

1. The following table gives the probability distribution of the number of people per car crossing a toll bridge:

Number of people (including the driver)	1	2	3	4	5	6
Probability	0.05	0.43	0.27	0.12	0.09	0.04

- (a) (2 marks) Let X be the number of people in the car including the driver. Compute the mean and standard deviation of X .
- (b) (2 marks) What is the expected revenue per car if you charge \$0.30 per person in the car? What is the standard deviation of the revenue per car?
- (c) (2 marks) What is the expected revenue per car if you charge \$0.50 for the car and driver together and \$0.25 for each passenger? What is the standard deviation of the revenue per car?
- (d) (1 mark) Would more money be raised by charging \$1.00 per car or by charging \$0.30 per person, or by charging \$0.50 for the car and driver and \$0.25 for each passenger?
- (e) (2 marks) Suppose that bridge A charges according to the method given in (b) and that bridge B charges according to the method given in (c). Suppose also that the traffic patterns and the operations in the two bridges are independent of each other. What is the standard deviation of the total revenue per car for the two bridges?
- (f) (1 mark) What is the standard deviation for the total revenue obtained from 8 cars using bridge B?
2. Monthly food expenditure for families of three in a large city follows a distribution with mean \$420 and standard deviation \$80. Assume that the monthly food expenditure is normally distributed.
- (a) (2 marks) What is the probability that a randomly selected family of three from this city has monthly food expenditures less than \$352?
- (b) (2 marks) Nine families of three from this city are randomly selected. What is the probability that the average monthly food expenditures of these nine families is less than \$352?
- (c) (2 marks) Five families of three from this city are randomly selected. Use the result in (a) to compute the probability that at most two of these five families have monthly food expenditures less than \$352?

3. Let

$$f(x) = \begin{cases} cx & 0 < x \leq 1 \\ c(3-x)/2 & 1 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

- (a) (1 mark) Sketch the function $f(x)$.
- (b) (2 marks) Find c such that $f(x)$ is a probability distribution.
- (c) (2 marks) Calculate $P(0.5 < X < 1.5)$.
4. It is known from examining a woman's family tree that she has a 50% chance of being a carrier of the hereditary disease hemophilia. If she is a carrier, there is a 50% chance that any particular son will inherit the disease, with different sons being independent of one another. If she is not a carrier, no son can inherit the disease.
- (a) (2 marks) What is the probability that her first son did not inherit the disease?
- (b) (1 mark) What is the probability that her second son did not inherit the disease?
- (c) (3 marks) If neither of her two sons inherited the disease, what is the probability that she is a carrier of the disease?

5. Use the continuity correction whenever applicable, in this problem

Consider an acceptance-sampling scheme in which a factory takes delivery of a batch of components if a random sample of 400 components contains fewer than M defectives. Otherwise the batch is returned to the supplier. Let \hat{p} be the sample proportion of defectives.

- (a) (3 marks) Suppose that the batch has 5% defectives. What is the value of c for which

$$P(\hat{p} > c) = .10?$$

- (b) (1 mark) What value should M have?
- (c) (4 marks) Suppose that the supplier makes a batch in which the true proportion of defectives is only 2%. If the cutoff value M computed in (b) is adopted, what is the probability that the batch will be sent back?

6. The following table lists the number of employees (in hundreds) and the profit (in million dollars) for 16 publishing firms in 1990.

Firm	1	2	3	4	5	6	7	8
Employees	9.4	6.3	10.7	7.4	17.1	21.2	36.8	28.5
Profit	33.5	31.4	25.0	23.1	14.2	11.7	7.8	10.5

Firm	9	10	11	12	13	14	15	16
Employees	10.7	9.9	26.1	70.5	14.8	21.3	14.6	26.8
Profit	9.8	9.1	8.5	8.3	4.8	3.2	2.7	-9.5

A business student wants to examine if there is a relationship between the profit and the number of employees. Given below is a *MINITAB* printout of the descriptive statistics and the correlation coefficient of the two variables.

Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Employee	16	20.76	15.95	18.24	15.88	3.97
Profit	16	12.13	9.45	12.14	11.17	2.79

Variable	Minimum	Maximum	Q1	Q3
Employee	6.30	70.50	10.10	26.62
Profit	-9.50	33.50	5.55	20.80

Correlation of Employees and Profit = -0.404

- (2 mark) Using the $1.5 \times IQR$ criterion, determine whether the number of employees of the 12th firm is a possible outlier.
- (2 marks) Determine the regression equation of Profit on the number of employees.
- (2 marks) What is the absolute value of the difference in the expected profit between firms that differ by 5,000 employees.
- (2 marks) Give the residual from the regression for the 12th firm.
- (1 mark) If the profit figures were given in thousands of dollars, what would happen to the correlation coefficient?
- (2 marks) What percentage of the total variation in the profit is accounted for by the regression on the number of employees? In your view is the number of employees a good explanatory variable of the profit?

7. The stemplot below shows the weight X (in lbs), of 27 ground beef packages in a supermarket meat display:

Stem-and-leaf of Weights N = 27

Leaf Unit = 0.01

```

1      7 5
6      8 37999
(8)    9 13666789
13     10 3688
9      11 04578
4      12 48
2      13 8
1      14 1

```

- (3 marks) Calculate the mean \bar{x} in lbs.
- (2 marks) Find the first and third quartiles, Q_1 and Q_3 respectively, of the data.
- (1 mark) Without calculating the median M , say whether M is closer to Q_1 or closer to Q_3 , or

midway between the two? Why?

8. A bank interested in expanding its customer base requested a market survey. Customers were asked to rate overall service on a 3-point scale from 1 (Poor), 2 (Average), 3 (Good)

The results are summarized in the following table, where we have also included the age of the respondents.

	Good	Average	Poor	Total
Under 40	266	32	24	322
40 and over	277	52	13	342
Total	543	84	37	664

- (a) (2 marks) Draw side by side the bar graphs (in percentages) for the conditional distributions of the response given age.
- (b) (2 marks) Are the two variables (response and age) independent? Why?

9. **In this problem do not use the continuity correction**

In this problem, we study the sampling behaviour of a coin-tossing experiment in which a fair coin (i.e. $P(\text{Head}) = .5$) is tossed an increasing number n of times.

- (a) (2 marks) To which value does the sample proportion \hat{p} of heads in the n tosses get closer and closer to, as n increases? Justify your answer.
- (b) (2 marks) How big does n have to be so that the probability that \hat{p} is between .499 and .501 is equal to .95?

10. **In this problem, use the continuity correction where applicable**

The number of accidents per week at a hazardous intersection varies with mean $\mu = 2.2$ and standard deviation $\sigma = 1.4$. This distribution takes only integer values. So it is certainly not normal.

- (a) (2 marks) Let \bar{X} be the average number of accidents per week at the intersection during a year (52 weeks). What is the approximate distribution of \bar{X} ? Justify your answer.
- (b) (2 marks) What is the approximate probability that \bar{X} is less than 2?

- (c) (2 marks) What is the approximate probability that there are fewer than 100 accidents at the intersection in a year?

11. (4 marks) The voters in a large US city are 40% black, 40% white and 20% hispanic. A black mayoral candidate anticipates attracting 30% of the White votes, 90% of the Black votes, and 50% of the Hispanic votes. What percent of the overall vote does the candidate expect to get?

