Political Science Ph.D. Math Camp[*]

Babak RezaeeDaryakenari (srezaeed@ASU.edu)
http://www.babakrezaee.com

Contents

[*] WORK IN PROGRESS: Please do not circulate or cite without permission

1 LaTeX

For high-quality typesetting, we can use LaTeX. It is used mostly in medium-to-large technical or scientific documents. There are several packages in $\mathcal{R}$ and STATA that generate LaTeX codes for publishing nice and professional statistical results. LaTeX is also compatible with $\mathcal{R}$, and they can be knitted together for a better management and documentation of your scientific research (Quantitative and Qualitative).

There are different ways of implementing LaTeXcodes. If you learn to code in LaTeX, you easily can move across the different *Tex* editors to see which one works better for you. Since I am a big fan of working and benefiting cloud-based platforms, I explain coding in LaTeXusing *Overleaf* project[1]. *Overleaf* is an online and free real-time collaborative writing and publishing tool which uses LaTeXcommands. There are thousands of templates saved on this online platform which can facilitate your work. Many academic journals, including *American Political Science Review* and *British Journal of Political Science*, provided their LaTeXtemplate on this website. During this course, I will give you homework to practice what is covered in class. Everyone should submit her or his solutions typed in LaTeX; I will not accept handwritten solutions, or documents which are typed in other word processing software such as MS Word.

Let's start with a simple document. Go to `www.overleaf.com`, and sign up an account[2]. Go to *My Projects*, and then click on *New Project*, and open a blank project.

Now, go to `https://www.overleaf.com/read/xcqpdsfgtjvg`, and find the template that I prepared for this class. We will discuss basic LaTeXcommands using this template.

[1] ShareLaTeX Joined Overleaf in June 2017 to improve the quality of conducting shared projects with you co-authors/teammates.

[2] Overleaf has been used by well-known academic publishers such as Cambridge and Oxford for sharing their manuscript templates. You usually receive tips and suggestions from this website, and I barely can remember they sent marketing or spam emails to me. Thus, I suggest using this platform. If you are not comfortable signing up an account with this website, please let me know so I suggest an alternative platform.

2 Logic

> The fundamental thesis of the following pages, that mathematics and logic are identical, is one which I have never since seen any reason to modify. *Bertrand Russell* in *Principles of Mathematics (1903)*

> Mathematics is a logical method ... Mathematical propositions express no thoughts. In life it is never a mathematical proposition which we need, but we use mathematical propositions only in order to infer from propositions which do not belong to mathematics to others which equally do not belong to mathematics. *Ludwig Wittgenstein* in *Tractatus Logico Philosophicus (1922)*

Theory is a set of *internally consistent* sentences, which include as a set of *assumptions* and *axioms*. *Axiom* is defined as a rule or principle that is generally considered to be true.

Our job, or one of our jobs, as a political scientist is to develop consistent theories and evaluate/review our peer's scientific research for logical consistency. Thus, learning about logical rules can improve our skills for this job.

A mathematical proposition, statement is either true or false; that is, a statement whose certainty or falsity can be ascertained; we call this the truth value of the statement. Thus, a statement can have only one two truth values: it can be either true, denoted by T, or it can be false, denoted by F. For example, the statement "Arizona is a state in the United States of America" is either true or false. On the other hand, "A week has seven days" is not a statement or scientific statements, since we cannot evaluate its validity. It cannot be either T or F; it is always true. A more related example is Ludwig von Mises' argument of rationality.

The statements that we make in political science, general social science[3], are conditional statement:

- *If* food price increases, the likelihood of urban unrest increases.
- If the quality of political institutions improves, GDP per capita increases.
- What else?

In math, we usually write these conditional statements as follow:

$$\text{if } A \text{ then } B\text{, or}$$

[3] Even more generally, in science

A implies B, or more formally
$$A \Rightarrow B$$

where A and B stand for statements. You may saw another format of writing this conditional statement before as sufficient and necessary conditions:

A is a sufficient condition for B, *and*
B is a necessary condition for A.

Example: True or False?

- Being a mammal is a sufficient condition for being human.
- Answer:

- Being human is a sufficient condition for being a mammal.
- Answer:

$A \Rightarrow B$, also can be written as $B \Leftarrow A$, is read B is implied by A. In fact, $A \Rightarrow B$ is equivalent to $B \Leftarrow A$.

Finally, the symbol $\Leftrightarrow$ means "if and only if", "iff", or "implies and is implied by". In other words, $A \Leftrightarrow B$ means $A \Rightarrow B$ and $A \Leftarrow B$. Some textbooks also use $\equiv$ to show that statement A and statement B are equivalent.

Now, to negate a statement, we write not(A), or more formally $\neg A$. There are two rules regarding the mathematical negation of a statement:

1. If $A \Rightarrow B$, then $\neg B \Rightarrow \neg A$
2. If $\neg(A\ and\ B)$, then $\neg A$ or $\neg B$. It also can be written as $\neg(A \wedge B) \Rightarrow \neg A \vee \neg B$.

Two other *quantifiers* that are used widely in mathematics is $\forall$ and $\exists$ which are read respectively "for all" or "for any" and "for some", "there is", or "there exists".

Example: Formalize below statement using quantifiers:

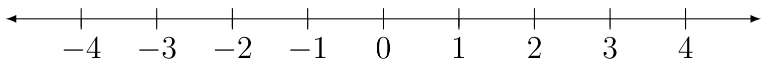
- *There exists an integer, n, such that n^2 is even.*
- Answer:

3 Functions on $\mathbb{R}^1$

To express the relationships between political variables, we use function. We are interested to study how a variable affects another variable. The simplest form of this association is a single variable function. For example, how does *GDP* affect *democracy*? How does insurgent group cohesion influence violence against civilians? We will discuss single variable, multi variable, linear, and non-linear functions in this course. In regression analysis, we estimate these different types of functions using different estimation methods.

3.1 Dimensionality

$\mathbb{R}^1$ is the set of real numbers extending from $-\infty$ to $+\infty$ i.e. the real number line. $\mathbb{R}^n$ is an n-dimensional space (often referred to as Euclidean space), where each of the *n* axes extends from $-\infty$ to $+\infty$. We usually use geometry and graphs to represent functions. The *number line* is a line that extends from $-\infty$ to $+\infty$. 0 is the *origin* point. The numbers on the right side of the origin point represent positive points, and the numbers on the left side of the origin point represent negative points.



Examples:

- $\mathbb{R}^1$ is a one dimensional line.
- $\mathbb{R}^2$ is a two dimensional plane.
- $\mathbb{R}^3$ is a three dimensional space.

Points in $\mathbb{R}^n$ are ordered n-tuples, where each element of the n-tuple represents the coordinate along that dimension.

- $\mathbf{R}^1$: (3)
- $\mathbf{R}^2$: (-5,5)
- $\mathbf{R}^3$: (4,2,5)

A function is like a machine; it receives an input, processes it, and produce an outcome. In fact, a single variable function $f(.)$ maps a number from $\mathbb{R}^1$ to number in $\mathbb{R}^1$. More formally, $f : \mathbb{R}^1 \to \mathbb{R}^1$. Here are several single variable functions:

- $f(x) = x$

- $f(x) = x + 1$
- $f(x) = 2x$
- $f(x) = -2x + 2$

Usually, we use variable x for the input variable, also known as exogenous variable, and variable y for the output variable, also known is endogenous variable.

3.2 Polynomials

Monomials functions: $f(x) = ax^k;\ k \in \mathbb{Z}^+$. Examples,

$$f_1(x) = x^2;\ \ f_2(x) = 5x^4;\ \ f_3(x) = -2x \tag{1}$$

k is called the degree of monomial, and a is called a coefficient. A function formed by adding monomials is called a polynomial:

$$g(x) = 3x^5 + 6x^3 - 8x + 10 \tag{2}$$

We usually write the monomial terms of a polynomial in order of decreasing degree. The highest degree of any polynomials is called the degree of polynomial.

3.3 Graphs

We usually use a graph to study geometric patterns of a function. The graph of a function is all the (x, y) points which satisfies $y = f(x)$ in the Cartesian system.

