TELFER

Mid-term Exam

ADM2303, Winter 2018
STATISTICS FOR MANAGEMENT I
Date: March 10th, 2018; 9:30-11:00

Professor: Schillo Duration: 90 min

INSTRUCTIONS
1. Books and notes **are not** permitted. **One** sheet of notes, 8.5"x11", is allowed as per instructions in course outline.
2. Calculators **are** permitted.
3. Use the space on the question paper for rough work.

Good Luck !!

GIVEN NAME: _____ FAMILY NAME_____

SECTION: ____ STUDENT #: _____

ANSWER ALL QUESTIONS ON THE SCANTRON SHEET

ALSO PUT YOUR ANSWER AND ALL INTERMEDIATE CALCULATION AND REASONING ON THIS QUESTION BOOKLET.

Statement of Academic Integrity
The School of Management does not condone academic fraud, an act by a student that may result in a false academic evaluation of that student or of another student. Without limiting the generality of this definition, academic fraud occurs when a student commits any of the following offences: plagiarism or cheating of any kind, use of books, notes, mathematical tables, dictionaries or other study aid unless an explicit written note to the contrary appears on the exam, to have in his/her possession cameras, radios (radios with head sets), tape recorders, pagers, cell phones, or any other communication device which has not been previously authorized in writing.

Statement to be signed by the student:
I have read the text on academic integrity and I pledge not to have committed or attempted to commit academic fraud in this examination.

Signed:_____

Note: an examination copy or booklet without that signed statement will not be graded and will receive a final exam grade of zero.

Note: Numbers in brackets, e.g. [2] indicate how many points the question is worth.

Questions 1-3

Suppose a multinational company is interested in obtaining feedback from its employees regarding the working relationships between its employees. For each of the following, identify the sampling method.

Q1 [2] A sample that consists of the entire population is called a

a) Simple sample
b) Comprehensive sample
c) Census
d) Parameter
e) Survey

Q2 [2] There are six categories of employees (administration, software developers, consultants, managers, IT/infrastructure, and other). Randomly select twenty individuals from each category.

a) convenience sampling
b) simple random sampling
c) stratified sampling
d) cluster sampling
e) systematic random sampling

Q3 [2] Each employee has an ID number. Randomly select 150 numbers.

a) convenience sampling
b) simple random sampling
c) stratified sampling
d) cluster sampling
e) systematic random sampling

Q4 [2] Randomly select three countries where the company has a presence and survey all of the employees who work in that country.

a) convenience sampling
b) simple random sampling
c) stratified sampling
d) cluster sampling
e) systematic random sampling

Q5 [2] The multinational company hires an external consultant to conduct interviews with clients. The company specifically asks the consultant to make sure they only share the overall results from the interviews, and no information that would identify respondents. The company also asks the consultant to make sure the results of the study are representative of the company's workforce. The consultant picks the three closest locations for interviews and prepares the report.
Are there ethical concerns with regards to this situation?

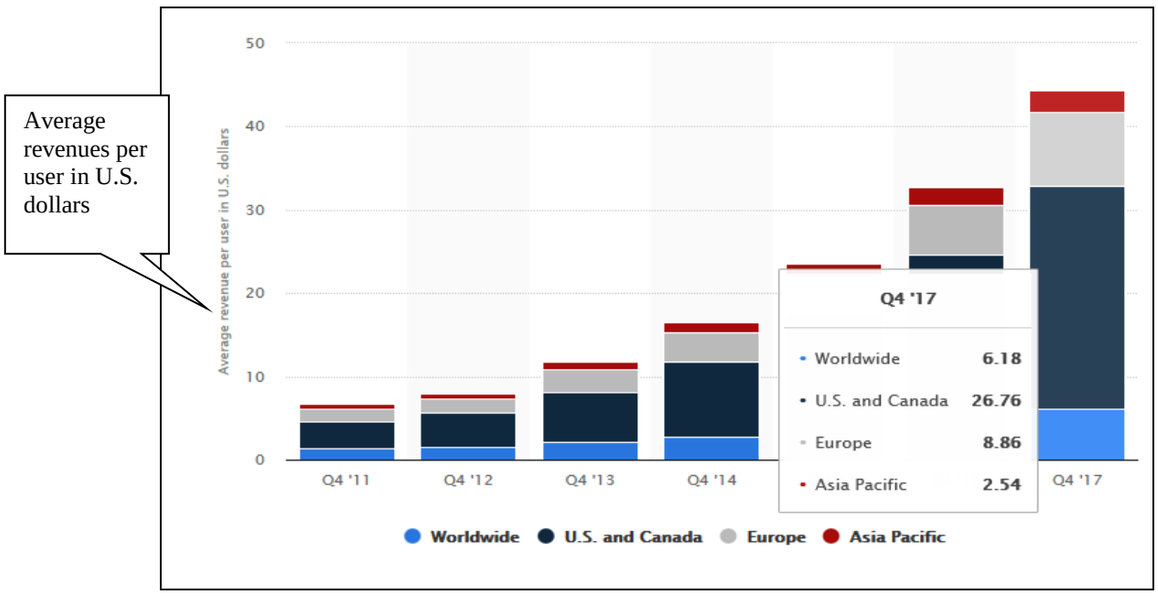
a) Yes, Guideline A: Professionalism is violated
b) Yes, Guideline C: Responsibilities in Publications and Testimony is violated
c) Yes, Guideline H: Responsibilities of Employers is violated
d) Yes, Guidelines C and H are violated
e) No, there are no ethical concerns

