Homework 2

Jiaqi Li

1.

> treatment.not.random=c(rep("none",4),rep("low",4),rep("medium",4),rep("high",4))

> treatment.random=sample(treatment.not.random)

> plot.ID=1:length(treatment.not.random)

> data.frame(plot.ID,treatment.not.random,treatment.random)

plot.ID treatment.not.random treatment.random

1 1 none none

2 2 none none

3 3 none low

4 4 none none

5 5 low medium

6 6 low high

7 7 low low

8 8 low medium

9 9 medium medium

10 10 medium low

11 11 medium high

12 12 medium high

13 13 high low

14 14 high medium

15 15 high high

16 16 high none

By randomization, we will have table below:

plot.ID treatment.not.random treatment.random

1 1 none none

2 2 none none

3 3 none low

4 4 none none

5 5 low medium

6 6 low high

7 7 low low

8 8 low medium

9 9 medium medium

10 10 medium low

11 11 medium high

12 12 medium high

13 13 high low

14 14 high medium

15 15 high high

16 16 high none

2.

> treatment.random2=sample(treatment.not.random)

> data.frame(plot.ID,treatment.not.random,treatment.random2)

plot.ID treatment.not.random treatment.random2

1 1 none low

2 2 none medium

3 3 none none

4 4 none low

5 5 low low

6 6 low high

7 7 low low

8 8 low high

9 9 medium high

10 10 medium high

11 11 medium medium

12 12 medium medium

13 13 high none

14 14 high none

15 15 high none

16 16 high medium

By doing randomization again, we have the new table:

plot.ID treatment.not.random treatment.random2

1 1 none low

2 2 none medium

3 3 none none

4 4 none low

5 5 low low

6 6 low high

7 7 low low

8 8 low high

9 9 medium high

10 10 medium high

11 11 medium medium

12 12 medium medium

13 13 high none

14 14 high none

15 15 high none

16 16 high medium

3.

> r.not.random=c(rep("r1=3",3),rep("r2=5",5),rep("r3=5",5))

> r.random=sample(r.not.random)

> plot.ID.r=1:length(r.not.random)

> data.frame(plot.ID.r,r.not.random,r.random)

plot.ID.r r.not.random r.random

1 1 r1=3 r1=3

2 2 r1=3 r2=5

3 3 r1=3 r1=3

4 4 r2=5 r3=5

5 5 r2=5 r3=5

6 6 r2=5 r1=3

7 7 r2=5 r3=5

8 8 r2=5 r2=5

9 9 r3=5 r2=5

10 10 r3=5 r3=5

11 11 r3=5 r2=5

12 12 r3=5 r3=5

13 13 r3=5 r2=5

By randomization, we have the table below:

