Probability

1. Probabilities from Tables of Counts

The table below contains information for sales of 4 bedroom houses in Auckland City in June, July and August, 2004.

Exercise 1

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under $300,000	39	31	15	85
$300,000 - $600,000	35	45	4	84
Over $600,000	8	4	0	12
Total	82	80	19	181

Let C be the event that a house sold for under $300,000,
D be the event that a house sold for between $300,000 and $600,000 (inclusive),
E be the event that a house sold for over $600,000,
F be the event that the sale was made in less than 30 days,
G be the event that the sale was made in 30 to 90 days (inclusive), and
H be the event that the sale was made in more than 90 days.

1. For a sale selected at random from these 181 sales, find the probability that the sale was:

   (a) under \$300,000 $\frac{85}{181}$

   (b) made in less than 30 days $\frac{82}{181}$

   (c) under \$300,000 and made in less than 30 days $\frac{39}{181}$

   (d) under \$300,000 or made in less than 30 days $\frac{85+82-39}{181}$

   $= \frac{128}{181}$

   (e) \$300,000 or more $\frac{84+12}{181}$

   $= \frac{96}{181}$

2. For a sale selected at random from these 181 sales, find the probability that the sale was:

(a) over $600,000

$$\mathbf{P(E) = \frac{12}{181}}$$

(b) made in 30 to 90 days (inclusive)

$$\mathbf{P(G) = \frac{80}{181}}$$

(c) over $600,000 and made in 30 to 90 days (inclusive)

$$\mathbf{P(E\ and\ G) = \frac{4}{181}}$$

(d) over $600,000 or made in 30 to 90 days (inclusive)

$$\mathbf{P(E\ or\ G) = \frac{12 + 80 - 4}{181} = \frac{88}{181}}$$

(e) over $600,000 and made in more than 90 days

$$\mathbf{P(E\ and\ H) = \frac{0}{181} = 0}$$

(f) over $600,000 or made in more than 90 days

$$\mathbf{P(E\ or\ H) = \frac{12 + 19 - 0}{181} = \frac{31}{181}}$$

(g) made in at least 30 days

$$\mathbf{P(not\ F)\ or\ P(G\ or\ H) = \frac{80 + 19}{181} = \frac{99}{181}}$$

3. The following is a student's answer to Question 1(d).

$$\begin{aligned} P(C \text{ or } F) &= \frac{85+82}{181} \\ &= \frac{167}{181} \\ &= 0.9227 \text{ (to 4 d.p.)} \end{aligned}$$

Describe why this answer is not correct.

The 39 houses that sold for under $300,000 and sold in less than 30 days have been counted twice.

4. (a) Consider the event C or G (That is, house sold for under $300,000 or the sale was made in 30 to 90 days (inclusive).

Fill in the gaps:

$$\begin{aligned} P(C \text{ or } G) &= \frac{85+80-31}{181} \\ &= \frac{85}{181}+\frac{80}{181}-\frac{31}{181} \\ &= P(\mathbf{C}) + P(\mathbf{G}) - P(\mathbf{C\ and\ G}) \end{aligned}$$

(b) Consider the event D or H (That is, house sold for between \$300,000 and \$600,000 inclusive or the sale was made in more than 90 days).

Fill in the gaps: $P(\text{D or H}) = \frac{84+19-4}{181}$

$= \frac{84}{181} + \frac{19}{181} - \frac{4}{181}$

= P(**D**) + P(**H**) – P(**D and H**)

(c) Fill in the gaps: For any two events A and B

P(A or B) = **P(A)** + **P(B)** – **P(A and B)**

5. Look at your answer to Question 2(e).

(a) Comment on the events E and H.

There are no house sales that are in both event E and event H.

(b) E and H are called **mutually exclusive** events or **disjoint** events.

Fill the gap: For any two mutually exclusive events A and B

P(A and B) = **0**

Discussion Exercise

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under \$300,000	39	31	15	85
\$300,000 - \$600,000	35	45	4	84
Over \$600,000	8	4	0	12
Total	82	80	19	181

For the 181 house sales in the above table, what proportion of the houses that sold for over \$600,000 were on the market for less than 30 days?

$\frac{8}{12}$

2. Conditional Probability

Exercise 2

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under $300,000	39	31	15	85
$300,000 - $600,000	35	45	4	84
Over $600,000	8	4	0	12
Total	82	80	19	181

1. For the 181 house sales in the above table:

   (a) What proportion of the houses that sold for under $300,000 were on the market for less than 30 days?