Multiple choice questions:

- M1. (3 marks) If a random variable X is symmetric with mean -0.36 and standard deviation of 0.2 , and $Y = 5X + 1.8$, then the distribution of Y can be described as
- (A) skewed with mean 0.36 and standard deviation -0.2.
 - (B) symmetric with mean 0.36 and standard deviation 1
 - (C) symmetric with mean 0 and standard deviation 1.
 - (D) symmetric with mean 0 and standard deviation 2.8.
- M2. (3 marks) Event A occurs with probability 0.2. Event B occurs with probability 0.8. If A and B are independent, then
- (A) $P(A \text{ or } B) = 1$
 - (B) $P(A \text{ and } B) = 0.016$
 - (C) $P(A \text{ or } B) = .84$
 - (D) $P(A \text{ or } B) = .92$

- M3. (3 marks) Suppose that, at an intersection having flashing red lights in all directions, only 20% of all drivers come to a complete stop, when no other cars are visible. We assume that drivers behave independently from each other. Then the number 0.3704 represents
- (A) the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions, at most 5 will come to a complete stop.
 - (B) the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions, exactly 5 will come to a complete stop.
 - (C) the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions, at least 5 will come to a complete stop.
 - (D) the expected number of drivers, among the next 20 drivers coming to an intersection under these conditions, who will come to a complete stop.

- M4. (3 marks) Suppose you are going to roll a fair die 80 times and record \hat{p} , the proportion of times that a number divisible by three is showing. The mean of the sampling distribution of \hat{p} is
- (A) 4.5
 - (B) 1/2
 - (C) 1/3
 - (D) 1/6

- M5. (3 marks) Given that $P(A) = 0.5$, $P(B) = 0.4$ and $P(A|B) = 0.3$. What is $P(A^c \text{ and } B^c)$?

- | | |
|-----|------|
| (A) | 0.90 |
| (B) | 0.78 |
| (C) | 0.12 |
| (D) | 0.22 |

- M6. (3 marks) One hundred volunteers who suffer from severe depression are available for a study. Fifty are selected at random and are given a new drug thought to be particularly effective in treating severe depression. The other fifty are given an existing drug for treating severe depression. A psychiatrist evaluates the symptoms of all volunteers after four weeks in order to determine if there has been substantial improvement in the severity of the depression. The levels of the factor in this study are

- | | |
|-----|--|
| (A) | which treatment the volunteers receive. |
| (B) | the use of randomization and the fact that this was a comparative study. |
| (C) | the extent to which the depression was reduced. |
| (D) | the use of a psychiatrist to evaluate the severity of depression. |

M7. (3 marks) A pharmaceutical company interested in measuring how often physicians prescribe a certain drug has selected a simple random sample from each of two groups: medical doctors and osteopathic doctors. What is this type of sampling called?

- | | |
|-----|-----------------------------|
| (A) | Simple random sampling. |
| (B) | Multistage sampling. |
| (C) | Stratified random sampling. |
| (D) | None of the above. |

M8. (3 marks) If the histogram of a variable show that the distribution is symmetric, which of the following is generally **incorrect**.

- (A) It has a normal distribution.
 (B) The mean and the median are approximately equal.
 (C) The proportion of data to the left of the median is equal to the proportion of data to the right of the median.
 (D) The left and right whiskers of the box-plot are approximately equal.

M9. (3 marks) Two variables in a study are said to be confounded if

- | | |
|-----|---|
| (A) | one cannot separate their effects on a response variable. |
| (B) | they are highly correlated. |
| (C) | they do not have a normal distribution. |
| (D) | one of them is a placebo. |

M10. (3 marks) The ten people listed below are enrolled in an internet statistics course.

Moore	Casella	Santner
Goel	Jones	Klein
Anderson,	Draper	Fisher
Pearson		

List the names in alphabetical order and give them successive numerical labels starting with **1** for the first name in the alphabetical order. Use the following random digits to choose a simple random sample of three to be interviewed in detail about the quality of the course.

27110 256057 55892 33063 40148

What is the sample you obtain?

- (A) Casella, Klein, Anderson
- (B) Casella, Moore, Klein
- (C) Casella, Goel, Jones
- (D) Casella, Anderson, Goel