Trata fator quantitativo

Panosso AR & Oliveira JA

2023-11-06

library(tidyverse)

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --  
## v dplyr 1.1.2 v readr 2.1.4  
## v forcats 1.0.0 v stringr 1.5.0  
## v ggplot2 3.4.2 v tibble 3.2.1  
## v lubridate 1.9.2 v tidyr 1.3.0  
## v purrr 1.0.1   
## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

Exemplo de dados

trat <- gl(4,2,labels = c("A","B","C","D"))  
dose <- rep(1:4,rep(2,4))  
y <- c(8.2, 7.8, 9.8, 10.4, 12.5, 11.5, 10.8, 11.2)  
dados <- tibble(trat, dose, y)  
dados

## # A tibble: 8 x 3  
## trat dose y  
## <fct> <int> <dbl>  
## 1 A 1 8.2  
## 2 A 1 7.8  
## 3 B 2 9.8  
## 4 B 2 10.4  
## 5 C 3 12.5  
## 6 C 3 11.5  
## 7 D 4 10.8  
## 8 D 4 11.2

# Modelo de médias

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_med = mean(y),  
 desvio = y - mod\_med,  
 desvio2 = desvio\*desvio)

## # A tibble: 8 x 7  
## # Groups: dose [4]  
## trat dose y y2 mod\_med desvio desvio2  
## <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 A 1 8.2 67.2 8 0.200 0.0400  
## 2 A 1 7.8 60.8 8 -0.200 0.0400  
## 3 B 2 9.8 96.0 10.1 -0.300 0.0900  
## 4 B 2 10.4 108. 10.1 0.300 0.0900  
## 5 C 3 12.5 156. 12 0.5 0.25   
## 6 C 3 11.5 132. 12 -0.5 0.25   
## 7 D 4 10.8 117. 11 -0.200 0.0400  
## 8 D 4 11.2 125. 11 0.200 0.0400

mod <- lm(y~trat)  
anova(mod)

## Analysis of Variance Table  
##   
## Response: y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## trat 3 17.415 5.805 27.643 0.003909 \*\*  
## Residuals 4 0.840 0.210   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_med = mean(y),  
 desvio = y - mod\_med,  
 desvio2 = desvio\*desvio) %>%   
 ungroup() %>%   
 summarise(  
 SQT = sum(y2) - sum(y)^2/length(y),  
 SQD = sum(desvio2),  
 SQE = SQT-SQD,  
 )

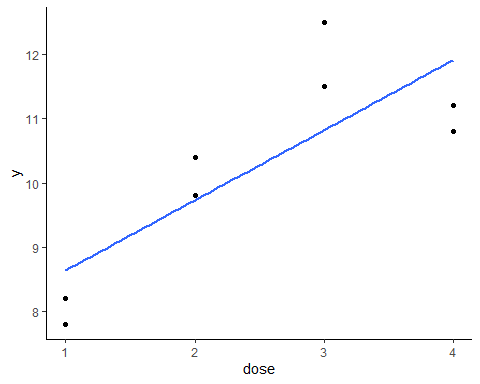
## # A tibble: 1 x 3  
## SQT SQD SQE  
## <dbl> <dbl> <dbl>  
## 1 18.3 0.840 17.4

# Modelo Reta

Vamos ajustar uma reta

dados %>%   
 ggplot(aes(x=dose, y=y)) +  
 geom\_point() +   
 theme\_classic() +   
 geom\_smooth(method = "lm",se=FALSE)

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



n <- length(dose)  
X <- matrix(0, n, 2)  
X[, 1] <- 1  
X[, 2] <- dose  
  
# Calculando os coeficientes (parâmetros) manualmente  
coefs <- solve(t(X) %\*% X) %\*% t(X) %\*% y  
  
# Exibindo os coeficientes  
coefs

## [,1]  
## [1,] 7.55  
## [2,] 1.09

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_reta = 1.09\*dose+7.55,  
 desvio = y - mod\_reta,  
 desvio2 = desvio\*desvio)

