Chapter 6  Probability Distributions

1. A stockbroker makes 7 calls to clients each morning recommending stocks to purchases. Each call is a success or failure with a probability of success of 25% and the broker maintains this success rate. Using Excel (or similar spreadsheet) complete:

(i) A table of discrete probabilities and cumulative probabilities

(ii) A chart of the distribution

(iii) If his success rate increased to 40% what is the probability of at least 4 successes on a particular morning.

(i) 7 Number of calls
    0.25 Probability of success
    x Number of successes

<table>
<thead>
<tr>
<th>x</th>
<th>p(X = x)</th>
<th>p(X ≤ x)</th>
<th>p(X &lt; x)</th>
<th>p(X &gt; x)</th>
<th>p(X ≥ x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1335</td>
<td>0.1335</td>
<td>0.0000</td>
<td>0.8665</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.3115</td>
<td>0.4449</td>
<td>0.1335</td>
<td>0.5551</td>
<td>0.8665</td>
</tr>
<tr>
<td>2</td>
<td>0.3115</td>
<td>0.7564</td>
<td>0.4449</td>
<td>0.2436</td>
<td>0.5551</td>
</tr>
<tr>
<td>3</td>
<td>0.1730</td>
<td>0.9294</td>
<td>0.7564</td>
<td>0.0706</td>
<td>0.2436</td>
</tr>
<tr>
<td>4</td>
<td>0.0577</td>
<td>0.9871</td>
<td>0.9294</td>
<td>0.0129</td>
<td>0.0706</td>
</tr>
<tr>
<td>5</td>
<td>0.0115</td>
<td>0.9987</td>
<td>0.9871</td>
<td>0.0013</td>
<td>0.0129</td>
</tr>
<tr>
<td>6</td>
<td>0.0013</td>
<td>0.9999</td>
<td>0.9987</td>
<td>0.0001</td>
<td>0.0013</td>
</tr>
<tr>
<td>7</td>
<td>0.0001</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.0000</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

(ii) Chart of the distribution
(iii) 7  Number of calls
  0.4  Probability of success
  x  Number of successes

<table>
<thead>
<tr>
<th>x</th>
<th>p(X = x)</th>
<th>p(X ≤ x)</th>
<th>p(X &lt; x)</th>
<th>p(X &gt; x)</th>
<th>p(X ≥ x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0280</td>
<td>0.0280</td>
<td>0.0000</td>
<td>0.9720</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.1306</td>
<td>0.1586</td>
<td>0.0280</td>
<td>0.8414</td>
<td>0.9720</td>
</tr>
<tr>
<td>2</td>
<td>0.2613</td>
<td>0.4199</td>
<td>0.1586</td>
<td>0.5801</td>
<td>0.8414</td>
</tr>
<tr>
<td>3</td>
<td>0.2903</td>
<td>0.7102</td>
<td>0.4199</td>
<td>0.2898</td>
<td>0.5801</td>
</tr>
<tr>
<td>4</td>
<td>0.1935</td>
<td>0.9037</td>
<td>0.7102</td>
<td>0.0963</td>
<td>0.2898</td>
</tr>
<tr>
<td>5</td>
<td>0.0774</td>
<td>0.9812</td>
<td>0.9037</td>
<td>0.0188</td>
<td>0.0963</td>
</tr>
<tr>
<td>6</td>
<td>0.0172</td>
<td>0.9984</td>
<td>0.9812</td>
<td>0.0016</td>
<td>0.0188</td>
</tr>
<tr>
<td>7</td>
<td>0.0016</td>
<td>1.0000</td>
<td>0.9984</td>
<td>0.0000</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

2. What are the similarities and differences between the normal and the t-distributions?

Both are symmetric about the mean and for sample sizes of 30 or more the t-distribution converges to the normal distribution. For small samples, less than 30, the t-distribution is 'fatter' in the tail areas than the normal and it is lower than the normal at the mean or middle of the distribution.

