**ID: 2731
HW02
BIOST 518 – Winter 2015**

Question 1

**Methods:** In order to provide baseline descriptives for these data, we categorized levels of c-reactive protein (crp) with “high crp” defined as crp > 3 mg/L per the Mayo Clinic Guidelines for cardiovascular risk. Crp was collected as a continuous variable, so a categorical was created to represent “high” crp status, “average” crp status (1-3 mg/L), and “low” crp status (<1 mg/L). Within the below table the mean, standard deviation (SD), and range are shown for continuous variables (age, body mass index (BMI), cholesterol, and serum fibrinogen) while percentages are shown for binary variables (male, smoking history, and prior cardiovascular disease (CVD)). All computing was done with Stata version 13.1.

In order to investigate potential confounders, we stratified the overall sample by history of prior CVD which includes previous angina, myocardial infarction (MI), transient ischemic attack (TIA), and/or stroke. Two-way scatters for crp and serum fibrinogen were fitted with the least-squares line as well as the Lowess smoothed curve.

**Results:**  Data was available for 5000 participants; however, 101 participants had missing serum fibrinogen and/or crp measures. These individuals were not included in the subsequent analyses.

Among 4899 participants, 426 had low crp, 1167 had high crp, and 3306 had average crp. Table 1 provides descriptive statistics within these groups and for all participants.

On average, individuals with high crp were less likely to be females and more likely to be smokers, and have a history of CVD. These individuals also showed trends for higher BMI and serum fibrinogen level than individuals with lower crp levels. The mean crp was 372.68 (mg/dL) among those with high crp compared with 311.05 and 279.81 mg/dL for participants with average and low crp, respectively.

The figures below depict the relationship between serum fibrinogen and crp overall and stratified by the history of prior CVD. As evidenced by some differences in Lowess smooths, there does appear to be some effect of prior CVD on the relationship between crp and serum fibrinogen.

|  |  |  |
| --- | --- | --- |
|   |   | Level of crp |
|   | Overall | Low crp  | Average crp | High crp |
| Number | 4899 | 426 | 3306 | 1167 |
| Age | 72.80 (5.56, 65-100) | 73.47 (5.81, 65-94) | 72.72 (5.52, 65-100) | 72.76 (5.57, 65-93) |
| Male (%) | 41.80% | 45.31% | 43.10% | 36.85% |
| BMI | 26.66 (4.72, 14.7-58.8) | 23.80 (3.63, 15.6-38.6) | 26.39 (4.31, 14.7-53.2) | 28.46 (5.45, 15.3-58.8) |
| Cholesterol  | 211.70 (39.22, 73-430) | 205.95 (40.55, 109-407) | 212.84 (38.53, 73-363) | 210.58 (40.45, 97-430) |
| Smoker (%) | 12.12% | 9.65% | 10.94% | 16.37% |
| History of CVD | 22.90% | 18.31% | 21.45% | 28.71% |
| Fibrinogen level (mg/dL) | 323.02 (67.35, 109-872) | 279.81 (50.55, 172-540) | 311.05 (53.18, 109-592) | 372.68 (80.96, 132-872) |
| \*All baseline characteristics are mean (SD, range) unless otherwise indicated. |
|  |



Question 2

**Methods (All parts):** In order to evaluate the relationship between prior CVD and our outcome of interest, serum fibrinogen, we compared the mean fibrinogen levels between participants with and without history of prior CVD. Two sample t tests were performed, which provided an estimate of the difference in fibrinogen as well as the 95% confidence intervals (CI). We tested both assumptions of equal variances for groups and unequal variances (using Satterthwaite degrees of freedom). We used a significance level of 0.05 for these statistical tests.

Simple linear regression analyses were undertaken, where fibrinogen was the outcome of interest and prior CVD was a binary predictor of interest.

**Results (for Part A and Part B):** A total of 4915 participants had valid fibrinogen measures and data on CVD history. Among the 3791 participants with no history of CVD, the mean fibrinogen (SD) was 319.57 (67.76) mg/dL. Participants with history of CVD had a mean fibrinogen level of 334.46 (74.06) mg/dL. With 95% confidence, the observed data are consistent with participants with a history of CVD having a mean serum fibrinogen between 10.42 and 19.35 mg/dL higher than with no history of CVD (difference in means: 14.89 mg/dL). In a two sample t test assuming equal variance, the difference was statistically significant with a two-sided p-value of 0.0000. We thus rejected the null hypothesis of no difference in the mean fibrinogen between these two groups. There is evidence to suggest that the distributions of fibrinogen level are different between those with and without a history of CVD.

This relationship could also be explored in linear regression analysis. The use of use of classical linear regression gives inference that is the same as a two sample t test assuming equal variances. The estimate of the difference in means (14.8851 mg/dL) is the same as the estimate given from linear regression (14.8851 mg/dL). The variances are also the same (2.2756).

