PsyConnect Statistics Guide
v0.1

Ho Han Sheng[1]

October 15, 2022

[1] https://github.com/ho-han-sheng

Dedicated to M, B and BB.

Contents

List of Figures

List of Tables

Preface

A common aversion faced by psychology undergraduates around the world is the need to study statistics. Given the relative infancy of psychology compared to other hard science fields or even the humanities, it is imperative that we as psychologists are able to read, comprehend and incorporate the most recent research as part of our continuous learning journey. With strong statistical foundations, one would be able to discern sound and rigourous statistical analyses from misleading or erroneous methods used in a fair percentage of psychological research (Bakker & Wicherts, 2011).

In some universities, the compulsory statistics modules are taught by mathematics professors. This is logical given their area of expertise, although it may further exacerbate the "expert blind spot" effect (Nathan & Petrosino, 2003). As experts, professors perform many abstract and symbolic reasoning steps automatically, "jumping" from one train of thought to another seamlessly as they would perceive. However, as undergraduates attempting to learn these concepts from scratch, it is exceptionally difficult to be able to "see" these connections automatically or even at all. Additionally, as psychology majors, certain "mathematical connections" inculcated in students majoring in fields requiring substantial mathematical foundations (e.g. engineering, computer science, mathematics) would be difficult to develop in short duration. What professors feel is trivial and does not require explicit instruction then, is what leads to students falling through the cracks.

In the case of my university (Singapore University of Social Sciences), the psychology program is fully part-time. The typical student profile then consists of mid-career switchers, mature adults and others who have gone years since their last interaction with mathematics. Assuming one leaves the education system with a GCE "O" Levels at 16 years old and pursues a non-mathematically heavy diploma in a polytechnic, it would be 6 years (!) before one can enrol in a part-time course with 2 years of working experience at 22 years old. That is akin to spending 6 years in a sedentary lifestyle with minimal exercise, making it difficult to return to a certain level of fitness.

What this book is for

This book serves to hopefully bridge the gap between expert knowledge and novice learning. Content will first focus on topics expected of students at "O" Levels, following the syllabus document published by the Singapore Examinations and Assessment Board. Afterwards, topics related to or typically required to understand statistical concepts at the undergraduate level will be discussed.

By attempting to keep technical jargon to a minimum, this book should require minimal effort to comprehend. It would also provide a sort of "warm-up" for students' minds to prepare them to think mathematically, serving as pre-reading before they embark on their statistics modules.

What this book is for me

My handwriting is terrible. When I first started my statistics modules, I needed a way to write my notes such that they were at least legible and easily written (i.e. typed out because I'm lazy). I was already using Notion to develop my own notes for my psychology modules and a quick google bestowed upon me the masterpiece that is LaTeX. Notion allowed me to typeset mathematical formulae with LaTeX commands and what followed was the development of my summary notes that were shared to my fellow psychology peers.

This book is then a personal project for me to further develop my skills in working with LaTeX, from typesetting everything within the document, learning the uses of various packages (TikZ, tabulary, BibLaTeX etc.) to creating my own graphics and figures in LaTeX.

About the companion website

The website[1] for this file contains:

- The (freely downlodable) latest version of this document.
- The LaTeX source code for this document.
- A way to propose amendments through pull requests.

[1] https://github.com/ho-han-sheng/psyconnect-statistics-guide

Acknowledgements

A special word of thanks goes to

- The PsyConnect EXCO Team
-

Ho Han Sheng

Chapter 1

Numbers and Algebra

1.1 Numbers and their operations

1.1.1 Number systems

What do these funny looking letters $\mathbb{R}$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ mean? These symbols denote which sets a particular number falls under (real, natural, integer etc.). Later on we will look at how these symbols may be used in set theory. Table 1.1 on page 6 outlines the main terms used.

1.1.2 Less than, greater than or equal to

Also called inequality symbols, they are used to represent the relationship between two variables or expressions.

The value of a is less than the value of b:

$$a < b$$

The value of a is greater than the value of b:

$$a > b$$

The value of a is less than or equal to the value of b:

$$a \leq b$$

The value of a is greater than or equal to the value of b:

$$a \geq b$$

Useful Tip: Imagine the inequality symbol as a crocodile's mouth and that the crocodile prefers eating larger numbers.

