



UNIVERSITAS ISLAM INDONESIA
FAKULTAS MIPA
Jl. Kaliurang Km. 14,5 Jogjakarta

FINAL EXAM ODD ACADEMIC 2019/2020

Course : Data Intelligent
Study Program/Class : Statistika/ A dan B
Date : Tuesday, 12 January 2021
Time : 10:15 – 10:30 WIB
Duration : 1x24 hours
Dosen Penguji : Arum Handini Primandari
Method : Take home
Media : Google form, Google Docs

CPL : PPa (Intelligence) Understanding the concepts of probability theory and statistics, mathematics, calculus, elementary linear algebra, statistical analysis methods, and elementary computer programming

Section 1, problem number 1-13 [Poin 100, weight: 80%]
CPMK PPa3 Students are able to compose intelligent computer programming for data analysis.
Indicator: - Students employ python library - Students analyze data using python
Section 2, problem number 1 [Poin: 15, weight: 20%]
CPMK PPa3 Students are able to compose intelligent computer programming for data analysis.
Indicator: - Students formulate an information of the data using descriptive analysis in python

Instruction:
1. Use your UII account to log in to your browser. 2. Open Google Classroom of Data intelligence, under the post "UAS/Final Semester Examination." 3. The test is divided into two sections: [section 1] multiple choices and [section 2] essay. 4. Read carefully every problem because we use two datasets, "Student Performance Dataset" and "Data Covid-19 INA".

5. Please note that you will be asked to collect your jupyter notebook at the end of section 1, so make sure that you employ python for calculation.
6. The form will be closed automatically when the exam time has expired.

The verification of Final Exam, Odd Academic Year 2020/2021				
The suitability of the problems with CO	The Completeness of the problems information	Verifier	Verification date	Sign
Suitable/ Less suitable / Not Suitable	Complete/ Less Complete / Not Complete	Science cluster coordinator	7-Jan-2021	

Statement:

I hereby confirm that I completed the exam alone, without consulting with or the assistance and cooperation of any other person. (Dengan ini saya menyatakan bahwa: Saya mengerjakan ujian akhir berikut secara mandiri, tanpa bantuan orang lain.)

☐ Yes

☐ No

SECTION 1: Multiple Choices

Use "Student Performance Data Set" provided in the posting of UAS. The data is student achievement in secondary education of two Portuguese schools for mathematics subject. The metadata is attached in the same post.

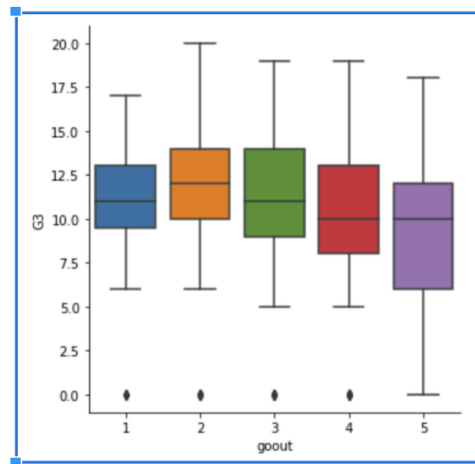
	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10

5 rows × 33 columns

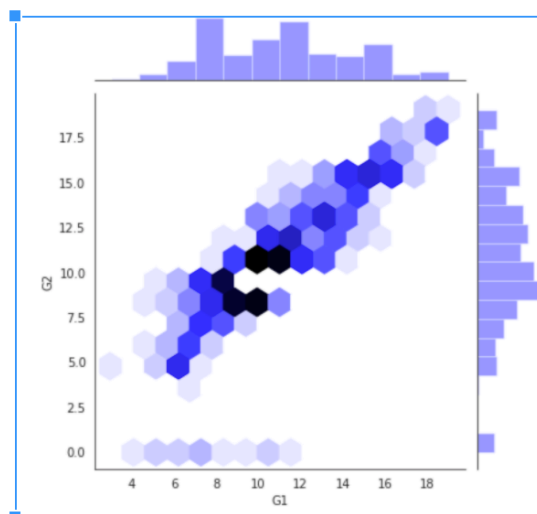
Figure 1. A Glimpse of Student Performance Dataset

1. What is the library used to read a XLSX-file?
 - a. numpy
 - b. pandas
 - c. read_excel
 - d. read_csv
2. How many variables in the data identified as "object" (presented as string) types?
 - a. 5
 - b. 10
 - c. 15
 - d. more than 15
3. What is the data size?
 - a. 395 of rows, 33 of columns
 - b. 394 of rows, 33 of columns
 - c. 394 of rows, 32 of columns
 - d. none above
4. What is the percentage ratio of female and male students?
 - a. 47% : 53%
 - b. 53% : 47%

- c. 48% : 52%
 - d. 52% : 48%
5. Utilize the information on family size and parent's cohabitation status: what is the probability that student is not the only child in their family?
- a. 0.67
 - b. 0.87
 - c. 0.66
 - d. 0.71
6. The boxplot shows the final grade and the frequency of students going out with their friends. Select all the conclusions that suitable for the graph. Pick 2 answer that apply.



