Report

Analyzing the consummation of coffee in university student

Submitted by

1. Worawalan Chatlatanagulchai 6510545713
2. Chananthida Sopaphol 6510545331
3. Napasorn Tevarut 6510545519
4. Thunyanan Tangpipatpong 6510545501

Group11

present

Faculty of Engineering Kasetsart University

1/2022

01204216 Probability and Statistics

executive summary

Analyzing the consummation of coffee in university student

Overview

This project Focus on analyzing the consumption of coffee among university students. By using Collected data about relation of student and consummation of coffee in university student. We can collect data from 68 students who do our survey.

Problem

- What is relation between coffee drinking and university student behavior?

Solution

- Collected data from university student and use calculate data by using statistic method.

Objective

- to analyze the consumption of coffee among university students.
- For study relation of university student and drinking coffee.

Data collection

Objective:

- collect information about coffee
- collect information about student
- collect information that relate and tell relation of student and drinking coffee

representative sample:

- all university student (age 18 – 20)

We used survey that contain 16 question to collected data through google form.

Question:

1. How old are you? (Years old)
2. What is your GPAX (Overall GPA: 0.00 - 4.00)?
3. Do you drink coffee? if not, what do you drink instead?
4. Average cup of coffee you consume per week during term break? (One cup of coffee is 250 ml)
5. Average cup of coffee you consume per week in university during open semester? (One cup of coffee is 250 ml)
6. Average amount of money you spend on coffee during term break per week?
7. Average amount of money you spend on coffee in university during semester per week?
8. How many average hours do you sleep during term break?
9. How many average hours do you sleep during open semester?
10. What time do you usually drink coffee? (1.00-24.00)
11. How many hours do you spend on studying per week?
12. How many hours do you spend on playing games per week during term break?
13. How many hours do you spend on playing games per week during open semester?
14. On which day you tend to drink coffee the most?
15. On which day you tend to drink coffee the least?
16. Why do you drink coffee? (Ex. To stay fresh during the day/ Just a personal favor)

Data type:

- Nominal/Ordinal data: question 3,15,16,17
- Numerical data: question 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,15
- Pair data: question: 4-5, 6-7, 8-9, 12-13

We have receiving 68 respond from survey.

Example of received Data

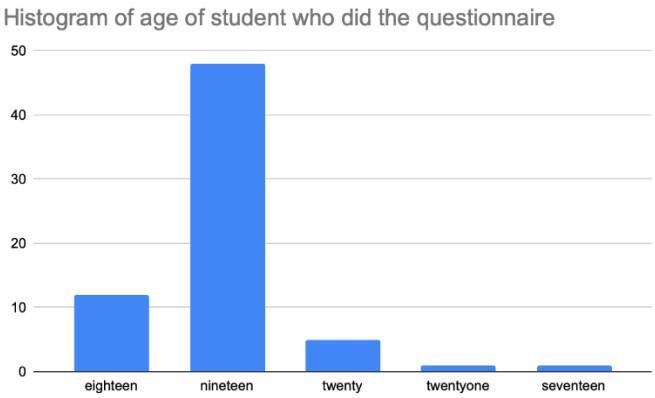
A	B	C	D	E	F	G	H
ประทับเวลา	คุณอายุเท่าไร (ปี)	เรียนยาลัย (GPAX)	คุณดื่มกาแฟหรือไม่	ช่างวัน/ความชอบส่วนตัว/ทั้งคู่)	กดิคุณดื่มกาแฟเวลาไหน	งสัปดาห์ที่คุณดื่มกาแฟมากที่สุด	ณดื่มกาแฟน้อยที่สุด
10/4/2023, 8:39:	18	3.67	ดื่ม	เพื่อความสดชื่นระหว่างวัน	9	วันจันทร์	วันพุธ
11/4/2023, 9:15:	18	2.86	ดื่ม	เพื่อความสดชื่นระหว่างวัน	10	วันพุธ	วันอาทิตย
11/4/2023, 9:21:	18	3.57	ดื่ม	ความชอบส่วนตัว	13	วันพฤหัสบดี	วันอาทิตย
11/4/2023, 10:00	19	3.36	ดื่ม	เพื่อความสดชื่นระหว่างวัน	9	วันจันทร์	วันเสาร์
11/4/2023, 12:12	20	3.39	ดื่ม	เพื่อความสดชื่นระหว่างวัน	7.5	วันพุธ	วันศุกร์
11/4/2023, 13:02	20	3.73	ดื่ม	ความชอบส่วนตัว	6	วันเสาร์	วันจันทร์
11/4/2023, 14:47	19	2.71	ดื่ม	ื่อความสดชื่น และ ความชอบ	8	วันจันทร์	วันเสาร์

