Introduction to biostatistics and machine learning

Olga Dethlefsen, Eva Freyhult, Bengt Sennblad, Payam Emami

2020-10-29

Contents

Preface

This "bookdown" book contains teaching and learning materials prepared and used during "Introduction to biostatistics and machine learning" course organised by NBIS, National Bioinformatics Infrastructure Sweden. The course is open for PhD students, postdoctoral researcher and other employees in need of biostatistical skills within Swedish universities. The materials are geared towards life scientists wanting to be able to understand and use basic statistical and machine learning methods. It may also suits those already applying biostatistical methods but who have never gotten a chance to reflect on the basic statistical concepts, such as the commonly misinterpreted p-value.

More about the course https://nbisweden.github.io/workshop-mlbiostatistics/

Part I

Preliminary Mathematics

Chapter 1

Mathematical notations

Aims

- to recapitulate the basic notations and conventions used in mathematics and statistics

Learning outcomes

- to recognize natural numbers, integrals and real numbers
- to understand the differences between variables and constants
- to use symbols, especially Sigma and product notations, to represent mathematical operations

1.1 Numbers

- **Natural numbers, N**: numbers such as 0, 1, 3, ...
- **Integers, Z**: include negative numbers ..., -2, -1, 0, 1, 2
- **Rational numbers**: numbers that can be expressed as a ratio two integers, i.e. in a form $\frac{a}{b}$, where a and b are integers, and $b \neq 0$
- **Real numbers, R**: include both rational and irrational numbers
- **Reciprocal** of any number is found by diving 1 by the number, e.g. reciprocal of 5 is $\frac{1}{5}$
- **Absolute value** of a number can be viewed as its distance from zero, e.g. the absolute value of 6 is 6, written as $|6| = 6$ and absolute value of -5 is 5, written as $|-5| = 5$
- **Factorial** of a non-negative integer number n is denoted by $n!$ and it is a product of all positive integers less than or equal to n, e.g. $4! = 4{\cdot}3{\cdot}2{\cdot}1 = 24$

1.2 Variables, constants and letters

Mathematics gives us a precise language to communicate different concepts and ideas. To be able to use it it is essential to learn symbols and understand how they are used to represent physical quantities as well as understand the rules and conventions that have been developed to manipulate them.

- **variables**: things that can vary, e.g. temperature and time
- **constants**: fixed and unchanging quantities used in certain calculations, e.g. 3.14159
- in principle one could freely choose letters and symbols to represent variables and constants, but it is helpful and choose letters and symbols that have meaning in a particular context. Hence, we
- x, y, z, the end of the alphabet is reserved for variables
- a, b, c, the beginning of the alphabet is used to represent constants
- π, ω and Greek letters below are used frequently used to represent common constant, e.g. $\pi = 3.14159$

Table 1.1: Uppercase and lowercase letters of the Greek alphabet

Letter	Upper case	Lower case	Letter	Upper case	Lower case
Alpha	A	α	Nu	N	ν
Beta	B	β	Xi	Ξ	ξ
Gamma	Γ	γ	Omicron	O	o
Delta	Δ	δ	Pi	Π	π
Epsilon	E	ϵ	Rho	P	ρ
Zeta	Z	ζ	Sigma	Σ	σ
Eta	H	η	Tau	T	τ
Theta	Θ	θ	Upsilon	Y	υ
Iota	i	ι	Phi	Φ	ϕ
Kappa	K	κ	Chi	Γ	γ
Lambda	Γ	γ	Psi	Ψ	ψ
Mu	M	μ	Omega	Ω	ω

1.3 A precise language

- Mathematics is a precise language meaning that a special attention has to be paid to the exact position of any symbol in relation to other.
- Given two symbols x and y, xy and x^y and x_y can mean different things
- xy stands for multiplication, x^y for superscript and x_y for subscript

1.4 Using symbols

If the letters x and y represent two numbers, then:

- their **sum** is written as $x + y$
- subtracting y from x is $x - y$, known also as **difference**
- to multiply x and y we written as $x \cdot y$ or also with the multiplication signed omitted as xy. The quantity is known as **product of x and y**
- multiplication is **associative**, e.g. when we multiply three numbers together, $x \cdot y \cdot z$, the order of multiplication does not matter, so $x \cdot y \cdot z$ is the same as $z \cdot x \cdot y$ or $ycdotz \cdot x$
- division is denoted by $\frac{x}{y}$ and mans that x is divided by y. In this expression x, on the top, is called **numerator** and y, on the bottom, is called **denominator**
- division by 1 leaves any number unchanged, e.g. $\frac{x}{1} = x$ and division by 0 is not allowed

Equal sign

- the equal sign $=$ is used in **equations**, e.g. $x - 5 = 0$ or $5x = 1$
- the equal sign $=$ can be also used in **formulae**. Physical quantities are related through a formula in many fields, e.g. the formula $A = \pi r^2$ relates circle area A to its radius r and the formula $s = \frac{d}{t}$ defines speed as distance d divided by time t
- the equal sign $=$ is also used in identities, expressions true for all values of the variable, e.g. $(x-1)(x-1) = (x^2 - 1)$
- opposite to the equal sign is "is not equal to" sign $\neq$, e.g. we can write $1 + 2 \neq 4$

Sigma and Product notation

- the Σ notation, read as **Sigma notation**, provides a convenient way of writing longs sums, e.g. the sum of $x_1 + x_2 + x_3 + ... + x_{20}$ is written as $\sum_{i=1}^{i=20} x_i$
- the Π notation, read as **Product notation**, provides a convenient way of writing longs products, e.g. $x_1 \cdot x_2 \cdot x_3 \cdot ... \cdot x_{20}$ is written as $\prod_{i=1}^{i=20} x_i$

1.5 Inequalities

Given any two real numbers a and b there are three mutually exclusive possibilities:

- $a > b$, meaning that a is greater than b
- $a < b$, meaning that a is less than b
- $a = b$, meaning that a is equal to b

Strict and weak

- inequality in $a > b$ and $a < b$ is **strict**
- as oppose to **weak** inequality denoted as $a \geq b$ or $a \leq b$

Some useful relations are:

- if $a > b$ and $b > c$ then $a > c$
- if $a > b$ then $a + c > b$ for any c
- if $a > b$ then $ac > bc$ for any positive c
- if $a > b$ then $ac < bc$ for any negative c

1.6 Indices and powers

- **Indices**, also known as **powers** are convenient when we multiply a number by itself several times
- e.g. $5 \cdot 5 \cdot 5$ is written as 5^3 and $4 \cdot 4 \cdot 4 \cdot 4 \cdot 4$ is written as 4^5
- in the expression x^y, x is called the *base* and y is called *index* or *power*

The laws of indices state:

- $a^m \cdot a^n = a^{m+n}$
- $\frac{a^m}{a^n} = a^{m-n}$
- $(a^m)^n = a^{m \cdot n}$

