Yohandi
assignment 8

1a. $f_X(x) = \sum_{y=1}^{4} f(x,y)$ $\quad f_Y(y) = \sum_{x=1}^{2} f(x,y)$

$f_X(1) = f(1,1) + f(1,2) + f(1,3) + f(1,4)$
$= \frac{14}{32}$

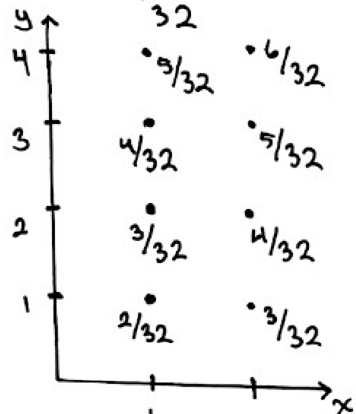
$f_X(2) = f(2,1) + f(2,2) + f(2,3) + f(2,4)$
$= \frac{18}{32}$

$f_Y(1) = f(1,1) + f(2,1)$
$= \frac{5}{32}$

$f_Y(2) = f(1,2) + f(2,2)$
$= \frac{7}{32}$

$f_Y(3) = f(1,3) + f(2,3)$
$= \frac{9}{32}$

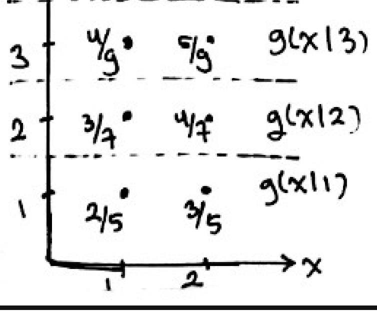
$f_Y(4) = f(1,4) + f(2,4)$
$= \frac{11}{32}$



Joint and marginal pmfs

b. $g(x|y) = \frac{f(x,y)}{f_Y(y)}$

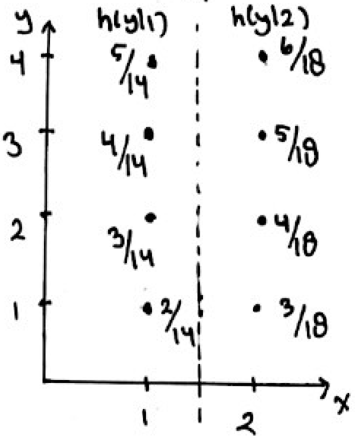
$g(1|1) = \frac{2}{5}$ $\quad g(2|1) = \frac{3}{5}$
$g(1|2) = \frac{3}{7}$ $\quad g(2|2) = \frac{4}{7}$
$g(1|3) = \frac{4}{9}$ $\quad g(2|3) = \frac{5}{9}$
$g(1|4) = \frac{5}{11}$ $\quad g(2|4) = \frac{6}{11}$



Conditional pmfs of x given y

c. $h(y|x) = \frac{f(x,y)}{f_X(x)}$

$h(1|1) = \frac{2}{14}$ $\quad h(1|2) = \frac{3}{18}$
$h(2|1) = \frac{3}{14}$ $\quad h(2|2) = \frac{4}{18}$
$h(3|1) = \frac{4}{14}$ $\quad h(3|2) = \frac{5}{18}$
$h(4|1) = \frac{5}{14}$ $\quad h(4|2) = \frac{6}{18}$



