130/195

Homework #2

20 January 2015

1. Methods: Descriptive statistics are displayed in both tabular and graphical formats. In Table 1, mean and standard deviation of fibrinogen levels are presented by serum C reactive protein (CRP) level and previous history of cardiovascular disease (CVD). CRP level was defined as low, intermediate, or high based on the Mayo Clinic recommendations for risk of CVD. In the figure below, points are color-coded by history of CVD (blue for those with prior CVD and orange for those without). Lowess smooths are also displayed, both for groups defined by prior CVD and for the combined sample. Subjects missing CRP, fibrinogen, or previous history of disease variables were excluded.

Results: Serum CRP measurements were missing for 68 subjects (18 with prior history of CVD, 49 without prior history of CVD, 1 with a missing value for previous CVD). Fibrinogen measurements were missing for 86 subjects (25 with a prior history of CVD, 60 without a prior history of CVD, and 1 with a missing value). The 4899 subjects with non-missing CRP, fibrinogen, and history of CVD are included in both table 1 and the figure below. It appears that those with a prior history of CVD tend to have a slightly higher mean fibrinogen level across CRP levels. Higher CRP level also appears to correspond to a higher mean fibrinogen level, regardless of disease history. The consistent upward trend in mean fibrinogen across increasing CRP across both disease categories suggests that prior history of CVD is not an effect modifier and that there is a potential association between CRP and fibrinogen. Similarities in the lowess smooth by disease group also do not suggest any effect modification and the upward slope of those lowess smooths suggests a potential positive association between CRP and fibrinogen. Good!

**Table 1**. Descriptive statistics of Fibrinogen Levels by CRP level and History of CVD.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Fibrinogen Levelsa (mg/dL)** | | |
| CRP Level | **Prior history of CVD** | **No prior history of CVD** | **Overall** |
| Lowb | 290 (57.9), n=78 | 277 (48.5), n=348 | 280 (50.5), n=426 |
| Intermediateb | 315 (55.6), n=709 | 310 (52.5), n=2597 | 311 (53.2), n=3306 |
| Highb | 386 (84.5), n=335 | 367 (78.9), n=832 | 373 (81.0), n=1167 |
| Overall | 334 (74.1), n=1122 | 320 (64.8), n=3777 | 323 (67.4), n=4899 |

aValues shown are mean (SD), count; bLow CRP level is <3 mg/L, intermediate CRP level is 1-3 mg/L, and high CRP level is >3 mg/L.



. – 1 point deducted for absent Y axis label

2. a. Methods: Mean fibrinogen levels were compared between those subjects who had a prior history of CVD and those who did not using a t-test assuming equal variances. 95% confidence intervals were also calculated assuming equal variances. Those subjects missing fibrinogen levels or history of disease were not included in this analysis.

Results: 4915 subjects were included in this analysis. The sample mean fibrinogen level in those with a prior history of disease (N=?) was 334.5 mg/dL and the sample standard deviation was 74.06 mg/dL. The sample mean fibrinogen level in those without a prior history of CVD (N=?) was 319.6 mg/dL and the sample standard deviation was 64.76 mg/dL. The estimated mean fibrinogen is 14.89 mg/dL higher in subjects with a history of CVD. Given the 95% confidence interval, this value would not be considered unusual if the true difference in means is between 10.42 and 19.35 mg/dL. A two-sided p-value < 0.0001 indicates we can reject the null hypothesis of no difference in the mean of fibrinogen levels in these two groups. – 1 deducted not detailing samples size for each group

b. Methods: Classical linear regression was used to evaluate the association between mean fibrinogen level and history of CVD. Robust standard errors were not used, so this regression analysis assumes equal variances. Subjects missing fibrinogen or history of disease were excluded.-1

Results: The estimate for the slope in this regression analysis is 14.89 mg/dL, which is exactly the same as the difference in means from part a. The coefficient for the intercept is the same as the sample mean for the group coded as zero (without a history of disease), 319.6 mg/dL. The 95% confidence intervals and the two-sided p-value are also the same (95% CI: 10.42, 19.35; p<0.0005). Based on a two-sided p-value<0.0005 we reject the null hypothesis of no linear trend in mean fibrinogen levels across history of CVD . – 2 t statistic and SE missing

c. Methods: Mean fibrinogen levels were compared between those subjects who had a prior history of CVD and those who did not using a t-test that allows for unequal variances. 95% confidence intervals were also calculated allowing for equal variances. Those subjects missing fibrinogen levels or history of disease were excluded from this analysis.