plot.ID.r r.not.random r.random

1 1 r1=3 r1=3

2 2 r1=3 r2=5

3 3 r1=3 r1=3

4 4 r2=5 r3=5

5 5 r2=5 r3=5

6 6 r2=5 r1=3

7 7 r2=5 r3=5

8 8 r2=5 r2=5

9 9 r3=5 r2=5

10 10 r3=5 r3=5

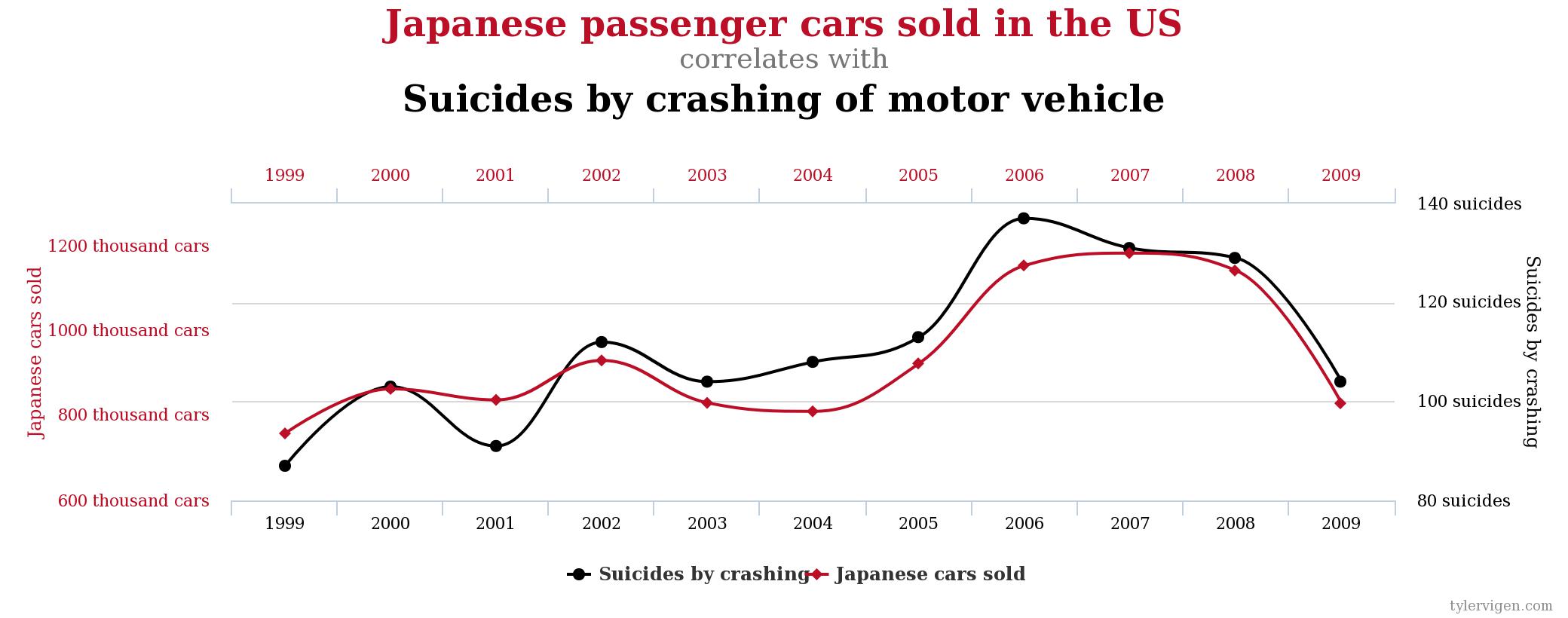
11 11 r3=5 r2=5

12 12 r3=5 r3=5

13 13 r3=5 r2=5

4.

From webside of <http://www.tylervigen.com/spurious-correlations>,



In this chart, the correlation between “Japanese passenger cars sold in the US” and “Suicides by crashing of motor vehicle” is 93.75%, which is quite high. However, I think these two observed quantities are not causally related because suicides do not necessarily relate to Japanese cars sold. Also, most of suicides by crashing of motor vehicle may not happen in US, which means Japanese passenger cars sold in US does not necessarily relate to suicides by crashing of motor vehicle.

I think American passenger cars sold in US may have a causative link with Japanese passenger cars sold in US, because Japanese cars are sold a lot in US, which could influence the sales of cars made by American in US.

5. X ∼ N(2, 6), Y ∼ N(-3, 2), Z ∼ N(0, 1), X,Y,Z are independent of each other.

(a) The distribution of W is normal distribution

X + Y + Z ∼ N(2-3+0, 6+2+1) ⇒ X + Y + Z ∼ N(-1, 9) ⇒ W ∼ N(-1, 9)

E(W) = -1, Var(W) = 9

(b) The distribution of Q is normal distribution

2 × Y ∼ N(2 × (-3), × 2) ⇒ 2Y ∼ N(-6, 8) ⇒ Q ∼ N(-6, 8)

(c) The distribution of P is normal distribution

-2 × X ∼ N(-2 × 2, × 6) ⇒ -2X ∼ N(-4, 24)

-2X+4 ∼ N(-4+4, 24) ⇒ -2X+4 ∼ N(0, 24) ⇒ P ∼ N(0, 24)

(d) b × 2 + a = 0

× 6 = 1

Solve the linear system beyond, we have a = -, b=.