**Results (for Part C and D):** A total of 4915 participants had valid fibrinogen measures and data on CVD history. Among the 3791 participants with no history of CVD, the mean fibrinogen (SD) was 319.57 (67.76) mg/dL. Participants with history of CVD had a mean fibrinogen level of 334.46 (74.06) mg/dL. With 95% confidence, the observed data are consistent with participants with a history of CVD having a mean serum fibrinogen between 10.09 and 19.68 mg/dL higher than with no history of CVD (difference in means: 14.89 mg/dL). In a two sample t test assuming unequal variance, the difference was statistically significant with a two-sided p-value of 0.0000. We thus rejected the null hypothesis of no difference in the mean fibrinogen between these two groups. There is evidence to suggest that the distributions of fibrinogen level are different between those with and without a history of CVD.

This relationship could also be explored in linear regression analysis using robust standard errors. Linear regression assuming unequal variance gives inference that ALMOST the same as a two sample t test assuming unequal variances. The estimate of the difference in means (14.8851 mg/dL) is the same as the estimate given from linear regression (14.8851 mg/dL). The t test’s standard error for the difference in means (2.4467) is slightly different than the regression standard error for the slope (2.4463). The confidence interval from the t test is slightly wider than that from regression (10.0861-19.6841 and 10.08926-19.68091). Difference in CIs is attributed to both the difference in variance and the differences in the critical value used in analyses. The t test used a critical value based on Satterthwaite's degrees of freedom = 1664.57 while the regression used a criterial value based on 4913 degrees of freedom.

**Part E.** The t tests assuming equal and unequal variances will have the same point estimate for the difference in means but will differ in their SE, CI, and, possibly, their p-values. The t test that assumes equal variance is comparing two groups where the group with prior CVD has a smaller sample size and greater variance than the group with no prior CVD. Thus, we would expect that the reported p-value and confidence interval will be smaller than that from the t test assuming unequal variance. So, we expect the p-value could be greater than p=0.0000 and the CI will be wider than 8.9224 mg/dL (distance between upper and lower bounds). If the CI is wider we also expect that the t-statistic will be larger than -6.5412 and the SE will be larger than 2.2756. Thus, the t test assuming equal variance provides for anti-conservative inference, that is, the p-value is too small and the CI is too narrow.

Question 3

**Methods (All parts):** In order to evaluate the relationship between a predictor of interest, crp level, and outcome of interest, serum fibrinogen, we performed linear regression with robust standard errors. Participants with missing data for crp or serum fibrinogen were excluded from the analysis.

**Results (Part A):**  From linear regression analysis on serum fibrinogen using Huber-White estimates of the standard error, the estimated intercept is 304 mg/dL. Thus, among elderly individuals with a mean crp value that is equal to 0 mg/L, the estimated mean value of serum fibrinogen would be 304 mg/dL. With 95% confidence, the observed data are consistent with a mean serum fibrinogen between 302 and 307 mg/dL.

**Results (Part B):** From linear regression analysis on serum fibrinogen using Huber-White estimates of the standard error, the estimated slope is 5.25 mg/dL. Thus, for every one mg/L difference in crp between two groups of elderly individuals, the mean serum fibrinogen would be 5.25 mg/dL higher in the population with higher crp. With 95% confidence, the observed data are consistent with a mean serum fibrinogen between 4.60 and 5.90 mg/dL higher for every one 1 mg/L difference in crp level.

**Results (Part C):** A total of 4899 participants had valid measures of crp and serum fibrinogen. Linear regression analysis of serum fibrinogen and crp using Huber-White estimates of the standard error suggest that for every one mg/L difference in crp between two groups of elderly individuals, the mean serum fibrinogen would be 5.25 mg/dL higher in the population with higher crp. With 95% confidence, the observed data are consistent with a mean serum fibrinogen between 4.60 and 5.90 mg/dL higher for every one 1 mg/L difference in crp level. Based on a two-sided p-value of 0.0000, we reject the null hypothesis for no linear relationship between serum fibrinogen and crp.

**Results (Part D):**

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Crp Level (mg/L)** | **Question 3** | **Question 4** | **Question 5** | **Question 6** |
| Summary measure | mean | mean | geometric mean | geometric mean |
| 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 |

Question 4

**Methods (All parts):** In order to evaluate the relationship between a predictor of interest, the log transformed crp level, and outcome of interest, serum fibrinogen, we performed linear regression with robust standard errors. Crp was collected as integer data, with a crp level of 1 mg/L representing the lower limit of detection. In order to account for participants with crp 0 mg/L (N=21) in log transformation, we created a new crp value such that all zeros were assigned half of the lower limit of detection, that is 0.5 mg/L, and all other values remained the same. Participants with missing data for crp or serum fibrinogen were excluded from the analysis.

