Year 12 Mathematics EAS 2.12

Probability Methods

Robert Lakeland & Carl Nugent

Contents

- Achievement Standard 2
- Theoretical Probability 3
- Relative Frequency 4
- Informal Conditional Probability 6
- Multiplicative Principle and Expected Number 11
- Tree Diagrams 16
- Risk, Absolute and Relative 22
- The Standard Normal Distribution 28
- Inverse Normal – Calculating X Values 43
- Inverse Normal – Calculating Parameters 48
- Experimental Distributions 52
- Practice External Assessment 55
- Answers 62
- Normal Distribution Table 70
- Order Form 71

νuLake Ltd
Innovative Publisher of Mathematics Texts

Probability Methods 2.12

This achievement standard involves applying probability methods in solving problems.

Achievement	Achievement with Merit	Achievement with Excellence
• Apply probability methods in solving problems.	• Apply probability methods, using relational thinking, in solving problems.	• Apply probability methods, using extended abstract thinking, in solving problems.

- ◆ This achievement standard is derived from Level 7 of The New Zealand Curriculum and is related to the achievement objectives
  - ❖ evaluate statistically based reports
    - interpreting risk and relative risk
  - ❖ investigate situations that involve elements of chance
    - comparing theoretical continuous distributions, such as the normal distribution, with experimental distributions
    - calculating probabilities, using such tools such as two-way tables, tree diagrams

  in the Statistics strand of the Mathematics and Statistics Learning Area.

- ◆ Apply probability methods in solving problems involves:
  - ❖ selecting and using methods
  - ❖ demonstrating knowledge of probability concepts and terms
  - ❖ communicating using appropriate representations.

- ◆ Relational thinking involves one or more of:
  - ❖ selecting and carrying out a logical sequence of steps
  - ❖ connecting different concepts or representations
  - ❖ demonstrating understanding of concepts;

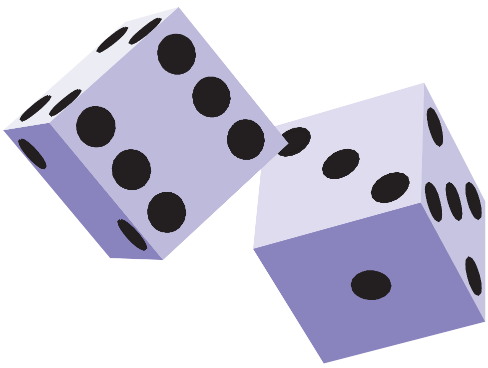
  and also relating findings to a context or communicating thinking using appropriate statements.

- ◆ Extended abstract thinking involves one or more of:
  - ❖ devising a strategy to investigate or solve a problem
  - ❖ identifying relevant concepts in context
  - ❖ developing a chain of logical reasoning
  - ❖ making a statistical generalisation;

  and also where appropriate, using contextual knowledge to reflect on the answer.

- ◆ Problems are situations which provide opportunities to apply knowledge or understanding of mathematical and statistical concepts. Situations will be set in real-life or statistical contexts.

- ◆ Methods include a selection from those related to:
  - ❖ risk and relative risk
  - ❖ the normal distribution
  - ❖ experimental distributions
  - ❖ relative frequencies
  - ❖ two-way tables
  - ❖ probability trees.



Theoretical Probability

Theoretical Probability

Theoretical probability is where we have prior knowledge of possible outcomes and the physical situation allows us to deduce or infer the relevant probabilities.

E.g. The probability of drawing an Ace from a pack of cards is $\frac{4}{52} = \frac{1}{13}$ as there are four aces out of 52 cards.

Similarly, if we throw a fair coin, we can predict with certainty that it will fall with heads or tails uppermost. We say that the coin has a probability of $\frac{1}{2}$ of landing with heads uppermost, because there are two equally likely outcomes.

To calculate the probability of an event occurring we begin by considering an 'experiment', namely throwing a die.

The possible outcomes or **sample space** when throwing a die are {1, 2, 3, 4, 5, 6}. The number of equally likely outcomes is six.

If we wish to calculate the probability of getting a number greater than four (i.e. 5 or 6) when throwing a die, then the number of correct outcomes is two.

Therefore

$$P(\text{result} > 4) = \frac{\text{Number of correct outcomes}}{\text{Total number of outcomes}}$$

$$= \frac{2}{6} \text{ or } \frac{1}{3}$$

Probabilities vary from 0 (impossible) to 1 absolutely certain.

