475.101 / 475.102 Semester 1 2000
Assignment 2 Solutions

Question 1.

(a) (i) 0.25 **(ii)** 25%

(b) (i) $\text{pr}(X = 4) = 0.236$.

(ii) $\text{pr}(X \geq 6) = 1 - \text{pr}(X \leq 5) = 1 - 0.753 = 0.247$.

or $\text{pr}(X \geq 6) = \text{pr}(X=6) + \ldots + \text{pr}(X=10) = 0.147+0.070+0.023+0.005+0.001 = 0.246$.

(iii) $\text{pr}(X < 8) = \text{pr}(X \leq 7) = 0.971$.

or $\text{pr}(X<8) = 1 - \text{pr}(X=8) - \text{pr}(X=9) - \text{pr}(X=10) = 1 - 0.023 - 0.005 - 0.001 = 0.971$.

or $\text{pr}(X < 8) = \text{pr}(X = 0) + \ldots + \text{pr}(X = 7) = 0.004 + \ldots + 0.070 = 0.971$.

(iv) $\text{pr}(4 \leq X \leq 7) = \text{pr}(X \leq 7) - \text{pr}(X \leq 3) = 0.971 - 0.296 = 0.675$.

or $\text{pr}(4 \leq X \leq 7) = \text{pr}(X = 4) + \ldots + \text{pr}(X = 7) = 0.236 + 0.221 + 0.147 + 0.070 + = 0.674$.

(c) (i) $\text{pr}(X \leq 4) = 0.313$

or $\text{pr}(X \leq 4) = \text{pr}(X = 0) + \ldots + \text{pr}(X = 4) = 0.003+0.018+0.051+0.098+0.143 = 0.313$.

(ii) $\text{pr}(X > 2) = 1 - \text{pr}(X \leq 2) = 1 - 0.072 = 0.928$.

or $\text{pr}(X > 2) = \text{pr}(X = 3) + \ldots + \text{pr}(X = 15) = 0.098+0.143+ \ldots +0.001 = 0.929$.

(iii) $\text{pr}(X = 11) = 0.019$.

(iv) $\text{pr}(8 \leq X \leq 12) = \text{pr}(X \leq 12) - \text{pr}(X \leq 7) = 0.993 - 0.771 = 0.222$.

or $\text{pr}(8 \leq X \leq 12) = \text{pr}(X = 8) + \ldots + \text{pr}(X = 12) = 0.096+0.062+0.036+0.019+0.009 = 0.222$.

(d) (i) 1.00 should have read 0.10. (Probabilities must be in the range 0 to 1, probabilities for a distribution must sum to 1.)

(ii) The probability that X is at least 7 = $\text{pr}(X \geq 7) = 0.37 + 0.24 + 0.09 = 0.70$.

(iii) The probability that X is no more than 8 = $\text{pr}(X \leq 8) = 0.10 + 0.20 + 0.37 + 0.24 = 0.91$.

or $\text{pr}(X \leq 8) = 1 - \text{pr}(X = 9) = 1 - 0.09 = 0.91$.

(iv) $E[X] = 5 \times 0.10 + 6 \times 0.20 + 7 \times 0.37 + 8 \times 0.24 + 9 \times 0.09 = 7.02$

$$\text{sd}(X) = \sqrt{[5-7.02]^2 \times 0.10 + [6-7.02]^2 \times 0.20 + [7-7.02]^2 \times 0.37 + [8-7.02]^2 \times 0.24 + [9-7.02]^2 \times 0.09}$$

$$= \sqrt{1.1996} \approx 1.0953.$$

Question 2.

(a) X is an <u>approximate</u> Poisson distribution where $X \sim \text{Poisson}\ \lambda = \frac{4}{5} = 0.8$.

The assumptions for the Poisson distribution to check include that the events (dead possums on a given kilometre of the Napier to Taupo highway) occur at a constant average rate (here, 0.8 per kilometre.) Here we are assuming that this rate is constant throughout the highway. This seems unreasonable as the number of dead possums could vary depending on environments (e.g. season, dense forest, farmland or towns). The second assumption is that occurrences are independent of one another. This may not be the case due to the possibility of small family groups crossing the road together. The third assumption is that as the space interval gets smaller the likelihood of finding two or more dead possums diminishes. So this assumption is true.

(b) X does not have a Binomial or Poisson distribution.

(c) X is an <u>approximate</u> Binomial distribution where $X \sim \text{Binomial}(n=200, p=0.20)$.

There is a fixed number of hotels (200) selected without replacement from all hotels in America (i. e. 200 trials). Each hotel (trial) has two outcomes - it either does not offer complimentary shampoo (a success) or it does, (a failure). The selection of the 200 hotels can be viewed as follows:

(i) The first hotel is randomly selected from all hotels in America.
(ii) The second hotel (2nd trial) is randomly selected from all hotels in America minus 1 etc.

The conditions for each trial are not identical, but are not too different. A Binomial model assumes the conditions are identical for each trial. Here because of the small difference in the conditions for each trial, a Binomial model will be a reasonable approximation.

Question 3.

(a)

	Brand		
Type	Coke	Pepsi	Total
Diet	0.2895	0.1684	0.4579
Regular	0.3158	0.2263	0.5421
Total	0.6053	0.3947	1.0000

Note: pr(Coke and Regular) = $0.6053 \times 0.5217 \approx 0.3158$.
pr(Pepsi and Diet) = $0.3947 \times 0.4267 \approx 0.1684$.

(b) The probability that a randomly chosen person who drinks diet soft drinks, prefers Coke $= \frac{0.2895}{0.4579} \approx 0.6322$.