   $P(F \mid C) = \frac{39}{85}$

   (b) Given that a house sold for between $300,000 and $600,000 (inclusive), what proportion were on the market for more than 90 days?

   $P(H \mid D) = \frac{4}{84}$

   (c) If a house was on the market for less than 30 days, what proportion sold for over $600,000?

   $P(E \mid F) = \frac{8}{82}$

   (d) Of those on the market for more than 90 days, what proportion sold for under $300,000?

   $P(C \mid H) = \frac{15}{19}$

2. For the 181 house sales in the above table:

(a) What proportion of the houses took more than 90 days to sell?

$P(H) = \frac{19}{181}$

(b) What proportion sold for under $300,000 given that they were on the market for 30 to 90 days (inclusive)?

$P(C \mid G) = \frac{31}{80}$

(c) What proportion took less than 30 days to sell or sold for between $300,000 and $600,000 (inclusive)?

$P(F \text{ or } D) = \frac{82 + 84 - 35}{181} = \frac{131}{181}$

(d) Of those that sold for over $600,000, what proportion were on the market for more than 90 days?

$P(H \mid E) = \frac{0}{12} = 0$

(e) What proportion sold for between $300,000 and $600,000 (inclusive) and took from 30 to 90 days (inclusive) to sell?

$P(D \text{ and } G) = \frac{45}{181}$

(f) What proportion of the houses took 90 days or less to sell?

$P(\text{not } H) \text{ or } P(F \text{ or } G) = \frac{82 + 80}{181} = \frac{162}{181}$

(g) What proportion of the houses that were on the market for more than 90 days sold for over $600,000?

$P(E \mid H) = \frac{0}{19} = 0$

(h) What proportion sold for over $600,000 and took more than 90 days to sell?

$P(E \text{ and } H) = 0$

3. Comment on your answers to Questions 2(d) and 2(g).

E and H are mutually exclusive events. Both conditional probabilities are zero. This is because if E has happened then H cannot occur and vice versa.

4. (a) Fill in the gaps:

Probability a house sold for under \$300,000 given that it sold in less than 30 days

$= P(C \mid F)$

$$= \frac{39}{82}$$

$$= \frac{39/181}{82/181}$$

$$= \frac{P(C \text{ and } F)}{P(F)}$$

(b) Fill in the gaps:

Probability a house sold for between \$300,000 and \$600,000 (inclusive) given that it was on the market for between 30 and 90 days (inclusive)

$= P(D \mid G)$

$$= \frac{45}{80}$$

$$= \frac{45/181}{80/181}$$

$$= \frac{P(D \text{ and } G)}{P(G)}$$

(c) Fill in the gaps: For any two events A and B,

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

3. Risk

Exercise 3

Study 1

In 1988 the results of the Physicians' Health Study Research Group study were reported in the *New England Journal of Medicine*. In this study 22 071 male physicians (aged from 40 to 84) were randomly assigned to two groups. One group took an aspirin every second day and the other group took a placebo, a pill with no active ingredient which looked just like an aspirin. The participants did not know whether they were taking aspirin or the placebo.

After five years the number of participants in each group who had had a heart attack was recorded. The results are shown in the table below.

Treatment	Heart attack	No heart attack	**Total**
Aspirin	104	10 933	11 037
Placebo	189	10 845	11 034
Total	293	21 778	22 071

1. All parts of this question apply to Study 1.

   (a) For those in the aspirin group:

   (i) The proportion who had a heart attack $= \frac{104}{11037}$

   = 0.00942

   (ii) The probability that a randomly selected participant had a heart attack

   = 0.00942

   (iii) The percentage who had a heart attack = 0.942%

   (iv) The risk of having a heart attack = 0.00942

   (v) Write this risk as a rate per 1000 participants 9.42 per 1000

   (vi) Write this risk as a rate per 10 000 participants 94.2 per 10 000

(b) (i) For those in the placebo group, the risk of having a heart attack

$$= \frac{189}{11034}$$

= 0.01713

(ii) Write this risk as a rate per 1000 participants **17.13 per 1000**

(iii) Write this risk as a rate per 10 000 participants **171.3 per 10 000**

(c) (i) Calculate the relative risk of having a heart attack using the risk for the placebo group as the denominator (i.e., as the baseline risk).

$$\text{Relative risk} = \frac{0.00942}{0.01713}$$

= 0.55

(ii) Interpret this relative risk.

