Question 8

Chloe Moore

2024-09-11

#### Part a(i)

data = read.csv("WiscNursingHome.csv")  
  
cor(data$TPY, log(data$TPY))

## [1] 0.9288011

The correlation value is quite close to 1, which means that TPY and it’s log value are strongly positively correlated.

#### Part a(ii)

cor(data[c("TPY", "NUMBED", "SQRFOOT")], use = "complete.obs")

## TPY NUMBED SQRFOOT  
## TPY 1.0000000 0.9836241 0.8219443  
## NUMBED 0.9836241 1.0000000 0.8136223  
## SQRFOOT 0.8219443 0.8136223 1.0000000

All the correlations we see here are close to 1. This means TPY, NUMBED, and SQRFOOT are all positively correlated with each other.

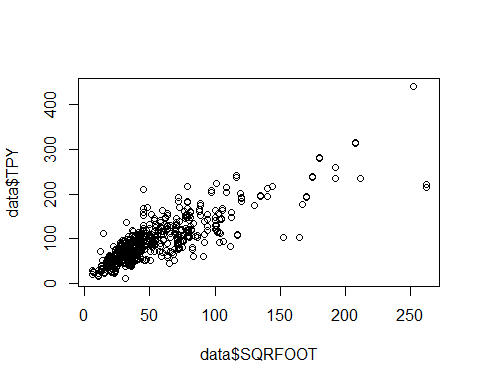
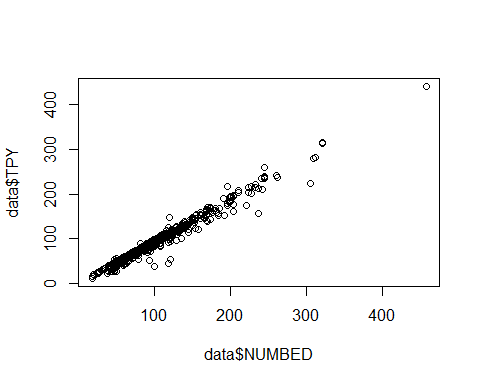
#### Part a(iii)

cor(data$TPY, data$NUMBED/10)

## [1] 0.9837719

This number is very close the the original correlation value for TPY and NUMBED (0.9836241). This could indicate that the positive correlation between these variables remains strong, regardless of the value of NUMBED.

#### Part b)

 The TPY and NUMBED plot has the data points clustered closely together along a linear trend line. This suggest a strong correlation between these variables, and this is supported by the correlation value we found earlier. The TPY and SQRFOOT plot also showed a linear trend, but the data was more scattered. This increased variance suggests the correlation between TPY and SQRFOOT is not as strong (but still definitely present!) as NUMBED. This is supported by the correlation values we found earlier too.

#### Part c(i)

fit1 = lm(TPY ~ NUMBED, data)  
summary(fit1)

##   
## Call:  
## lm(formula = TPY ~ NUMBED, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -63.249 -2.067 0.866 3.967 37.064   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.877827 0.692503 -1.268 0.205   
## NUMBED 0.927191 0.006324 146.611 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.538 on 715 degrees of freedom  
## Multiple R-squared: 0.9678, Adjusted R-squared: 0.9678   
## F-statistic: 2.149e+04 on 1 and 715 DF, p-value: < 2.2e-16

R^2 = 0.9678, t-value = 146.611. These results suggest a very good fit, since the R^2 is close to 1 and t-value is relatively high.

#### Part c(ii)

fit2 = lm(TPY ~ SQRFOOT, data)  
summary(fit2)

##   
## Call:  
## lm(formula = TPY ~ SQRFOOT, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -114.501 -15.391 -2.426 15.615 126.599   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.54754 1.78645 18.78 <2e-16 \*\*\*  
## SQRFOOT 1.11786 0.02917 38.32 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 27.14 on 705 degrees of freedom  
## (10 observations deleted due to missingness)  
## Multiple R-squared: 0.6756, Adjusted R-squared: 0.6751   
## F-statistic: 1468 on 1 and 705 DF, p-value: < 2.2e-16

R^2 = 0.6756, t-value = 38.32. The model using NUMBED fit better, as it had a R^2 score closer to 1.

#### Part c(iii)

fit3 = lm(log(TPY) ~ log(NUMBED), data)  
summary(fit3)

##   
## Call:  
## lm(formula = log(TPY) ~ log(NUMBED), data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.89233 -0.01923 0.01741 0.05768 0.29396   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.163315 0.036045 -4.531 6.88e-06 \*\*\*  
## log(NUMBED) 1.015739 0.008038 126.372 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1058 on 715 degrees of freedom  
## Multiple R-squared: 0.9571, Adjusted R-squared: 0.9571   
## F-statistic: 1.597e+04 on 1 and 715 DF, p-value: < 2.2e-16

#### Part c(iv)

fit4 = lm(log(TPY) ~ log(SQRFOOT), data)  
summary(fit4)

##   
## Call:  
## lm(formula = log(TPY) ~ log(SQRFOOT), data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.73615 -0.15269 0.02281 0.19078 1.06551   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.83032 0.06797 26.93 <2e-16 \*\*\*  
## log(SQRFOOT) 0.68163 0.01800 37.87 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2925 on 705 degrees of freedom  
## (10 observations deleted due to missingness)  
## Multiple R-squared: 0.6704, Adjusted R-squared: 0.67   
## F-statistic: 1434 on 1 and 705 DF, p-value: < 2.2e-16