Probability Concepts

We use the word **probability** to represent the **chance** of an event occurring, founded on certain evidence.

We begin by identifying some basic **probability** facts.

- All probabilities vary from 0 to 1 inclusive.
- An impossible event has a probability of 0.
- An event which is a certainty has a probability of 1.
- We denote the probability of an event A occurring, by using the notation P(A). If A represents the event occurring then A′ is the event not occurring. The probability of it not occurring is

$$P(A') = 1 - P(A)$$

Extra Notes

When we write probabilities they can be represented using fractions or decimals (although people commonly also use percentages).

E.g. $P(\text{event}) = \frac{1}{2}$ (0.5) (50%)

Achievement – Answer the following questions.

3. An insurance company analysed 909 policies to see if there was a pattern to the claims per year.

	Claims			
	Nil	1 – 2	3+	Total
Under 25	175	64	17	256
25 and over	520	112	21	653
Total	695	176	38	909

a) What is the probability of a policy holder under 25 years old making one or more claims?

b) What is the probability that a policy holder making 3+ claims per year was 25 years or older?

c) What is the probability that a policy holder making 1–2 claims per year was under 25 years?

4. Long term departures (migrants) from New Zealand.
Figures from Statistics New Zealand June 2016.

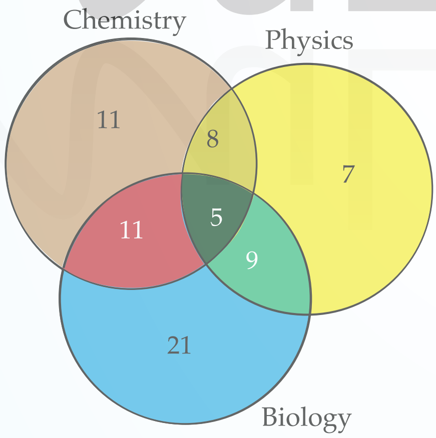
	2015	2016	Total
Australia	25 246	23 770	49 016
Pacific Is.	1 756	1 717	3 473
Asia	8 840	8 198	17 038
Europe	13 717	14 227	27 944
Americas	5 444	5 845	11 289
Other	2 393	2 209	4 602
Total	57 396	55 966	113 362

a) Find the probability that a migrant went to Australia in either year.

b) What is the probability that a migrant in 2015 went to the Pacific Islands?

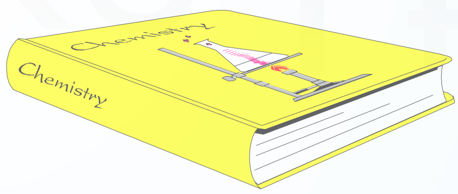
c) If a migrant went to Australia, what is the probability that they went in 2015?

5. A school has 72 students who study at least one science subject as shown below.



a) Find the probability that one of these students takes biology.

b) What is the probability that a physics student also takes chemistry?

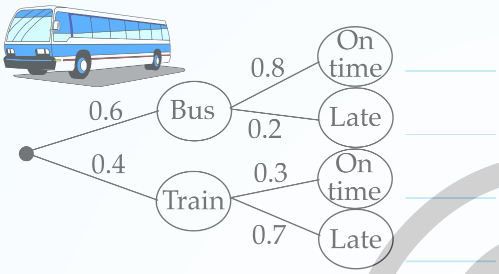


c) What is the probability that a biology student also takes chemistry?

d) What is the probability that a biology student also takes physics and chemistry?

Achievement /Merit – Answer the following tree diagram problems.

23. A person can travel by train or bus. Their probabilities and outcomes are on the tree diagram below.



a) Complete the probability tree.

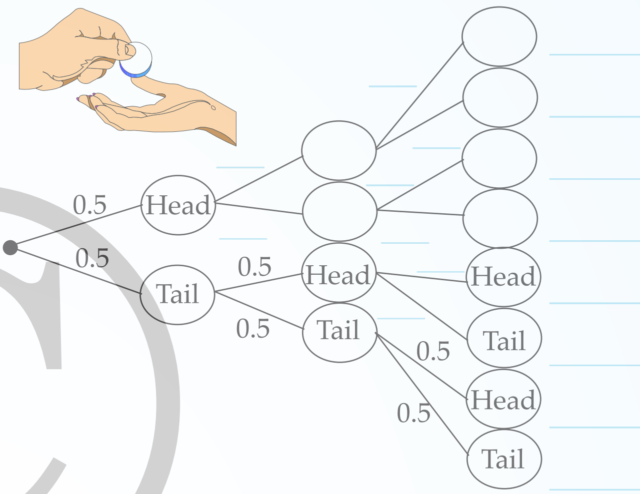
Find the probability of a person

b) travelling by bus and arriving late.

c) travelling by train and being on time.

d) being on time by bus or train.