Rules derived from the laws of indices:

- $a^0 = 1$
- $a^1 = a$

Negative and fractional indices:

- $a^{-m} = \frac{1}{a^m}$ e.g. $5^{-2} = \frac{1}{5^2} = \frac{1}{25}$ for negative indices
- e.g. $4^{\frac{1}{2}} = \sqrt{4}$ or $8^{\frac{1}{3}} = \sqrt[3]{8}$ for fractional indices

1.7 Exercises: notations

Exercise 1.1. Classify numbers as natural, integers or real. If reall, specify if they are rational or irrational.

a) $\frac{1}{3}$
b) 2
c) $\sqrt{4}$
d) 2.3
e) π
f) $\sqrt{5}$
g) -7
h) 0
i) 0.25

Exercise 1.2. Classify below descriptors as variables or constants. Do you know the letters or symbols commonly used to represent these?

a) speed of light in vacuum
b) mass of an apple
c) volume of an apple

d) concentration of vitamin C in an apple
e) distance from Stockholm central station to Uppsala central station
f) time on the train to travel between the above stations
g) electron charge

Exercise 1.3. Write out explicitly what is meant by the following:

a) $\sum_{i=1}^{i=6} k_i$

b) $\prod_{i=1}^{i=6} k_i$

c) $\sum_{i=1}^{i=6} i^k$

d) $\prod_{i=1}^{i=3} i^k$

e) $\sum_{i=1}^{n} i$

f) $\sum_{i=1}^{i=4} (i+1)^k$

g) $\prod_{i=1}^{i=4} (k+1)^i$

h) $\prod_{i=0}^{n} i$

Exercise 1.4. Use Sigma or Product notation to represent the long sums and products below:

a) $1+2+3+4+5+6$
b) $2^2+3^2+4^2+5^2$
c) $4 \cdot 5 \cdot 6 \cdot 7 \cdot 8$
d) $1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+...+\frac{1}{n}$
e) $2-2^2+2^3-2^4+...+2^n$
f) $3+6+9+12+\cdots+60$
g) $3x+6x^2+9x^3+12x^4+...+60x^{20}$
h) $3x \cdot 6x^2 \cdot 9x^3 \cdot 12x^4 \cdot ... \cdot 60x^{20}$

Answers to selected exeercises (notations)

Exr. 1.1

a) real, rational
b) natural and integers, integers include natural numbers
c) $\sqrt{4}=2$ so it is a natural number and/subset of integers
d) real number, rational as it could be written as $\frac{23}{10}$
e) real number, irrational as it cannot be explained by a simple fraction
f) real number, irrational as it cannot be explained by a simple fraction
g) integer, non a natural number as these do not include negative numbers
h) natural number, although there is some argument about it as some define nautural numbers as positive integers starting from 1, 2 etc. while others include 0.

i) real, rational number, could be written as $\frac{25}{100}$

Exr. 1.2

a) constant, speed of light in vacuum is a constant, denoted c with $c = 299792458\frac{m}{s^2}$
b) variable, mass of an apple is a variable, different for different apple sizes, for instance 138 grams, denoted as $m = 100g$
c) variable, like mass volume can be different from apple to apple, denoted as V, e.g. $V = 200cm^3$
d) variable, like volume and mass can vary, denoted as ρ_i and defined as $\rho_i = \frac{m}{V}$. So given 6.3 milligrams of vitamin C in our example apple, we have $\rho_i = \frac{0.0063}{2}\frac{g}{cm^3} = 0.0000315\frac{g}{cm^3}$ concentration of vitamin D
e) constant, the distance between Stockholm and Uppsala is fixed; it could be a variable though if we were to consider an experiment on a very long time scale; distance is often denoted in physics as d
f) variable, time on the train to travel between the stations varies, often denoted as t with speed being calculated as $s = \frac{d}{t}$
g) constant, electron charge is $e = 1.60217663 \cdot 10^{-19}C$

Exr. 1.3

a) $\sum_{i=1}^{i=6} k_i = k_1 + k_2 + k_3 + k_4 + k_5 + k_6$

b) $\prod_{i=1}^{i=6} k_i = k_1 \cdot k_2 \cdot k_3 \cdot k_4 \cdot k_5 \cdot k_6$

c) $\sum_{i=1}^{i=3} i^k = 1^k + 2^k + 3^k$

d) $\prod_{i=1}^{i=3} i^k = 1^k \cdot 2^k \cdot 3^k$

e) $\sum_{i=1}^{n} i = 1 + 2 + 3 + ... + n$ we are using dots (...) to represent all the number until n. Here, thanks to Gauss we can also write $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$, i.e. Gauss formula for sum of first n natural numbers

Exr. 1.4

a) $1 + 2 + 3 + 4 + 5 + 6 = \Sigma_{k=1}^{6} k$

b) $2^2 + 3^2 + 4^2 + 5^2 = \Sigma_{x=2}^{5} x^2$

c) $4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 = \prod_{x=4}^{8} x$

d) $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + ... + \frac{1}{n} = \sum_{n=1}^{n} \frac{1}{n}$

Chapter 2

Sets

Aims

- to introduce sets and basic operations on sets

Learning outcomes

- to be able to explain what a set is
- to be able to construct new sets from given sets using the basic set operations
- to be able to use Venn diagrams to shows all possible logical relations between two and three sets

2.1 Definitions

- **set**: a well-defined collection of distinct objects, e.g. $S = \{2, 4, 6\}$
- **elements**: the objects that make up the set are also known as **elements** of the set
- if x is an element of S, we say that x belongs to S and write $x \in S$ and if x is not an element of S we say that x does not belong to S and write $x \notin S$
- a set may contain **finitely** many or **infinitely** many elements

- **subset,** $\subseteq$: if every element of set A is also in B, then A is said to be a subset of B, written as $A \subseteq B$ and pronounced A is contained in B, e.g. $A \subseteq B$, when $A = \{2, 4, 6\}$ and $\$ = B = \{2, 4, 6, 8, 10\}$. Every set is a subset if itself.
- **superset**: for our outs A and B we can also say that B is a **superset** of A and write $B \supset A$

- **cardinality**: the number of elements within a set S, denoted as $|S|$

- **empty set,** $\emptyset$: is a unique set with no members, denoted by $E = \emptyset$ or $E = \{\}$. The empty set is a subset of very set.

