Good job. 185/195.

1.

**Methods:** All analyses were performed in R using standard regression packages and the “psych” library with the “describe.by” function. 101 participants had missing CRP or fibrinogen and were excluded, leaving 4899 total subjects with complete data for analyses. Indicator variables were created for three defined levels of C - reactive protein (CRP <1 mg/dl, 1-3 mg/dl, and >3 mg/dl) and for prior history of cardiovascular disease (CVD). Mean fibrinogen (with standard deviation and range in parentheses) is reported for the subset of the study with previously diagnosed CVD (n=1122), those without CVD (n=3777), and the overall cohort (n=4899), stratified by the aforementioned CRP groupings. Plots of CRP by fibrinogen were made in R and had lowess lines fitted using default settings. Good. Well-written and thorough. I would explain what the default settings are for those not familiar with R.

**Results:** An approximate linear trend was observed between increasing CRP levels and increasing fibrinogen levels across participants with CVD (n=1122, mean fibrinogen of 290.23 to 314.84 to 386.29 for low, medium, and high CRP levels), participants with no prior CVD (n=3777, mean fibrinogen of 277.48 to 310.02 to 367.20 for low, medium, and high CRP levels), and the overall cohort (n=4899, mean fibrinogen of 279.81 to 311.05 to 372.68 for low, medium, and high CRP levels). We note that there is a slightly higher mean fibrinogen for each CRP subset in the group of participants with prior CVD as compared to those without prior CVD. Plots of the data show an increasing relationship between CRP and fibrinogen, while also noting the outliers in the CRP distribution.

|  |  |  |
| --- | --- | --- |
|  |  | **History of CVD** |
|  |  | **Prior CVD** **(n=1122)** | **No Prior CVD** **(n=3777)** | **Overall****(n=4899)** |
|  |  | N | Fibrinogen\* | N | Fibrinogen\* | N | Fibrinogen\* |
| **CRP Levels** | <1mg/L | 78 | 290.23 (57.93; 180-540) | 348 | 277.48 (48.52; 172-436) | 426 | 279.81 (50.55; 172-540) |
| 1-3mg/L | 709 | 314.84 (55.6; 138-592) | 2597 | 310.02 (52.46; 109-562) | 3306 | 311.05 (53.18; 109-592) |
| >3mg/L | 335 | 386.29 (84.5; 175-695) | 832 | 367.20 (78.88; 132-872) | 1167 | 372.68 (80.96; 132-872) |

\* Mean (Standard Deviation; Range) reported



12/15 points – 3 points for the discussion (-1 for not discussing whether CVD could be an effect modifier or confounder, -1 for not commenting on the apparent non-linearity in the data), 4 points for table (-1 for not showing patterns in missingness), 5 points for scatter plots.

2a.

**Methods:** The mean levels of fibrinogen were compared for participants with prior history of CVD (n=1149) and those without prior CVD (n=3791). 85 total participants with missing fibrinogen data were excluded from analyses. A two-sided t-test assuming equal variances was used to compare the differences in mean fibrinogen between the two groups. A 95% confidence interval (CI) for the difference in population mean fibrinogen was constructed under the assumption of equal variances. The R package, “uwIntroStats” and the function “ttest” were used for this analysis.

Overall good methods write-up.

**Results:** The mean fibrinogen level in those without prior CVD (n=3791) was 319.6 with a 95% CI of 317.5-321.6. In contrast, the mean fibrinogen level in those with known CVD (n=1149) was 334.5 with a 95% CI of 330.1-338.8. The 14.9 unit increase in fibrinogen in those with prior CVD compared to those without known CVD would not be deemed unusual if the true population difference in mean fibrinogens was between 10.4 and 19.3 based on a 95% CI calculated assuming equal variances. The reported t-statistic is equal to -6.541. This observation is statistically significant (two-sided P = 6.72x10-11) at the 0.05 level; thus we can reject the null hypothesis that there is no difference in mean fibrinogen levels between those with known CVD and those without.

Good. 10 points.

2b. The same analysis presented in 2a could have also been performed using classical linear regression, as it is the exact same as the t-test assuming equal variances with a binary predictor. The p-values and t statistics calculated would be the same in both tests. Moreover, the difference in means would equal the absolute value of the slope (as the direction of effect can depend on which group is deemed the “reference” group in the t-test) and the 95% CI would have their absolute value be the same in both tests.

Good: 4.5/5 points (-0.5 points because it is not *exactly* the same since the SE for the intercept is not the same as the SE for the mean in the t-test).

2c.

