Lesson 7 R Activity 2

Rick Davila

5/10/2020

Lesson 7, Windmill Example - Install packages

Install necessary packages using library()

Perform data housekeeping - upload, name columns, display to make sure it reads properly, etc.

```
knitr::opts_chunk$set(echo = TRUE)

library(e1071)
library(xtable)
library("xlsx") # Needed to read data
```

```
## Warning: package 'xlsx' was built under R version 4.0.3
```

```
library(psych) # For geometric mean in Example 5.3
```

```
## Warning: package 'psych' was built under R version 4.0.3
```

```
rm(list = ls())
```

Read data file (data-ex-5-2.xlsx)

```
ex5_2 <- read.xlsx("data-ex-5-2.xlsx",
                   sheetIndex = 1,
                   colIndex = c(2,3),
                   as.data.frame = TRUE,
                   header = TRUE)
```

Assign labels to data columns using names() and attach() commands

```
names(ex5_2) <- c("Wind_Velocity", "DC_Output")
attach(ex5_2)
```

Output data to make sure it reads properly

```
out <- as.data.frame(c(ex5_2))
colnames(out) <- c("Wing_Velocity", "DC_Output")
tab <- (xtable(out, digits=c(0,2,3)))
print(tab, type="html")
```

	Wing_Velocity	DC_Output

1	5.00	1.582
2	6.00	1.822
3	3.40	1.057
4	2.70	0.500
5	10.00	2.236
6	9.70	2.386
7	9.55	2.294
8	3.05	0.558
9	8.15	2.166
10	6.20	1.866
11	2.90	0.653
12	6.35	1.930
13	4.60	1.562
14	5.80	1.737
15	7.40	2.088
16	3.60	1.137
17	7.85	2.179
18	8.80	2.112
19	7.00	1.800
20	5.45	1.501
21	9.10	2.303
22	10.20	2.310
23	4.10	1.194
24	3.95	1.144
25	2.45	0.123

```
# Output data structure and dimensions
str(ex5_2)
```

'data.frame': 25 obs. of 2 variables: $ Wind_Velocity: num 5 6 3.4 2.7 10 9.7 9.55 3.05 8.15 6.2 … $ DC_Output : num 1.58 1.82 1.06 0.5 2.24 …

