**Question 1**

(a)

**> model\_1.1 = lm(log(Displacement)~Length + Beam,data = destroyers)**

**> summary(model\_1.1)**

Call:

lm(formula = log(Displacement) ~ Length + Beam, data = destroyers)

Residuals:

Min 1Q Median 3Q Max

-0.40664 -0.05649 0.05264 0.09835 0.23408

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.6504403 0.1299365 35.790 < 2e-16 \*\*\*

Length 0.0039436 0.0009044 4.361 0.000141 \*\*\*

Beam 0.0352421 0.0066055 5.335 9.04e-06 \*\*\*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Residual standard error: 0.1487 on 30 degrees of freedom

Multiple R-squared: 0.967, Adjusted R-squared: 0.9648

F-statistic: 439 on 2 and 30 DF, p-value: < 2.2e-16

(b)

**> model\_1.2 = lm(log(Displacement)~I(Length^2)+ I(Beam^2) + Length\*Beam,data = destroyers)**

**> summary(model\_1.2)**

Call:

lm(formula = log(Displacement) ~ I(Length^2) + I(Beam^2) + Length \*

Beam, data = destroyers)

Residuals:

Min 1Q Median 3Q Max

-0.144297 -0.032167 -0.007257 0.028312 0.160636

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.919e+00 4.639e-01 6.292 9.84e-07 \*\*\*

I(Length^2) 4.801e-05 2.683e-05 1.789 0.08481 .

I(Beam^2) 2.844e-03 1.524e-03 1.866 0.07290 .

Length 6.895e-04 5.723e-03 0.120 0.90500

Beam 1.569e-01 4.459e-02 3.520 0.00155 \*\*

Length:Beam -8.733e-04 4.175e-04 -2.092 0.04601 \*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Residual standard error: 0.08167 on 27 degrees of freedom

Multiple R-squared: 0.991, Adjusted R-squared: 0.9894

F-statistic: 596.5 on 5 and 27 DF, p-value: < 2.2e-16

(c)

Beam and Length:Beam appear to be significant. Based on hierarchy, we should maintain Length and only remove I(Length^2) and I(Beam^2).

(d)

**> anova(model\_1.1,model\_1.2)**

Analysis of Variance Table

Model 1: log(Displacement) ~ Length + Beam

Model 2: log(Displacement) ~ I(Length^2) + I(Beam^2) + Length \* Beam

Res.Df RSS Df Sum of Sq F Pr(>F)

1 30 0.66311

2 27 0.18009 3 0.48303 24.14 8.388e-08 \*\*\*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

At least one of the additional term in model\_1.2 is significantly not zero. So we conclude that the second-order polynomial model is better.

(e)

**> drop1(model\_1.2, test="F")**

Single term deletions

Model:

log(Displacement) ~ I(Length^2) + I(Beam^2) + Length \* Beam

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 0.18009 -159.96

I(Length^2) 1 0.021352 0.20144 -158.26 3.2013 0.08481 .

I(Beam^2) 1 0.023231 0.20332 -157.95 3.4830 0.07290 .

Length:Beam 1 0.029179 0.20927 -157.00 4.3748 0.04601 \*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

**> model\_1.2 <- update(model\_1.2, . ~ . - I(Length^2))**

**> drop1(model\_1.2, test="F")**

Single term deletions

Model:

log(Displacement) ~ I(Beam^2) + Length + Beam + Length:Beam

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 0.20144 -158.26

I(Beam^2) 1 0.001884 0.20332 -159.95 0.2619 0.61285

Length:Beam 1 0.027870 0.22931 -155.98 3.8739 0.05902 .

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

**> model\_1.2 <- update(model\_1.2, . ~ . - I(Beam^2))**

**> drop1(model\_1.2, test="F")**

Single term deletions

Model:

log(Displacement) ~ Length + Beam + Length:Beam

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 0.20332 -159.95

Length:Beam 1 0.45979 0.66311 -122.94 65.58 6.255e-09 \*\*\*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

My final model contains Length, Beam and the interaction term Length\*Beam as predictors

**Question 2**

(a)

**> model\_2.1<-lm(weight~feed,data=chickwts)**

**> summary(model\_2.1)**

Call:

lm(formula = weight ~ feed, data = chickwts)

Residuals:

Min 1Q Median 3Q Max

-123.909 -34.413 1.571 38.170 103.091

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 323.583 15.834 20.436 < 2e-16 \*\*\*

feedhorsebean -163.383 23.485 -6.957 2.07e-09 \*\*\*

feedlinseed -104.833 22.393 -4.682 1.49e-05 \*\*\*

feedmeatmeal -46.674 22.896 -2.039 0.045567 \*

feedsoybean -77.155 21.578 -3.576 0.000665 \*\*\*

feedsunflower 5.333 22.393 0.238 0.812495

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Residual standard error: 54.85 on 65 degrees of freedom