Set	Symbol	Definition	Example
Natural numbers	$\mathbb{N}$	Also called counting numbers, these are numbers we use to count items	$1, 2, 3, 4, 5, \ldots$
Whole numbers	No official symbol	Includes the set of natural numbers and 0	$0, 1, 2, 3, 4, 5, \ldots$
Integers	$\mathbb{Z}$	Consists of whole numbers, their opposites and 0	$\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$
Rational numbers	$\mathbb{Q}$	All numbers that can be expressed as a fraction	$44, -\frac{17}{5}, 0.1\dot{2} = \frac{11}{90}$
Irrational numbers	No official symbol	Numbers that cannot be expressed as a fraction	$\pi, \sqrt{2}, -3\pi$, Euler's number
Real numbers	$\mathbb{R}$	Encompasses all rational and irrational numbers. Every number that can exist on a typical number line	$\ldots, -5, -\pi, 0, \sqrt{17}, 75$
Imaginary numbers	No official symbol	Any real number multiplied by $\sqrt{-1} = i$	$2i, \pi i, -3i$
Complex numbers	$\mathbb{C}$	Includes the set of real numbers. Any number that has a real part and imaginary part.	$(2+3i), (\pi-6i)$

Table 1.1: Number Set Notations

1.1.3 Approximation and estimation

Significant figures refer to the number of digits in a value that are necessary and contribute to the degree of accuracy of the quantity of something. Insufficient significant figures during your calculations can lead to an inaccurate final answer. As a rule of thumb, *five* significant figures are used for intermediate calculations and *three* significant figures are used for your final answer.

We start by counting the digits from left to right, beginning with the first *non-zero* digit. To round off to *three* significant figures, we have to look at the fourth non-zero digit from the left.

$$0.0\underset{\substack{\uparrow\\ \text{1st}}}{7}56\underset{\substack{\uparrow\\ \text{4th}}}{9}$$

$$\underset{\substack{\uparrow\\ \text{1st}}}{7}89\underset{\substack{\uparrow\\ \text{4th}}}{0}43.0889$$

$$0.0\underset{\substack{\uparrow\\ \text{1st}}}{4}59\underset{\substack{\uparrow\\ \text{4th}}}{6}0$$

$$\underset{\substack{\uparrow\\ \text{1st}}}{3}00\underset{\substack{\uparrow\\ \text{4th}}}{2}.01$$

The value of this fourth digit determines how we round off the number. If the value of the fourth digit is ≥ 5 then we will add 1 to the third digit.

$$0.075\underset{\substack{\uparrow\\ \text{Add 1}}}{6}9 \approx 0.0757$$

$$78\underset{\substack{\uparrow\\ \text{Keep value}}}{9}043.0889 \approx 789000$$

$$0.045\underset{\substack{\uparrow\\ \text{Add 1}}}{9}60 \approx 0.0460$$

$$30\underset{\substack{\uparrow\\ \text{Keep value}}}{0}2.01 \approx 3000$$

If the value of the fourth digit is ≤ 4, then we will keep the value of the third digit. Note that the subsequent digits will become 0. For *five* significant figures then, we would look at the sixth non-zero digit from the left.

1.1.4 Scientific notation

Also known as standard form. This is a way to express numbers that are either very large or very small.

$$A \times 10^n$$

Where n is an integer[1] and $1 \leq A < 10$

The advantage of scientific notation lies in its lack of ambiguity in the number of significant figures. Refer to the number 3000, which is valid for 1, 2, 3 or 4 significant figures. As seen in Table 1.2, 3.251×10^6 is clearly 4 significant figures. Some calculators display E in place of $\times 10$. 3.251×10^6 would then appear as 3.251E6.

Decimal Notation	**Scientific Notation**
3	3×10^0
3000	3×10^3
3251000	3.251×10^6
0.02	2×10^{-2}
0.0000451	4.51×10^{-5}
0.000060075	6.0075×10^{-5}

Table 1.2: Decimal and Scientific Notation

Useful Tip: n can be thought of as the number of times we move the decimal point left or right[2].

$$3.251 \times 10^6 = 3.\underbrace{2\,5\,1\,0\,0\,0}_{\text{6 places}} = 3251000$$

$$6.0075 \times 10^{-5} = 0\,\underbrace{0\,0\,0\,0\,6}_{\text{5 places}}.0075 = 0.000060075$$

[1]Do you remember what an integer is? Refer back to Table 1.1 if you need to.
[2]Only applicable for base 10.