2.2 Basic set operations

- **union of two sets,** $\cup$: two sets can be "added" together, the union of A and B, written as $A \cup B$, e.g. $\{1, 2\} \cup \{2, 3\} = \{1, 2, 3\}$ or $\{1, 2, 3\} \cup \{1, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\}$
- **intersection of two sets,** $\cap$: a new set can be constructed by taking members of two sets that are "in common", written as $A \cap B$, e.g. $\{1, 2, 3, 4, 5, 6\} \cap \{2, 3, 7\} = \{2, 3\}$ or $\{1, 2, 3\} \cap \{7\} = \emptyset\}$

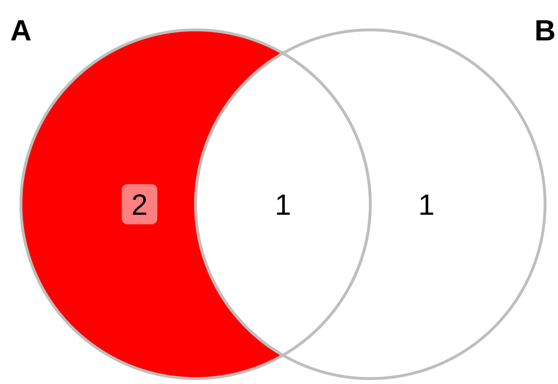
- **complement of a set,** A', A^c: are the elements not in A
- **difference of two sets,** : two sets can be "substracted", denoted by $A \quad B$, by taking all elements that are members of A but are not members of B, e.g. $\{1, 2, 3, 4\} \quad \{1, 3\} = \{2, 4\}$. This is also in other words a relative complement of A with respect to B.

- **partition of a set**: a partition of a set S is a set of nonempty subset of S, such that every element x in S is in exactly one of these subsets. That is, the subset are pairwise *disjoint*, meaning no two sets of the partition contain elements in common, and the union of all the subset of the partition is S, e.g. Set $\{1, 2, 3\}$ has five partitions: i) $\{1\}, \{2\}, \{3\}$, ii) $\{1, 2\}, \{3\}$, iii) $\{1, 3\}, \{2\}$, iv) $\{1\}, \{2, 3\}$ and v) $\{1, 2, 3\}$

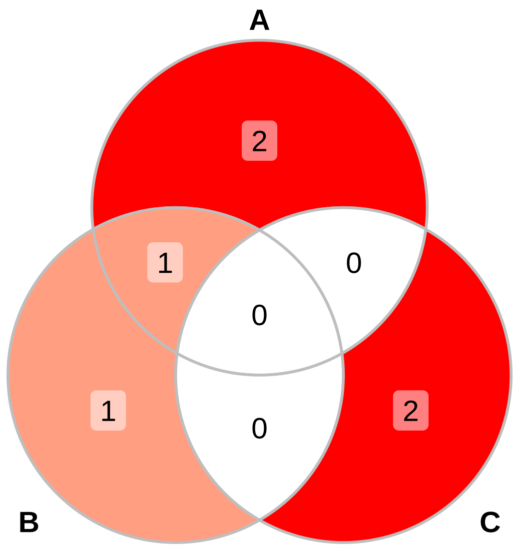
2.3 Venn diagrams

Venn diagram is a diagram that shows all possible logical relations between a finite collection of different sets. A Venn diagrams shows elements as points in the plane, and sets as regions inside closed curves. A Venn diagram consists of multiple overlapping closed curves, usually circles, each representing a set.

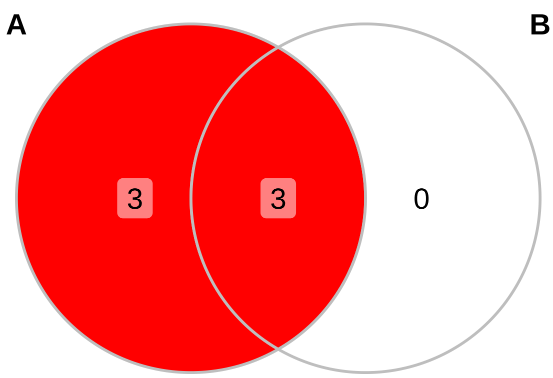
E.g. given $A = \{1, 2, 5\}$ and $B = \{1, 6\}$ Venn diagram of A and B:



And given $A = \{1, 2, 5\}$, $B = \{1, 6\}$ and $C = \{4, 7\}$ Venn diagram of A, B and C:



And given $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{2, 4, 6\}$ Venn diagram of A and B:



2.4 Exercises: sets

Exercise 2.1. Given set $S = \{1, 2, 3, 4, 5, 6\}$:

a) what is the subset T of S consisting of its even elements?
b) what is the complement T^c of T?
c) what is the subset U of S containing of the prime numbers in S?
d) What is the intersection $T \cap U$?
e) What is the union of $T \cup U$?
f) What is the set difference $U \quad T$?

Exercise 2.2. Given set $A = \{cat, elephant, dog, turtle, goldfish, hamster, parrot, tiger, guineapig, lion\}$

a) what is the subset D of A consiting of domesticated animals?
b) what is the subset C of A consiting of Felidae (cat) family?
c) what is the interection of D and C?
d) what is the complement of D, D^c?
e) what is the union of D and C?
f) what is the set difference of $A \quad C$?
g) can you draw Venn diagram showing relationship between D and C?

Answers to selected exercises (sets)

Exr. 2.1

a) $T = \{2, 4, 6\}$
b) $T^c = \{1, 3, 5\}$, i.e. T^c contains all the elements of S not in T
c) $U = \{2, 3, 5\}$, the primes in S
d) $T \cap U = \{2\}$, common elements of T and U, i.e. the even and prime numbers
e) $T \cup U = \{2, 3, 4, 5, 6\}$
f) $U \quad T = \{3, 5\}$, consisting of the elements of U that are not in T

Chapter 3

Functions

Aims

- to revisit the concept of a function

Learning outcomes

- to be able to explain what function, function domain and function range are
- to be able to identify input, output, argument, independent variable, dependent variable
- to be able to evaluate function for a given value and plot the function

3.1 Definitions

knitr::include_graphics(figures/precourse/math-functions-definition-02.png)

- A **function**, $f(x)$, can be viewed as a rule that relates input x to an output $f(x)$
- In order for a rule to be a function it must produce a single output for any given input
- Input x is also known as **argument** of the function
- **Domain of a function**: the set of all values that the function "maps"
- **Range**: the set of all values that the function maps into

Many names are used interchangeably

Functions have been around for a while and there are many alternative names and writing conventions are being used. Common terms worth knowing:

3.2 Evaluating function

To evaluate a function is to replace (substitute) its variable with a given number or expression. E.g. given a rule (function) that maps temperature measurements from Celsius to Fahrenheit scale:

$$f(x) = 1.8x + 32$$

where x is temperature measurements in Celsius and $f(x)$ is the associated value in Fahrenheit, we can find for a given temperature in Celsius corresponding temperature in Fahrenheit. Let's say we measure 10 Celsius degrees one autumn day in Uppsala and we want to share this information with a friend in USA. We can find the equivalent temperature in Fahrenheit by evaluating our function at 10, $f(10)$, giving us

$$f(10) = 1.8 \cdot 10 + 32 = 50$$

3.3 Plotting function

Function graphs are a convenient way of showing functions, by looking at the graph it is easier to notice function's properties, e.g. for which input values functions yields positive outcomes or whether the function is increasing or decreasing. To graph a function, one can start by evaluating function at different values for the argument x obtaining $f(x)$, plotting the points by plotting the pairs $(x, f(x))$ and connecting the dots. E.g. evaluating our above temperature rule at -20, -10, 0, 10, 20, 30 Celsius degrees results in:

x (Celsius degrees)	evaluates	f(x) (Farenheit degress)
-20	$f(-20) = 1.8 \cdot (-20) + 32$	-4
-10	$f(-20) = 1.8 \cdot (-10) + 32$	14
0	$f(-20) = 1.8 \cdot (0) + 32$	32
10	$f(-20) = 1.8 \cdot (10) + 32$	50
20	$f(-20) = 1.8 \cdot (20) + 32$	68
20	$f(-20) = 1.8 \cdot (30) + 32$	86

3.4 Standard classes of functions

Algebraic function: functions that can be expressed as the solution of a polynomial equation with integer coefficients, e.g.

- constant function $f(x) = a$
- identity function $f(x) = x$
- linear function $f(x) = ax + b$
- quadratic function $f(x) = a + bx + cx^2$
- cubic function $fx() = a + bx + cx^2 + dx^3$

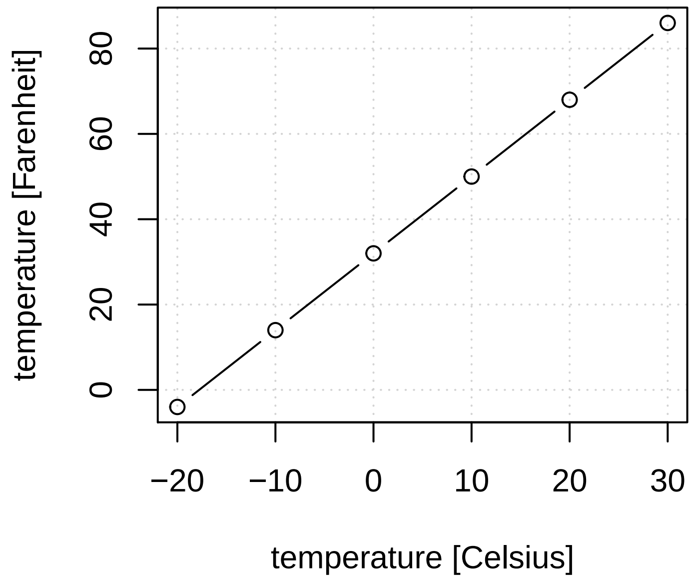


Figure 3.1: Graph of f(x) for the temeprature rule

Transcedental functions: functions that are not algebraic, e.g.

- exponential function $f(x) = e^x$
- logarithimic function $f(x) = log(x)$
- trigonometric function $f(x) = -3sin(2x)$

3.5 Piecewise functions

A function can be in pieces, i.e. we can create functions that behave differently based on the input x value. They are useful to describe situations in w which a rule changes as the input value crosses certain "boundaries". E.g. a function value could be fixed in a given range and equal to the input value (identify function) for input values outside this range

$$f(x) = \begin{cases} 2 & \text{if } x \leq 1 \\ x & \text{if } x > 1 \end{cases} \tag{3.1}$$

The function can be split in many pieces, e.g. the personal training fee in SEK may depend whether the personal trainer is hired for an hour, two hours or three or more hours:

$$f(h) = \begin{cases} 500 & \text{if } h \leq 1 \\ 750 & \text{if } 1 < h \leq 2 \\ 500 + 250 \cdot h & \text{if } h > 2 \end{cases} \tag{3.2}$$

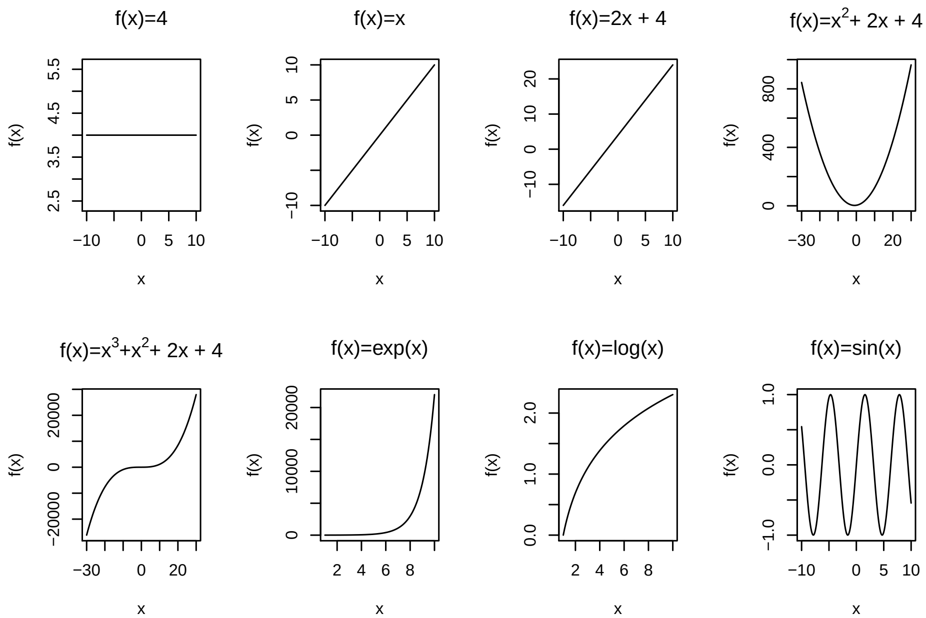


Figure 3.2: Examples of the standard classess of functions

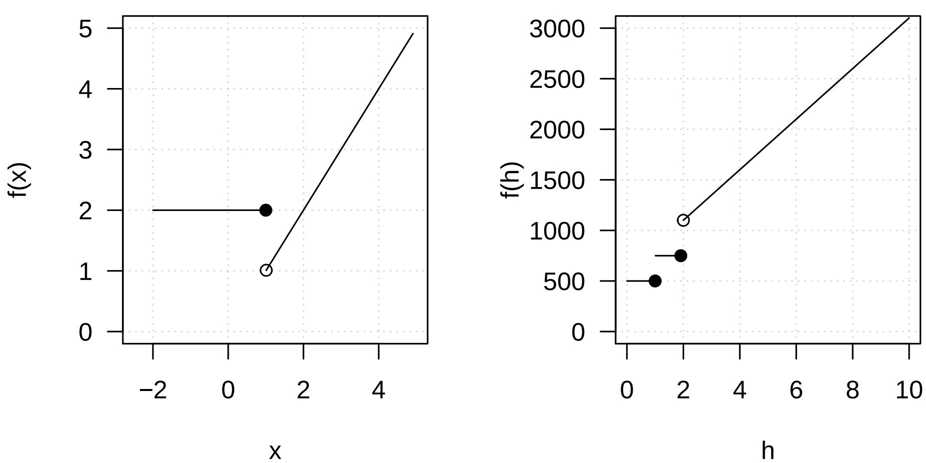


Figure 3.3: Examples of piece-wise functions

3.6 Exercises: functions

Exercise 3.1. Given the function for the personal trainer costs:

$$f(h) = \begin{cases} 500 & \text{if } h \leq 1 \\ 750 & \text{if } 1 < h \leq 2 \\ 500 + 250 \cdot h & \text{if } h > 2 \end{cases} \tag{3.3}$$

How much would you pay

a) for a 4-hours session? Evaluate function f(h) for value 4.
b) for a 2-hour session? Evalue function f(h) for value 2.

