STATISTICS 3.14

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

Estimating probabilities using simulations

Example

Air Picton runs small nine-seater planes across Cook Strait. People must buy their ticket on the internet. *Air Picton* has found that in general 10% of the people who buy their tickets on the internet do not show up. The airline therefore 'overbooks' each flight by selling 10 tickets for each flight. Design a simulation to estimate the probability that more than nine people show up for any flight (one ticketed passenger has no seat available).

Solution

Problem: We wish to determine the likelihood that a ticketed passenger turns up for a flight to find no seat available, in other words the probability of a total of more than 9 passengers turning up.

Tool: the probability of a passenger not showing up is 10% $= \frac{1}{10}$, so use the calculator to generate random whole numbers from 1–10 (press 10 Ran# + 1 and take the whole number part) so that each number is equally likely.

Assigning outcomes: Let 1 represent a passenger not turning up; the numbers 2–10 will represent a passenger who turns up.

Trials: Generate a set of 10 numbers to represent the 10 passengers booked on a flight on a 9-seater plane. Repeat this trial 30 times.

Displaying results: Record the numbers in a table. For example, a trial with the numbers:

2,4,2,5,7,1,3,9,8,9 would represent nine passengers turning up (there was one 1, representing a passenger who did not turn up) so all passengers have seats.

Trial number	Numbers obtained	Number of passengers	More than 9 passengers
1	2,5,6,1,1,5,8,9,10,3	8	No
2	4,9,3,5,4,6,5,8,2,1	9	No
3	1,9,9,10,5,5,7,1,10,4	8	No
4	7,1,10,9,10,2,1,5,2,1	7	No
5	1,5,7,3,8,8,3,8,2,5	9	No
6	8,8,4,5,9,7,5,5,8,8	10	Yes
7	10,2,6,5,3,10,1,8,7,1	8	No
8	8,7,1,8,1,5,6,1,7,7	7	No
9	4,5,2,2,10,4,8,5,4,10	10	Yes
10	4,4,6,5,9,1,8,10,10,1	8	No
11	2,6,8,7,2,1,8,5,9,2	9	No
12	7,10,7,6,5,10,10,7,5,3	10	Yes
13	4,3,8,10,4,1,3,9,5,2	9	No
14	3,1,5,1,2,4,9,8,5,5	8	No
15	8,7,8,5,10,9,6,9,5,10	10	Yes
16	1,6,3,5,4,5,3,3,6,4	9	No
17	7,2,1,8,1,2,1,10,8,1	6	No
18	2,7,8,6,4,5,5,3,4,4	10	Yes
19	5,7,1,10,7,7,4,9,1,4	8	No
20	1,8,2,1,9,4,6,9,4,3	8	No
21	6,2,5,7,1,4,10,9,3,8	9	No
22	1,1,7,8,4,5,3,8,9,9	8	No
23	1,10,10,8,6,9,7,6,9,3	9	No
24	10,9,4,710,2,2,6,9,2	10	Yes
25	5,3,6,5,9,6,6,8,7,7	10	Yes
26	10,6,4,3,7,5,9,9,8,10	10	Yes
27	4,8,1,10,5,6,1,2,9,4	8	No
28	7,10,6,2,3,2,1,5,8,9	9	No
29	3,7,6,10,10,4,9,1,5	9	No
30	1,9,2,7,8,9,9,2,8,7	9	No

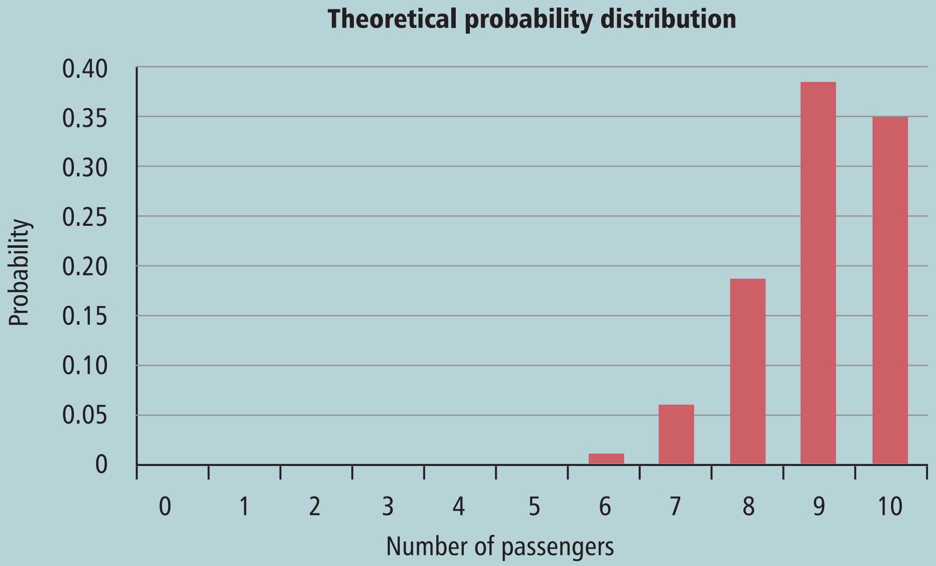
Calculate the proportion of trials for which ten customers showed up for a 9-seater flight (8 trials out of 30) to get the experimental probability that a ticketed passenger fails to get a seat on a flight $= \frac{8}{30}$ or 0.2667.