**Methods:** The mean levels of fibrinogen were compared for participants with prior history of CVD (n=1149) and those without prior CVD (n=3791). 85 total participants with missing fibrinogen data were excluded from analyses. A two-sided t-test allowing for unequal variances was used to compare the differences in mean fibrinogen between the two groups. A 95% confidence interval (CI) for the difference in population mean fibrinogen was constructed under the allowance of unequal variances. The R package, “uwIntroStats” and the function “ttest” were used for this analysis.

**Results:** The mean fibrinogen level in those without prior CVD (n=3791) was 319.6 with a 95% CI of 317.5-321.6. In contrast, the mean fibrinogen level in those with known CVD (n=1149) was 334.5 with a 95% CI of 330.1-338.8. The 14.9 unit increase in fibrinogen in those with prior CVD compared to those without known CVD would not be deemed unusual if the true population difference in mean fibrinogens was between 10.1 and 19.7 based on a 95% CI calculated allowing for unequal variances. The reported t-statistic is equal to -6.084. This observation is statistically significant (two-sided P = 1.45x10-9) at the 0.05 level; thus we can reject the null hypothesis that there is no difference in mean fibrinogen levels between those with known CVD and those without.

10/10 points. Similar comments as above RE reporting details of coding in R and reporting the p-value as p<0.0001 instead of scientific notation.

2d. The analysis presented in 2c could have been *approximated* using linear regression with robust standard errors (Huber-White sandwhich estimator). If performing a linear regression using robust standard errors, the slope estimate and 95% CI of the slope match up to the difference in means and the 95% CI for the difference in population means from the t-test allowing for unequal variances. The two tests differ on their calculated p-value, however, as the linear regression model with Huber-White estimates gives a p=1.26x10-9, while the t-test with unequal variances provides a p-value of 1.45x10-9.

Good: 10/10 points. As the key notes, the 95% CIs won’t be exactly equal, but they do ‘match up’ in this case.

2e. The t-test assuming equal variances is the superior test only when the variances between the two populations are actually equal (which happens very rarely). When the variances are not equal, then the t-test assuming equal variances is anti-conservative (estimates are too high and the inferences are too strong). Given this, I would have predicted *a priori* that the analyses in part c would find a weaker association (more conservative). This was the case, as both the t-test (decreased to -6.084 from -6.541) and p-value (increased to 1.45x 10-9 from 6.72x10-11) represented a weaker association.

Good: 5/5 points.

3.

**Methods:** All analyses were performed in R using standard regression packages and the “uwIntroStats” library. 101 participants had missing CRP or fibrinogen and were excluded, leaving 4899 total subjects with complete data for analyses. Linear regression comparing group means was used to assess the relationship between fibrinogen (outcome) and CRP (predictor) with standard errors calculated allowing for unequal variances (Huber-White estimates).

**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **N** | **Intercept** | **Beta Coefficient** | **95% CI** | **P-Value** |
| Overall  | 4899 | 304.0 | 5.251 | 4.604 – 5.898 | <0.001 |

3a. The estimated intercept from the fitted regression model is 304.0. This is interpreted to mean that, for those subjects with a CRP = 0, their estimated mean fibrinogen levels are 304.0.

Good: 4/5 points (Minus 1 point because the question asked you to report fitted values to 3 sig figs not 4 for all of these).

3b. The estimated slope from the fitted regression model is 5.251, which is interpreted to mean that for each 1 unit increase in CRP levels, the difference in mean fibrinogen levels is 5.251 units.

5/5 points.

3c. From linear regression analysis using robust standard error calculations, we estimate that for each 1 unit increase in CRP levels, the difference in mean fibrinogen levels is 5.251 units. The 95% CI calculated with robust standard errors suggests that this finding would not be deemed unusual if the true population difference in mean fibrinogen was between 4.604 and 5.898 units for each 1 unit increase in CRP levels. As the two-sided p-value is P < 0.001, we reject the null hypothesis that there is no association between CRP and fibrinogen levels in the overall population.

Good: 10/10 points.

3d. See table after problem 6.

4.

**Methods:** All analyses were performed in R using standard regression packages and the “uwIntroStats” library. 101 participants had missing CRP or fibrinogen and were excluded, leaving 4899 total subjects with complete data for analyses. Natural log transformation (ln) was performed on CRP prior to analyses, with those with CRP = 0 being adjusted to have CRP = 0.5, or ½ the value of the lowest non-zero observation (CRP=1). Linear regression comparing group means was used to assess the relationship between fibrinogen (outcome) and ln(CRP) (predictor) with standard errors calculated allowing for unequal variances (Huber-White estimates).

Good.