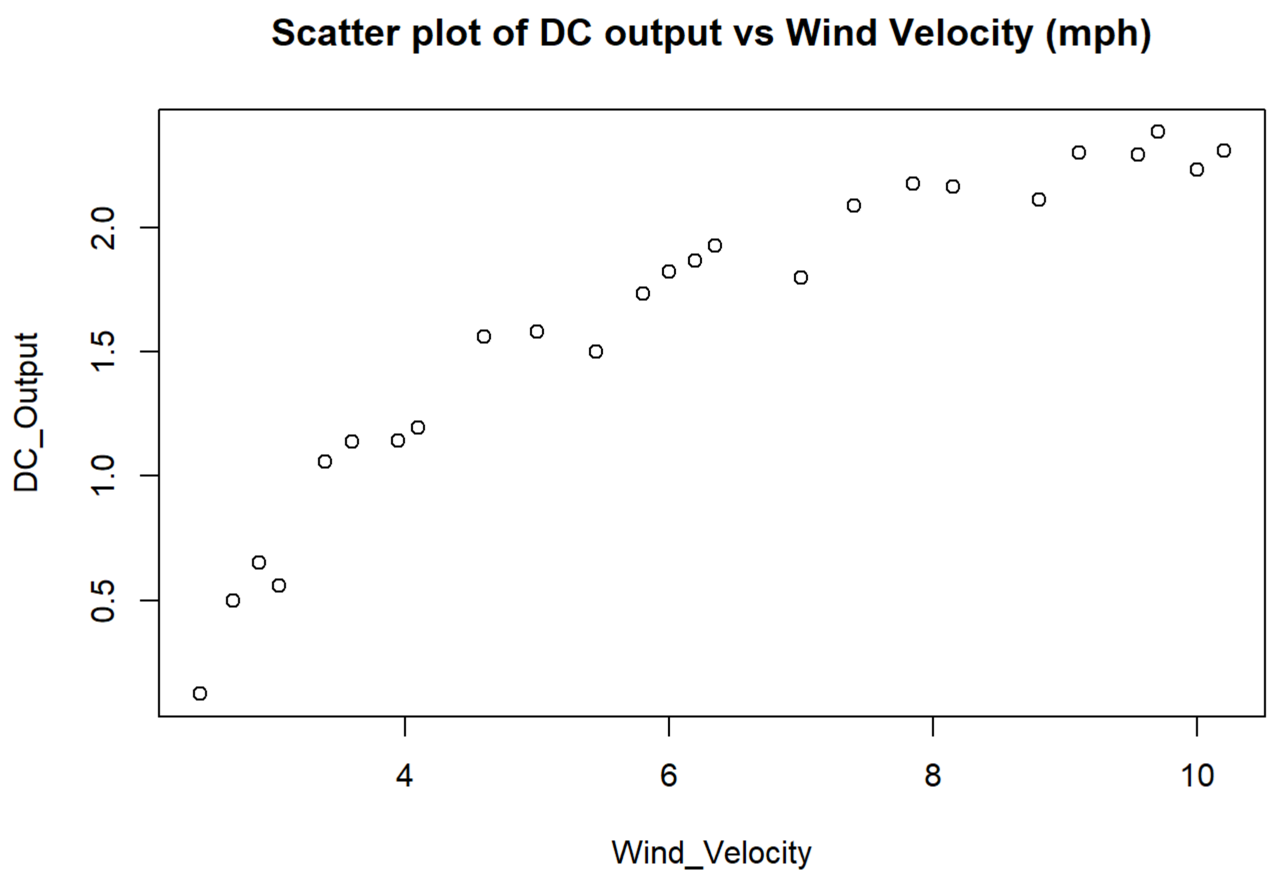
```
dim(ex5_2)
```

[1] 25 2

Example 5.2 (p. 178-181)

Create scatterplot to visualize data

```
plot(Wind_Velocity,DC_Output, main = "Scatter plot of DC output vs Wind Velocity (mph)")
```



Fit linear model first; compare results to p. 178

```
# Obtain regression estimates using lm command
model <- lm(DC_Output ~ Wind_Velocity)
beta_0 <- model$coefficients[1]
beta_1 <- model$coefficients[2]

# show beta estimates, std. error, tvalues and p-values
xtable(model)
```

	Estimate <dbl>	Std. Error <dbl>	t value <dbl>	Pr(>|t|) <dbl>
(Intercept)	0.1308751	0.1259894	1.038779	3.097053e-01
Wind_Velocity	0.2411489	0.0190492	12.659268	7.545525e-12

2 rows

```
xtable(summary(aov(model)))
```

	Df <dbl>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
Wind_Velocity	1	8.929615	8.92961484	160.2571	7.545525e-12

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Residuals	23	1.281573	0.05572057	NA	NA

2 rows

The equation for the linear regression model:

$$\hat{y} = (0.1308751) + (0.2411489)x$$

```
# ANOVA table
# summary(aov)
out <- anova(model)

xtable(out)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
Wind_Velocity	1	8.929615	8.92961484	160.2571	7.545525e-12
Residuals	23	1.281573	0.05572057	NA	NA

2 rows

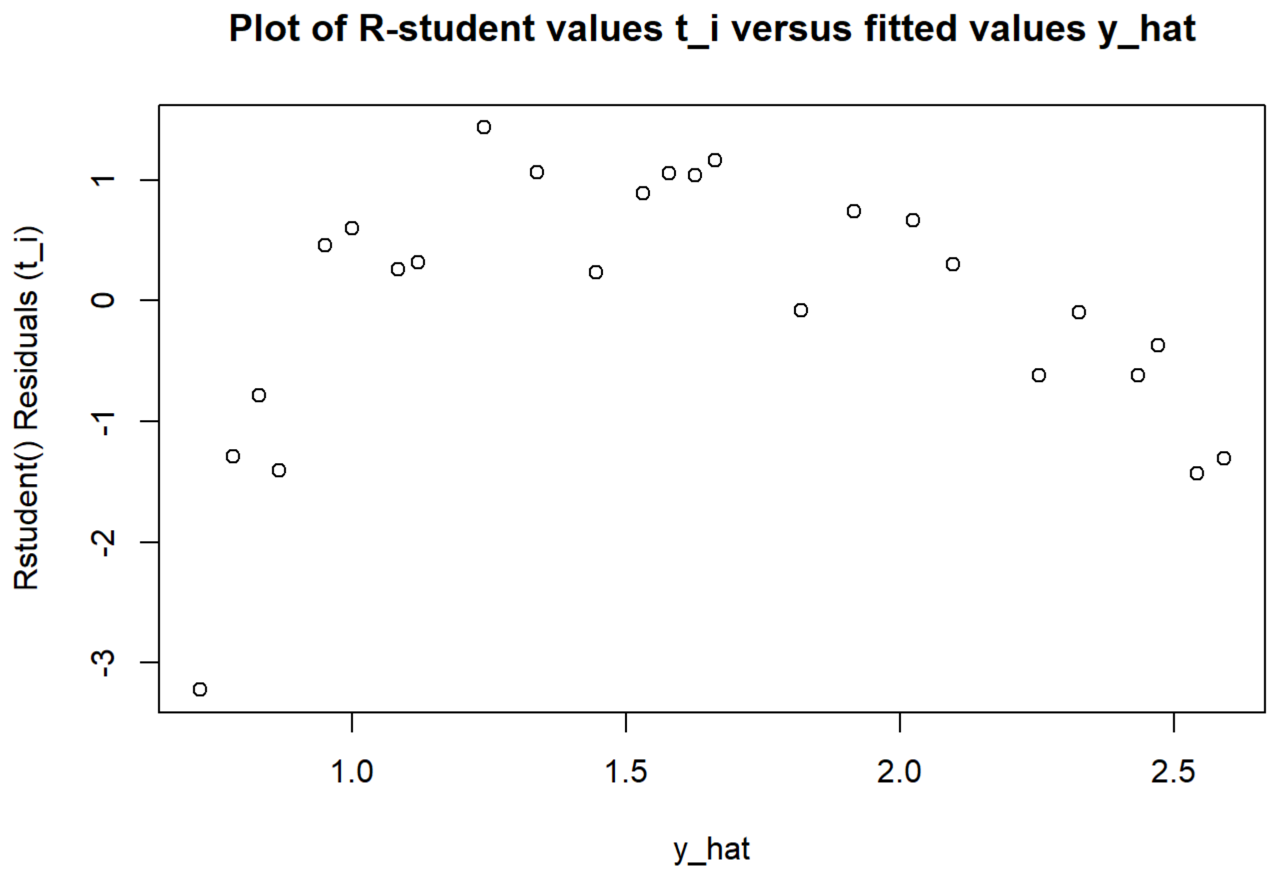
Create residual plots - R-Student and raw residuals versus fitted values

```
# DC_Output, given Wind_Velocity
model <- lm(DC_Output ~ Wind_Velocity)
beta_0 <- model$coefficients[1]
beta_1 <- model$coefficients[2]