**Results (Part A):**  From linear regression analysis on serum fibrinogen using Huber-White estimates of the standard error, the estimated intercept is 296 mg/dL. Thus, among elderly individuals with a mean log-transformed crp that is equal to 0 units, the estimated mean value of serum fibrinogen would be 296 mg/dL. With 95% confidence, the observed data are consistent with a mean serum fibrinogen between 294 and 297 mg/dL.

**Results (Part B):** From linear regression analysis on serum fibrinogen using Huber-White estimates of the standard error, the estimated slope is 36.8 mg/dL. Thus, for every one-unit difference in log-transformed crp between two groups of elderly individuals, the mean serum fibrinogen would be 36.8 mg/dL higher in the population with higher log-crp. With 95% confidence, the observed data are consistent with a mean serum fibrinogen between 34.6 and 39.1 mg/dL higher for every one 1-unit difference in log-transformed crp level.

**Results (Part C):** A total of 4899 participants had valid measures of crp and serum fibrinogen. Linear regression analysis of serum fibrinogen and log-transformed crp using Huber-White estimates of the standard error suggest that for every one unit difference in crp between two groups of elderly individuals, the mean serum fibrinogen would be 36.8 mg/dL higher in the population with higher crp. With 95% confidence, the observed data are consistent with a mean serum fibrinogen between 34.6 and 39.1 mg/dL higher for every one 1-unit difference in log-transformed crp level. Based on a two-sided p-value of 0.0000, we reject the null hypothesis for no linear relationship between serum fibrinogen and log-crp.

**Results (Part D):** See table above in Question 3.
Question 5

**Methods (All parts):** In order to evaluate the relationship between a predictor of interest, crp level, and outcome of interest, the log transformed serum fibrinogen, we performed linear regression with robust standard errors. Participants with missing data for crp or serum fibrinogen were excluded from the analysis.

**Results (Part A):**  From linear regression analysis on the log transformation of serum fibrinogen using Huber-White estimates of the standard error, the estimated intercept is 5.71 units. Thus, among elderly individuals with a mean crp that is equal to 0 mg/L, the estimated mean log-transformed serum fibrinogen would be 5.71. With 95% confidence, the observed data are consistent with a geometric mean log-transformed serum fibrinogen between 5.70 and 5.71 units. Back-transforming gives an intercept of 301 mg/dL serum fibrinogen.

**Results (Part B):** From linear regression analysis on the log transformation of serum fibrinogen using Huber-White estimates of the standard error, the estimated slope is 0.0139 units.. Thus, for every 1 mg/L difference in crp between two groups of elderly individuals, the geometric mean serum fibrinogen would be 0.0139 units higher in the population with higher crp. With 95% confidence, the observed data are consistent with a mean log-transformed serum fibrinogen between 0.0122 and 0.0157 higher for every one 1 mg/L difference in crp level.

**Results (Part C):** A total of 4899 participants had valid measures of crp and serum fibrinogen. Linear regression analysis of log-transformed serum fibrinogen and crp using Huber-White estimates of the standard error suggest that for every 1 mg/L difference in crp between two groups of elderly individuals, the geometric mean serum fibrinogen would be 1.40% higher in relative terms in the population with higher crp. With 95% confidence, the observed data are consistent with a geometric mean serum fibrinogen between 1.22% and 1.58% higher for every one 1 mg/L difference in crp level. Based on a two-sided p-value of 0.0000, we reject the null hypothesis for no linear relationship between log-transformed serum fibrinogen and crp.

**Results (Part D):** See table above in Question 3.
Question 6

**Methods (All parts):** In order to evaluate the relationship between a predictor of interest, the log-transformed crp level, and outcome of interest, the log transformed serum fibrinogen, we performed linear regression with robust standard errors. Participants with missing data for crp or serum fibrinogen were excluded from the analysis.

**Results (Part A):**  From linear regression analysis on the log transformation of serum fibrinogen using Huber-White estimates of the standard error, the estimated intercept is 5.68 units. Thus, among elderly individuals with a mean log-transformed crp that is equal to 0 units, the estimated mean log-transformed serum fibrinogen would be 5.68. With 95% confidence, the observed data are consistent with a geometric mean log-transformed serum fibrinogen between 5.67 and 5.68 units. Back-transforming gives an intercept of 293 mg/dL serum fibrinogen.

**Results (Part B):** From linear regression analysis on the log transformation of serum fibrinogen using Huber-White estimates of the standard error, the estimated slope is 0.105 units. Thus, for every 1 unit difference in log-transformed crp between two groups of elderly individuals, the geometric mean log-transformed serum fibrinogen would be 0.105 units higher in the population with higher log-transformed crp. With 95% confidence, the observed data are consistent with a geometric mean log-transformed serum fibrinogen between 0.100 and 0.111 higher for every 1 unit difference in log-transformed crp.