24. Three coins are flipped and some of the outcomes are shown on the probability tree below.



a) Complete the probability tree.

Find the probability of getting

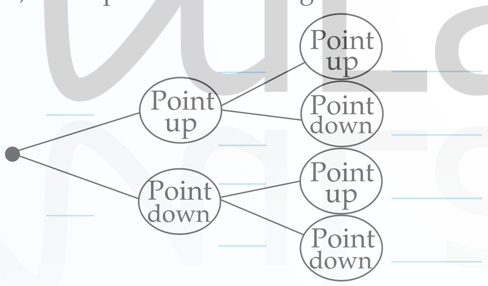
b) three heads.

c) two heads and one tail (in any order).

d) at least one head.

25. A drawing pin has a probability of landing point up of 0.7. Two drawing pins are dropped.

a) Complete the tree diagram.



Find the probability of the following results

b) that a single drawing pin will land point down.

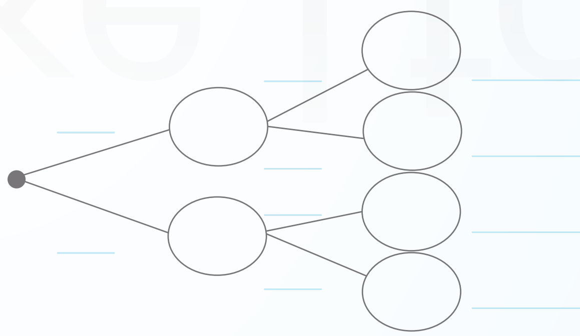
c) two will land point up.

d) two will land point down.

e) two will land with 1 point up and 1 point down (in any order).

f) A handful of similar drawing pins had a probability of all landing point up of 0.1681. How many drawing pins were dropped at once?

26. The probability that a basketball player gets his first free throw in is 0.6. If the first goes in, the probability the second goes in is 0.7 otherwise the probability of the second going in drops to 0.5.



a) Complete the probability tree to represent this situation.

Find the probability of getting

b) both throws in.

c) neither throw in.

d) at least one throw in.

Risk

Risk

In statistics, risk is the probability or likelihood of an event occurring. Risk is often expressed as a decimal or percentage.

When you attempt something that could result in harm or injury you are putting yourself at risk. Risk is generally considered to have this negative connotation so in situations where this is not the case you should look to use another term to avoid confusing readers. For example, if you were looking at the data for two groups gaining employment it would be inappropriate even if mathematically correct to use the term 'risk of getting a job' as you are likely to confuse your audience. You could just use the term 'probability of getting a job'. There are many different kinds of risk: business risk, economic risk, health risk etc.

In probability we can define two types of risk: absolute risk and relative risk.

Absolute Risk

Absolute risk is the probability or percentage of subjects in any group that experience an outcome. For example, the absolute risk of a person developing lung cancer over their lifetime could be expressed as 0.11 or 11% or 1 in 9.

Relative Risk

Relative risk is used to compare the risk in two sub groups of our population. For example, we could measure the risk of people getting lung cancer as a consequence of smoking by comparing the risk of getting lung cancer for people who smoke with the risk of getting lung cancer for the group that do not smoke.

We can calculate the relative risk of getting lung cancer as a consequence of smoking (or the risk ratio) by using the formula

$$RR = \frac{\text{Risk or Prob. with condition present}}{\text{Risk or Prob. with no condition present}}$$

Where RR is the relative risk.

Consider the table below which shows that the risk of getting lung cancer among smokers is 25% while only 2% among non-smokers.

	Lung cancer	No lung cancer
Smoker	0.25	0.75
Non-smoker	0.02	0.98

To calculate the relative risk of getting lung cancer as a consequence of smoking we use the formula given above. The probability of a smoker being exposed to cancer is 0.25 while the probability of a non-smoker being exposed to cancer is 0.02.

Therefore the relative risk $= \frac{0.25}{0.02} = 12.5$.

So smokers are 12.5 times more likely than non-smokers to develop lung cancer.

