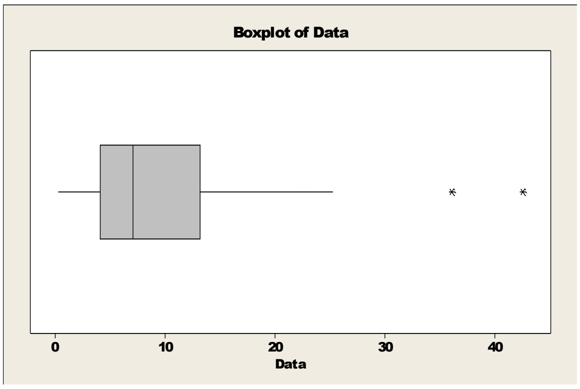
STAT 2507 and BIT 2000 A
MIDTERM TEST
Fall 2012

Name	Student number	Section

INSTRUCTIONS: The test consists of 10 multiple-choice questions and 5 long-answer questions. The total of marks for questions is 50. Marks for individual questions are given in [ ]. This is a **closed-book** test. You are allowed to use **non-programmable** calculator. **Formula sheet**, **Poisson table**, and a blank page for rough calculations are attached to the test paper.

Multiple-Choice Questions. Please circle one answer only.

The first five questions are based on the following boxplot.



[2] 1. Which one of the following statements about the distribution above is correct?

(a) Symmetric and has no outlier.
(b) Skewed to the left and has outliers(s).
(c) Skewed to the right and has outlier(s).
(d) Symmetric but the mean and the median are different.

[2] 2. Which one of the following statements is false?

(a) The mean is larger than the median.
(b) The median is approximately equal 7.
(c) The maximum value is approximately 43.
(d) The maximum value is approximately 25.

[2] 3. Let X be a random variable with $E(X) = 4$ and $E(X^2) = 21$. Let $h(X) = 2X - 6$. What is the value of $E(h(X))$?

(a) 36 **(b)** 20 **(c)** 65 **(d)** 2

[2] 4. Refer to the previous question. What is the value of $\text{Var}(h(X))$?

(a) 6 **(b)** 34 **(c)** 20 **(d)** 30

[2] 5. You are told that the equation of the least-squares regression line for predicting y from x based on $n = 10$ pairs of observations is $y = 1.73 - 2x$. If $s_x^2 = 1$ and $s_y^2 = 16$ then the equation of the least-squares regression line for predicting x from y, based on the same $n = 10$ pairs of observations is $x = a + by$, where the slope b equals

(a) $-1/16$ **[(b)]** $-1/8$ **(c)** $-3/16$ **(d)** $-1/4$

[2] 6. Suppose you roll a pair of fair dice. What is the probability that the number of dots on the two dice sum to either 5 or 10?

(a) 5/36   **[(b)]** 7/36   **(c)** 11/36   **(d)** 4/36

[2] 7. Suppose that $P(A) = 0.4$, $P(B) = 0.3$, and $P(A \cap B) = 0$. Which one of the following statements correctly defines the relationship between the events A and B?

(a) The events A and B are independent, but not mutually exclusive.
[(b)] The events A and B are mutually exclusive, but not independent.
(c) The events A and B are neither mutually exclusive nor independent.
(d) The events A and B are both mutually exclusive and independent.

[2] 8. A telephone survey of Canadian families is conducted to determine the number of children in the average Canadian family. Past experience has shown that 30% of the families who are telephoned will refuse to respond to the survey. Which of the following is a binomial random variable?

[(a)] The number of families out of 50 who respond to the survey.
(b) The percentage of families which refuse to respond to the survey.
(c) The number of children in a family which responds to the survey.
(d) The age of the first children in a family which responds to the survey.

[2] 9. According to a survey of adults, 64% have money in a regular savings account. If we plan on surveying 50 randomly selected adults, find the mean number of adults who would have regular savings accounts.

(a) 12   **(b)** 22   **[(c)]** 32   **(d)** 42

[2] 10. You would rather use a hypergeometric probability distribution than a binomial distribution in case of sampling

(a) with replacement from a finite population;
[(b)] without replacement from a finite population of size N when the size N is small in relation to the sample size n; namely, $n/N \geq 0.05$;
(c) without replacement from an infinite population;
(d) with replacement from an infinite population.

Long-Answer Questions

[6] 11. Suppose an experiment involving five subjects is conducted to determine the relationship between the percentage of a certain drug in the bloodstream and the length of time it takes to react to a stimulus. The results are shown in the table below.