# Calculate y_hat
y_hat <- beta_0 + beta_1*Wind_Velocity

# Calculate R-Student residuals
residuals <- rstudent(model)

plot(y_hat, residuals, main='Plot of R-student values t_i versus fitted values y_hat', xlab = 'y
_hat', ylab = 'Rstudent() Residuals (t_i)')
```

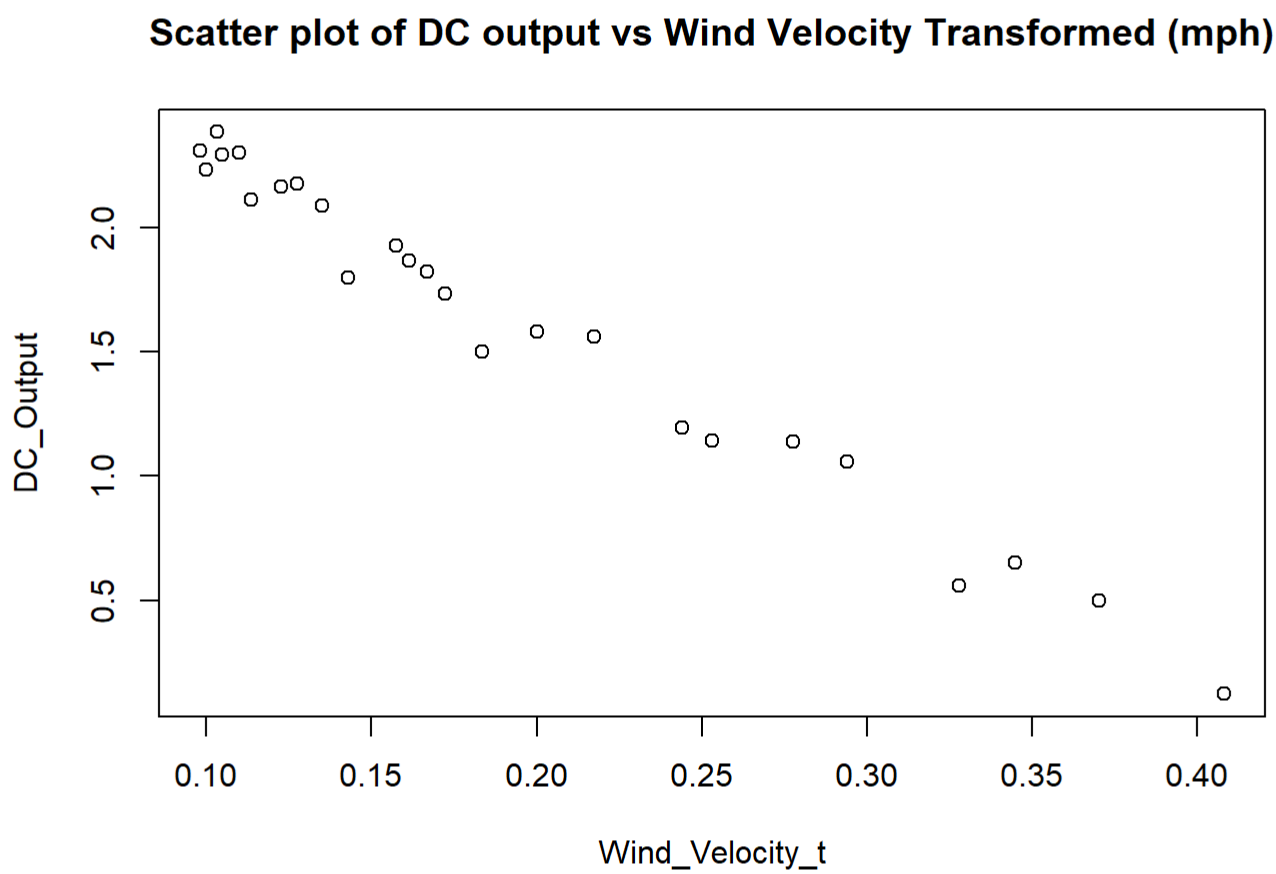


Residual plots show clear nonlinearity. Attempt to rectify through transformation: x' = 1/x

```
Wind_Velocity_t <- 1/Wind_Velocity
```

Create scatterplot and model using transformed data; compare to p. 180 and 181

```
plot(Wind_Velocity_t,DC_Output, main = "Scatter plot of DC output vs Wind Velocity Transformed
 (mph)")
```




```
# create lm() model with transformed wind velocity
model_t <- lm(DC_Output ~ Wind_Velocity_t)

#
xtable(model_t)
```

	Estimate <dbl>	Std. Error <dbl>	t value <dbl>	Pr(>|t|) <dbl>
(Intercept)	2.978860	0.04490227	66.34096	8.917745e-28
Wind_Velocity_t	-6.934547	0.20643355	-33.59215	4.742548e-21

2 rows