Relative risk is a value that identifies how much something you do, such as maintaining a healthy weight, can change your risk compared to the risk if you're very overweight.

Values for relative risk (RR) can be any number greater than zero. If your relative risk (RR) value is 1 it means the outcome is not influenced by the factor present, i.e. the likelihood of lung cancer is not affected by smoking.

If your relative risk (RR) value is greater than 1 it means the outcome is greater when the factor is present, i.e. you are more likely to get lung cancer if you smoke.

If your relative risk (RR) is less than 1 it means the outcome is less likely when the factor is present. An example may be the relative risk of having a heart attack when taking a low dose of aspirin compared to having a heart attack and not taking any aspirin.

	Heart Attack	No heart attack
Takes aspirin	0.94%	99.06%
No medication	1.71%	98.29%

If you now calculated the relative risk of having a heart attack with or without a low dose of aspirin then

$$RR = \frac{\text{Risk of heart attack with aspirin}}{\text{Risk of heart attack with no aspirin}}$$

$$RR = \frac{0.94}{1.71}$$

$$RR = 0.55$$

A relative risk less than 1 is correct but difficult to interpret so often the problem is restated so that the increased risk is the numerator and the relative risk is then greater than 1.

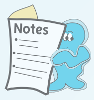
In this case we would state that the relative risk of having a heart attack by taking no medication compared to taking a low dose of aspirin is

$$RR = \frac{\text{Risk of heart attack with no aspirin}}{\text{Risk of heart attack with aspirin}}$$

$$RR = \frac{1.71}{0.94}$$

$$RR = 1.82$$

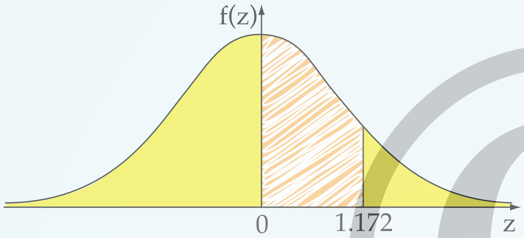
The risk of a heart attack by taking no medication is 1.8 times the risk compared to taking a low dose of aspirin.



Using Standard Normal Tables (see Page 70)

The standard normal table gives the area under the curve between 0 and Z.

For example, to calculate the probability of a result from $0 < Z < 1.172$ first we sketch a quick bell shaped curve to make sure we are calculating the correct area.



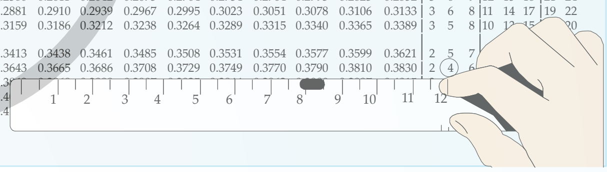
We then place a ruler on the tables corresponding to the row that starts 1.1. If we mark the ruler with a black mark we can position this under the 7 column (0.3790). Now from the difference column we get the difference corresponding to the digit 2 which is a 4 and we add this to 0.3790 to get 0.3794.

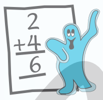
We can calculate all other areas under the normal curve by using the symmetry of the curve and the fact that the total area under the curve is 1.
(Note: Half the area is 0.5)

For example, to calculate the probability of a result $Z > 1.172$ we need the area outside of 1.172. To get this we subtract from 0.5 the area up to 1.172.

$$\begin{aligned} P(Z > 1.172) &= 0.5 - P(0 < Z < 1.172) \\ &= 0.5 - 0.3794 \\ &= 0.1206 \end{aligned}$$

z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	4	8	12	16	20	24	28	32
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	4	8	12	16	20	24	28	32
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	4	8	12	15	19	23	27	31
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	4	8	11	15	19	23	26	30
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	4	7	11	14	18	22	25	29
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	3	7	10	14	17	21	24	27
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	3	6	10	13	16	19	23	26
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2793	0.2823	0.2852	3	6	9	12	15	18	21	24
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	3	6	8	11	14	17	19	22
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	3	5	8	10	[illegible]	[illegible]	[illegible]	20
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2	5	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2	4	6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



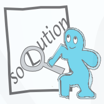


Example (using Standard Normal Tables)

