Chapter 4

BPR

6/16/2020

c(1, 2, 3, 4, 5) -> integers  
integers

## [1] 1 2 3 4 5

mean(integers)

## [1] 3

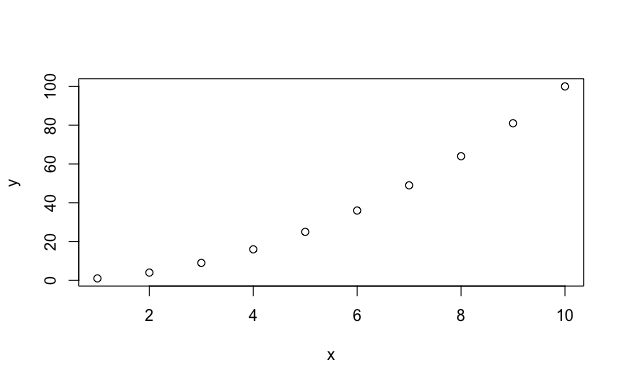
read.csv("https://raw.githubusercontent.com/barneyricca/CSRforSS/master/Data/sample.csv",  
 header = TRUE) -> sample\_data  
sample\_data

## x y  
## 1 1 2.1  
## 2 2 4.3  
## 3 3 6.1  
## 4 4 8.3  
## 5 5 10.1  
## 6 6 12.3  
## 7 7 14.1

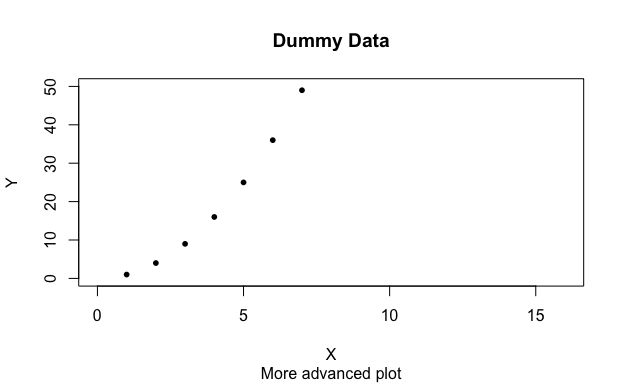
lm(y~x, data = sample\_data)

##   
## Call:  
## lm(formula = y ~ x, data = sample\_data)  
##   
## Coefficients:  
## (Intercept) x   
## 0.1857 2.0000

1:10 -> x  
x^2 -> y  
plot(x,y)

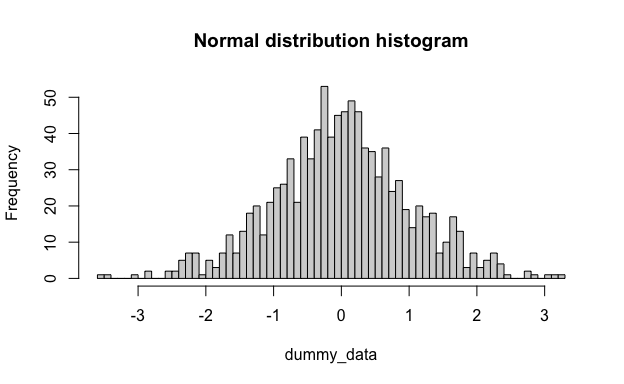


plot(x, y,  
 main = "Dummy Data", # Change the title above the plot  
 sub = "More advanced plot", # Subtitle  
 xlab = "X", # set the x-axis label  
 ylab = "Y", # set the y-axis label  
 xlim = c(0,16), # set the x-axis range   
 ylim = c(0,50), # set the y-axis range  
 pch = 16, # Use a solid dot, not an open  
 # circle  
 cex = 0.8) # make the dots a little smaller;



# cex = 1 is the default size

rnorm(n = 1000,  
 mean = 0,  
 sd = 1) -> dummy\_data  
hist(dummy\_data,  
 breaks = 50,  
 main = "Normal distribution histogram")



2 + 3

## [1] 5

8 - 5

## [1] 3

8 \* 7

## [1] 56

16 / 2

## [1] 8

sqrt(81)

## [1] 9

2 ^ 5

## [1] 32

exp(3)

## [1] 20.08554

log(8, 2)

## [1] 3

log(1000, 3)

## [1] 6.28771

mean(1:10)

