101 / 102 / 107 / 108 Semester 2 2000
Assignment 2 Solutions

Question 1.

(a) **(i)** 0.734375 **(ii)** 73.4375%

(b) **(i)** pr(X = 10) = 0.031

(ii) pr(X > 12) = pr($X \geq$ 13) = 1 – pr($X \leq$ 12) = 1 – 0.228 = 0.772

or pr(X > 12) = pr(X= 13) + … + pr(X = 20) = 0.164 + … + 0.001 = 0.773

(iii) pr(X < 11) = pr($X \leq$ 10) = 0.048

or pr(X < 11) = pr(X= 0) + … + pr(X = 10)

= 0.001 + 0.004 + 0.012 + 0.031 = 0.048

(iv) pr(13 $\leq X \leq$ 17) = pr($X \leq$ 17) – pr($X \leq$ 12) = 0.965 – 0.228 = 0.737

or pr(13 $\leq X \leq$ 17) = pr(X = 13) + ... + pr(X = 17)

= 0.164 + 0.192 + 0.179 + 0.130 + 0.072 = 0.737

(c) **(i)** pr($X \leq$ 6) = 0.060

or pr($X \leq$ 6) = pr(X= 0) + … + pr(X = 6) = 0.001 + … + 0.033 = 0.061

(ii) pr($X \geq$ 16) = 1 – pr(X $\leq$ 15) = 1 – 0.878 = 0.122

or pr($X \geq$ 16) = pr(X = 16) + … + pr(X = 20) = 0.045 + … + 0.012 = 0.122

(iii) pr(X = 5) = 0.017

(iv) pr(9 $\leq X \leq$ 14) = pr($X \leq$ 14) – pr($X \leq$ 8) = 0.815 – 0.191 = 0.624

or pr(9 $\leq X \leq$ 14) = pr(X =9) +...+ pr(X =14)

= 0.098+0.113+0.118+0.113+0.100+0.082 = 0.624

(d) **(i)** The probability that X is at most 2 = pr($X \leq$ 2) = 0.150 + 0.333 + 0.350 = 0.833

(ii) The probability that X is more than 3 = pr(X > 3) = 0.067 + 0.010 = 0.077

(iii) What is the probability that X is between 1 and 4 (inclusive)

= pr(1 $\leq X \leq$ 4) = 0.333 + 0.350 + 0.090 + 0.067 = 0.840

or 1 – pr(X =0) – pr(X= 5) = 1 – 0.150 – 0.010 = 0.840

(iv) $E[X] = 0 \times 0.150 + 1 \times 0.333 + 2 \times 0.350 + 3 \times 0.090 + 4 \times 0.067 + 5 \times 0.010 = 1.621$

$\text{sd}(X) = \sqrt{[0-1.621]^2 \times .150 + [1-1.621]^2 \times .333 + ... + [5-1.621]^2 \times .010} = \sqrt{1.237358999} \approx 1.1124$

Question 2.

(a) X does not have a Poisson or Binomial distribution (as $\frac{n}{N} = \frac{7}{52} = 0.13$. i.e. > 0.1 and there is no rate at which events occur).

(b) X is an <u>approximate</u> Binomial distribution where X ~ Binomial(n=13, p=0.60).

There is a fixed number of service calls (13) monitored without from service calls (i. e. n =13 trials). Each service call (trial) has two outcomes – the repair person will reach the customer’s home within the acceptable waiting period (a success) or not, (a failure). The probability that the repair person will reach the customer’s home within the acceptable waiting period is assessed to be 60% (i. e. p =0.6), based on a number of years data. Whether this probability is constant for a particular month may be in doubt as busy periods or weather may effect the response time. Independence of response times could also be questionable for consecutive calls, as a particularly long job may cause delays in response to several other jobs. As long as these issues are not known to be too major, we can still approximate X with a Binomial model.

Note: Viewing the selection of the 13 service calls as sampling without replacement from that particular months service calls will mean the conditions for each trial are not identical, as getting a success on the first call will alter the chance of getting a success on the second call. However, the conditions are not too different as the population size N of all service calls for the electricity industry within the whole of New Zealand is likely to be more than 130 a month thus n/N is likely to be less than 0.1.

(c) X is an <u>approximate</u> Poisson distribution where X ~ Poisson (λ = 55) customers.

The assumptions to check for the Poisson distribution include that the events (customers visiting the bank during the 12:00pm to 1:00pm lunch hour) occur at a constant average rate (here, 55). Here we are assuming that this rate is constant over all lunch hours. This may not be the case as the number of customers visiting the bank could differ depending on the day of the week. I.e., the rate of visits may differ for Mondays compared with Tuesdays. Also, the rate of customer visiting the bank during the lunch hour may not be constant as more customers may visit at the start of the lunch hour than the end of the lunch hour. The second assumption is that occurrences are independent of one another. This may not be the case due to the possibility of customers (e.g work-mates) influencing one another’s decision to visit the bank. The third assumption is that as the time interval gets smaller the likelihood of two or more customers visiting the bank diminishes. So this assumption is true.

Question 3.

