STAT461 HW2

Xiangyu Ren

9/11/2020

## 1.Suppose that you are planning to run an experiment with one treatment factor having four levels: “none”, “low”, “medium”, and “high”, and you have the resources to conduct the experiment on 20 experimental units. Assign at random 20 experimental units to the 4 levels of the treatment, so that each treatment is assigned 5 units. Your answer should include your R code used.

set.seed(01109)  
  
exp.unit = 1:20  
treat.nr = c(rep('none',5), rep('low',5), rep('medium',5), rep('high',5))  
treat = sample(treat.nr, replace = FALSE)  
CRD.table = data.frame(exp.unit, treat)  
print(CRD.table)

## exp.unit treat  
## 1 1 medium  
## 2 2 none  
## 3 3 medium  
## 4 4 none  
## 5 5 low  
## 6 6 low  
## 7 7 low  
## 8 8 medium  
## 9 9 low  
## 10 10 none  
## 11 11 high  
## 12 12 none  
## 13 13 medium  
## 14 14 high  
## 15 15 high  
## 16 16 low  
## 17 17 medium  
## 18 18 high  
## 19 19 high  
## 20 20 none

## 2.Repeat question 1 to obtain a second experimental design assigning the 20 units to the 4 levels of the treatment.

set.seed(01109)  
exp.unit = 1:20  
treat.nr = c(rep('none',5), rep('low',5), rep('medium',5), rep('high',5))  
treat = sample(treat.nr, replace = FALSE)  
CRD.table = data.frame(exp.unit, treat)  
print(CRD.table)

## exp.unit treat  
## 1 1 medium  
## 2 2 none  
## 3 3 medium  
## 4 4 none  
## 5 5 low  
## 6 6 low  
## 7 7 low  
## 8 8 medium  
## 9 9 low  
## 10 10 none  
## 11 11 high  
## 12 12 none  
## 13 13 medium  
## 14 14 high  
## 15 15 high  
## 16 16 low  
## 17 17 medium  
## 18 18 high  
## 19 19 high  
## 20 20 none

## 3.Suppose that you are planning to run an experiment with one treatment factor having three levels. It has been determined that . Assign at random 13 experimental units to the 3 treatments so that the first treatment is assigned 3 units and the other two treatments are each assigned 5 units.

r\_exp.unit = 1:13  
r\_treat.nr = c(rep('r1',3), rep('r2',5), rep('r3',5))  
r\_treat = sample(r\_treat.nr, replace = FALSE)  
r\_CRD.table = data.frame(r\_exp.unit, r\_treat)  
print(r\_CRD.table)

## r\_exp.unit r\_treat  
## 1 1 r2  
## 2 2 r3  
## 3 3 r3  
## 4 4 r2  
## 5 5 r3  
## 6 6 r2  
## 7 7 r1  
## 8 8 r3  
## 9 9 r3  
## 10 10 r1  
## 11 11 r1  
## 12 12 r2  
## 13 13 r2

## 4.Visit <http://www.tylervigen.com/spurious-correlations> (or some other website of your choosing) and find an example of two observed quantities that are correlated, but you think are not causally related. Clearly show the data (you could download an image), and describe why you think the two quantities are not causally related. Give an example of another factor (not measured) which you think could have a causative link with one or both of the quantities shown. Give some explanation for why this not measured factor could be causally linked to one or both of the quantities.

I picked this one, US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation.

Weird Correlation

Weird Correlation

I think these two quantities are not causally related because I think the amount of suicide cases is not a response variable nor an explanatory variable to the amount of money US spent on science. People might suicide because of working stresses or other reasons, but I can’t see any direct relations to do with science, space, or technology, it’s too far.

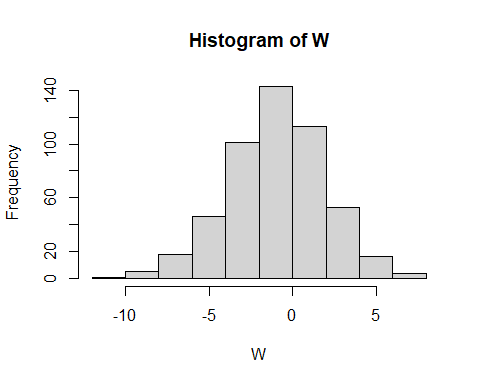
An example of another factor which I think could have a causative link with US spending on science, space, and technology is populations living on Earth. I think there is a link because with increasing populations, we might run into short of food, water, or even living spaces, therefore when we have more people, we would spend more on technology, science and space to handle these problems or find a new planet to live.

## 5.Let X∼N(2,6)and Y∼N(−3,2)and Z∼N(0,1). All three random variables are independent of each other. Do the following. Show all work.

### (a) What is the distribution of W=X+Y+Z? What are E(W) and Var(W)?

Since X,Y,Z are all R.V, and they are all independent, so W is normal and E(W) = mean(X)+mean(Y)+mean(Z)=-1, Var(W)=var(X)+var(Y)+var(Z)=9

N = 500  
X = rnorm(N, mean = 2, sd = sqrt(6))  
Y = rnorm(N, mean = -3, sd = sqrt(2))  
Z = rnorm(N, mean = 0, sd = 1)  
  
W = X+Y+Z  
alldata = data.frame(X,Y,Z,W)  
hist(W)



print(mean(W))

## [1] -0.9941789

print(var(W))

## [1] 8.417299

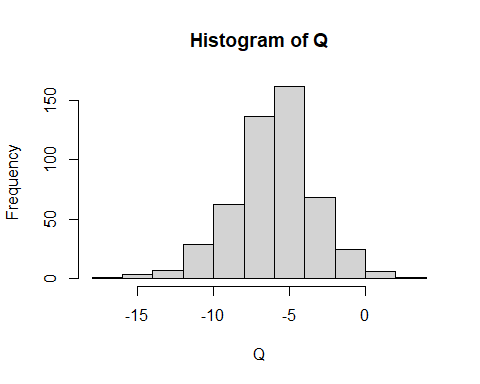
print(sd(W))

## [1] 2.901258

### (b) What is the distribution of Q=2Y?

Q is also normal distribution.

