Week2.HW.data

Bryan Aguiar

Sunday, April 05, 2015

## Week 2 Homework

The data file contains the systolic blood pressure (SBP), body size (QUET), age (AGE), and smoking history (SMK = 0 if nonsmoker, SMK = 1 if a current or previous smoker) for a hypothetical sample of 32 white males over 40 years old from the town of Angina.

##Clear the environment  
rm(list=ls())  
  
##Turn off scientific notations for numbers  
options(scipen = 999)   
  
##Set locale  
Sys.setlocale("LC\_ALL", "English")

## [1] "LC\_COLLATE=English\_United States.1252;LC\_CTYPE=English\_United States.1252;LC\_MONETARY=English\_United States.1252;LC\_NUMERIC=C;LC\_TIME=English\_United States.1252"

##Set seed for reproducibility  
set.seed(2345)  
  
DF<- read.csv("D:/Data/week2-HW-data.csv")  
summary(DF)

## Person SBP QUET AGE   
## Min. : 1.00 Min. :120.0 Min. :2.368 Min. :41.00   
## 1st Qu.: 8.75 1st Qu.:134.8 1st Qu.:3.022 1st Qu.:48.00   
## Median :16.50 Median :143.0 Median :3.381 Median :53.50   
## Mean :16.50 Mean :144.5 Mean :3.441 Mean :53.25   
## 3rd Qu.:24.25 3rd Qu.:152.0 3rd Qu.:3.776 3rd Qu.:58.25   
## Max. :32.00 Max. :180.0 Max. :4.637 Max. :65.00   
## SMK   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :1.0000   
## Mean :0.5312   
## 3rd Qu.:1.0000   
## Max. :1.0000

## Exercise One and Two

Determine the ANOVA table for the following regressions. You should see an ANOVA table, as well as a t-statistic for the slope coefficient below, and an F-statistic of model fit.

Use the ANOVA tables to perform the F test for the significance of each straight-line  
regression. For each of the following, we are testing whether the slope coefficient is equal to zero. In other words, we are testing if the independent variable contributes significantly to  
the model. Notice that, because there is only one independent variable in the model,  
the F-test for the overall model yields the same significance as the t-test for the slope  
coefficient.

1. SPB (Y) vs. QUET (X)

##   
## Call:  
## lm(formula = DF$SBP ~ DF$QUET)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.231 -7.145 -1.604 7.798 22.531   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 70.576 12.322 5.728 0.00000299 \*\*\*  
## DF$QUET 21.492 3.545 6.062 0.00000117 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.812 on 30 degrees of freedom  
## Multiple R-squared: 0.5506, Adjusted R-squared: 0.5356   
## F-statistic: 36.75 on 1 and 30 DF, p-value: 0.000001172

1. SBP (Y) vs. SMK (X)

##   
## Call:  
## lm(formula = DF$SBP ~ DF$SMK)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.824 -9.056 -2.812 11.200 32.176   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 140.800 3.661 38.454 <0.0000000000000002 \*\*\*  
## DF$SMK 7.024 5.023 1.398 0.172   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.18 on 30 degrees of freedom  
## Multiple R-squared: 0.06117, Adjusted R-squared: 0.02988   
## F-statistic: 1.955 on 1 and 30 DF, p-value: 0.1723

1. QUET (Y) vs. AGE (X)

##   
## Call:  
## lm(formula = DF$QUET ~ DF$AGE)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.54248 -0.16601 -0.02293 0.10368 0.71998   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.38645 0.41769 0.925 0.362   
## DF$AGE 0.05736 0.00778 7.373 0.0000000325 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3013 on 30 degrees of freedom  
## Multiple R-squared: 0.6444, Adjusted R-squared: 0.6326   
## F-statistic: 54.37 on 1 and 30 DF, p-value: 0.00000003253

1. SBP (Y) vs. AGE (X)

##   
## Call:  
## lm(formula = DF$SBP ~ DF$AGE)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.548 -6.990 -2.481 5.765 23.892   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 59.0916 12.8163 4.611 0.000069782 \*\*\*  
## DF$AGE 1.6045 0.2387 6.721 0.000000189 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.245 on 30 degrees of freedom  
## Multiple R-squared: 0.6009, Adjusted R-squared: 0.5876   
## F-statistic: 45.18 on 1 and 30 DF, p-value: 0.0000001894

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details, see [R Markdown](http://rmarkdown.rstudio.com).