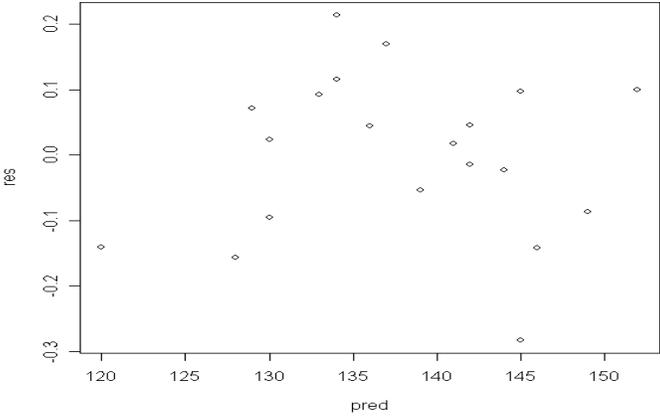
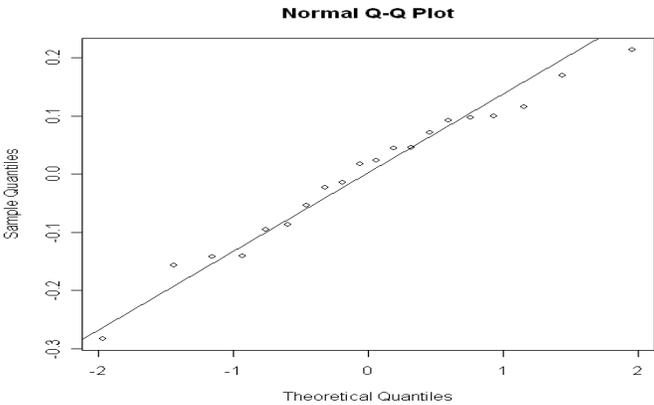
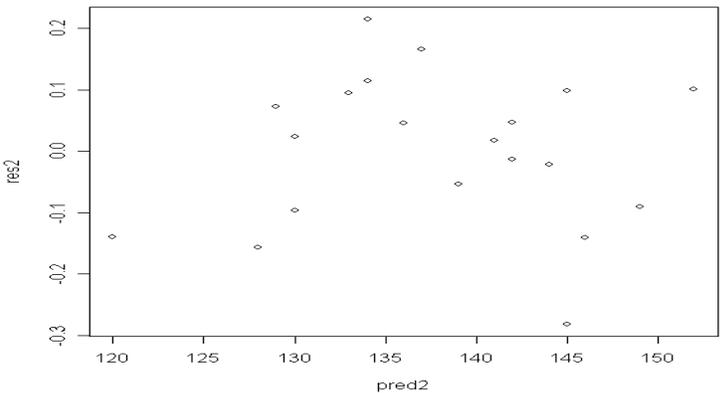


Stat 241/251

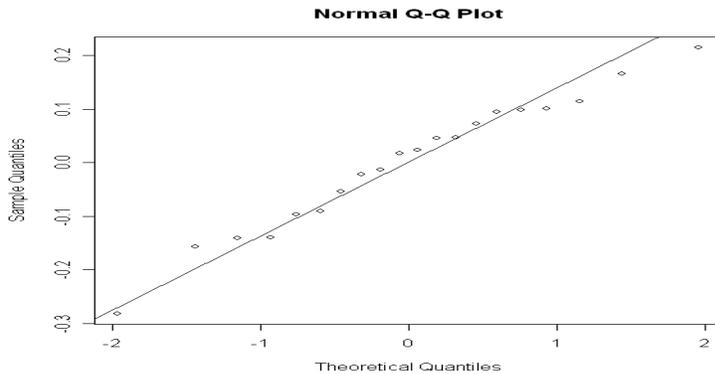
Lab 5: Multivariate Linear Regression and Regression Analysis

Ehsan. Karim (<http://www.stat.ubc.ca/People/Home/index.php?person=e.karim>)

#	R console Command	Details
1	<pre>> bp = read.table("http://ehsan.karim.googlepages. com/blood_pressure.txt", head=T)</pre>	Read data set from web
2	<pre>> attach(bp)</pre>	Attach the data set in R
3	<pre>> names(bp) [1] "bpress" "height" "age"</pre>	Look-up the names of the variables in the data set
4	<pre>> multireg = lm(bpress ~ height + age)</pre>	Run multiple linear regression using two covariates (model 1)
5	<pre>> res = residuals(multireg)</pre>	Get the residuals (model 1)
6	<pre>> pred = predict(multireg)</pre>	Get fitted values (model 1)
7	<pre>> plot(pred, res)</pre>	Plot the fitted vs. residual (model 1) Seems random (no trend): assumption satisfied.
	 <p>A scatter plot showing residuals on the y-axis (ranging from -0.3 to 0.2) against predicted values on the x-axis (ranging from 120 to 150). The data points are scattered around the zero line, indicating that the residuals are randomly distributed, which supports the assumption of no trend in the residuals.</p>	
8	<pre>> qqnorm(res) > qqline(res)</pre>	Check normality assumption (model 1) Assumption approximately satisfied.
	 <p>A Normal Q-Q Plot with 'Sample Quantiles' on the y-axis (ranging from -0.3 to 0.2) and 'Theoretical Quantiles' on the x-axis (ranging from -2 to 2). The data points closely follow a diagonal reference line, suggesting that the residuals are approximately normally distributed.</p>	
9	<pre>> summary(multireg)</pre>	Get summary (model 1): notice that the significant value of