Q5 [2] The best way for the consultant to protect the survey from measurement errors is to

a) Use a simple random sample
b) Use a representative sample
c) Conduct a pilot test
d) Conduct a voluntary response survey
e) Nothing can be done to prevent measurement errors

Questions 6-7

The website statista.com presents what I assume to be real figures for Facebook's average revenue per user as of 4th quarter 2017, by region (in U.S. dollars) (https://www.statista.com/statistics/251328/facebooks-average-revenue-per-user-by-region/)



Q6 [2] These data are

a) Qualitative
b) Quantitative – discrete
c) Quantitative – continuous
d) Qualitative and Quantitative – discrete
e) Qualitative and Quantitative – continuous

Q7 [2] Using a stacked bar graph for these particular data is

a) A good choice because it allows to show the proportions of each category
b) A good choice because it shows the differences between the bars
c) A bad choice because the data are averages
d) A bad choice because the data are time series
e) Both c and d

Questions 8 – 9

The Green Tree Lumber Company is considering expanding its product line using a new bio-based composite material called BGW (BioGlueWood). It has the choice of building a small plant that can just produce BGW lumber from 1x3 to 2x12 in size, it can build a larger plant that can also produce 4x8 sheets of BGW, or it can decide to not invest in this opportunity. The company's marketing department develops two scenarios, one for favourable market conditions, and one of unfavourable market conditions (see the table). It considers favourable market conditions more likely, with a probability of 0.6, and unfavourable market conditions less likely, with a probability of 0.4.

Projections of profits

	Favourable market conditions ($)	Unfavourable market conditions ($)
Large plant	200,000	-180,000
Small plant	100,000	-20,000
No plant	0	0

Q8 [4] What is the expected value for profits from the large plant?

a) -$20,000
b) $10,000
c) $20,000
d) $48,000
e) $52,000

Q9 [4] On the basis of the expected value, which option should green tree lumber choose?

a) The large plant, because the expected value is lowest.
b) The large plant, because the expected value is highest.
c) The small plant, because the expected value is lowest.
d) The small plant, because the expected value is highest.
e) It should not invest in this opportunity.

Questions 10-12

Before moving ahead, the CEO of Green Tree Lumber asks the marketing department to conduct a survey to gain greater certainty about the market conditions. Green Tree Lumber has introduced many products over the years and has good information about the reliability of its surveys. It knows that under favourable market conditions, the probability of getting positive results from the survey is 0.7 and under unfavourable market conditions the probability of getting positive results from the survey is 0.2.

Q10 [2] What kind of probabilities are these (the 0.7 and 0.2) ?

a) Marginal probabilities
b) Joint (AND) probabilities
c) Union (OR) probabilities
d) Conditional probabilities
e) Bayes probabilities

Q11 [5] The survey comes back negative. What is the probability that market conditions are unfavourable?

a) Based on the survey, we can be sure the market conditions are unfavourable.
b) 0.40
c) 0.64
d) 0.78
e) 0.80

Q12 [2] What approach did you use to answer the previous question?

a) The answer was obvious from the text
b) AND rules
c) OR rules
d) Bayes
e) A combination of b-d

Questions 13-19

VillageNets is a microfinance NGO (non-governmental organization). They provide financing to micro (very small) businesses. VillageNets assesses the credit risk of loan applicants by assessing the applicant's internet-based social-network profile (to be categorized into one of three categories, A, B, C). Historical records show that borrowers who did not repay their loans (i.e. who defaulted) fall into these categories with the following probabilities: A – 0.1, B – 0.2 and C – 0.7. The same records show that the proportions among borrows who paid back their loans (good loans) was A – 0.6, B – 0.3, and C – 0.1. Records show that 10 percent of Village Nets loan applicants will not pay their loans back (i.e. default).