3. Why does both the Chi-square and F-distributions have a lower bound of zero?

The Chi-square distribution is formed from the ratio of the sample variance to the population variance. A variance is obtained by squaring deviations, hence a variance will be non-negative. The ratio of two numbers that are non-negative will also be non-negative, i.e. $\chi^2 \geq 0$.

The F-distribution is formed from the ratio of two sums of squares, or the ratio of two Chi-square distributions, hence $F \geq 0$.

4. Why is the CLT (central limit theorem) so important for statistical theory?

The sampling distribution of the sample mean converges to a normal distribution regardless of the distribution of the original population, if the sample is large. The CLT enables us to analyse the means of many different random variables even when we don't know the actual population distributions of these variables. This enables statistical tests to be carried out using the normal distribution and distributions derived from the normal.

5. The mean success rate of an auctioneer in Melbourne during any month is 24 sales with a standard deviation of 6. Assuming that the auctioneer's performance can be described by a normal distribution.
(i) What is the probability that he successfully actions 30 properties in a given month?

\[ Z = \frac{30 - 24}{6} = 1 \]

The area from:

0 to 1 is 0.3413

total area is:

0.5 + 0.3413 = 0.8413

(ii) If the probability of successfully auctioning \( x \) properties is 0.67, how many properties will be sold?

From the normal tables the area required is: 0.67 - 0.5 = 0.17.
This corresponds to a \( Z \) value of 0.44, hence,

\[ Z = 0.44 = \frac{X - 24}{6} \quad \text{giving} \quad X = 24 + 0.44(6) = 26.64 \quad (27 \text{ properties}) \]

(iii) Using Excel (or similar spreadsheet) to create a random variable describing the number of properties sold in a given month. The variable should extend to at least 3 standard deviations of either side of the mean and also generate the corresponding density value. Use this information to chart the distribution.
6. Before adjustment, measurement errors for the height of aeroplanes above the earth (as measured by a certain altimeter) are normally distributed with a mean of zero and a standard deviation of one metre. These errors will be negative if the measured altitude is too low and positive if too high. Find the probability that for a given altimeter the error will be between:

(i) 0 and 1.65 metre  Probability = 0.451
(ii) 0 and 1.96 metres  Probability = 0.475
(iii) 0 and 2.33 metres  Probability = 0.4901
(iv) 0 and 2.58 metres  Probability = 0.4951
(v) -1.65 and 0 metres  Probability = 0.4505
(vi) -1.96 and 0 metres  Probability = 0.475
(vii) -2.33 and 0 metres  Probability = 0.4901
(viii) -2.58 and 0 metres  Probability = 0.4951
(ix) -1.65 and 1.65 metres  Probability = 2(0.4505) = 0.9010
(x) -1.96 and 1.96 metres  Probability = 2(0.4750) = 0.950
(xi) -2.33 and 2.33 metres  Probability = 2(0.4901) = 0.9802
(xii) -2.58 and 2.58 metres  Probability = 2(0.4951) = 0.9902

7. The location of a new hospital has been narrowed to a choice between two possible sites - I and II. An important consideration in site selection is the probability distribution of the travelling time involved in getting critically ill patients to the hospital. An analysis of the sites indicates that travelling time is normally distributed with a mean of 10.8 minutes (12 minutes) and a standard deviation of 3.1 minutes (2 minutes) for Site I (Site II).

It is considered crucial that travelling times be less than 15 minutes. What proportion of the travelling times would be less than 15 minutes for each of the two sites?

Site I:  \[ z = \frac{15 - 10.8}{3.1} = 1.355 \]  Probability = 0.5 + 0.4082 = 0.9082
Site II: \[ z = \frac{15 - 12}{2} = 1.50 \] Probability = 0.5 + 0.4332 = 0.9332

8. The time involved in commuting by tram from Flemington Racecourse to Latrobe Street (Melbourne) is a normally distributed variable. 95% of such commuter journeys take more than 26.75 minutes and 99% take less than 47.90 minutes. Determine the mean and standard deviation of this distribution.

