STAT 572 **Hw4** **Suggested Solution**

Problem 1: Part (a):

```
> calib = read.table("calib.txt", header=T)
> str(calib)
> lm1 = lm(new~old, data=calib)
> summary(lm1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.748    114.732   0.015    0.988
old            0.984      0.115   8.553 2.69e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 169.5 on 8 degrees of freedom
Multiple R-Squared: 0.9014,     Adjusted R-squared: 0.8891
F-statistic: 73.15 on 1 and 8 DF,  p-value: 2.691e-05
```

Hence our linear model is : new = 1.748 + 0.984*old
For $H_0 : \beta_0 = 0,\ t - ratio = 0.015,\ df = 8,\ p - value = 0.988$**, we do not reject** H_0.

Part(b):

```
lm1b = lm(new~old-1, calib) ##fit a linear model with intercept
summary(lm1b)
Coefficients:
    Estimate Std. Error t value Pr(>|t|)
old  0.98551    0.05067   19.45 1.16e-08 ***
---
```

```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 159.8 on 9 degrees of freedom
Multiple R-Squared: 0.9768,     Adjusted R-squared: 0.9742
F-statistic: 378.3 on 1 and 9 DF,  p-value: 1.161e-08
```

Hence our linear model is : new = 0.986*old

For H_0: $\beta_0 = 1,\ t-ratio = \frac{(\hat{\beta}_1 - 1)}{SE(\hat{\beta}_1)} = (0.98551 - 1)/0.05067) = -0.286,\ \ df = 9,\ p - value = 0.7813$ **do not reject** H_0**.**

Part(c):

From (b), we do not reject the hypothesis that the slope is 1. Hence there is no evidence of a difference between the two techniques.

Part(d):

```
> sum(lm1a$resid)
[1] -3.197442e-14
>sum(lm1b$resid)
[1] 3.813304
```

The residuals from part(a) sum to 0.
The residuals from part(b) sum to 3.813.
We lose the feature of residuals summing to 0 when forcing a regression line through the origin.

Problem 2:

larch = read.table("larch.txt", header=T)

lm2 = lm(height ~ . , larch)

Part(a)

```
>best.aic = step(lm2)
>summary(best.aic)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -193.07      35.78  -5.396 2.03e-05 ***
nitro         107.80      22.92   4.702 0.000109 ***
phos          304.24     165.17   1.842 0.078998 .
potas         143.13      44.13   3.243 0.003731 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 38.05 on 22 degrees of freedom
Multiple R-Squared: 0.8602,     Adjusted R-squared: 0.8412
F-statistic: 45.14 on 3 and 22 DF,  p-value: 1.435e-09
```

The model prefered by AIC is:

Height = -193.07 + 107.80 * nitro + 304.24 * phos + 143.13 * potas.

Part(b)

```
n = length(larch \$ height) # number of observations
best.bic = step(lm2, k=log(n))
summary(best.bic)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -193.07      35.78  -5.396 2.03e-05 ***
```

```
nitro         107.80      22.92   4.702 0.000109 ***
phos          304.24     165.17   1.842 0.078998 .
potas         143.13      44.13   3.243 0.003731 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 38.05 on 22 degrees of freedom
Multiple R-Squared: 0.8602,     Adjusted R-squared: 0.8412
F-statistic: 45.14 on 3 and 22 DF,  p-value: 1.435e-09
```

The model prefered by BIC is:

Height = -193.07 + 107.80*nitro + 304.24 * phos + 143.13 * potas

Part(c)

Except **phos**, all other terms are significant at the 0.05 level.

Note: There is an outlier (ID=7) in the data. If deleted, the best model should be: Height = -217.82 + 114.75*nitro + 463.81 * phos + 100.33 * potas. All terms in this model are significant at the 0.05 level. The model plots seem better than previous model plots

```
larch1 = subset(larch, id != 7) ## subset without observation ID=7.
lm22 = lm(height~., data=larch1)
best.aic = step(lm22)
summary(best.aic)
best.bic = step(lm22, k=log(n))
summary(best.bic)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -217.82      31.92  -6.824 9.54e-07 ***
nitro         114.75      19.87   5.774 9.89e-06 ***
```

```
phos          463.81     152.12   3.049   0.0061 **
potas         100.33      40.66   2.467   0.0223 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 32.75 on 21 degrees of freedom
Multiple R-Squared: 0.9008,     Adjusted R-squared: 0.8866
F-statistic: 63.56 on 3 and 21 DF,  p-value: 1.049e-10
```

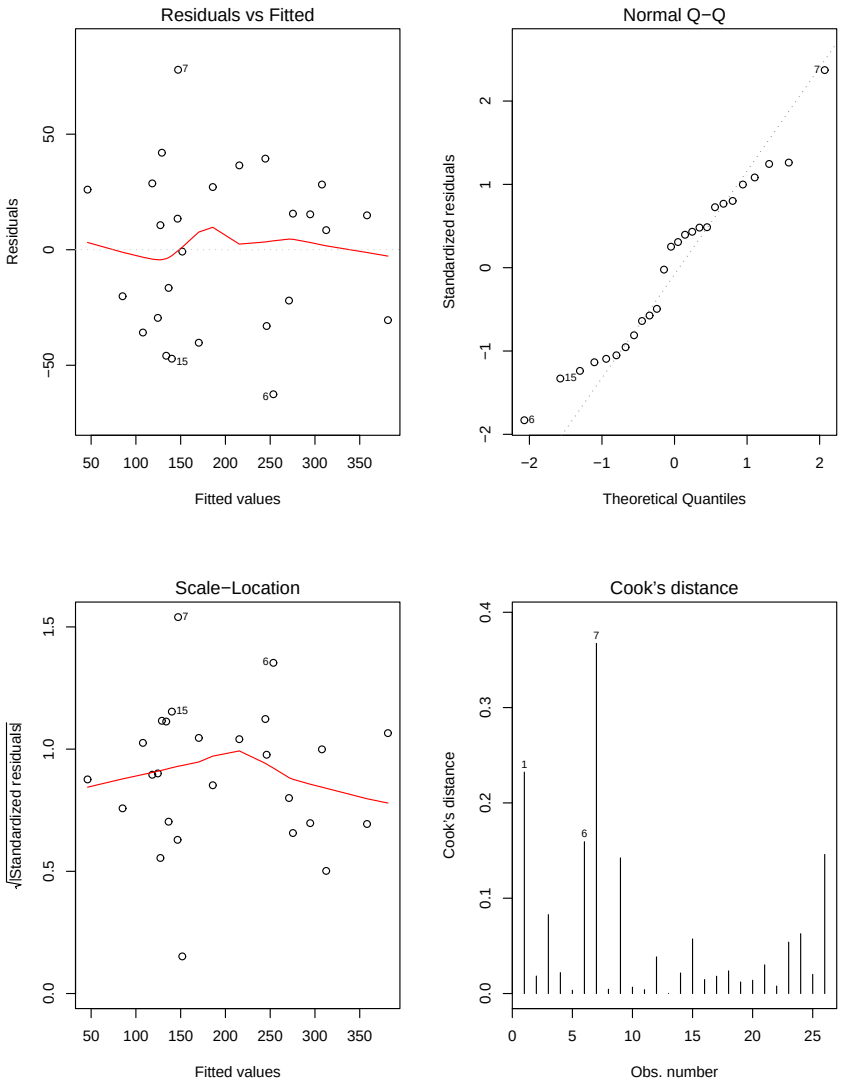

Figure 1: Model Checking Plots for Problem 2