STAT 2507D
MIDTERM TEST-Version 1
Fall 2017

INSTRUCTIONS: The test consists of 16 multiple-choice questions. Please use **scantron sheet** to answer the questions. The total of marks for questions is 40; each question has 2.5 marks. This is a **80**-minute **closed-book** test. You are allowed to use **non-programmable** calculator. **Formula sheet** and a blank page for rough calculations are provided.

1. Given $P(A) = 0.4,\ P(B) = 0.5$ and $P(A \cap B) = 0.2$. What is $P(A|B')$?

   (a) $P(B)$ **(b)** $P(A)$ **(c)** $P(B')$ **(d)** $P(A')$
   (b)

2. Among the 24 invoices prepared by billing department, 4 contain errors while the others do not. If we randomly check 2 of these invoices, what is the probability that neither will contain an error?

   (a) $\frac{5}{6}$ **(b)** $\frac{95}{120}$ **(c)** $\frac{95}{138}$ **(d)** $\frac{19}{23}$
   (c)

3. Evidence shows that the chance of a driver will be involved in a serious accident during a given year is 0.01. We take a sample of 100 drivers, find the probability that exactly two drivers will be involved in a serious accident during the coming year?

   (a) $\frac{e^{-2}}{2}$ **(b)** $\frac{e^{-1}}{2}$ **(c)** e^{-2} **(d)** e^{-1}
   (b)

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

5. Let X be a random variable with $E(X) = 3$ and $E(X^2) = 15$. Let $h(X) = 3X - 4$, what are the values of $E(h(X))$ and $\text{Var}(h(X))$?

   (a) 9 and 45 **(b)** 5 and 54 **(c)** 9 and 14 **(d)** 5 and 48
   (b)

6. If the interval $(\mu - 2\sigma,\ \mu + 2\sigma)$ is equal to (1, 5), then the interval $(\mu - \sigma,\ \mu + \sigma)$ must be

   (a) (2,5) **(b)** (5,2) **(c)** (2,4) **(d)** Impossible to compute
   (c)

7. Which of the following is/are not correct about the normal distribution?

   (i) The total area under the curve is 1.
   (ii) The curve is symmetric about the mean.
   (iii) The value of the mean is always greater than the value of the standard deviation.
   (iv) About 68% of the probability fall within 1 standard deviation of the mean.

   (a) Only (ii) is not correct. **(b)** Only (iii) is not correct.
   (c) Only (iv) is not correct. **(d)** Only (iii) and (iv) are not correct.
   (b)

8. A Professor gives her students 8 questions a week before an exam and announces that 5 of these questions will be chosen for the exam. If Sarah, a student in the class, knows the solutions to 6 of the questions, then what is the probability that Sarah will receive a perfect mark on her exam next week?

   (a) 5/6 **(b)** 6/8 **(c)** 5/56 **(d)** 6/56
   (d)

9. Dan walks to work 20% of days and the rest of the days he takes the bus to work. 40% of the days that he walks to work he arrives late while he arrives late only 5% of the days he takes the bus. If we know that Dan was late today, what is the probability that he walked to work?

   (a) 2/3 **(b)** 1/3 **(c)** 2/12 **(d)** 1/12
   (a)

10. Assume that the annual salaries of part-time bank tellers is normally distributed with a mean of \$22,500 and a standard deviation of \$1,000. Ten percent of the tellers have annual salary of greater than what value?

    (a) 22372 **(b)** 21220 **(c)** 23780 **(d)** 22628
    (c)

11. 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 upper quartile, Q_3, and the median are, respectively:

    (a) 31.5 and 24.5 **(b)** 36 and 24 **(c)** 30.75 and 25 **(d)** 33 and 25
    (d)

12. Suppose you have a data set that contains a small fraction of the observations that are very large or very small. Which of the following numerical measures would be most appropriate for describing variation in the data set?

(a) Sample standard deviation **(b)** Sample range
(c) Sample interquartile range **(d)** All of the above
(c)

13. 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 per 100 km). The consumer took a random sample of 7 cars and recorded the following information:

$$\sum_{i=1}^{7} x_i = 40, \quad \sum_{i=1}^{7} y_i = 77, \quad s_{xy} = 8, \quad s_x = 1.799, \quad s_y = 4.583.$$

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

14. 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
$p(x)$	0.20	?	0.25	0.15	0.20	?

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

15. Given two points (2 ,3) and (3.5, b), where $b < 2.3$, the value of the correlation coefficient

(a) is always $+1$,

(b) is always -1,

(c) is always either $+1$ or -1, depending on the value of b,

(d) can be any number in the range -1 and $+1$, depending on the values of b.
(b)

16. Mass Index (BMI) is normal for specific gender and age groups. For females aged 30-39, the mean BMI is 24.5 with standard deviation of 3.3. What proportion of females aged 30-39 has a BMI over 30?

(a) 0.0475 **(b)** 0.9525 **(c)** 0.025 **(d)** 0.6254]
(a)