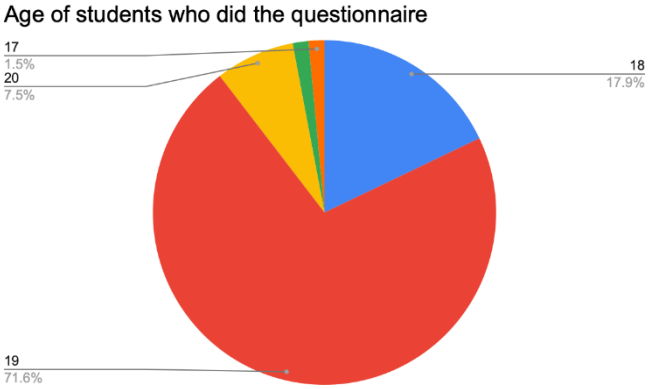
Methodology

Single Data

1. Age of people who did the questionnaire

- -Average age = 18.882
- Mode = 19
- Median = 19
- Max = 21
- Min = 17
- Standard deviation = 0.6116





The statistic tells that people who do the survey almost is 19 years old.

Hypothesis test

sample mean = 18.882, sample standard deviation = 0.612

claim: Claim that average age of university student is more than 18

Proof whether claim is true at significance level = 0.01

1,2,3) u = average age, u0 = 18

H0 = u = 18, Ha = u > 18

4) test statistics = 11.89744913

5) significance level = 0.01

z0.01 = 2.33, rejection region z>= 2.33

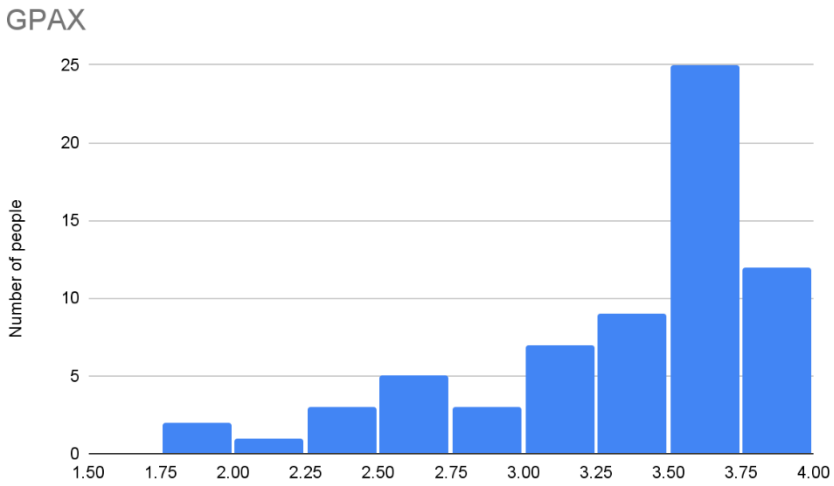
6) reject null hypothesis

7) Average age > 18, upper-tailed test

It mean average student in university have age more than 18 year old

2. GPAX of people who did the questionnaire

- Average GPAX = 3.34
- Mode = 3.5
- Median = 3.5
- Min = 1.75
- Max = 4
- Standard deviation = 0.528
- GPAX is left skewed distribution



Hypothesis test

sample mean = 3.335, sample standard deviation = 0.528

claim: Claim that average GPAX of university student is more than 3

Proof whether claim is true at significance level = 0.01

1,2,3) u = average age, u0 = 3

H0 = u = 3, Ha = u > 3

4) test statistics = 5.236767851

5) significance level = 0.01

z0.01 = 2.33, rejection region z>= 2.33

6) reject null hypothesis

7) Average GPAX < 3, upper-tailed test

The statistic is tell us that true average of student have GPAX less than 3

Goodness of fit test

Test GPAX of university students at =0.01 whether, is normally distributed with mean μ=3.335, σ=0.528 or not

