345 | 0.0056 | 0.0427 |
| 4-1 | 7.2793 | 51.0615 | 0.0193 | 0.1461 |
| 4-2 | 3.6396 | 25.5308 | 0.0096 | 0.0731 |
| 6-3 | 3.6396 | 25.5308 | 0.0096 | 0.0731 |
| 8-4 | 3.6396 | 25.5308 | 0.0096 | 0.0731 |
| 9-6 | 2.1291 | 14.9345 | 0.0056 | 0.0427 |
| 9-8 | 0.6185 | 4.3383 | 0.0016 | 0.0124 |
| 12-6 | 3.6396 | 25.5308 | 0.0096 | 0.0731 |
|   |   |   |   |   |
| Ratios |   |   |   |   |
| 2/1 | 1.0120 | 1.0864 | 1.0017 | 1.0129 |
| 3/2 | 1.0069 | 1.0465 | 1.0010 | 1.0074 |
| 4/1 | 1.0239 | 1.1728 | 1.0034 | 1.0257 |
| 6/3 | 1.0117 | 1.0760 | 1.0017 | 1.0126 |
| 8/4 | 1.0117 | 1.0737 | 1.0017 | 1.0125 |
| 9/6 | 1.0068 | 1.0413 | 1.0010 | 1.0073 |
| 9/8 | 1.0020 | 1.0117 | 1.0003 | 1.0021 |
| 12/6 | 1.0116 | 1.0706 | 1.0017 | 1.0125 |

1. 1. Question 4.
	2. Question 3.
	3. Question 6.
	4. Question 5.
2.

It depends on scientifically what is happening between fibrinogen and CRP. Biological data is often computed on the log scale and therefore the geometric means are likely the best potential analysis to investigate. This allows for more precision at the level that these markers are often measured. First it would be important to investigate what you know about fib and CRP in the body. We know that a CRP over 3 is considered high risk, but with each point of increase is the risk of CRP increase?