Remedial Measures in SLR Tutorial 2

For this second example, we will go over the windmill example from the textbook. The textbook describes the data as:

A research engineer is investigating the use of a windmill to generate electricity. He has collected data on the DC output from his windmill and the corresponding wind velocity.

Download the data file, `windmill.txt` and read the data in.

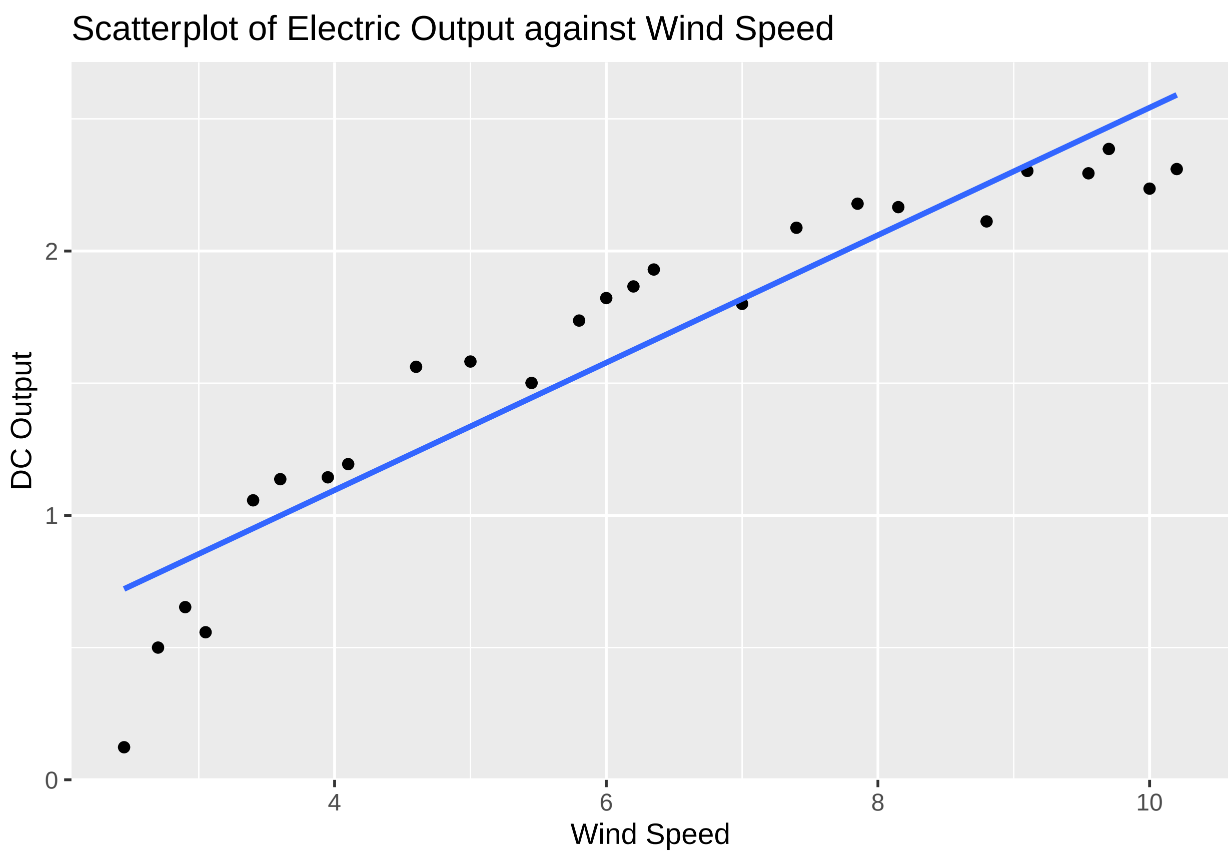
```
Data<-read.table("windmill.txt", header=TRUE, sep="")
```

1. Model Diagnostics with Scatterplots

We can use a scatterplot of the response variable against the predictor to assess assumptions 1 and 2.

```
library(tidyverse)

##scatterplot, and overlay regression line
ggplot2::ggplot(Data, aes(x=wind,y=output))+
  geom_point()+
  geom_smooth(method = "lm", se=FALSE)+
  labs(x="Wind Speed", y="DC Output",
       title="Scatterplot of Electric Output against Wind Speed")
```