Hypothesis:

H0: GPAX is distributed with normal distribution with =3.335, = 0.528

Ha: GPAX is not distributed with normal distribution with =3.335, = 0.528

Test statistics = 27.347, k=9, m=0, α = 0.01, P = 1 - =1 - 0.01 = 0.99

Degree of freedom 1 = k-1-m = 9-1-0 = 8

Cutoff of non-rejection region 1 = 21.666

Degree of freedom 1 = k-1 = 9-1 = 8

Cutoff of rejection region 2 = 21.666

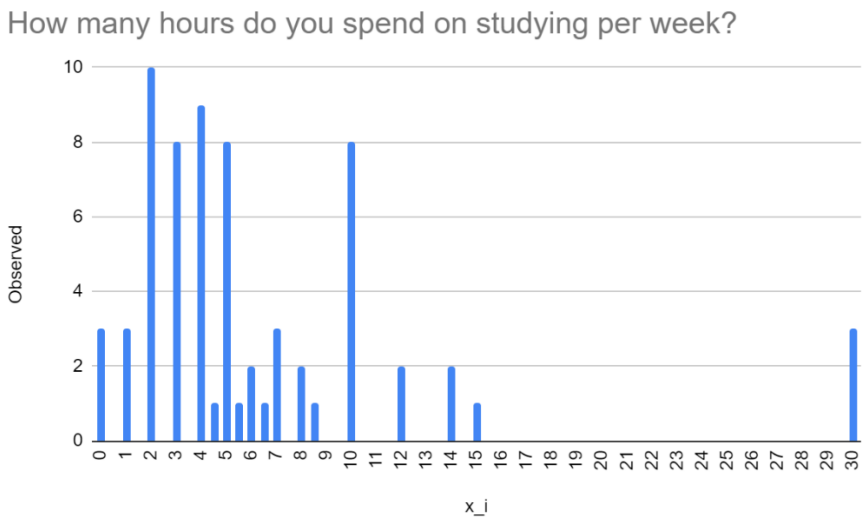
Rejection region: X^2 >= X^2 = 21.666

We reject null hypothesis

Conclusion: GPAX is not distributed with normal distribution with μ =3.335, σ= 0.528

3. Hours spend on studying per week

- Average study hour = 6.446969697
- mode = 10
- Median = 5
- Min = 0
- Max = 30
- Standard deviation = 6.316723096
- Study hour is right skewed distribution



Goodness of fit test

Test study hour of university students at =0.01 whether is normally distributed with mean μ= 6.446969697, σ= 6.316723096 or not

H0: Data is distributed with normal distribution.

Ha: Data is not distributed with normal distribution.

