Ch 6 Example - Dwaine Studio Sales

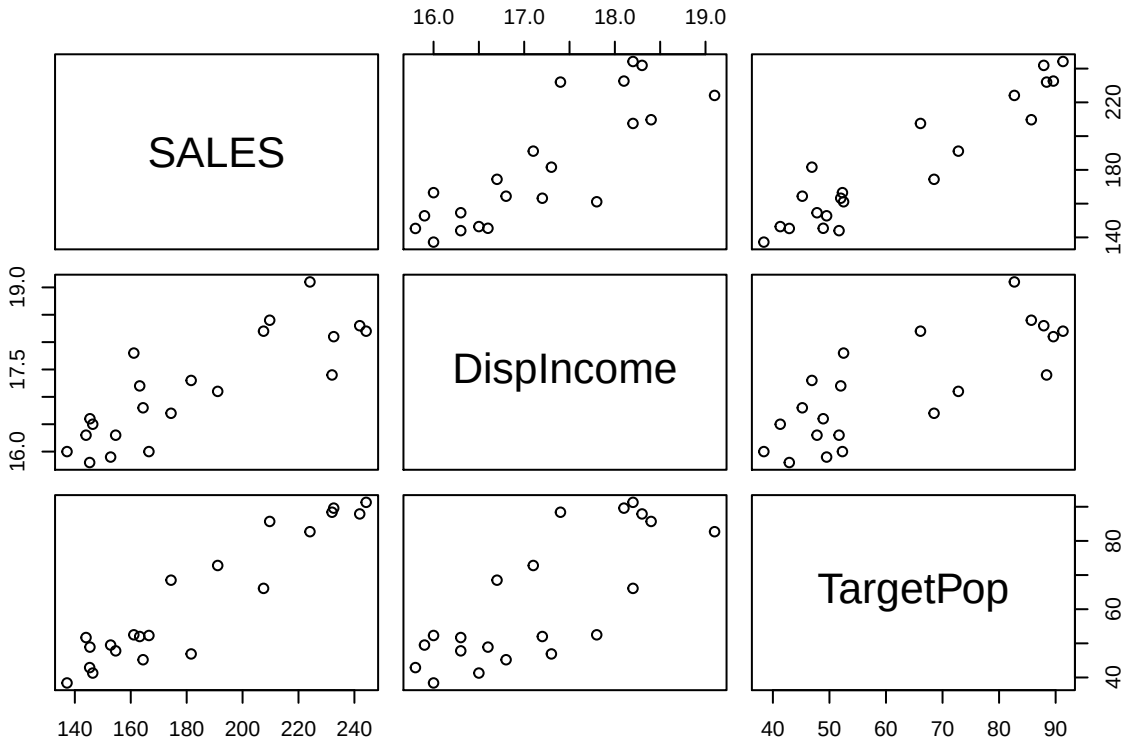
Adam McQuistan

Sunday, April 03, 2016

```
library("ggplot2")
setwd("C:\\Users\\AdamMcQuistan\\Documents\\ISQA 8340\\ch 6")
df <- read.csv("data/ch 6 sales.csv")
df$id = NULL
df <- with(df, data.frame(SALES, DispIncome, TargetPop))
```

View General Linearity Plots Among Predictors and Outcome Vars

```
pairs(df, pch=21)
```



Sales appears to have a linear relationship with the target population (individuals less than or equal to 16 yrs old) and the disploable income of the cities.

Building the Model

```r
result <- lm(SALES ~ DispIncome + TargetPop, data=df)
result_smry <- summary(result)
F_stat <- round(as.numeric(result_smry$fstatistic["value"]),1)
F_crit <- round(qf(0.95, df1=2, df2=result_smry$df[2]),1)
result_smry
```

```
##
## Call:
## lm(formula = SALES ~ DispIncome + TargetPop, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.4239  -6.2161   0.7449   9.4356  20.2151
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -68.8571    60.0170  -1.147   0.2663
## DispIncome    9.3655     4.0640   2.305   0.0333 *
## TargetPop     1.4546     0.2118   6.868    2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.01 on 18 degrees of freedom
## Multiple R-squared:  0.9167, Adjusted R-squared:  0.9075
## F-statistic:  99.1 on 2 and 18 DF,  p-value: 1.921e-10
```

Analyze and test for the relatedness of sales to the target population and per capita disposable income. F critical is 3.6.

H_0 : $\beta_1 = 0$ and $\beta_2 = 0$ for F* <= F critical

H_a : not both β_1 and β_2 equal zero for F* > F critical

Since F* (99.1) > F critical (3.6) we conclude that sales are related to target population and per capital income but, the usefullness of the model for making predictions have yet to be determined.

