STAT 2507 and BIT 2000 A
SOLUTIONS TO MIDTERM TEST (Version 2)
Fall 2014

1. Which of the following statements is always true?

   (a) Discrete variable can take on only a finite number of possible values.
   (b) Continuous variable can take on a finite number of possible values.
   [(c)] Both discrete and continuous variables can take on an infinite number of possible values.
   (d) None of the above statements are true.

   Solution: Both discrete and continuous variables can take on an infinite number of possible values, and the set of possible values of a continuous variable cannot be finite. Therefore the correct answer is (c).

2. From a random sample of 1,000 high school students, sample mean and sample standard deviation of their systolic blood pressures were found to be $\bar{x} = 110$ mmHg and $s = 10$ mmHg. Which of the following would always be true about the sample?

   (a) At least 250 students have their systolic blood pressures outside the range 90 mmHg – 130 mmHg.
   [(b)] At most 250 students have their systolic blood pressures outside the range 90 mmHg – 130 mmHg.
   (c) At least 125 students have their systolic blood pressures higher than 130 mmHg.
   (d) At most 125 students have their systolic blood pressures higher than 130 mmHg.

   Solution: According to Chebyshev's inequality, at most 25% of sample data fall outside of the internal $\bar{x} \pm 2s$. In this problem, $(\bar{x} - 2s, \bar{x} + 2s) = (90, 130)$. Therefore the correct answer is (b).

3. The estimated slope in a fitted least-squares line of Y on X is $b = 0.85$. If the sample variance of X is greater than that of Y, how would you describe the sample correlation coefficient between X and Y?

   (a) The sample correlation coefficient is the same as 0.85.
   (b) The sample correlation coefficient is greater than 0 but less than 0.85.
   [(c)] The sample correlation coefficient is greater than 0.85 but less than or equal to 1.
   (d) The sample correlation coefficient is not related to the estimated slope.

   Solution: We know that the slope b of the least-squares line satisfies $b = r\dfrac{s_y}{s_x}$, and hence $r = b\dfrac{s_x}{s_y}$. Since $0 < s_y = \sqrt{s_y^2} < \sqrt{s_x^2} = s_x$, it follows that $r > b = 0.85$. At the same time, $r \leq 1$ by the property of a correlation coefficient.

4. Which of the following diagrams or measures loses the largest amount of information when summarizing and/or describing the data?

   [(a)] Median **(b)** Boxplot **(c)** Histogram **(d)** Stem-and-leaf plot

   Solution: Boxplot, histogram, and stem-and-leaf plot display the distribution of a data set, whereas the median, a single numerical measure, does not.

5. Suppose you have a data set that contains a small fraction of the observations that are very large. Which of the following numerical measures would be most appropriate for indicating the centre of the data set?

   (a) Sample mean **[(b)]** Sample median
   (c) Largest observation **(d)** None of the above

   Solution: The sample mean and the largest observation are too much sensitive to the presence of unusually large observations. At the same time, the sample median is robust to the presence of outliers. Therefore, the sample median would be most appropriate as a measure of centre.

6. Consider the following stem-and-leaf plot of 23 observations.

   Leaf unit=1

   0 | 012

   1 | 2248

   2 | 1134599

   3 | 00036

   4 | 457

   5 | 1

   The lower quartile, Q_1, and the median are, respectively:

   [(a)] 14 and 25 **(b)** 13.5 and 24.5 **(c)** 12 and 24 **(d)** 12 and 29

   Solution: The position of the median is $0.5(n+1) = 0.5(24) = 12$, and the 12th element in the ordered data set is 25. The position of the lower quartile is $0.25(n+1) = 0.25(24) = 6$, and the 6th element in the ordered data set is 14.

7. Which of the following randomly selected measurements, X, may be considered to be an outlier if it was selected from the given population?

   (a) $X = 10$ from a population with $\mu = 3$ and $\sigma^2 = 16$.
   (b) $X = -1$ from a population with $\mu = 1$ and $\sigma^2 = 25$.
   [(c)] $X = 3/2$ from a population with $\mu = -3$ and $\sigma^2 = 1$.
   (d) $X = 0$ from a population with $\mu = -2$ and $\sigma^2 = 3$.

   Solution: We know that observations with z-scores equal to or greater than 3 are considered very unlikely. The only observation among (a)–(d) that has the z-score in excess of 3 is $X = 3/2$.

8. If the interval $\bar{x} \pm 2s$ is equal to (3,7) then the interval $\bar{x} \pm s$ is

   (a) (4.5,5.5) **[(b)]** (4,6) **(c)** (3.5,6.5) **(d)** Impossible to compute

   Solution: By the statement of the problem, $\bar{x} - 2s = 3$ and $\bar{x} + 2s = 7$. Thus we have two linear equations with two unknowns. Solving this equations yields $\bar{x} = 5$ and $s = 1$. Therefore, $\bar{x} - s = 5 - 1 = 4$ and $\bar{x} + 2s = 5 + 1 = 6$.

9. An interior decorator must furnish two offices that already have desk and chair. Each office needs 1 file cabinet and 2 bookcases. At a local office furniture store, there are 6 models of file cabinets, and 7 models of bookcases all of which are compatible. How many choices does the decorator have if he wants to select 2 file cabinets and 4 bookcases but he does not want to select more than one of any model?

   (a) 525 **(b)** 315 **(c)** 42 **(d)** 252

   Solution: There are C_2^6 different ways in which 2 models of file cabinets can be selected by the decorator from the 6 models available, and, regardless of which particular 2 models of file cabinets have been chosen by the decorator, there are C_4^7 different ways in which 4 models of bookcases can be selected by the decorator from the 7 models available. Therefore, by the Multiplication Rule, the total number of choices that the decorator can have is $C_2^6 \times C_4^7 = 525$.

