COVER PAGE

STAT 608 Homework 05 Summer 2017

Please TYPE your name and email address below, then convert to PDF and attach as the first page of your homework upload.

NAME: RAJAN KAPOOR

EMAIL: r.kapoor@tamu.edu

HOMEWORK NUMBER: 5

21/21

(Solution 1)

```
adRevenue <- read.csv("AdRevenue.csv", header=TRUE)
#attach(AdRevenue)

library(alr3)
summary(powerTransform(cbind(AdRevenue,Circulation)~1,data=adRevenue)
)
```

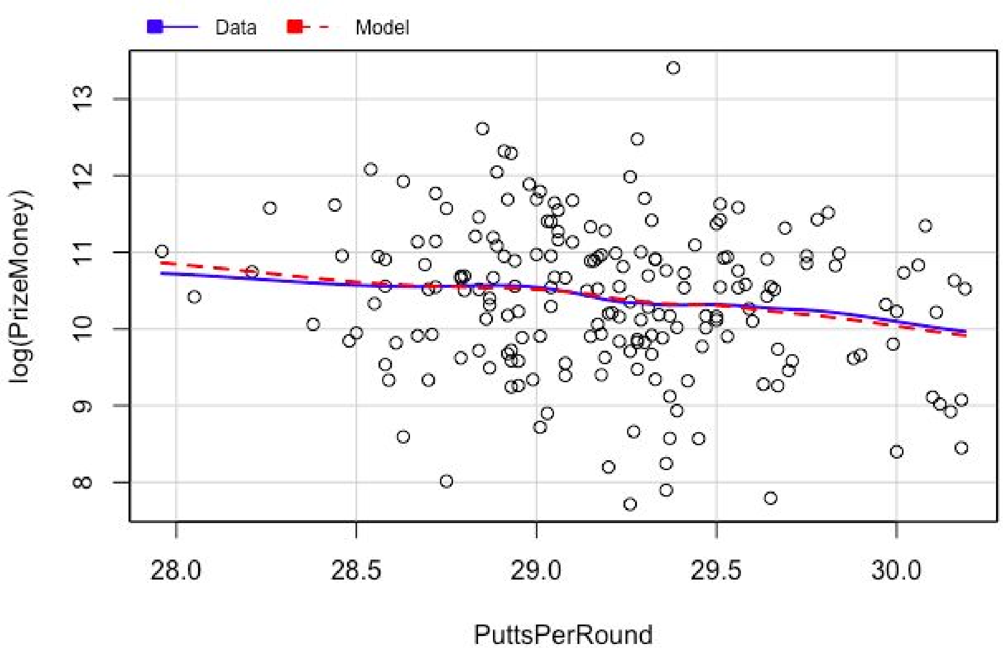
R output:

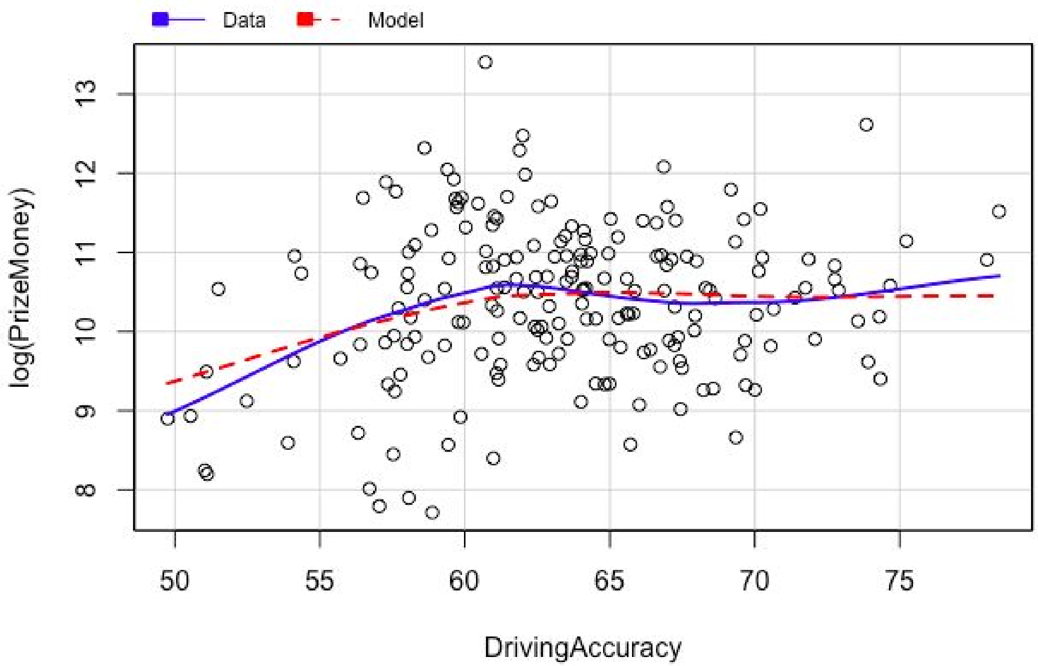
```
bcPower Transformations to Multinormality
            Est.Power Std.Err. Wald Lower Bound Wald Upper Bound
AdRevenue     -0.5873   0.1709          -0.9222          -0.2524
Circulation   -0.2428   0.0828          -0.4051          -0.0805

Likelihood ratio tests about transformation parameters
                                     LRT df        pval
LR test, lambda = (0 0)        13.809568  2 0.001002976
LR test, lambda = (1 1)       249.214715  2 0.000000000
LR test, lambda = (-0.5 -0.33)  4.224511  2 0.120964845
```

