HW 4 #2

Jacob S Townson

March 5, 2018

a) Given the definitions defined in problem #2 of Homework #4, compute the MLE of $\beta_0, \beta_1, ..., \beta_9$ and σ^2.

```
setwd("D:/obewa/Documents/My real documents/University of Louisville/public_work/Classes/Homework/Stati
yield=read.table("YieldData.txt",sep="\t",header=TRUE)
```

Now that we have the data as a dataframe in R, we can get down to business. First Let's make life easier for making the linear model by adding in new columns to the dataframe containing the values we want:

```
require(dplyr)
```

```
yield2 = mutate(yield, x4 = x1*x1)
yield2 = mutate(yield2, x5 = x2*x2)
yield2 = mutate(yield2, x6 = x3*x3)
yield2 = mutate(yield2, x7 = x1*x2)
yield2 = mutate(yield2, x8 = x1*x3)
yield2 = mutate(yield2, x9 = x2*x3)
```

Using the mutate function in dplyr, I have made new columns containing the values we want to use in our model, which will make using the lm function very easy.

```
require(knitr)
```

```
yield2.model = lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9, data = yield2)
coefs = summary(yield2.model)$coefficients
kable(coefs)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2659.8665969	1092.5943546	-2.4344502	0.0377063
x1	13.4927677	6.0815387	2.2186437	0.0536827
x2	11.4364072	6.1252679	1.8670869	0.0947375
x3	18.6128169	11.9956712	1.5516278	0.1551662
x4	-0.0160884	0.0086430	-1.8614405	0.0955938
x5	0.0009565	0.0178709	0.0535242	0.9584835
x6	0.0343860	0.0788924	0.4358598	0.6732056
x7	-0.0339549	0.0140041	-2.4246511	0.0383172
x8	-0.0557019	0.0386738	-1.4403020	0.1836474
x9	-0.0138611	0.0531443	-0.2608192	0.8001018

Here in this summary, we can see the maximum likelihood estimates for each β_i coefficient where β_i corresponds to x_i.

As for the estimate of σ^2, we use the following bit of code:

```
sig = summary(yield2.model)$sigma
sig^2
```

```
## [1] 10.13782
```

Thus the estimate of σ^2 is 10.13782.

b) Perform and F-test of $H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$ at level .05.

So in this situation we want to compare the model we created in part (a) above, and a new model without the extra elements added to it. This new model will be made below:

```
yield.model = lm(y~x1+x2+x3, data = yield)
```

Notice this model only relies on the first three x_i values, so we will only use β_1, β_2, and β_3. To do this comparison and F-test, we will use the anova function.

```
anova(yield.model,yield2.model)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x1 + x2 + x3
## Model 2: y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     15 279.71
## 2      9  91.24  6    188.47 3.0984 0.0624 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here in this table, we see that our $F_{\text{obs}} = 3.0984$ and $P(F > F_{\text{obs}}) = 0.0624 > 0.05$. Thus we cannot yet reject H_0, which leads us to believe further testing is required, especially with a probablility as small as 0.0624.