The P-value is < 0.05 further indicating the rejection of H_o.

Part C. Calculation coefficient of multople determination.

The Multiple R^2 value is 0.9167 which translates into the model explaining about 92% of the variation in sales. The Adjusted R^2 value of 0.9075 shows the correction for more than on predictor variable overfitting which is a very small change from the standard multiple r-squared value indicating little effects of overfitting.

To analyze and test the individual parameters of the regression model we use the students t-test such that:

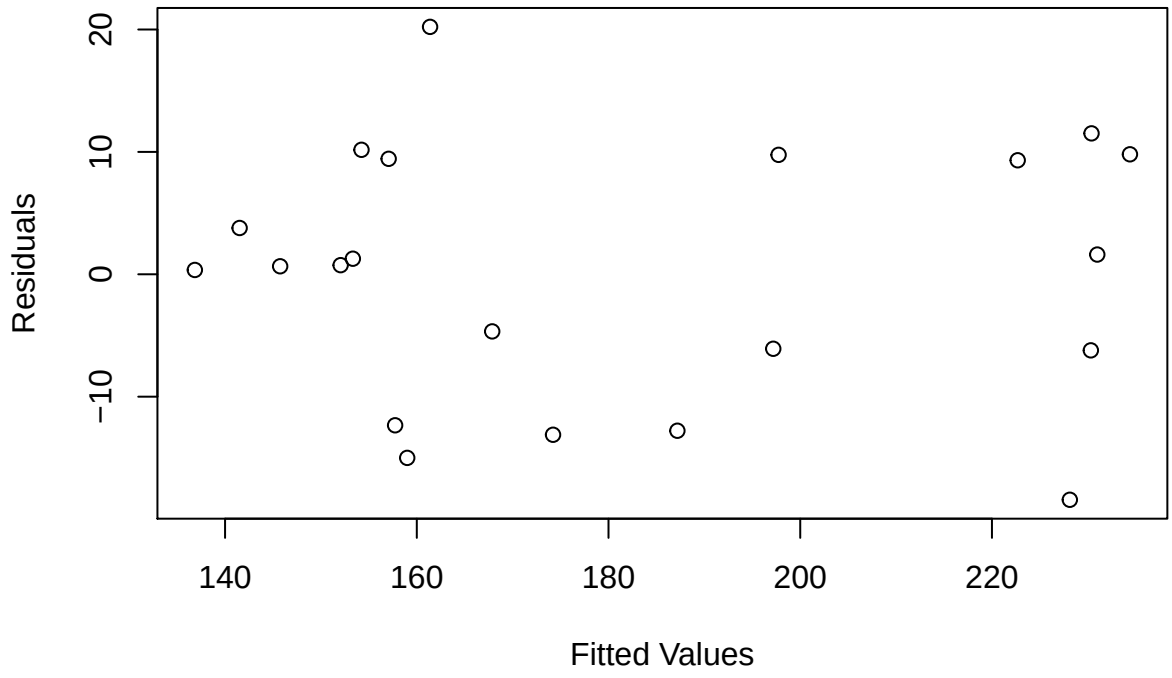
H_0 : $\beta_p = 0$ indicating no relationship

H_a : $\beta_p \neq 0$ indicating there is a relationship

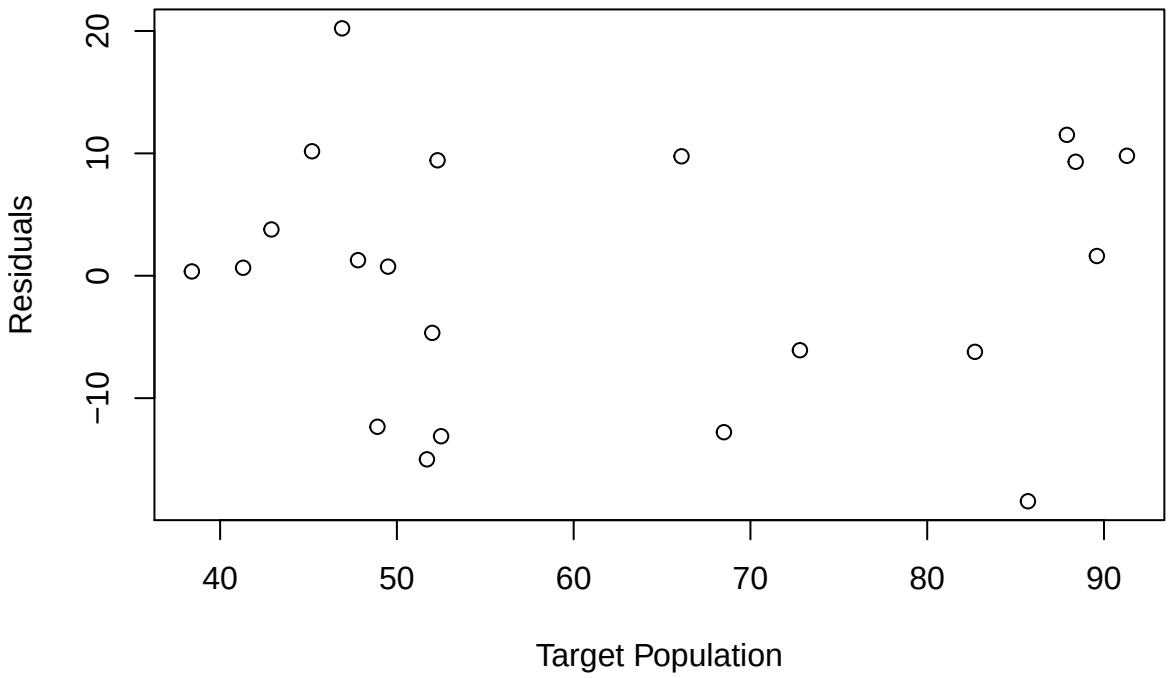
Since the p-values for both parameters (Disposable income and Target population) are less than 0.05 we conclude each individual parameter is significant and related to sales.

