# Ch 8 - Soap Production Lines Example

*Adam McQuistan*

*Saturday, April 30, 2016*

## The dataset

| Variable | Description |
|---|---|
| Y | Amount of scrap produced |
| X1 | Quantitative variable. Line Speed |
| X2 | Qualitative variable. 1 for line 1, 0 for line 2 |

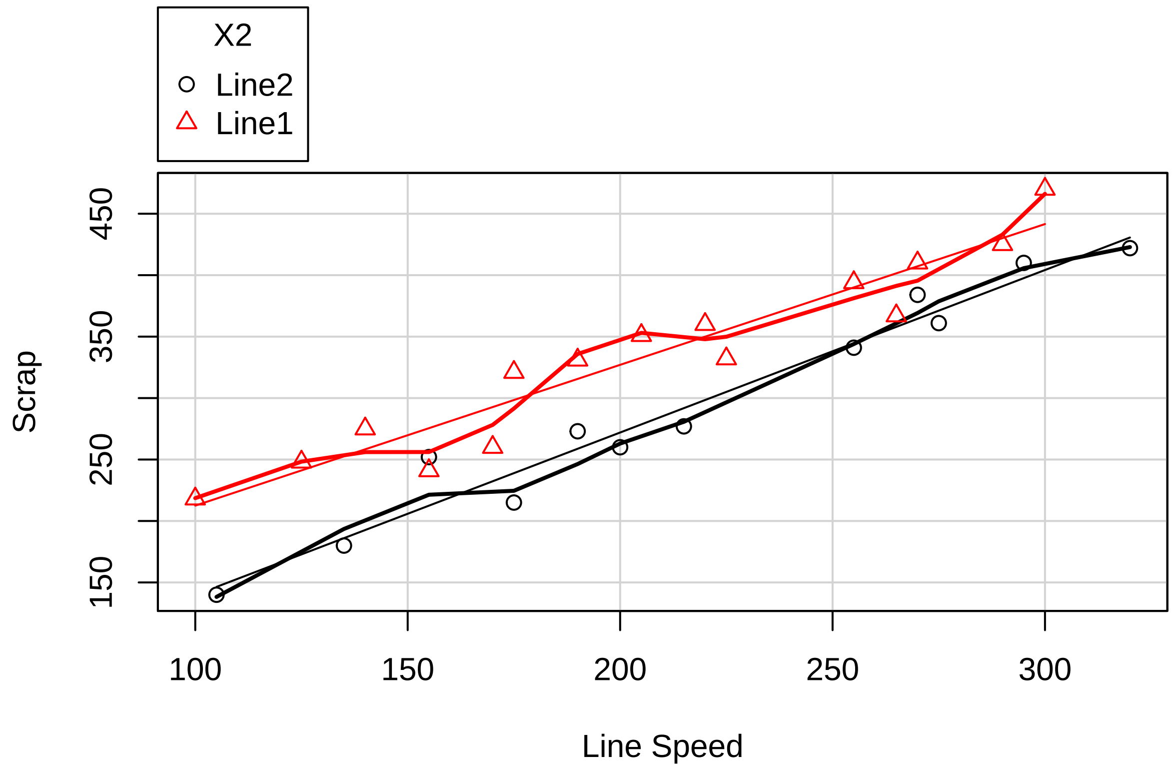
```
lines <- readLines("CH08TA05.txt")
Y <- vector()
X1 <- vector()
X2 <- vector()

library(gdata)
for(line in lines){
  line <- trim(line)
  lineAry <- unlist(strsplit(line, "  "))
  Y <- c(Y, lineAry[1])
  X1 <- c(X1, lineAry[2])
  X2 <- c(X2, lineAry[3])
}

df <- data.frame(Y=as.numeric(Y),
                 X1=as.numeric(X1),
                 X2=factor(X2, levels=c(0,1), labels=c("Line2", "Line1")))

str(df)
```

```
## 'data.frame':    27 obs. of  3 variables:
##  $ Y : num  218 248 360 351 470 394 332 321 410 260 ...
##  $ X1: num  100 125 220 205 300 255 225 175 270 170 ...
##  $ X2: Factor w/ 2 levels "Line2","Line1": 2 2 2 2 2 2 2 2 2 2 ...
```

```
summary(df)
```

```
##        Y               X1            X2
##  Min.   :140.0   Min.   :100.0   Line2:12
##  1st Qu.:256.0   1st Qu.:162.5   Line1:15
##  Median :331.0   Median :205.0
##  Mean   :315.5   Mean   :210.2
##  3rd Qu.:375.5   3rd Qu.:267.5
##  Max.   :470.0   Max.   :320.0
```

```r
library(car)
scatterplot(Y ~ X1 | X2, data=df,
            ylab="Scrap",
            xlab="Line Speed")
```




**Tenative Model**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

```r
result1 <- lm(Y ~ X1 + X2 + I(X1 * as.numeric(X2)), data=df)
result1_smry <- summary(result1)
df$residuals <- result1_smry$residuals
print(result1_smry)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + I(X1 * as.numeric(X2)), data = df)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -34.50 -11.06   2.78  14.82  39.51
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)               7.5745    20.8697   0.363  0.71996
## X1                        1.4987     0.2058   7.284 2.06e-07 ***
## X2Line1                  90.3909    28.3457   3.189  0.00409 **
## I(X1 * as.numeric(X2))  -0.1767     0.1288  -1.371  0.18355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.75 on 23 degrees of freedom
## Multiple R-squared:  0.9447, Adjusted R-squared:  0.9375
## F-statistic: 130.9 on 3 and 23 DF,  p-value: 1.341e-14
```

```
result1_aov <- fullRegressionAnova(anova(result1))
```

```
##          VariationSource DF          SS          MS    F_stats
## 1             Regression  3 169164.6838  56388.2279 130.949292
## 2                     X1  1 149660.9825 149660.9825 347.554808
## 3                     X2  1  18694.0787  18694.0787  43.412898
## 4 I(X1 * as.numeric(X2))  1    809.6226    809.6226   1.880171
## 5              Residuals 23   9904.0569    430.6112         NA
## 6                  Total 26 179068.7407          NA         NA
```

## Brown-Forsythe Test for Equal Error Variance in the Lines

$H_o$: equal variances

$H_a$: the two variances are not equal

```
leveneTest(df$residuals, df$X2, center="median")
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##       Df F value Pr(>F)
## group  1  0.4047 0.5304
##       25
```

Since pvalue $> 0.05$ conclude equal variances which leads us to conclude the data is approriate for the regression model.

## Inferences about Two Regression Lines

$H_o : \beta_2 = \beta_3 = 0$

$H_a$ : Not both $\beta_2 = 0$ and $\beta_3 = 0$

```
term1 <- (result1_aov$SS[3] + result1_aov$SS[4]) / (result1_aov$DF[3] + result1_aov$DF[4])
term2 <- result1_aov$SS[5] / result1_aov$DF[5]
F_stat <- term1 / term2

F_stat
```

```
## [1] 22.64653
```

```r
F_crit <- qf(0.99, 2,23)

msg <- paste("F stat: ", F_stat, "\nF crit: ", F_crit, sep="")
result <- ifelse(F_stat > F_crit, "Conclude Ha, not identical", "Conclude Ho, they are identical")
cat(msg, "\n", result, sep="")
```

```
## F stat: 22.646534296706
## F crit: 5.66369876809604
## Conclude Ha, not identical
```