Q13 [3] What proportion of all applicants will fall into Category C?

a) 0.07
b) 0.16
c) 0.7
d) 0.4
e) 0.09

Q14 [3] What proportion will fall into either Category A or B?

a) 0.91
b) 0.6
c) 0.84
d) 0.3
e) 0.93

Q15 [3] Given that a potential customer is known to fall into category C, what is the probability that they would default on a loan (if loan given).

a) 0.7
b) 0.07
c) 0 because independent
d) 0 because disjoint
e) 0.4375

Q16 [3] Consider a group of 5 Category C customers. What is the approximate probability that none of them default?

a) 0.7
b) 0.0024
c) 0.0563
d) 0.0160
e) 0 because disjoint

Q17 [3] The events of falling into Category C and defaulting on the loan are

a) Disjoint and independent
b) Disjoint and dependent
c) Not disjoint and independent
d) Not disjoint and dependent
e) It is not possible to determine this with the information provided

Questions 18- 21

Statistics Canada (http:www.statcan.ca) provides data about the average number of days lost per worker by cause of each province. Listed below are some summary statistics for the distributions of days lost for all provinces because of "Personal or Family Responsibility" (P or F) and "Illness or Disability" (I or D) for 2004.

	Days lost	
Sample statistics	**P or F**	**I or D**
Mean	1.62	7.98
Median	1.75	8
Mode	1.9	8
Standard deviation	0.25	1.31
Interquartile range	0.4	1.9

Q18 [3] Comment on the symmetry of these two distributions.

a) Both distributions are probably approximately symmetric.
b) 'P or F' is probably approximately symmetric; 'I or D' is not symmetric.
c) 'P or F' is not symmetric; 'I or D' is probably approximately symmetric.
d) Neither distribution is symmetric.
e) This information does not allow us to tell.

Q19 [3] For the Personal or Family Responsibility distribution, which would be a more appropriate summary of the centre, the mean, the median, or the mode?

a) The mean would be most appropriate.
b) The median would be most appropriate.
c) The mode would be most appropriate.
d) Mean or mode are equally appropriate.
e) Mean or median are equally appropriate.

Q20 [4] For the Illness or Disability distribution, use the following summary statistics: What can we say about outliers?

Minimum	Maximum	Lower quartile, Q1	Upper quartile, Q3
5.7	9.9	6.9	8.8

a) There is at least one outlier at the high end.
b) There is at least one outlier at the low end.
c) There is at least one outlier at each ends.
d) There are no outliers.
e) It is impossible to say anything about outliers.

Q21 [3] Which distribution ('P or F' or 'I or D') has the largest relative spread (coefficient of variation)? Justify your answer with appropriate statistical measure calculations.

a) 'P or F'
b) 'I or D'
c) Both are approximately equal
d) This cannot be calculated
e) Not applicable

Q22 [3] Suppose that in 2010, data values for days lost because of Personal or Family Responsibility reason have increased by 20% compared to 2004 data values. By what % do the statistics below increase?

a) The interquartile range increases by 20%, the mean does not change
b) The interquartile range increases by 20%, the mean increases by 20%
c) The interquartile range does not change, the mean increases by 20%
d) This cannot be calculated
e) None of the above

Questions 23 - 25

As a commercial property real-estate broker your annual revenue stream depends on the number of properties sold. The expected number of sales is 7 sales/year, but the number of sales varies, having a standard deviation of 2 sales/year. Your annual costs are also variable; having an expected value of $20,000/year and a standard deviation of $5,000/year. Your costs are correlated with the number of sales per year with a correlation coefficient of 0.3. Assume that the commission per sale is fixed at $20,000/sale (all dollar amounts are in Canadian dollars).