**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **N** | **Intercept** | **Beta Coefficient** | **95% CI** | **P-Value** |
| Overall  | 4899 | 295.6 | 36.83 | 34.58 – 39.09 | <0.001 |

4a. The estimated intercept from the fitted regression model is 295.6 This is interpreted to mean that, for those subjects with a ln(CRP) = 0, their estimated mean fibrinogen levels are 295.6.

5/5 points. It would be better to report “ln(CRP)” interms of units of CRP (ln(0)=1).

4b. The estimated slope from the fitted regression model is 36.83, which is interpreted to mean that for each 1 unit increase in ln(CRP) levels, the difference in mean fibrinogen levels is 36.83 units.

4/5 points. Minus 1 point because it is not a 1-unit increase – if you used natural log, it is a 2.718-fold increase..

4c. From linear regression analysis using robust standard error calculations, we estimate that for each 1 unit increase in ln(CRP) levels, the difference in mean fibrinogen levels is 36.83 units. The 95% CI calculated with robust standard errors suggests that this finding would not be deemed unusual if the true population difference in mean fibrinogen was between 34.58 and 39.09 units for each 1 unit increase in ln(CRP) levels. As the two-sided p-value is P < 0.001, we reject the null hypothesis that there is no association between ln(CRP) and fibrinogen levels in the overall population.

10/10 points. Same comment as above, but I won’t take off points twice.

4d. See table after problem 6.

5.

**Methods:** All analyses were performed in R using standard regression packages and the “uwIntroStats” library. 101 participants had missing CRP or fibrinogen and were excluded, leaving 4899 total subjects with complete data for analyses. Linear regression comparing group geometric means (using the fnctl = “geometric mean” in the uwIntroStats “regress” function) was used to assess the relationship between geometric mean fibrinogen (outcome) and CRP (predictor) with standard errors calculated allowing for unequal variances (Huber-White estimates).

**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **N** | **Intercept\*** | **Beta Coefficient\*** | **95% CI\*** | **P-Value** |
| Overall  | 4899 | 300.9 | 1.014 | 1.012 – 1.016 | <0.001 |

\*Note that the intercept, beta coefficient, and 95% CIs have been exponentiated [exp()] from the geometric mean calculations.

5a. The estimated exponentiated intercept from the fitted regression model is 300.9. This is interpreted to mean that, for those subjects with a CRP = 0, their estimated **geometric** mean fibrinogen levels are 300.9 after exponentiation.

4/5 points. Minus 1 because it is the geometric mean, not the mean.

5b. The estimated exponentiated slope from the fitted regression model is 1.012, which is interpreted to compare the ratio of geometric means across groups differing by 1 unit of CRP. In other words, it is used to compare groups with a k-fold difference in their measured CRP such that k1.012 represents the ratio of geometric means for two groups who differ in CRP by k-fold amount (e.g., 10% is k = 1.1).

5/5 points.

5c. From linear regression analysis using robust standard error calculations, we estimate that the ratio of geometric means is 1.014 across groups with varying CRP. The 95% CI calculated with robust standard errors suggests that this finding would not be deemed unusual if the true population ratio of geometric means was between 1.012 and 1.016 units for a given proportional increase in CRP levels. As the two-sided p-value is P < 0.001, we reject the null hypothesis that there is no association between CRP and fibrinogen levels in the overall population.

Good. 10/10 points

5d. See table after problem 6.

6.

**Methods:** All analyses were performed in R using standard regression packages and the “uwIntroStats” library. 101 participants had missing CRP or fibrinogen and were excluded, leaving 4899 total subjects with complete data for analyses. Natural log transformation (ln) was performed on CRP prior to analyses, with those with CRP = 0 being adjusted to have CRP = 0.5, or ½ the value of the lowest non-zero observation (CRP=1). Linear regression on geometric means (using the fnctl = “geometric mean” in the uwIntroStats “regress” function) was used to assess the relationship between geometric mean fibrinogen (outcome) and ln(CRP) (predictor) with standard errors calculated allowing for unequal variances (Huber-White estimates).

**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **N** | **Intercept\*** | **Beta Coefficient\*** | **95% CI\*** | **P-Value** |
| Overall  | 4899 | 292.5 | 1.111 | 1.105 – 1.118 | <0.001 |

\*Note that the intercept, beta coefficient, and 95% CIs have been exponentiated [exp()] from the geometric mean calculations.

6a. The estimated intercept from the fitted regression model is 292.5. This is interpreted to mean that, for those subjects with a ln(CRP) = 0, their estimated geometric mean fibrinogen levels are 292.5 after exponentiation.