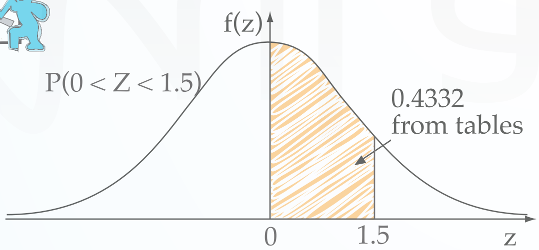
Using the normal distribution table, calculate the area under the standard normal curve for each of the following. Illustrate each one by drawing a sketch of the required area on the normal curve.

a) $P(0 < Z < 1.5)$ b) $P({}^-0.6 < Z < 1.5)$

c) $P(Z > 1.5)$ d) $P(1.2 < Z < 1.5)$

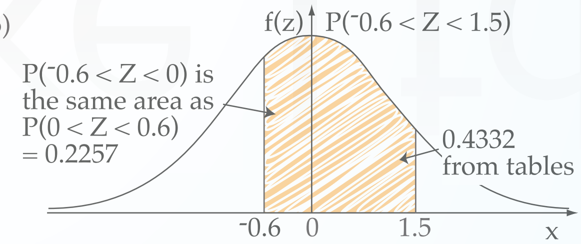


a)



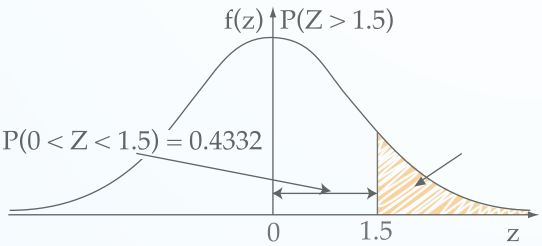
$P(0 < Z < 1.5) = 0.4332$

b)



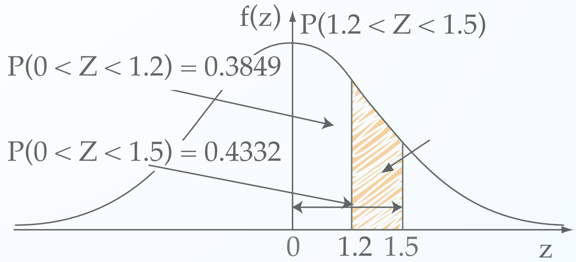
$$\begin{aligned} P({}^-0.6 < Z < 1.5) &= P({}^-0.6 < Z < 0) + P(0 < Z < 1.5) \\ &= 0.2257 + 0.4332 \\ &= 0.6589 \end{aligned}$$

c)

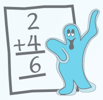


$$\begin{aligned} P(Z > 1.5) &= P(Z > 0) - P(Z < 1.5) \\ &= 0.5 - 0.4332 \\ &= 0.0668 \end{aligned}$$

d)



$$\begin{aligned} P(1.2 < Z < 1.5) &= P(0 < Z < 1.5) - P(0 < Z < 1.2) \\ &= 0.4332 - 0.3849 \\ &= 0.0483 \end{aligned}$$



Example (Using graphics calculator)

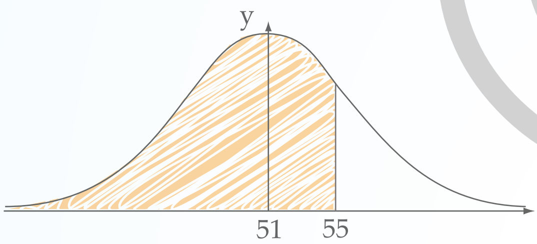
The mean mark in a school's Year 12 mathematics exam was 51% with a standard deviation of 16%. Assuming the marks are normally distributed find the probability that a student selected at random scored

a) less than 55%.

b) between 35% and 45%. Hence find how many out of 250 students you expect to get between 35% and 45%.

Using your TI-84 Plus

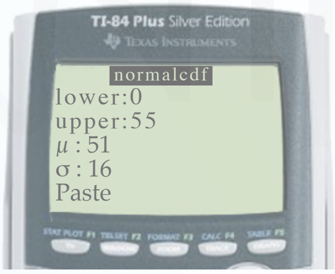
Even when using your calculator sketch at least one normal curve so you can visualise the area you require. We can see we expect an answer over 0.5.



Access the distribution menu and either scroll down or select 2 to get normalcdf(.

| 2nd | DISTR | 2 |

a) **With P(X < 55) and μ = 51, σ = 16, enter lower : 0, upper : 55, μ : 51 and σ : 16.**



This gives 0.5979886673 or 0.5980 (4 dp).

b) **The sketch for part b) is on the right. For P(35 < X < 45) and μ = 51, σ = 16. Enter lower : 35, upper : 45, μ : 51 and σ : 16.**

This gives 0.1952 (4 dp).

