

STAT W4201 001, Homework 6

Brian Weinstein (bmw2148)

Mar 9, 2016

Code is attached here and also posted at <https://github.com/BrianWeinstein/advanced-data-analysis>. Where relevant, code snippets and output are included in-line.

Problem 1: Ramsey 9.14

- (a) *Draw a matrix of scatterplots of the four variables. Construct it so that the bottom row of plots all have heart on the vertical axis. If you do not have this facility, draw scatterplots of heart versus each of the other variables individually.*

A matrix of pairwise scatterplots is shown in Figure 1.

- (b) *Obtain the least squares fit to the linear regression of heart on bank, walk, and talk.*

```
> lm1 <- lm(Heart ~ Bank + Walk + Talk, data=paceData)
> summary(lm1)

Call:
lm(formula = Heart ~ Bank + Walk + Talk, data = paceData)

Residuals:
    Min       1Q   Median       3Q      Max
-8.4014 -3.0263  0.0602  2.6748  8.4646

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.1787     6.3369   0.502  0.6194
Bank           0.4052     0.1971   2.056  0.0480 *
Walk           0.4516     0.2009   2.248  0.0316 *
Talk          -0.1796     0.2222  -0.808  0.4249
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.805 on 32 degrees of freedom
Multiple R-squared:  0.2236, Adjusted R-squared:  0.1509
F-statistic: 3.073 on 3 and 32 DF,  p-value: 0.04162
```

- (c) *Plot the residuals versus the fitted values. Is there evidence that the variance of the residuals increases with increasing fitted values or that there are any outliers?*

The residual plot is shown in Figure 2. There does not seem to be evidence that the variance of the residuals increases with increasing fitted values, or that there are any extreme outliers.

