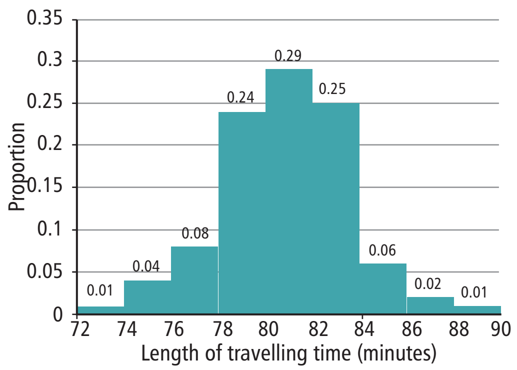
Statistics 3.14

Relating to page 264 of *Level 3 Statistics Learning Workbook*

Practice assessment task

1. The times taken to drive in morning traffic between Otaihanga and Central Wellington can be modelled by the normal distribution, with a mean of 49.2 minutes and a standard deviation of 4.1 minutes.
   a. Calculate the percentage of drives that could be expected to last less than 46 minutes or longer than 55 minutes.

Data on the length of travelling times to travel by car (to get to the station) then travel by train between Otaihanga and Central Wellington was collected over a long period of time. The histogram below displays this information.



For this data, the mean length of travelling times is 81.3 minutes and the standard deviation is 2.4 minutes.

   b. Explain whether a normal distribution would be an appropriate model for the distribution of travelling times between Otaihanga and Central Wellington using a car (to get to the station) and train as the mode of travel.

   As part of your explanation, describe the features of the distribution and include at least one calculation.

   c. Two commuters (B and C) read books as they travel by train between Paraparaumu and Wellington railway stations.

   On these trips, the number of words that commuter B reads can be modelled by a normal distribution, with mean 12 430 words and standard deviation 64.3 words.

   The number of words commuter C reads on these trips can be also be modelled by a normal distribution, with mean 12 430 words. Commuter C expects to read more than 12 500 words on 17% of the train journeys.

   For the numbers of words read by commuters B and C on these train trips between Paraparaumu and Wellington railway stations, compare the standard deviation of commuter B with the standard deviation of commuter C.

2. a. The table below shows the probability distribution of the random variable N, the number of takeaway pizzas ordered by a customer.

n	1	2	3	4	5	6
$P(N = n)$	0.18	0.26	0.36	0.13	0.04	0.03

i. Calculate the mean number of takeaway pizzas ordered by customers.

The pizzas are packaged in cardboard boxes.

The cost to make each pizza, including the cost of the material and labour, is \$4.50.

The cost of the cardboard box that holds one pizza is \$0.20, the cost of a cardboard box that holds two pizzas is \$0.25 and the cost of a cardboard box that holds three or more pizzas is \$0.50.

ii. Calculate the expected cost of each takeaway pizza order.

b. The time each customer spends in the pizza restaurant waiting for their order is somewhere between 5 minutes and 25 minutes.

i. Using an appropriate model, determine the maximum length of time that the restaurant could expect 90% of the customers to be waiting for their orders.

ii. Justify your selection of an appropriate distribution model.

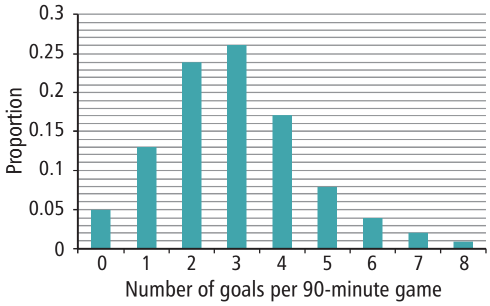
c. Over the previous three months, the pizza restaurant owner recorded the number of orders received in each 10-minute interval (between the hours of 6 p.m. and 8 p.m.). She found that in 92% of such periods there was at least one customer.

Calculate the probability that the takeaway bar will receive more than 8 orders in any given 30-minute period.

3. Shinty is an ancient Scottish game (similar to field hockey) which is played with sticks (camans) and a ball.

Shinty is played in two 45-minute halves and there are seldom any breaks in the play. Goals are fairly rare, occur independently and are approximately equally likely to happen during any minute of the game.

a. Recordings of the number of goals scored by a team during a 90-minute shinty game, over a large number of games, resulted in the following graph.



i. Describe the key features of the distribution, and obtain an estimate for the mean number of goals scored by this team per 90-minute game (rounded to one decimal place).

After the first 45-minute half of one shinty game this team played, the team had scored no goals.

ii. Using your answer to part **i.** and an appropriate distribution to model this situation, calculate the probability of the team scoring at least two goals in the second half of the game.

In your answer you should justify your choice of this distribution, and state any assumption(s) you make.

b. For another shinty team, the mean number of goals scored per 90-minute game is 1.2.

i. Calculate the probability that during three different 90-minute games, the team scores fewer than four goals in each game.

State any assumption(s) you make.

ii. Using an appropriate distribution, find the probability that this team will score at least four goals in at least three of the next six 90-minute games.