Test statistics = 1754.372216 , k=60, m=2, α = 0.01, P = 1 - =1 - 0.01 = 0.99

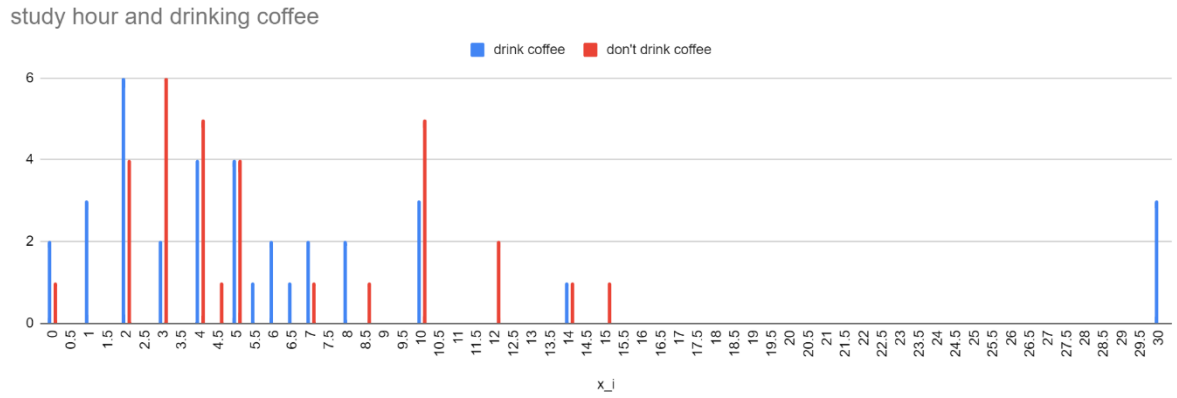
Cutoff of rejection region: z = 84.733

Rejection region: X^2 >= X^2 = 21.666

We reject null hypothesis

Conclusion: Data is not distributed with normal distribution with parameter μ= 6.446969697, σ= 6.316723096

Hypothesis test



From study hour and drinking coffee it tells us that drinking or not is not tell much about their study hour so we decide to perform hypothesis test with null hypothesis: average hour that spend on study of student who drink coffee is more than 5 hour per week to test that drink coffee is have relation or not with average study hour

1) parameter of interest: u = average hour spends on study

2) null value: u = 6

3) H0: average hour that spend on study of student who drink coffee is more than 5 hour per week

Ha: average hour that spend on study of student who drink coffee is less than 5 hour per week

4) test statistics = 1.40479654

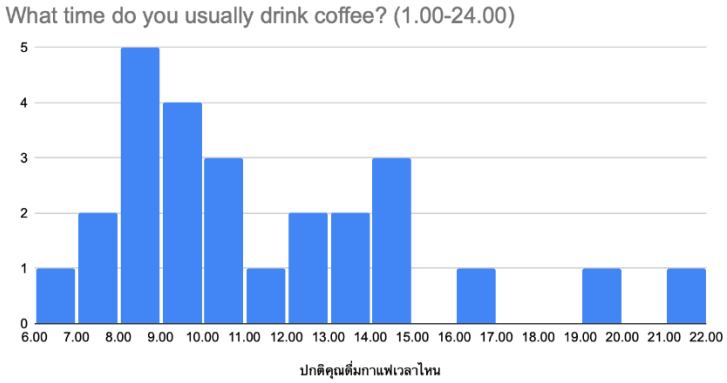
5) significance level = 0.01 z0.01 = 2.33, rejection region z>= 2.33

6) not reject null hypothesis

7) use upper-tailed test
conclusion: average hour that spend on study of student who drink coffee is more than 5 hour per week

4. Time that usually drinks coffee

- min = 6
- max = 21
- median = 10
- mode = 8
- data is right continuous skewed distribution



<u>Goodness of fit test</u>

Test Time that university student usually drink coffee at =0.01 whether is normally distributed with mean μ= 11, σ= 3.690528417 or not

H0: Data is distributed with normal distribution.

Ha: Data is not distributed with normal distribution.

Test statistics = 1746.596756, k=40, m=2, α = 0.01, P = 1 - =1 - 0.01 = 0.99

Cutoff of rejection region: z = 72.443

Rejection region: X^2 >= X^2 = 72.443

We reject null hypothesis

Conclusion: Data is not distributed with normal distribution with parameter μ= 11, σ= 3.690528417

<u>Hypothesis test</u>

1) parameter of interest: u = average time that usually drink coffee

2) null value: u = 6

3) H0: Average time that usually drink coffee is between 9-12

Ha: Average time that usually drink coffee is not between 9-124) test statistics = 1.40479654