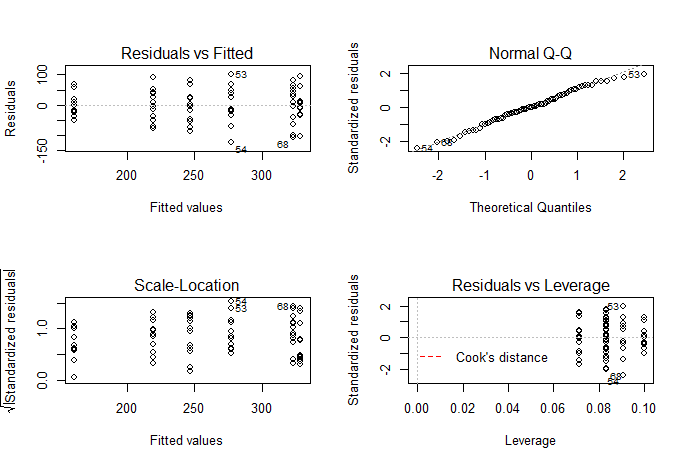
Multiple R-squared: 0.5417, Adjusted R-squared: 0.5064

F-statistic: 15.36 on 5 and 65 DF, p-value: 5.936e-10

(b)

**> par(mfrow=c(2,2))**

**> plot(model\_2.1, add.smooth=FALSE)**



The diagnostic plot indicates a good fit of model. It does not show any problems of heteroscedasticity or non-normality or outliers

(c)

**> anova(model\_2.1)**

Analysis of Variance Table

Response: weight

Df Sum Sq Mean Sq F value Pr(>F)

feed 5 231129 46226 15.365 5.936e-10 \*\*\*

Residuals 65 195556 3009

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

(d) Since the p-value of F-test is 5.936e-10, we conclude that there are differences between mean

weights of the groups at 5% significant level.

(e)

**> TukeyHSD(aov(model\_2.1, data=chickwts))**

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = model\_2.1, data = chickwts)

$feed

diff lwr upr p adj

horsebean-casein -163.383333 -232.346876 -94.41979 0.0000000

linseed-casein -104.833333 -170.587491 -39.07918 0.0002100

meatmeal-casein -46.674242 -113.906207 20.55772 0.3324584

soybean-casein -77.154762 -140.517054 -13.79247 0.0083653

sunflower-casein 5.333333 -60.420825 71.08749 0.9998902

linseed-horsebean 58.550000 -10.413543 127.51354 0.1413329

meatmeal-horsebean 116.709091 46.335105 187.08308 0.0001062

soybean-horsebean 86.228571 19.541684 152.91546 0.0042167

sunflower-horsebean 168.716667 99.753124 237.68021 0.0000000

meatmeal-linseed 58.159091 -9.072873 125.39106 0.1276965

soybean-linseed 27.678571 -35.683721 91.04086 0.7932853

sunflower-linseed 110.166667 44.412509 175.92082 0.0000884

soybean-meatmeal -30.480519 -95.375109 34.41407 0.7391356

sunflower-meatmeal 52.007576 -15.224388 119.23954 0.2206962

sunflower-soybean 82.488095 19.125803 145.85039 0.0038845

(f) The following pairs of feeds have significantly different means: horsebean-casein; linseed-casein; soybean-casein; meatmeal-horsebean; soybean-horsebean; sunflower-horsebean; sunflower-linseed; sunflower-soybean

**Question 3**

(a) 6 treatment groups with 9 experiment units in each group.

(b)

**> model\_3.1<-lm(breaks~wool \* tension, data=warpbreaks)**

**> summary(model\_3.1)**

Call:

lm(formula = breaks ~ wool \* tension, data = warpbreaks)

Residuals:

Min 1Q Median 3Q Max

-19.5556 -6.8889 -0.6667 7.1944 25.4444

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 44.556 3.647 12.218 2.43e-16 \*\*\*

woolB -16.333 5.157 -3.167 0.002677 \*\*

tensionM -20.556 5.157 -3.986 0.000228 \*\*\*

tensionH -20.000 5.157 -3.878 0.000320 \*\*\*

woolB:tensionM 21.111 7.294 2.895 0.005698 \*\*

woolB:tensionH 10.556 7.294 1.447 0.154327

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Residual standard error: 10.94 on 48 degrees of freedom