Exercise 3.2. A museum charges 50 SEK per person for a guided tour with a group of 1 to 9 people or a fixed $500 SEK fee for a group of 10 or more people. Write a function relating the number of people n to the cost C.

Exercise 3.3. Given function

$$f(x) = \begin{cases} x^2 & \text{if } x \leq 1 \\ 3 & \text{if } 1 < h \leq 2 \\ x & \text{if } h > 2 \end{cases} \tag{3.4}$$

a) sketch a graph of a function for x range (-4, 4)
b) evaluate function at f(1)
c) evaluate function at f(4)

Answers to selected exercises (functions)

Exr. 3.1

a) $f(4) = 500 + 250 \cdot 4 = 1500$
b) $f(2) = 750$ as $h \leq 2$ means less or equal to 2, that is including 2

Chapter 4

Differentiation

Aims

- introduce the concept of differentiation and rules of differentiation

Learning outcomes

- to be able to explain differentiation in terms of rate of change
- to be able to find derivatives in simple cases

4.1 Rate of change

- We are often interested in the rate at which some variable is changing, e.g. we may be interested in the rate at which the temperature is changing or the rate of water levels increasing
- Rapid or unusual changes may indicate that we are dealing with unusual situations, e.g. global warming or a flood
- Rates of change can be positive, negative or zero corresponding to a variable increasing, decreasing and non-changing

The function $f(x) = x^4 - 4x^3 - x^2 - e^{-x}$ changes at different rates for different values of x, e.g.

- between $x \in (-10, -9)$ the $f(x)$ is increasing at slightly higher pace than $x \in (5, 6)$
- between $x \in (-7, -5)$ the $f(x)$ is decreasing and
- between $x \in (0, 1)$ the $f(x)$ is not changing
- to be able to talk more precisely about the rate of change than just saying "large and positive" or "small and negative" change we need to quantify the changes, i.e. assign the rate of change an exact value
- **Differentiation** is a technique for calculating the rate of change of any function

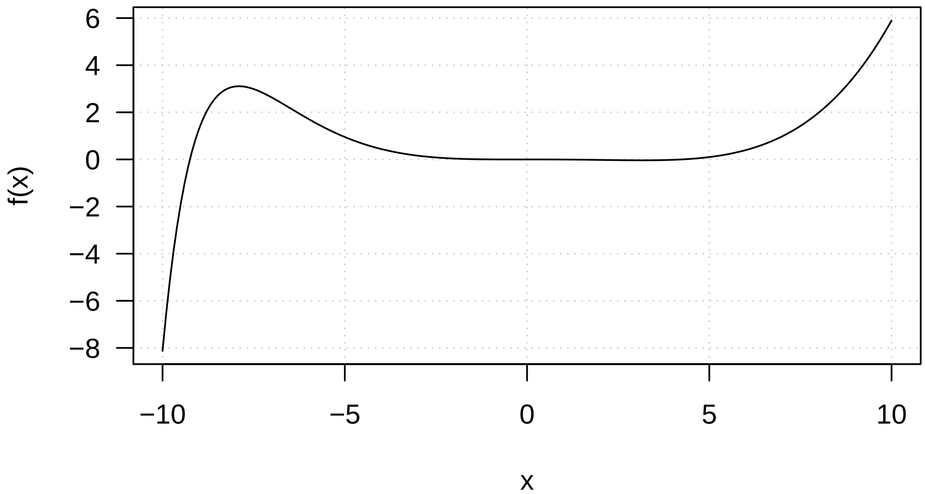


Figure 4.1: The function $f(x)$ changes at different rates for different values of x

4.2 Average rate of change across an interval

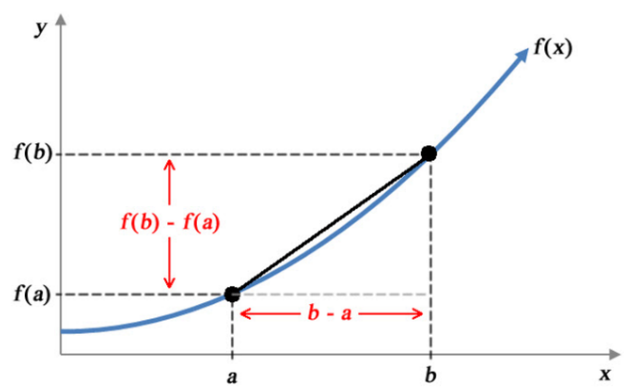


Figure 4.2: The average rate of change of $f(x)$ with respect to x over $[a, b]$ is equal to the slope of the secant line (in black)

To dive further into calculating the rate of change let's look at Figure 4.2 and define the *average rate of change* of a function across an interval. Figure 4.2 shows a function $f(x)$ with two possible argument values a and b marked and their corresponding function values $f(a)$ and $f(b)$.

Consider that x is increasing from a to b. The change in x is $b - a$, i.e. as x increases from a to b the function $f(x)$ increase from $f(a)$ to $f(b)$. The change in $f(x)$ is $f(b) - f(a)$ and the average rate of change of y across the $[a, b]$ interval is:

$$\frac{change\ in\ y}{change\ in\ x} = \frac{f(b) - f(a)}{b - a} \tag{4.1}$$

E.g. let's take a quadratic function $f(x) = x^2$ and calculate the average rate of change across the interval $[1, 4]$.