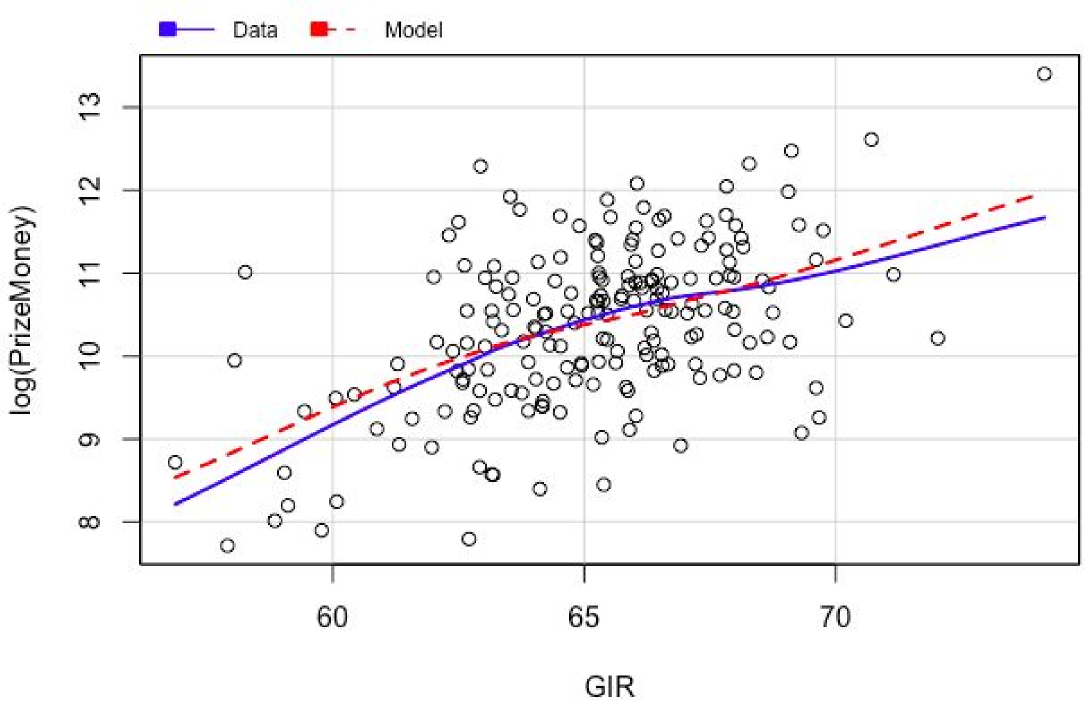
These results indicate `lambda = 0` i.e. log transformation for both Ad Revenue and Circulation is appropriate.

(Solution 2)