1.1.5 Indices

An index (Plural: indices) is the small superscript number that appears above a number. It denotes the number of times that number is multiplied by itself. 10^4 means 10 multiplied by itself 4 times, $10 \times 10 \times 10 \times 10 = 10000$. We can read this as "10 to the power of 4" or "10 raised to the power of 4".

Index

$$a^m$$

Base

There exists index laws and rules that allow us to simplify expressions when working with indices.

$$a^m \times a^n = a^{m+n} \tag{1.1}$$

$$2^4 + 2^3 = 2^{4+3} = 2^7 = 128$$

1.1. If 2 numbers with the same base are multiplied together, the result is equal to the same base raise to the power of the sum of the 2 indices.

$$(a^m)^n = a^{m \times n} \tag{1.2}$$

$$(2^4)^3 = 2^{4 \times 3} = 2^{12} = 4096$$

1.2. If a number a^m is itself raised to another power, the result is equal to the same base raised to the power of the 2 indices multiplied together.

$$\frac{a^m}{a^n} = a^{m-n} \tag{1.3}$$

$$\frac{2^5}{2^3} = 2^{5-3} = 2^2 = 4$$

$$\frac{2^5}{2^3} = 2^5/2^3$$

1.3. The result of the division of 2 indices is equal to the same base raised to the power of the subtraction of the denominator's index from the numerator's index.

$$a^{-m} = \frac{1}{a^m} \tag{1.4}$$

$$2^{-3} = \frac{1}{2^3}$$

1.4. The result of a number raised to a negative power is equal to the reciprocal[3] of the same number raised to a positive number of the same value.

$$a^0 = 1 \tag{1.5}$$

$$2^0 = 381^0 = 0^0 = 1$$

1.5. Any base raised to the power of 0 will always give a value of 1.

$$a^{\frac{m}{n}} = \sqrt[n]{a^m} \tag{1.6}$$

$$4^{\frac{2}{3}} = \sqrt[3]{4^2}$$

1.6. If the index is a fraction, the denominator n indicates the nth root and the numerator indicates the power to raise the base by.

$$a^m \times b^m = ab^m \tag{1.7}$$

$$2^3 \times 4^3 = (2 \times 4)^3 = 8^3 = 512$$

1.7. The result of the multiplication of 2 numbers with different bases but with the same index is equal to the multiple of both bases raised to the same power.

$$a^m/b^m = \frac{a^m}{b^m} = \left(\frac{a}{b}\right)^m \tag{1.8}$$

$$\frac{2^5}{3^5} = \left(\frac{2}{3}\right)^5 = \frac{32}{243}$$

1.8. The division of 2 numbers with different bases but with the same index yields a result equal to the fraction (or division) of the two bases raised to the same power.

[3]The reciprocal of $X = \frac{1}{X}$

1.2 Percentages

We frequently deal with percentages in statistics; From deriving the probability of an event occurring from a given proportion to determining our level of significance for hypothesis testing. It is imperative then for us to understand how to work with percentages and how to derive them when required.

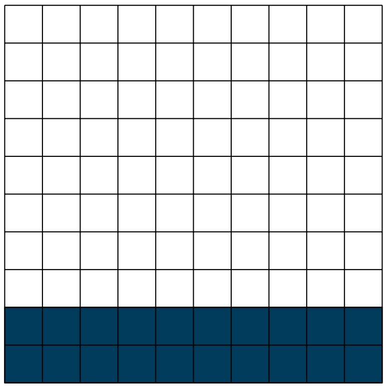

Figure 1.1: 20 Shaded Squares out of 100, 20%

The dictionary definition of "percent" is "of each 100" (Merriam-Webster, n.d.). Percentages can also be expressed as fractions or decimals. For example, 20 out of 100 can be expressed as:

$$\frac{20}{100} = \frac{2}{10} = 0.2 = 20\%$$

If the total quantity is not out of 100, we can (1) divide the two quantities to obtain the decimal form of the percentage or (2) multiply the result by 100 to obtain the percentage.

$$\frac{25}{47} = 0.53191$$

$$\frac{25}{47} \times 100\% = 53.191\% \approx 53.2\% \text{ (3 s.f.)}$$

Note that even before computing the result of this division, we have already obtained the fractional form of a percentage.