For male physicians aged 40 to 84, **the risk of having a heart attack for those taking aspirin every second day is 0.55 the risk for those taking a placebo.**

(iii) Calculate the relative risk of having a heart attack using the risk for the aspirin group as the baseline risk.

$$\text{Relative risk} = \frac{0.01713}{0.00942}$$

= 1.82

(iv) Interpret this relative risk.

For male physicians aged 40 to 84, **the risk of having a heart attack for those taking a placebo is about 1.8 times the risk for those taking aspirin every second day.**

(v) Which group is more appropriate as the baseline group? Briefly explain.

Placebo. It makes more sense to compare the risk for a treatment group to that for a non-treatment (control) group.

(d) (i) Do male physicians aged 40 to 84 who take aspirin every second day have an **increased** or **decreased** risk of having a heart attack compared to those who take a placebo?

Circle one: **increased** [✓ **decreased**]

(ii) Calculate the percentage change in risk relative to the baseline (placebo group) risk.

Percentage change in risk = $\dfrac{0.00942 - 0.01713}{0.01713} \times 100\%$

= -45%

(iii) Interpret this percentage change in risk.

For male physicians aged 40 to 84 there is a 45% decrease in the chances of having a heart attack for those taking aspirin every second day (compared to those taking a placebo).

Study 2

In 2006 the results of a study carried out among 132 271 Jewish children born in Israel during 6 consecutive years in the 1980s were published in the *Archives of General Psychiatry*. The objective of the study was to examine the relationship between father's age at birth of child (offspring) and their risk of autism.

The offspring were assessed for autism at age 17 years. The results are shown in the table below.

Father's Age Group	Autism	No autism	**Total**
15 – 29	34	60 654	60 688
30 – 39	62	67 211	67 273
≥ 40	14	4 296	4 310
Total	110	132 161	132 271

2. All parts of this question apply to Study 2.

(a) For offspring from fathers aged 15 to 29 at the birth of their child:

(i) The proportion who had autism $= \frac{34}{60688}$

$= 0.00056$

(ii) The probability that a randomly selected offspring had autism

$= 0.00056$

(iii) The percentage who had autism $= 0.056\%$

(iv) The risk of having autism $= 0.00056$

(v) Write this risk as a rate per 10 000 offspring **5.6 per 10 000**

(b) (i) For offspring from fathers aged 30 to 39 at the birth of their child, the risk of having autism

$= \frac{62}{67273}$

$= 0.00092$

(ii) Write this risk as a rate per 10 000 offspring **9.2 per 10 000**

(c) (i) For offspring from fathers aged 40 or more at the birth of their child, the risk of having autism

$= \frac{14}{4310}$

= 0.00324

(ii) Write this risk as a rate per 10 000 offspring 32.4 per 10 000

(d) Using the risk for fathers in the 15 – 29 year age group as the baseline:

(i) Calculate the relative risk for fathers in the 30 – 39 year age group of having autistic offspring.

Relative risk $= \frac{0.00092}{0.00056}$

= 1.64

(ii) Interpret this relative risk.

Men aged 30 to 39 are about 1.6 times more likely to father an autistic child than those aged 15 to 29.

(iii) Calculate the relative risk for fathers in the 40 or more year age group of having autistic offspring.

Relative risk $= \frac{0.00324}{0.00056}$

= 5.79

(iv) Interpret this relative risk.

Men aged 40 or over are about 6 times more likely to father an autistic child than those aged 15 to 29.

In parts (e) and (f), use the 15 – 29 year age group as the baseline group.

(e) (i) Do fathers aged 30 to 39 have an **increased** or **decreased** risk of having autistic offspring compared to those aged 15 to 29?

Circle one: [✓ **increased**] **decreased**

(ii) Calculate the percentage change in risk relative to the baseline risk.

$$\text{Percentage change in risk} = \frac{0.00092 - 0.00056}{0.00056} \times 100\%$$

$$= 64\%$$

(iii) Interpret this percentage change in risk.

There is a 64% increase in the chances of fathering an autistic child for men aged 30 to 39 compared to those for men aged 15 to 29.

(f) (i) Do fathers aged 40 or more have an **increased** or **decreased** risk of having autistic offspring compared to those aged 15 to 29?

Circle one: [✓ **increased**] **decreased**

(ii) Calculate the percentage change in risk relative to the baseline risk.

$$\text{Percentage change in risk} = \frac{0.00324 - 0.00056}{0.00056} \times 100\%$$

$$= 479\%$$

(iii) Interpret this percentage change in risk.