- The change in x is $4-1$ and the change in $f(x)$ is $f(4)-f(1) = 4^2-1^2 = 16-1 = 15$. So the average rate of change is $\frac{15}{3} = 5$. What does this mean? It means that across the interval $[1, 4]$ on average the $f(x)$ value increases by 5 for every 1 unit increase in x.
- If we were to look at the average rate of change across the interval $[-2, 0]$ we would get $\frac{f(-2)-f(1)}{-2-0} = \frac{-2^2-0}{-2-0} = \frac{4}{-2} = -2$. Here, over the $[-2, 0]$ on average the $f(x)$ value decreases by 2 for every 1 unit increase in x.
- Looking at the graph of $f(x) = x^2$ verifies our calculations

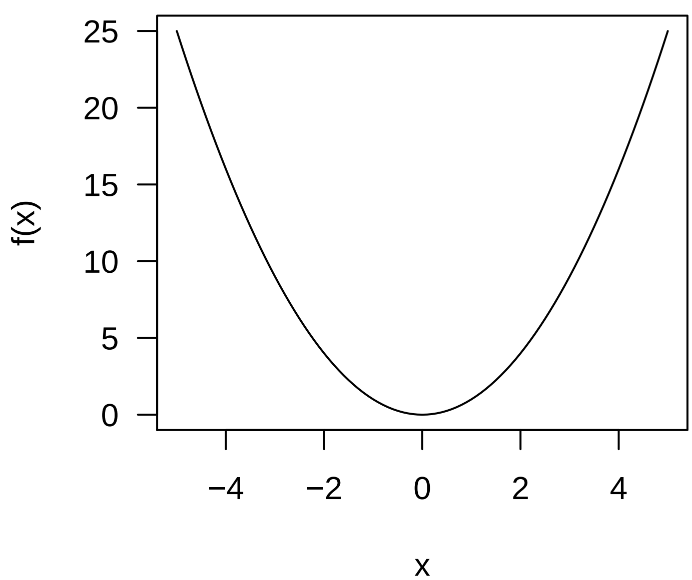


Figure 4.3: Example function $f(x) = x^2$

4.3 Rate of change at a point

- We often need to know the rate of change of a function at a point, and not simply an average rate of change across an interval.
- Figure 4.4, similar to Figure 4.2, shows, instead of two points a and b, point a and a second point defined in terms of its distance from the first

point a. Thus, the two points are now a and $a+h$ and the distance between the two points is equal to h.

- Now we can write that:

$$\frac{change\ in\ y}{change\ in\ x} = \frac{f(a+h)-f(a)}{a+h-a} = \frac{f(a+h)-f(a)}{h}$$

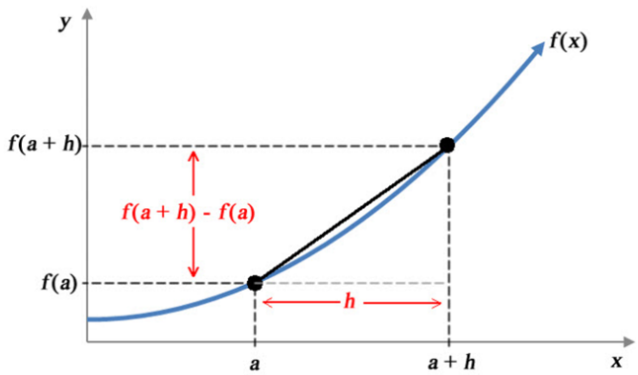


Figure 4.4: The average rate of change of $f(x)$ with respect to x over $[a, b]$ is equal to the slope of the secant line (in black)

Further:

- if we assume that the second point $a+h$ is really close to a, meaning that h approaches 0, denoted as $h \to 0$, we can find the rate of change at the point a
- the distance between the two points a and $a+h$ is getting smaller and so is the difference of the function values $f(a+h) - f(a)$. We denote these small differences as δa and $\delta f(x) = \delta(y)$ respectively
- the term δ reads as "delta" and represents a small change

We can thus continue and write that a **rate of change of a function at a point** is given by

$$\frac{small\ change\ in\ y}{small\ change\ in\ x} = \lim_{x\to 0} \frac{f(a+h)-f(a)}{h} \tag{4.2}$$

E.g. let's look at the linear function $f(x) = 2x + 3$. We can find the rate of change at any point of a function by:

$$\frac{small\ change\ in\ y}{small\ change\ in\ x} = \lim_{x\to 0} \frac{f(a+h)-f(a)}{h} = \frac{f(x+h)-f(x)}{x+h-x} = \lim_{x\to 0} \frac{2(x+h)+3-(2x+3)}{x+h-x} = \lim_{x\to 0} \frac{2h}{h} =$$

It mean that the $f(x)$ value increases by 2 for every small increase, h, in x. This increase is the same for all the values of x, does not depend on x. It could be the case, for instance if we look at the quadratic $f(x) = x^2$ function, we get:

$$\frac{\textit{small change in y}}{\textit{small change in x}} = \lim_{x \to 0} \frac{f(a+h) - f(a)}{h} = \frac{f(x+h) - f(x)}{x+h-x} = \lim_{x \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = \lim_{x \to 0} \frac{2xh + h^2}{h} = 2x+h$$

This means that:

- the rate of change for the function $f(x)$ at a point x is $2x$
- the $f(x)$ value increases by $2x$ for every small increase, h, in x
- the rate of change along a quadratic function is changing constantly according to the value of x we are looking at, it is a function of x
- and finally that the rate of change does not give us any information about the rate of change globally.

4.4 Terminology and notation

- **differentiation** is the process of finding the rate of change of a given function
- the function is said to be **differentiated**
- the rate of change of a function is also known as the **derivative** of the function
- given a function $f(x)$ we say that we differentiate function in respect to x and write:

$$\lim_{x \to 0} \frac{\delta y}{\delta x} = \frac{dy}{dx}$$

or use the "prime"

$$f'(x)$$

4.5 Table of derivatives

- in practice, there is no need to compute $\lim_{x \to 0} \frac{\delta y}{\delta x}$ every time when we want to find a derivative of a function
- instead, we can use patterns of the common functions and their derivatives

Table 4.1: Common functions and their derivatives

Function $f(x)$	Derivative $f'(x)$
$constant$	0
x	1
kx	k
x^n	nx^{n-1}
kx^n	knx^{n-1}
e^x	e^x

Function $f(x)$	Derivative $f'(x)$
$e^k x$	ke^{kx}
$log(x)$	$\frac{1}{x}$
$log(kx)$	$\frac{k}{x}$

We can use the Table 4.1 to find derivatives of some of the functions e.g.