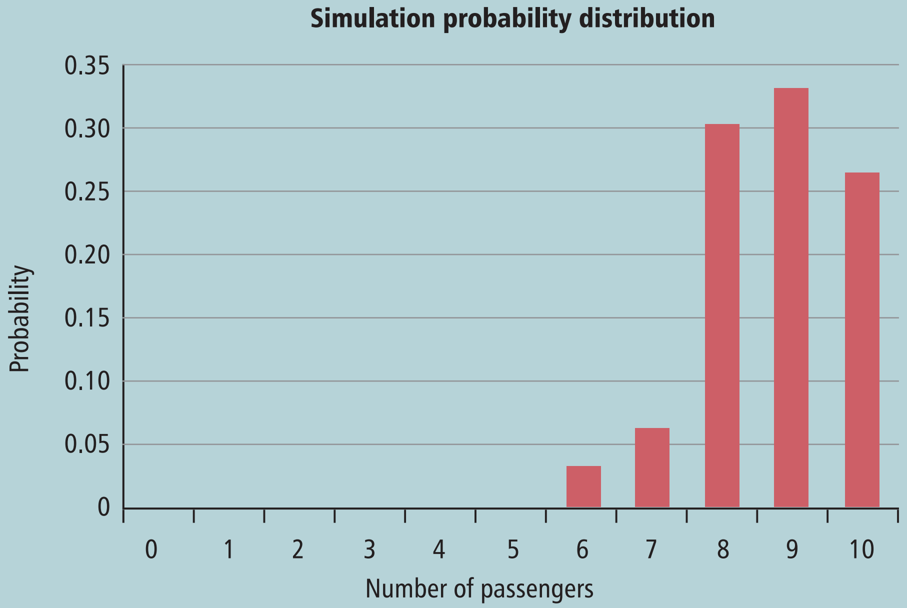
Assumptions and limitations: It is assumed that P(customer doesn't show up for flight) = 10% for each customer independently. This may not be accurate, if passengers have booked as a couple or group, as one customer not turning up makes it more likely that another customer will also not turn up). To improve the simulation, investigate the frequency of couple or group bookings, and introduce this element to the simulation. Other factors influencing customer 'no-shows' could also be investigated, e.g. weather, time since the booking was made, etc. as these may alter the 10% probability.

Theoretical probability: Modelling this situation with the binomial distribution (X = number of 'no shows'; $n = 10$, $p = 0.1$), the theoretical probability of a flight having ten customers is $P(X = 0) = 0.3487$. By comparison, the simulation probability of 0.2667 is somewhat lower than expected.

This variability is to be expected with a relatively small sample size of 30 (repeating the simulation would result in a different probability being obtained). If this simulation were to be repeated a large number of times (using computers it is feasible to carry out simulations with thousands of repeats), the average probability obtained from the simulations would be expected to be closer to the theoretical probability.

The distribution of simulation probabilities can be compared graphically with the theoretical probability distribution.





The simulation has a probability distribution similar in shape to that of the theoretical (binomial) probability distribution, with some variations in individual probabilities (notably the simulation has a higher probability of 8 passengers than does the theoretical probability).

This variability is to be expected with only 30 trials – another simulation would give a different distribution.

Exercise: Simulations

In the following investigations:

- Design and describe a simulation to model the investigation. You must include details of the tool you chose and why, relating this to the theoretical probabilities.
- Carry out the simulation.
- Discuss your findings, giving the probabilities determined by your investigation.
- Discuss any limitations and assumptions that were made and suggest any possible improvements to your simulation.
- Compare your experimental probabilities with theoretical probabilities, using calculations and graphs (where appropriate).

1. Design and analyse a simulation to determine the probability that a three-child family will have exactly one girl.

2. Design and analyse a simulation to determine the probability that in a group of 20 people, two or more people have their birthdays on the same day.