Compute the following for your net (before tax) revenue (= commission * sales - costs):

Q23 [3] The expected value is

a) 7
b) $120,000
c) $137,000
d) $140,000
e) $160,000

Q24 [5] The variance is

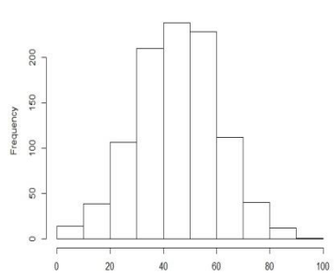
a) 4
b) 1,505,000,000 $Cdn^2
c) 1,575,000,000 $Cdn^2
d) 1,625,000,000 $Cdn^2
e) 1,745,000,000 $Cdn^2

Q25 [3] If the $ amounts had to be expressed in US dollars (note 1CN$ = 0.8US$) what would be the new values for Variance (Var) and Coefficient of Variation (CV)?

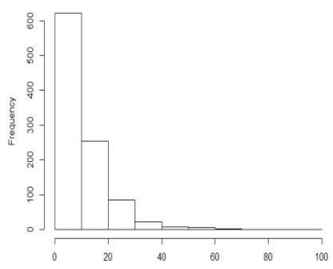
a) Variance unchanged; Coefficient of Variation unchanged
b) Variance = 0.8 * Var; Coefficient of Variation unchanged
c) Variance = 0.8^2 * Var; Coefficient of Variation unchanged
d) Variance = 0.8 * Var; Coefficient of Variation = 0.8*CV
e) Variance = 0.8^2 * Var; Coefficient of Variation=0.8*CV

Questions 26 – 27

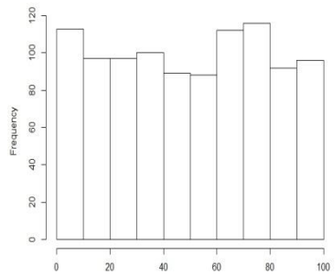
Q26 [2] Match the histograms to boxplots.



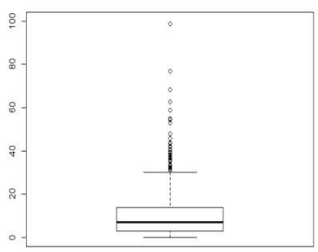
a)



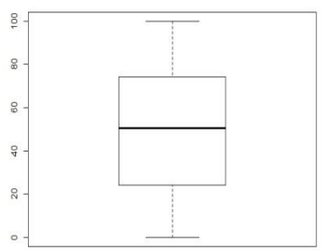
b)



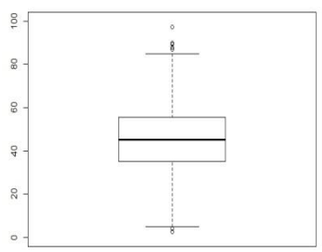
c)



1)



2)



3)

a) (a) to (1); (b) to (2); (c) to (1)
b) (a) to (2); (b) to (1); (c) to (3)
c) (a) to (2); (b) to (3); (c) to (1)
d) (a) to (3); (b) to (2); (c) to (1)
e) (a) to (3); (b) to (1); (c) to (2)

Q27 [2] One might choose a stem-and-leaf display rather than a boxplot because it …

a) Reveals the shape of a distribution
b) Is better for large data sets
c) Displays the actual data values
d) a, b, and c
e) the boxplot is always better than a stem-and-leaf display

Questions 28 – 30

IKEA sells a broad range of household products and offers a low-cost breakfast in its store restaurants to attract customers into the stores early in the day. Suppose you are the manager of a store like IKEA and are making a loss on the breakfasts. You need to know whether the breakfast attracts customers to shop for the more profitable household items, and you decide to calculate the correlation between the number of customers taking breakfast and the number buying household items.

The following data are collected within one week:

Day	**Number of Customers at Breakfast**	**Number of Customers Buying Household Items**
Monday	73	54
Tuesday	52	45
Wednesday	56	26
Thursday	81	57
Friday	83	34

Q28 [2] Are the conditions for calculating correlations met?

a) No, the variables are not quantitative
b) No, the relationship is not linear
c) No, there are outliers
d) b and c
e) Yes, the conditions are met

Q29 [5] Just of this exam, please ignore whether the conditions are met or not and calculate the correlation coefficient. Which of the following numbers is closest?

a) 0.2993
b) 0.3377
c) 0.3742
d) 0.4677
e) 0.7531