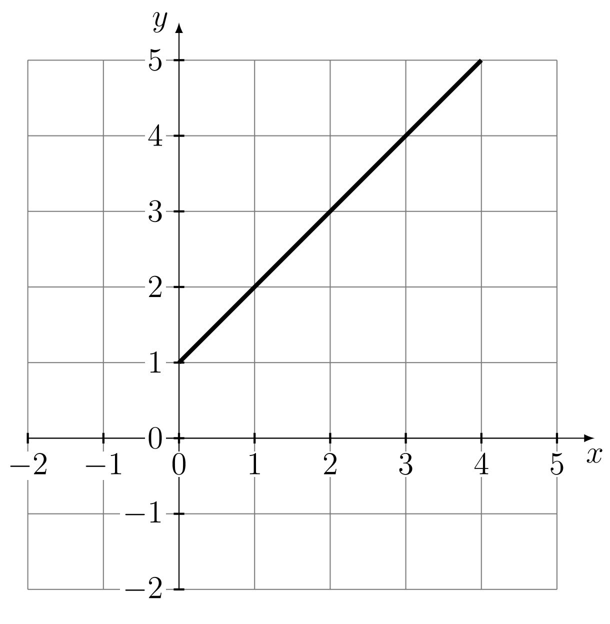


Figure 1: $y = x + 1$

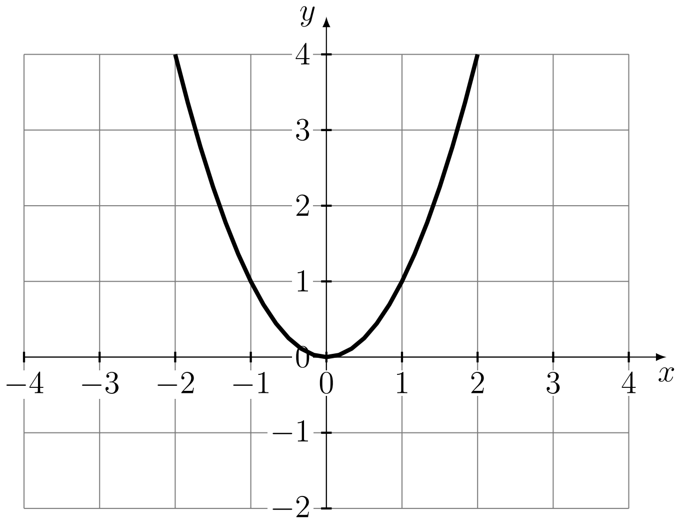


Figure 2: $y = x^2$

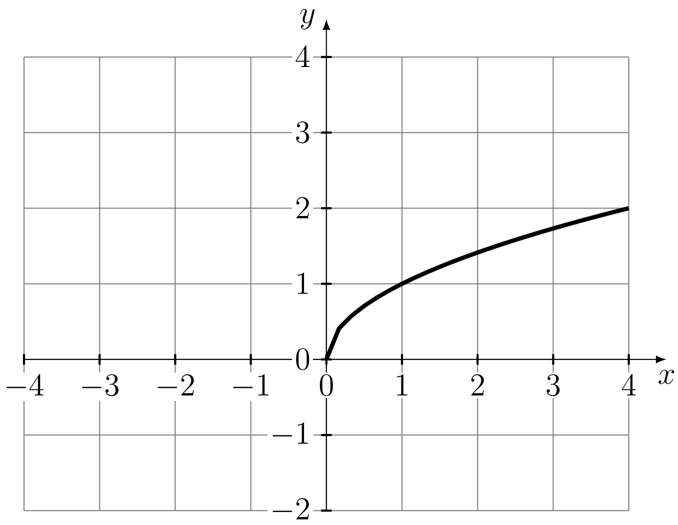


Figure 3: $y = x^{.5}$

3.4 Increasing and Decreasing Functions

A function is *increasing* if its graph moves upward from left to right. More formally,

$$\forall\ x_1\ and\ x_2\ s.t.\ x_2 > x_1 \Rightarrow f(x_2) > f(x_1)$$

A function is *decreasing* if its graph moves downward from left to right. More formally,

$$\forall\ x_1\ and\ x_2\ s.t.\ x_2 > x_1 \Rightarrow f(x_2) < f(x_1)$$

Examples: For each of the following functions, where is the function increasing and where is it decreasing?

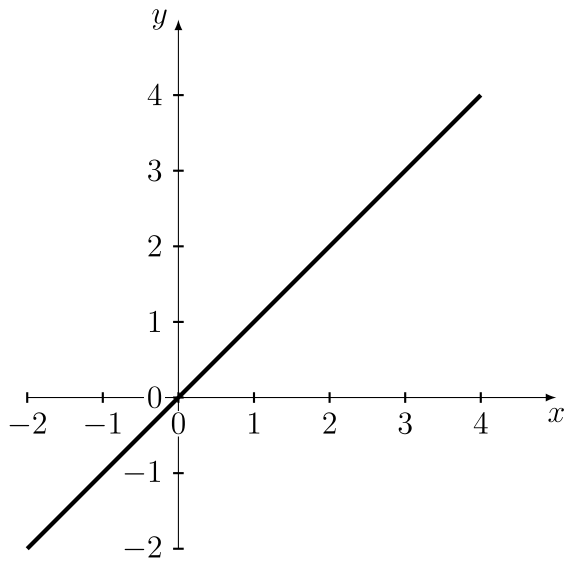


Figure 4: $y = x + 1$

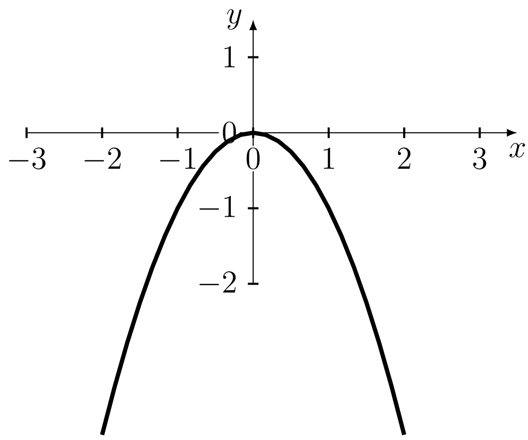


Figure 5: $y = -x^2$

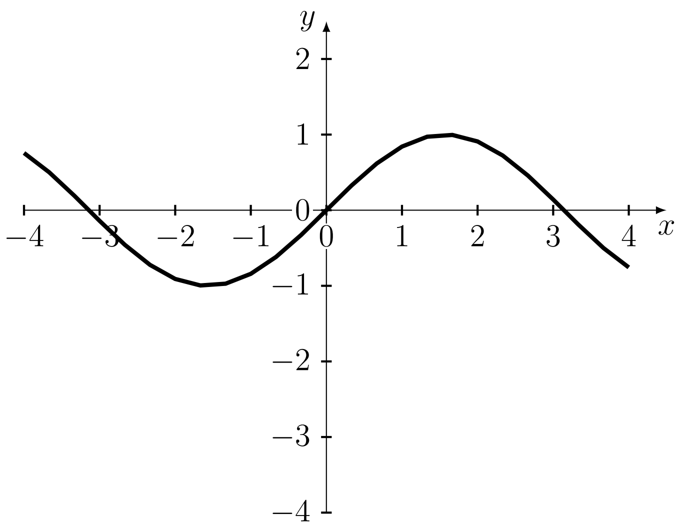


Figure 6: $y = sin(x)$

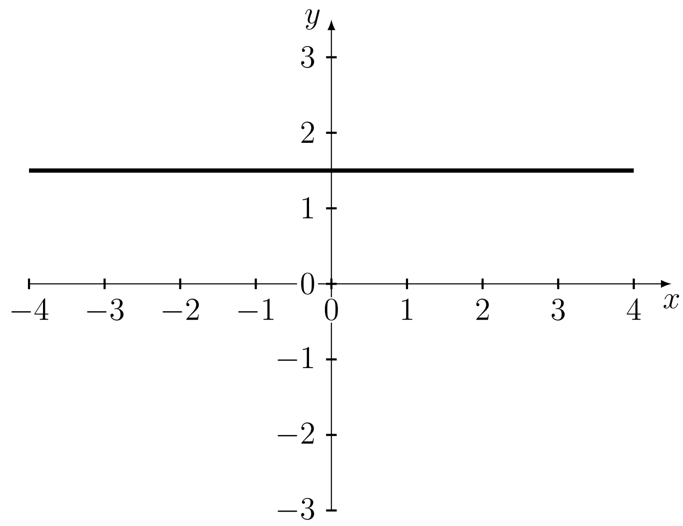


Figure 7: $y = 1.5$

3.5 Domain

Some functions are defined on a subset of $\mathbb{R}^1$. Given a function $f(.)$, the set of numbers x at which $f(x)$ is defined is called the *domain* of $f(.)$.

Example: What is the domain of each of the following functions:

- $y = \sqrt{x}$
- Answer:

- $y = \frac{1}{x-1}$
- Answer:

- $y = \frac{1}{\sqrt{x^2+1}}$
- Answer:

3.6 Interval Notation

Since the domain of a function can be a subset of $\mathbb{R}^1$, it would be helpful to review intervals. For given two real numbers a and b, the set of all numbers between a and b is called an *interval*. More formally, we can write an open interval as follow:

$$(a, b) = \{x \in \mathbb{R}^1 : a < x < b\}$$

A closed closed interval also can be written as follow:

$$[a, b] = \{x \in \mathbb{R}^1 : a \leq x \leq b\}$$

If only one endpoint is included, the interval is called *half-open* or *half-closed*, i.e. $(a, b]$ and $[a, b)$. There are also five types of *infinite intervals*:

$$(a, \infty) = \{x \in \mathbb{R}^1 : x > a\}$$

$$[a, \infty) = \text{............................}$$

$$(-\infty, a) = \text{............................}$$

$$(-\infty, a] = \text{............................}$$

$$(-\infty, \infty) = \mathbb{R}^1$$

4 Linear Functions

The simplest interesting functions are the polynomials of degree one:

$$f(x) = ax + b$$

These functions are called linear functions because their graphs are straight lines.

4.1 The slope of a line in the plane

The steepness of a graph, which is called *slope*, is one the main characteristics of linear functions. One way of measuring the slope of a linear function is that we move from one given point on *x-axis* to another one on the *x-axis* by one unit, and then calculate how the value of function, i.e. y changed.

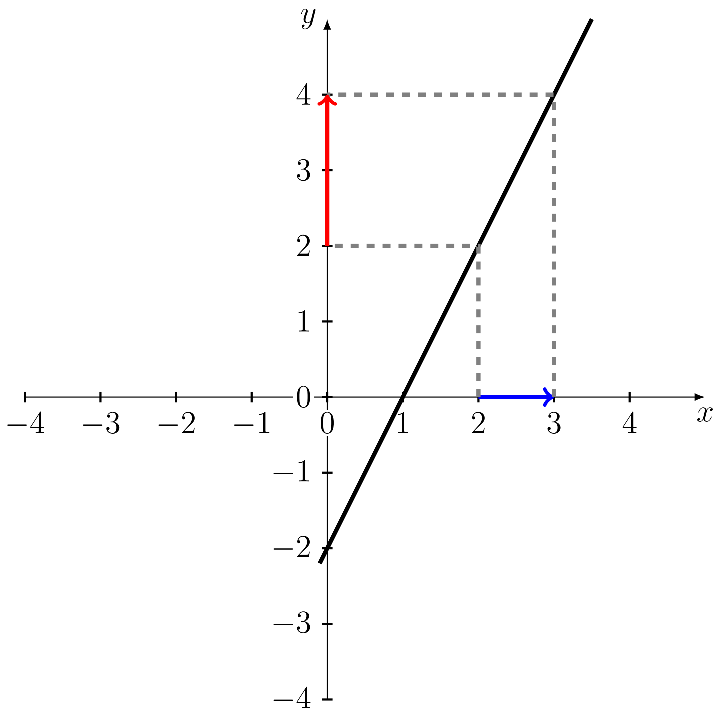


Figure 8: The slope of a function in the plane.