```
xtable(summary(aov(model_t)))
```

	Df <dbl>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
Wind_Velocity_t	1	10.0072184	10.007218353	1128.433	4.742548e-21
Residuals	23	0.2039696	0.008868246	NA	NA

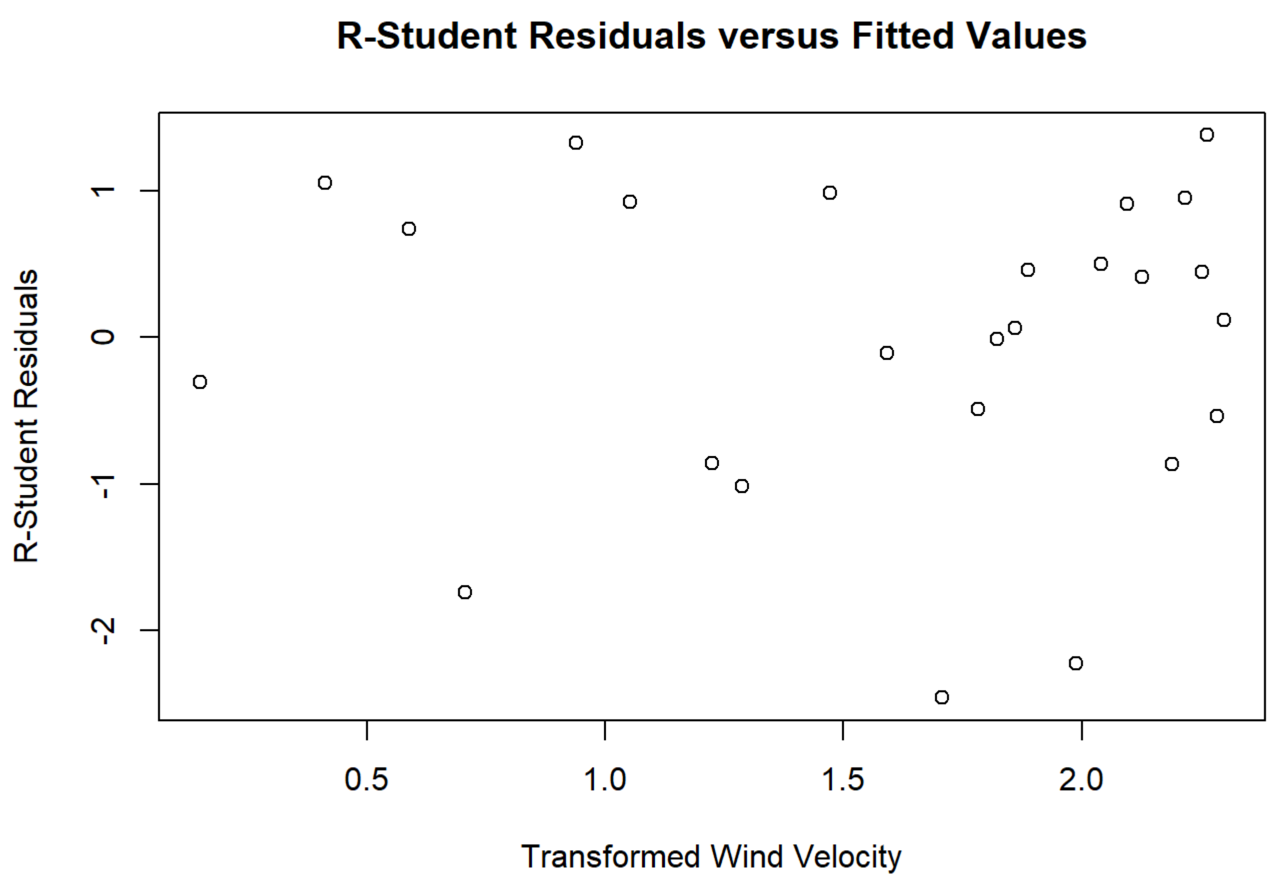
2 rows

Plot residuals versus fitted values; compare to p. 182

```
# Calculate r_student residuals
R_Student_Residuals_t <- rstudent(model_t)

y_hat_t <- model_t$fitted.values

plot(y_hat_t, R_Student_Residuals_t, ylab = "R-Student Residuals", xlab = "Transformed Wind Velo
city", main = "R-Student Residuals versus Fitted Values")
```



Example 5.4 (p. 187)

Determine alpha value to transform independent variable using Box-Tidwell Procedure

Step 1.0: Define alpha_0 = 1 and run simple regression model (performed in Example 5.1 above)

```
# Obtain regression estimates using lm command, for alpha_0 = 0
model <- lm(DC_Output ~ Wind_Velocity)
beta_0 <- model$coefficients[1]
beta_1 <- model$coefficients[2]

# show beta estimates, std. error, tvalues and p-values
xtable(model)
```

	Estimate	Std. Error	t value	Pr(>|t|)
	<dbl>	<dbl>	<dbl>	<dbl>
(Intercept)	0.1308751	0.1259894	1.038779	3.097053e-01
Wind_Velocity	0.2411489	0.0190492	12.659268	7.545525e-12

2 rows

```
xtable(summary(aov(model)))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Wind_Velocity	1	8.929615	8.92961484	160.2571	7.545525e-12
Residuals	23	1.281573	0.05572057	NA	NA

2 rows

For step 1.0 we have, the equation for the linear regression model is:

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1}x = (0.1308751) + (0.2411489)x$$

Step 1.1: Define 2 = x ln x and add to model as an independent variable; compare output to p. 187

```
# Then defining w = x ln x, we fit Eq. (5.8) and obtain
w <- Wind_Velocity*log(Wind_Velocity)          #  w = xlog(x)
model <- lm(DC_Output ~ Wind_Velocity + w)
beta_0_star <- model$coefficients[1]
beta_1_star <- model$coefficients[2]
gamma_1 <- model$coefficients[3]

# show beta estimates, std. error, tvalues and  p-values
xtable(model)
```