Number of students = 0.1952 x 250
= 48 (rounding down)

Using your Casio 9750GII

The sketch for part a) is on the left. It is important that you sketch at least one normal curve so you can visualize the area you require.

From the Main menu select STAT then DIST, NORM and Ncd.

STAT	DIST	NORM	Ncd
2	F5	F1	F2

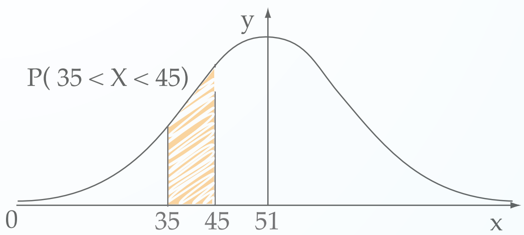
For the Casio use the format Lower, Upper, σ and μ.

a) **With P(X < 55) with σ = 16 and μ = 51. enter Lower : 0, Upper : 1.5, σ : 16 and μ : 51.**

| 0 | EXE | 5 | 5 | EXE | 1 | 6 |
| EXE | 5 | 1 | EXE | EXE |

getting the probability = 0.5980 (4 dp).

b) **For P(35 < X < 45)**



Enter Lower : 35, Upper : 45, σ : 16 and μ : 51.

| 3 | 5 | EXE | 4 | 5 | EXE |
| 1 | 6 | EXE | 5 | 1 | EXE | EXE |

This gives 0.1952 (4 dp).

Number of students = 0.1952 x 250
= 48 (rounding down)

Experimental Distributions

Probability Distributions from Experiments

A probability experiment is a repeated set of trials in which outcomes are collected. An example would be the result of flipping five coins or the score of a large group in a test.

The data from the experiment (the outcomes) can be used to determine the probabilities of certain outcomes.

For example, say your experiment was based on Yahtzee. With five dice and up to three throws, what is the probability of getting a particular result? To collect data the trial (throwing five dice, holding any sixes and throwing the others and repeating the previous step with any dice which are not six) would be repeated a large number of times.

On the basis of the recorded outcomes we make probability predictions.

Outcome	0	1	2	3	4	5
Freq.	3	15	29	37	11	5

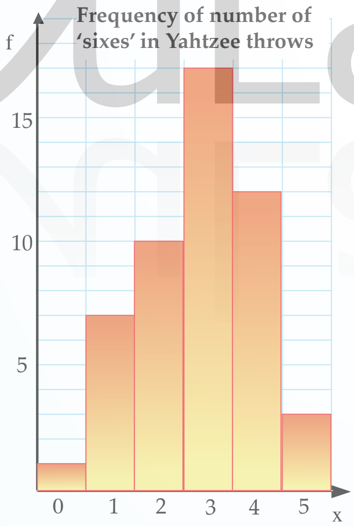
On the basis of this experiment the probability that you would get a result of four or five 'sixes' in three Yahtzee would be

$$P(4 \text{ or } 5 \text{ 'sixes'}) = \frac{16}{100}$$
$$P(4 \text{ or } 5 \text{ 'sixes'}) = 0.16$$

We use the outcome of the experiment to determine probabilities rather than approximate the results with a model such as the normal distribution.

Example

A class was interested in the outcomes when attempting to roll 'sixes' according to the Yahtzee rules. Five dice are rolled, the 'sixes' removed and the remaining dice rolled again. Any new 'sixes' are removed and the remaining dice rolled a third time. The total number of 'sixes' is recorded. The 25 class members each did this twice and recorded all 50 results.



What proportion of throws had at least 4 'sixes'?

There are 50 different outcomes and 12 + 3 outcomes of 4 'sixes' or more.

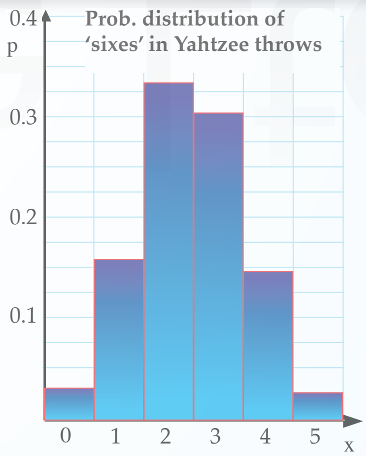
$$P(4 \text{ plus 'sixes'}) = \frac{15}{50}$$
$$P(4 \text{ plus 'sixes'}) = 0.3$$

Example cont...

