BIOS 645 Spring 2023 Homework #2

Maya Krishnamoorthy

1. Compute univariate descriptives for these variables.   
   Chart, histogram

   Description automatically generatedTable

   Description automatically generated

Chart, histogram

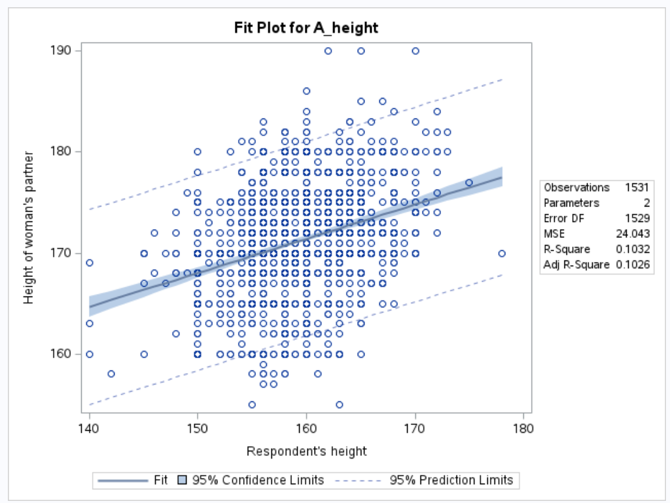
Description automatically generatedTable

Description automatically generated

1. Regress A onto R, and R onto A. Report the parameter estimates; with simple algebra   
   and arithmetic, show that these are not equivalent. Why aren’t they?  
    Table

   Description automatically generatedTable

   Description automatically generated

Chart, scatter chart

Description automatically generated

When we regress A onto R, our parameter estimates are that the slope is 0.33851 (95% CI: (0.28845, 0.38557)) and the intercept is 117.26 (95% CI: (109.28, 125.24)). When we regress R onto A, our parameter estimates are that the slope is 0.30482 (95% CI: (0.25975, 0.34990)) and the intercept is 107.12 (95% CI: (99.40, 114.84)).

The reason that our parameter estimates are not equivalent is because we are fixing different variables for each regression. The predicted values depend on the variable that is getting fixed, thus presenting us with a different regression slope and intercept.

1. Correlate A and R, turn them into z-scores, and regress z(A) onto z(R). Show that the   
   slope value is equivalent to the correlation. Why is that?

Graphical user interface, application

Description automatically generatedTable

Description automatically generated

Chart, scatter chart

Description automatically generated

The correlation value between A\_height and R\_height that we get from computing the Pearson test gives us 0.32123. The slope that we get from regressing z(A) onto z(R) is 0.32123, as well.

In a simple regression equation, , where is the slope of the line. We also know that . Thus, when both A\_height and R\_height are standardized, the value of the correlation coefficient and the slope of the regression line are equal.

1. Regress A onto R suppressing the intercept; with simple algebra and arithmetic, show   
   that the regression line goes through (mean of x, mean of y) for the above case (#2), but   
   not when the intercept is suppressed.   
   Chart, scatter chart

   Description automatically generatedTable

   Description automatically generated

The regression line we got in #2 that included the intercept is in the following form of an equation: . We can plug in the mean value of R\_height and we will see that the equation returns the value of A\_height, meaning that the point (R\_height, A\_height) is on our line: .

However, when we plug into the regression equation we solved for when the intercept is suppressed, we see the following: .

Table

Description automatically generatedChart, scatter chart

Description automatically generated

* 1. Report the parameter estimates with substantive interpretations in terms of the   
     meanings of the variables in question.

We get a parameter estimate for the slope from regressing A\_height onto R\_height equal to 0.33851. This means that as the respondent’s height increases by one unit, their partner’s height increases proportionally by 0.33851. However, this data does not have meaning without the context of an intercept, because only then do we have an accurate estimate of the partner’s height compared to the respondent’s.

* 1. Report the p-values associated with the t-tests of the parameter estimates. Interpret the p-values, from both Fisher’s original perspective, and whether or not they are ‘significant’ using a conventional alpha of .05.

Both p-values associated with the t-tests of the parameter estimates are less than 0.0001, for the intercept, as well as the regression slope. For the p-value associated with R\_height, the t-value is 13.26. The p-value indicates that the probability of seeing a slope equal to or more extreme than 0.33851 is less than 0.01%. This is significant because the p-value is less than the alpha 0.05.

* 1. Compute and report a 95% confidence interval for the slope.

The 95% CI for the slope is (0.28845, 0.38857). This means that if we were to recreate numerous trials of this, 95% of the time, the true population slope would be inside this confidence interval.

* 1. Report the ANOVA table for the model, and compare the p-value derived from the ANOVA table to the p-value that came from the t-test of the slope. Is it the same or different? Why?

The p-value derived from the ANOVA table is also less than 0.001, so the p-values are the same. They are the same because both tests are comparing the sample means for each of two groups, where (at least) one group has proven to be significant.

* 1. Include the scatterplot of your model with the confidence band.

Chart, scatter chart

Description automatically generated

1. ANOVA Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | df | SS | MS | F | p |
| Regression |  | SSreg = 6.995 |  |  |  |
| Residual |  | SSres = 45.922 |  |  |  |
| Total |  | SStot = 52.917 |  |  |  |