Reaction Time versus Drug Percentage

Subject	Percent x of Drug	Reaction Time y (seconds)
1	1	1
2	2	1
3	3	2
4	4	2
5	5	4

Using the fact that $\sum x_i = 15$, $\sum y_i = 10$, $\sum x_i^2 = 55$, $\sum y_i^2 = 26$, and $\sum x_i y_i = 37$, find the correlation coefficient for the reaction time and the amount of drug in the bloodstream. What is the relationship (if any) between the reaction time and the amount of drug in the bloodstream?

Solution: Using the computing formulas for the sample covariance and the sample variance,

$$\begin{aligned} s_{xy} &= \sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right) = 37 - \frac{(15)(10)}{5} = 37 - 30 = 7, \\ s_x^2 &= \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 55 - \frac{(15)^2}{5} = 55 - 45 = 10, \\ s_y^2 &= \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 26 - \frac{(10)^2}{5} = 26 - 20 = 6. \end{aligned}$$

Therefore

$$r = \frac{s_{xy}}{s_x s_y} = \frac{7}{\sqrt{(10)(6)}} = 0.904.$$

The fact that r is positive and near 1 indicates that there is a strong positive linear relationship between x and y. In other words, the reaction time tends to increase as the amount of drug in the bloodstream increases.

[6] 12. Thirty students in a experimental psychology class use various techniques to train a rat to move through a maze. At the end of the course, each student's rat is timed as it negotiates the maze. The sample mean and the sample standard deviation of times of 30 rats running through a maze (in minutes) were found to be $\bar{x} = 3.74$ and $s = 2.20$. Using Chebyshev's inequality, at least how many of the 30 running times will fall in the interval $(-0.66, 8.14)$?

Solution: Observe that $\bar{x} - 2s = -0.66$ and $\bar{x} + 2s = 8.14$. Therefore by Chebyshev's inequality at least $1 - 1/2^2 = 3/4$ of the measurements will be between -0.66 and 8.14. Since $3/4 \times 30 \approx 22.5$, it follows that at least 22 of the running times will fall within the interval $(-0.66, 8.14)$.

[6] 13. If 5% of men and 0.25% of women are colour blind, what is the probability that a randomly selected person is colour blind? We assume that it is equally likely that a selected person will be a man or a woman.

Solution: Let us introduce the following events:

$$\begin{aligned} M &= \{\text{selected person is a man}\}, \\ W &= \{\text{selected person is a women}\}, \\ C &= \{\text{selected person is colour blind}\}. \end{aligned}$$

By assumption,

$$\mathbf{P}(C|M) = 0.05, \quad \mathbf{P}(C|W) = 0.0025, \quad \mathbf{P}(M) = \mathbf{P}(W) = 0.5.$$

Therefore, using the Law of Total Probability,

$$\mathbf{P}(C) = \mathbf{P}(M)\mathbf{P}(C|M) + \mathbf{P}(W)\mathbf{P}(C|W) = 0.5 \times 0.05 + 0.5 \times 0.0025 = 0.026.$$

The result is rounded using three decimal places.

[6] 14. Suppose that in a large population the proportion of people that have a certain disease is 0.02. Determine the approximate probability that in a random group of 100 people at least four people will have the disease.

Solution: We can assume that the exact distribution of the number of people having the disease among the 100 people in the random group is a binomial distribution with parameters $n = 100$ and $p = 0.02$. Therefore this distribution can be approximated by a Poisson distribution for which the mean is $\mu = np = 2$. If $X \sim Bin(100, 0.02)$ and $Y \sim Poi(2)$, then it can be found from the table of the Poisson distribution that

$$\mathbf{P}(X \geq 4) \approx \mathbf{P}(Y \geq 4) = 1 - \mathbf{P}(Y \leq 3) = 1 - 0.857 = 0.143.$$

[6] 15. A salesman of small-business computer systems will contact three customers during a week. Each contact can result in either a sale, with probability 0.3, or no sale, with probability 0.7. Assume that customer contacts are independent. If X denotes the number of computer systems sold during the week, find $P(X \geq 2)$.

Solution: The sample space S consists of 2^3 elementary outcomes:

$$S = \{sss, ssn, sns, nss, snn, nsn, nns, nnn\},$$

where we use s for 'sale' and n for 'no sale'. Therefore X can take on four values 0,1,2, and 3 with the respective probabilities

$$p(0) = \mathbf{P}(X = 0) = (0.7)^3 = 0.343, \quad p(1) = \mathbf{P}(X = 1) = \binom{3}{1}(0.3)(0.7)^2 = 0.441,$$
$$p(2) = \mathbf{P}(X = 2) = \binom{3}{2}(0.3)^2(0.7) = 0.189, \quad p(3) = \mathbf{P}(X = 3) = (0.3)^3 = 0.027.$$

From this,

$$\mathbf{P}(X \geq 2) = p(2) + p(3) = 0.189 + 0.027 = 0.216.$$