(c) The percentage of regular drinkers preferring Pepsi $= \frac{0.2263}{0.5421} \times 100 \approx 41.75\%$.

(d) The results from this study suggest that overall people prefer Coke to Pepsi, with approximately 60.5% preferring Coke compared to approximately 39.5% preferring Pepsi. In addition roughly 63% of those who drink diet soft drinks and 58% (100% – 42%) of those who drink regular soft drinks prefer Coke.

Question 4.

(a) (i) Prestigious. **(ii)** Defensive.

(b) The percentage of respondents who describe themselves as prestigious drivers that selected Alphasud or the Spider as his or her model of choice $= \frac{56+80}{204} \times 100 = \frac{136}{204} \times 100 \approx 66.67\%$.

(c) (i) The proportion of respondents who said the Spider was his or her model of choice and described him or herself as an aggressive driver $= \frac{89}{669} \approx 0.1330$.

(ii) The proportion of respondents who selected either the Alphasud or the Giulia as their model of choice $= \frac{133+194}{669} = \frac{327}{669} \approx 0.4888$.

(d) Among respondents who chose the Giulia or Spider as his or her model of choice, the proportion who describe him or herself as either an "enjoying" or "prestigious" type of driver $= \frac{42+68+96+80}{194+342} = \frac{286}{536} \approx 0.5336$.

(e) The probability that a respondent selects the Alphasud as his or her model of choice given that he or she describes him or herself as a defensive driver $= \frac{22}{138} \approx 0.1594$.

Question 5.

(a) An approximate Binomial distribution is an appropriate model as:

There is a fixed number of pilots (25) selected without replacement from all of the company's pilots. (i. e. 25 trials). Each pilot (trial) has two outcomes – he or she is either over 40 years of age (success) or is not (failure). The selection of the 25 pilots can be viewed as follows:

(i) The first pilot is randomly selected from all the airline's pilots.
(ii) The second pilot (2nd trial) is randomly selected from all the airline's pilots minus 1 etc.

The conditions for each trial are not identical, but are not too different. A Binomial model assumes the conditions are identical for each trial. Here because of the small difference in the conditions for each trial, a Binomial model will be a reasonable approximation.

Note: Any conclusions obtained when using the Binomial model will only be as good as the validity of the assumptions that have been made.

(b) Number of trials $n = 25$
Probability of success $p = 0.35$
$X \sim \text{Binomial}(n = 25, p = 0.35)$.

(c) **(i)** $\text{pr}(X > 7) = 1 - \text{pr}(X \leq 7) = 1 - 0.306078 = 0.693922$.
(ii) $\text{Pr}(X = 5) = 0.050576$.
(iii) $\text{pr}(13 \leq X \leq 18) = \text{pr}(X \leq 18) - \text{pr}(X \leq 12) = 0.999966 - 0.939555 = 0.060411$.

(d) Expected value for the number of pilots over 40 years of age,
$E[X] = np = 25 \times 0.35 = 8.75$.
Standard deviation for the number of pilots over 40 years of age,
$\text{sd}[X] = \sqrt{np(1-p)} = \sqrt{25 \times 0.35 \times 0.65} = 2.384848$

(e) The airline company may be interested in the variable "over 40 years of age" for timing & cost reasons. For instance age may indicate level of experience and thus it may be quicker (and cost less) to train a pilot over 40 years of age (assuming they are more experienced). On the other-hand the older the pilot is the longer it may take to train him or her, hence increasing training costs. Another reason the company may be interested in the variable "over 40 years of age" is that it may not be worth training pilots in this age group as they may tend to retire sooner than pilots "40 years of age or under".

Another variable the airline company may be interested in measuring could be experience. The level of experience could determine the speed of training and thus have cost implications.

Question 6.

(a) The assumptions for the Poisson distribution to check include that the events (the number of "hits" on the stage I web page) occur at a constant average rate (here, 7.5 "hits' per day) Here we are assuming that this rate is constant every day. This may not be the case as certain days may be more "popular". For instance when assignment solutions, test answers or grades are posted. The second assumption is that occurrences are independent of one another. This also may not be the case as fellow students may together, decide to view the stage I web page using different computers. The third assumption is that as the time interval gets smaller the likelihood of getting two or more "hits" diminishes. So this assumption is true.

Note: Any conclusions obtained when using the Poisson model will only be as good as the validity of the assumptions that have been made.

(a) The parameter for this distribution is $\lambda = 7.5$ "hits" on the stage I web page.
$X \sim \text{Poisson}(\lambda = 7.5)$.

(c) $\text{pr}(X \leq 9) = 0.776408 \approx 0.776$.

(d) The parameter of this distribution is $\lambda = 15$ "hits" on the stage I web page. $X \sim \text{Poisson}(\lambda = 15)$.
(i) $\text{pr}(Y = 12) = 0.082859$.
(ii) $\text{pr}(11 \leq Y \leq 19) = \text{pr}(Y \leq 19) - \text{pr}(Y \leq 10) = 0.875219 - 0.118464 = 0.756755$.

(e) Expected value $E[Y] = \lambda = 15$ "hits" on the stage I web page in a given two-day period.

Standard deviation for the number of hits" on the stage I web page in a given two-day period,
$\text{sd}[Y] = \sqrt{\lambda} = \sqrt{15} = 3.87298$.

(f) The range of values, no further than three standard deviations away from the mean, in which the number of $\lambda = 15$ "hits" on the stage I web page in a given two day period is very likely to be is:
$$E[Y] \pm 3 \times \text{sd}[Y] \approx 15 \pm 3 \times 3.87298 \approx 15 \pm 11.6189 \approx [3.38, 26.62].$$
It is very likely that between 3 and 27 "hits" on the stage I web page occur in a given two day period