The teacher put together a spread sheet to repeat the Yahtzee throws for 'sixes' 5000 times and the result was:

Sixes	0	1	2	3	4	5
Prob.	0.0314	0.165	0.3296	0.3032	0.1458	0.025

Compare the frequency histogram of 50 throws (on the left) to this probability distribution. You should comment on the comparative shape, centre and spread of the two distributions.



The outcomes from 50 throws were skewed left (tail to left) with a large peak of results for 3. The probability distribution is almost symmetrical around the 2 – 3 boundary. The experimental probability of 4 plus 'sixes' was 0.3 but this dropped to 0.1708 for the probability distribution. The median number of throws has dropped from 3 to 2 but the spread of outcomes is similar for both experiments.

Page 26 Q42 cont...

d) RR = 4.1

e) RR = 0.24

f) An obese person is 6.6 times more likely to have diabetes than a person who is not obese. There appears to be causal link between obesity and diabetes.

43. RR = 9

44. a) Cycling

b) They have a fatality rate less than cycling.

c) 3.3 times

Page 27

45. a) i) $\frac{306}{435}$ (0.70)

ii) 7034 per 10 000

b) $\frac{180}{225}$ (0.8)

c) $\frac{126}{210}$ (0.6)

d) $\frac{0.6}{0.8}$ (0.75)

e) The risk of having an asthma attack for those taking the trialled drug is 0.75 times the risk of those taking the placebo.

f) $\frac{0.8}{0.6}$ (1.33 (3 sf))

g) The risk of having an asthma attack for those on the placebo is 1.3 times the risk of those taking the trialled drug.

h) Placebo. It is better to compare the risk of the drug group (treatment) with the placebo group (non-treatment).

i) $\frac{0.6-0.8}{0.8} \times 100\% = {}^{-}25\%$

j) There is a 25% decrease in the chance of a patient having an asthma attack if they are taking the trialled drug compared to the placebo.

Page 31

Answer from tables (calculator Ans. in brackets).

46. a) 0.4554

b) 0.4222 + 0.4032 = 0.8254

Page 32 Q46 cont...

c) 0.5 – 0.4538 = 0.0462

d) 0.5 – 0.4268 = 0.0732 = (0.0733)

e) 0.5 + 0.2197 = 0.7197

f) 0.4505 – 0.4066 = 0.0439

47. a) 0.5 – 0.2190 = 0.2810

b) 0.4896 – 0.4525 = 0.0371 = (0.0370)

Page 33 Q47 cont...

c) 0.5 + 0.4694 = 0.9694 = (0.9693)

d) 0.4332 + 0.4332 = 0.8664

e) 0.5 + 0.2245 = 0.7245 = (0.7244)

f) 0.4756 + 0.1646 = 0.6402

48. a) 0.4500

b) 0.0500

c) 0.4276

d) 0.3530

e) 0.3149

f) 0.9483

Page 34 Q48 cont...

g) 0.2180

h) 0.4712

i) 0.8901 (0.8900)

j) 0.1806

Page 37

49. a) 0.4522

b) 0.0478

c) 0.3944

d) 0.2906 (0.2907)

e) 0.9522

50. a) 0.2412

b) 0.2247

c) 0.2016 (0.2017)

d) 0.5932 (0.5934)

e) 0.9559 (0.9560)

51. a) Z = 1.000

$P(85 < X < 115) = 0.6826$ (0.6827)

b) Z = 1.333 to Z = 2

$P(120 < X < 130) = 0.0685$

c) P = 0.0478. Expect 16 or 17.

d) P = 0.0310. Expect 10 or 11.

Page 38

52. a) p = 0.3545 (0.3546)

b) p = 0.0391 (0.0392)

c) p = 0.0391 (0.0392)

d) p = 0.8545 (0.8546)

e) p = 0.2752

53. a) p = 0.4136 (0.4137)

b) p = 0.0864 (0.0863)

c) p = 0.7932 (0.7934)

54. a) p = 0.0863

b) p = 0.3247

c) p = 0.3132

d) $p^2 = 0.0075$

e) p = 0.0363

f) $p^3 = 0.0342$

g) Expect 38 or 39.