HW1 Question 2

Saksham Goel

February 3, 2018

Question 2

Processing Time of Invoices

Fitting a linear model

```r
invoiceTime = read.table("invoices.txt", header = T)
mod = lm(invoiceTime$Time ~ invoiceTime$Invoices )
mod
```

```
##
## Call:
## lm(formula = invoiceTime$Time ~ invoiceTime$Invoices)
##
## Coefficients:
##          (Intercept)  invoiceTime$Invoices
##              0.64171               0.01129
```

```r
summary(mod)
```

```
##
## Call:
## lm(formula = invoiceTime$Time ~ invoiceTime$Invoices)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59516 -0.27851  0.03485  0.19346  0.53083
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.6417099  0.1222707   5.248 1.41e-05 ***
## invoiceTime$Invoices 0.0112916  0.0008184  13.797 5.17e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3298 on 28 degrees of freedom
## Multiple R-squared:  0.8718, Adjusted R-squared:  0.8672
## F-statistic: 190.4 on 1 and 28 DF,  p-value: 5.175e-14
```

Fitting a model and observing the values of the parameters $\hat{\beta}_0, \hat{\beta}_1$ we say $\hat{\beta}_0 = 0.6417099$ $\hat{\beta}_1 = 0.0112916$

The final fitted model looks as follows:

Y = 0.6417099 + 0.0112916 * X

where X = Number of Invoices Y = Total Processing Time

Finding a 95% Confidence Inrerval β_0

The confidence interval for β_0 is given as follows:

'"insert formula here'"'define all the terms"'

Using the following snipped of R code we find the Confidence Interval for β_0

```
bet0_hat = 0.6417099
x_col = invoiceTime$Invoices
y_col = invoiceTime$Time
x_bar = mean(x_col)
y_bar = mean(y_col)
sxx = sum((x_col-x_bar)^2)
sxy = sum((x_col-x_bar)*(y_col-y_bar))
n = length(x_col)
s = 0.3298
t_mult = 2.048
beta0CIlower = bet0_hat - ( t_mult * s * sqrt( (1/n) + ((x_bar^2)/(sxx)) ) )
beta0CIupper = bet0_hat + ( t_mult * s * sqrt( (1/n) + ((x_bar^2)/(sxx)) ) )
beta0CIlower
```

```
## [1] 0.3912792
```

```
beta0CIupper
```

```
## [1] 0.8921406
```

The values we found are as follows:

The 95% confidence interval can thus be given as follows : (0.3912792, 0.8921406).

Hypothesis Testing for β_1

The Hypothesis Testing for β_1 is can be done as follows

'"insert formula here'"'define all the terms"'

Here the Null Hypothesis is given by: '"insert null hypothesis"' And the Alternate Hypothesis is given by:"'insert alternate hypothesis"'

Using the following snippet of R code we do the hypothesis testing for β_1

```
bet1_hat = 0.0112916
x_col = invoiceTime$Invoices
y_col = invoiceTime$Time
x_bar = mean(x_col)
y_bar = mean(y_col)
sxx = sum((x_col-x_bar)^2)
sxy = sum((x_col-x_bar)*(y_col-y_bar))
n = length(x_col)
s = 0.3298
t_mult = 2.048
beta1test = ( ( bet1_hat - 0.01 )/( s/sqrt( sxx ) ) )
beta1test
```

```
## [1] 1.57807
```

```
2*pt(-abs(beta1test), df = n-2)
```

```
## [1] 0.1257819
```

The values we found are as follows: p_val = 0.1257819

The Hypothesis test for β_1 gives us a p-value = 0.1257819 which is greater than 0.05 which means that we cannot reject the null hypothesis.

Point Estimate for new Y

Find the point estimate of Y using the fitted model we get: As Y = 0.6417099 + 0.0112916 * X and we have to find Y|X = 130 we get

```
X = 130
Y = 0.6417099 + 0.0112916 * X
Y
```

```
## [1] 2.109618
```

So Y = 2.109618 hours = 127 mins.

Prediction Interval for new Y

Finding the Prediction interval for the new Y we get:

```
#fitted model =
# Y = 0.6417099 + 0.0112916 * X
x_curr = 130
y_hat = 0.6417099 + (0.0112916 * x_curr)
mse = mean(mod$residuals^2)

x_col = invoiceTime$Invoices
x_bar = mean(x_col)
sxx = sum((x_col-x_bar)^2)
n = length(x_col)
s = 0.3298
t_mult = 2.048

yPILower = y_hat - (t_mult * sqrt(mse) * sqrt(1 + (1/n) + (((x_curr - x_bar)^2)/sxx) ))
yPIUpper = y_hat + (t_mult * sqrt(mse) * sqrt(1 + (1/n) + (((x_curr - x_bar)^2)/sxx) ))
yPILower
```

```
## [1] 1.446357
```

```
yPIUpper
```

```
## [1] 2.772878
```

The values we found are as follows: yPILower = 1.446357 yPIUpper = 2.772878