MORRIS

William Morris

4/16/2019

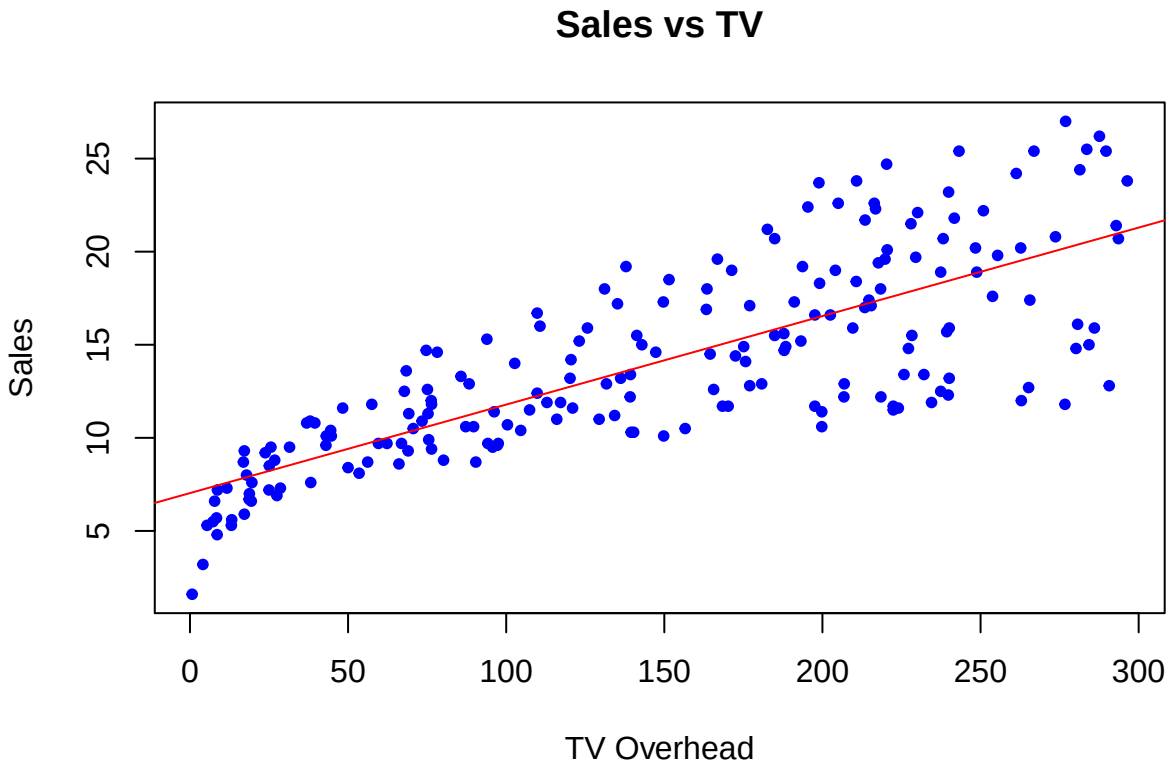
1.

```
cat("\014")
```

```
adv<-read.csv("Advertising.csv")
plotModel<-lm(Sales~TV,data=adv)
plotLM<-summary(plotModel)
print(plotLM)
```

```
##
## Call:
## lm(formula = Sales ~ TV, data = adv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3860 -1.9545 -0.1913  2.0671  7.2124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.032594   0.457843   15.36   <2e-16 ***
## TV          0.047537   0.002691   17.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.259 on 198 degrees of freedom
## Multiple R-squared:  0.6119, Adjusted R-squared:  0.6099
## F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

```
plot(adv$TV,adv$Sales,pch=20,col="blue",xlab="TV Overhead",ylab="Sales",main="Sales vs TV")
abline(plotModel,col="red")
```



2.