3. A multiple-choice test consists of 10 questions, each of which has one correct answer and three incorrect answers. In order to pass the test you must get at least 5 questions right. Unfortunately you know nothing about the subject being tested and have to guess your answers.

Design a simulation to estimate your chances of passing.

4. The siren has gone to mark the end of a basketball match; however, just before the siren, the shooter of Team A was fouled and so wins two free throws. In order to win the match, one of the free throws must be successful. From the shooter's past experience it is known that two thirds of his free throws are successful.

Design a simulation to estimate the probability of Team A winning.

5. In a *Fear Factor* challenge three almost identical keys are given to a contestant, who must then dive underwater to release two locks in succession (i.e. he cannot release the second lock until he has unlocked the first) to free his partner.

Design a simulation to determine the probability that he undoes each lock on the first attempt. (Assume that the key that unlocks the first lock may be used again, i.e. there are three choices of key for both locks.)

6. On 10% of flights into Wellington Airport a jet flies into a flock of birds resulting in the chance of engine failure due to birdstrike. The jet has four engines: if two suffer birdstrike then the plane has to make an emergency landing; if three or more suffer birdstrike then the plane will crash.

Assuming that if the plane flies into a flock of birds the probability of birdstrike is 0.5, and that birdstrike is independent for each engine, design a simulation to determine the probability that a jet will have to make an emergency landing and the probability that a jet will crash.

7. At a bank there are three tellers operating. Customers form a single queue and are served as a teller becomes free. At a particular bank, customers arrive randomly at a rate of between 1 and 6 per minute and each teller takes 1 minute to deal with a customer

Assuming there are no customers to start with, design a simulation to determine the probability that during a 20-minute period, there will be two or more people in the queue at the end of a minute period.

Answers

Exercise: Simulations

Answers will vary; examples follow.

1. Outcomes: child is a boy or child is a girl.

   Assume the gender of a child is independent of the gender of the other children and that the theoretical probability that a child is a girl is $\frac{1}{2}$

   Tool: use the random number generator of a calculator (or use a coin). Press 2Ran# + 1 and take the whole-number part (this ensures only the numbers 1 and 2 are generated).

   Assign 1 to represent a girl and 2 to represent a boy.

   Trial: generate three random numbers to represent the three children in the family. A trial is a 'success' if there is exactly one girl, e.g. if the results of a trial are (1, 1, 2) this would represent 2 girls and 1 boy (and so the trial would be labelled 'unsuccessful'). Run 30 trials.

   Calculate the probability by dividing the number of successes by the number of trials, e.g. if there were 8 successful trials (out of 30), then the probability of a three-child family having exactly one girl would be $\frac{8}{30}$.

   In theory, the probability of having exactly one girl is $\frac{3}{8}$ (using a probability tree or the binomial distribution) hence the probability derived from the simulation would be less than that predicted theoretically.

   Graphs of the experimental and theoretical probability distributions could be included and comments made on their features and similarities.

2. Work out on what day of the year the birthday falls for each person in the group (numbers from 1–365), e.g. a person whose birthday is on February 5th was born on day 36. The probability a birthday is on a particular day is $\frac{1}{365}$.

   Tool: use a calculator to generate random numbers between 1 and 365 by pressing 365 Ran# + 1, and taking the whole-number part.

   Trial: generate 20 random whole numbers to represent the days of the year for the birthdays of the 20 people in the group. A trial is a 'success' if the same number appears more than once in a trial, as this will represent two people having their birthday on the same day.

   Run 30–50 trials.

   Calculate the probability by dividing the number of successes by the number of trials.

   Results will vary for simulations.

   The theoretical probability of two or more people out of 20 having the same birthday is 0.4114 (this can be worked out by finding the probability that no people have a matching birthday, then subtracting the result from 1 – or you can research the net). Compare this with your results from the simulation.

   As there were only 30–50 trials, results from the simulation are more affected by the random nature of probability. If this simulation were repeated a very large number of times, the probability found from the simultation would tend to be close to that found theoretically.

   We have assumed that each birth date is equally likely, that no birthday falls on 29 February (in a leap year), and that the group does not have some connection (e.g. twins) that would make it less likely that birthday dates are independent.

3. The probability of getting a question correct is $\frac{1}{4}$.

   Tool: use a calculator to generate random whole numbers between 1 and 4 (press 4 Ran# + 1, taking whole-number part).

   Trial: generate 10 whole numbers between 1 and 4 to represent the answers to a test of 10 questions.

   Assign: the number 1 will represent getting a question correct, the numbers 2, 3, 4 will represent getting the question incorrect.

