STAT 512 – Assignment 2

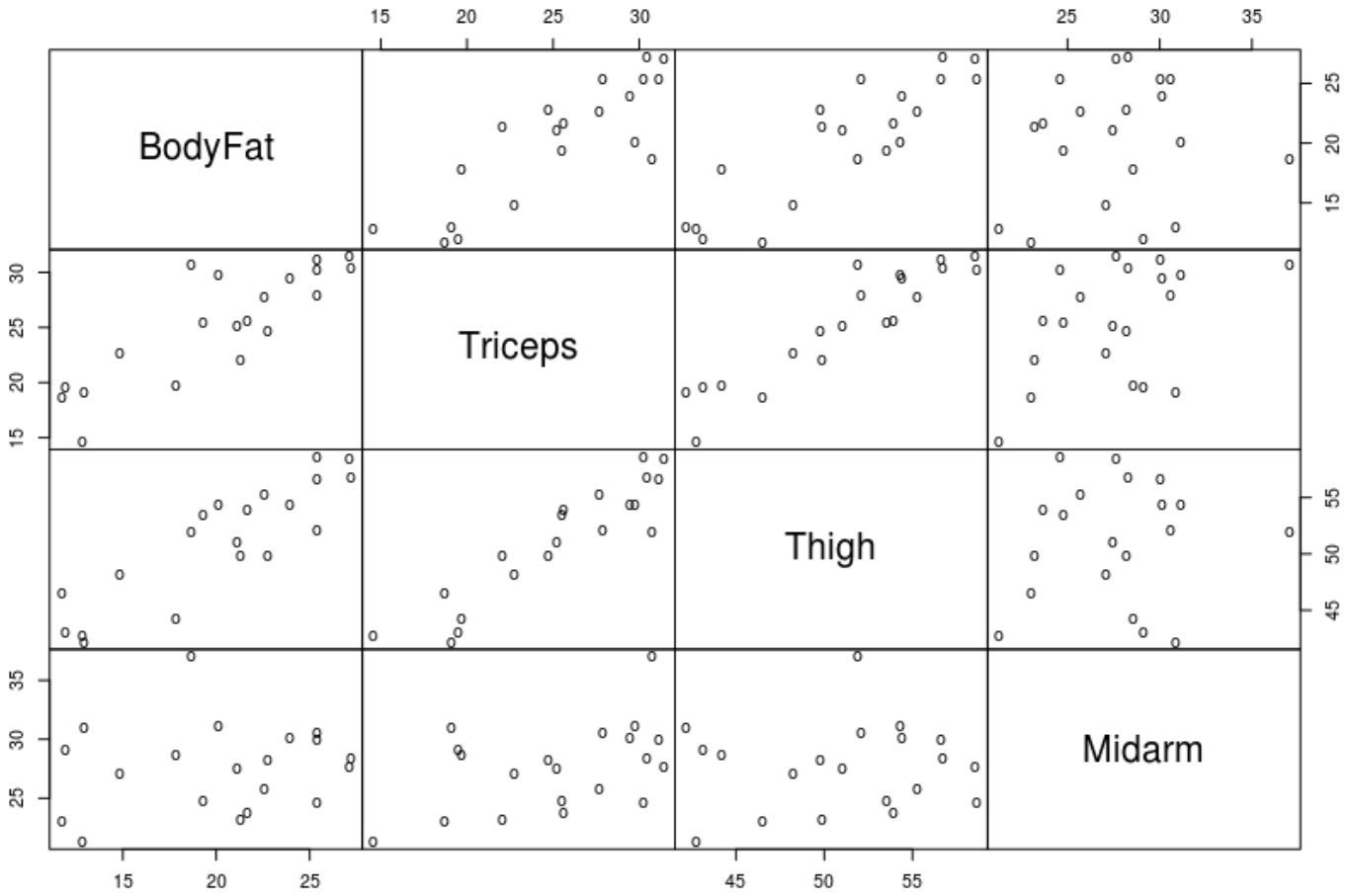
Vignesh M. Pagadala
Vignesh.Pagadala@ColoState.Edu

Feb 18, 2019

1. Calculate pairwise (Pearson) correlations between the 4 variables (BodyFat and each of the predictors). You should also briefly examine the pairwise scatterplots, but you do NOT need to include them in your assignment.

ANSWER:

	BODYFAT	TRICEPS	THIGH	MIDARM
BODYFAT	1.0000000	0.8432654	0.8780896	0.1424440
TRICEPS	0.8432654	1.0000000	0.9238425	0.4577772
THIGH	0.8780896	0.9238425	1.0000000	0.0846675
MIDARM	0.1424440	0.4577772	0.0846675	1.0000000



2. Fit the "full" model using BodyFat as the response and including all 3 predictors. Include the parameter estimate information ("Coefficients" table) and R 2 value for the full model in your assignment.