Conditional pmfs of y given x

d. $p(1 \le y \le 3 \mid X=1) = \sum_{y=1}^{3} h(y|1) = \frac{9}{14}$

$p(y \le 2 \mid X=2) = \sum_{y=1}^{2} h(y|2) = \frac{7}{18}$

$p(X=2 \mid y=3) = g(2|3) = \frac{5}{9}$

e. $E(Y|X=1) = \sum_{y=1}^{4} y\,h(y|1) = \frac{20}{7}$

$Var(Y|X=1) = \sum_{y=1}^{4} y^2 h(y|1) - E(Y|X=1)^2 = \frac{55}{49}$

2a. $X \sim B(n,p)$

$n = 400$

$p = \frac{n(\{(W,W),(R,W),(W,R)\})}{4} = \frac{3}{4}$

$\therefore X \sim B(400, \frac{3}{4})$

b. $E(X) = np = 300$

$Var(X) = np(1-p) = 75$

c. $Y \sim B(X, p')$

$X = 300$

$p' = \frac{n(\{(R,W),(W,R)\})}{3} = \frac{2}{3}$

d. $E(Y|X=300) = Xp' = 200$

$Var(Y|X=300) = Xp'(1-p') = \frac{200}{3}$

3a. $f(x,y) = P(X=x,\ Y=y)$

$= \frac{\binom{30}{x}\binom{30-x}{y} 4^{30-x-y}}{6^{30}}$

b. $f_Y(y) = P(Y=y)$

$= \frac{\binom{30}{y} 5^{30-y}}{6^{30}}$

$g(x|y) = \frac{f(x,y)}{f_Y(y)}$

$= \frac{\binom{30}{x}\binom{30-x}{y}}{\binom{30}{y}} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{30-x-y}$

$= \binom{30-y}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{(30-y)-x}$

$g(x|y) \sim B(30-y, \frac{1}{5})$

c. $E(X^2 - 4XY + 3Y^2)$

$= E(X^2) + 3E(Y^2) - 4E(XY)$

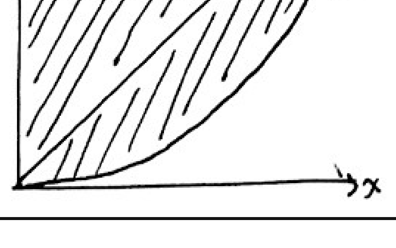
$= 4(Var(X) + E(X)^2)$

$-4(\rho \sigma_x \sigma_y + E(X)^2)$

$= 4\left(30(\tfrac{1}{6})(\tfrac{5}{6}) - \left(-\sqrt{\frac{(\frac{1}{6})^2}{(\frac{5}{6})^2}}\right) 30(\tfrac{1}{6})(\tfrac{5}{6})\right)$

$= 20$

4a,



b. $P(X>Y) = \int_0^1 \left(\frac{4}{3}\right)(x - x^3)\,dx = \frac{4}{3}\left(\frac{1}{2}x^2 - \frac{1}{4}x^4\right)\Big|_{x=0}^{1} = \frac{1}{3}$

5. $\int_2^{2.5}\int_2^{2.3} c\,dy\,dx = 1$

$\Rightarrow c = \frac{20}{3}$

$f(y \mid |x-y| \le 0.1) = \int_2^{2.1} \left(\frac{20}{3}\right)((y+0.1)-2)\,dy + \int_{2.1}^{2.3} \left(\frac{20}{3}\right) 0.2\,dy$

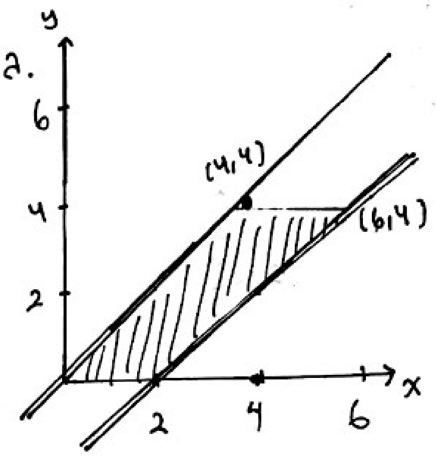
$= \frac{11}{30}$

6. for every $Y \sim U(x-0.1, x+0.1)$,

$E(Y) = \frac{(x-0.1)+(x+0.1)}{2} = x$

$E(Y|X) = E(X) = \int_{0.2}^{\infty} x\,(2e^{-2(x-0.2)})\,dx = \frac{7}{10}$ (thousand dollars)

7a.



b. $f_X(x) = \frac{1}{8}(\min(x,4) - \max(x-2, 0))$

c. $f_Y(y) = \frac{1}{8}((y+2)-y) = \frac{1}{4}$

d. $h(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\min(x,4)-\max(x-2,0)}, \; x \in (0,6)$