- a. the more frequent of going out with friends, the worse the final grade
 - b. students who have more often going out always get worse final grade
 - c. the student who gets the lowest final grade is the one who has the most frequent of hanging out
 - d. the student who gets the highest final grade is not the one who has the least frequent of hanging out
7. The hexbin plot portrays the first and second grades. Select all the conclusions that apply.



- a. the higher the first grade, the higher the second grade
 - b. there are probably some outlier contained in the data
 - c. most of the students have the first grade between 8 to 12
 - d. the first grade distribution is more likely to be normal than the second grade distribution
8. What can be inferred from the two tails paired t-test for first and second grade? Use $\alpha = 5\%$.
- a. there is no significant difference between the first and second grade
 - b. there is significant difference between the first and second grade
 - c. the first grade is significantly higher than the second grade
 - d. the first grade is significantly lower than the second grade
9. What can be inferred from the Kendall Tau correlation test for study time and frequency of going out with friends? Use $\alpha = 5\%$.
- a. there is a significant linear relationship between the study time and frequency of going out with friends
 - b. there is no significant linear relationship between the study time and frequency of going out with friends
 - c. the correlation between the study time and frequency of going out with friends is strong and positive
 - d. the correlation between the study time and frequency of going out with friends is strong, but negative value
10. What can be inferred from the normality test for the final grade? Use $\alpha = 5\%$.
- a. the population of final grade is normally distributed
 - b. the population of final grade is not normally distributed
 - c. the population of final grade is exponentially distributed
 - d. the population of final grade is not exponentially distributed
11. What can be inferred from one way ANOVA test for the grades?
- a. the means for all grades are equal
 - b. the means for all grades are not equal
 - c. there is grade that is different from others
 - d. the means for all grades are equal to zero
12. Find the error in the following list comprehension to create a new list of the reason for choosing the school in which the term "course" substitute with "curriculum".

```
list_baru = [baris if "course" in baris else "curriculum" for baris in mat['reason']]
print(list_baru)
```

- a. the expression should be "not in"
- b. the iterable should be "in mat"
- c. the condition should be "!=course"
- d. the syntax should used regex

13. Below is the information about students' parents. Please give the values (T/F) for each statement. Read the metadata to find variable definitions.
- a. The number of mothers who work in health care is more than the number of fathers who work in the same field. (T/F)
 - b. Most of the father have higher education. (T/F)
 - c. Fewer than 10% of mother have primary education or below. (T/F)

Section 2: Essay

Perform a clustering analysis for Covid-19 Cases in Indonesia. This data contains Case, RR (Recovery Rate), CFR (Case Fatality Rate), PR (Positivity Rate), and TR (Test Rate). Use this sheet to write about the results of your clustering analysis that is cluster profiling. You can include both table and graphic to emphasize the description. Please note that you only allowed writing the answer for a max of 2 pages.

	Provinsi	Case	RR	CFR	PR	TR
0	Aceh	8776	0.817912	0.041	0.0437	22.9
1	Bali	18263	0.907354	0.029	0.0588	17.0
2	Banten	19161	0.552320	0.023	0.0358	27.9
3	Bangka Belitung	2638	0.750569	0.016	0.0956	10.5
4	Bengkulu	3779	0.790156	0.031	0.2049	4.9

Figure 2. A Glimpse of Covid-19 INA Dataset

Answer:

Matrix score:

Items/score	4	3	2	1	0
Plagiarism	no flag	1 flag	2 flag	3 flag	more than 3 flag
Table	there is a table about descriptive analysis each cluster, with a good explanation/insight	there is a table about descriptive analysis, but lacking of table explanation	there is a table about descriptive analysis, but there is no explanation	there is a table about descriptive analysis, but lacking of information	there is no table
Graph	there is an appropriate graph with a good description		there is graph, but lacking of description		there is no graph
Cluster profile	there are a clear explanation of cluster profile	there are explanation of cluster profile, but not correct	there are explanation of cluster profile, but not clear	there are explanation of cluster profile, but not related to the result	there is no cluster profile

