Slay the Word

Mark Andrews

13 July 2018

library(tibble)  
library(dplyr)  
library(ggplot2)  
library(pander)

A univariate normal, or Gaussian, linear model is defined as follows. Assume that our data consists of independent pairs of observations:

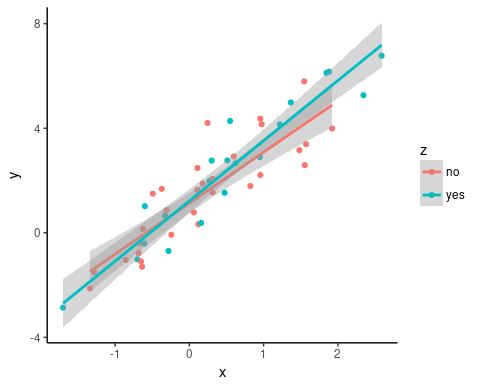
where each and . We then model this data as follows:

# Demonstration

Here, we generate some data.

N <- 50  
Df <- tibble(x = rnorm(N),  
 y = 1.25 + 2.25\*x + rnorm(N),  
 z = sample(c('yes','no'), size=N, replace=T)  
)

Df %>% ggplot(mapping=aes(x=x, y=y, col=z)) +  
 geom\_point() +  
 stat\_smooth(method = 'lm') +  
 theme\_classic()



Scatterplot with line of best fit.

Here, we fit the model with maximum likelihood estimation:

M <- lm(y ~ x + z, data=Df)  
pander(summary(M))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **(Intercept)** | 1.102 | 0.179 | 6.156 | 1.562e-07 |
| **x** | 2.127 | 0.1426 | 14.92 | 1.894e-19 |
| **zyes** | 0.2112 | 0.2844 | 0.7428 | 0.4613 |

Fitting linear model: y ~ x + z

|  |  |  |  |
| --- | --- | --- | --- |
| Observations | Residual Std. Error |  | Adjusted |
| 50 | 0.9694 | 0.833 | 0.8259 |