$h(y|0) = h(y|6) = 0$

e. $g(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{2}$

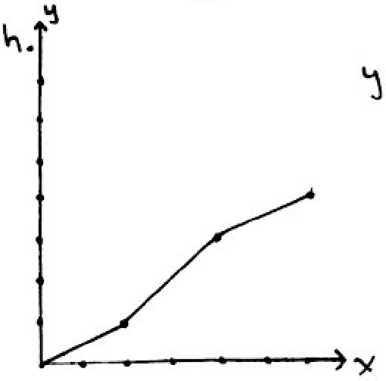
f. $E(y|x) = \begin{cases} \int_0^x \frac{y}{x}\,dy = \frac{x}{2}, & x \in (0,2) \\ \int_{x-2}^x \frac{y}{2}\,dy = x-1, & x \in [2,4] \\ \int_{x-2}^4 \frac{y}{6-x}\,dy = \frac{x+2}{2}, & x \in (4,6) \end{cases}$

g. $E(x|y) = \int_{-\infty}^{\infty} x\, g(x|y)\, dx$

$= \int_y^{y+2} \frac{x}{2}\, dx$

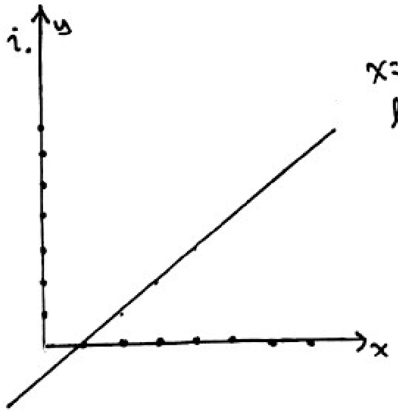
$= y+1$

h.



$y = E(y|x)$ is not linear

i.



$x = E(x|y)$ is linear

8a. $P(-5<X<5) = \phi\left(\frac{5-(-3)}{5}\right) - \phi\left(\frac{-5-(-3)}{5}\right) \approx 0.6006$

b. $E(X|Y=13) = -3+\left(\frac{3}{5}\right)\frac{5}{3}(13-10) = 0$

$\sigma^2_{X|Y} = 25-\left(\frac{3}{5}\right)^2(25) = 16$

$P(-5<X<5 \mid Y=13) = \phi\left(\frac{5-0}{4}\right) - \phi\left(\frac{-5-0}{4}\right) \approx 0.7888$

c. $P(7<Y<16) = \phi\left(\frac{16-10}{3}\right) - \phi\left(\frac{7-10}{3}\right) \approx 0.8185$

d. $E(Y|X=2) = 10+\left(\frac{3}{5}\right)\frac{3}{5}(2-(-3)) = \frac{59}{5}$

$\sigma^2_{Y|X} = 9-\left(\frac{3}{5}\right)^2 9 = \frac{144}{25}$

$P(7<Y<16 \mid X=2) = \phi\left(\frac{16-\frac{59}{5}}{\frac{12}{5}}\right) - \phi\left(\frac{7-\frac{59}{5}}{\frac{12}{5}}\right) \approx 0.9371$

9a. $P(18.5<Y<25.5) = \phi\left(\frac{25.5-22.7}{3.5}\right) - \phi\left(\frac{18.5-22.7}{3.5}\right) \approx 0.6730$

b. $E(Y|x) = 22.7+(0.78)\left(\frac{3.5}{4.2}\right)(x-22.7) = 7.945+0.65x$

c. $Var(Y|x) = 12.25-(0.78)^2\, 12.25 = 4.7971$

d. $P(18.5<Y<25.5 \mid X=23)$

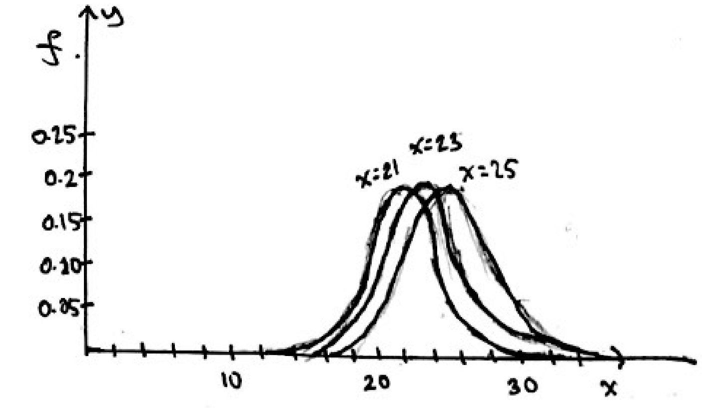
$= \phi\left(\frac{25.5-(7.945+0.65\cdot 23)}{\sqrt{4.7971}}\right) - \phi\left(\frac{18.5-(7.945+0.65\cdot 23)}{\sqrt{4.7971}}\right)$