There is a 480% increase in the chances of fathering an autistic child for men aged 40 or more compared to those for men aged 15 to 29.

4. Independent Events

If A and B are independent events then P(A | B) = P(A)

Exercise 4

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under $300,000	39	31	15	85
$300,000 - $600,000	35	45	4	84
Over $600,000	8	4	0	12
Total	82	80	19	181

1. Recall: C is the event that a house sold for under $300,000 and
F is the event that the sale was made in less than 30 days.
Are events C and F independent?

$P(C \mid F) = \frac{39}{82} = 0.4756$

$P(C) = \frac{85}{181} = 0.4696$

C and F are not independent because $P(C \mid F) \neq P(C)$

2. Recall: E is the event that a house sold for over $600,000 and
H is the event that the sale was made in more than 90 days.
Are events E and H independent?

$P(E \mid H) = \frac{0}{19} = 0$

$P(E) = \frac{12}{181}$

E and H are not independent because $P(E \mid H) \neq P(E)$

Discussion Exercise

Fill the gaps: For events A and B, $P(A|B) = \frac{P(\text{ A and B })}{P(\text{ B })}$

If A and B are independent events, $P(A|B) = P(\text{ A })$

So if A and B are independent events, $\frac{P(\text{ A and B })}{P(\text{ B })} = P(\text{ A })$

That is, if A and B are independent events $P(\text{ A and B }) = P(\text{ A }) \times P(\text{ B })$

Exercise 5

1. For events A and B, P(A) = 0.8, P(B) = 0.5 and P(A or B) = 0.9.

   (a) Calculate P(A and B)

   P(A or B) = P(A) + P(B) – P(A and B)

   0.9 = 0.8 + 0.5 – P(A and B)

   P(A and B) = 0.4

   (b) Calculate P(A | B)

   $$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

   $$= \frac{0.4}{0.5}$$

   = 0.8

   (c) Are A and B independent events? Justify your answer.

   Yes, because P(A | B) = P(A)

2. For events A and B, P(A) = 0.3, P(B) = 0.4 and P(A | B) = 0.

   (a) Are A and B independent events? Justify your answer.

   No, because P(A | B) ≠ P(A)

   (b) Are A and B mutually exclusive events? Justify your answer.

   Yes, because P(A | B) = 0

3. Events A and B are independent. Also P(A | B) = 0.4 and P(B) = 0.6.

   (a) State P(A)

   P(A) = 0.4 (by independence, P(A | B) = P(A))

   (b) Calculate P(A and B)

   P(A and B) = P(A) × P(B)

   = 0.4 × 0.6

   = 0.24

4. Events A and B are mutually exclusive. Also P(A) = 0.4 and P(B) = 0.3.

(a) State or calculate P(A | B)

P(A | B) = 0

(b) Calculate P(A or B)

P(A or B) = P(A) + P(B)

= 0.4 + 0.3

= 0.7

5. Tables of Counts and Probability Trees

Blood Group Systems Example: Example 4.7.3 from Chance Encounters (p 179)

There are a large number of genetically based blood group systems that have been used for typing blood. Two of these are the Rh system (with blood types Rh+ and Rh–) and the Kell system (with blood types K+ and K–). It is found that any person's blood type in any one system is independent of his or her blood type in any other.

It is known that, for Europeans in New Zealand, about 81% are Rh+ and about 8% are K+. If a European New Zealander is chosen at random, what is the probability that he or she is either positive in both systems or negative in both systems?