4/5 points. Again, this is the geometric mean.

6b. The estimated exponentiated slope from the fitted regression model is 1.111, which is interpreted to compare the ratio of geometric means across groups differing by 1 unit of CRP. In other words, it is used to compare groups with a k-fold difference in their measured ln(CRP) such that k1.111 represents the ratio of geometric means for two groups who differ in ln(CRP) by k-fold amount (e.g., 10% is k = 1.1).

4/5 points. This is not a 1-unit difference in CRP, it is a 2.718-fold difference.

6c. From linear regression analysis using robust standard error calculations, we estimate that the ratio of geometric means is 1.111 across groups with varying ln(CRP). The 95% CI calculated with robust standard errors suggests that this finding would not be deemed unusual if the true population ratio of geometric means was between 1.105 and 1.118 units for a given proportional increase in ln(CRP) levels. As the two-sided p-value is P < 0.001, we reject the null hypothesis that there is no association between ln(CRP) and fibrinogen levels in the overall population.

10/10 points.

6d. See table after problem 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** | 309.251 | 295.6 a | 305.186 b | 292.657 c |
| **2 mg/L** | 315.502 | 321.129 a | 309.464 b | 314.838 c |
| **3 mg/L** | 319.753 | 336.062 a | 313.801 b | 328.584 c |
| **4 mg/L** | 325.004 | 346.657 a | 318.200 b | 338.700 c |
| **6 mg/L** | 335.506 | 361.591 a | 327.183 b | 353.489 c |
| **8 mg/L** | 346.008 | 372.186 a | 336.420 b | 364.371 c |
| **9 mg/L** | 351.259 | 376.524 a | 341.136 b | 368.923 c |
| **12 mg/L** | 367.012 | 387.119 a | 335.683 b | 380.281 c |

a Calculated using ln(x) and the coefficients in #4, where x = CRP level values in the first column.

b Calculated using non-exponentiated coefficients and then taking exp() to obtain geometric mean for given x.

c Calculated using ln(x) and non-exponentiated coefficients in #6 (where x = CRP level values in the first column) and then taking exp()to obtain geometric mean for given ln(x).

Good. A few values were slightly off from the key, but mostly that I imagine is from the precision with which they were estimated.

7.

|  |  |
| --- | --- |
|  | **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 | 25.529 a | 4.278 b | 22.181 c |
| **3 mg/L – 2 mg/L** | 5.251 | 14.933 a | 4.338 b | 13.747 c |
| **4 mg/L – 1 mg/L** | 15.753 | 51.057 a | 13.014 b | 46.044 c |
| **4 mg/L – 2 mg/L** | 10.502 | 25.529 a | 8.736 b | 23.862 c |
| **6 mg/L – 3 mg/L** | 15.753 | 25.529 a | 13.382 b | 24.904 c |
| **8 mg/L – 4 mg/L** | 21.004 | 25.529 a | 18.220 b | 25.671 c |
| **9 mg/L – 6 mg/L** | 15.753 | 14.933 a | 13.952 b | 15.434 c |
| **9 mg/L – 8 mg/L** | 5.251 | 4.338 a | 4.716 b | 4.552 c |
| **12 mg/L – 6 mg/L** | 31.506 | 25.529 a | 28.499 b | 26.792 c |
| ***Ratios*** |
| **2 mg/L / 1 mg/L** | 1.017 | 1.086 a | 1.014 b | 1.076 c |
| **3 mg/L / 2 mg/L** | 1.017 | 1.047 a | 1.014 b | 1.044 c |
| **4 mg/L / 1 mg/L** | 1.051 | 1.173 a | 1.043 b | 1.157 c |
| **4 mg/L / 2 mg/L** | 1.033 | 1.079 a | 1.028 b | 1.076 c |
| **6 mg/L / 3 mg/L** | 1.049 | 1.076 a | 1.043 b | 1.076 c |
| **8 mg/L / 4 mg/L** | 1.065 | 1.074 a | 1.057 b | 1.076 c |
| **9 mg/L / 6 mg/L** | 1.047 | 1.041 a | 1.043 b | 1.044 c |
| **9 mg/L / 8 mg/L** | 1.015 | 1.012 a | 1.014 b | 1.012 c |
| **12 mg/L / 6 mg/L** | 1.094 | 1.071 a | 1.087 b | 1.076 c |

a Calculated using ln(x) and the coefficients in #4, where x = CRP level values in the first column.

b Calculated using non-exponentiated coefficients in #5 and then taking exp() to obtain geometric mean for given x.

c Calculated using ln(x) and non-exponentiated coefficients in #6 (where x = CRP level values in the first column) and then taking exp()to obtain geometric mean for given ln(x).