≈ 0.8608

e. $P(18.5<Y<25.5 \mid X=25)$

$= \phi\left(\frac{25.5-(7.945+0.65\cdot 25)}{\sqrt{4.7971}}\right) - \phi\left(\frac{18.5-(7.945+0.65\cdot 25)}{\sqrt{4.7971}}\right)$

≈ 0.7210



$\Rightarrow X \sim N(0,1)$

$\cdot)$ similarly we can prove that Y also follows normal distribution with mean $= 0$ and standard deviation $= 1$ (swap x and y)

10. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}} \left(1 + xye^{-\left(\frac{x^2+y^2-2}{2}\right)}\right) dx\,dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2}}\,dx\,dy + \frac{1}{2\pi}\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xye^{-\left(\frac{x^2+y^2-2}{2}\right)}\,dx\,dy}_{\text{odd function} \Rightarrow \text{the integral equals to zero}}$$

$$= \frac{1}{2\pi}\underbrace{\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy}_{\sqrt{2\pi}}\ \underbrace{\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx}_{\sqrt{2\pi}}$$

$$= 1$$

$\Rightarrow$ this shows that $f(x,y)$ is a joint pdf

$$\cdot)\ f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \left(1 + xye^{-\frac{x^2+y^2-2}{2}}\right) dy$$

$$= \frac{1}{2\pi} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} \Big(\underbrace{e^{-\frac{y^2}{2}}}_{\sqrt{2\pi}} + \underbrace{xy \cdot e^{-\frac{y^2}{2}} \cdot e^{-\frac{x^2+y^2-2}{2}}}_{\text{odd function} \Rightarrow \text{the integral equals to zero}}\Big) dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

yohandi - assignment 0 (computer-based)

for $X \sim U(-L, +L)$,

$$E(X) = \frac{(-L)+L}{2} = 0$$

$$Var(X) = \frac{(L-(-L))^2}{12} = \frac{L^2}{3}$$

given that $Var(X) = 1$;

$$\Rightarrow \frac{L^2}{3} = 1$$

$$\Rightarrow L = \pm\sqrt{3}$$

In the same way, we can generate independent samples of a zero-mean, uniform random variable Y with $Var(Y) = 1$.

suppose we have coefficient ρ as the correlation of both X and Y distributions,

for $Z = \rho X + \sqrt{1-\rho^2}\, Y$,

$\Rightarrow E(Z) = E(\rho X + \sqrt{1-\rho^2}\, Y)$

$$= E(\rho X) + E(\sqrt{1-\rho^2}\, Y)$$

$$= \rho \underbrace{E(X)}_{0} + \sqrt{1-\rho^2}\, \underbrace{E(Y)}_{0}$$

$$= 0$$

$\Rightarrow Var(Z) = Var(\rho X + \sqrt{1-\rho^2}\, Y)$

$$= \rho^2 \underbrace{Var(X)}_{1} + (1-\rho^2)\underbrace{Var(Y)}_{1} + 2\, \underbrace{Cov(X,Y)}_{0} \cdot \rho\sqrt{1-\rho^2}$$

$E[(X-\mu_X)(Y-\mu_Y)]$

since X and Y are independent

$\underbrace{E(\underbrace{X-\mu_X}_{0})\, E(\underbrace{Y-\mu_Y}_{0})}_{0}$

$$= \rho^2 + (1-\rho^2)$$

$$= 1$$

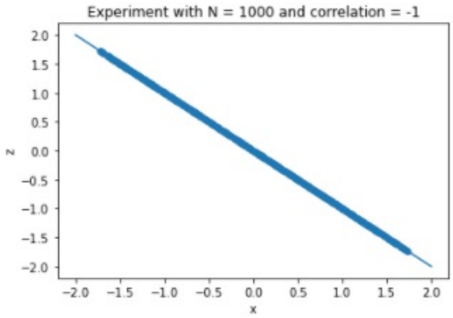
$\Rightarrow E(XZ) = E(\rho X^2 + \sqrt{1-\rho^2}\, XY)$