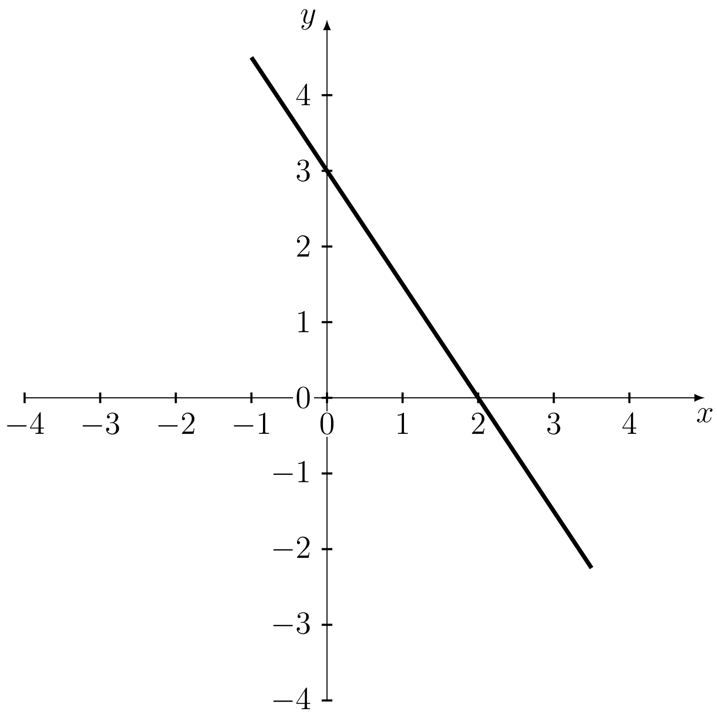


Figure 9: The slope of a function in the plane.

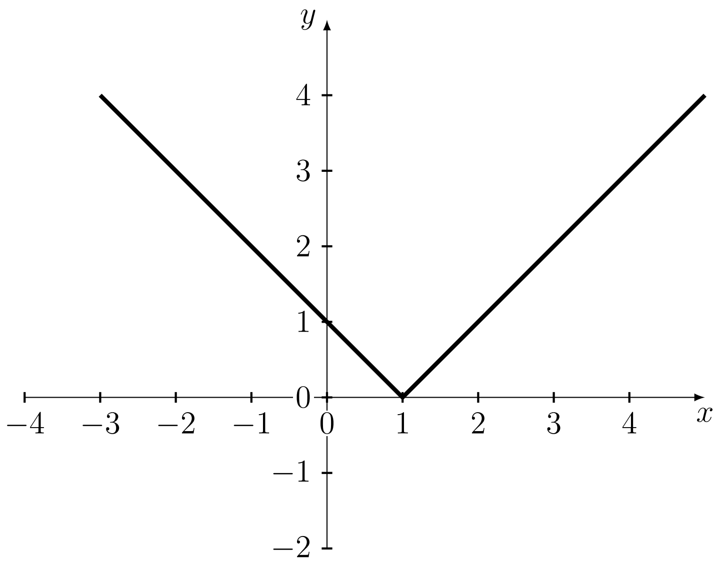


Figure 10: The slope of a function in the plane.

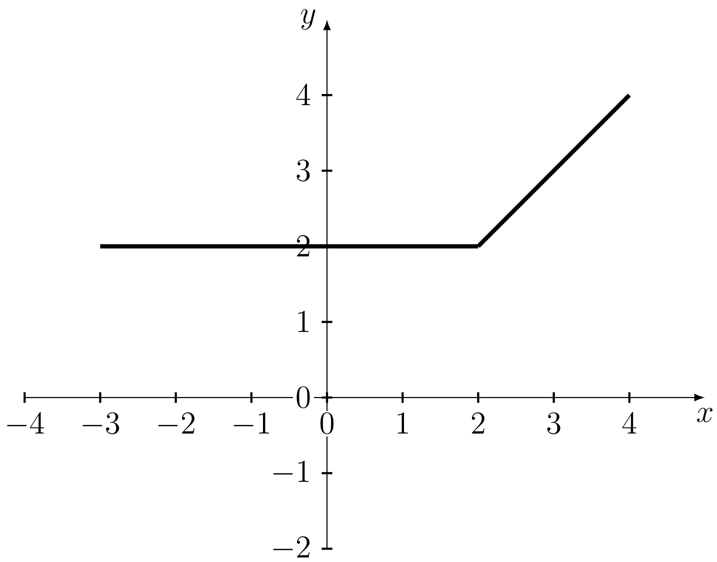


Figure 11: The slope of a function in the plane.

Question: Show the slope of a linear function is independent of the starting point. Could you suggest a mathematical proof for it? (*tip*: Use fundamental results of plane geometry.)
Answer:

Let (x_0, y_0) and (x_1, y_2) be arbitrary points on a linear function. Then the slope of the function is:

$$m = \frac{y_1 - y_0}{x_1 - x_0}$$

Can you prove this mathematically? (*tip*: Use fundamental results of plane geometry.)

Example: Find the slope of following functions:

- $y = x$

- $y = 3x + 2$

- $y = -5x + 6$

- $y = .25x - 1$

4.2 The equation of a line

The second component of a linear function is the $y - intercept$, i.e. b. The coordinate of the $y - intercept$ on the plane is $(0, b)$ For any given slope and $y - intercept$, we can find the equation of its associated linear function. Using (x, y), $(0, b)$, and m, we can find the equation as follow:

$\frac{y-b}{x-0} = m \Rightarrow y - b = m(x - 0) \Rightarrow y - b = mx \Rightarrow y = mx + b$

Example: Use two given coordinates and find the equation of its linear function.

- $(0, 3)$ and $(5, 5.5)$

- $(4, -1)$ and $(7, -7)$

- $(2, .5)$ and $(6, -.5)$

- $(2, 4)$ and $(5, 14.5)$

- (x, y) and (x_0, y_0)

4.3 Interpreting the slope of a linear function

The slope of a linear function is a key concept. It shows how much the value of function, the outcome/dependent variable, changes for one unit increase in the size of input variable, independent variable. The estimated coefficients in a regression model are the slope of function for any given regressor. For linear functions, the slope of function does not change by changes in the independent variables. However, this does not hold for nonlinear functions such as Logit and Probit. The slope of a function also is known as the *marginal effect* of independent variable.

5 The slope of non-linear functions

The slope of a linear function as a measure of its marginal effect is a key concept for linear functions. However, in political science, we also use and work with non-linear functions. How do we measure the marginal effects of these nonlinear functions? Suppose that our nonlinear function is $y = f(x)$, and we are at point $(x_0, f(x_0))$ on the graph of f, as shown in Figure 14.

We want to measure the rate of change of f or the steepness of the graph of f when $x = x_0$. A natural solution to this problem is to draw the tangent line to the graph at x_0, Figure 14.

We should note that for nonlinear functions, unlike linear functions, the slope of the tangent line will vary from point to point.

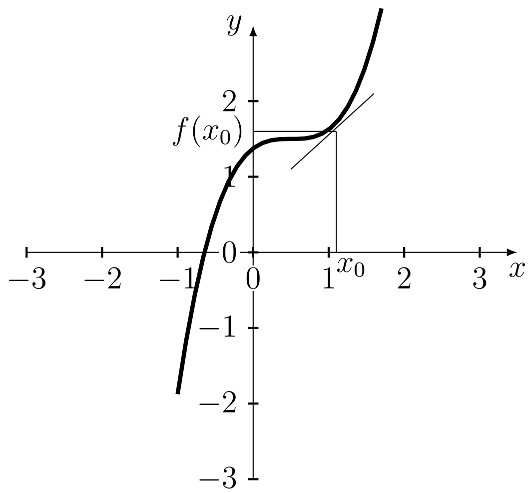


Figure 12: The graph of a nonlinear function.

We call the slope of the tangent line to the graph of f at $(x_0, f(x_0))$ the derivative of f at x_0, and we write it as

$$f'(x_0) \text{ or } \frac{df}{dx}|_{x=x_0}$$

This is similar to the slope formula we discussed before, i.e. $\frac{\Delta y}{\Delta x}$. Before extending our discussion about the derivative of a function, we need to learn more about *limits*.

6 Limits and open sets

Studying the marginal effects of an independent variable, i.e. x, on the outcome variable, i.e. y, requires studying these effects as small changes around a specified nearby, i.e. $x = x_0$. Thus, we need to learn about the concepts of *small changes* and *nearby*. These concepts help us to understand the concept of limits better which is one of the main concepts required for learning the derivative of functions, known as marginals effects among political scientists.

6.1 Sequences of real numbers

The *natural numbers*, also called the *positive integers*, are just the usual counting numbers: $1, 2, 3, 4, ...$

A *sequence* of real numbers is an assignment of real numbers is an assignment of a real number to each natural number. A sequence is usually written as $\{x_1, x_2, x_3, ..., x_n, ...\}$, where x_1 is the real number assigned to the natural number 1, the first number in the sequence; x_2 is the real number assigned to 2, the second number in the sequence, and so on.

Example: Some examples of a sequence of real numbers are:

- $\{1, 2, 3, 4, ...\}$
- $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...\}$
- $\{1, \frac{1}{2}, 4, \frac{1}{8}, 16, ...\}$
- $\{0, -\frac{1}{2}, \frac{2}{3}, -\frac{3}{4}, \frac{4}{5}, ...\}$
- $\{3.1, 3.14, 3.141, 3.1415, ...\}$
- $(-1)^n =$
- $\frac{1}{n} =$
- $(-1)^n \frac{1}{n} =$
- $\frac{n}{n+1} =$
- $\sqrt{n} =$

We sometimes write a typical sequence $\{x_1, x_2, x_3, ...\}$ as $\{x\}_{n=1}^{\infty}$.