## # A tibble: 8 x 7  
## # Groups: dose [4]  
## trat dose y y2 mod\_reta desvio desvio2  
## <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 A 1 8.2 67.2 8.64 -0.440 0.194   
## 2 A 1 7.8 60.8 8.64 -0.840 0.706   
## 3 B 2 9.8 96.0 9.73 0.0700 0.00490  
## 4 B 2 10.4 108. 9.73 0.67 0.449   
## 5 C 3 12.5 156. 10.8 1.68 2.82   
## 6 C 3 11.5 132. 10.8 0.68 0.462   
## 7 D 4 10.8 117. 11.9 -1.11 1.23   
## 8 D 4 11.2 125. 11.9 -0.710 0.504

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_reta = 1.09\*dose+7.55,  
 desvio = y - mod\_reta,  
 desvio2 = desvio\*desvio) %>%   
 ungroup() %>%   
 summarise(  
 SQT = sum(y2) - sum(y)^2/length(y),  
 SQerro\_puro = 0.83999,  
 SQDfalta\_ajus = sum(desvio2) - SQerro\_puro,  
 SQreta = SQT-SQerro\_puro-SQDfalta\_ajus,  
 R2\_lin = SQreta/SQT,  
 R2falta\_ajus = SQDfalta\_ajus/SQT,  
 )

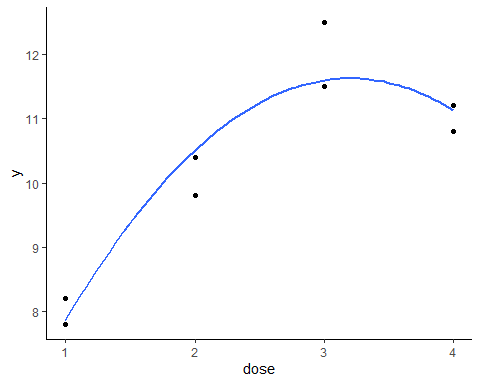
## # A tibble: 1 x 6  
## SQT SQerro\_puro SQDfalta\_ajus SQreta R2\_lin R2falta\_ajus  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 18.3 0.840 5.53 11.9 0.651 0.303

# Modelo Quadrático (parábola)

Vamos ajustar uma parábola

dados %>%   
 ggplot(aes(x=dose, y=y)) +  
 geom\_point() +   
 theme\_classic() +   
 stat\_smooth(method = "lm",   
 formula = y ~ x + I(x^2),   
 size = 1,  
 se=FALSE)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## i Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



n <- length(dose)  
X <- matrix(0, n, 3)  
X[, 1] <- 1  
X[, 2] <- dose  
X[, 3] <- dose^2  
  
# Calculando os coeficientes (parâmetros) manualmente  
coefs <- solve(t(X) %\*% X) %\*% t(X) %\*% y  
  
# Exibindo os coeficientes  
coefs

## [,1]  
## [1,] 3.675  
## [2,] 4.965  
## [3,] -0.775

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_quad = -0.7750\*dose^2 + 4.9650\*dose+3.6750,  
 desvio = y - mod\_quad,  
 desvio2 = desvio\*desvio)

## # A tibble: 8 x 7  
## # Groups: dose [4]  
## trat dose y y2 mod\_quad desvio desvio2  
## <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 A 1 8.2 67.2 7.86 0.335 0.112   
## 2 A 1 7.8 60.8 7.86 -0.0650 0.00422  
## 3 B 2 9.8 96.0 10.5 -0.705 0.497   
## 4 B 2 10.4 108. 10.5 -0.105 0.0110   
## 5 C 3 12.5 156. 11.6 0.905 0.819   
## 6 C 3 11.5 132. 11.6 -0.0950 0.00902  
## 7 D 4 10.8 117. 11.1 -0.335 0.112   
## 8 D 4 11.2 125. 11.1 0.0650 0.00423