$$= E(\rho X^2) + E(\sqrt{1-\rho^2}\, XY)$$

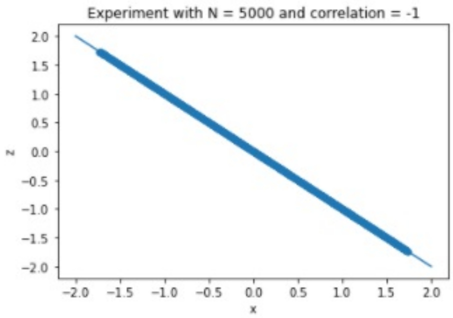
$$= \rho E(X^2) + \sqrt{1-\rho^2}\, E(XY)$$

$$= \rho(\underbrace{Var(X)}_{1} + \underbrace{E(X)^2}_{0}) + \sqrt{1-\rho^2}\, \underbrace{E(X)}_{0}\, \underbrace{E(Y)}_{0}$$

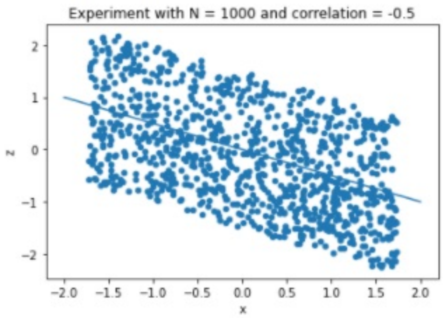
$$= \rho$$



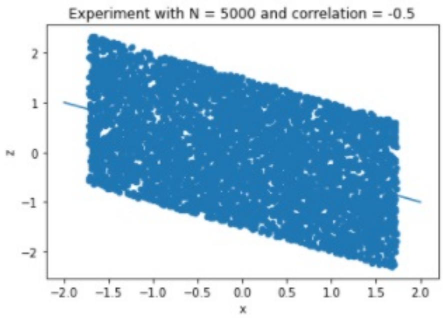

```
For the case where:
  n = 1000
  correlation = -1
 => the sigma/N equals to = -0.9624647902682075
```



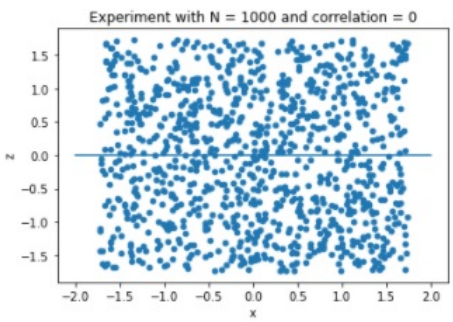

```
For the case where:
  n = 5000
  correlation = -1
 => the sigma/N equals to = -1.0089446126922659
```



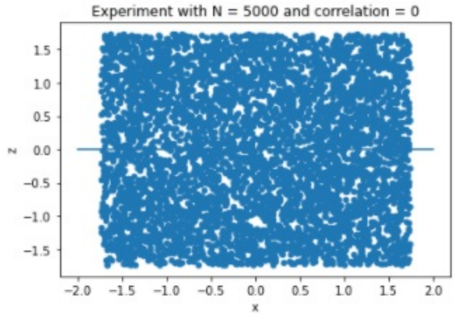

```
For the case where:
  n = 1000
  correlation = -0.5
 => the sigma/N equals to = -0.5291048157620475
```



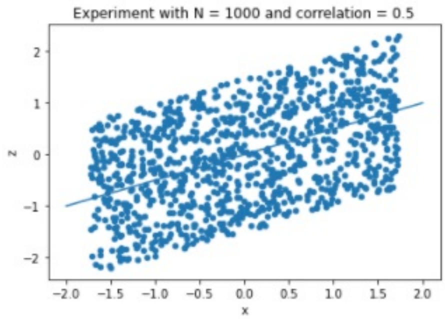

```
For the case where:
  n = 5000
  correlation = -0.5
 => the sigma/N equals to = -0.4874085318910722
```