6.2 Limits of a sequence

There are generally three kinds of sequences:

1. Sequences like $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...\}$ and $\{3.1, 3.14, 3.141, 3.1415, ...\}$ in which the entries get closer and closer and stay close to some limiting value;(The sequence converges)
2. Sequences like $\{1, 2, 3, 4, ...\}$ and $\{1, 4, 9, 16, ...\}$ in which the entries increase without bound, (The sequence diverges) and
3. Sequences like $\{1, \frac{1}{2}, 4, \frac{1}{8}, 16, ...\}$ and $\{0, -\frac{1}{2}, \frac{2}{3}, -\frac{3}{4}, \frac{4}{5}, ...\}$ in which neither behavior occurs so that the entries jump back and forth on the number line.

We are mostly interested in the first type of sequence. The entries in this kind of sequence *approach* arbitrarily close and *stay* close to some real number, called the *limit* of the sequence. We need both parts of this statement. The entries in $\{1, \frac{1}{2}, 4, \frac{1}{8}, 16, ...\}$ approach to some number, i.e. zero, but they do not stay there. What about $\{0, -\frac{1}{2}, \frac{2}{3}, -\frac{3}{4}, \frac{4}{5}, ...\}$?

To define the concept of a limit carefully, we need to formalize the notion that number s is close to number r if s lies in some small interval around r. More precisely, let ϵ denote a small, positive real number, as is the custom in mathematics. Then, the ϵ–interval around the number r is defined to be the interval

$$l - \epsilon(r) = \{s \in \mathbb{R} : |s - r| < \epsilon\} \tag{3}$$

In interval notation, $I_{\epsilon(r)} = (r - \epsilon, r + \epsilon)$. Intuitively speaking, if s is in $l_\epsilon(r)$ and if ϵ is small, then s is 'close' to r. The smaller ϵ is, the closer s is to r.

Definition: Let $\{x_1, x_2, x_3, ...\}$ be a sequence of real numbers and let r be a real number. We say that r is the limit of this sequence if for any (small) positive number ϵ, there is a positive integer N such that for all $n \geq N$, x_n is in the ϵ–interval about r; that is

$$|x_n - r| < \epsilon$$

In this case, we say that the sequence *converges* to r, and we write

$$\lim_{n \to \infty} x_n = r \quad \text{or} \quad \text{simply} \quad x_n \longrightarrow r$$

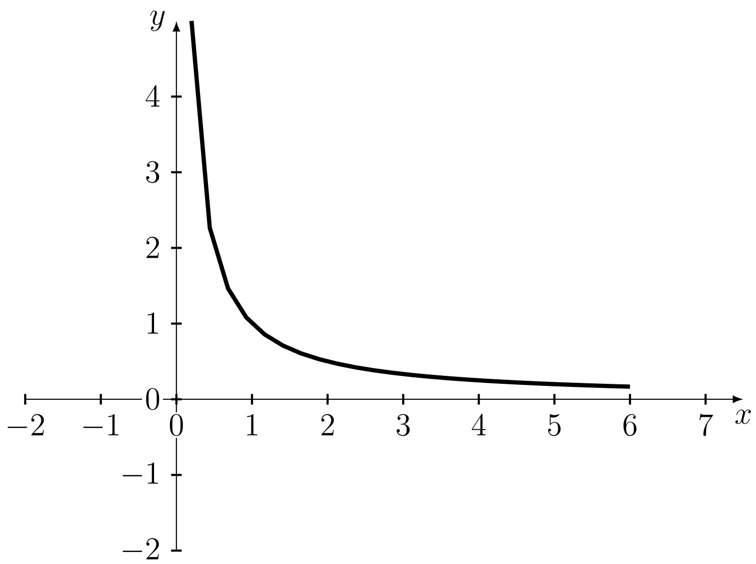


Figure 13: The limit of a function.

Example: Discuss the limit of following sequences

- $\{1, 0, \frac{1}{2}, 0, \frac{1}{3}, 0, ...\}$

- $\{1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, ...\}$

- $\{1, 3, \frac{1}{2}, \frac{3}{2}, \frac{1}{3}, \frac{3}{3}, \frac{1}{4}\}$

Remember that convergence does not need to be monotonic.

Theorem: A sequence can have at most one limit.
Proof:

6.3 Algebraic properties of limits

Theorem: Let $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ be sequence with limits x and y, respectively. Then, the sequence $\{x_n + y_n\}_{n=1}^{\infty}$ converges to the limit $x + y$.

Theorem: Let $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ be sequence with limits x and y, respectively. Then, the sequence $\{x_n y_n\}_{n=1}^{\infty}$ converges to the limit xy.

7 Computing derivatives

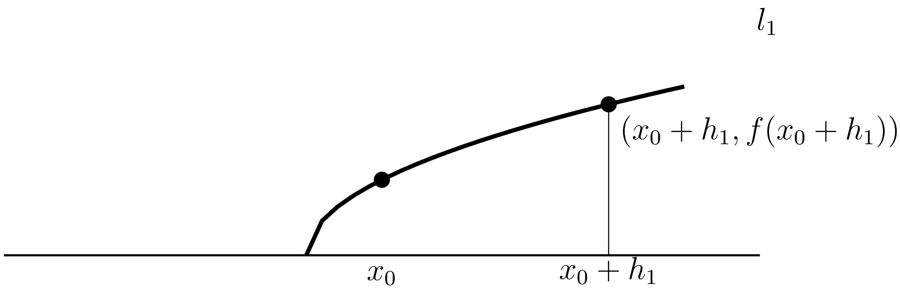


Figure 14: Approximating the tangent line by a sequence of secant lines.

Since l_n passes through the two points $(x_0, f(x_0))$ and $(x_0 + h_n, f(x_n + h_n))$, its slope is

$$\frac{f(x_0+h_n)-f(x_0)}{(x_0+h_n)-x_0} = \frac{f(x_0+h_n)-f(x_0)}{h_n}. \tag{4}$$

Therefore, the slope of the tangent line is the limit of this process as h_n converges to 0.

Definition Let $(x_0, f(x_0))$ be a point on the graph of y=f(x). The derivative of f at x_0, written

$$f'(x_0) \text{ or } \tfrac{df}{dx}(x_0) \text{ or } \tfrac{dy}{dx}(x_0),$$

is the slope of the tangent line to the graph of f at $(x_0, f(x_0))$. Analytically,

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h)-f(x_0)}{h} \tag{5}$$

Example: Use above formula to compare the derivative of the simplest nonlinear function, $f(x) = x^2$, at the point $x_0 = 2$. (*tip:*Choose a sequence of h_n's converging to zero.)

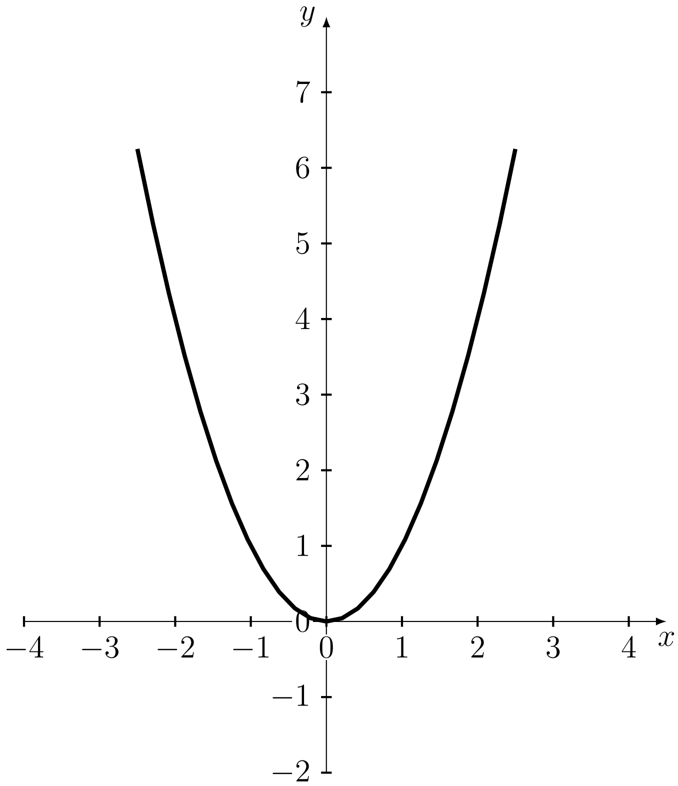


Figure 15: $y = x^2$

h_n	$x_0 + h_n$	$f(x_0 + h_n)$	$\frac{f(x_0+h_n)-f(x_0)}{h_n}$

7.1 Differentiability and continuity

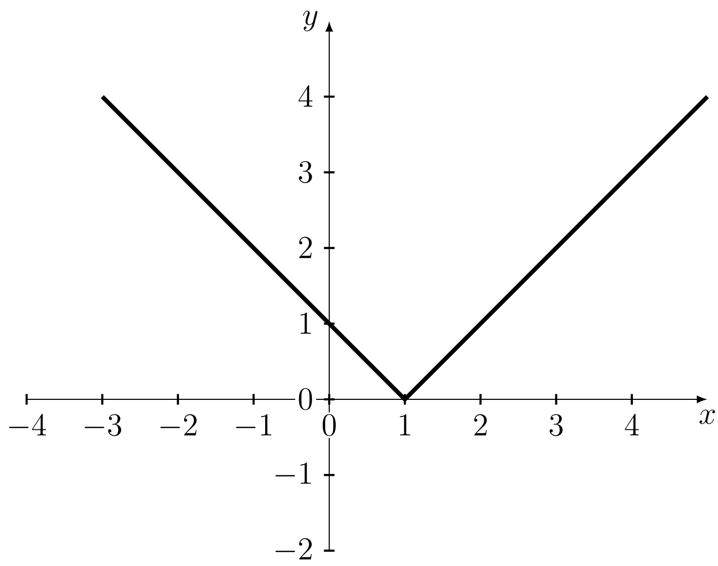
As we saw above, a function f is differentiable at x_0 if, geometrically speaking, its graph has a tangent line at $(x_0, f(x_0))$, or analytically speaking, the limit