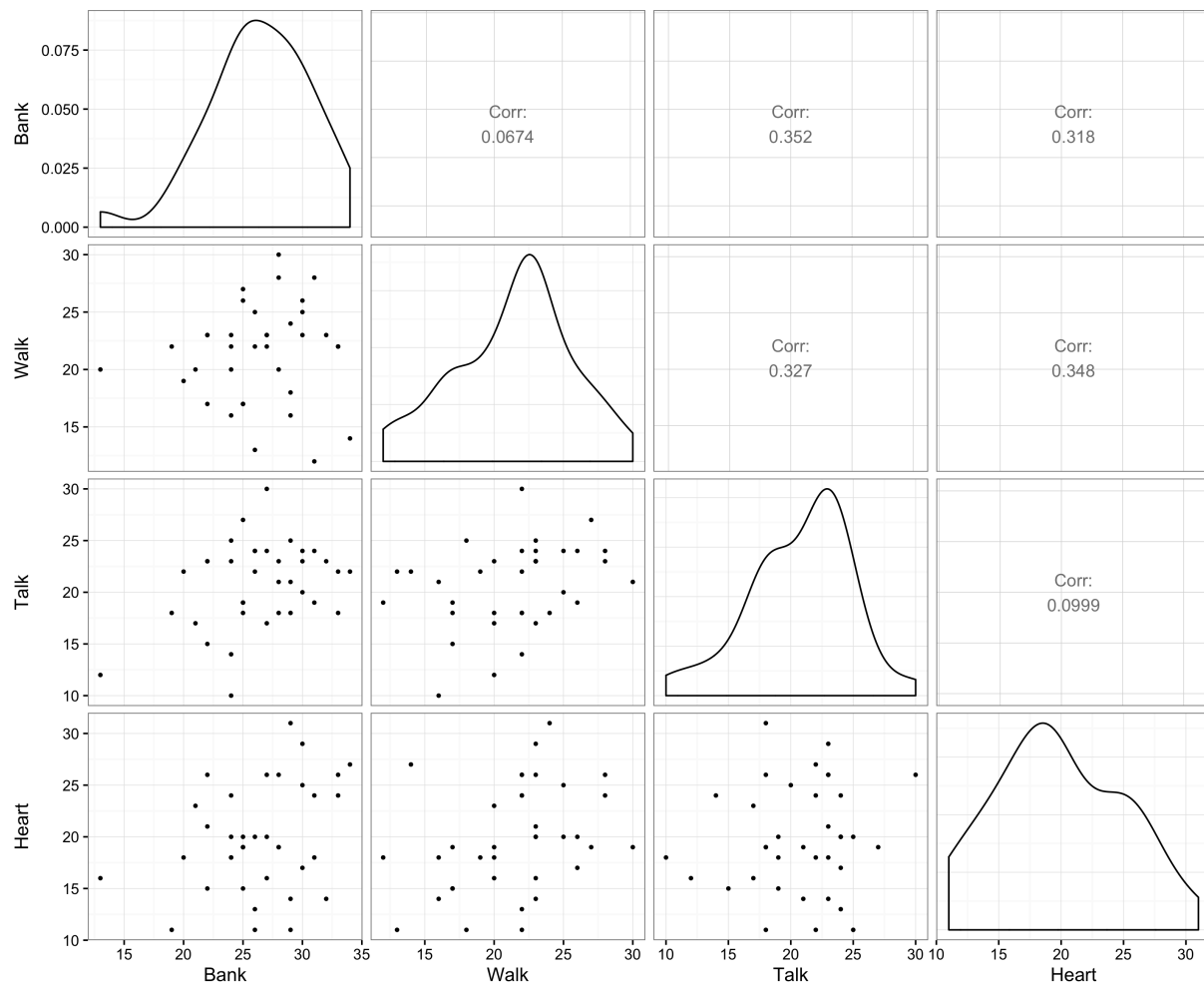


Figure 1: Pairwise scatterplots of the variables in the “Pace of Life and Heart Disease” dataset.

- (d) Report a summary of the least squares fit. Write down the estimated equation with standard errors below each estimated coefficient.

Under the parallel lines regression model, the age-adjusted death rate due to heart disease (**Heart**) increases by 0.4052 for every one unit increase in the bank clerk speed (**Bank**) (95% confidence interval from 0.0037 to 0.8067). Similarly, **Heart** increases by .4516 for every one unit increase in the pedestrian walking speed (**Walk**) (95% confidence interval from 0.0424 to 0.8608). The data provides no evidence that **Heart** is associated with postal clerk talking speed (**Talk**) (two sided p-value = 0.4249 for a test that the **Talk** coefficient is zero).

Problem 2: Ramsey 9.16

Problem 3: Ramsey 9.18

Problem 4: Ramsey 9.20

Problem 5: Ramsey 10.19

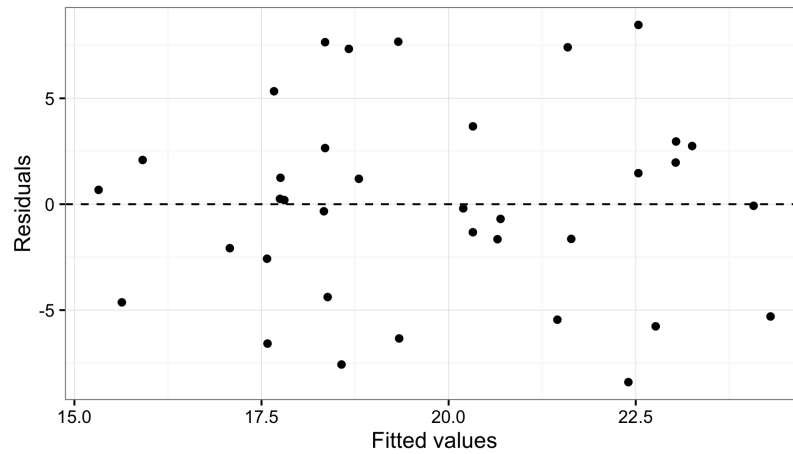


Figure 2: Residual plot for the fitted model from part (b).

Problem 6: [Ramsey 10.28](#)