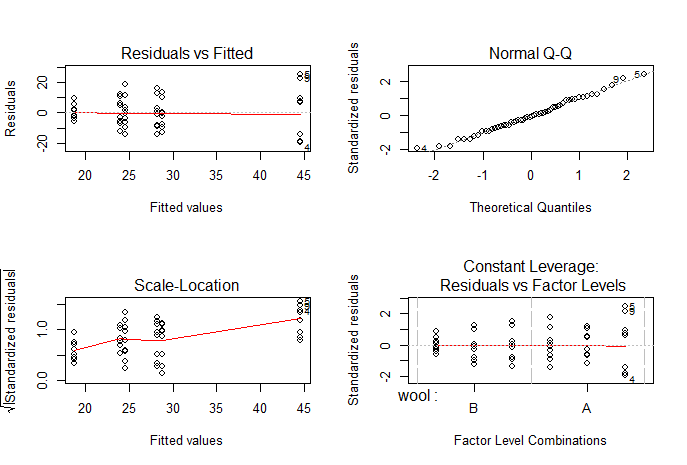
Multiple R-squared: 0.3778, Adjusted R-squared: 0.3129

F-statistic: 5.828 on 5 and 48 DF, p-value: 0.0002772

(c)

**> par(mfrow=c(2,2))**

**> plot(model\_3.1)**



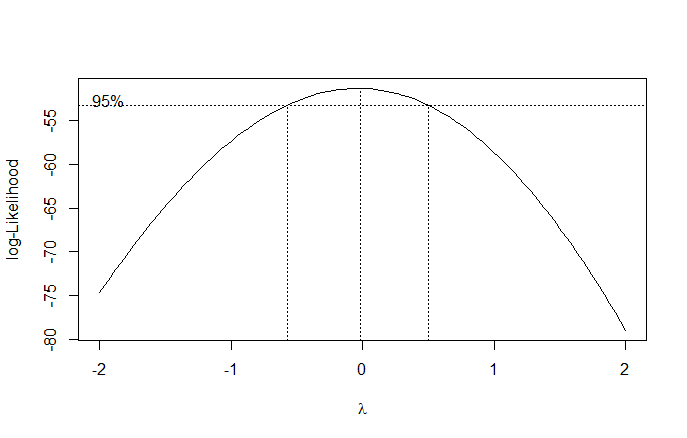
The diagnostic plot indicates heteroscedasticiy

(d)

**> library(MASS)**

**> par(mfrow=c(1,1))**

**> boxcox(model\_3.1)**



The simple transformation of y is log(y) based on lamda = 0

(e)

**> model\_3.2<-lm(log(breaks)~wool \* tension, data=warpbreaks)**

**> summary(model\_3.2)**

Call:

lm(formula = log(breaks) ~ wool \* tension, data = warpbreaks)

Residuals:

Min 1Q Median 3Q Max

-0.81504 -0.27885 0.04042 0.27319 0.64358

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.7179 0.1247 29.824 < 2e-16 \*\*\*

woolB -0.4356 0.1763 -2.471 0.01709 \*

tensionM -0.6012 0.1763 -3.410 0.00133 \*\*

tensionH -0.6003 0.1763 -3.405 0.00134 \*\*

woolB:tensionM 0.6281 0.2493 2.519 0.01514 \*

woolB:tensionH 0.2221 0.2493 0.891 0.37749

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

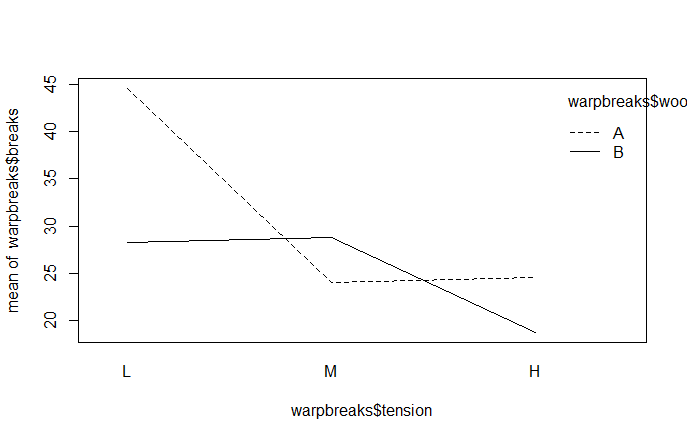
Residual standard error: 0.374 on 48 degrees of freedom

Multiple R-squared: 0.3363, Adjusted R-squared: 0.2672

F-statistic: 4.864 on 5 and 48 DF, p-value: 0.001116

(f)

**> interaction.plot(warpbreaks$tension, warpbreaks$wool, warpbreaks$breaks)**



(g)

**> anova(model\_3.2)**

Analysis of Variance Table

Response: log(breaks)

Df Sum Sq Mean Sq F value Pr(>F)

wool 1 0.3125 0.31253 2.2344 0.141511

tension 2 2.1762 1.08808 7.7792 0.001185 \*\*

wool:tension 2 0.9131 0.45657 3.2642 0.046863 \*

Residuals 48 6.7138 0.13987

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

(h)

There is interaction between wool and tension, which effects the number of breaks in yarn during

weaving.

(i)

Since the interaction is already significant, based on hierarchy, there is no need to test for main effect.