Cairo University
Faculty of Computers and Artificial Intelligence

Midterm Exam
Model 1

Department: Operations Research and Decision Support
Course Name: Data Analytics
Course Code: DS342
Instructor(s): Dr.Marwa Sabry – Dr Hayam Gamal

Date: November 23, 2023
Duration: 1 hour
Total Marks: 40 Marks

تعليمات هامة

- حيازة التليفون المحمول مفتوحا داخل لجنة الإمتحان يعتبر حالة غش تستوجب العقاب وإذا كان ضرورى الدخول بالمحمول فيوضع مغلق فى الحقائب.
- لا يسمح بدخول سماعة الأذن أو البلوتوث.
- لايسمح بدخول أي كتب أو ملازم أو أوراق داخل اللجنة والمخالفة تعتبر حالة غش.

Part I – Multiple Choice Questions (MCQ) ***[15 marks]***

***In the "MCQ / T or F" section of the bubble sheet,* fill in the letter of the choice that best completes each sentence.**

1. Tables created through Excel's table feature allow users to filter ____ by different values.
   a. rows. b. columns. c. sections. d. values.
2. Which of the following would *not* be included under data analysis?
   a. Measuring uncertainty b. Data description
   c. Data inference d. Search for relationships
3. To examine relationships between two categorical variables, we can use
   a. counts and corresponding charts of the counts. b. scatter plots.
   c. histograms. d. boxplots.
4. ____ is the correct syntax of IF() Function.
   a. =IF (logical_test, TRUE([value_if_true]), FALSE([value_if_false]))
   b. =IF (logical_test, [value_if_true], [value_if_false])
   c. =IF (logical_test, {[value_if_true]}, {[value_if_false]})
   d. =IF (logical_test: [value_if_true], [value_if_false])
5. Examples of comparison problems include
   a. salary broken down by male and female subpopulations.
   b. cost of living broken down by region of a country.
   c. recovery rate for a disease broken down by patients who have taken a drug and patients who have taken a placebo.
   d. all these choices.
6. A sample, selected from a population, taken at one particular point in time is categorized as
   a. categorical. b. discrete. c. cross-sectional. d. time-series.
7. Creating tables is as easy as highlighting cells that have already been filled in appropriately and then clicking on the insert tab and then clicking on the table button.
   a. No, highlighting cells disables the ability to create tables.
   b. Yes, it is that simple, but the highlighted cells must have numerical data.
   c. Yes, it is that simple, but the highlighted cells should not have blank cells.
   d. No, you go to the home tab and then click on the table button.

8. The limitation of covariance as a descriptive measure of association is that it
   a. only captures positive relationships.
   b. does not capture the units of the variables.
   c. is very sensitive to the units of the variables.
   d. is invalid if one of the variables is categorical.
9. Gender and states of residence are examples of ____ data.
   a. Discrete b. Categorical c. Continuous d. Ordinal
10. The length of the box in the box plot portrays the
   a. mean. b. median. c. range. d. interquartile range.
11. In the following MS Excel spreadsheet, you are given a list of 100 customers. Column A is their names, B is for customer category, C for payment category (0 means discounted price, 1 means full price), and D indicates price that customers pay.
   Which of the following formula correctly counts all customers who are adults and get discounted price?
   a. = countif (B2:B101, " =Adult", C2:C101, " =0")
   b. = countif (B2:B101, =Adult, C2:C101, " =0")
   c. = countifs (B2:B101, " =Adult", C2:C101, " =0")
   d. = countifs (B2:B101, " =0", C2:C101, " = Adult")

	A	B	C	D
1	Customer Name	Customer Category	Payment category	Price (Rs)
2	Raman	Child	0	0
3	Dinesh	Adult	1	5
4	Jashan	Adult	0	2
:	:	:	:	:

12. Which of the following is true?
   a. When entering an expression of text into an excel function, it must be enclosed in double quotes.
   b. A spreadsheet model should always include input numbers, rather than cell references, in formulas.
   c. If we enter A1:A5 as part of an Excel function, this refers to cells A2, A3, and A4…the cells that are between A1 and A5, exclusive.
   d. All of these statements are true.
13. In Excel, the model outputs are
   a. the numeric values that result from combinations of inputs and decision variables through the use of logical formulas.
   b. useful for making formulas more readable.
   c. the variables a decision maker has control over to obtain the best solutions.
   d. useful for finding a particular value based on a comparison.
14. Where will you find "time" on a time series graph?
   a. horizontal axis b. first column c. vertical axis d. last column
15. Categorizing a numeric age variable as "young," "middle-aged," and "elderly" is an example of
   a. counting. b. ordering. c. quantifying. d. binning.

<u>Part II – True or False</u> ***[10 marks]***

***In the "MCQ / T or F" section of the bubble sheet*, circle A for true or false.**

16. Excel's IF function can be used to determine if an expression is true or false.
17. Spreadsheet modeling is the process of entering the outputs into a spreadsheet and then relating them appropriately, by means of formulas, to obtain the decision variables.
18. To form a scatterplot of X versus Y, X and Y must be paired variables.
19. When creating a spreadsheet model, it is important to keep sensitivity in mind, because other people will be reading and trying to make sense out of your spreadsheet models.

