# PA1 - Homework 2

*Kristiyan Dimitrov*

*10/7/2019*

## Exercise 3.12

### Part (a)

```
data=read.csv("/Users/kristiyan/Documents/MSiA 401 - Predictive 1/Homeworks/Cobb-Douglas.csv",stringsAsI
str(data)
```

```
## 'data.frame':    569 obs. of  4 variables:
##  $ No.    : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ capital: num  2.61 1.32 22.09 10.74 1.16 ...
##  $ labor  : int  184 91 426 72 46 7642 48 108 35 72 ...
##  $ output : num  9.25 3.66 28.78 4.12 2.89 ...
```

```
#Calculating log of the three variables (capital, labor, and output)

log_data <- data %>%
  mutate(log_output=log(data$output),log_capital=log(data$capital),log_labor=log(data$labor)) %>%
  .[5:7] #Keep only the last 3 columns with log data

head(log_data)
```

```
##   log_output log_capital log_labor
## 1   2.224706   0.9580326  5.214936
## 2   1.298640   0.2800809  4.510860
## 3   3.359733   3.0952921  6.054439
## 4   1.416979   2.3737750  4.276666
## 5   1.061308   0.1495957  3.828641
## 6   6.174079   5.6232833  8.941415
```

```
# Now we fit a LS model through the data
log_lm <- lm(log_output ~ log_capital + log_labor, data=log_data)
summary(log_lm)
```

```
##
## Call:
## lm(formula = log_output ~ log_capital + log_labor, data = log_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7604 -0.2665 -0.0694  0.1926  3.7975
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.71146    0.09671  -17.70   <2e-16 ***
## log_capital  0.20757    0.01719   12.08   <2e-16 ***
## log_labor    0.71485    0.02314   30.89   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4781 on 566 degrees of freedom
## Multiple R-squared:  0.8378, Adjusted R-squared:  0.8373
## F-statistic:  1462 on 2 and 566 DF,  p-value: < 2.2e-16
```

```
print(paste0("We see that the capital elasticity is: ", round(log_lm$coefficients[2],4)))
```

```
## [1] "We see that the capital elasticity is: 0.2076"
```

```
print(paste0("We see that the labor elasticity is: ", round(log_lm$coefficients[3],4)))
```

```
## [1] "We see that the labor elasticity is: 0.7148"
```

### Part (b) - T-Test

```
# Calculating the Variance-Covariance matrix of the elasticities
vcov_mat <- vcov(log_lm)
var_capital_elas <- vcov_mat[2,2]
var_labor_elas <- vcov_mat[3,3]
covar_capital_labor_elas <- vcov_mat[2,3]
print(paste("The variance of capital elasticity is",var_capital_elas))
```

```
## [1] "The variance of capital elasticity is 0.000295412270078063"
```

```
print(paste("The variance of capital elasticity is",var_labor_elas))
```

```
## [1] "The variance of capital elasticity is 0.000535537185242537"
```

```
print(paste("The covariance of capital & labor elasticities is",covar_capital_labor_elas))
```

```
## [1] "The covariance of capital & labor elasticities is -0.000267477018380779"
```

### Part (c) - Extra SS & F-Test

```
anova(log_lm)
```

```
## Analysis of Variance Table
##
## Response: log_output
##              Df Sum Sq Mean Sq F value    Pr(>F)
## log_capital   1 450.23  450.23 1969.96 < 2.2e-16 ***
## log_labor     1 218.08  218.08  954.19 < 2.2e-16 ***
## Residuals   566 129.36    0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# This shows us that the SSE for the full model is 129.36
```

```
# Partial model
# First we generate the data:

new_data <- log_data %>%
  mutate(response=log_data$log_output-log_data$log_labor) %>% #New response = log(output) - log(labor)
  mutate(input_one=log_data$log_capital-log_data$log_labor) %>%
  .[3:5]

head(new_data)
```

```
##   log_labor  response input_one
```

```
## 1  5.214936 -2.990230 -4.256903
## 2  4.510860 -3.212219 -4.230779
## 3  6.054439 -2.694706 -2.959147
## 4  4.276666 -2.859687 -1.902891
## 5  3.828641 -2.767333 -3.679046
## 6  8.941415 -2.767336 -3.318131
```

```r
full_model <- lm(data=new_data, response ~ input_one + log_labor)
summary(full_model)
```

```
##
## Call:
## lm(formula = response ~ input_one + log_labor, data = new_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7604 -0.2665 -0.0694  0.1926  3.7975
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.71146    0.09671 -17.697  < 2e-16 ***
## input_one    0.20757    0.01719  12.077  < 2e-16 ***
## log_labor   -0.07758    0.01720  -4.509  7.9e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4781 on 566 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2366
## F-statistic: 89.03 on 2 and 566 DF,  p-value: < 2.2e-16
```

```r
anova(full_model) # We see that the SSE is still 129.36
```

```
## Analysis of Variance Table
##
## Response: response
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## input_one   1  36.048  36.048 157.727 < 2.2e-16 ***
## log_labor   1   4.648   4.648  20.335 7.905e-06 ***
## Residuals 566 129.358   0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
drop1(full_model,test="F")
```

```
## Single term deletions
##
## Model:
## response ~ input_one + log_labor
##           Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                 129.36 -836.86
## input_one  1    33.333 162.69 -708.40 145.849 < 2.2e-16 ***
## log_labor  1     4.648 134.01 -818.77  20.335 7.905e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We see that for log_labor the RSS is 134.01
```

### <<<<< EXPERIMENTATION >>>>>

```
# Let's try doubling all the data (this is ok, based on the assumption of H_0 null hypothesis of consta
# Then we'll do a t.test() between the regular data and the log data
double_log_data <- log_data %>%
  mutate(double_log_output=2*log_data$log_output, double_log_capital=2*log_data$log_capital,double_log_l
  .[4:6]

head(double_log_data)
```

```
##   double_log_output double_log_capital double_log_labor
## 1          4.449411          1.9160653        10.429872
## 2          2.597280          0.5601617         9.021719
## 3          6.719467          6.1905843        12.108879
## 4          2.833958          4.7475500         8.553332
## 5          2.122616          0.2991915         7.657283
## 6         12.348157         11.2465666        17.882829
```

```
head(log_data)
```

```
##   log_output log_capital log_labor
## 1   2.224706   0.9580326  5.214936
## 2   1.298640   0.2800809  4.510860
## 3   3.359733   3.0952921  6.054439
## 4   1.416979   2.3737750  4.276666
## 5   1.061308   0.1495957  3.828641
## 6   6.174079   5.6232833  8.941415
```

```
t.test(log_data$log_output,double_log_data$double_log_output,paired=T)
```

```
##
##  Paired t-test
##
## data:  log_data$log_output and double_log_data$double_log_output
## t = -33.714, df = 568, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.772486 -1.577329
## sample estimates:
## mean of the differences
##               -1.674907
```

```
# The t statistic is -33.714, therefore we should reject  H_0
```