Results: 4915 subjects were included in this analysis. The sample mean fibrinogen level in those with a prior history of disease was 334.5 mg/dL and the sample standard deviation was 74.06 mg/dL. The sample mean fibrinogen level in those without a prior history of CVD was 319.6 mg/dL and the sample standard deviation was 64.76 mg/dL. The estimated mean fibrinogen is 14.89 mg/dL higher in subjects with a history of CVD. Given the 95% confidence interval, this value would not be considered unusual if the true difference in means is between 10.09 and 19.68 mg/dL. A two-sided p-value < 0.0001 indicates that at a 0.05 level of statistical significance we can reject the null hypothesis of no difference in the mean fibrinogen levels in these two groups (-1 as per Scott’s key: in favor of the hypothesis that the mean fibrinogen is higher in patients with prior CVD.)

d. Methods: Linear regression with robust standard errors allowing for unequal variances (Huber-White sandwich estimator) was used to evaluate the association between mean fibrinogen level and history of CVD. Subjects missing fibrinogen or history of disease were excluded.-1

Results: 4915 subjects were included in this analysis. The estimate for the slope coefficient was exactly the same as the estimate of difference in means in part c (14.89 mg/dL). The intercept from linear regression is exactly equal to the sample mean in the group with no prior history of CVD (319.6 mg/dL). The 95% confidence intervals (10.09, 19.68) are also approximately the same as those calculated using the t-test allowing for unequal variances and thus the interpretation is the same. Based on a two-sided p-value<0.0005 we reject the null hypothesis of no linear trend in mean fibrinogen levels across history of CVD. – 2 t statistic and SE missing

e. In part a, we used a t test presuming equal variances and in part c we used a t test that allowed for unequal variances. While the estimated difference in means from these two tests will be the same, the standard error of that estimate is different. The group with the smaller sample size (history of CVD, n=1124) has a larger variance, so the t test presuming equal variances will be anti-conservative with p values that are too small and a confidence interval that is too narrow.

Allowing for unequal variances will lead to a wider CI and a larger p value, as well as a more negative t statistic. Good!

3. a. The intercept (β0) represents the estimated mean fibrinogen level among those with a CRP of zero, and is estimated to be 304 mg/dL

b. The estimated slope (β1) is 5.25 and represents the estimated (-1) difference in mean fibrinogen level among those who differ in CRP by 1 mg/L.

c. Methods: Linear regression with robust standard errors was used to estimate the difference in mean fibrinogen level for every 1 mg/L increase in CRP. A two-sided p value was used and alpha was 0.05. Those subjects missing fibrinogen or CRP were excluded from this analysis.

Results: 4899 subjects were included in this analysis. The estimated slope (β1) is 5.25 and represents the estimated difference in mean fibrinogen level among those who differ in CRP by 1 mg/L. A 95% confidence interval suggests that this observed 5.25 mg/dL difference is not unusual if the true difference in means is between 4.60 and 5.90 mg/dL. A two-sided p-value<0.0005 indicates that we can reject the null hypothesis of no linear trend in mean fibrinogen across CRP levels.

d. See table 2 below

4. a. The intercept (β0) represents the estimated mean fibrinogen level among those with a log-transformed CRP of zero, and is estimated to be 296. (-5)

b. The estimated slope (β1) is 36.8 mg/dL and represents the estimated difference in mean fibrinogen level among those who differ in log of CRP by 1. -5

c. Methods: Linear regression with robust standard errors was used to evaluate the association between mean fibrinogen across groups defined by CRP, with CRP log transformed. Observations where CRP was equal to zero were replaced with 0.5. Those subjects missing fibrinogen or CRP were excluded from this analysis.