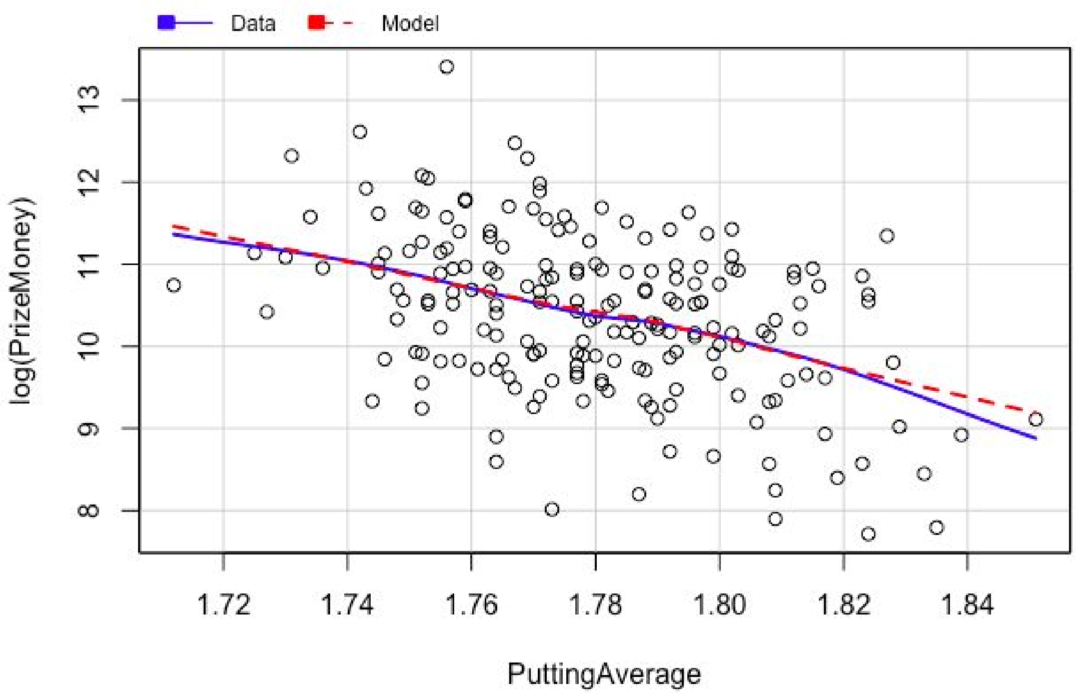


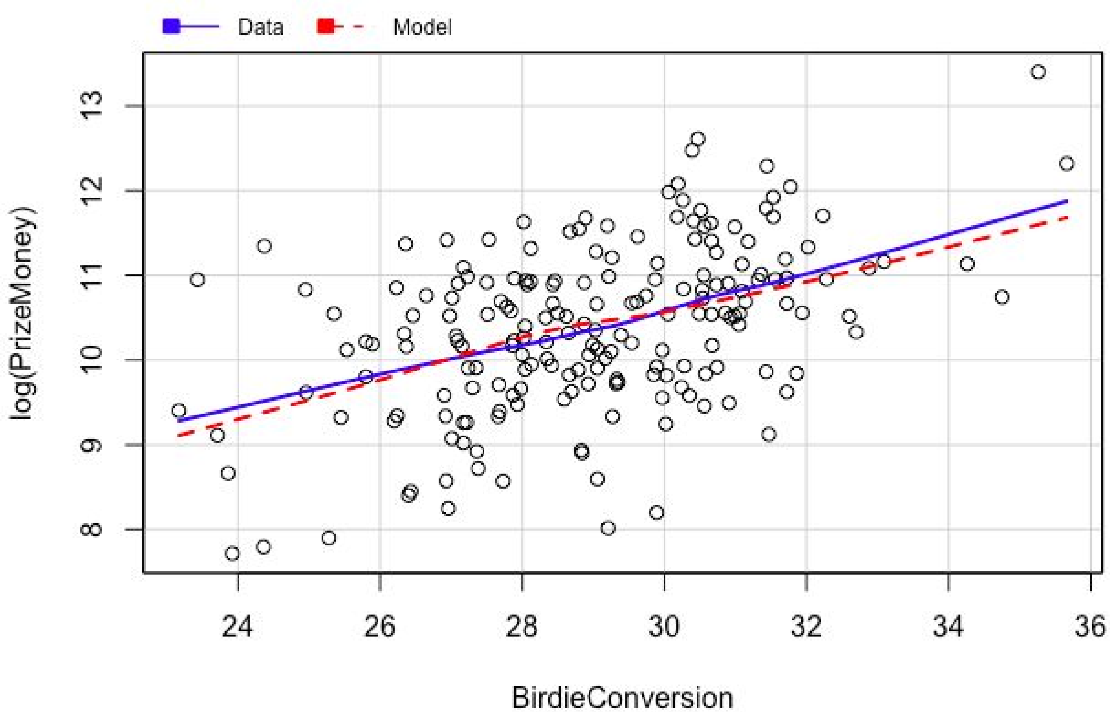


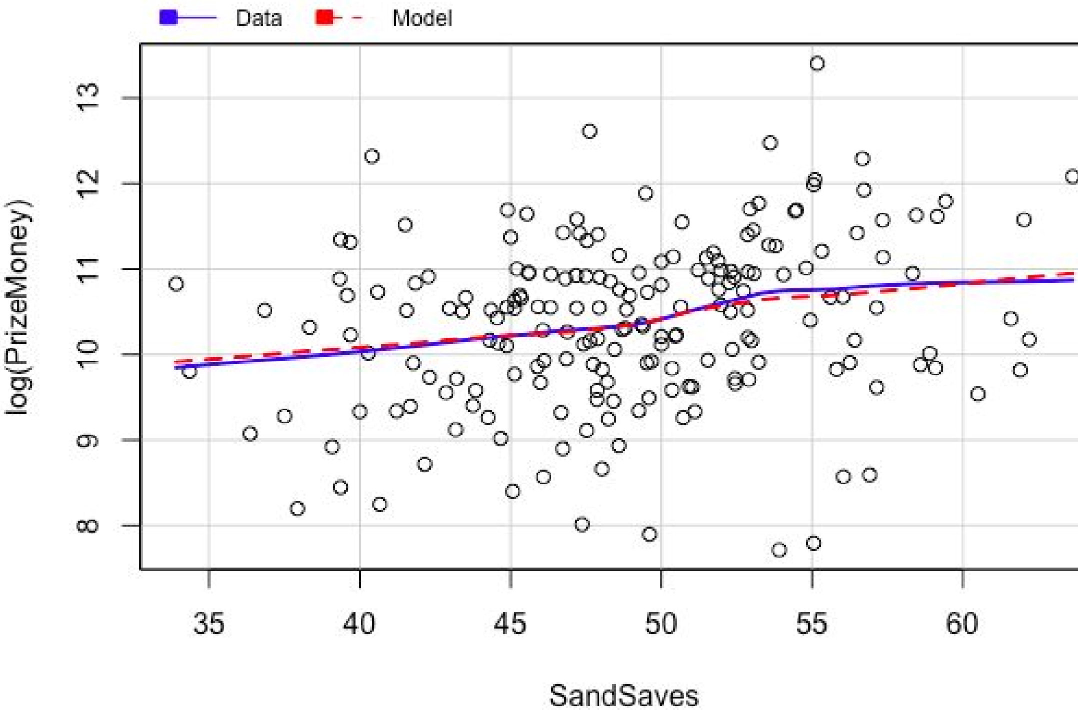
Data
Model
log(PrizeMoney)