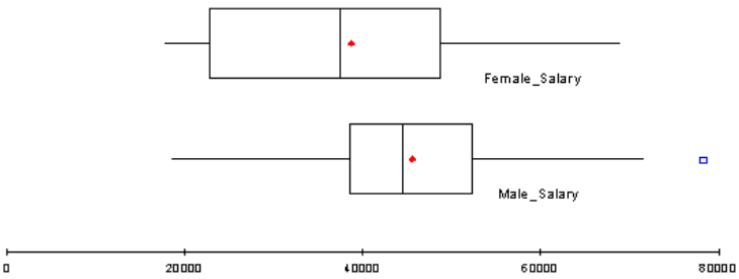
20. All nominal data may be treated as ordinal data.
21. Correlation and covariance can be used to examine relationships between numerical variables and categorical variables that have been coded numerically.
22. Relationships between two variables are more evident when counts are expressed as percentages of row totals or column totals.
23. There are four quartiles that divide the values in a data set into four equal parts.
24. Unlike histograms, box plots depict only one aspect of a variable.
25. A distribution with a flattened peak has almost all its observations within three standard deviations of the mean.
26. If you change the data, the chart will change simultaneously.
27. Strongly related variables may have a correlation close to zero if the relationship is nonlinear.
28. We can use side-by-side boxplots to compare at most 2 distributions of numeric data.
29. A PivotTable is comprised of three areas: ROW, COLUMN and DATA.
30. Data analysis includes data description, data visualization, data inference, and the search for relationships in data.
31. An example of a joint category of two variables is the count of all non-drinkers who are also nonsmokers.
32. Correlation is not useful for describing the strength and direction of linear relationships.
33. Counts for a categorical variable are often expressed as percentages of the total.
34. The filters field of a pivot table contains the data that you want summarized.
35. When a formula is copied into another cell, the relative references in the formula keep their relative positions.

<u>Part III – Analytical Multiple-Choice Questions (MCQ)</u> ***[15 marks]***

***In the "MCQ / T or F" section of the bubble sheet,* fill in the letter of the choice that best completes each sentence.**

For questions 23 to 27, answer the following:

36. A manager for Marko Manufacturing, Inc. has recently been hearing some complaints that women are being paid less than men for the same type of work in one of their manufacturing plants. The box plots shown below represent the annual salaries for all salaried workers in that facility (40 men and 34 women).



Based upon the boxplots, does there seem to be reason to conclude that there is a difference between the salaries of women and men in this plant?
a. True b. False

37. Upon your answer in question (23):
a. the mean of men is higher than that of women. b. there exist outliers.
c. middle 50% of salaries for men is above the median for women.
d. all of the previous.

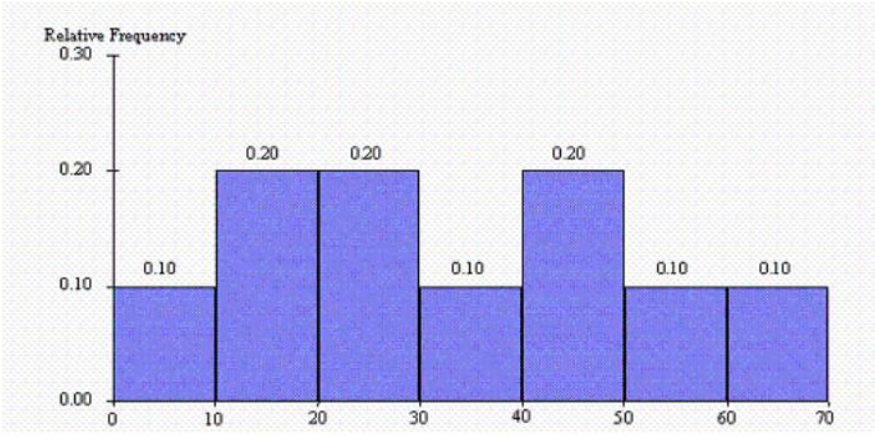
38. Approximately, how large must a male's salary be to qualify as an outlier on the high side?
    a. $60,000 b. $70,000 c. $80,000 d. None of previous

39. Which shape is the distribution of annual salary for males that work at Marko Manufacturing, Inc.
    a. Left Skewed b. Right Skewed c. Symmetric
    d. Cannot tell

40. Which shape is the distribution of annual salary for females that work at Marko Manufacturing, Inc.
    a. Left Skewed b. Right Skewed c. Symmetric
    d. Cannot tell

For questions 28 to 32, answer the following:

41. The histogram below represents scores achieved by 250 job applicants on a personality profile.



    What percentage of the job applicants scored between 30 and 40?
    a. 60% b. 10% c. %90 d. None of previous

42. What percentage of the job applicants scored below 60?
    a. 60% b. 10% c. %90 d. None of previous

43. How many job applicants scored between 10 and 30?
    a. 50 b. 100 c. 125 d. None of previous

44. Seventy percent of the job applicants scored above what value?
    a. 45 b. 40 c. 20 d. None of previous

45. Half of the job applicants scored below what value?
    a. 30 b. 40 c. 35 d. None of previous

Good Luck

Dr. Marwa Sabry