	Estimate	Std. Error	t value	Pr(>|t|)
	<dbl>	<dbl>	<dbl>	<dbl>
(Intercept)	-2.4168438	0.28511523	-8.476726	2.228595e-08
Wind_Velocity	1.5344349	0.14189388	10.813961	2.853452e-10
w	-0.4625963	0.05065422	-9.132435	6.127920e-09

3 rows

```
xtable(summary(aov(model)))
```

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Wind_Velocity	1	8.9296148	8.92961484	734.40491	2.147909e-18
w	1	1.0140756	1.01407560	83.40137	6.127920e-09
Residuals	22	0.2674976	0.01215898	NA	NA

3 rows

For step 1.1 we have, the equation for the linear regression model is:

$$\hat{y} = \hat{\beta}_0^* + \hat{\beta}_1^* x + \hat{\gamma} w = (-2.4168438) + (1.5344349)x + (-0.4625963)w$$

Step 1.2: Using Equation 5.10, calculate alpha_1; compare to p. 187

```
alpha_1 <- (gamma_1/beta_1) + 1
```

For step 1.2, we have:

$$\alpha_1 = \frac{\hat{\gamma}}{\hat{\beta}_1} + 1 = (-0.9183019)$$

Step 2.0: Define x’ = x^alpha_1 and regress model with alpha_0 = 0; compare to p. 187

```
# Obtain regression estimates using lm command, for alpha_0 = 1
Wind_Velocity_t1 <- Wind_Velocity^(alpha_1)
model <- lm(DC_Output ~ Wind_Velocity_t1)
beta_0_t <- model$coefficients[1]
beta_1_t <- model$coefficients[2]

# show beta estimates, std. error, tvalues and  p-values
xtable(model)
```

	Estimate	Std. Error	t value	Pr(>|t|)
	<dbl>	<dbl>	<dbl>	<dbl>
(Intercept)	3.103860	0.04747509	65.37871	1.245836e-27
Wind_Velocity_t1	-6.678423	0.19538144	-34.18146	3.203122e-21

2 rows

```
xtable(summary(aov(model)))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Wind_Velocity_t1	1	10.0140562	10.014056206	1168.372	3.203122e-21
Residuals	23	0.1971318	0.008570948	NA	NA

2 rows

For step 2.0, $x' = x^{-0.9183019}$ and the equation for the linear regression model is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x' = (3.10386) + (-6.6784226)x'$$

Step 2.1: Define w' = x' ln x', add as independent variable, and run model

```
# Obtain regression estimates using lm command
w_t1 <- Wind_Velocity_t1*log(Wind_Velocity_t1) # w = x'log(x')
model <- lm(DC_Output ~ Wind_Velocity_t1 + w_t1)
beta_0_star_t1 <- model$coefficients[1]
beta_1_star_t1 <- model$coefficients[2]
gamma_1_t1 <- model$coefficients[3]

# show beta estimates, std. error, tvalues and  p-values
xtable(model)
```

	Estimate <dbl>	Std. Error <dbl>	t value <dbl>	Pr(>|t|) <dbl>
(Intercept)	3.2409424	0.2707985	11.9680954	4.184077e-11
Wind_Velocity_t1	-6.4444831	0.4962074	-12.9874786	8.580514e-12
w_t1	0.5994122	1.1651551	0.5144484	6.120681e-01

3 rows

```
xtable(summary(aov(model)))
```

	Df <dbl>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
Wind_Velocity_t1	1	10.014056206	10.014056206	1131.0176009	2.070529e-20
w_t1	1	0.002343281	0.002343281	0.2646572	6.120681e-01
Residuals	22	0.194788513	0.008854023	NA	NA

3 rows

For step 2.1 we have, the equation for the linear regression model is:

$$\hat{y} = \hat{\beta_0}^* + \hat{\beta_1}^* x' + \hat{\gamma} w' = (3.2409424) + (-6.4444831)x' + (0.5994122)w'$$

Step 2.2: Using equation 5.10, calculate alpha_2; compare to value on p. 187. Pretty close to -1, which was the transformation in Example 5.2.

```
alpha_2 <- (gamma_1_t1/beta_1_t) + alpha_1
```

$$\alpha_2 = \frac{\hat{\gamma}}{\hat{\beta}_1} + \alpha_1 = (-1.0080555)$$