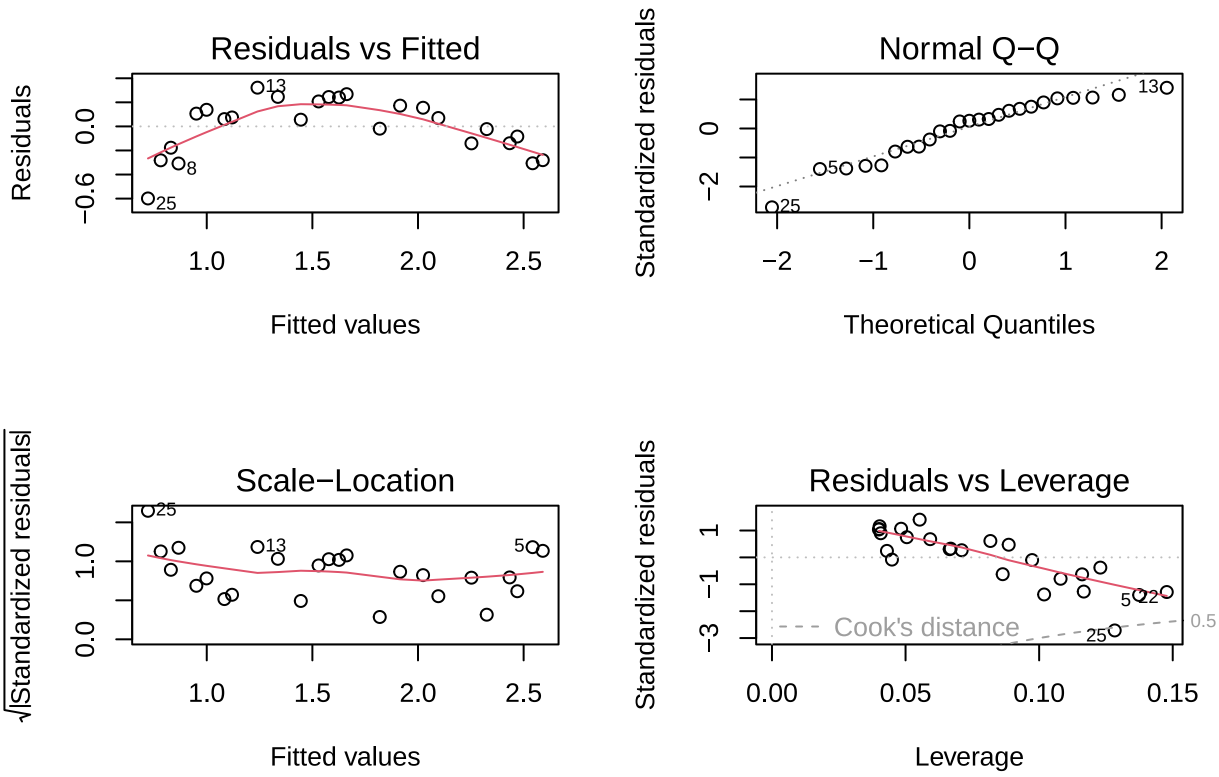
To assess assumption 1, the data points should be evenly scattered on both sides of the regression line, as we move from left to right. This plot looks nonlinear. For wind velocity between 4 and 8, plots are above the regression line. Otherwise, the plots are below the regression line. Assumption 1 is not met.

To assess assumption 2, the vertical spread of the data points should be constant as we move from left to right. There are not enough data points to assess this assumption with this plot.

2. Model Diagnostics with Residual Plots

Fairly often, when assessing regression assumptions, a residual plot is easier to visualize than a scatterplot.

```
result<-lm(output~wind, data=Data)
par(mfrow = c(2, 2))
plot(result)
```



From the residual plot (top left), we see a curved pattern, so we have a nonlinear relationship. Assumption 1 is not met.

From both plots (top left, bottom left), the vertical variance of the plots appear to be fairly constant, so assumption 2 is met.

Now that we know that assumption 1 is not met, we need to transform the predictor variable.