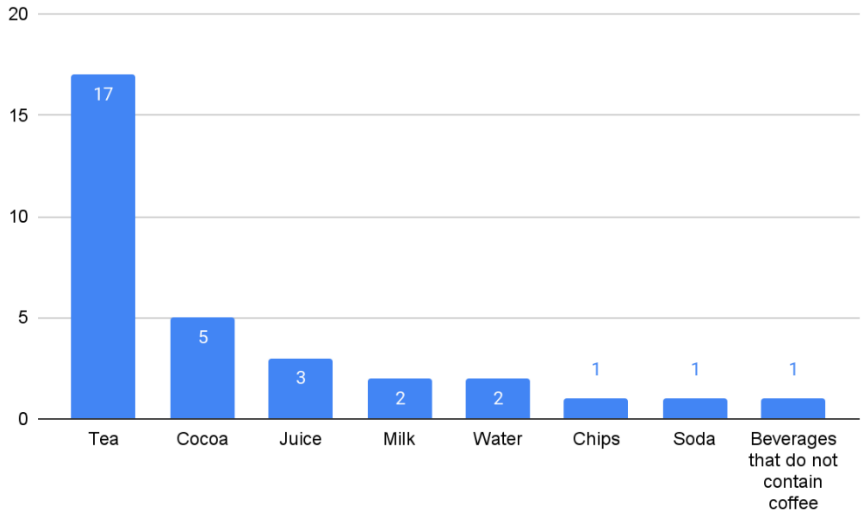
5) significance level = 0.01 z0.005 = 2.5758, rejection region z<= - 2.5758 or z >= 2.5758

6) not reject null hypothesis

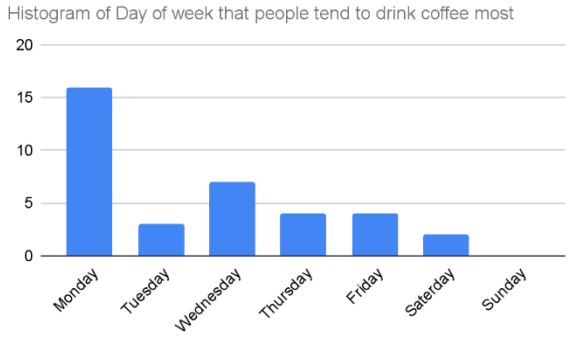
7) use two-tailed test.

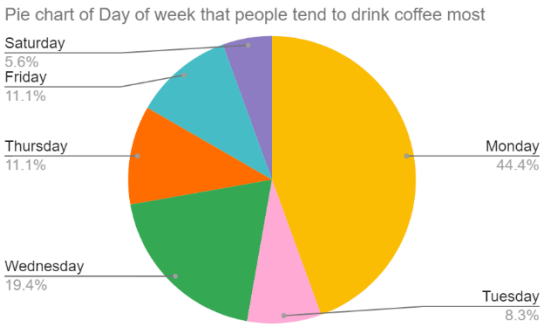
conclusion: Average time that usually drink coffee is between 9-12

5. the drink that student who not drink coffee drink instead

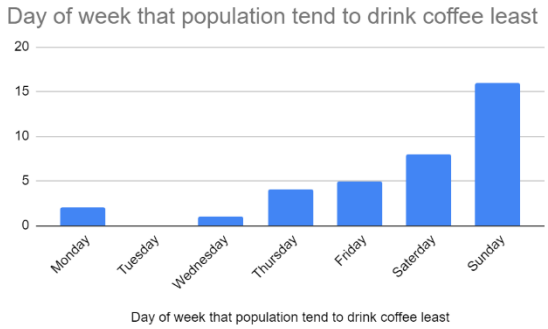


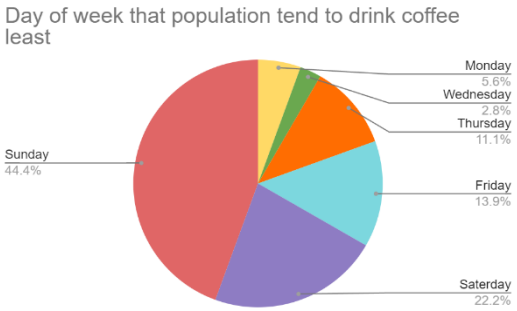
6. Day of the week that people tend to drink coffee most
    - Mode = Monday (1)
    - Median = Tuesday (2)
    - Max = Saturday (6)
    - Min = Monday (1)





7. Day of week that population tend to drink coffee least
    - Average = 5.722
    - Mode = Sunday (7)
    - Median = Saturday (6)
    - Min = Monday (1)
    - Max = Sunday (7)