	<pre> Residuals: Min 1Q Median 3Q Max -0.28230 -0.08877 0.02083 0.09414 0.21347 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.000e+02 4.828e-01 207.218 <2e-16 *** height -3.834e-04 7.054e-03 -0.054 0.957 age 9.997e-01 3.839e-03 260.428 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.1307 on 17 degrees of freedom Multiple R-Squared: 0.9998, Adjusted R-squared: 0.9997 F-statistic: 3.659e+04 on 2 and 17 DF, p-value: < 2.2e-16 </pre>	<p>“height” covariate is more than 0.10: which means this covariate is not significant. Also note the R-square value.</p>
10	<pre>> multireg2 = lm(bpress ~ age)</pre>	<p>Run multiple linear regression using one covariates, excluding "height" (model 2)</p>
11	<pre>> summary(multireg2)</pre> <pre> Residuals: Min 1Q Median 3Q Max -0.28150 -0.09176 0.02020 0.09538 0.21468 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.000e+02 1.387e-01 721.1 <2e-16 *** age 9.997e-01 3.592e-03 278.3 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.127 on 18 degrees of freedom Multiple R-Squared: 0.9998, Adjusted R-squared: 0.9998 F-statistic: 7.747e+04 on 1 and 18 DF, p-value: < 2.2e-16 </pre>	<p>Get summary (model 2): notice that the significant value of covariate is less than 0.10, and hence significant in the model</p>
12	<pre>> res2 = residuals(multireg2) > pred2 = predict(multireg2) > plot(pred2, res2)</pre> 	<p>Get the residuals (model 2) Get fitted values (model 2) Plot the fitted vs. residual (model 2) Seems random (no trend): assumption satisfied.</p>

```
13 > qqnorm(res2)
> qqline(res2)
```



Check normality assumption (model 2)
Assumption approximately satisfied.

```
14 > bpress2 = log(bpress)
> multireg3 = lm(bpress2 ~ height + age)
> summary(multireg3)
```

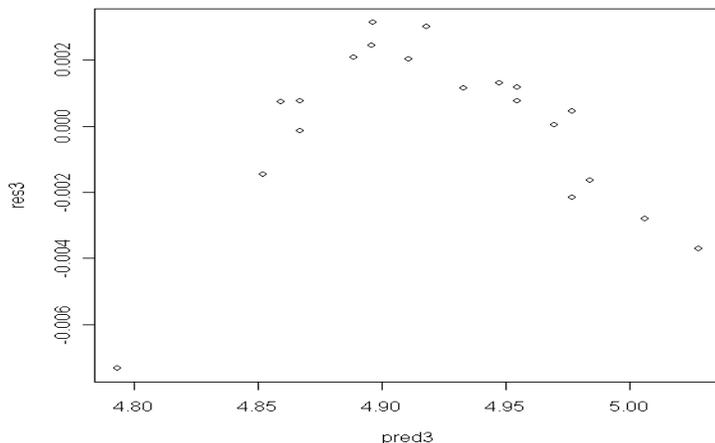
```
Residuals:
    Min       1Q   Median       3Q      Max
-0.0073140 -0.0014941  0.0007544  0.0014895  0.0031373

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.648e+00  9.992e-03  465.137  <2e-16 ***
height      -9.751e-06  1.460e-04  -0.067   0.948
age          7.330e-03  7.945e-05  92.265  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Residual standard error: 0.002705 on 17 degrees of
freedom
Multiple R-Squared:  0.9982,    Adjusted R-squared:
 0.9979
F-statistic: 4591 on 2 and 17 DF,  p-value: < 2.2e-16
```

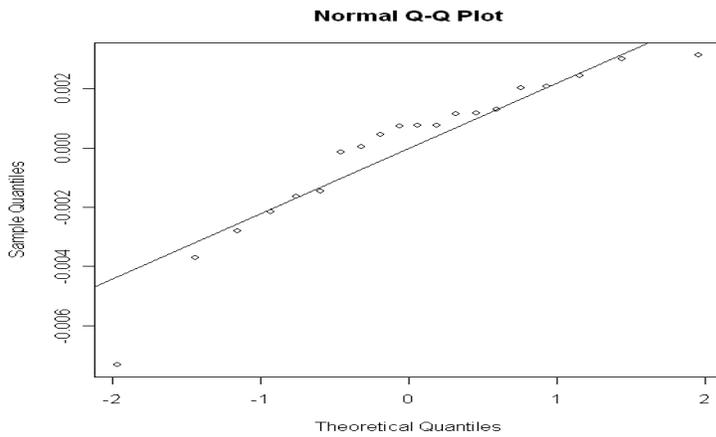
Run multiple linear regression by (log) transforming the dependent variable (model 3)
Get summary (model 3):
notice that the significant value of “height” covariate is more than 0.10: which means this covariate is not significant. Also note the R-square value.
(overall scenario just like model 1)

```
15 > res3 = residuals(multireg3)
> pred3 = predict(multireg3)
> plot(pred3, res3)
```



Get the residuals (model 3)
Get fitted values (model 3)
Plot the fitted vs. residual (model 3)
Does not seem random (trend present): assumption violated. This model is thus not acceptable due to violation of one assumption.

16 > qqnorm(res3)
> qqline(res3)



Check normality assumption (model 3)
Assumption approximately satisfied.

Conclusion:

#	All covariates significant? (p-value less than 0.10)	Assumptions satisfied?	R-square	Model Acceptable?
Model 1: covariate height, age	No, height covariate is not significant	Yes	0.9998	No
Model 2: covariate age	Yes	Yes	0.9998	Yes
Model 3: covariate height, age, but dependent variable log transformed	No, height covariate is not significant	No, fitted vs. residual plot does not seem random	0.9982	No