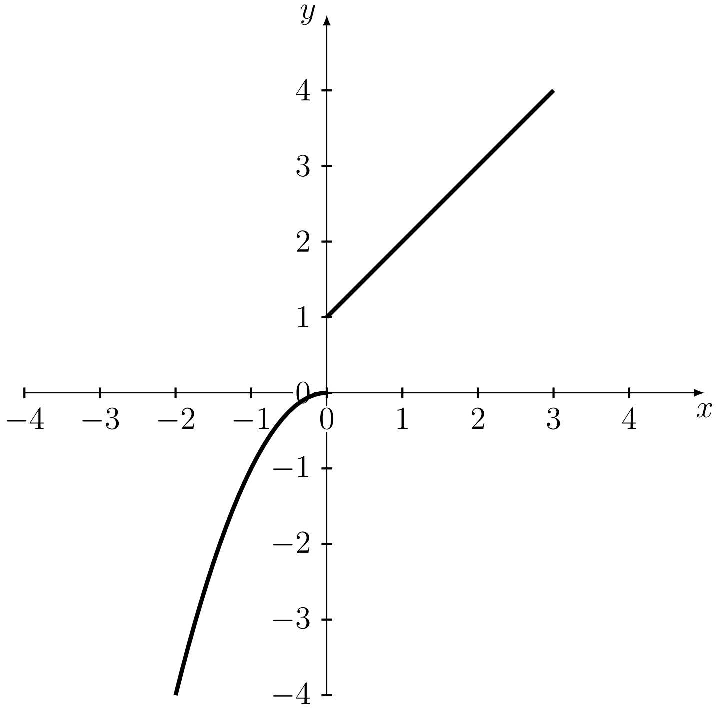
$$\lim_{h_n \to 0} \frac{f(x_0 + h_n) - f(x_0)}{h_n} \tag{6}$$

exists and is the same for every sequence h_n which converges to 0. if a function is differentiable at every point x_0 in its domain D, we say that the function is *differentiable*.

Only functions whose graphs are 'smooth curves' have tangent lines everywhere; in fact, mathematicians commonly use the word 'smooth' in place of the word 'differentiable'.

Example: Show following functions are not differentiable on their domain:




y
x
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4

7.2 Rules for computing derivatives

Theorem: For any positive integer k, the derivative of $f(x) = x^k$ at x_0 is $f'(x_0) = kx_0^{k-1}$.

7.2.1 Common derivatives

- $\frac{d}{dx}(x) = 0$
- $\frac{d}{dx}(e^x) = e^x$
- $\frac{d}{dx}(a^x) = a^x.ln(a)$
- $\frac{d}{dx}(log_a x) = \frac{1}{x.ln(a)}$
- $\frac{d}{dx}(lnx) = \frac{1}{x}$

Theorem: Suppose that k is an arbitrary constant and that f and g are differentiable functions at $x = x_0$. Then,

- $(f + g)'(x_0) = f'(x_0) + g'(x_0)$,
- $(f - g)'(x_0) = f'(x_0) - g'(x_0)$,
- $(kf)'(x_0) = k.f'(x_0)$,
- $(\frac{f}{g}(x_0))' = \frac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}$
- $((f(x))^n)' = n(f(x))^{n-1}.f'(x)$

Examples: Find the derivative of the following functions:

- $-7x^3$
- $3x^{-\frac{3}{2}}$
- $3x^2 - 9x + 7x^{\frac{2}{5}} - 3x^{\frac{1}{2}}$
- $(x^2 + 1)(x^2 + 3x + 2)$
- $\frac{x-1}{x+1}$
- $\frac{1-x}{x+1}$
- $\sqrt{x}$
- $\sqrt[3]{x}$

8 One Variable Calculus: Applications

8.1 Using the first derivative for graphing

The derivative of a function carries much information about the important properties of the function. In this section, we will see that knowing the signs of a function's first and second derivatives and the location of only few points on its usually enable us to draw an accurate graph of the function.

Theorem: Let f be a continuously differentiablefunction on domain $D \subset \mathbb{R}^1$

(a) If $f' > 0$ on interval $(a, b) \subset D$, then f is increasing on (a, b).

(b) If $f' < 0$ on interval $(a, b) \subset D$, then f is decreasing on (a, b).

(c) If f is increasing on (a, b), then $f' \geq 0$ on (a, b).

(d) If f is decreasing on (a, b), then $f' \leq 0$ on (a, b).

To use the above theorem to sketch the graph of a given function f, we need to find the intervals where $f' > 0$ and the intervals where $f' < 0$. To accomplish this:

1. First find the points at which $f'(x) = 0$ or f' is not defined. Such points are called *critical points* of f.
2. Evaluate the function at each of these critical points $x_1, x_2, ..., x_k$, and plot the corresponding points on the graph.
3. Then, check the sign of f' on each of the intervals
$$(-\infty, x_1), (x_1, x_2), ..., (x_{k-1}, x_k), (x_k, \infty)$$
4. On any one of these intervals, f' is defined and nonzero. Since $f'(x) = 0$ only when $x = x_1, ..., x_k$ and since f' is continuous, f' cannot change sign on any of these intervals; it must be either always negative or always positive on each. To see whether f' is positive or negative on any one of these intervals, one need only check the sign of f' at one convenient point in that interval.
5. If $f' > 0$ on interval I, draw the graph of f increasing over I. If $f' < 0$ on I, draw a decreasing graph over I.

Example: Consider the cubic function $f(x) = x^3 - 3x$, and graph it.

8.2 Second derivatives and convexity

We need to know more about the shape of the graph than where it is increasing and where decreasing. Some of the functions that we work with them are decreasing return to scale, meaning the marginal effects decreases as the independent variable increases (learning function, utility function, welfare function).

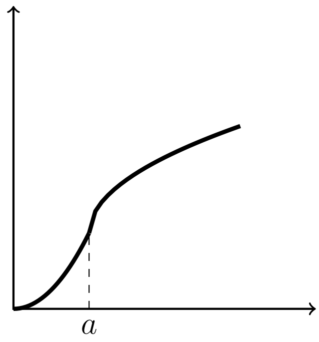


Figure 16: A decreasing return to scale function.

For $x \in (0, a)$ in Figure 16, the slope of $f'(x)$ is increasing function. In fact, the derivative of f', i.e. $f''(x)$, is nonnegative there: $f''(x) \geq 0$ on $(0, a)$. For $x > a$ in Figure 16, f' is decreasing function x; so $f''(x) \leq 0$ on (a, ∞).

A differentiable function f for which $f''(x) \geq 0$ on an interval I (f' is increasing on I) is said to be *concave up* or *convex* on I.

A differentiable function f for which $f''(x) \leq 0$ on an interval I (f' is decreasing on I) is said to be *concave down* on I.

Increasing and decreasing functions can be concave up or concave down on their interval of increase and decrease, respectively.

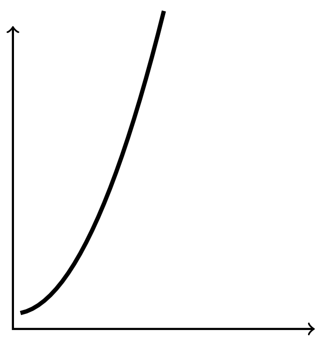


Figure 17: An increasing concave up function.

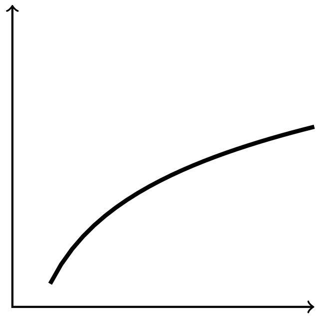


Figure 18: An increasing concave down function.

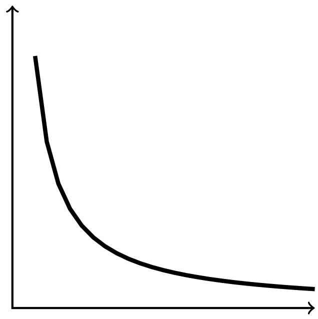


Figure 19: A decreasing convex function.

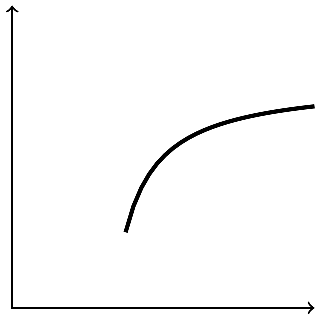


Figure 20: A decreasing concave function.

Example: For below functions, compute the regions of decreasing, increasing, convexity, and concavity.

- $x^3 + 3x$

- $x^4 - 8x^3 + 18x^2 - 11$

- $\frac{1}{3}x^3 + 9x + 3$

- $x^7 - 7x$

8.3 Maxima and Minima

One of the major uses of calculus in mathematical models is to find and characterize maxima and minima of functions.

A function f has a local or relative maximum at x_0 if $f(x) \leq f(x_0)$ for all x in some open interval containing x_0; f has a global or absolute maximum at x_0 if $f(x) \leq f(x_0)$ for all x in the domain of f.

The function f has a local or relative minimum at x_0 if $f(x) \geq f(x_0)$ for all x in some open interval containing x_0; f has a global or absolute minimum at x_0 if $f(x) \geq f(x_0)$ for all x in the domain of f.

If f has a local maximum (minimum) at x_0, we will simply say that x_0 is a max (min) of f. If we want to emphasize that f has a global maximum (minimum) at x_0, we will say that x_0 is a *global max* (*global min*) of f.

8.4 Local maxima and minima on the boundary and in the interior

A max or min of a function can occur at an endpoint of the domain of f or at a point which is not an endpoint–in the *interior* of the domain.