```
cat("\014")
```

```
data1<-read.csv("MyData1.csv")
model<-lm(Var3~Var1+Var2,data=data1)
LM<-summary(model)
print(LM)
```

```
##
## Call:
## lm(formula = Var3 ~ Var1 + Var2, data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61003 -0.12660  0.01984  0.14131  0.49530
##
## Coefficients:
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -0.0132746  0.0408846   -0.325    0.746
## Var1         2.0002672  0.0005034 3973.674   <2e-16 ***
## Var2         2.9996557  0.0005314 5644.370   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.202 on 197 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
```

```
## F-statistic: 2.236e+07 on 2 and 197 DF,  p-value: < 2.2e-16
```

```
################################################
cat("The Residual Standard Error is: ",LM$sigma, "\n")
```

```
## The Residual Standard Error is:  0.2020083
```

```
cat("The coefficient of determination is: ",LM$r.squared,"\n")
```

```
## The coefficient of determination is:  0.9999956
```

```
cat("The observed standard errors are: ",LM$coefficients[,"Std. Error"],"\n")
```

```
## The observed standard errors are:  0.04088459 0.0005033797 0.0005314421
```

```
################################################
CI<-confint(model,level=.95)
print(model)
```

```
##
## Call:
## lm(formula = Var3 ~ Var1 + Var2, data = data1)
##
## Coefficients:
## (Intercept)         Var1         Var2
##    -0.01327      2.00027      2.99966
```

```
print(CI)
```

```
##                   2.5 %     97.5 %
## (Intercept) -0.09390223 0.06735304
## Var1         1.99927445 2.00125986
## Var2         2.99860763 3.00070372
```

```
################################################
newInput<-data.frame(Var1=12.55,Var2=20.32)
alpha=.05
ciMean<-predict(model,newdata=newInput,interval="confidence",level=1-alpha)
ciPred<-predict(model,newdata=newInput,interval="prediction",level=1-alpha)
print(ciMean)
```

```
##        fit      lwr      upr
## 1 86.04308 85.98483 86.10133
```

```
print(ciPred)
```

```
##        fit      lwr      upr
## 1 86.04308 85.64047 86.44569
```

3.

```
cat("\014")
```

```
houseData<-read.csv("kc_house_data.csv")
houseData$id<-NULL
houseData$date<-NULL
houseData$sqft_basement<-NULL
houseData$floors<-NULL
```

```
model<-lm(data=houseData)
LM<-summary(model)
print(LM)
```

```
##
## Call:
## lm(data = houseData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1295113   -99168    -9551    77465  4317351
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.114e+06  2.885e+06   2.119  0.03411 *
## bedrooms      -3.921e+04  1.968e+03 -19.921  < 2e-16 ***
## bathrooms      4.382e+04  3.145e+03  13.935  < 2e-16 ***
## sqft_living    1.497e+02  4.186e+00  35.758  < 2e-16 ***
## sqft_lot       1.242e-01  4.788e-02   2.595  0.00946 **
## waterfront     5.817e+05  1.735e+04  33.528  < 2e-16 ***
## view           5.271e+04  2.138e+03  24.650  < 2e-16 ***
## condition      2.606e+04  2.346e+03  11.109  < 2e-16 ***
## grade          9.573e+04  2.144e+03  44.642  < 2e-16 ***
## sqft_above     3.461e+01  3.907e+00   8.857  < 2e-16 ***
## yr_built      -2.602e+03  7.088e+01 -36.704  < 2e-16 ***
## yr_renovated   1.994e+01  3.648e+00   5.467 4.64e-08 ***
## zipcode       -5.796e+02  3.282e+01 -17.660  < 2e-16 ***
## lat            6.032e+05  1.069e+04  56.423  < 2e-16 ***
## long          -2.168e+05  1.308e+04 -16.583  < 2e-16 ***
## sqft_living15  2.100e+01  3.424e+00   6.134 8.71e-10 ***
## sqft_lot15    -3.930e-01  7.317e-02  -5.371 7.90e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 201100 on 21595 degrees of freedom
## Multiple R-squared:  0.7002, Adjusted R-squared:    0.7
## F-statistic:  3152 on 16 and 21595 DF,  p-value: < 2.2e-16
```

So, the "id" and "date" data were clearly insignificant as they are index data only. "sqft_basement" kept coming up in the summary as a non-numeric argument. I was not able to locate the datum that caused this, so I removed it. "floors" was the only qualitative data that I removed because of its insignificance.

I was happy to leave "sqft_lot" in the model, as it's still 99% significant, even if that's substantially less that many of the other data.