Results: Suggest you check page 10 of Scott’s reading "Use of Ratios and Logarithms in Statistical Regression Models" 4899 subjects were included in this analysis. The estimated slope (β1) is 36.8 and represents the estimated difference in mean fibrinogen level among those who differ in log of CRP by 1 expressed as XX fold difference (as per logarithm base i.e. 2.718) between groups.since you used natural log. The fold difference of natural log (2.718) is not meaningful and will need to be converted to understandable logs e.g. log 10 “10 fold” or log 2 “2 fold”. The conversion =X (36.8) estimate from regression model\* log 10/ log (2.718) OR X estimate from regression model\* log 2/ log (2.718) for 10 fold or 2 fold interpretation. A 95% confidence interval suggests that this observed 36.8 difference is not unusual if the true difference in means is between 34.6 and 39.1 per XXX fold proportional/relative difference between two groups with group with higher CRP having higher fibrinogen level. A two-sided p-value<0.0005 indicates that we can reject the null hypothesis of no linear trend in mean fibrinogen across log of CRP levels. -5

d. See table 2 below

5. a. The intercept (β0) represents the estimated log of the geometric mean fibrinogen level among those with a CRP of zero, and is estimated to be 5.71. -5

b. The estimated slope (β1) is 0.0139 and represents the estimated difference in log of geometric mean fibrinogen level among those who differ in CRP by 1 mg/L.-5

c. Methods: Linear regression with robust standard errors was used to evaluate the association between the geometric mean fibrinogen across groups defined by CRP, with CRP as a continuous, untransformed variable. Those subjects missing fibrinogen or CRP were excluded from this analysis.

Results: 4899 subjects were included in this analysis. The estimated difference in geometric mean fibrinogen level among those who differ in CRP by 1 mg/L is 1.014. A 95% confidence interval suggests that this observed 5.25 mg/dL difference is not unusual if the true difference in means is between 1.012 and 1.016 mg/dL. A two-sided p-value<0.0005 indicates that we can reject the null hypothesis of no linear trend in geometric mean fibrinogen across CRP levels. - 5

d. See table 2 below

6. a. The intercept (β0) represents the estimated log of the geometric mean fibrinogen level among those with a log-transformed CRP of zero, and is estimated to be 5.68.-5

b. The estimated slope (β1) is 0.105 and represents the estimated difference in log of geometric mean fibrinogen level among those who differ in log of CRP by 1.-5

c. Methods: Linear regression with robust standard errors was used to evaluate the association between the geometric mean fibrinogen across groups defined by log transformed CRP. Observations where CRP was equal to zero were replaced with 0.5 before log transforming CRP. Those subjects missing fibrinogen or CRP were excluded from this analysis.

Results: 4899 subjects were included in this analysis. The estimated difference in geometric mean fibrinogen level among those who differ in log of CRP by 1 is 1.11. A 95% confidence interval suggests that this difference is not unusual if the true difference in means is between 1.10 and 1.12 mg/dL. A two-sided p-value<0.0005 indicates that we can reject the null hypothesis of no linear trend in geometric mean fibrinogen across log transformed CRP levels. - 5

d. See table 2 below

**Table 2**. Fitted values of fibrinogen from regression models fitted in problems 3-6.

Fitted equations missing for all models (- 4, method not clear how these figures were derived)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** | | | |
| **CRP (mg/L)** | **Problem 3:**  **(Mean)** | **Problem 4:**  **(Mean)** | **Problem 5:**  **(Geometric Mn)** | **Problem 6:**  **(Geometric Mn)** |
| 1 | 309 | 296 | 305 | 293 |
| 2 | 315 | 321 | 309 | 315 |
| 3 | 320 | 336 | 314 | 328 |
| 4 | 325 | 347 | 318 | 339 |
| 6 | 336 | 362 | 327 | 353 |
| 8 | 346 | 372 | 336 | 364 |
| 9 | 351 | 376 | 341 | 369 |
| 12 | 367 | 387 | 356 | 380 |