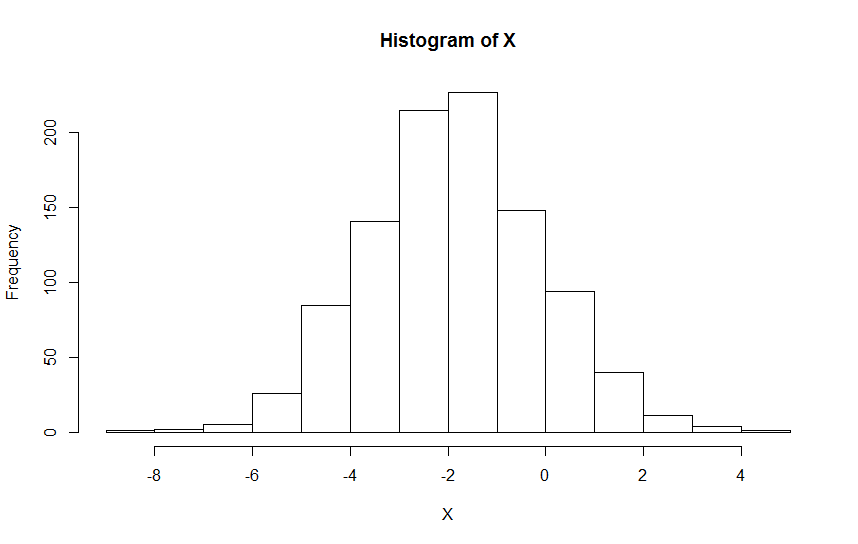
6.

(a)

> N=1000

> X=rnorm(N,mean=-2,sd=sqrt(3))

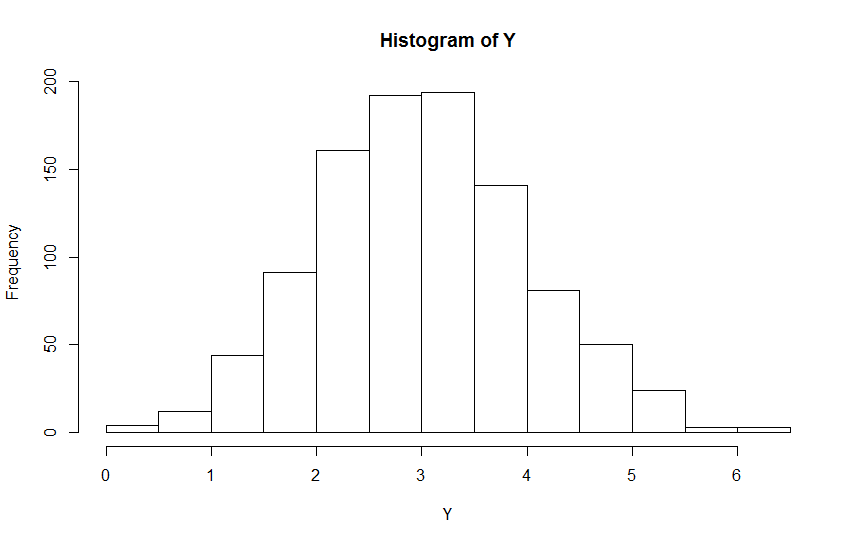
> hist(X)



(b)

> Y=rnorm(N,mean=3,sd=1)

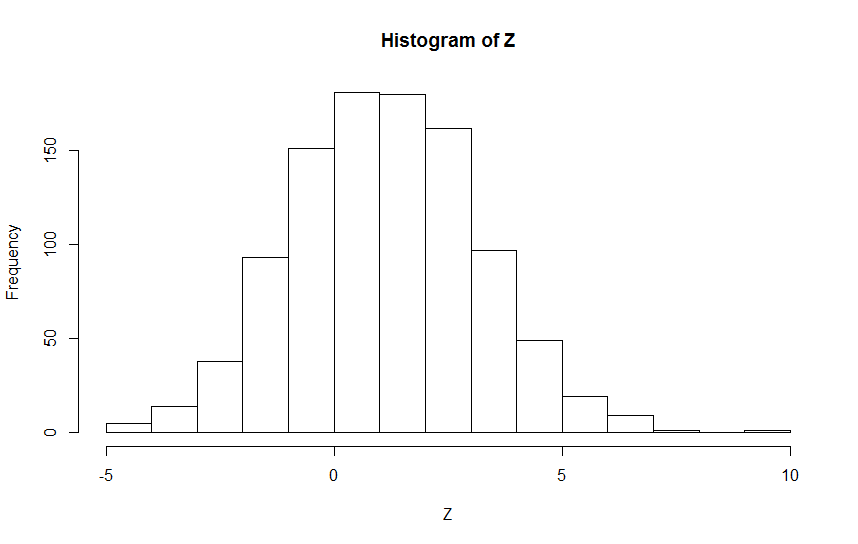
> hist(Y)