In your answer, identify the distribution and its parameter(s), and state any assumptions(s) you make.

Answers

1. a. Normal distribution $\mu = 49.2$, $\sigma = 4.1$
   $P(X < 46) + P(X > 55) = 0.2961$ (4 s.f.)
   29.61% of all drives will take less than 46 minutes or longer than 55 minutes.

   b. The normal distribution would be an appropriate model because
   - the shape of the distribution appears to be symmetrical
   - the shape of the distribution appears to be bell-shaped
   - the variable is continuous (length of drives measured in minutes)
   - it is possible for the lengths of drives to be higher or lower than the bounds given.

   Using the normal distribution with $\mu = 81.3$ and $\sigma = 2.4$, the probability that the travel time is between 76 and 84 minutes is $P(76 < X < 84) = 0.8561$

   Reading off the histogram $P(76 < X < 84) = 0.86$ which is very similar to the theoretical probability obtained using the normal distribution.

   c. Using the inverse normal distribution ($\mu = 0$, $\sigma = 1$, right tail $= 0.17$) gives the value $x = 12\,500$ a z-score of 0.95417.

   Solving $\frac{12\,500 - 12\,430}{\sigma} = 0.95417$ gives $\sigma = 73.4$

   The standard deviation of the distribution of the number of words read by commuter C is larger than the standard deviation of the distribution of words read by commuter B (i.e. the number of words read by commuter C is more variable than the number of words read by commuter B).

2. a. i. $E(N) = 1 \times 0.18 + 2 \times 0.26 + 3 \times 0.36 + 4 \times 0.13 + 5 \times 0.04 + 6 \times 0.03$
   $= 2.68$ pizzas

   ii. Let C = cost of pizza order (in dollars), n = number of pizzas

n	1	2	3	4	5	6
Cost c	4.70	9.25	14.00	18.50	23.00	27.50
$P(C = c)$	0.18	0.26	0.36	0.13	0.04	0.03

   $E(C) = 4.7 \times 0.18 + 9.25 \times 0.26 + 14 \times 0.36 + 18.5 \times 0.13 + 23 \times 0.04 + 27.5 \times 0.03$
   $E(C) = \$12.44$

   b. i. Uniform distribution where $20h = 1$, i.e. $h = \frac{1}{20}$

   P(Waiting time $< x$) $= 0.9$ gives $\frac{1}{20}x = 0.9$, so $x = 18$ minutes

   So maximum time 90% of customers wait for orders is $5 + 18 = 23$ minutes

   ii. A uniform distribution can be used because there is no indication that one interval of wait times is any more likely than another interval:
   - Minimum and maximum values are given.
   - No mode is given.

   c. Poisson distribution, with $P(X \geq 1) = 0.92$ gives
   $P(X = 0) = 0.08$
   Solving $e^{-\lambda} = 0.08$ gives an average of $\lambda = 2.52573$ orders per 10 minutes
   Tripling gives $\lambda = 7.5772$ orders per 30 minutes
   $P(X > 8) = 1 - P(X \leq 8) = 0.3487$
   There is a 35% chance that the takeaway bar will receive more than 8 orders in any given 30-minute period.

3. a. i. Key features of distribution:
   - Discrete random variable
   - Lowest value 0, highest value 8 (for the data given but could be higher)
   - Shape is skewed to the right
   - Mode is 3

   Mean = 2.9 goals (1 d.p.)

   ii. Poisson distribution with $\lambda = 2.9$

   Apply this distribution because the variable is discrete (number of goals) over a 'continuous' interval of 90 minutes, with no upper limit on the number of goals that it is possible to score
   - occurrences occur randomly at a constant rate
   - two goals cannot occur at the same time
   - each goal is independent of any other goal
   - the probability of a goal in an interval of time is proportional to the size of the interval

   Using Poisson with $\lambda = 1.45$ (for 45 minutes)
   $P(X \geq 2) = 0.4253$ (4 s.f.)

   b. i. Let X = number of goals scored by this team per 90-minute game.
   Using Poisson distribution ($\lambda = 1.2$) gives
   $P(X < 4) = P(X \leq 3) = 0.9662$
   So $P(X < 4$ in each of three games$) = 0.9662^3$
   $= 0.902$ (3 s.f.)
   Assumption is that, for any game, the event 'fewer than four goals scored in game' is independent of the event 'fewer than four goals scored in a game' in any other game.

   ii. Let X = number of goals scored by this team per 90-minute game.
   X is Poisson ($\lambda = 1.2$) so $P(X \geq 4) = 1 - P(X \leq 3) = 0.0338$
   Let Y = number of games in which at least four goals are scored.
   Y is binomial ($n = 6$, $p = 0.0338$)
   $P(Y \geq 3) = 0.0007$
   Assumptions
   - probability of scoring at least four goals in a 90-minute game remains the same for all six games
   - independence – the scoring of at least four goals in a 90-minute game is not affected by whether at least four goals were scored (or not) in any other 90-minute game.