STAT40720 Intro. to Data Analytics

Assignment 1

Name	Kevin O'Connell
Student ID	#########
Submission Date	5th. October 2015

Question 1

Chebyshev's Theorem provides the proportion of a dataset lying within **k** standard deviations of the mean is given by:

$$1 - \frac{1}{k^2}$$

So, for 75% of the population:

$$\begin{aligned} 1 - \tfrac{1}{k^2} &= 0.75 \\ \tfrac{1}{k^2} &= 0.25 \\ k^2 &= 4 \\ k &= 2 \end{aligned}$$

75% of the population will lie between 2 standard deviations of the mean, yielding the range:

$$R = €528{,}000 \pm 2 \times €57{,}000$$

75% of the houses will sell for a price between €414,000 and €642,000.

Question 2

(a) At most 2 lines are in use

$$\begin{aligned} P(x \le 2) &= P(0) + P(1) + P(2) \\ &= 0.14 + 0.21 + 0.2 \\ &= 0.55 \end{aligned}$$

(b) At least 3 lines are in use

$$\begin{aligned} P(x \ge 3) &= P(3) + P(4) \\ &= 0.25 + 0.2 \\ &= 0.45 \end{aligned}$$

(c) Between 2 and 3 line, inclusive, are not in use

$$\begin{aligned} P(!\,2\ or\ 3) &= 1 - \big(P(2) + P(3)\big) \\ &= 1 - (0.2 + 0.25) \\ &= 0.55 \end{aligned}$$

(d) At least 3 lines are not in use

$$
\begin{aligned}
P(x \leq 1) &= P(0) + P(1) \\
&= 0.14 + 0.21 \\
&= 0.35
\end{aligned}
$$

Question 3

(a) What is the value of *i*?

$$
\begin{aligned}
\sum_{x=1}^{5} P(x) &= 1 \\
\sum_{x=1}^{5} ix &= 1 \\
i \sum_{x=1}^{5} x &= 1 \\
i(1+2+3+4+5) &= 1 \\
i &= \frac{1}{15}
\end{aligned}
$$

(b) What is the probability that at most 2 documents are required?

$$
\begin{aligned}
P(x \leq 2) &= P(1) + P(2) \\
&= i\times 1 + i\times 2 \\
&= \frac{1}{15}\times 3 \\
&= 0.2
\end{aligned}
$$

Question 4

(a) Determine P(X ≤ 2)?

$$
\begin{aligned}
\sum_{X=0}^{2} \frac{n!}{X!(n-X)!} p^X q^{n-X} &= \frac{20!}{0!(20-0)!} 0.05^0 0.95^{20} + \frac{20!}{1!(20-1)!} 0.05^1 0.95^{20-1} \\
&\quad + \frac{20!}{2!(20-2)!} 0.05^2 0.95^{20-2} \\
&= 0.358486 + 0.377354 + 0.188677 \\
&= 0.924517
\end{aligned}
$$

(b) Determine P(1 ≤ X ≤ 4)?

$$\sum_{X=1}^{4} \frac{n!}{X!(n-X)!} p^X q^{n-X} = \frac{20!}{1!(20-1)!} 0.05^1 0.95^{20-1} + \frac{20!}{2!(20-2)!} 0.05^2 0.95^{20-2}$$

$$+ \frac{20!}{3!(20-3)!} 0.05^3 0.95^{20-3} + \frac{20!}{4!(20-4)!} 0.05^4 0.95^{20-4}$$

$$= 0.37735 + 0.18868 + 0.059582 + 0.013328$$

$$= 0.638941$$

(b) Determine P(0)?

$$X = 0, \frac{n!}{X!(n-X)!} p^X q^{n-X} = \frac{20!}{0!(20-0)!} 0.05^0 0.95^{20-0}$$

$$= 0.358486$$

Question 5

(a) Determine P(x < 30km/h)?

$$P(x < 30) = P\left(z < \frac{30-35}{16}\right)$$

$$= P\left(z < -\frac{5}{16}\right)$$

But the Normal distribution is symmetric about zero, yielding:

$$P(x < 30) = 1 - P\left(z < \frac{5}{16}\right)$$

$$= 1 - P(z < 0.3125)$$

$$= 1 - 0.62265$$

$$= 0.37735$$

Note, the value for $P(z < 0.3125)$ above was taken by linear interpolation between probabilities for 0.31 and 0.32 in the NCST tables.

(b) Determine P(x > 45km/h)?

$$
\begin{aligned}
P(x > 45) &= 1 - P(x < 45) \\
&= 1 - P\left(z < \tfrac{45-35}{16}\right) \\
&= 1 - P(z < 0.625) \\
&= 1 - 0.73405 \\
&= 0.26595
\end{aligned}
$$

(c) Determine P(40km/h < x < 60km/h)?

$$
\begin{aligned}
P(40 < x < 60) &= P(x < 60) - P(x < 40) \\
&= P\left(z < \tfrac{60-35}{16}\right) - P\left(z < \tfrac{40-35}{16}\right) \\
&= P(z < 1.5625) - P(z < 0.3125) \\
&= 0.9409 - 0.62265 \\
&= 0.31825
\end{aligned}
$$