Q = 2\*Y  
y\_data = data.frame(Y,Q)  
hist(Q)



print(mean(Q))

## [1] -6.067319

print(var(Q))

## [1] 7.410732

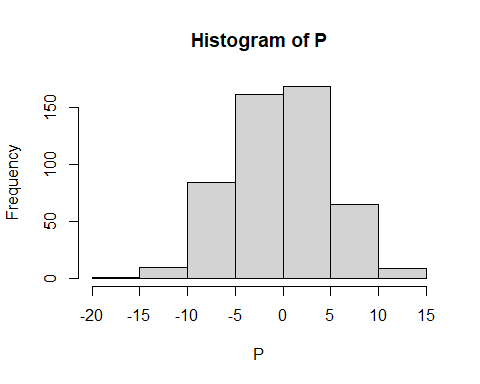
print(sd(Q))

## [1] 2.722266

### (c) What is the distribution of P=−2X+4?

P is also normal distribution.

P = -2\*X+4  
x\_data = data.frame(X,P)  
hist(P)



print(mean(P))

## [1] -0.1937621

print(var(P))

## [1] 24.00086

print(sd(P))

## [1] 4.899067

### (d) Find a and b so that M=a+bX is distributed as a standard Normal distribution.

To get a standard Normal distribution, we need E(X)=mean(X)=0, and var(X)=1. With scaling,

b = sqrt(1/6)  
b

## [1] 0.4082483

By shifting,

a = -2\*b  
a

## [1] -0.8164966

Testing:

W = a+b\*X  
print(mean(W))

## [1] 0.03955152

print(var(W))

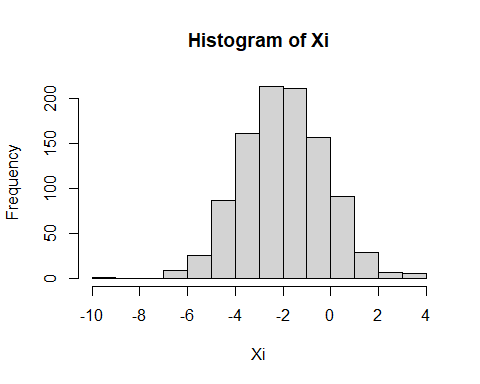
## [1] 1.000036

So, we get a = -0.8164966, b = 0.4082483.

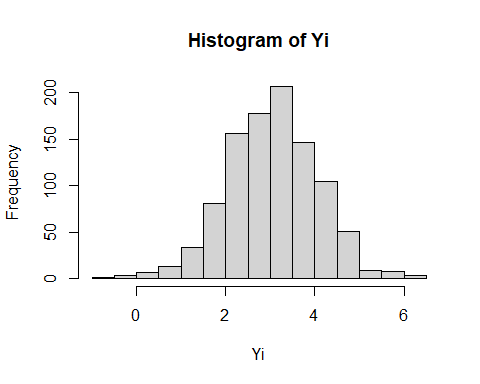
## 6. Do the following

### (a) Use R to simulate 1000 iid random variables {Xi} with Xi∼N(−2,3). Plot a histogram of your simulated values.

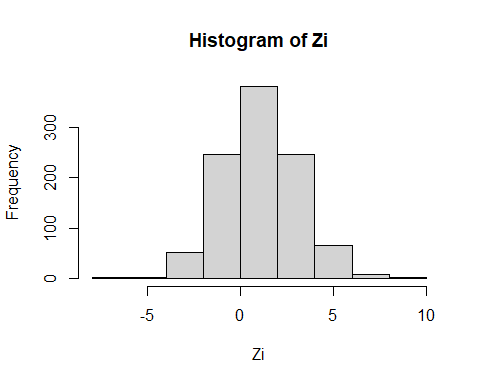
Xi = rnorm(1000, mean = -2, sd = sqrt(3))  
hist(Xi)

 ### (b) Also simulate 1000 iid random variables {Yi} with Yi∼(3,1). Plot a histogram of your simulated values.

Yi = rnorm(1000, mean = 3, sd = 1)  
hist(Yi)

 ## (c) Finally, plot a histogram of{Zi}, where Zi=Xi+Yi.

Zi = Xi + Yi  
hist(Zi)

 ## (d) Is Zi independent of Xi? Explain your answer.

Zi is not independent of Xi, because when Xi changes, Zi would also change since it’s Xi+Yi.

## (e) Find the sample mean and variance of the Zis you simulated, and compare them with the true, theoretical mean and variance.

Theoretical mean should be -2+3 = 1, and var should be 3+1 = 4

print(mean(Zi))

## [1] 1.080926

print(var(Zi))

## [1] 4.007769

Comparing the sample mean and variance with the theoretical ones, they are really close to each other with little difference.