(c)

> Z=X+Y

> hist(Z)



(d)

Z = X + Y

By the equation above, we can see that Z is not independent to X because Z is composed by X and Y, which means changes in X or Y can influence Z.

(e)

> mean(Z)

[1] 1.140892

> var(Z)

[1] 4.155833

Thus, the sample mean of Z is 1.140892, the sample variance of Z is 4.155833.

The theoretical mean is 3 – 2 = 1, the theoretical variance is 3 + 1 = 4.

Compare the sample mean and variance to the theoretical mean and variance of Z, we can see that sample mean is close to the theoretical mean but not same, and the sample variance is also close to the theoretical variance.

R code:

treatment.not.random=c(rep("none",4),rep("low",4),rep("medium",4),rep("high",4))

treatment.random=sample(treatment.not.random)

plot.ID=1:length(treatment.not.random)

data.frame(plot.ID,treatment.not.random,treatment.random)

treatment.random2=sample(treatment.not.random)

data.frame(plot.ID,treatment.not.random,treatment.random2)

r.not.random=c(rep("r1=3",3),rep("r2=5",5),rep("r3=5",5))

r.random=sample(r.not.random)

plot.ID.r=1:length(r.not.random)

data.frame(plot.ID.r,r.not.random,r.random)

N=1000

X=rnorm(N,mean=-2,sd=sqrt(3))

hist(X)

Y=rnorm(N,mean=3,sd=1)

hist(Y)

Z=X+Y

hist(Z)

mean(Z)

var(Z)

R code after run:

> treatment.not.random=c(rep("none",4),rep("low",4),rep("medium",4),rep("high",4))

> treatment.random=sample(treatment.not.random)

> plot.ID=1:length(treatment.not.random)

> data.frame(plot.ID,treatment.not.random,treatment.random)

plot.ID treatment.not.random treatment.random

1 1 none medium

2 2 none none

3 3 none none

4 4 none medium

5 5 low none

6 6 low high

7 7 low none

8 8 low low

9 9 medium low

10 10 medium low

11 11 medium high

12 12 medium low

13 13 high medium

14 14 high high

15 15 high medium

16 16 high high

>

> treatment.random2=sample(treatment.not.random)

> data.frame(plot.ID,treatment.not.random,treatment.random2)

plot.ID treatment.not.random treatment.random2

1 1 none medium

2 2 none high

3 3 none none

4 4 none none

5 5 low none

6 6 low medium

7 7 low low

8 8 low high

9 9 medium low

10 10 medium medium

11 11 medium medium

12 12 medium low

13 13 high none

14 14 high high

15 15 high high

16 16 high low

>

> r.not.random=c(rep("r1=3",3),rep("r2=5",5),rep("r3=5",5))

> r.random=sample(r.not.random)

> plot.ID.r=1:length(r.not.random)

> data.frame(plot.ID.r,r.not.random,r.random)

plot.ID.r r.not.random r.random

1 1 r1=3 r3=5

2 2 r1=3 r2=5

3 3 r1=3 r3=5

4 4 r2=5 r2=5

5 5 r2=5 r2=5

6 6 r2=5 r3=5

7 7 r2=5 r3=5

8 8 r2=5 r2=5

9 9 r3=5 r1=3

10 10 r3=5 r3=5

11 11 r3=5 r1=3

12 12 r3=5 r1=3

13 13 r3=5 r2=5

>

> N=1000

> X=rnorm(N,mean=-2,sd=sqrt(3))

> hist(X)

> Y=rnorm(N,mean=3,sd=1)

> hist(Y)

> Z=X+Y

> hist(Z)

> mean(Z)

[1] 1.140892

> var(Z)

[1] 4.155833