Graphical Assessment of the Model (Residual Analysis)

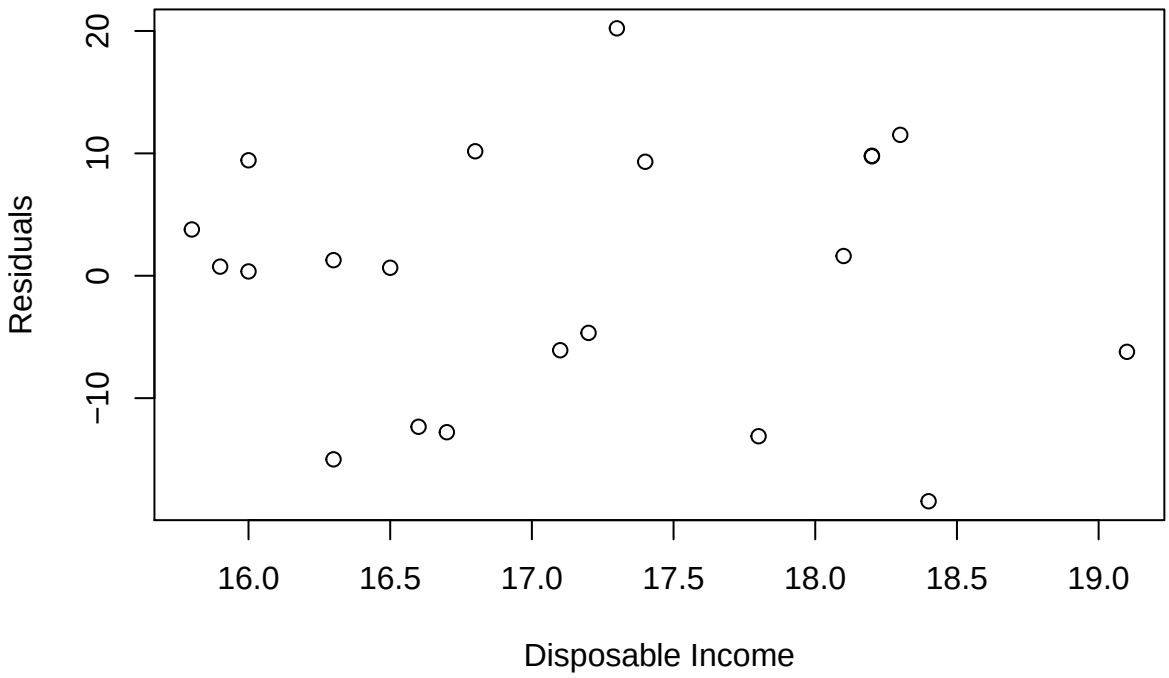
```
df_model <- result$model[,1:3]
df_model$FittedVals <- result$fitted.values
df_model$Residuals <- result$residuals

plot(x=df_model$FittedVals, y=df_model$Residuals, xlab="Fitted Values", ylab="Residuals")
```




```
plot(x=df_model$TargetPop, y=df_model$Residuals, xlab="Target Population", ylab="Residuals")
```




```
plot(x=df_model$DispIncome, y=df_model$Residuals, xlab="Disposable Income", ylab="Residuals")
```


Residuals
20
10
0
−10
16.0
16.5
17.0
17.5
18.0
18.5
19.0
Disposable Income