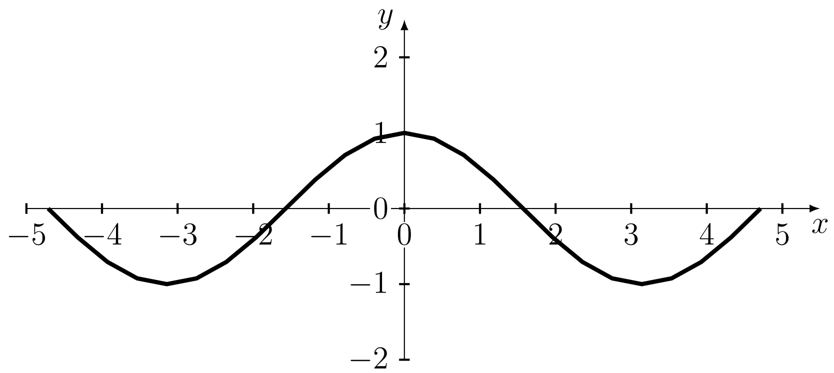


Figure 21

Theorem: If x_0 is an interior max or min, then x_0 is a critical point of f.

8.5 Second order conditions

If x_0 is a critical point of a function f, how can we use calculus to decide whether critical point x_0 is a max, a min, or neither? The answer to this question lies in the *second* derivative of f at x_0.

Theorem:

(a) If $f'(x_0) = 0$ and $f''(x_0) < 0$, then x_0 is a max of f.

(b) If $f'(x_0) = 0$ and $f''(x_0) > 0$, then x_0 is a min of f.

(c) If $f'(x_0) = 0$ and $f''(x_0) = 0$, the second derivative does not provide any significant information about max or min of function at x_0.

8.6 Global maxima and minima

In general, it is difficult to find a global max of a function or even to prove that a given local max is a global max. These are, however, three situations in which this problem is somewhat easier:

- when f has only one critical point in its domain.
- when $f'' > 0$ or $f'' < 0$ throughout the domain of f, and
- when the domain of f is a closed finite interval.

8.7 Functions with only one critical point

Theorem: Suppose that:

(a) the domain of f is an interval I (finite or infinite) in $\mathbb{R}^1$

(b) x_0 is a local maximum of f, and

(c) x_0 is the only critical point of f on I.

Then, x_0 is the global maximum of f on I.

8.8 Functions with nowhere-zero second derivatives

Theorem: If f is a C^2 function whose domain is an interval I and if f'' is never zero on I, then f has at most one critical point in I. This critical point is a global minimum if $f'' > 0$ and a global maximum if $f'' < 0$.

9 Exponents and Logarithms

In this section, we learn about exponential and logarithmic functions as they are used in some quantitative and formal models of political science.

9.1 Exponential functions

We learned about polynomial functions and used them in previous parts of this course. In this section, we study exponential and logarithmic functions, their properties, and derivatives.

For exponential functions, the variable x appears as an *exponent*. A simple example is $f(x) = 2^x$, a function whose domain is all the real numbers. In this case, 2 is called the *base* of the exponential function. To understand this exponential function better, let's draw its graph. Since we do not know how to take the derivative of 2^x yet, we will have to plot points.

x	2^x
-3	$\frac{1}{8}$
-2	$\frac{1}{4}$
-1	$\frac{1}{2}$
0	1
1	2
2	4
3	8

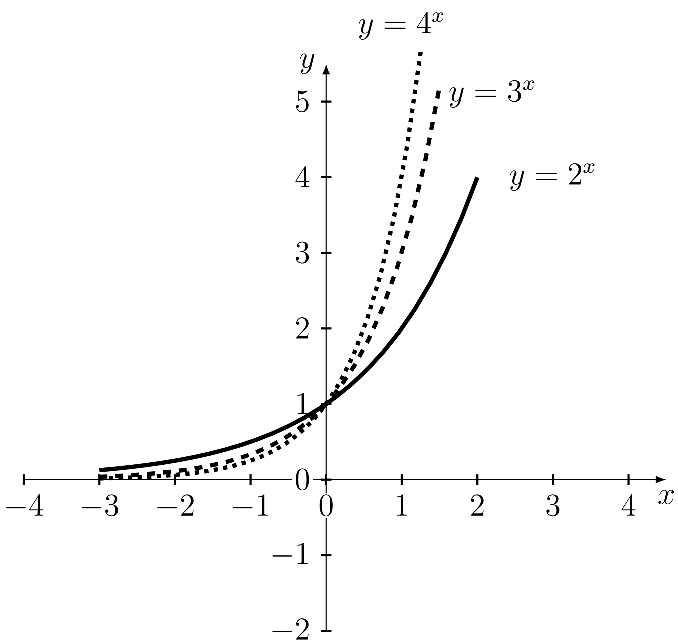


Figure 22: $y = 2^x$

Figure 23 show different exponential functions for different bases, henceforth b, where $b > 1$. However, when $0 < b < 1$, the exponential function shows different behavior.

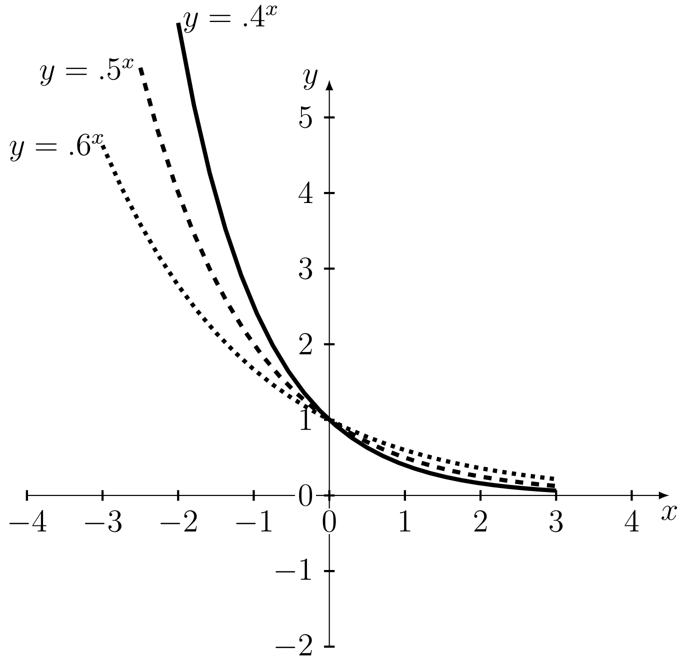


Figure 23

9.2 The number e

Here, we introduce a number which is one of the most important base for an exponential function, the number $e = 2.7182818284$. The exponential functions with base e are usually used for studying the growth of a variable. It also is used in modeling models with discrete dependent variable.

9.3 Logarithms

Consider a general exponential function, $y = a^x$, with base $a > 1$. Such an exponential function is a strictly increasing function:

$$x_1 > x_2 \Longrightarrow a^{x_1} > a^{x_2} \tag{7}$$

Logarithm is the inverse function of $y = a^x$. Consider the linear function $y = mx + b$, the inverse of this function is $y = \frac{x-b}{m}$. To find the inverse of a function, we just replace y and x, then we solve the equation for y. However, for an exponential function, after replacing x and y, we have: $x = a^y$, and we cannot solve this equation for y. Thus, we use logarithm as follow:

$$y = log_a x \iff a^y = x \tag{8}$$

The logarithm of x, by definition, is the power to which one must raise a to yield x. This definition implies that:

$$log_a a^x = x \quad and \quad a^{log_a x} = x \tag{9}$$

9.4 Base 10 logarithms

Let's first work with base $a = 10$. The logarithmic function for base 10 us such a commonly used logarithm that it is usually written as $y = Logx$ with an uppercase L, or even $y = logx$:

$$y = logx \iff 10^y = x$$

Example:

- log10
- log100,000
- log1
- log625

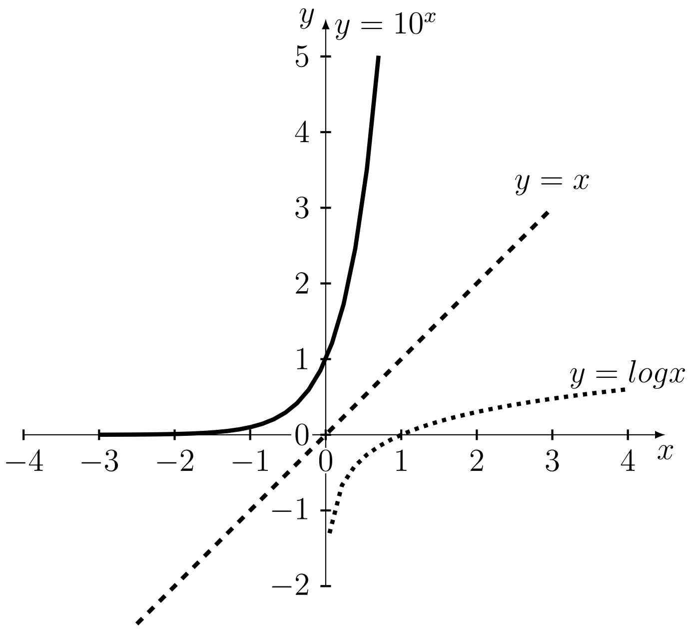


Figure 24

9.5 Base e logarithms/ Natural logarithm

Since the exponential function $exp(x) = e^x$ has all the properties that 10^x has, it also has an inverse. Its inverse works the same way that $logx$ does. Mirroring the fundamental role that e plays in applications, the inverse of e^x is called the *natural logarithm* function and is written as lnx. Formally,

$$lnx = y \Longleftrightarrow e^y = x \tag{10}$$

From the definition, we have

$$e^{lnx} = x \longleftrightarrow lne^x = x \tag{11}$$

9.6 Properties of Exp and Log

Exponential functions have the following five basic properties:

1. $a^r.a^s = a^{r+s}$
2. $a^{-r} = \frac{1}{a^r}$
3. $\frac{a^r}{a^s} = a^{r-s}$
4. $(a^r)^s = a^{rs}$, and
5. $a^0 = 1$

The five properties of exponential functions are mirrored by five corresponding properties of the logarithmic functions:

1. $log(r.s) = log\ r + log\ s$
2. $log(\frac{1}{s}) = -log\ s$
3. $log(\frac{r}{s}) = log\ r - log\ s$
4. $log\ r^s = s\ log\ r$
5. $log1 = 0$

9.7 Derivatives of Exp and Log

Theorem: The functions e^x and lnx are continuous functions on their domains and have continuous derivatives of every order. Their first derivatives are given by

(a) $(e^x)' = e^x$

(b) $(lnx) = \frac{1}{x}$

If $u(x)$ is differentiable function, then

(c) $(e^u(x))' = e^u(x).u'(x)$

(d) $(lnu(x))' = \frac{u'(x)}{u(x)}$

10 linear Algebra

The analysis of some political science models reduces to the study systems of equations. Below, we study the simplest possible system of equations, linear systems.