ANSWER:

Coefficients:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
Triceps        4.334      3.016   1.437    0.170
Thigh         -2.857      2.582  -1.106    0.285
Midarm        -2.186      1.595  -1.370    0.190

Multiple R-squared:  0.8014
```

3. Based on the "full" model, test the null hypothesis that all three of the partial regression coefficients are simultaneously zero. In other words, test H 0 : β 1 = β 2 = β 3 = 0. Give the F-statistic and p-value and make a conclusion about the test. (4 pts)

ANSWER:

Linear hypothesis test

```
Hypothesis:
Triceps = 0
Thigh = 0
Midarm = 0

Model 1: restricted model
Model 2: BodyFat ~ Triceps + Thigh + Midarm

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     19 495.39
2     16  98.40  3    396.98 21.516 7.343e-06 ***
```

Since the p-value is low (<0.05) we can reject the null hypothesis for this model. This means, at least one predictor variable contributes positively in this model.

4. Based on the "full" model, test the null hypothesis that the partial regression coefficient for Triceps equals 2.0 versus a two-sided alternative. In other words, test H 0 : β 1 = 2 versus H A : β 1 ≠ 2. Give a test statistic, p-value and conclusion. (4 pts)

ANSWER:

```
Linear hypothesis test

Hypothesis:
Triceps = 2

Model 1: restricted model
Model 2: BodyFat ~ Triceps + Thigh + Midarm

  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     17 102.090
2     16  98.405  1    3.6848 0.5991 0.4502
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This indicates that the predictor Triceps is not useful for the model.

5. Based on the "full" model, test the null hypothesis that the partial regression coefficients for Thigh and Midarm are simultaneously zero. In other words, test H 0 : β 2 = 0 AND β 3 = 0. Give a test statistic, p-value and conclusion. (4 pts)

ANSWER:

```
Linear hypothesis test

Hypothesis:
Thigh = 0
Midarm = 0

Model 1: restricted model
Model 2: BodyFat ~ Triceps + Thigh + Midarm

  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     18 143.120
2     16  98.405  2    44.715 3.6352 0.04995 *
```

Here the p-value is lesser than 0.05, so we reject the null hypothesis.

6. Now we will sequentially eliminate any terms from the model that are not significant at the 0.05 level. Starting from the "full" model, eliminate the least significant predictor variable (highest p-value) and rerun the regression. Continue that process until all predictor variables are significant at the 0.05 level. Include the parameter estimate information ("Coefficients" table) and R 2 value for the final model in your assignment. (4 pts) We will use this "final" model for the remaining questions.

ANSWER:

In the full model, the predictor with the highest p value is Thigh. So we eliminate that first.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.7916     4.4883   1.513   0.1486
Triceps       1.0006     0.1282   7.803 5.12e-07 ***
Midarm       -0.4314     0.1766  -2.443   0.0258 *

Multiple R-squared:  0.7862
```

Now, both predictors are significant, with p values less than 0.05.

7. In the initial inspection of the pairwise correlations and plots (question 1) it appeared that there was a relationship between BodyFat and Thigh; however, Thigh was dropped from the multiple regression because it was not significant. Speculate about why this is the case.

ANSWER:

This was because, Thigh seemed too correlated with Triceps. In multiple regression, predictors with too much correlation can be disruptive. This correlation is not present between Triceps and Midarmn and Midarm and Thigh.

8. Working from the "final" model, look at the residual plots, paying particular attention to the (A) plot of residuals versus fitted values and (B) qqplot of residuals. Discuss each of these plots and whether the regression assumptions appear to be satisfied. You do not need to include the graphs in your assignment, just discuss your findings and conclusions. (4 pts)

ANSWER:

A: The plot of residuals vs fitted values look great. We can see that no trend exists, and the values are equally spread around. This ensures that the equal variance assumption and the linearity assumption stand validated.

B: The QQ plots line up in the straight line more or less, hence, we can be assured that the normality assumption holds.

9. Consider a subject with Triceps = 20 and Midarm = 25. Working from the "final" model, give (A) predicted body fat for this subject, (B) 95% confidence interval for the mean BodyFat of subjects with the same values and (C) 95% prediction interval for the predicted BodyFat for a new subject with these values. (4 pts)

ANSWER:

A:
16.01728

B:
14.25175<= B <= 17.7828

C:
10.46253 <= B <= 21.57202

10. Working from the "final" model, identify the largest RStudent residual and do an outlier test for that value. Give the test statistic, unadjusted p-value and Bonferonni adjusted p-value. Based on the Bonferonni adjusted p-value, can we conclude this observation is an outlier? Note: The outlierTest() function from the car package can be used for this question, but may return an NA for the Bonferonni p-value. I still want the Bonferonni adjusted p-value! (4 pts)

ANSWER

```
No Studentized residuals with Bonferonni p < 0.05

Largest |rstudent|:
     rstudent unadjusted p-value Bonferonni p
13 -1.818309           0.087787           NA
```