Assignment 2

Chloe Moore

2024-10-22

library(tidyverse)  
library(ggplot2)

### Question 11. Exercise 2.20

nursing\_homeDF = read.csv("WiscNursingHome.csv")  
nursing\_home2001 = subset(nursing\_homeDF, CRYEAR == 2001) %>%  
 mutate(  
 LOGTPY = log(TPY),  
 LOGNUMBED = log(NUMBED)  
 )

#### Part a)

summary(nursing\_home2001$LOGTPY) # summary statistics LOGTPY

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.511 4.041 4.396 4.368 4.700 6.088

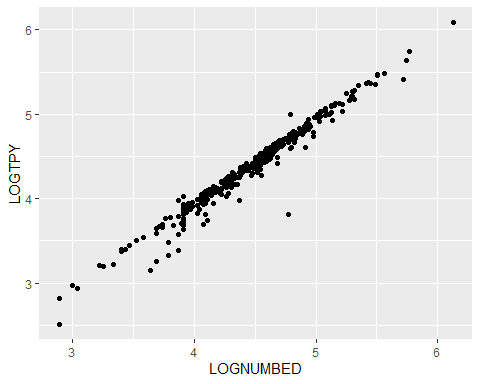
summary(nursing\_home2001$LOGNUMBED) # summary statistics LOGNUMBED

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.890 4.094 4.500 4.457 4.779 6.125

cor(nursing\_home2001$LOGTPY, nursing\_home2001$LOGNUMBED) # correlation statistic

## [1] 0.9830461

ggplot(nursing\_home2001, aes(x=LOGNUMBED, y=LOGTPY)) +  
 geom\_point()



Comments: The correlation statistic is very close to 1, which indicates LOGTPY and LOGNUMBED are closely, positively correlated. The scatter plot chart appears to show a linear trend.

#### Part b)

fit1 = lm(LOGTPY ~ LOGNUMBED, nursing\_home2001)  
summary1 = summary(fit1)  
summary1

##   
## Call:  
## lm(formula = LOGTPY ~ LOGNUMBED, data = nursing\_home2001)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.87482 -0.02201 0.01517 0.05316 0.28862   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.17469 0.04537 -3.85 0.00014 \*\*\*  
## LOGNUMBED 1.01923 0.01012 100.73 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09373 on 353 degrees of freedom  
## Multiple R-squared: 0.9664, Adjusted R-squared: 0.9663   
## F-statistic: 1.015e+04 on 1 and 353 DF, p-value: < 2.2e-16

Comments: R^2 = 0.9664 (adjusted R^2=0.9663). Regression coefficient LOGNUMBED = 1.01923. t-statistic = 100.73

#### Part c)

b1 = summary1$coefficients[2, 1]  
seb1 = summary1$coefficients[2, 2]  
tcrit = 2 # for t\_(353, 0.975)

(I) vs . We reject the null hypothesis because our t-statistic is greater than the critical t-value. The p-value is less than our alpha level of 0.05, which supports this rejection.

tstat = b1/seb1 # 100.730193  
pval = 2\*(1-pt(abs(tstat), 353)) # 0.0000000...

(II) vs . We fail to reject the null hypothesis because our t-statistic is less than the critical t-value (however, it’s very close!). The p-value is greater than our alpha level of 0.05, which supports this conclusion.

tstat = (b1-1)/seb1 # 1.900567  
pval = 2\*(1-pt(abs(tstat), 353)) # 0.058173

(III) vs . We might fail to reject the null hypothesis because our t-statistic is less than the critical t-value… But the p-value is less than our alpha level of 0.05, so in this case, we will actually reject the null hypothesis.

tstat = (b1-1)/seb1 # 1.900567  
pval = (1-pt(abs(tstat), 353)) # 0.029087

(IV) vs . We might fail to reject the null hypothesis because our t-statistic is less than the critical t-value… And the p-value is greater than our alpha level of 0.05, so we will indeed fail to reject the null hypothesis.

tstat = (b1-1)/seb1 # 1.900567  
pval = pt(abs(tstat), 353) # 0.970913