- $f(x) = 3x,\ f'(x) = 3$
- $f(x) = 2x^4,\ f'(x) = 2 * 4x^{4-1} = 8x^3$
- $f(x) = e^{2x},\ f'(x) = 2e^{2x}$
- $f(x) = log(4x),\ f'(x) = \frac{4}{x}$

4.6 Exercises (differentiation)

Exercise 4.1. # Find derivatives of the functions

a) $f(x) = 2$
b) $f(x) = 2x + 1$
c) $f(x) = 5x^2$
d) $f(x) = 4x^3 + x^2$
e) $f(x) = \sqrt{(}x)$
f) $f(x) = \log(2x)$
g) $f(x) = e^x$
h) $f(x) = \frac{9}{x^2} + log(4x)$
i) $f(x) = 4x - 6x^6$
j) $f(x) = \frac{3}{x^2}$

Answers to selected exercises (differentiation)

Exr. 4.1

a) $f(x) = 2,\ f'(x) = 0$
b) $f(x) = 2x + 1,\ f'(x) = 2$
c) $f(x) = 5x^2,\ f'(x) = 10x$
d) $f(x) = 4x^3 + x^2,\ f'(x) = 12x^2 + x$
e) $f(x) = \sqrt{(}x) = x^{\frac{1}{2}},\ f'(x) = \frac{1}{2}x^{\frac{1}{2}-1} = \frac{1}{2}x^{-\frac{1}{2}}$
f) $f(x) = \log(2x),\ f'(x) = \frac{2}{x}$
g) $f(x) = e^x,\ f'(x) = e^x$

Chapter 5

Integration

Aims

- to introduce the concept of integration

Learning outcomes

- to be able to explain what integration
- to be able to explain the relationship between differentiation and integration
- to use integration to find functions based on derivatives in simple cases
- to use integration to calculate the area under the curve

5.1 Reverse to differentiation

- when a function $f(x)$ is known we can differentiate it to obtain the derivative $f'(x)$
- the reverse process is to obtain $f(x)$ from the derivative
- this process is called **integration**
- apart from simple reversing differentiation integration comes very useful in finding **areas under curves**, i.e. the area above the x-axis and below the graph of $f(x)$, assuming that $f(x)$ is positive
- the symbol for integration is $\int$ and is known as "integral sign"

E.g. let's take a function $f(x) = x^2$. Suppose we only have a derivative, which is $f'(x) = 2x$ and we would like to find the function given this derivative. Formally we write:

$$\int 2x dx = x^2 + c$$

.

where:

- the term $2x$ within the integral is called the **integrand**
- the term dx indicates the name of the variable involved, here x
- c is **constant of integration**

5.2 What is constant of integration?

- Integration reverses the process of differentiation, here, given our example function $f(x) = x^2$ that we pretended we do not know, we started with the derivative $f\,'(x) = 2x$ and via integration we obtained back the very function
$$\int 2xdx = x^2$$
- However, many function can result in the very same derivative since the derivative of a constant is 0 e.g. a derivatives of $f(x) = x^2$, $f(x) = x^2 + 10$ and $f(x) = x^2 + \frac{1}{2}$ all equal to $f'(x) = 2x$
- We have to take this into account when we are integrating, i.e. reverting differentiation. As we have no way of knowing what the original function constant is, we add it in form of c, i.e. unknown constant, called the constant of integration.

5.3 Table of integrals

Similar to differentiation, in practice we can use tables of integrals to be able to find integrals in simple cases

Table 5.1: Common functions and their integrals

Function $f(x)$	Integral $\int f(x)dx$
$constantk$	$kx + c$
x	$\frac{x^2}{2} + c$
kx	$k\frac{x^2}{2} + c$
x^n	$\frac{x^{n+1}}{n+1} + c$
kx^n	$k\frac{x^{n+1}}{n+1} + c$
e^x	$e^x + c$
e^kx	$\frac{e^{kx}}{k} + c$
$\frac{1}{x}$	$log(x) + c$

E.g.

- $\int 4x^3dx = \frac{4x^{3+1}}{3+1} = x^4 + c$
- $\int (x^2 + x)dx = \frac{x^3}{3} + \frac{x^2}{2} + c$ (note: we can evaluate integrals separately and add them as integration as differentiation is linear)

5.4 Definite integrals

- the above examples of integrals are **indefinite integrals**, the result of finding an indefinite integral is usually a function plus a constant of integration
- we have also **definite integrals**, so called because the result is a definite answer, usually a number, with no constant of integration
- definite integrals are often used to areas bounded by curves or, as we will cover later on, estimating probabilities
- we write:

$$\int_a^b f(x)dx$$

  where:
- $\int_a^b f(x)dx$ is called the definite integral of $f(x)$ from a to b
- the numbers a and b are known as lower and upper limits of the integral

E.g. let's look at the function $f(x) = x$ plotted below and calculate a definite integral from 0 to 2.

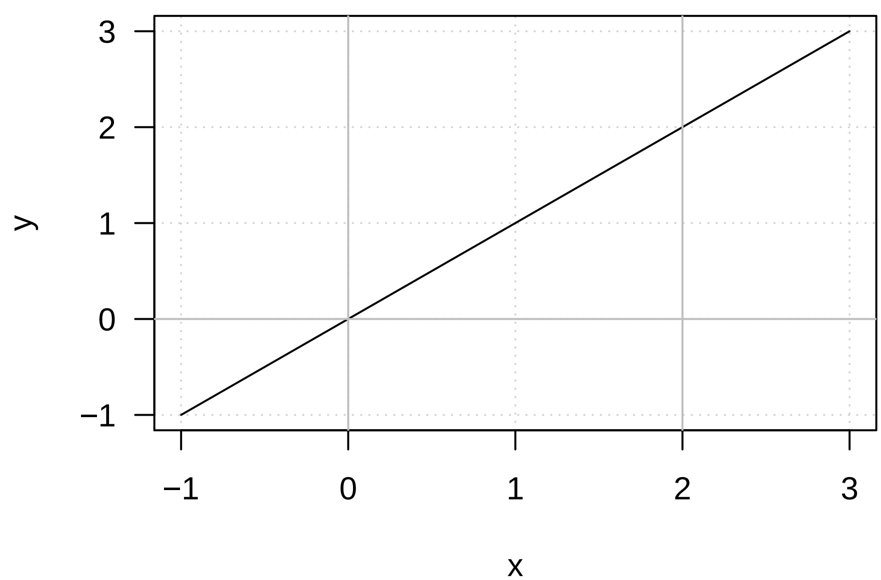


Figure 5.1: Graph of function $f(x) = x$

We write

$$\int_0^2 f(x)dx = \int_0^2 xdx = \Big[\frac{1}{2}x^2\Big]_0^2 = \frac{1}{2}(2)^2 - \frac{1}{2}(0)^2 = 2$$

so first find the integral and then we evaluate it at upper limit and subtracting the evaluation at the lower limit. Here, the result it 2. What would be the result if you tried to calculate the triangle area on the above plot, area defined by the gray vertical lines drawn at 0 and 2 and horizontal x-axis? The formula for the triangle area is $Area = \frac{1}{2} \cdot base \cdot height$ so here $Area = \frac{1}{2} \cdot 2 \cdot 2 = 2$ the same result as achieved with integration.