The probability area from 26.79 to +\(\infty\) = 95%  
The area to the left of 0, in left tail, = 0.95 - 0.5 = 0.45  
This corresponds to a z value of -1.645

The area of the distribution up to 47.90 = 99%  
The area to the right of 0, in the right tail, = 0.99 - 0.5 = 0.49  
This corresponds to a z value of 2.33

This produces two simultaneous equations with unknows \(\mu\) and \(\sigma\):

(i) \[-1.645 = \frac{26.79 - \mu}{\sigma}\]

(ii) \[2.33 = \frac{47.90 - \mu}{\sigma}\]

From (i) \[-1.645\sigma = 26.79 - \mu\], giving \(\mu = 26.79 + 1.645\sigma\)

From (ii) \[2.33\sigma = 47.90 - \mu\], giving \(\mu = 47.90 - 2.33\sigma\)

Equating (i) and (ii): \[26.79 + 1.645\sigma = 47.90 - 2.33\sigma\]

\[47.90 - 26.79 = (2.33 + 1.645)\sigma\]

giving \[\sigma = \frac{47.90 - 26.79}{2.33 + 1.645} = 5.31\]

and \[\mu = 26.79 + 1.645(5.31) = 35.53\]
9. Management at the Speedy Freight Company is planning its budget for next quarter's overtime wages. The amount of time (X) for which the long haul drivers will exceed their trip schedules is normally distributed with a mean of 200 hours and a standard deviation of 40 hours. 

Use Excel to find:
(i) \( P(200 < X < 250) \)
(ii) \( P(180 < X < 200) \)
(iii) \( P(230 > X > 180) \)
(iv) \( P(190 > X > 150) \)

10. The probability that wiring is faulty in houses that were built prior to 1940 is 0.15. 

Required:
(i) Suppose one wished to model the probability distribution of the number of houses with faulty wiring from a set of 10 houses built prior to 1940. Then, under what circumstances would the Binomial probability model be an appropriate one to employ?
(ii) If the circumstances set out in (i) are satisfied, specify the mathematical form of the Binomial model that would apply in this instance.
(iii) Indicate the variance and standard deviation of the Binomial probability distribution identified (ii)
(iv) Evaluate the model presented in (ii) to determine the probability that 3 of the 10 house possess defective wiring.
(v) Use Excel to confirm the validity of your answer in (iv)
(vi) Use Excel to determine the probability that 3 or less of these 10 houses have defective wiring.
(vii) Use Excel to determine the probability that more than 3 of these 10 houses have defective wiring.

11. Management at a supermarket is considering whether it is worthwhile expanding the number of check out points. Data made available from the supermarket's internal records reveal that an average of 10 customers are served per check out point - hour. 

Required:
(i) Suppose one wished to model the probability distribution for the number of customers that are served per check out point - hour. Then, under what circumstances would the Binomial probability model be an appropriate one to employ?
(ii) If the circumstances set out in (i) are satisfied, specify the mathematical form of the Poisson model that would apply in this instance.
(iii) Indicate the variance and standard deviation of the Poisson probability distribution identified in (ii)
(iv) Evaluate the model presented in (ii) to determine the probability that 12 customers are served per check out point - hour.
(v) Use Excel to confirm whether your answer to (iv) is correct.

A Spreadsheet Approach to Business Quantitative Methods
(vi) Use Excel to determine the probability that 12 or less customers are served per check out point - hour.

(vii) Use Excel to determine the probability that more than 12 customers are served per check out point - hour.

12. A large agribusiness in Queensland owns 10000 hectares of land that it has set aside for the cultivation of a high cash crop. Based on historical experience the yield per hectare parcel is normally distributed with a mean of 120 bushels and a standard deviation of 36 bushels.

