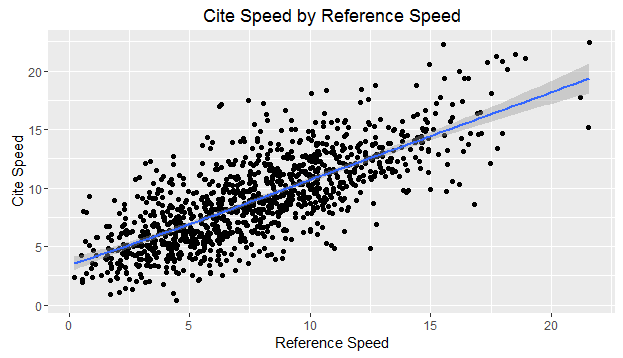
1. Windmill companies are always looking at new sites where they might invest in building a new wind farm. In order to determine if a "candidate site" is a good place to build, they need to be able to know that the windspeed at that site is high enough to create enough energy to pay off the costs of building. However, the data collection period can be costly in both time and money. Using statistical modeling, we can avoid the high costs in data by using windspeeds at "reference sites" to predict the windspeeds at a candidate site. A reference site is a nearby site at which the company is already monitoring the windspeed and it should be similar to the candidate site.

2.

The plot above shows a linear relationship between the reference speed and the cite speed. Generally, with the exception of a few outliers, as the reference speed goes up so does the cite speed. With a correlation of .756, we can see that there is a strong positive relationship between the reference speed and the site speed. This adds validity to the approach of using the reference speed to predict the cite speed. With these findings, an SLR model is appropriate for this analysis.

3.

Cspd = β0 + β1(RSpd) + ε

We are assuming in this model that each observation is independent of previous observations. In other words, one observed cite speed has no effect on the next observed cite speed. Also, we are assuming that the errors follow a normal distribution and that there is equal variance throughout the data.

CSpd = The predicted cite speed

RSpd = the speed at the reference site

β0 = The average predicted cite speed when the reference speed is 0.

β1 = the average increase in predicted cite speed when the reference speed increases by 1 unit.

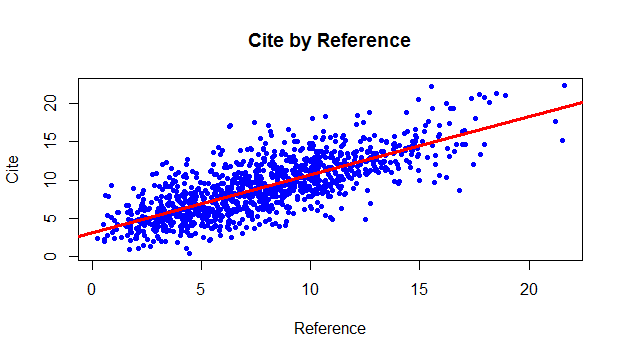
ε = the random deviation from the true value, known as stochastic error

After fitting this model to the data, we will be able to use a give reference speed to predict the speed at the nearby cite location. This will allow windmill companies to get a good estimate at the cite speed without having to spend time or money in the data collection stage.

4. The fitted model results look as follows:

Cspd = 3.14+ .756×(RSpd) + ε

With β0 = 3.14, we can say that 3.14 is the average predicted cite speed when the reference speed is 0. With β1 = .756, we infer that as the reference speed increases by 1 unit, the predicted cite speed will increase by .756 units.



The plot above is the data plotted with a fitted regression line from our model.

5. Using our fitted model it is easy to obtain a prediction of windspeed at the candidate site given the windspeed at the reference site. You simply plug the reference speed into the following equation:

Cspd = 3.14+ .756×(RSpd)

For example, if the reference speed was 12 m/s.

Cspd = 3.14+ .756×(12) = 12.212

The predicted cite speed would be 12.212 m/s.

6. There are limitations to this model. We really only have solid data for reference cites with speeds from 1-2 m/s up to 15-16 m/s. Making predictions about the cite speed outside of this range will most likely not yield valid results because we really don't have any data to fit our model to in order to ensure that our coefficients are effective.

For example, if the reference speed was 30 m/s.

Cspd = 3.14+ .756×(30) = 25.82

This prediction causes me to be more wary of the model that we are using because it is so low. Trying to predict outside of the range of your data is called extrapolation and we must avoid it.

## Appendix (R code)

#### WINDMILL ANALYSIS ####

setwd("~/3Fall2016/stat330/Homework")

wind <- read.table("Windmill.txt", header=TRUE)

head(wind)

## 2 ##

require(ggplot2)

windplot <- ggplot(wind,aes(x=RSpd,y=CSpd))

windplot + geom\_point() + geom\_smooth() + theme\_grey() +

ggtitle("Cite Speed by Reference Speed") + xlab("Reference Speed") +

ylab("Cite Speed")

cor(wind$CSpd,wind$RSpd)

## 4 ##

windmod <- with(wind,{lm(CSpd~RSpd)})

names(windmod)

summary(windmod)

windmod$coefficients

with(wind,{plot(RSpd,CSpd,pch=20,col='blue',xlab="Reference",ylab="Cite",main="Cite by Reference")})

abline(reg=windmod,col='red',lwd=3)

## 5 ##

cspd <- function(x){3.14+(x\*.756)}

cspd(12)

## 6 ##

cspd(30)