**67-357: Lab 5 – Linear Regression**

**Solving the Bovine TB problem**

This code shows an alternate way of building and analyzing the problem

**install.packages("UsingR")**

**library(UsingR)**

**library(ggplot2)**

**# -------Load the dataset and plot ----------------------------------------------**

**TBCasesinUK <- read.csv("TBCasesinUK.csv", stringsAsFactors=FALSE)**

**View(TBCasesinUK)**

**RegressedTB<- lm(NoOfBTBCases ~ Year, data=TBCasesinUK) #Build regression model**

**summary(RegressedTB) # Get Summary Statistics**

**par(mfrow=c(1,1)) #plot in one row and one column**

**plot(TBCasesinUK$Year, TBCasesinUK$NoOfBTBCases, pch = 8, xlab = "Year", ylab = "Case notifications for TB from Cattle")**

**#pch is used for specifying the character for plotting**

**abline(RegressedTB) #Draw the regression line**

**preds <- data.frame( Year=62.0)** # what is no of cases for Year 2012

**predict(RegressedTB, newdata = preds)** # predict no of cases from the regression model

**Your answer should include the following (these instructions will not be given to you, so remember these expectations for your final exam):**

1. Scatter plot and comments about the scatter plot

Positive slope (1.1154) and Negative intercept (-2214.8077) Good Fit of regression between all the points.

1. The regression equation (y=ax+b)

NoofTBCases = 1.1154 \*Year -2214.8077

1. The significance of the regression (discuss p-value, F-statistics, Adj-RSquared)

P value = 0.00205 (P-value less than cut off point of 0.05, making year a good predictor of no of TB cases)  
F-Statistic P-value = 0.002047 (P-value less than cut off point of 0.05, making the model a good fit and statistically significant)   
R-squared = 0.557 (55.7% of the change in NoTBCases can be predicted by change in Year)

1. Predicted value for Years 2012 and 2013  
   2012 = 29 patients  
   2013 = 30 patients