8
9
10
11
12
13
50
55
60
65
70
75
DrivingAccuracy



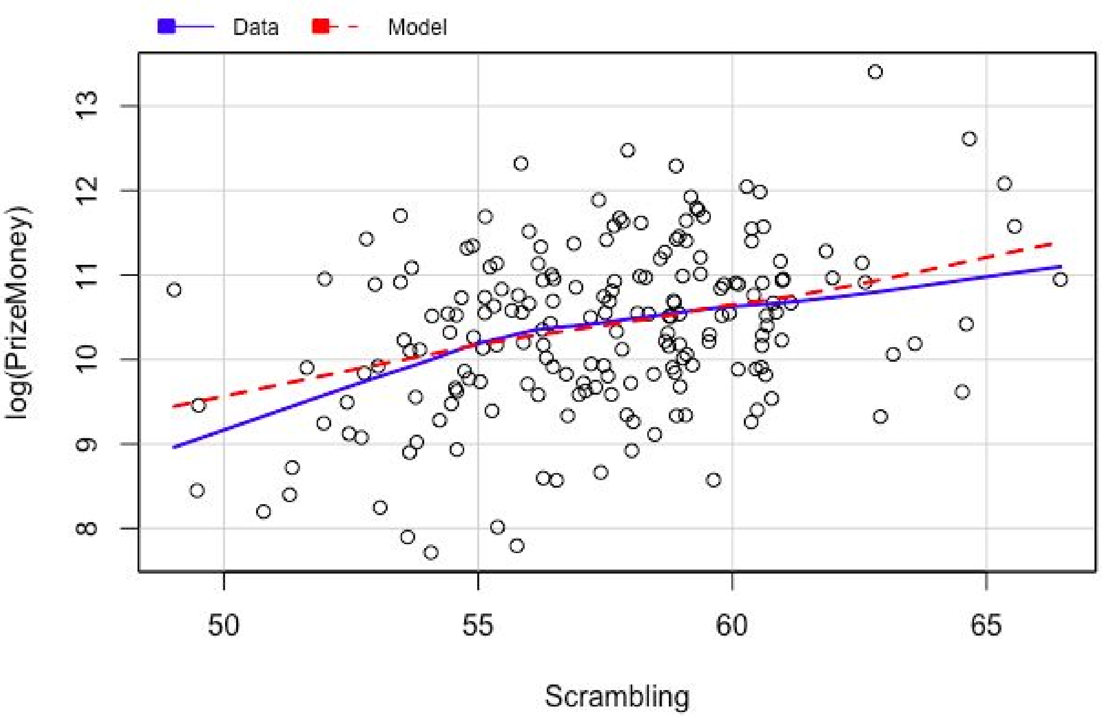
Data
Model
log(PrizeMoney)
8
9
10
11
12
13
60
65
70
GIR