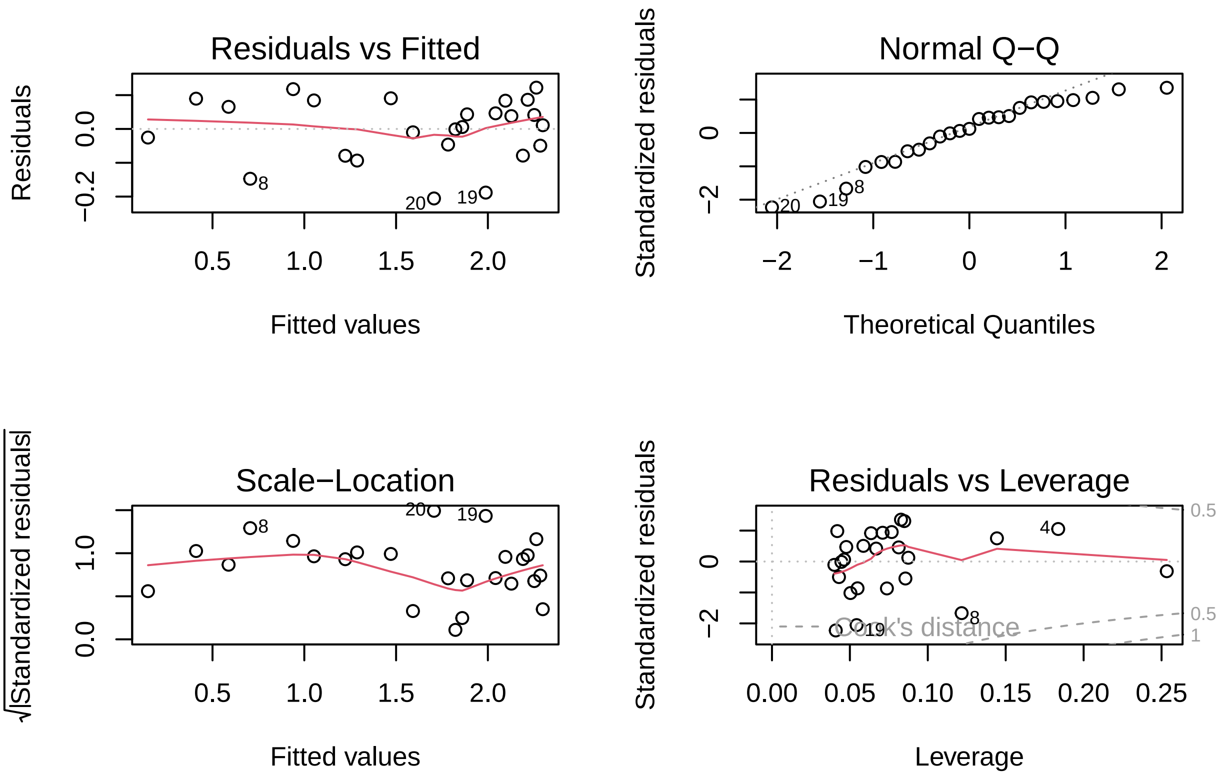
3. Transformation on x

Going back to the original scatterplot, the patterns appears to be an inverse shape, so we consider an inverse transform on the predictor. Note that the shape is similar to a log or square root so those transformations may also work. It turns out that the inverse makes sense for this engineering problem, as it is known that DC output is inversely proportional to wind speed.

```
Data$xstar<-1/(Data$wind)
result.xstar<-lm(output~xstar, data=Data)
```

We create the diagnostic plots just like earlier

```
par(mfrow = c(2, 2))
plot(result.xstar)
```



We need to reassess assumptions 1 and 2.

We see an improvement. Residuals more evenly scattered across horizontal axis in the residual plot (top left). Vertical spread appears constant as well. Assumptions 1 and 2 both met.

Let's report the estimated regression coefficients:

```
result.xstar
```

```
##
## Call:
## lm(formula = output ~ xstar, data = Data)
##
## Coefficients:
## (Intercept)        xstar
##       2.979       -6.935
```

$\hat{y} = 2.979 - 6.935x^*$, where $x^* = \frac{1}{x}$. Since we performed a transformation that is not a log transform, we cannot interpret the slope of the regression.