car.R

cris

2023-10-03

library(data.table)  
dados <- fread(input = paste0("car\_base.csv"), header = T, na.strings = "NA", data.table = FALSE, dec=",")  
names(dados)

## [1] "drivewheel" "carwidth" "price"

# Classificação das variáveis  
dados$drivewheel <- as.factor(dados$drivewheel)  
dados$carwidth <- as.numeric(dados$carwidth)  
dados$price <- as.numeric(dados$price)  
  
# A)  
# Limpar dados  
dados <- dados[dados$drivewheel != "4wd", ]   
  
# B)  
# Modelo  
modelo <- lm(price ~ drivewheel\*carwidth, data=dados)  
summary(modelo)

##   
## Call:  
## lm(formula = price ~ drivewheel \* carwidth, data = dados)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.7120 -2.0507 -0.4600 0.9686 23.0225   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -102.0301 17.7772 -5.739 3.66e-08 \*\*\*  
## drivewheelrwd -48.4992 24.0548 -2.016 0.0452 \*   
## carwidth 1.7103 0.2732 6.261 2.45e-09 \*\*\*  
## drivewheelrwd:carwidth 0.8210 0.3640 2.256 0.0252 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.698 on 192 degrees of freedom  
## Multiple R-squared: 0.6701, Adjusted R-squared: 0.6649   
## F-statistic: 130 on 3 and 192 DF, p-value: < 2.2e-16

library(car)

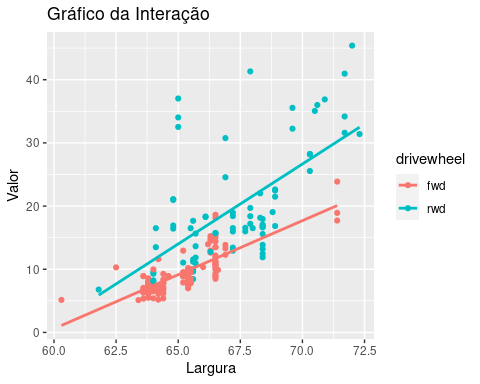
## Carregando pacotes exigidos: carData

Anova(modelo)

## Anova Table (Type II tests)  
##   
## Response: price  
## Sum Sq Df F value Pr(>F)   
## drivewheel 1126.0 1 51.0080 1.854e-11 \*\*\*  
## carwidth 3197.7 1 144.8559 < 2.2e-16 \*\*\*  
## drivewheel:carwidth 112.3 1 5.0875 0.02523 \*   
## Residuals 4238.3 192   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#C)  
#ŷ = -102,0301 +1,7103carwidth -48,4992drivewheelrwd + 0.8210drivewheelrwd\*carwidth   
# Ao nível de significancia de 5%, a relção é relevante 0.0252  
  
#D)  
# Gráficos  
library(ggplot2)  
ggplot(data = dados, aes(x = carwidth, y = price, color = drivewheel)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE) +  
 labs(title = "Gráfico da Interação",  
 x = "Largura",  
 y = "Valor")

## `geom\_smooth()` using formula = 'y ~ x'



#E)  
# Intervalo de confiança e intervalo de predição  
pred = data.frame(drivewheel='fwd',carwidth=70)  
confidence\_interval <- predict(modelo, pred, interval="confidence",level=0.95)  
pred\_interval <- predict(modelo, pred, interval="predict",level=0.95)  
  
# Intervalo de valores dentro do qual está contido o parâmetro populacional com determinada confiança  
# Estimativa de um intervalo de valores, a partir de dados observados, dentro do qual novos dados do mesmo contexto estarão contidos, com determinada probabilidade