   A trial is 'successful' if there are at least five 1's among the ten numbers.

   Repeat the trial 30–50 times.

   Calculate the probability by dividing the number of successes by the number of trials.

   Results will vary for simulations.

   Theoretically: using the binomial distribution, the theoretical probability of getting at least 5 questions correct is 0.07813.

   Assumptions: that the questions are answered randomly and independently, and that there is no pattern to the letters of the answers.

   Graphs of the experimental and theoretical probability distributions could be included and comments made on their features and similarities.

4. The probability of a successful free throw is $\frac{2}{3}$.

   Tool: use a calculator to generate random whole numbers from 1 to 3 (press 3 Ran# + 1, taking whole-number part).

   Assign: 1 and 2 to represent scoring a goal and 3 to represent missing.

   Trial: generate two random numbers to represent the two shots at goal. A successful trial will be one in which either a 1 or a 2 appears at least once.

   Repeat the trial 30–50 times.

   Calculate the probability by dividing the number of successes by the number of trials.

   Results will vary for simulations.

   Theoretically: using a probability tree (or the binomial distribution), the theoretical probability of success (i.e. at least one successful shot) is $\frac{8}{9}$.

   Assumptions: shots are independent – success or failure on the first attempt does not impact on the probability of success/failure in the second attempt. Probability of a success is correctly estimated from previous success rates, and is constant.

   Graphs of the experimental and theoretical probability distributions could be included and comments made on their features and similarities.

5. The probability of choosing a correct key is $\frac{1}{3}$.

   Tool: use a calculator to generate random whole numbers from 1 to 3 to represent the key chosen (press 3 Ran# + 1, taking whole-number part).

   Assign: 1 to represent getting the correct key and 2 and 3 to represent getting the wrong key.

   Trial: Generate a random number:

   - if it is 2 or 3, then the first lock has not been unlocked and so this is immediately an unsuccessful trial and no further action is required
   - if it is a 1, then this will represent opening the first lock and so generate a second random number (3Ran# +1 as before) to represent the attempt at the second lock – again if this is a 1 this will represent successfully opening the second lock. Generating two 1's in succession will represent a successful trial – i.e. opening both locks first time.

   Repeat trials 30–50 times.

   Calculate the probability by dividing the number of successes by the number of trials.

   Results will vary for simulations.

   Using the multiplication rule, the probability of success first time for each lock is $\frac{1}{9}$.

   Graphs of the experimental and theoretical probability distributions could be included and comments made on their features and similarities.

6. The probability of a jet flying into a flock of birds is $\frac{1}{10}$.

   Tool: use the calculator to generate random numbers 1–10 (press 10 Ran# + 1, taking whole-number part).

   Assign: let 1 represent flying into a flock of birds, and 2–10 represent not flying into a flock.

   Trial: generate a number to represent birdstrike or not for the plane. If 2–10 is obtained, then no birdstrike occurs and the trial finishes.

   If a 1 is obtained, then birdstrike has occurred, so generate another four random numbers to represent birdstrike or not for the four engines (press 2 Ran# +1, taking whole number part; assign 1 to represent birdstrike (engine failure), assign 2 to represent no birdstrike – as the probability of an engine failing is 0.5).

   If, in the four numbers generated, three or more 1's are obtained the plane will crash. If two 1's are obtained the plane will make an emergency landing.

   Calculate the probabilities by dividing the number of 'successes' by the number of trials.

   The theoretical probability of the plane making an emergency landing is 0.0375, while the probability of crashing is 0.03125.

   Assumptions: independence of birdstrike for an engine – in reality if an engine suffers from birdstrike, it would be more likely that the other engines will also be hit.

7. As each potential number of customers arriving is equally likely, each number of customers has a probability of $\frac{1}{6}$.

   Tool: use a calculator to generate random numbers 1–6 (press 6 Ran# + 1, taking the whole-number part).

   Assign: 1 represents 1 customer arriving, 2 represents 2 customers arriving, etc.

   Trial: for each minute, generate a single random number to represent the number of customers arriving by the end of that minute – if the number generated is greater than 3 then there will be a queue forming – the number in the queue will need to be carried forward to the next minute (the numbers reduce by 1 per minute per teller as they are served). A trial is a success if there are at least 2 people in the queue at the end of a minute. Repeat 20 times to represent the 20-minute period.

   Calculate the probability by dividing the number of successes by the number of trials.

   Results will vary for simulations.

   Assumptions: that customers are arriving independently – the arrival of one customer has no effect on the arrival of another customer; that arrival rate stays constant – customers are not put off by a long queue.