(a)

	Grade			
Smokeless tobacco	**10**	**11**	**12**	**Total**
Uses	0.0499	0.0755	0.0644	0.1898
Does not use	0.3119	0.2908	0.2075	0.8102
Total	0.3618	0.3663	0.2719	1.0000

Note: pr(Does not use and 11) = $\frac{262}{330} \times \frac{330}{901} \approx 0.7939 \times 0.3663 \approx 0.2908$

pr(Uses and 12) = $0.1898 \times 0.3392 \approx 0.0644$

pr(Does not use and 10) = $0.8102 \times 0.3850 \approx 0.3119$

(b) The percentage of randomly selected students who use smokeless tobacco:

(i) grade 10 students = $\frac{0.0499}{0.3618} \approx 0.1379 \approx 13.79\%$

(ii) grade 11 students = $\frac{0.0755}{0.3663} \approx 0.2061 \approx 20.61\%$

(iii) grade 12 students = $\frac{0.0644}{0.2719} \approx 0.2369 \approx 23.69\%$

(c) The results from this study suggest that in these two Arkansas communities roughly 20% of high school students use smokeless tobacco. There also appears to be increasing use of smokeless tobacco as the grades go higher, with roughly 14% of grade 10 students, 21% of grade 11 students and finally 24% of grade 12 students using smokeless tobacco.

(d) This study is not representative of the smokeless tobacco use among high school students in Arkansas. Instead we would expect it to just represent the smokeless tobacco use among high school students in the two Arkansas communities. Thus it is not valid to rely on any conclusions reached from this study when discussing smokeless tobacco use among high school students in Arkansas.

Question 4.

(a) (i) 45 – 54.

(ii) 25 – 34, as $\dfrac{19,587+2,444}{39,354}\times 100 = 55.98\%$

(b) (i) The percentage of Americans in this survey who did not attend university
$= \dfrac{29,559+58,174}{172,214}\times 100 = \dfrac{87,733}{172,214}\times 100 = 50.94\%$

(ii) The percentage of Americans in this survey who only completed high school or were aged between 35 and 44 years $= \dfrac{58,174+44,460-15,136}{172,214}\times 100 = \dfrac{87,498}{172,214}\times 100 = 50.81\%$

(c) The probability that a randomly chosen American who had only completed high school, is at most 54 years of age $= \dfrac{12,569+15,136+10,943}{58,174} = \dfrac{38,648}{58,174} = 0.6644$

(d) The proportion of 25 and 34 year old Americans in this survey who did not attend university
$= \dfrac{4,754+12,569}{39,354} = \dfrac{17,323}{39,354} = 0.4402$

Question 5.

(a) The Binomial distribution assumption of "constant probability of getting a success" is violated here as the probability of solving a problem is not the same for all of the problems due to the differing degrees of difficulty.

(b) The Binomial distribution assumption of "independence" is violated here as the result of the previous shot may affect the result of the current shot.

(c) (i) The Poisson distribution assumption of "events occur at a constant average rate per unit of time" is violated here as the passengers tend to arrive close to the time the bus is due. The Poisson distribution assumption of "independent occurrences" is violated here as groups of people (such as friends etc) can decide to arrive at the bus stop at particular times.

(ii) - more than one person arrives at the bus stop in a:

10 minute interval	LIKELY
2 minute interval.	LESS LIKELY
30 second interval.	EXTREMELY UNLIKELY

The above suggests that the Poisson distribution assumption of "more than one occurrence cannot happen at the same time" appears to be valid.

Question 6.

(a) The parameter for this distribution is $\lambda = 10.2$ people arrive at the bus stop in the half hour prior to the bus arriving.

$X \sim \text{Poisson}(\lambda = 10.2)$.

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

(b) (i) $\text{pr}(X \le 6) = 0.11802629 \approx 0.1180$

(ii) $\text{pr}(X = 0) = 3.717\text{E-}05 \approx 0.0000$

(iii) $\text{pr}(11 \le X \le 19) = \text{pr}(X \le 19) - \text{pr}(X \le 10) = 0.9957289 - 0.55803425$
$= 0.43769465 \approx 0.4377$

(c) (i) Expected value $\text{E}[X] = \lambda = 10.2$ people arrive at the bus stop in the half hour prior to the bus arriving

(ii) Standard deviation for the number of people arriving at the bus stop in the half hour prior to the bus arriving $\text{sd}[X] = \sqrt{\lambda} = \sqrt{10.2} = 3.193743885 \approx 3.1937$.

(d) The range of values, no further than three standard deviations away from the mean, in which the number of people arriving at the bus stop in the half hour prior to the bus arriving is very likely to be:
$\text{E}[X] \pm 3 \times \text{sd}[X] \approx 10.2 \pm 3 \times 3.1937 \approx 10.2 \pm 9.5811 \approx [0.619, 19.781]$.
It is very likely that between 1 and 20 people arriving at the bus stop in the half hour prior to the bus arriving.

Question 7

(a) Number of trials $n = 3$
Probability of success $p = 0.62$
$X \sim \text{Binomial}(n = 3, p = 0.62)$.

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) (i) $\text{pr}(X = 3) = 0.238328 \approx 0.2383$

(ii) $\text{pr}(X \ge 2) = 1 - \text{pr}(X \ge 1) = 1 - 0.323456 = 0.676544 \approx 0.6765$
or $\text{pr}(X \ge 2) = 1 - \text{pr}(X = 0) - \text{pr}(X = 1) = 1 - 0.054872 - 0.268584 = 0.676544 \approx 0.6765$

(iii) $\text{pr}(0 < X \le 2) = \text{pr}(1 \le X \le 2) = \text{pr}(X \le 2) - \text{pr}(X = 0) = 0.761672 - 0.054872 = 0.7068$
or $\text{pr}(0 < X \le 2) = \text{pr}(1 \le X \le 2) = \text{pr}(X = 1) + \text{pr}(X = 2) = 0.268584 - 0.438216 = 0.7068$

(c) Expected value for the number of shots that would hit the bulls-eye,
$\text{E}[X] = np = 3 \times 0.62 = 1.86$

Standard deviation for the number of shots that would hit the bulls-eye,
$\text{sd}[X] = \sqrt{np(1-p)} = \sqrt{3 \times 0.62 \times 0.38} = \sqrt{0.7068} \approx 0.8407$