CSC 423 Homework 2

Akhil Kumar Ramasagaram

April 14, 2016

Feeding behavior of blackbream fish.

```
load("rdata/BLACKBREAM.Rdata")
lm_model <- lm(STRIKES ~ AGE, BLACKBREAM)
lm_coeff <- lm_model$coefficients
lm_coeff
```

```
## (Intercept)         AGE
## 175.7033300  -0.8194806
```

a) Write the equation of a straight-line model relating number of strikes (y) to age of fish (x)

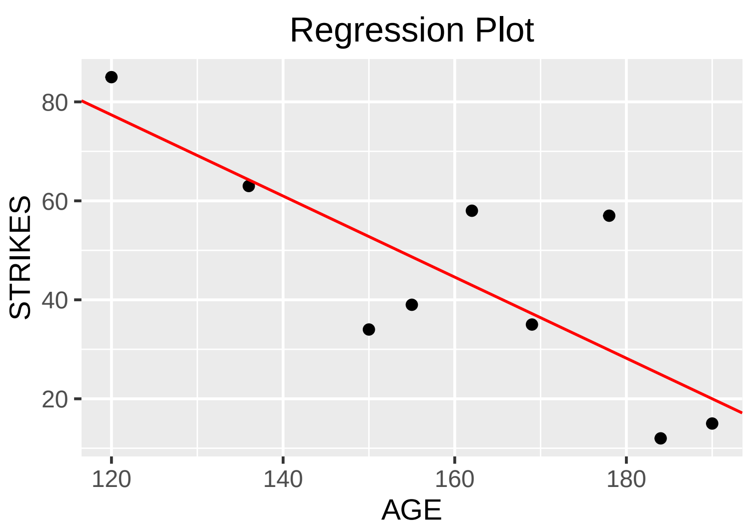
$$STRIKE = \beta_0 + \beta_1 * AGE$$

b) Fit the model to the data using the method of least squares and give the least squares prediction equation.

$$STRIKE = 175.70 - 0.81 * AGE$$

c&d) Give a practical interpretation of the value of $\hat{\beta}_0$ & $\hat{\beta}_1$.

```
library(ggplot2)
ggplot(BLACKBREAM, aes(y = STRIKES, x = AGE)) +geom_point() +
  geom_abline(intercept = lm_coeff[1], slope = lm_coeff[2], colour = "red") +
  ggtitle("Regression Plot")
```



β_0 is the intercept, which is the y value when x = 0, in this case when age is zero the strikes will be 175 & β_1 is the slope of the regression line.

Extending the life of an aluminum smelter pot.

```
load("rdata/SMELTPOT.Rdata")
lm_model <- lm(POROSITY ~ DIAMETER, SMELTPOT)
lm_coeff <- lm_model$coefficients
lm_coeff
```

```
## (Intercept)    DIAMETER
##   6.3518117   0.9498247
```

```
summary(lm_model)
```

```
##
## Call:
## lm(formula = POROSITY ~ DIAMETER, data = SMELTPOT)
##
## Residuals:
##       1       2       3       4       5       6
##  0.2503  2.7349  3.0145 -4.4859  0.3951 -1.9089
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.3518     3.9046   1.627   0.1791
## DIAMETER      0.9498     0.3563   2.666   0.0561 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.184 on 4 degrees of freedom
## Multiple R-squared:  0.6398, Adjusted R-squared:  0.5498
## F-statistic: 7.106 on 1 and 4 DF,  p-value: 0.05606
```

```
summary(lm_model)[6]
```

```
## $sigma
## [1] 3.184034
```

The equation for the above model will be $POROSITY = 6.35 + 0.94 * AGE$. a) An estimate of the above model standard deviation is $\sigma = 3.184034$ b) The error of prediction will be 2*σ which will be 2x3.18 = 6.36

Massage therapy for boxers.

```
load("rdata/BOXING2.Rdata")
lm_model <- lm(RECOVERY ~ LACTATE, BOXING2)
summary(lm_model)
```

```
##
## Call:
## lm(formula = RECOVERY ~ LACTATE, data = BOXING2)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.577 -3.752  0.060  3.067  8.043
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.7967     4.9838   0.561   0.5836
## LACTATE       2.5667     0.9883   2.597   0.0211 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.28 on 14 degrees of freedom
## Multiple R-squared:  0.3251, Adjusted R-squared:  0.2769
## F-statistic: 6.744 on 1 and 14 DF,  p-value: 0.0211
```

As $\alpha = 0.10$ and our p-value is 0.0211, blood lactate level is linearly related to perceived recovery.