Exercise 5.1. Integrate:

a) $\int 2 \cdot dx$
b) $\int 2x \cdot dx$
c) $\int (x^4 + x^2 + 1) \cdot dx$
d) $\int e^x \cdot dx$
e) $\int e^{2x} \cdot dx$
f) $\int \frac{2}{x} \cdot dx$
g) $\int_0^4 2x \cdot dx$
h) $\int_0^4 x^2 + 1 \cdot dx$
i) $\int x^4 + \frac{2}{x} + e^{2x} \cdot dx$
j) $\int_0^4 x^4 + 1 \cdot dx$

Answers to selected exercises (integration)

Exr. 5.1

a) $\int 2 \cdot dx = 2x + c$
b) $\int 2x \cdot dx = \frac{2x^2}{2} = x^2 + c$
c) $\int (x^4 + x^2 + 1) \cdot dx = \frac{x^5}{5} + \frac{x3}{3} + x + c$
d) $\int e^x \cdot dx = e^x + c$
e) $\int e^{2x} \cdot dx = \frac{1}{2} e^2 x$
f) $\int \frac{2}{x} \cdot dx = \int 2x^{-2} \cdot dx = -2x^{-1} + c$
g) $\int_2^4 2x \cdot dx = \left[x^2 \right]_0^4 = 16 - 4 = 12$
h) $\int_0^4 x^2 + 1 \cdot dx = \left[\frac{x^3}{3} + x \right]_0^4 = \frac{4^3}{3} + 4 - 0 = \frac{64}{3} + 4 = \frac{76}{3}$

Chapter 6

Vectors and Matrices

Aims

- to revisit the concept of a function

Learning outcomes

- to be able to explain what function, function domain and function range are
- to be able to identify input, output, argument, independent variable, dependent variable
- to be able to evaluate function for a given value and plot the function

Part II

Probability and Statistics

Chapter 7

Probability theory

Aims

Learning outcomes

- understand the concept of random variables and probability
- use resampling to compute probabilities

7.1 Random variables

The outcome of a random experiment can be described by a *random variable*.

Whenever chance is involved in the outcome of an experiment the outcome is a random variable.

A random variable can not be predicted exactly, but the probability of all possible outcomes can be described.

A random variable is usually denoted by a capital letter, $X, Y, Z, \dots$. Values collected in an experiment are *observations* of the random variable, usually denoted by lowercase letters $x, y, z, \dots$.

The *population* is the collection of all possible observations of the random variable. Note, the population is not always countable.

A *sample* is a subset of the population.

Example random variables:

- The weight of a random newborn baby
- The smoking status of a random mother
- The hemoglobin concentration in blood
- The number of mutations in a gene
- BMI of a random man

- Weight status of a random man (underweight, normal weight, overweight, obese)
- The result of throwing a die

Part III

Other

Chapter 8

Probability: reasoning under uncertainty

Learning outcomes

- understand the concept of probability
- manipulate probabilities by their rules
- assign probabilities in very simple cases

8.1 Introduction

Some things are more likely to occur than others. Compare:

- the chance of the sun rising tomorrow with the chance that no-one is infected with COVID-19 tomorrow
- the chance of a cold dark winter in Stockholm with the chance of no rainy days over the summer months in Stockholm

We intuitively believe that the chance of sun rising or dark winter occurring are enormously higher than COVID-19 disappearing over night or having no rain over the entire summer. **Probability** gives us a scale for measuring the likeliness of events to occur. **Probability rules** enable us to reason about uncertain events. The probability rules are expressed in terms of sets, a well-defined collection of distinct objects.

8.2 Basic set definitions

- **set**: a well-defined collection of distinct objects, e.g. $A = \{2, 4, 6\}$
- **subset,** $\subseteq$: if every element of set A is also in B, then A is said to be a subset of B, written as $A \subseteq B$ and pronounced A is contained in B,

e.g. $A \subseteq B$, when $B = \{2, 4, 6, 8, 10\}$. Every set is a subset if itself.
- **empty set,** $\emptyset$: is a unique set with no members, denoted by $E = \emptyset$ or $E = \{\}$. The empty set is a subset of very set.

8.3 Basic set operations

- **union of two sets,** $\cup$: two sets can be "added" together, the union of A and B, written as $A \cup B$, e.g. $\{1, 2\} \cup \{2, 3\} = \{1, 2, 3\}$ or $\{1, 2, 3\} \cup \{1, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\}$
- **intersection of two sets,** $\cap$: a new set can be constructed by taking members of two sets that are "in common", written as $A \cap B$, e.g. $\{1, 2, 3, 4, 5, 6\} \cap \{2, 3, 7\} = \{2, 3\}$ or $\{1, 2, 3\} \cap \{7\} = \emptyset\}$
- **complement of a set,** A', A^C: are the elements not in A
- **difference of two sets,** : two sets can be "substracted", denoted by $A \;\; B$, by taking all elements that are members of A but are not members of B, e.g. $\{1, 2, 3, 4\} \;\; \{1, 3\} = \{2, 4\}$. This is also in other words a relative complement of A with respect to B.
- **partition of a set**: a partition of a set S is a set of nonempty subset of S, such that every element x in S is in exactly one of these subsets. That is, the subset are pairwise *disjoint*, meaning no two sets of the partition cotain elements in common, and the union of all the subset of the partition is S, e.g. Set $\{1, 2, 3\}$ has five partitions: i) $\{1\}, \{2\}, \{3\}$, ii) $\{1, 2\}, \{3\}$, iii) $\{1, 3\}, \{2\}$, iv) $\{1\}, \{2, 3\}$ and v) $\{1, 2, 3\}$

8.4 Exercises

Exercise 8.1. Here is my exercise.

8.5 Answers to exercises

Chapter 9

Probability: random variables

Learning outcomes

- understand the concept of random discrete and continous variables
- to be able to use probability density/mass functions and cumulative distribution functions and to understand the relationship between them
- describe properties of binomial, geometric, Poisson, uniform, exponential and normal distributions and identify which distributions to use in practical problems

9.1 Random variables

The outcome of a random experiment can be described by a random variable.

Example random variables:

- The weight of a random newborn baby
- The smoking status of a random mother
- The hemoglobin concentration in blood
- The number of mutations in a gene
- BMI of a random man
- Weight status of a random man (underweight, normal weight, overweight, obese)
- The result of throwing a die

Whenever chance is involved in the outcome of an experiment the outcome is a random variable.

A random variable is usually denoted by a capital letter, $X, Y, Z, \ldots$. Values collected in an experiment are observations of the random variable, usually denoted by lowercase letters $x, y, z, ...$

A random variable can not be predicted exactly, but the probability of all possible outcomes can be described.

The population is the collection of all possible observations of the random variable. Note, the population is not always countable.

A sample is a subset of the population.

9.2 Discrete random variables

A discrete random variable can be described by its *probability mass function.*

–>

–>

Chapter 10

Summarising and visualising data

Chapter 11

Linear regression

11.1 Simple regression

11.2 Multiple regression