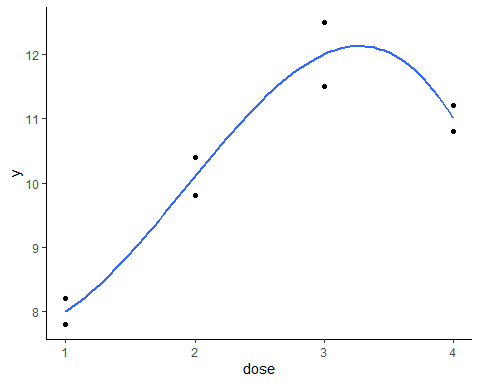
dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_quad = -0.7750\*dose^2 + 4.9650\*dose+3.6750,  
 desvio = y - mod\_quad,  
 desvio2 = desvio\*desvio) %>%   
 ungroup() %>%   
 summarise(  
 SQT = sum(y2) - sum(y)^2/length(y),  
 SQerro\_puro = 0.83999,  
 SQDfalta\_ajus = sum(desvio2) - SQerro\_puro,  
 SQquad = SQT-SQerro\_puro-SQDfalta\_ajus,  
 R2\_quad = SQquad/SQT,  
 R2falta\_ajus = SQDfalta\_ajus/SQT,  
 )

## # A tibble: 1 x 6  
## SQT SQerro\_puro SQDfalta\_ajus SQquad R2\_quad R2falta\_ajus  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 18.3 0.840 0.729 16.7 0.914 0.0399

# Modelo Cúbico

Vamos ajustar um polinômio de grau 3

dados %>%   
 ggplot(aes(x=dose, y=y)) +  
 geom\_point() +   
 theme\_classic() +   
 stat\_smooth(method = "lm",   
 formula = y ~ x + I(x^2) + I(x^3),   
 size = 1,  
 se=FALSE)



n <- length(dose)  
X <- matrix(0, n, 4)  
X[, 1] <- 1  
X[, 2] <- dose  
X[, 3] <- dose^2  
X[, 4] <- dose^3  
# Calculando os coeficientes (parâmetros) manualmente  
coefs <- solve(t(X) %\*% X) %\*% t(X) %\*% y  
  
# Exibindo os coeficientes  
coefs

## [,1]  
## [1,] 8.40  
## [2,] -2.55  
## [3,] 2.60  
## [4,] -0.45

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_cub = -0.4500\*dose^3 +2.6000\*dose^2 -2.5500\*dose+8.4000,  
 desvio = y - mod\_cub,  
 desvio2 = desvio\*desvio)

## # A tibble: 8 x 7  
## # Groups: dose [4]  
## trat dose y y2 mod\_cub desvio desvio2  
## <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 A 1 8.2 67.2 8 0.200 0.0400  
## 2 A 1 7.8 60.8 8 -0.200 0.0400  
## 3 B 2 9.8 96.0 10.1 -0.300 0.0900  
## 4 B 2 10.4 108. 10.1 0.300 0.0900  
## 5 C 3 12.5 156. 12 0.500 0.250   
## 6 C 3 11.5 132. 12 -0.500 0.250   
## 7 D 4 10.8 117. 11 -0.200 0.0400  
## 8 D 4 11.2 125. 11 0.200 0.0400

dados %>%   
 mutate(  
 y2 = y \* y) %>%   
 group\_by(dose) %>%   
 mutate(mod\_cub = -0.4500\*dose^3 +2.6000\*dose^2 -2.5500\*dose+8.4000,  
 desvio = y - mod\_cub,  
 desvio2 = desvio\*desvio) %>%   
 ungroup() %>%   
 summarise(  
 SQT = sum(y2) - sum(y)^2/length(y),  
 SQerro\_puro = 0.83999,  
 SQDfalta\_ajus = sum(desvio2) - SQerro\_puro,  
 SQcub = SQT-SQDfalta\_ajus,  
 R2\_cub = SQcub/SQT,  
 R2falta\_ajus = SQDfalta\_ajus/SQT,  
 )

## # A tibble: 1 x 6  
## SQT SQerro\_puro SQDfalta\_ajus SQcub R2\_cub R2falta\_ajus  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 18.3 0.840 0.0000100 18.3 1.00 0.000000548