Recalling student names.

```r
load("rdata/NAMEGAME2.Rdata")
lm_model <- lm(RECALL ~ POSITION, NAMEGAME2)
predict(lm_model, newdata = data.frame("POSITION" = 5), interval="confidence", level=.99)
```

```
##         fit       lwr       upr
## 1 0.7025529 0.6459537 0.7591521
```

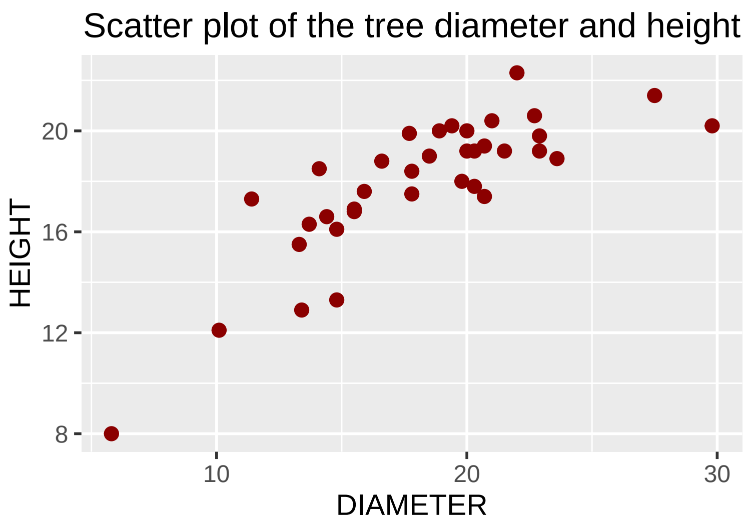
```r
predict(lm_model, newdata = data.frame("POSITION" = 5), interval="prediction", level=.99)
```

```
##         fit        lwr      upr
## 1 0.7025529 0.03656847 1.368537
```

Confidence interval refers to a sample of observations where as a prediction interval refers to a single observations. In this case prediction interval is much wider than confidence interval and prediction interval will always be wider than confidence interval.

Predicting heights of spruce trees.

```r
load("rdata/WHITESPRUCE.Rdata")
ggplot(WHITESPRUCE, aes(x=DIAMETER ,y=HEIGHT)) +
  geom_point(size = 2, col = "darkred") +
  ggtitle("Scatter plot of the tree diameter and height")
```

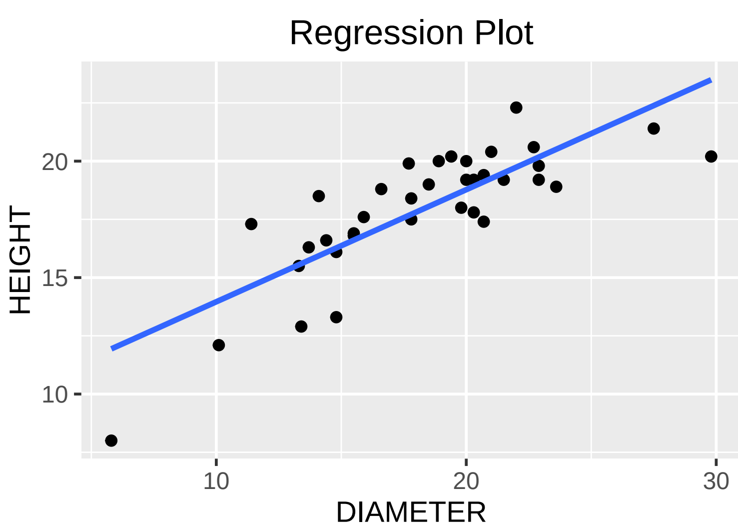



```
lm_model <- lm(HEIGHT ~ DIAMETER, WHITESPRUCE)
lm_coeff <- lm_model$coefficients
lm_coeff
```

```
## (Intercept)    DIAMETER
##   9.1468390   0.4814743
```

The equation for the above model is $HEIGHT = 9.14 + 0.48 * DIAMETER$ With $\alpha = 0.05$, our p-values from the model is 2.089e-09 which is almost 0 hence we can conclude that there is sufficient evidence to indicate that the breast height diameter x contributes information for the prediction of tree height y

```
ggplot(WHITESPRUCE, aes(x = DIAMETER, y = HEIGHT)) +
  geom_point() + geom_smooth(method = "lm", se = F) +
  ggtitle("Regression Plot")
```




```
predict(lm_model, newdata = data.frame("DIAMETER" = 20), interval="confidence", level=.90)
```

```
##        fit      lwr      upr
## 1 18.77632 18.26972 19.28293
```

with 90% confidence that for a sample of observations with mean diameter = 20 the height will be in the following interval [18.26972, 19.28293]