1.2.1 Expressing quantities as percentages

We typically wish to express a certain quantity x as a percentage of another quantity y. We accomplish this by expressing the 2 values as a fraction and multiplying the result by 100. Do note the respective positions of x and y as the numerator and denominator.

$$\begin{array}{r l} \text{Numerator} \rightarrow & \\ & \dfrac{x}{y} \times 100\% \\ \text{Denominator} \rightarrow & \end{array}$$

For example, I wish to express the weight of the phone (196g) in my pocket as a percentage of the total weight (60kg) measured on a weighing scale. Because the units of measurement are different (grams vs kilograms), we have to convert both quantities to the same units first.

$$\frac{0.196}{60} \times 100\% = 0.32667\% \approx 0.327\% \text{ (3 s.f.)}$$

Note that the decimal form for this percentage is 0.0032667 and not 0.32667. Therefore, we can say that the phone is 0.327% of the total weight (and is most likely not the cause of my weight gain).

1.2.2 Comparing two quantities with percentages

Sometimes we wish to compare quantities with differing total parts. For instance, test scores with different total marks.

$$\frac{28}{50} \text{ vs } \frac{71}{125}$$

It is not immediately apparent which score is better overall. Percentages help "transform" the quantities into a *proportion*, making it easier to differentiate.

$$\frac{28}{50} \times 100\% = 56.0\%$$
$$\frac{71}{125} \times 100\% = 56.8\%$$

1.2.3 Percentage change

When dealing with experimental data (or any data really), we are usually interested in how much a certain value has changed. This change in quantity can be expressed as a percentage change of the original value.

For example, the life expectancy at birth for Singaporeans born in 2011 was 81.9 years. At 2021, the preliminary life expectancy at birth was 83.5 years (Department of Statistics, 2022b). So how much of an increase is this? We can determine the percentage change by expressing the difference in years as a fraction of the original value at 2011.

$$\text{Percentage change} = \frac{\overset{\text{New value}\downarrow}{83.5} - \overset{\downarrow\text{Original value}}{81.9}}{81.9} \times 100\% = 1.9536\% \approx 1.95\% \text{ (3 s.f.)}$$

This change can also be negative. For example, the fertility rate (per female) in Singapore was 1.6 in 2000 and 1.1 in 2020 (Department of Statistics, 2022a). Note that we always subtract the original value from the new value.

$$\text{Percentage change} = \frac{\overset{\text{New value}\downarrow}{1.1} - \overset{\downarrow\text{Original value}}{1.6}}{1.6} \times 100\% = -31.25\% \approx -31.3\% \text{ (3 s.f.)}$$

If we wish to find the percentage *decrease*, note that the negative sign is removed. This is because the word *decrease* implies diminution, the process of becoming less. Think of it as an invisible negative sign that exists when referring to decreases. We also have to ensure that we subtract the new number from the original number instead to obtain a positive value.

$$\text{Decrease} = \overset{\text{Original value}\downarrow}{1.6} - \overset{\downarrow\text{New value}}{1.1} = 0.5$$

$$\text{Percentage decrease} = \frac{0.5}{1.6} \times 100\% = 31.25\% \approx 31.3\% \text{ (3 s.f.)}$$

If we determined the decrease with the usual method (new value − original value), we would obtain a negative number (−31.25%). If we (wrongly) declare this to be a percentage decrease then:

$$\begin{aligned}\text{Percentage decrease} &= -31.3\% \\ \text{Percentage change} &= (-) - 31.3\% = 31.3\%\end{aligned}$$

We would actually be referring to a percentage increase (or a positive percentage change)!

1.2.4 Reverse percentages

Just a fancy way of saying that we can calculate the original quantity given the percentage change and value of the new quantity. For example, after government subsidies, the price of a Build-To-Order HDB flat is 60% cheaper at $\$350,000$ (one can dream). What was the original price of the HDB without subsidies?

$$
\begin{aligned}
(100-60)\% = 40\% &= \$350000 \\
1\% &= \frac{\$350000}{40} = \$8750 \\
100\% &= \$8750 \times 100 = \$875000
\end{aligned}
$$

1.3 Algebraic expressions and formulae

Perhaps the most important foundation you need to tackle statistics. Most of the time, we are interested in discovering what value fits into a given function or expression. There could be a single solution or an infinitely large range of solutions. Algebra allows us to work towards the solution by replacing unknown and indefinite values with letters. *Variables* are letters that may represent any given numerical value (e.g., x, y).