Table of Counts Approach

By independence, 81% of 800 or 8% of 8100

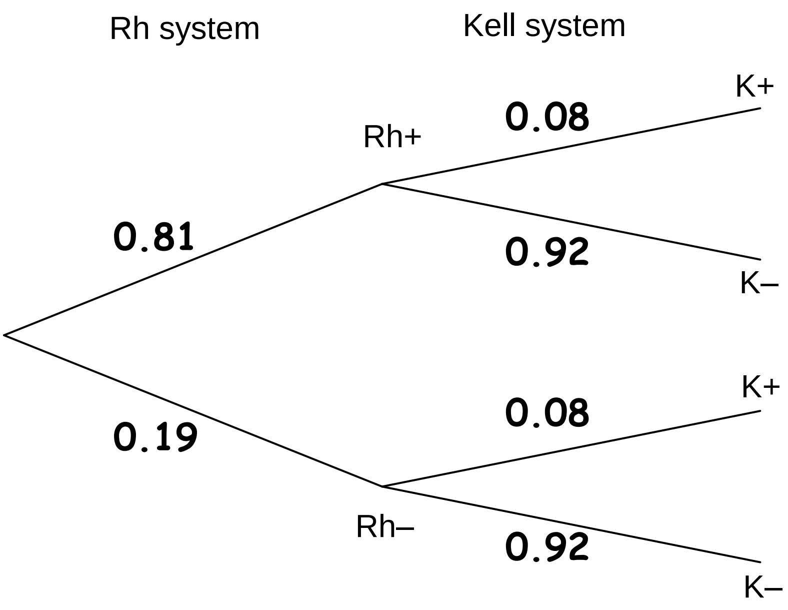
	Kell system		
Rh system	K+	K–	**Total**
Rh+	648	7452	8100
Rh–	152	1748	1900
Total	800	9200	10 000

Probability = P(Rh+ and K+) + P(Rh- and K-)

$$= \frac{648}{10000} + \frac{1748}{10000}$$

= 0.2396

Probability Tree Approach



Outcome	Probability
Rh+ and K+	0.0648
Rh+ and K-	0.7452
Rh- and K+	0.0152
Rh- and K-	0.1748

Probability = P(Rh+ and K+) + P(Rh- and K-)

= 0.0648 + 0.1748

= 0.2396

Imperfect Testing Procedures Example:

The ELISA test is used as a screening test for HIV. The test, however, is **not** a perfect one.

For people with HIV: 99.7% test positive

For people without HIV: 0.3% test positive

It is estimated that 0.1% of the New Zealand population have HIV.

Question: Suppose that a person is picked at random from New Zealand.
What is the probability of having HIV given that the test is positive?

Table of Counts Approach

99.7% of the 1000 with HIV test positive

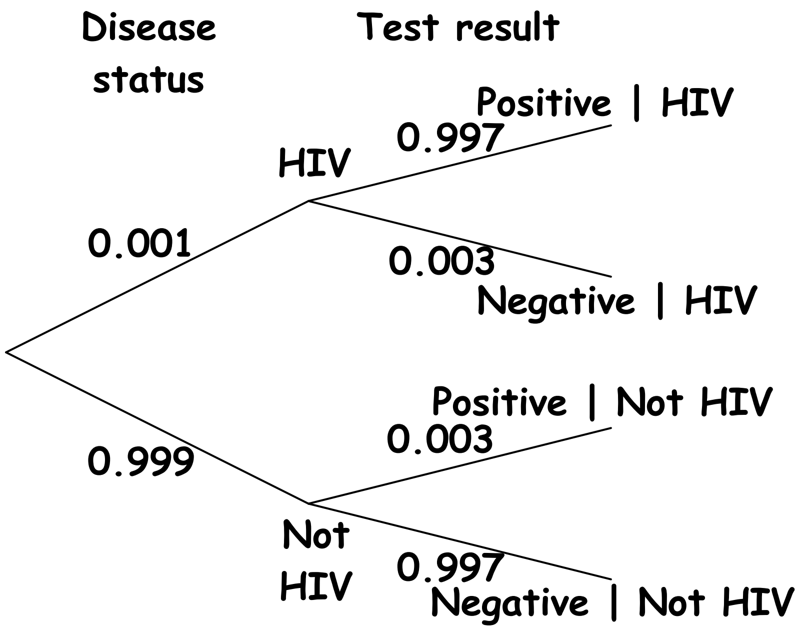
	Test result		
Disease status	Positive	Negative	**Total**
HIV	997	3	1000
Not HIV	2997	996 003	999 000
Total	3994	996 006	1 000 000

0.3% of the 999 000 without HIV test positive

$$P(\text{HIV} \mid \text{Positive}) = \frac{997}{3994}$$

$$= 0.2496$$

Probability Tree Approach



Outcome	Probability
HIV and Positive	0.000997
HIV and Negative	0.000003
Not HIV and Pos	0.002997
Not HIV and Neg	0.996003

$$P(\text{HIV} \mid \text{Positive}) = \frac{P(\text{HIV and Positive})}{P(\text{Positive})}$$

$$= \frac{0.000997}{0.000997 + 0.002997}$$

$$= \frac{0.000997}{0.003994}$$

$$= 0.2496$$