In Class Assignment 5

Rahul Avadhoot

October 25, 2017

fev2 <- read.csv('http://faculty.washington.edu/tathornt/Biost509/DataSets/fev2.csv')

## Question 1

t.test(fev~sex, data=fev2)

##   
## Welch Two Sample t-test  
##   
## data: fev by sex  
## t = 5.5037, df = 575.75, p-value = 5.604e-08  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.2323494 0.4902038  
## sample estimates:  
## mean in group 1 mean in group 2   
## 2.812446 2.451170

p-value for the t-test = 5.604e-08

There is a significant difference in the means of males and females. As the p-value is <0.05, we reject the null hypothesis (difference between means is 0). This is sufficient evidence to support the alternate hypothesis (true difference in means is not equal to 0).

## Question 2

lm1<-lm(fev~sex,data=fev2)  
summary(lm1)

##   
## Call:  
## lm(formula = fev ~ sex, data = fev2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.01645 -0.69420 -0.06367 0.58233 2.98055   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.1737 0.1041 30.484 < 2e-16 \*\*\*  
## sex -0.3613 0.0664 -5.441 7.5e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8487 on 652 degrees of freedom  
## Multiple R-squared: 0.04344, Adjusted R-squared: 0.04197   
## F-statistic: 29.61 on 1 and 652 DF, p-value: 7.496e-08

confint(lm1, parm="sex", level=0.95)

## 2.5 % 97.5 %  
## sex -0.491653 -0.2309002

Interpretatioon of coefficient of sex = The coefficient of sex is -0.3613. This means that as the value of sex increases from 1 to 2 (male to female), the value of fev decreases.

95% CI for coefficient of sex = [-0.491653,-0.2309002]

There is significant evidence that fev differs by sex (p-value of coefficient of sex = 7.5e-08)

R-squared value = 0.04197

mean(predict(lm1,fev2[fev2$sex==1,]))

## [1] 2.812446

mean(predict(lm1,fev2[fev2$sex==2,]))

## [1] 2.45117

The mean fev values for males are females calculated from the regression model is identical to that obtained from the t-test

## Question 3

lm2<-lm(fev~height,data=fev2)  
summary(lm2)

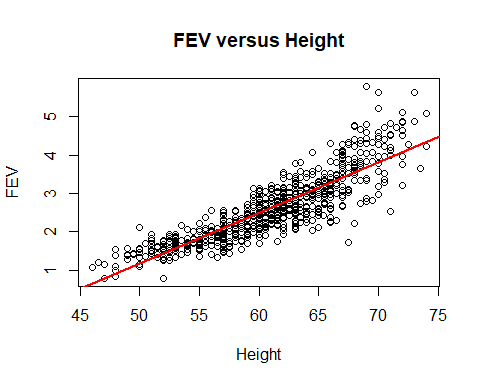
##   
## Call:  
## lm(formula = fev ~ height, data = fev2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.75167 -0.26619 -0.00401 0.24474 2.11936   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.432679 0.181460 -29.94 <2e-16 \*\*\*  
## height 0.131976 0.002955 44.66 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4307 on 652 degrees of freedom  
## Multiple R-squared: 0.7537, Adjusted R-squared: 0.7533   
## F-statistic: 1995 on 1 and 652 DF, p-value: < 2.2e-16

fev and height are positively correlated with the coefficient of height 0.131976. For every unit increase in height, the fev increases by 0.131976.

The p-value for the coefficient of height is 2e-16. Hence we reject the null hypothesis (coefficient is zero). This provides sufficient evidence of an association between fev and height.

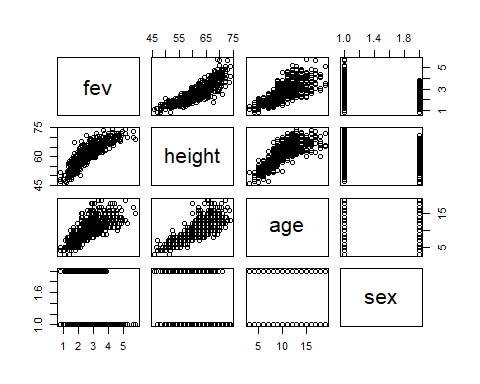
## Question 4

plot(fev~height,data=fev2,main="FEV versus Height",xlab="Height",ylab="FEV")  
abline(lm2,col='red',lw=2)



## Question 5

pairs(fev2[,c("fev","height","age","sex")])



lm3<-lm(fev ~ height + age + sex, data=fev2)  
summary(lm3)

##   
## Call:  
## lm(formula = fev ~ height + age + sex, data = fev2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.37613 -0.24834 0.01051 0.25748 1.94538   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.126335 0.241896 -17.058 < 2e-16 \*\*\*  
## height 0.104560 0.004756 21.986 < 2e-16 \*\*\*  
## age 0.061364 0.009069 6.766 2.96e-11 \*\*\*  
## sex -0.161112 0.033125 -4.864 1.45e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4126 on 650 degrees of freedom  
## Multiple R-squared: 0.7746, Adjusted R-squared: 0.7736   
## F-statistic: 744.6 on 3 and 650 DF, p-value: < 2.2e-16

All the 3 coefficients are significant in the model as the p-values are <0.05 The p-values of the coefficients are: height = <2e-16 age = 2.96e-11 sex = 1.45e-06