## [1] 5.5

sd(1:10)

## [1] 3.02765

median(1:10)

## [1] 5.5

min(1:10)

## [1] 1

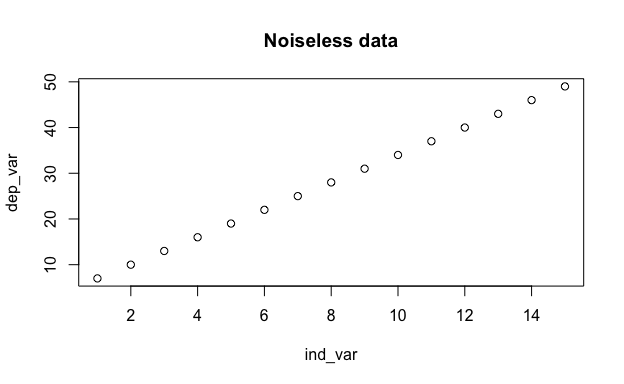
max(1:10)

## [1] 10

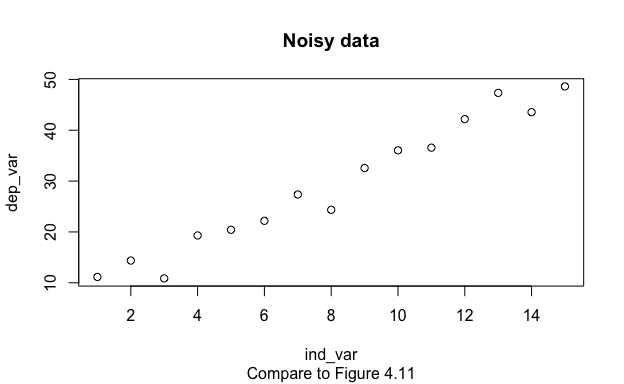
1:15 -> data1  
seq(from = 2,  
 to = 30,  
 by = 2) -> data2  
data1 + data2

## [1] 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45

1:15 -> ind\_var  
3 \* ind\_var + 4 -> dep\_var  
plot(ind\_var, dep\_var,  
 main = "Noiseless data")



set.seed(42)  
runif(n = 15,  
 min = -5,  
 max = 5) + dep\_var -> dep\_var  
plot(ind\_var, dep\_var,  
 main = "Noisy data",  
 sub = "Compare to Figure 4.11")



lm(dep\_var ~ ind\_var) -> y\_lm  
summary(y\_lm)

##   
## Call:  
## lm(formula = dep\_var ~ ind\_var)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7812 -1.1594 0.6106 1.6199 4.0616   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.4764 1.3682 4.734 0.000391 \*\*\*  
## ind\_var 2.8314 0.1505 18.816 8.19e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.518 on 13 degrees of freedom  
## Multiple R-squared: 0.9646, Adjusted R-squared: 0.9619   
## F-statistic: 354.1 on 1 and 13 DF, p-value: 8.189e-11

kable(coefficients(summary(y\_lm)),  
 digits = 3)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 6.476 | 1.368 | 4.734 | 0 |
| ind\_var | 2.831 | 0.150 | 18.816 | 0 |

{  
 plot(ind\_var, dep\_var, # Plot the data  
 main = "Dummy Data", # Change the title above the plot  
 xlab = "X", # set the x-axis label  
 ylab = "Y", # set the y-axis label  
 xlim = c(0,16), # set the x-axis range   
 ylim = c(0,50), # set the y-axis range  
 pch = 16, # Use a solid dot, not an open circle  
 cex = 0.8) # make the dots a little smaller  
 abline(a = y\_lm$coef[1], # Add a line to the plot, with  
 # intercept a  
 b = y\_lm$coef[2], # and slope b  
 col = "darkgreen") # and use a dark green color  
 text(1, 40, # Put some text info on the plot  
 # "paste" together the text and the slope,   
 # rounded to 2 digits. Put the lower left point of  
 # that text at the point (1,40) on the graph:  
 labels = paste0("slope = ",   
 round(y\_lm$coef[2], digits = 2)),   
 adj = c(0,0))  
 text(1, 35, # Put some more text on the plot  
 labels = paste("intercept = ",   
 round(y\_lm$coef[1], digits = 2)),  
 adj = c(0,0))  
}