1.3.1 Algebraic notations

The way we write algebraic expressions follow many of the same conventions used in basic arithmetic (addition, subtraction, multiplication, division). Table 1.3 lists some of the common ways we describe algebraic relationships.

Algebraic Expression	**Meaning**
$x + y$ or $y + x$	x plus y or y plus x
$x - y$	y subtracted from x
$x \times y$ or xy, $y \times x$ or yx	Product of x and y
$\frac{x}{y}$	x/y or $x \times \frac{1}{y}$
x^2	$x \times x$
x^n	$x_1 \times x_2 \times x_3 \times \ldots \times x_n$
$3a$	$3 \times x$
$2(x + y)$	$2 \times (x + y)$
$\frac{2+x}{3}$	$(2 + x)/3$ or $\frac{1}{3} \times (2 + x)$

Table 1.3: Algebraic Notation

Note that $x^2 \neq 2 \times x$ and $1 \times x$ is simply written as x. Any numerical coefficient (the static number next to a letter) is always placed infront of its corresponding letter (i.e., not $x3$).

1.3.2 Polynomial nomenclature

Before we dive in too deep, let us touch on the different names we have for identifying certain algebraic equations. The classes are based on the *degree* of the equation, that is the highest power amongst all the terms in an equation with a non-zero coefficient.

Name	Degree	Example
Linear equation	1	$ax + b = 0$
Quadratic equation	2	$ax^2 + bx + c = 0$
Cubic equation	3	$ax^3 + bx^2 + cx + d = 0$
Quartic equation	4	$ax^4 + bx^3 + cx^2 + dx + e = 0$
Quintic equation	5	$ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0$

Table 1.4: Types of Algebraic Equations and Polynomials

There are names for higher degree equations but I doubt that you would need to know the nomenclature for all of them. Note that the degree refers to the highest power we can see in the *simplest expanded* form of the expression.

For example:

$$\begin{aligned} 4x(x-2) - 8x + 1 &= 4x^2 - 8x - 8x + 1 \\ &= 4x^2 + 1 \end{aligned}$$

Therefore, the degree of this (quadratic) equation is 2.

1.3.3 Expansion and factorisation of algebraic expressions

Algebraic expressions will not always be in their optimal form, especially when multiple independent expressions are involved. Expansion and factorisation helps us get to the simplest forms for computation. Expansion involves expressing all terms with a single coefficient, essentially removing any brackets. Factorisation is the opposite of expansion, where we try to rewrite the algebraic expression as a product of its factors.

$$4(x + 2y) \underset{Factorisation}{\overset{Expansion}{\rightleftharpoons}} 4x + 8y$$

$$\begin{aligned} x \cdot (y + z) &= x \cdot y + x \cdot z \\ &= xy + xz \end{aligned} \tag{1.9}$$

1.9. The **distributive law** is one of the primary ways we expand expressions, especially with brackets involved. While there exist mnemonics such as FOIL (First, Outer, Inner, Last) to guide us in multiplying terms, they do have special exceptions in which they fail, like multiplying polynomials with more than 2 terms. The distributive law however, can be used to multiply polynomials with any number of terms, albeit tediously.

For example:

$$\begin{aligned} (a + b + c + d)(w + x + y + z) &= (a)(w + x + y + z) + (b)(w + x + y + z) \\ &\quad + (c)(w + x + y + z) + (d)(w + x + y + z) \\ &= aw + ax + ay + az + bw + bx + by + bz \\ &\quad + cw + cx + cy + cz + dw + dx + dy + dz \end{aligned}$$

MTH219TMAQ6JUL21

Suppose that three random variables X_1, X_2, X_3 form a random sample from the continuous random distribution on the interval [0,1]. Assume X_1, X_2, X_3 are independent, calculate the expectation of $E[(X_1 - 2X_2 + X_3)^2]$.

