Chapter 4 Exercise 5

Allen Church

10/9/2019

5a. Estimate model with income at age 33 as dependent variable and height at age 33 as indepdendent variable. Exclude observations with wages above 400 pounds/hr and height less than 40 inches. Interpret t stats on coefficients.

#Load data  
load("Ch3\_Exercise3\_Height\_and\_Wages\_UK.RData")  
data <- dta  
  
#Subset data exclusing wages above 400 pounds/hr and height less than 40 inches  
data <- subset(data, gwage33 < 400 & height33 > 40)  
  
#Run linear model and summarize results  
olsresults <- lm(data$gwage33 ~ data$height33)  
summary(olsresults)

##   
## Call:  
## lm(formula = data$gwage33 ~ data$height33)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.632 -3.835 -2.014 0.356 157.690   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.34591 5.01799 -1.862 0.062616 .   
## data$height33 0.26810 0.07199 3.724 0.000199 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.04 on 3667 degrees of freedom  
## Multiple R-squared: 0.003768, Adjusted R-squared: 0.003496   
## F-statistic: 13.87 on 1 and 3667 DF, p-value: 0.0001989

t stat = beta 1 hat / standard error of beta 1 hat

#Calculate t stats to verify using beta hat / standard error of beta hat, since they ask for t stats on the coefficients  
t\_stat\_beta0 <- -9.34591 / 5.01799  
t\_stat\_beta1 <- 0.26810 / 0.07199  
  
#Print results  
print(paste0("T stat for beta 0 is ",t\_stat\_beta0 ))

## [1] "T stat for beta 0 is -1.86248079410282"

print(paste0("T stat for beta 1 is ",t\_stat\_beta1 ))

## [1] "T stat for beta 1 is 3.72412835115988"

In the results above, the t statistic on the coefficient of height is 3.724. Since this is above 2, we can suggest these results are statistically significant.

5b. Explain the p values for the two estimated coefficients.

In the summary of the model above, the p value of 0.0001989 indicates we can reject the null hypothesis decisively - regardless if alpha is set at 0.01 or 0.05 - as the p value is below both of these alphas.

5c. Show how to calculate the 95% confidence interval for coefficient on height

#Use cofint function and specify olsresults with a level of 0.95 as parameters  
confidence <- confint(olsresults, level = 0.95)  
confidence

## 2.5 % 97.5 %  
## (Intercept) -19.1842427 0.4924295  
## data$height33 0.1269547 0.4092422

5d. Do we accept or reject the null hypothesis that Beta1 = 0 for alpha = 0.01 and a two-sided alternative?

In the model above, the estimated coefficient on height has a value of 0.26810, which suggests that a 1 unit increase in the height variable is associated with a wage increase of 0.26810 pounds/hr. I believe we can reject the null hypothesis that Beta1 = 0, because the p-value of 0.0001989 is less than the alpha of 0.01, which suggests we can reject the null. In addition to this, the t statistic of 3.724 suggests our results are statistically significant.

5e. Do we accept or reject the null hypothesis that Beta0 = 0 (the constant) for alpha = 0.01 and a two sided alternative?

In the model above, the estimated coefficient for Beta0 has a value of -9.34591. I believe we can reject the null hypothesis that Beta0 = 0, because the p-value of 0.0001989 is less than the alpha of 0.01, which suggests we can reject the null.

5f. Limit sample size to first 800 observations. Do we accept or reject the null that Beta1 = 0 for alpha = 0.01 and a two-sided alternative?

data800 <- data[1:800,]  
olsresults2 <- lm(data800$gwage33 ~ data800$height33)  
summary(olsresults2)

##   
## Call:  
## lm(formula = data800$gwage33 ~ data800$height33)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.340 -3.530 -1.770 0.493 148.070   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -10.0802 9.0750 -1.111 0.2670   
## data800$height33 0.2723 0.1307 2.083 0.0376 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.17 on 798 degrees of freedom  
## Multiple R-squared: 0.005408, Adjusted R-squared: 0.004161   
## F-statistic: 4.339 on 1 and 798 DF, p-value: 0.03757

In contrast to the full dataset with 3669 observations, limiting the number of observations to 800 increases the p value from 0.0001989 to 0.03757. Since the alpha = 0.01, we cannot confidently reject the null hypothesis with a p value greater than the alpha. Clearly, reducing the number of observations increased the p value. In addition to this, the standard error on Beta1 hat increased from 0.07199 to 0.1307. As we reduced the number of observations, the standard error of Beta1 hat increased - both of which indicate a low-power statistical test.