10. A probability distribution of a discrete random variable X is partially given in the following table, with the additional information that $p(1) = 3p(5)$. Determine the missing entries in the table.

x	0	1	2	3	4	5
$p(x)$	0.15	?	0.10	0.15	0.20	?

   (a) $p(1) = 0.05, \quad p(5) = 0.15$
   (b) $p(1) = 0.15, \quad p(5) = 0.05$
   (c) $p(1) = 0.30, \quad p(5) = 0.10$
   (d) $p(1) = 0.10, \quad p(5) = 0.30$

   Solution: By the normalization property of a pmf,

$$p(0) + p(1) + p(2) + p(3) + p(4) + p(5) = 1.$$

   Therefore, using the fact that $p(1) = 3p(5)$, we obtain

$$0.15 + 3p(5) + 0.10 + 0.15 + 0.20 + p(5) = 1.$$

   This gives $4p(5) = 0.40$, and hence $p(5) = 0.10$ and $p(1) = 3p(5) = 0.30$.

11. In a certain part of downtown Ottawa, a car that is illegally parked on a street will be fined \$25 if caught, and the chance of being caught is 60%. What is the expected fine for person who parked on this street?

   (a) \$12.00 **(b)** \$10.00 **(c)** \$18.00 **(d)** \$15.00

   Solution: Let X be the cost of fine. Then X takes on only two values: 25 with probability 0.6 and 0 with probability 0.4. Therefore $E(X) = 25(0.6) + 0(0.4) = 15$.

12. When the price of gasoline gets high, consumers become very concerned about the gas mileage obtained by their cars. One consumer was interested in the relationship between car engine size (number of cylinders) and gas mileage (litres/100 km). The consumer took a random sample of 7 cars and recorded the following information:

$$\sum_{i=1}^{7} x_i = 45, \quad \sum_{i=1}^{7} y_i = 72, \quad \sum_{i=1}^{7} x_i y_i = 510, \quad s_x = 1.899, \quad s_y = 5.628.$$

Fit the least-squares line relating car engine size, x, and fuel efficiency, y, and find the predicted fuel efficiency for a car with a 6-cylinder engine. Round all intermediate numbers using 3 decimal places.

(a) 20.12 litres per 100 km **(b)** 9.35 litres per 100 km
(c) 11.71 litres per 100 km **(d)** 16.25 litres per 100 km

Solution: The least-squares line is given by the formula

$$y = a + bx, \quad \text{where} \quad b = \frac{s_{xy}}{s_x^2} \quad a = \bar{y} - b\bar{x},$$

and

$$s_{xy} = \frac{\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)}{(n-1)} = \frac{510 - (45)(72)/7}{7-1} = 7.857,$$

$$b = \frac{s_{xy}}{s_x^2} = \frac{7.857}{(1.899)^2} = \frac{7.857}{3.606} = 2.179,$$

$$a = \bar{y} - b\bar{x} = (72)/7 - (2.179)(45)/7 = -3.723.$$

Therefore the least-squares line is given by

$$y = -3.723 + 2.179x,$$

and the predicted value of fuel efficiency for a car with a 6-cylinder engine is

$$-3.723 + 2.179(6) = 9.351 \approx 9.35 \text{ litres per km}.$$

13. Event A occurs with probability 0.7 and event B occurs with probability 0.4. If we know that the event A has occurred, the probability that B will occur is 0.4. If we know that event B has occurred then what is the probability that A will occur, i.e., what is $P(A|B)$?

(a) 0.66 **(b)** 0.28 **(c)** 0.70 **(d)** 0.40

Solution: Using the Multiplication Rule for conditional probability, we obtain

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{(0.4)(0.7)}{(0.4)} = 0.7.$$

14. Dan walks to work 10% of days and the rest of the days he takes the bus to work. 45% of the days that he walks to work he arrives late while he arrives late only 5% of the days he takes the bus. What is the probability that Dan arrives late in a day?

(a) 0.121 **(b)** 0.131 **(c)** 0.090 **(d)** 0.190

Solution: Let W and B be the events that Dan walks to work today and takes a bus to work today, respectively. Let L be the event that Dan arrives late today. By assumption

$$P(W) = 0.1, \quad P(B) = 0.9, \quad P(L|W) = 0.45, \quad P(L|B) = 0.05.$$

We need to find $P(L)$. Using the Law of Total Probability, we get

$$P(L) = P(L|W)P(W) + P(L|B)P(B) = (0.45)(0.1) + (0.05)(0.9) = 0.09.$$

15. Under the conditions of problem 14, if we know that Dan was late today, what is the probability that he walked to work?

(a) 0.70 **(b)** 0.75 **[(c)]** 0.50 **(d)** 0.64

Solution: Using the notation and result of problem 14 and applying Bayes' Theorem, we get that the required probability is equal to

$$P(W|L) = \frac{P(L|W)P(W)}{P(L)} = \frac{(0.45)(0.1)}{0.09} = \frac{0.045}{0.09} = 0.5.$$

16. A Professor gives her students 7 questions a week before an exam and announces that 4 of these questions will be chosen for the exam. If Sarah, a student in the class, knows the solutions to 5 of the questions, what is the probability that Sarah will receive a perfect mark on her exam next week.

(a) 1/4 **[(b)]** 1/7 **(c)** 1/8 **(d)** 1/14

Solution: Using the classical definition of probability,

$$P(\text{Sarah will receive a perfect mark}) = \frac{\text{number of favourable outcomes}}{\text{number of possible outcomes}} = \frac{C_4^5}{C_4^7} = \frac{\frac{5!}{4!1!}}{\frac{7!}{4!3!}} = \frac{1}{7}.$$