Data
Model
log(PrizeMoney)
8
9
10
11
12
13
1.72
1.74
1.76
1.78
1.80
1.82
1.84
PuttingAverage


Data
Model
log(PrizeMoney)
8
9
10
11
12
13
24
26
28
30
32
34
36
BirdieConversion



Data
Model
log(PrizeMoney)
SandSaves



Data
Model
log(PrizeMoney)
Scrambling

The data and model plots do not perfectly match indicating some transformation of predictor variables is required.

Solution 3

$$y_j = \beta_0 + \beta_1 x_j + e_j \qquad j = 1, \ldots, n$$

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j$$

(3.1)

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j$$

$$\Rightarrow \sum_{j=1}^{n} \hat{y}_j = \sum_{j=1}^{n} \hat{\beta}_0 + \sum_{j=1}^{n} \hat{\beta}_1 x_j$$

$$\Rightarrow \frac{\sum_{j=1}^{n} \hat{y}_j}{n} = \frac{\sum_{j=1}^{n} \hat{\beta}_0}{n} + \frac{\sum_{j=1}^{n} \hat{\beta}_1 x_j}{n}$$

$$\Rightarrow \bar{\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$= \bar{y}.$$

(3.2) By definition,

$$SS_{reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$= \sum (\hat{y}_i - \bar{y})(\hat{y}_i - \bar{y})$$

$$= \sum (\hat{y}_i - \bar{y})[(\hat{y}_i - y_i) + (y_i - \bar{y})]$$

$$= \sum (\hat{y}_i - \bar{y})(\hat{y}_i - y_i) + \sum (\hat{y}_i - \bar{y})(y_i - \bar{y})$$

Consider the first term

$$\sum (\hat{y}_i - \bar{y})(\hat{y}_i - y_i)$$

$$= \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})(\hat{\beta}_0 + \hat{\beta}_1 x_i - y_i)$$

$$= \sum \hat{\beta}_1 (x_i - \bar{x})(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - y_i)$$

$$= \sum \hat{\beta}_1 (x_i - \bar{x})\{(\bar{y} - y_i) + \hat{\beta}_1 (x_i - \bar{x})\}$$

$$= -\sum \hat{\beta}_1 (x_i - \bar{x})(y_i - \bar{y}) + \sum \hat{\beta}_1^2 (x_i - \bar{x})^2$$

$$= -\hat{\beta}_1 (SXY) + \hat{\beta}_1^2 (SXX)$$

$$= -\hat{\beta}_1 SXY + \beta_1 SXY \qquad \left(\because \hat{\beta}_1 = \frac{SXY}{SXX} \Rightarrow SXX \cdot \hat{\beta}_1 = SXY\right)$$

$$= 0$$

$$\therefore SSreg = \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y}) \qquad (\because \bar{\hat{y}} = \bar{y})$$

(3.3)

$$R^2 = \frac{SSreg}{SST}$$

$$= \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \qquad \text{(by definition)} - \text{Eq}①$$

and

$$\rho^2_{(y_i, \hat{y}_i)} = \frac{\left[\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right]^2}{\left[\sum_{i=1}^{n} (y_i - \bar{y})^2\right]\left[\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2\right]}$$

Multiplying Eq ① by $\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2\right]^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (\because \bar{\hat{y}}_i = \bar{y})$$

$$= \frac{SSreg^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \frac{\left[\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y} - \bar{\hat{y}})\right]^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (\text{Using } 3.2)$$

$$= \rho^2$$

(3.4) Adding more covariates reduces RSS so that more of the variation is explained by the regression model, thus increasing SSreg since the sum RSS + SSreg = SST is constant. Since $R^2 = \frac{SSreg}{SST}$, the coefficient of determination increases on adding more covariates and should be maximum at maximum no. of allowable covariates i.e. $m = 20$.