10.1 Linear systems

Typical linear equations are

$$x_1 + 2x_2 = 3 \text{ and } 2x_1 - 3x_2 = 8$$

As we discussed before, they are called linear because their graphs are straight lines. In general, an equation is *linear* if it has the form

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b \tag{12}$$

The letters $a_1, a_2, ..., a_n$, and b stands for fixed numbers such as 2, -3, and 8 in the second equation above. These are called *parameters*. The letters $x_1, x_2, ..., x_n$ stand for variables.

Although the real world is not linear, social scientists widely use linear models. However, this is not only because of manageability of linear systems, but we can learn about the behavior of nonlinear models via studying suitably chosen linear approximation to the system. For example, as shown before, the behavior a nonlinear function can be studied around a specific point using the tangent line which is a linear approximation.

10.2 Elimination of variables

We need to solve the systems if equations that we develop. One method that you might remember from high school is *substitution*. Another method which is most conducive to theoretical analysis is *elimination of variables*. Let's review them below:

$$x_1 - 2x_2 = 8$$
$$3x - 1 + x_2 = 3$$

Substitution:

Elimination of variables:

Example: Solve following linear system:

$$
\begin{gathered}
x_1 - .4x_2 - .3x_3 = 130 \\
-.2x_1 + .88x_2 - .14x_3 = 74 \\
-.5x_1 - .2x_2 + .95x_3 = 95
\end{gathered}
$$

11 Vectors

We use coordinates to describe locations in exactly the same way in higher dimensions. We can interpret n-tuples as displacements. This is a useful way of thinking about vectors for doing calculus.

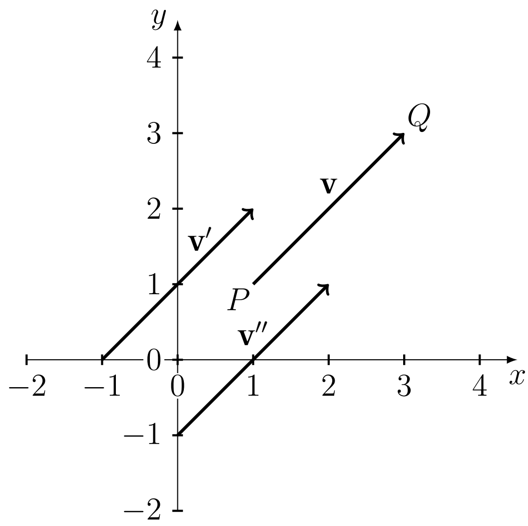


Figure 25: Vectors

For example, the tail of the displacement labeled **v** in Figure 28 is at the location $(1,1)$, and the head is at $(3,3)$. We will sometimes write $\vec{PQ}$ for the displacement whose tail is at the point P and head at the point Q. Although **v**, **v**′, and **v**″ represent displacements from different locations, they are equal since they have same length and direction. Thus, **v**, **v**′, and **v**″ are equivalent. We find the elements of a vector from point $P = (a_1, a_2, ..., a_n)$ to $Q = (b_1, b_2, ..., b_n)$ as follow:

$$\vec{PQ} = (b_1 - a_1, b_2 - a_2, ..., b_n - a_n) \tag{13}$$

Example: For points P and Q listed below, draw the corresponding displacement vector $\vec{PQ}$ and compute the corresponding n-tuple for $\vec{PQ}$:

a) $P(0,0)$ and $Q(2,-1)$

b) $P(3,2)$ and $Q(1,1)$

c) $P(3,2)$ and $Q(5,3)$

d) $P(0,1)$ and $Q(3,1)$

e) $P(0,0,0)$ and $Q(1,2,4)$

f) $P(0,1,0)$ and $Q(2,-1,3)$

12 The algebra of vectors

12.1 Addition and subtraction

We add two vectors just as we add two numbers. We simply add separately the corresponding coordinates of the two vectors. Thus,

$$(3, 2) + (4, 1) = (7, 3)$$

So,

$$(x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$$

$$(x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$$

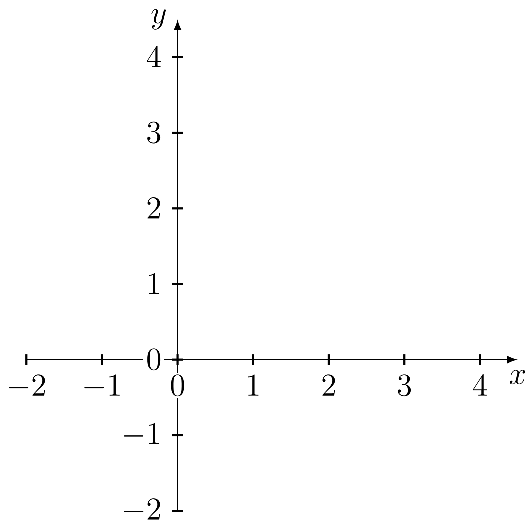


Figure 26: The geometry of $x + y = y + x$

$$(x_1, x_2, ..., x_n) - (y_1, y_2, ..., y_n) = (y_1 - x_1, y_2 - x_2, ..., y_n - x_n)$$

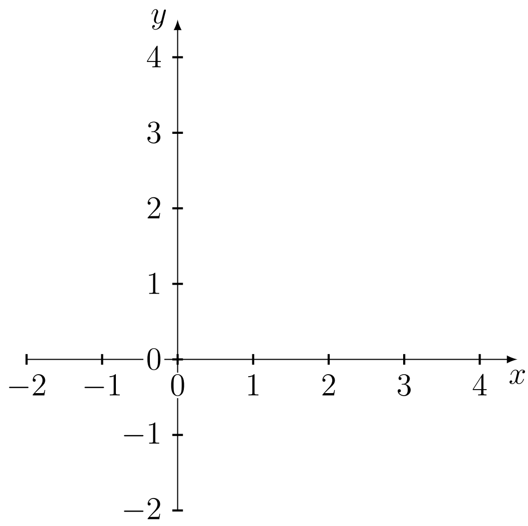


Figure 27: The geometry of $x - y$

12.2 Scalar multiplication

If r is a scalar and $\mathbf{x} = (x_1, ..., x_n)$ is a vector, then their product is

$$r.\mathbf{x} = (rx_1, ..., rx_n) \tag{14}$$

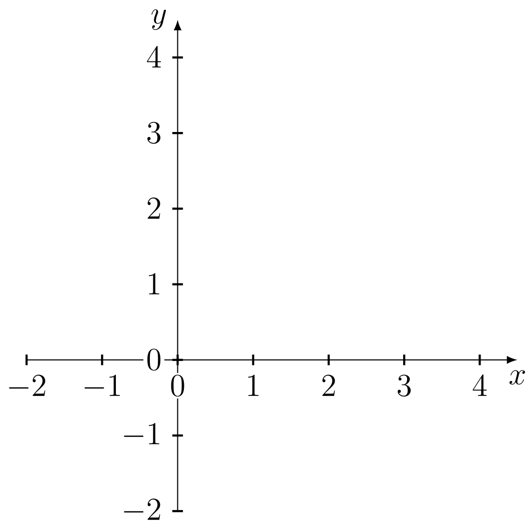


Figure 28: The geometry of $r.\mathbf{x}$

Review: The distributive laws:

$$a.(b + c) = ab + ac \text{ and } (a + b).c = ac + bc$$

The distributive laws apply also to vectors. For all scalars **r** and **s**, and vectors **u** and **v**, we have:

a) $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$

b) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$

12.3 Length and inner product in $\mathbb{R}^n$

$$|PQ| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + ... + (b_n - a_n)^2} \tag{15}$$

Theorem: $|r\mathbf{v}| = |r|.|v| \; \forall \; r \in \mathbb{R}^1$ and $\mathbf{v} \in \mathbb{R}^n$.
Example: Find the length of $(1, -2, 3) \in \mathbb{R}^3$.