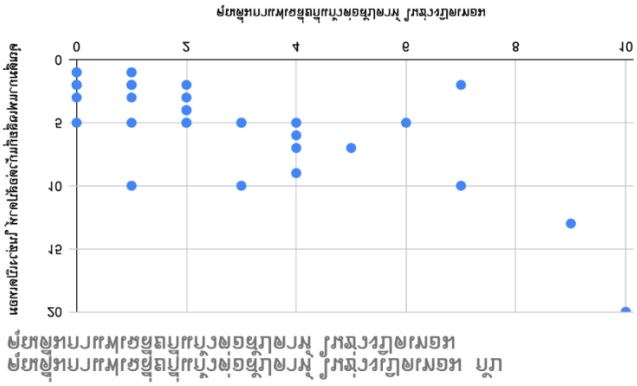




Pair Data

1) **Average cup of coffee consumed per week during open and break semester**
    - x = Average cup of coffee consumed per week during close semester.
    - y = Average cup of coffee consumed per week during open semester.
    - Average cup of coffee consumed per week during close semester = 2.61
    - Average cup of coffee consumed per week during open semester = 4.94
    - Median of X = 2
    - Median of Y = 4.5
    - Mode of X = 1
    - Mode of Y = 5
    - Covariance = 7.34

Scatter plot



Correlation between X and Y = 0.728106476

If one drinks coffee a lot during close semester, they will likely drink coffee a lot during open semester, too. So, it means close or open semester is not main attribute that make student drink more coffee. But if student drink coffee a lot in close semester that student tend to drink a lot of coffee in open semester too.

Conditional probability

Probability that cup of coffee consumed per week during open semester that student consume more than 2

P(Y>=2) = 22/36 = 0.61

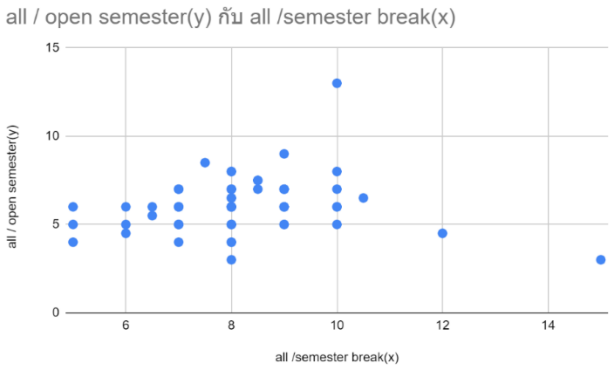
Interesting findings

Correlation	Cup of coffee during close semester	Cup of coffee during open semester
Age	0.049393917	0.1398721228
GPAX	-0.2076486487	-0.3543592501

The correlation of age and Cup of coffee during close semester and open semester is nearly to 0 so it is independence to each or other but GPAX and cup of coffee seems to have some negative correlation that if student drink a lot of cups of coffee, GPAX is tend to decrease.

2) **Average hours do you sleep during term break and open semester**
   - x = Average sleep hour per week during close semester**.**
   - y = Average sleep hour per week during open semester**.**
   - Average sleep hour per week during close semester = 8.146153846
   - Average sleep hour per week during close semester = 6.115384615
   - Median of X = 8
   - Median of Y = 6
   - Mode of X = 8
   - Mode of Y = 6
   - Covariance = 1083.378107

Scatter plot



Correlation = 0.09633860135

The correlation tells us that the data is independent it means sleep hour when you have to go to study and not depend on each person

Conditional probability

Probability when of student who drink coffee and not and their sleep hour in semester break

x,y	Drink coffee	don't drink coffee
5	0.02941176471	0.01470588235
5.5	0	0
6	0.01470588235	0.02941176471
6.5	0.02941176471	0.01470588235
7	0.07352941176	0.08823529412
7.5	0.01470588235	0.01470588235
8	0.1470588235	0.1617647059
8.5	0.01470588235	0.04411764706
9	0.1176470588	0.05882352941
9.5	0	0
10	0.04411764706	0.04411764706
10.5	0.01470588235	0
11	0	0
11.5	0	0
12	0.01470588235	0
12.5	0	0
13	0	0
13.5	0	0
14	0	0
14.5	0	0
15	0.01470588235	0