**Results (Part C):** A total of 4899 participants had valid measures of crp and serum fibrinogen. Linear regression analysis of log-transformed serum fibrinogen and log-transformed crp using Huber-White estimates of the standard error suggest that for every 10% difference in difference in crp between two groups of elderly individuals, the geometric mean fibrinogen would be 10.1% higher in the population with higher crp. With 95% confidence, the observed data are consistent with a geometric mean serum fibrinogen between 9.53% and 10.7% higher for every 10% difference in log-transformed crp. Based on a two-sided p-value of 0.0000, we reject the null hypothesis for no linear relationship between log-transformed serum fibrinogen and log-transformed crp.

**Results (Part D):** See table above in Question 3.

Question 7

|  |  |
| --- | --- |
|  | **Fitted Values for Fibrinogen (mg/dL)** |
| **Crp Level (mg/L)** | **Question 3** | **Question 4** | **Question 5** | **Question 6** |
|   | *Differences* |
| 2-1 | 5.2508 | 25.5308 | 4.2764 | 22.1702 |
| 3-2 | 5.2509 | 14.9345 | 4.3363 | 13.7396 |
| 4-1 | 15.7525 | 51.0616 | 13.01 | 46.0206 |
| 4-2 | 10.5017 | 25.5308 | 8.7336 | 23.8504 |
| 6-3 | 15.7526 | 25.5308 | 13.3772 | 24.8917 |
| 8-4 | 21.0035 | 25.5308 | 18.2135 | 25.658 |
| 9-6 | 15.7525 | 14.9346 | 13.9477 | 15.4263 |
| 9-8 | 5.2508 | 4.3383 | 4.7141 | 4.5492 |
| 12-6 | 31.5051 | 25.5308 | 28.49 | 26.7781 |
|   | *Ratios* |
| 2/1 | 1.0170 | 1.0864 | 1.0140 | 1.0758 |
| 3/2 | 1.0167 | 1.0465 | 1.0140 | 1.0437 |
| 4/1 | 1.0509 | 1.1728 | 1.0426 | 1.1573 |
| 4/2 | 1.0334 | 1.0795 | 1.0282 | 1.0758 |
| 6/3 | 1.0493 | 1.0760 | 1.0426 | 1.0758 |
| 8/4 | 1.0646 | 1.0737 | 1.0573 | 1.0758 |
| 9/6 | 1.0469 | 1.0413 | 1.0426 | 1.0437 |
| 9/8 | 1.0152 | 1.0117 | 1.0140 | 1.0125 |
| 12/6 | 1.0939 | 1.0706 | 1.0871 | 1.0758 |
|  |  |  |  |  |

Question 8

Part A. The analysis from question 3 gave the same differences in fitted values when comparing groups that differed on crp by an absolute increase in c units. Paired comparisons include:

1) 2-1 mg/L, 3-2 mg/L, and 9-8 mg/L (absolute difference of 1 unit)

2) 4-1 mg/L, 6-3 mg/L, and 9-6 mg/L (absolute difference of 3 units)

Part B. The analysis from question 6 gave the same ratios of fitted values when comparing groups that differed on crp by an absolute increase in c units. Paired comparisons include:

1) 2/1 mg/L, 4/2 mg/L, 6/3 mg/L, 8/4 mg/L, and 12/6 mg/L (ratio of 2)

2) 3/2 mg/L and 9/6 mg/L (ratio of 3/2)

Part C. The analysis from question 4 gave the same differences in fitted values when comparing groups that differed on crp by a relative c-fold increase in crp. Paired comparisons include:

1) 2-1mg/L, 4-2 mg/L, 6-3 mg/L, 8-4 mg/L, and 12-6 mg/L

2) 3-2 mg/L and 9-6 mg/L

Part D. The analysis from question 5 gave the same ratios of fitted values when comparing groups that differed on crp by a relative c-fold increase in crp. Paired comparisons include:

1) 2/1 mg/L, 3/2 mg/L, and 9/8 mg/L

2) 4/1 mg/L, 6/3 mg/L, and 9/6 mg/L

Question 9

I would base whether or not I transformed the predictor (crp) on scientific considerations, i.e. whether it is meaningful to talk about relative increases or decreases in crp rather than absolute increases. As seen below, the logcrp does not show a linear trend against crp over the range of values studied. Thus, it may be appropriate to log transform crp. Thus models from question 4 or 6 may be more appropriate.



I would base whether or not I modeled the geometric mean (that is, the log serum fibrinogen) based on the heteroscedasticity. Modeling the geometric mean of fibrinogen, as shown in the figures below, appears to reduce some of the heteroscedasticity that we saw in the original scatters. Moreover, a multiplicative model has somewhat of an easier interpretation in this case (as opposed to an additive model). For example, it is easier to talk about 10% increases in fibrinogen.

