Stat 415 Regression: Classwork/Lab 8

Frankie Tyndall

2022-07-22

## Head Circumference Dataset

tribble(~HEIGHT, ~WEIGHT, ~HEADCIRCUMFERENCE,   
 30, 339, 47,   
 26.25, 267, 42,   
 25, 289, 43,   
 27, 332, 44.5,   
 27.5, 272, 44,   
 24.5, 214, 40.5,   
 27.75, 311, 44,   
 25, 259, 41.5,  
 28, 298, 46,   
 27.25, 288, 44,  
 26, 277, 44,   
 27.25, 292, 44.5,  
 27, 302, 42.5,  
 28.25, 336, 44.5  
 ) -> Head\_circum  
Head\_circum

## # A tibble: 14 × 3  
## HEIGHT WEIGHT HEADCIRCUMFERENCE  
## <dbl> <dbl> <dbl>  
## 1 30 339 47   
## 2 26.2 267 42   
## 3 25 289 43   
## 4 27 332 44.5  
## 5 27.5 272 44   
## 6 24.5 214 40.5  
## 7 27.8 311 44   
## 8 25 259 41.5  
## 9 28 298 46   
## 10 27.2 288 44   
## 11 26 277 44   
## 12 27.2 292 44.5  
## 13 27 302 42.5  
## 14 28.2 336 44.5

## Use and show R code to construct a coefficient matrix. Is there multicollinearity involving any of the explanatory variables? If so, what variables are of concern? (Head Circumference is the response variable)

cor(Head\_circum)

## HEIGHT WEIGHT HEADCIRCUMFERENCE  
## HEIGHT 1.0000000 0.7847652 0.8708869  
## WEIGHT 0.7847652 1.0000000 0.7796990  
## HEADCIRCUMFERENCE 0.8708869 0.7796990 1.0000000

The predictor variables, height and weight both have high correlations with the dependent/response variable of Height circumference. There is multicollinearity between the predictor variables Height and Weight. The correlation coefficient between the two variables is 0.7847652, this indicates that both of the variables may not be needed in the model. The predicting power of the model is not improved if there is high correlation between the height and weight variables.

## Use and show R code to produce the full multi regression model where the response variable is Head Circumference and the two explanatory variables are Height and Weight.

ml\_headcir <- lm(HEADCIRCUMFERENCE~HEIGHT + WEIGHT, Head\_circum)  
ml\_headcir

##   
## Call:  
## lm(formula = HEADCIRCUMFERENCE ~ HEIGHT + WEIGHT, data = Head\_circum)  
##   
## Coefficients:  
## (Intercept) HEIGHT WEIGHT   
## 18.82425 0.78634 0.01281

The full regression model:

Head Circumference = 18.82425 + 0.78634*HEIGHT + 0.01281*WEIGHT

## Use and show R code to produce the summary table for your full model. Indicate the variable whose p value is not significant at the level of .05. And also indicate what proportion of the variability in Head Circumference is explained by your model.

summary(ml\_headcir)

##   
## Call:  
## lm(formula = HEADCIRCUMFERENCE ~ HEIGHT + WEIGHT, data = Head\_circum)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.42356 -0.55408 0.06335 0.44183 1.34134   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.82425 4.94099 3.810 0.00289 \*\*  
## HEIGHT 0.78634 0.26455 2.972 0.01269 \*   
## WEIGHT 0.01281 0.01160 1.105 0.29290   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8707 on 11 degrees of freedom  
## Multiple R-squared: 0.7826, Adjusted R-squared: 0.743   
## F-statistic: 19.79 on 2 and 11 DF, p-value: 0.0002266

The weight variable has a p value that is larger than 0.05, specifically it has a p value of 0.29290, this makes the explanatory variable weight not significant. Using the adjusted r-squared value of 0.743, this means that The model explains 74.3% of the variation in the dependent variable Head Circumference using only those predictor variables that have an impact on the response variable.

## Use and show R code that will produce a linear regression model with the explanatory variable of no significance removed.

removed\_variable <- lm(HEADCIRCUMFERENCE~HEIGHT, Head\_circum)  
removed\_variable

##   
## Call:  
## lm(formula = HEADCIRCUMFERENCE ~ HEIGHT, data = Head\_circum)  
##   
## Coefficients:  
## (Intercept) HEIGHT   
## 16.382 1.016

summary(removed\_variable)

##   
## Call:  
## lm(formula = HEADCIRCUMFERENCE ~ HEIGHT, data = Head\_circum)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3050 -0.5726 -0.1663 0.6315 1.2264   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 16.3818 4.4590 3.674 0.00318 \*\*   
## HEIGHT 1.0157 0.1655 6.138 5.04e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8787 on 12 degrees of freedom  
## Multiple R-squared: 0.7584, Adjusted R-squared: 0.7383   
## F-statistic: 37.68 on 1 and 12 DF, p-value: 5.035e-05

Linear Regression Model:

y-hat = 16.382 + 1.016\*Height

## Interpret the slope of this model

removed\_variable

##   
## Call:  
## lm(formula = HEADCIRCUMFERENCE ~ HEIGHT, data = Head\_circum)  
##   
## Coefficients:  
## (Intercept) HEIGHT   
## 16.382 1.016

The slope indicates that for every increase by 1 inch in height, a child’s head circumference (in centimeters) was found to have increased by 1.016 on average.

## What proportion of the variability of the response variable is not explained by the linear regression model?

1 - 0.7584 = 0.2416 or 24.16%

24.16% of the variability of the response variable Head Circumference is not explained by the linear regression model.

## Find the correlation coefficient r for the linear model

cor(Head\_circum$HEADCIRCUMFERENCE, Head\_circum$HEIGHT)

## [1] 0.8708869

The correlation coefficiebt r for the linear model is 0.8708869

## Does r suggest a linear relationship that is weak, moderate, strong, or very strong?

The linear relationship is strong with a correlation coefficient of 0.8708869

## What is the value of the standard error for the explanatory variable? Explain what it means in two or sentences.

summary(removed\_variable)

##   
## Call:  
## lm(formula = HEADCIRCUMFERENCE ~ HEIGHT, data = Head\_circum)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3050 -0.5726 -0.1663 0.6315 1.2264   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 16.3818 4.4590 3.674 0.00318 \*\*   
## HEIGHT 1.0157 0.1655 6.138 5.04e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8787 on 12 degrees of freedom  
## Multiple R-squared: 0.7584, Adjusted R-squared: 0.7383   
## F-statistic: 37.68 on 1 and 12 DF, p-value: 5.035e-05

The standard error of the explanatory variable is 0.1655, this means that 0.1655 is the typical distance a sample height coefficient and the true population height coefficient.