(i) Now suppose 9 of these one hectare parcels are selected randomly with replacement. What is the probability that the average yield on these parcels of land is:
- less than 132 bushels
- greater than 114 bushels
- between 114 and 138 bushels
- greater than 138 bushels
- less than 105 bushels

(ii) Recalculate the probabilities in (i) under the assumption that the 9 parcels were selected without replacement from a population of 100 rather than 10000 hectares.

(iii) Assuming a population size of 10000 hectare yields, construct an interval where the upper and lower bounds are equidistant from 120 bushels and the following statement is true:
There is a .99 probability that the mean yield of 9 one hectare parcels selected with replacement lies in the constructed interval.

(iv) Assuming a population size of 100 hectare yields, construct an interval where the upper and lower bounds are equidistant from 120 bushels and the following statement is true:
There is a .98 probability that the mean yield of 9 one hectare parcels selected without replacement lies in the constructed interval.

(v) Suppose the population of 10000 hectare yields is known to be normally distributed with a mean yield of 120 bushels. Its standard deviation however, is unknown. What then, is the probability that the average yield of 6 randomly selected one hectare parcels does not differ from 120 bushels by more than 2.02 estimated standard error units (i.e. units of \( s_x \) rather than \( \sigma_x \))

(vi) In (v) had the sample size been 26 rather than 6, find the probability that the average yield for the sample does not differ from 120 bushels by more than 1.71 estimated standard error units.

(vii) In (v) and (vi) what is observed regarding the relationship between the variability of the student t distribution and sample size?

(viii) For which sample size does the following statement hold with reasonable accuracy:
There is a .90 probability that the average yield per sampled parcel differs from 120 bushels (the population mean) by no more than 1.65 estimated standard error units.

(ix) Suppose, we have randomly selected 9 parcels with replacement from a population of 10000 and that their yields (measured in bushels) are given by: 110, 112, 115, 118, 120, 121, 122, 126, 127. Determine the sample mean $\bar{x}$, the sample standard deviation $s$ and $s_\bar{x}$ (the sample estimate of the true unknown standard error: $\sigma_\bar{x}$)

(x) Repeat (ix) under the assumption that $N = 100$ and sampling is undertaken without replacement.

13. A large nationwide supermarket chain is considering buying out Munchies which is a considerably smaller chain of 144 trendy snack bars. Before final negotiations are made for the buyout, management at the larger chain wishes to assure itself that these snack bars make good returns (i.e. after tax returns in excess of 12%). Accordingly, it intends to randomly select some of these snack bars and examine their after tax returns.

(i) Suppose a random sample of size 9 is chosen with replacement from the population of 144 snack bars. Moreover, assume that 64% ($= \pi$) of these snack bars earn after tax returns in excess of 12%. Then, determine the probability that the sample proportion is:
   - less than 72.8%
   - greater than 84.8%
   - greater than 52.0%
   - less than 60.0%
   - between 52.0% and 84.8%

(ii) Recalculate the above probabilities in (i) if a sample of 9 snack bars are randomly selected without replacement.

(iii) Assuming random sampling with replacement, construct an interval where upper and lower bounds are equidistant from $\pi = 64\%$ and the following statement is true:

   There is a .90 probability that the sample proportion of 9 snack bars lies in the constructed interval.

(iv) Assuming random sampling without replacement, construct an interval where upper and lower bounds are equidistant from $\pi = 64\%$ and the following statement is true:

   There is a .95 probability that the sample proportion of 9 snack bars lies in the constructed interval.

(v) Suppose we have randomly selected 9 snack bars with replacement from the population and that their after tax returns are given by: 11.1%, 11.6%, 11.8%, 12.1%, 12.07%, 13.0%, 13.5%, 13.6%, 13.8%. Determine the sample proportion of snack bars that have earned an after tax return in excess of 12% and use this sample proportion to obtain a sample point estimate ($s_p$) of the true unknown standard error of $p$ (i.e. $\sigma_p$).
(vi) Repeat (v) under the assumption that sampling is undertaken without replacement.