$$
\begin{aligned}
X_1, X_2, X_3 &\sim U(0,1) \\
E(X_1) = E(X_2) = E(X_3) &= \frac{0+1}{2} \\
&= 0.5
\end{aligned}
$$

$$
\begin{aligned}
V(X_1) = V(X_2) = V(X_3) &= \frac{(1-0)^2}{12} = \frac{1}{12} \\
V(X) &= E(X^2) - [E(X)]^2 \\
E(X^2) &= V(X) + [E(X)]^2 \\
E(X_1^2) = E(X_2^2) = E(X_3^2) &= \frac{1}{12} + 0.5^2 \\
&= \frac{1}{3}
\end{aligned}
$$

Let $X_1 = a,\ X_2 = b,\ X_3 = c$

$$
\begin{aligned}
(a - 2b + c)^2 &= (a - 2b + c)(a - 2b + c) \\
&= (a)(a - 2b + c) + (-2b)(a - 2b + c) \\
&\quad + (c)(a - 2b + c) \\
&= a^2 - 2ab + ac - 2ab + 4b^2 - 2bc \\
&\quad + ac - 2bc + c^2 \\
&= a^2 + 4b^2 + c^2 - 4ab + 2ac - 4bc
\end{aligned}
$$

$$
\begin{aligned}
E[(X_1 - 2X_2 + X_3)^2] &= E[X_1^2 + 4X_2^2 + X_3^2 - 4X_1X_2 + 2X_1X_3 - 4X_2X_3] \\
&= E(X_1^2) + 4E(X_2^2) + E(X_3^2) - 4E(X_1)E(X_2) \\
&\quad + 2E(X_1)E(X_3) - 4E(X_2)E(X_3) \\
&= \frac{1}{3} + 4\left(\frac{1}{3}\right) + \frac{1}{3} - 4(0.5)(0.5) + 2(0.5)(0.5) - 4(0.5)(0.5) \\
&= 2 - 1 + 0.5 - 1 \\
&= \mathbf{0.5}
\end{aligned}
$$

You don't have to understand the solution behind this question yet, this is just to demonstrate an example of how the distributive law can be applied.

Certain algebraic expressions give us a consistent result when expanded.

$$\begin{aligned}(a+b)^2 &= (a+b)(a+b)\\ &= a^2+ab+ba+b^2\\ &= \boldsymbol{a^2+2ab+b^2}\end{aligned} \tag{1.10}$$

$$\begin{aligned}(a-b)^2 &= (a-b)(a-b)\\ &= a^2+(-)ab+(-)ba+b^2\\ &= \boldsymbol{a^2-2ab+b^2}\end{aligned} \tag{1.11}$$

$$\begin{aligned}(a+b)(a-b) &= a^2+(-)ab+ab+(-)b^2\\ &= \boldsymbol{a^2-b^2}\end{aligned} \tag{1.12}$$

There are multiple ways to factorise algebraic expressions. Each of which approach the problem from a different perspective and may be more appropriate depending on the given expression.

For example:

$$6x^3y^3 + 45x^2y^2 + 21xy = (3xy)(2x^2y^2 + 15xy + 7)$$

The highest common factor of 6, 45 and 21 is 3. xy may also be factored out, giving us $3xy$ as one factor.

Of course, "eyeballing" it isn't always so easy, so we shall look at some other procedural ways or algorithms that make the process easier.

$$\begin{aligned} 4x^2 - 9 &= (2x)^2 - (3)^2 \\ &= (\mathbf{2x+3})(\mathbf{2x-3}) \qquad \text{Using } a^2 - b^2 = (a+b)(a-b) \end{aligned}$$

First, we should always look out for **difference of squares** $(a^2 - b^2)$, or any of the other "special cases" that we dealt with in expansion of algebraic expressions. With more practice, you would gradually develop your ability to recognise these identities.

$$\begin{aligned} x^4 - 16 &= (x^2)^2 - 4^2 \\ &= (x^2+4)(x^2-4) \qquad \text{Using } a^2 - b^2 = (a+b)(a-b) \\ &= (\mathbf{x^2+4})(\mathbf{x+2})(\mathbf{x-2}) \quad \text{Using } a^2 - b^2 = (a+b)(a-b) \text{ again} \end{aligned}$$

For trinomials (expressions with 3 terms):

$$\begin{aligned} x^2 + 10x + 25 &= x^2 + 10x + 5^2 \\ &= (x+5)^2 \qquad \text{Using } a^2 + 2ab + b^2 = (a+b)^2 \end{aligned}$$

$$\begin{aligned} x^2 - 6x + 9 &= x^2 - 6x + 3^2 \\ &= (x-3)^2 \qquad \text{Using } a^2 - 2ab + b^2 = (a-b)^2 \end{aligned}$$

Of course, not every trinomial can be factored into squares. A more generalised method for factorisation is the **cross method**.