12.4 The inner product

Definition: Let $\mathbf{u} = (u_1, ..., u_n)$ and $\mathbf{v} = (v_1, ..., v_n)$ be two vectors in $\mathbb{R}^n$. The inner product, also known as dot product, of **u** and **v**, is the number

$$\mathbf{u}.\mathbf{v} = u_1v_1 + u_2v_2 + ... + u_nv_n \tag{16}$$

Example: If **u**=(4,-1,2) and **v**=(6,3,-4), what is the inner product of **u** and **v**?
Theorem: Let u, v, and W be arbitrary versions in $\mathbb{R}^n$ and let r

be an arbitrary scalar. Then,

a) $\mathbf{u}.\mathbf{v} = \mathbf{v}.\mathbf{u}$

b) **u.(v+w)** $=$ **u.v+u.w**

c) $\mathbf{u}.(r\mathbf{v}) = r(\mathbf{u}.\mathbf{v}) = (r\mathbf{u}).\mathbf{v}$

d) $\mathbf{u}.\mathbf{u} \geq 0$

e) $\mathbf{u}.\mathbf{u} = 0$ implies **u=0**

f) $(\mathbf{u} + \mathbf{v}).(\mathbf{u} + \mathbf{v}) = \mathbf{u}.\mathbf{u} + 2(\mathbf{u}.\mathbf{v}) + \mathbf{v}.\mathbf{v}$

Example: Find the length of the following vectors. Draw the vectors for *a* through *g*:

a) $(3, 4)$

b) $(0, -3)$

c) $(1, 1, 1)$

d) $(3, 3)$

e) $(-1, -1)$

f) $(1, 2, 3)$

g) $(2, 0)$

h) $(1, 2, 3, 4)$

i) $(3, 0, 0, 0, 0)$

Example: Find the distance from P to Q, drawing the picture wherever possible:

a) $P(0, 0)$, $Q(3, -4)$

b) $P(1, -1)$, $Q(7, 7)$

c) $P(5, 2)$, $Q(1, 2)$

d) $P(1, 1, -1)$, $Q(2, -1, 5)$

e) $P(1, 2, 3, 4)$, $Q(1, 0, -1, 0)$

13 Matrix algebra

A *matrix* is simply a rectangular array of numbers. So, any table of data is a matrix. The size of a matrix is indicated by the number of its rows and the number of its columns. A matrix with k rows and n columns is called a $k \times n$ ("k by n") matrix. The number in row i and column j is called the (i, j)th entry, and is often written a_{ij}. Two matrices are *equal* if they both have the same size and if the corresponding entries in the two matrices are equal.

Matrices are in a sense generalized numbers. When the sizes are right, two matrices can be added, subtracted, multiplied, and even divided.

13.1 Addition

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & b_{ij} & \vdots \\ b_{k1} & \dots & b_{kn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & b_{ij}+b_{ij} & \vdots \\ a_{k1}+b_{k1} & \dots & a_{kn}+b_{kn} \end{bmatrix}$$

13.2 Subtraction

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix} - \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & b_{ij} & \vdots \\ b_{k1} & \dots & b_{kn} \end{bmatrix} = \begin{bmatrix} a_{11}-b_{11} & \dots & a_{1n}-b_{1n} \\ \vdots & b_{ij}-b_{ij} & \vdots \\ a_{k1}-b_{k1} & \dots & a_{kn}-b_{kn} \end{bmatrix}$$

13.3 Scalar multiplication

$$r\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \dots & a_{kn} \end{bmatrix} = \begin{bmatrix} ra_{11} & \dots & ra_{1n} \\ \vdots & ra_{ij} & \vdots \\ ra_{k1} & \dots & ra_{kn} \end{bmatrix}$$

13.4 Matrix multiplication

Not all pairs of matrices can be multiplied together, and the order in which matrices are multiplied can matter.

We can define the matrix product AB if and only if

number of columns of A=number of rows of B

For the matrix product to exist, A must be $k \times m$ and B must be $m \times n$. To obtain the (i, j)th entry of AB, multiply the ith row of A and the jth column of B as follows:

$$r \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{im} \end{bmatrix} \cdot \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{bmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}$$

For example,

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} aA + bC & aB + bD \\ cA + dC & cB + dD \\ eA + fC & eB + fD \end{bmatrix}$$

The $n \times n$ matrix

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

has the property that for any $m \times n$ matrix A,

$$AI = A$$

and for any $n \times 1$ matrix B,

$$IB = B$$

The matrix I is called the $n \times n$ *identity matrix* because it is a multiplicative identity for matrices just as the number 1 is for real numbers.

13.5 Laws of matrix algebra

- Associative laws: $(A + B) + C = A + (B + C)$, $(AB)C = A(BC)$
- Commutative law for addition: $A + B = B + A$
- Distributive laws: $A(B + C) = AB + AC$ $(A + B)C = AC + BC$

13.6 Transpose

There is one other operation on matrices which we shall frequently use. The *transpose* of $k \times n$ matrix A is the $n \times k$ matrix obtained by interchanging the rows and columns of A. This matrix is often written as A^T. The first row of A becomes the first column of A^T. The second row of A becomes the second column of A^T, and so on. For example,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

The properties of transpose:

- $(A + B)^T = A^T + B^T$
- $(A - B)^T = A^T - B^T$
- $(A^T)^T = A$
- $(rA)^T = rA^T$
- $(AB)^T = B^T A^T$

14 Statistics

14.1 Descriptive Statistics

The application of statistical thinking involves two sets of processes. First, there is the description and presentation of data. Second, there is the process of using the data, i.e. a sample, to make some inference about features of the environment from which the data were selected or about the underlying mechanism that generated the data, i.e. the population. The first is called *descriptive statistics* and the second *inferential statistics*.

14.2 Measures of Central Tendency

Often, data tends to group itself around some central value. There are three main measures of central tendency used by psychologists. They are the mean, the median, and the mode.

14.2.1 Mean

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \tag{17}$$

This statistics measure is also know as the average of a variable in a sample.

- Strengths: the most sensitive measure of central tendency.
- Weaknesses: easily distorted by outliers or skewed data (data that is clumped together near one extreme

Example: Find the mean of: 6, 8, 11, 5, 2, 9, 7, 8

14.2.2 Median

The median value of a set of data is the middle value of the ordered data. That is, the data must be put in numerical order first. ***Example:*** Find the median of the following:

a) 11, 4, 9, 7, 10, 5, 6

b) 1, 3, 0.5, 0.6, 2, 2.5, 3.1, 2.9

- Strengths: Not distorted by outliers. Best when describing data that is highly skewed
- Weaknesses: Can be distorted by small samples and is less sensitive as the mean.

14.2.3 Mode

Find the mode for: 2, 6, 3, 9, 5, 6, 2, 6

- Strengths: Not influenced by extreme scores. Useful to show the most common/popular value. • Weaknesses: Crude measure of central tendency. Not useful if there are many equal modes.

Example:Find the mode, median and mean of the following:

a) 3, 12, 11, 7, 5, 5, 6, 4, 10

b) 16, 19, 10, 24, 19

c) 8, 2, 8, 5, 5, 8

d) 28, 39, 42, 29, 39, 40, 36, 46, 41, 30

e) 133, 215, 250, 108, 206, 159, 206, 178

f) 76, 94, 76, 82, 78, 86, 90

g) 52, 61, 49, 52, 49, 52, 41, 58

14.3 Measures of dispersion

Almost all data sets demonstrate some degree of variability. In other words, data sets usually contain scores which differ from one another. This variability or "dispersion" of data cannot be captured or shown by measures of central tendency.

14.3.1 Range

The range is simply the difference between the highest and lowest scores in a distribution, and is found by subtracting the lowest score from the highest score.

- Strengths: Quick and easy to calculate
- Weaknesses: Distorted by outliers or extremes

14.3.2 Variance

A more informative measure of variability is the variance, which represents the degree to which scores tend to vary from their mean. This tends to be more informative because, unlike the range, the variance takes into account every score in the data set.

$$Var(x) = \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{n} \tag{18}$$

14.3.3 Standard Deviation

More informative still is the standard deviation, which is simply the square root of the variance. You may be asking yourself 'Why not simply use the variance?' One reason is that, unlike the variance, the standard deviation is in the same units as the raw scores themselves. This is what makes the standard deviation more meaningful.

- Strengths: The most sensitive measure of dispersion. Uses all the data available.
- Weaknesses: A little more time consuming to calculate than other measures of dispersion.

14.4 Probability distributions

We first define the term probability, using a relative frequency approach. Imagine a hypothetical experiment consisting of a very long sequence of repeated observations on some random phenomenon. Each observation may or may not result in some particular outcome. The probability of that outcome is defined to be the relative frequency of its occurence, in the long run. ***Definition:*** The *probability* of a particular outcome is the proportion of times that outcome would occur in a long run of repeated observations.

A simplified representation of such an experiment is a very long sequence of flips of a coin, the outcome of interest being that a head faces upwards. Any on flip may or may not result in a head. If the coin is balanced, then a basic result in probability, called *law of large numbers*, implies that the proportion of flips resulting in a head tends toward $\frac{1}{2}$ as the number of flips increases. Thus, the probability of a head in any single flip of the coin equals $\frac{1}{2}$.

Most of the time we are dealing with variables which have numerical outcomes. A variable which can take at least two different numerical values in a long run of repeated observations is called *random variable*.

14.4.1 Properties of probability

PROPERTY 1: For any event A, the probability of A is a number between 0 and 1. That is,

$$0 \leq P(A) \leq 1 \tag{19}$$

PROPERTY 2: The probability of sample space S is 1. Symbolically,

$$P(S) = 1 \tag{20}$$

PROPERTY 3: The probability of the union of disjoint events is equal to the sum of the probabilities of these events. For instance, if A and B are disjoint, then

$$P(A \cup B) = P(A) + P(B) \tag{21}$$

Addition Rule: For any events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{22}$$

Example: A certain retail establishment accepts either the American Express or the VISA credit card. A total of 22 percent of its customers carry an American Express card, 58 percent carry a VISA credit card, and 14 percent carry both. What is the probability that a customer will have at least one of these cards?

We are often interested in determining probabilities when some partial information concerning the outcome of the experiment is available. In such situations, the probabilities are called conditional probabilities. The conditional probability is denoted by

$$P(B|A) \tag{23}$$

Example: Suppose two dice are to be rolled. Suppose further that the first die lands on 4. Given this information, what is the resulting probability that the sum of the dice is 10? To determine this probability, we reason as follows. Given that the first die lands on4, there are 6 possible outcomes of the experiment,namely,

$$(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6)$$

The conditional probability of each of the outcomes should be $\frac{1}{6}$. Since in only one of the outcomes is the sum of the dice equal to 10, namely, the outcome $(4, 6)$, it follows that the conditional probability that the sum is 10,given that the first die lands on 4, is $\frac{1}{6}$.

A general formula for the conditional probability is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{24}$$

Example: As a further check of the preceding formula for the conditional probability, use it to compute the conditional probability that the sum of a pair of rolled dice is 10, given that the first die lands on 4.