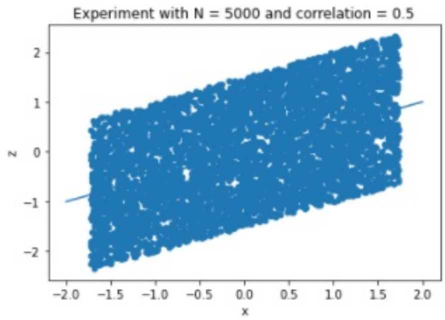

```
For the case where:
  n = 1000
  correlation = 0
 => the sigma/N equals to = -0.046926151395193766
```



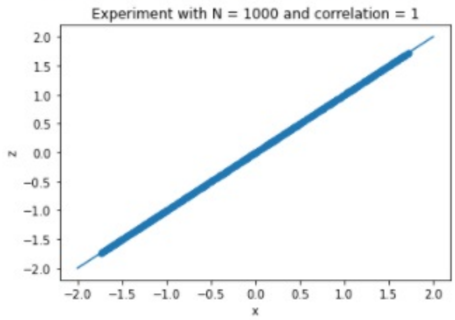

```
For the case where:
  n = 5000
  correlation = 0
 => the sigma/N equals to = -0.010972445393097222
```



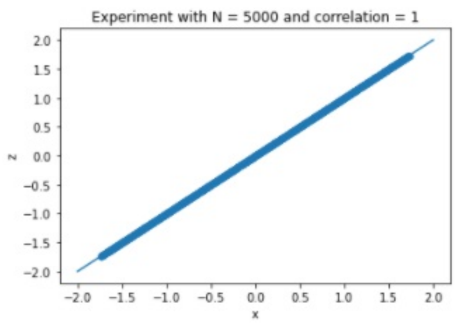

```
For the case where:
  n = 1000
  correlation = 0.5
 => the sigma/N equals to = 0.43632299016084247
```




```
For the case where:
  n = 5000
  correlation = 0.5
 => the sigma/N equals to = 0.5151775822280955
```




```
For the case where:
  n = 1000
  correlation = 1
 => the sigma/N equals to = 0.9842011780890118
```




```
For the case where:
  n = 5000
  correlation = 1
 => the sigma/N equals to = 1.0006543766963865
```

from those graphs,

-> we can observe that when $\rho=-1$ or $\rho=1$, the points are neatly scattered along the line $z=\rho x$

-> on the other hand the points $(x_i, z_i)_{i=1}^{N}$ are diffused for $\rho=-0.5$ or $\rho=0.5$.
Although, compared with the case where $|\rho|=1$, the points are widely spread along the line $z=\rho x$

-> in the case where $\rho=0$, the points are blindly spread evenly

-> in here, the larger N, the stronger the intensity of points plotted

those observations imply:

-> $\rho=0$ indicates no linearship between the RVs

-> $|\rho|=1$ indicates a perfect linear relationship between the RVs

from the computed sigma,

we can see that : $\rho \sim \frac{1}{N}\sum_{i=1}^{n} x_i z_i$

theoretically,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i z_i = \rho$$

(this can be observed by comparing the value error of each N=1000 and N=5000 with $\rho=\rho$)

```python
import math
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

def experiment(n, correlation):
    x = np.random.uniform(-math.sqrt(3), math.sqrt(3), n)
    y = np.random.uniform(-math.sqrt(3), math.sqrt(3), n)
    z = [correlation * x[i] + math.sqrt(1 - correlation ** 2) * y[i] for i in range(n)]
    df = pd.DataFrame({'x' : x, 'z' : z}, columns = ['x', 'z'])
    df.plot(x = 'x', y = 'z', kind = 'scatter')
    plt.plot(np.linspace(-2, 2, 100), correlation * np.linspace(-2, 2, 100))
    plt.title("Experiment with N = " + str(n) + " and correlation = " + str(correlation))
    plt.show()
    print("For the case where:")
    print("  n =", n)
    print("  correlation =", correlation)
    print(" => the sigma/N equals to =", sum([x[i] * z[i] for i in range(n)]) / n)

correlationList = [-1, -0.5, 0, 0.5, 1]
nList = [1000, 5000]
for correlation in correlationList:
    for n in nList:
        experiment(n, correlation)
```