Probability when of student who drink coffee and not and their sleep hour in open semester

x,y	Drink coffee	don't drink coffee
3	0.02941176471	0
3.5	0	0
4	0.02941176471	0.02941176471
4.5	0.01470588235	0.01470588235
5	0.1323529412	0.04411764706
5.5	0.02941176471	0.01470588235
6	0.1470588235	0.1617647059
6.5	0.04411764706	0
7	0.05882352941	0.1323529412
7.5	0	0.02941176471
8	0.01470588235	0.02941176471
8.5	0	0.01470588235
9	0.01470588235	0
9.5	0	0
10	0	0
10.5	0	0
11	0	0
11.5	0	0
12	0	0
12.5	0	0
13	0.01470588235	0

Expected value = 8.117647059

Expected value = 6.080882353

Expected value of sleep hour of drink coffee student = 7.522058824

Expected value of sleep hour of don't drink coffee student = 7.617647059

This mean it more likely to sleep in semester break than in open semester but sleep hour is not depended on drink coffee or not due to the expected value that vey nearly

Interesting findings

Drink coffee

Correlation(x,y)	Sleep hour (break)	Sleep hour (open)	age	GPAX
Sleep hour (break)	1	0.03706187376	0.06982296345	0.2148360634
Sleep hour (open)	0.03706187376	1	-0.4190115657	0.1116250593
age	0.06982296345	-0.4190115657	1	0.04994004399
GPAX	0.2148360634	0.1116250593	0.04994004399	1

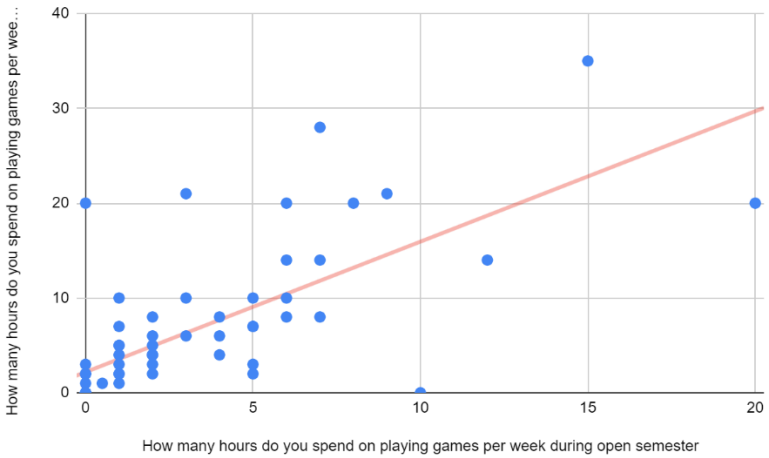
Don't drink coffee

Correlation(x,y)	Sleep hour (break)	Sleep hour (open)	age	GPAX
Sleep hour (break)	1	0.3375345804	0.1641519754	-0.05859686389
Sleep hour (open)	0.3375345804	1	0.07101013363	-0.2745558582
age	0.1641519754	0.07101013363	1	0.09584895028
GPAX	-0.05859686389	-0.2745558582	0.09584895028	1

Form above table interesting thing is student who drink coffee tend to have more GPAX when sleeping a lot. Conversely, people who do not drink coffee will have more grade when less sleeping

3) **Average playing games hour per week during term break**
   - x = Average hours spent on playing games per week during term break.
   - y = Average hours spent on playing games per week during open semester.
   - Average x = 6.2813
   - Average y = 2.9766
   - Median of X = 4
   - Median of Y = 1.5
   - Mode of X = 0
   - Mode of Y = 0
   - Covariance = 21.0457

   Scatter plot



Correlation between X and Y = 0.7089

this tell us that student who like to play game will play it even in open semester but the time will decrease so the correlation is positive

Conditional probability

Probability that student playing game for x ,y hour during open and close semester