Good. 10/10 points.

8a. The untransformed mean CRP gave a constant difference in fitted values for a given increase of *c* units in CRP levels.

-When CRP increased by 1 unit (2-1, 3-2, 9-8), mean fibrinogen increased by 5.251 units.

-When CRP increased by 2 units (4-2), mean fibrinogen increased by 10.502 units.

-When CRP increased by 3 units (4-1, 6-3, 9-6), mean fibrinogen increased by 15.753 units.

-When CRP increased by 4 units (8-4), mean fibrinogen increased by 21.004 units.

-When CRP increased by 6 units (12-6), mean fibrinogen increased by 31.506 units.

Good, but you were only supposed to show the differences that were the same for more than one contrast.

5/5 points.

8b. The geometric mean of fibrinogen on untransformed CRP gave a constant ratio in fitted values when comparing two groups that differed by an absolute increase in *c* units in CRP levels.

-When CRP increased by 1 unit (2/1, 3/2, 9/8), the ratio of fibrinogen geometric means = 1.014.

-When CRP increased by 2 units (4/2), the ratio of fibrinogen geometric means = 1.028.

-When CRP increased by 3 units (4/1, 6/3, 9/6), the ratio of fibrinogen geometric means = 1.043.

-When CRP increased by 4 units (8/4), the ratio of fibrinogen geometric means = 1.057.

-When CRP increased by 6 units (12/6), the ratio of fibrinogen geometric means = 1.087.

5/5 points.

8c. The regression of untransformed fibrinogen on ln(CRP) gave a constant difference in fitted values when comparing groups that differed by a *c-fold* increase in CRP levels.

 -When ln(CRP) increased by (ln(9)-ln(8)), the difference in mean fibrinogen was 4.338 units.

-When ln(CRP) increased by 1.5 fold, (ln(3)-ln(2), ln(9)–ln(6)), the difference in mean fibrinogen was 14.933 units.

-When ln(CRP) increased by 2 fold, (ln(2)-ln(1), ln(4)–ln(2), ln(6)–ln (3), ln(8)–ln(4), ln(12)–ln(6)), the difference in mean fibrinogen was 25.529units.

-When ln(CRP increased by 4 fold (ln(4)–ln (1)), the difference in mean fibrinogen was 51.057.

4.5/5 points. As for part b, you incorrectly stated that it is the mean ln(CRP) that has a difference – it is the mean fibrinogen that is different across levels of log-transformed CRP.

8d. The geometric mean of fibrinogen modeled on natural log transformed CRP gave a constant ratio in fitted values when comparing groups that differed by a *c-fold* increase in CRP levels.

-When ln(CRP) increased by (ln(9)-ln(8)), the ratio of fibrinogen geometric means = 1.012.

-When ln(CRP) increased by 1.5 fold, (ln(3)-ln(2), ln(9)–ln(6)), the ratio of fibrinogen geometric means = 1.044.

-When ln(CRP) increased by 2 fold, (ln(2)-ln(1), ln(4)–ln(2), ln(6)–ln (3), ln(8)–ln(4), ln(12)–ln(6)), the ratio of fibrinogen geometric means = 1.076.

-When ln(CRP increased by 4 fold (ln(4)–ln (1)), the ratio of fibrinogen geometric means = 1.157.

5/5 points.

9. The choice of analysis between the four potential ones to investigate the association between fibrinogen and CRP is dependent on two major factors: (1) accuracy and (2) ease of interpretation. In this case, while there are numerous subjects (n=4899) such that linear regression using robust standard errors should yield a fairly accurate and conservative estimate of the relationship between fibrinogen and CRP, there is concern about the outliers with extremely high CRP (see **Figure** in answer to #1). Thus, it is reasonable to consider the use of geometric means in regression as it would downweight outliers. With regard to ease of interpretation, it is arguable whether an additive (change in fibrinogen for absolute value change in CRP) vs. multiplicative model (change in fibrinogen for a proportional change in CRP) is easier to interpret, and in particular, which is clinically relevant. Given these considerations and the particular concern about outliers, I would have performed a regression allowing for unequal variances on the geometric means of untransformed CRP. It combines easily interpretable results (compared to normal regression with ln(CRP) or geometric mean regression with ln(CRP)) with more conservative and likely accurate estimates that downweight outliers (as compared to normal regression on untransformed CRP).

4/5 points. You state that you want to downweight outliers in CRP, but that you would model *untransformed* CRP? This is inconsistent and it is unclear what you are doing.