Given $2x^2 - x - 15$,

$$\begin{array}{ccc|c} 2x & + & 5 & 5x \\ x & - & 3 & -6x \\ \hline 2x^2 & - & 15 & -x \end{array}$$

Figure 1.2: Cross Method of Factorisation

Hence,

$$2x^2 - x - 15 = (2x + 3)(x - 5)$$

To start, we begin with a blank table.

$$\begin{array}{ccc|c} & & & \\ \hline & & & \end{array}$$

At the bottom left quadrant, we place the terms with a power of 2 and 0. The term with power 1 lies in the bottom right quadrant. Note that we include negative coefficients and not just their absolute values.

$$\begin{array}{ccc|c} & & & \\ \hline 2x^2 & - & 15 & -x \end{array}$$

We then deduce what terms lie in the top left quadrant. To obtain x^2, both terms need to contain x or one term needs to be x^2 and the other a (1). It should be clear that in order to obtain an x term with power 1, we should not introduce a power 2 into our factors. To obtain a 2 in $2x^2$, the only possible values are 2 and 1. Therefore,

$$\begin{array}{ccc|c} \text{2x} & & & \\ \text{x} & & & \\ \hline 2x^2 & - & 15 & -x \end{array}$$

To get -15, the possible factors are 1,15 and 3,5 and a negative sign in front of one of them (2 negatives is impossible as it would give us a positive value). Of course, we could trial and error this if we are unsure, but to save the trees lets assume that we chose 3,5 initially. Now the order in which we place 3 and 5 (top or bottom) matters, as you would soon see. Let's use a wrong example first.

$$\begin{array}{ccc|c} 2x & - & 5 & \\ x & + & 3 & \\ \hline 2x^2 & - & 15 & -x \end{array}$$

As the name suggests, we cross multiply the terms in the top left quadrant to obtain the values in the top right quadrant. The goal is to achieve 2 values which, when computed, give us the value in the bottom right quadrant ($-x$).

$$\begin{array}{ccc|c} 2x & - & 5 & -5x \\ x & + & 3 & 6x \\ \hline 2x^2 & - & 15 & -x \end{array}$$

$-5x + 6x = x \neq -x$, hence this set of factors is not a valid solution.

$$\begin{array}{ccc|c} 2x & + & 5 & 5x \\ x & - & 3 & -6x \\ \hline 2x^2 & - & 15 & -x \end{array}$$

Referring back to Figure 1.2, we know what the correct values are. To finalise our factors, instead of taking the cross multiplication, we simply take the two horizontal expressions $(2x + 5)(x - 3)$.

Now what if there are simply no possible solutions that we can derive with the cross method? Instead of trying to find a solution in the form of $(x+a)^2$, $(x+a)(x+b)$ or something similar, we can try to find a solution in the form of $(x+a)^2+b$. This method is called **completing the square**.

We begin with the simplest case of a monic polynomial, this is where the coefficient of the leading term in a polynomial is 1. i.e., (x^2+ax+b). Recall that $(a+b)^2 = a^2+2ab+b^2$. For us to factorise a polynomial into $(x+a)^2$, we must ensure that the polynomial we have follows this format:

$$(x+a)^2 = x^2+2ax+a^2$$

For example:

$$x^2+6x+7 \qquad\qquad \text{where } 2a = 6$$

To conform to the format, our constant value of a^2 would be:

$$\begin{aligned} 2a &= 6 \\ a &= \frac{6}{2} = 3 \\ a^2 &= 3^2 \end{aligned}$$

And to make things easier to compute, we will just casually slide in 3^2 between $6x$ and 7, giving us:

$$x^2+6x+3^2+7$$

Because we added a value to the equation, the overall "value" of this equation changed. Hence we need to subtract this extra value to maintain the value of this equation. This gives:

$$x^2+6x+3^2+7-3^2$$

Which we can now factorise and simplify into:

$$\begin{aligned} x^2+6x+3^2+7-3^2 &= (x+3)^2+7-3^2 \\ &= (x+3)^2+7-9 \\ &= \mathbf{(x+3)^2-2} \end{aligned}$$