	0<= x<1	1<= x<2	2<= x<3	3<= x<4	4<= x<5	5<= x<6	6<= x<7	7<= x<8	8<= x<9	9<= x<10	10<= x<11	11<= x<12	12<= x<13	13<= x<14	14<= x<15	x>= 15	P^(X)
0<= y<1	0.25	0	0	0	0	0	0	0	0	0	0.0156	0	0	0	0	0	0.2656
1<= y<2	0.0313	0.0156	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0469
2<= y<3	0.0469	0.0313	0.0156	0	0	0.0156	0	0	0	0	0	0	0	0	0	0	0.1094
3<= y<4	0.0156	0.0156	0.0156	0	0	0.0156	0	0	0	0	0	0	0	0	0	0	0.0625
4<= y<5	0	0.0156	0.0469	0	0.0156	0	0	0	0	0	0	0	0	0	0	0	0.0781
5<= y<6	0	0.0313	0.0156	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0469
6<= y<7	0	0	0.0156	0.0156	0.0156	0	0	0	0	0	0	0	0	0	0	0	0.0469
7<= y<8	0	0.0156	0	0	0	0.0031	0	0	0	0	0	0	0	0	0	0	0.0469
8<= y<9	0	0	0.0156	0	0.0156	0	0.0156	0.0156	0	0	0	0	0	0	0	0	0.0625
9<= y<10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
y>= 10	0.0156	0.0156	0	0.0313	0	0.0156	0.0469	0.0313	0.0156	0.0156	0	0	0.0156	0	0	0.0313	0.2344

Expected value = 4.836

Interesting findings

correlation	GPAX
during semester	0.3061
term break	0.2944
average	0.3200

From the table show us correlation between playing game hour and GPAX this tell us student who play game has a good grade

4) **Average amount of money you spend on coffee during term break and open semester**
   - X: Average amount of money spend on coffee during term break per week.
   - Y: Average amount of money spend on coffee in university during semester per week.
   - Average x = 217.92
   - Average y = 406.94
   - variance x = 157,343.39
   - variance y = 223,566.11
   - Covariance = 30,040.31

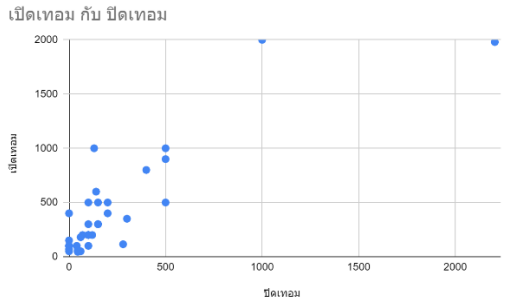
Correlation

	GPAX
term break	-0.05110271874
during semester	-0.1483980444
average	-0.1080225663

Correlation nearly to 0. Data is independence. So, it means that student spend money on coffee not has relation to their grade.

Correlation between X and Y = 0.1602

This correlation is nearly 0 too. student spend money on their interest not depend on time.



Conditional probability

Probability that student spend x ,y money for coffee during open and close semester

P^(X, Y)	0 <= Y < 100	100 <= Y < 200	200 <= Y < 300	300 <= Y < 400	400 <= Y < 500	500 <= Y < 600	600 <= Y < 700	700 <= Y < 800	800 <= Y < 900	900 <= Y < 1000	1000 <= Y < 2000	2000 <= X < 3000	P^(X)
0 <= X < 100	0.1389	0.1944	0.0278	0.0000	0.0278	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3889
100 <= X < 200	0.0000	0.0278	0.1111	0.0833	0.0000	0.0556	0.0278	0.0000	0.0000	0.0000	0.0278	0.0000	0.3333
200 <= X < 300	0.0000	0.0278	0.0000	0.0000	0.0278	0.0278	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0833
300 <= X < 400	0.0000	0.0000	0.0000	0.0278	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0278
400 <= X < 500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0278	0.0000	0.0000	0.0000	0.0278
500 <= X < 600	0.0000	0.0000	0.0000	0.0000	0.0000	0.0278	0.0000	0.0000	0.0000	0.0278	0.0278	0.0000	0.0833
600 <= X < 700	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
700 <= X < 800	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
800 <= X < 900	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
900 <= X < 1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1000 <= X < 2000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0278	0.0278
2000 <= X < 3000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0278	0.0000	0.0278

Expected value = 269.4444

Conclusion