7. Table 3 displays comparisons (both differences and ratios) of a selection of fitted values.

**Table 3**. Comparison of fitted values from regression models in problems 3-6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comparisons across CRP level** | **Problem 3: (mean)** | **Problem 4: (mean)** | **Problem 5: (Geom. Mn.)** | **Problem 6: (Geom. Mn.)** |
| *Differences* | | | | |
| 2 mg/L - 1 mg/L | 5.25 | 25.5 | 4.28 | 22.2 |
| 3 mg/L - 2 mg/L | 5.25 | 14.9 | 4.34 | 13.7 |
| 4 mg/L - 1 mg/L | 15.8 | 51.1 | 13.0 | 46.0 |
| 4 mg/L - 2 mg/L | 10.5 | 25.5 | 8.73 | 23.9 |
| 6 mg/L - 3 mg/L | 15.8 | 25.5 | 13.4 | 24.9 |
| 8 mg/L – 4 mg/L | 21.0 | 25.5 | 18.2 | 25.7 |
| 9 mg/L – 6 mg/L | 15.8 | 14.9 | 13.9 | 15.4 |
| 9 mg/L – 8 mg/L | 5.25 | 4.34 | 4.71 | 4.55 |
| 12 mg/L – 6 mg/L | 31.5 | 25.5 | 28.5 | 26.8 |
| *Ratios* | | | | |
| 2 mg/L / 1 mg/L | 1.02 | 1.09 | 1.01 | 1.08 |
| 3 mg/L / 2 mg/L | 1.02 | 1.05 | 1.01 | 1.04 |
| 4 mg/L / 1 mg/L | 1.05 | 1.17 | 1.04 | 1.16 |
| 4 mg/L / 2 mg/L | 1.03 | 1.08 | 1.03 | 1.08 |
| 6 mg/L / 3 mg/L | 1.05 | 1.08 | 1.04 | 1.08 |
| 8 mg/L / 4 mg/L | 1.06 | 1.07 | 1.06 | 1.08 |
| 9 mg/L / 6 mg/L | 1.05 | 1.04 | 1.04 | 1.04 |
| 9 mg/L / 8 mg/L | 1.02 | 1.01 | 1.01 | 1.01 |
| 12 mg/L / 6 mg/L | 1.09 | 1.07 | 1.09 | 1.08 |

8. a. When comparing two groups that differed by an absolute increase in c units in CRP levels, the analysis from Problem 3 gave constant differences: when c=1 the difference in means was equal to 5.25, and when c=3 the difference in means was equal to 15.8. – 2 “for not explicitly providing all those similar paired comparisons from the table”

b.When comparing two groups that differed by an absolute increase in c units in CRP levels, the analyses from Problem 3 and Problem 5 gave constant ratios of the fitted values. For problem 3, when c=1 the ratio of means was 1.02 and when c=3 the ratio of means was 1.05. For problem 5, when c=1 the ratio of geometric means was 1.01 and when c=3 the ratio of geometric means was 1.04. – 5 figures do not match, it is only problem 5 that is the correct answer

c.When comparing two groups that differ by a relative c-fold increase in CRP levels, the analysis from Problem 4 gave constant differences in the fitted values. For problem 4, when c=2 the difference in means was 25.5 and when c=1.5 the difference in means was 14.9. – 2 “for not explicitly providing all those similar paired comparisons from the table”

d. When comparing two groups that differ by a relative c-fold increase in CRP levels, the analysis from Problem 6 gave constant ratios of the fitted values. For problem 6, when c=2 the ratio of geometric means was 1.08 and when c=1.5 the ratio of geometric means was 1.04. – 2 “for not explicitly providing all those similar paired comparisons from the table”

9. The decision about which of these four potential analyses should be used to investigate associations between fibrinogen and CRP should be made *a priori*, before looking at the data. A multiplicative model using geometric means might be preferable for modeling fibrinogen and CRP due to the biochemistry of these blood proteins, though modeling the means rather than geometric means is simpler and probably more widely understood. Given this, a priori I would have chosen to use linear regression with robust standard errors using both the untransformed predictor (CRP) and untransformed outcome (fibrinogen), with the mean as the summary measure, as we did in Problem 3.