Now that you understand the concept behind completing the square, here is a quicker shortcut. Note that I am now referring to the coefficient of x as b for clarity sake.

$$x^2 + bx + \left(\frac{b}{2}\right)^2 = \left(x + \frac{b}{2}\right)^2$$

Given $x^2 + 10x + 8$, $b = 10$.

$$\begin{aligned}
x^2 + 10x + 8 &= \left(x + \frac{b}{2}\right)^2 + 8 - \left(\frac{b}{2}\right)^2 \\
&= \left(x + \frac{10}{2}\right)^2 + 8 - \left(\frac{10}{2}\right)^2 \\
&= (x+5)^2 + 8 - 25 \\
&= \mathbf{(x+5)^2 - 17}
\end{aligned}$$

What about non-monic polynomials, where the coefficient of $x^2 \neq 1$? In that case, we can simply divide the entire equation by the coefficient to ensure we get a monic polynomial.

For example,

$$\begin{aligned}
3x^2 - 16x + 34 &= \frac{3x^2 - 16x + 34}{3} \\
&= x^2 - \frac{16}{3}x + \frac{34}{3} \qquad\qquad \text{therefore, } b = \frac{16}{3} \\
&= \left(x - \frac{\frac{16}{3}}{2}\right)^2 + \frac{34}{3} - \frac{\frac{16}{3}}{2} \\
&= \left(x - \frac{16}{6}\right)^2 + \frac{34}{3} - \frac{16}{6} \\
&= \mathbf{\left(x - \frac{16}{6}\right)^2 + \frac{26}{3}}
\end{aligned}$$

We can also factor out the coefficient if possible.

$$\begin{aligned}
2x^2 + 8x + 28 &= 2(x^2 + 4x + 14) \\
&= 2\left[\left(x + \frac{4}{2}\right)^2 + 14 - \left(\frac{4}{2}\right)^2\right] \\
&= 2[(x+2)^2 + 14 - 2^2] \\
&= \mathbf{2[(x+2)^2 + 10]}
\end{aligned}$$

We can also choose to factor out the coefficient from just the first 2 terms.

$$\begin{aligned}
5x^2 - 25x + 12 &= 5(x^2 - 5x) + 12 \\
&= 5\left[\left(x - \frac{5}{2}\right)^2 - \left(\frac{5}{2}\right)^2\right] + 12 \\
&= 5(x - 2.5)^2 - 5(2.5)^2 + 12 \\
&= 5(x - 2.5)^2 - 31.25 + 12 \\
&= \mathbf{5(x - 2.5)^2 - 19.25}
\end{aligned}$$

Generally,

$$ax^2 + bx + c = a(x + d)^2 + e \tag{1.13}$$

Where,

$$d = \frac{b}{2a} \quad \text{and} \quad e = c - \frac{b^2}{4a}$$

The sum and difference of cubes can also be factored with their special identities.

$$a^3 + b^3 = (a + b)(a^2 - ab + b^2) \tag{1.14}$$

$$a^3 - b^3 = (a - b)(a^2 + ab + b^2) \tag{1.15}$$

1.3.4 Rational expressions

Factorising is useful as they help simplify complicated expressions. This is especially so when working with rational expressions. A **rational expression** refers to a polynomial expression that exists within a fraction; or when we try to find the quotient of 2 different polynomial expressions.

Bibliography

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*(3), 666–678. https://doi.org/10.3758/s13428-011-0089-5

Department of Statistics. (2022a). *Births and fertility rates* [Data set]. Ministry of Trade and Industry, Singapore. https://tablebuilder.singstat.gov.sg/table/TS/M810091

Department of Statistics. (2022b). *Complete life tables for Singapore resident population, 2020-2021*. Ministry of Trade and Industry, Singapore. https://www.singstat.gov.sg/-/media/files/publications/population/lifetable20-21.ashx

Merriam-Webster. (n.d.). Percent. In *Merriam-Webster.com dictionary*. Retrieved June 6, 2022, from https://www.merriam-webster.com/dictionary/percent

Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal*, *40*(4), 905–928. https://doi.org/10.3102/00028312040004905