adv.stats.mod10.R

hadeelmusallam

2022-11-03

#9.1  
library(base)  
install.packages("ISwR", repos = "http://cran.us.r-project.org")

##   
## The downloaded binary packages are in  
## /var/folders/17/1blphrc55\_l3\_j01yc9yrp5m0000gn/T//RtmpgnxOOW/downloaded\_packages

library(ISwR)  
cystfibr

## age sex height weight bmp fev1 rv frc tlc pemax  
## 1 7 0 109 13.1 68 32 258 183 137 95  
## 2 7 1 112 12.9 65 19 449 245 134 85  
## 3 8 0 124 14.1 64 22 441 268 147 100  
## 4 8 1 125 16.2 67 41 234 146 124 85  
## 5 8 0 127 21.5 93 52 202 131 104 95  
## 6 9 0 130 17.5 68 44 308 155 118 80  
## 7 11 1 139 30.7 89 28 305 179 119 65  
## 8 12 1 150 28.4 69 18 369 198 103 110  
## 9 12 0 146 25.1 67 24 312 194 128 70  
## 10 13 1 155 31.5 68 23 413 225 136 95  
## 11 13 0 156 39.9 89 39 206 142 95 110  
## 12 14 1 153 42.1 90 26 253 191 121 90  
## 13 14 0 160 45.6 93 45 174 139 108 100  
## 14 15 1 158 51.2 93 45 158 124 90 80  
## 15 16 1 160 35.9 66 31 302 133 101 134  
## 16 17 1 153 34.8 70 29 204 118 120 134  
## 17 17 0 174 44.7 70 49 187 104 103 165  
## 18 17 1 176 60.1 92 29 188 129 130 120  
## 19 17 0 171 42.6 69 38 172 130 103 130  
## 20 19 1 156 37.2 72 21 216 119 81 85  
## 21 19 0 174 54.6 86 37 184 118 101 85  
## 22 20 0 178 64.0 86 34 225 148 135 160  
## 23 23 0 180 73.8 97 57 171 108 98 165  
## 24 23 0 175 51.1 71 33 224 131 113 95  
## 25 23 0 179 71.5 95 52 225 127 101 195

attach(cystfibr)

## The following object is masked from package:ISwR:  
##   
## tlc

linearmodel<- lm(pemax~age+weight+bmp+fev1, data = cystfibr)  
summary(linearmodel)

##   
## Call:  
## lm(formula = pemax ~ age + weight + bmp + fev1, data = cystfibr)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -42.521 -10.885 3.003 15.488 41.767   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 179.2957 61.8855 2.897 0.00891 \*\*  
## age -3.4181 3.3086 -1.033 0.31389   
## weight 2.6882 1.1727 2.292 0.03287 \*   
## bmp -2.0657 0.8198 -2.520 0.02036 \*   
## fev1 1.0882 0.5139 2.117 0.04695 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.4 on 20 degrees of freedom  
## Multiple R-squared: 0.5918, Adjusted R-squared: 0.5101   
## F-statistic: 7.248 on 4 and 20 DF, p-value: 0.0008891

anova(linearmodel)

## Analysis of Variance Table  
##   
## Response: pemax  
## Df Sum Sq Mean Sq F value Pr(>F)   
## age 1 10098.5 10098.5 18.4385 0.0003538 \*\*\*  
## weight 1 945.2 945.2 1.7258 0.2038195   
## bmp 1 2379.7 2379.7 4.3450 0.0501483 .   
## fev1 1 2455.6 2455.6 4.4836 0.0469468 \*   
## Residuals 20 10953.7 547.7   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Interpretation of linear regression model   
#Variable= age  
#The estimate value of the intercept is 179.296 with significant p-value.  
#The p-value indicates that the result is not occurred because of chance,   
#so we can reject the null hypothesis. The variable of age is -3.4181,   
#which indicates that for a unit change in var age tends to change by -3.4181  
#units in dependent pemax with non-sig p-value because the p-value is greater  
#than 0.05.   
#Interpretation of ANOVA table  
#Variable = age  
#the p-value for the age variable is 0.0003538 which is less than   
#0.001. The null hypothesis can be rejected. We can interpret that the   
#effect of age variable coefficient on dependent variable pemax   
#does not equal 0.   
  
#9.2  
library(ISwR)  
linemodel<- lm(log(bwt)~ I(log(bpd)), data = secher)  
summary(linemodel)

##   
## Call:  
## lm(formula = log(bwt) ~ I(log(bpd)), data = secher)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.36478 -0.09725 0.01251 0.07703 0.51154   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.0862 0.9062 -7.819 4.35e-12 \*\*\*  
## I(log(bpd)) 3.3320 0.2017 16.516 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1488 on 105 degrees of freedom  
## Multiple R-squared: 0.7221, Adjusted R-squared: 0.7194   
## F-statistic: 272.8 on 1 and 105 DF, p-value: < 2.2e-16

linemodel2<- lm(log(bwt)~ I(log(ad)), data = secher)  
summary(linemodel2)

##   
## Call:  
## lm(formula = log(bwt) ~ I(log(ad)), data = secher)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.58560 -0.06609 0.00184 0.07479 0.48435   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.4446 0.5103 -4.791 5.49e-06 \*\*\*  
## I(log(ad)) 2.2365 0.1105 20.238 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1275 on 105 degrees of freedom  
## Multiple R-squared: 0.7959, Adjusted R-squared: 0.794   
## F-statistic: 409.6 on 1 and 105 DF, p-value: < 2.2e-16

#we can see that the log(bpd) is a significant predictor variable   
#for log(ad). The model interprets that 79.5% of the variation of log(bwt)  
  
linemodel3<- lm(log(bwt)~ I(log(ad))+ I(log(bpd)), data = secher)  
summary(linemodel3)

##   
## Call:  
## lm(formula = log(bwt) ~ I(log(ad)) + I(log(bpd)), data = secher)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.35074 -0.06741 -0.00792 0.05750 0.36360   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.8615 0.6617 -8.859 2.36e-14 \*\*\*  
## I(log(ad)) 1.4667 0.1467 9.998 < 2e-16 \*\*\*  
## I(log(bpd)) 1.5519 0.2294 6.764 8.09e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1068 on 104 degrees of freedom  
## Multiple R-squared: 0.8583, Adjusted R-squared: 0.8556   
## F-statistic: 314.9 on 2 and 104 DF, p-value: < 2.2e-16

#We see that both variables are significant and the Rsquare increases  
#from 79.5% to 85.8%. The sum of the two coefficients equal to 3.0186.   
#Which fits the interpretation of our analysis.