Math Review - MIDS W203 - Statistics

w203 Staff

20 August 2019

This document is designed to provide a quick refresher on exponents and logarithms, functions, differential and integral calculus, histograms, and density functions for MIDS W203 students.

This document is made available as both a PDF and an R Markdown document. It is full of examples of how to format reports using R Markdown and LaTeX, formally LaTeX(pronounced "LAY-Tek"). I encourage you to keep the source file handy through the semester as a mini-cookbook for making nice reports with many of the mathematical expressions you will use.

1. Mathematical Symbols and Nomenclature

wrt	with respect to
$\mathbb{R}$	The set of Real numbers
$\mathbb{Z}$	The set of Integers
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\forall$	For all (members of a set)
$\mid$	Such that (the pipe symbol, more common)
$\ni$	Such that (less common)
$\in$	In (a set)
$\approx$	Approximately equal to
$\{x \in A \mid 0 \leq Pr(x) \leq 1\}$	The set of all x in A such that the probability of x is greater than
	or equal to 0 and less than or equal to 1
$\{x \in \mathbb{R} : x > 0\}$	The set of all x in the set of Real numbers such that x is greater than zero

1

2. Rules of Logarithms and Exponents

$$a^n \times a^m = a^{n+m}$$
$$a^n \times b^n = (a \cdot b)^n$$
$$a^n / a^m = a^{n-m}$$
$$a^n / b^n = (a/b)^n$$
$$(b^n)^m = b^{n \cdot m}$$
$$b^{n^m} = b^{m^n}$$
$$\sqrt[m]{b^n} = b^{n/m}$$

$$log(uv) = log(u) + log(v)$$
$$log(u/v) = log(u) - log(v)$$
$$log(u^n) = n \cdot log(u)$$

3. Discrete Data and Summation Notation

Some data can assume only specific values or they form a series of values which are countable (although the series may be infinite). For discrete data, we common use subscripts to denote individual instances of a variable. Examples that will become very familiar include:

The sample mean:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

The sample variance:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

4. Functions and Functional Notation

a. What is a function?

A function is a mathematical rule that maps one or more input (independent) variables to a unique value of an output (dependent) variable. Here are some examples of discrete and continuous functions:

$$Pr(x) = \begin{cases} p(1-p)^{x-1} & \text{x = 1, 2, 3,...} \\ 0 & \text{otherwise.} \end{cases}$$

Can you think of a real-world relationship that this equation might represent?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

This is the equation for the multivariate linear regression model. You will know a great deal about this equation by the end of the term...
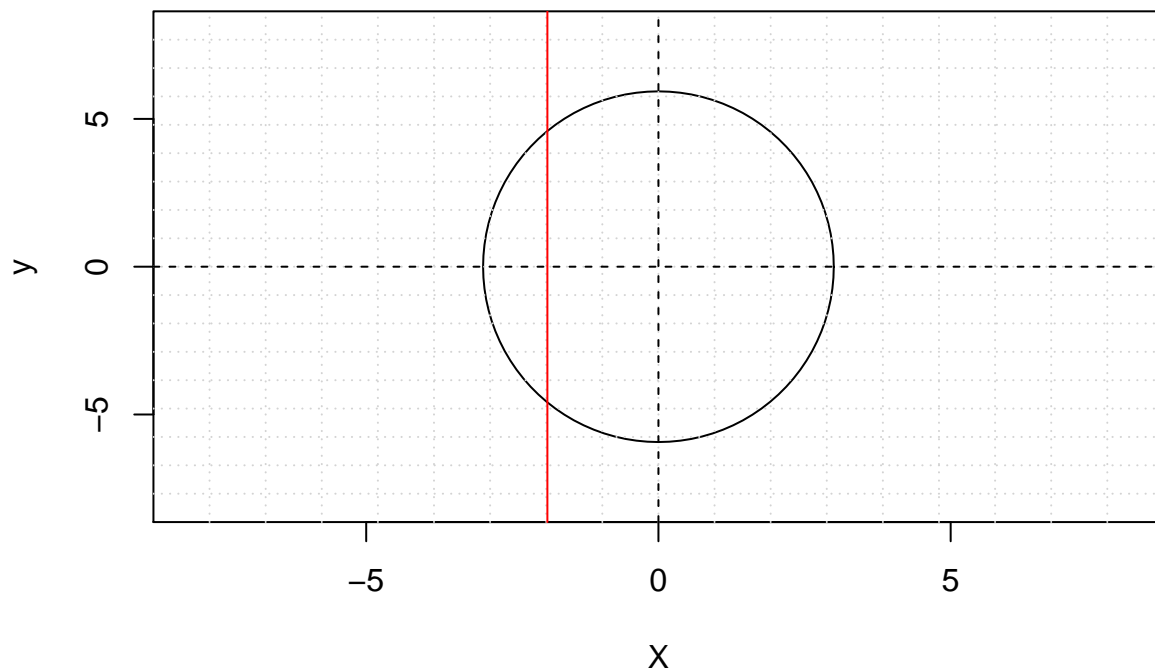
b. What is not a function:

The following rule is not a function:

$$9 = x^2 + y^2$$

It is the equation for a circle of radius 3.

```
## Warning: package 'plotrix' was built under R version 3.6.2
```

Circle Graph

The circle not a function because for $x = -2$, there are two values of y that satisfy the equation. In general, if you can draw a vertical line through two or more points on a graph, it's not a *function*. The problem arises from the squared term on y. Here y can take on the positive or negative value, making the outcome variable not a single value.

**c. Consider the line:** $y = 2x - 4$
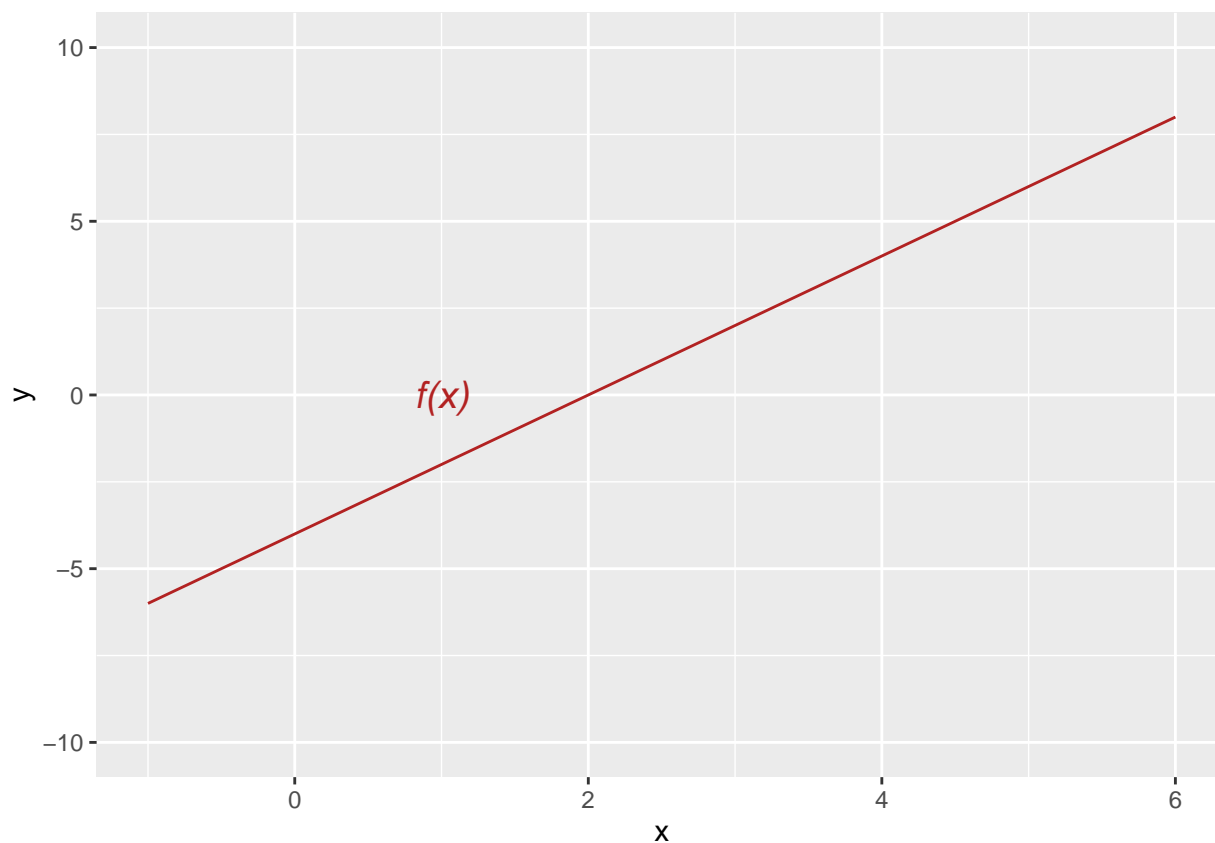
We can also write this funciton as $f(x) = 2x - 4$ since f is a function of x.

d. Let's add a second-degree polynomial function, $g(x)$, and we'll write it in two different ways:

$$g(x) = (x - 2)(x - 5) = x^2 - 7x + 10$$

What's the benefit of seeing the equation in these two ways? $(x - 2)(x - 5)$ and $x^2 - 7x + 10$

(Seriously, this is where you think about this insightful question. Why did we take the time and effort to write this equation in two different two ways? Each way has its own merit.)
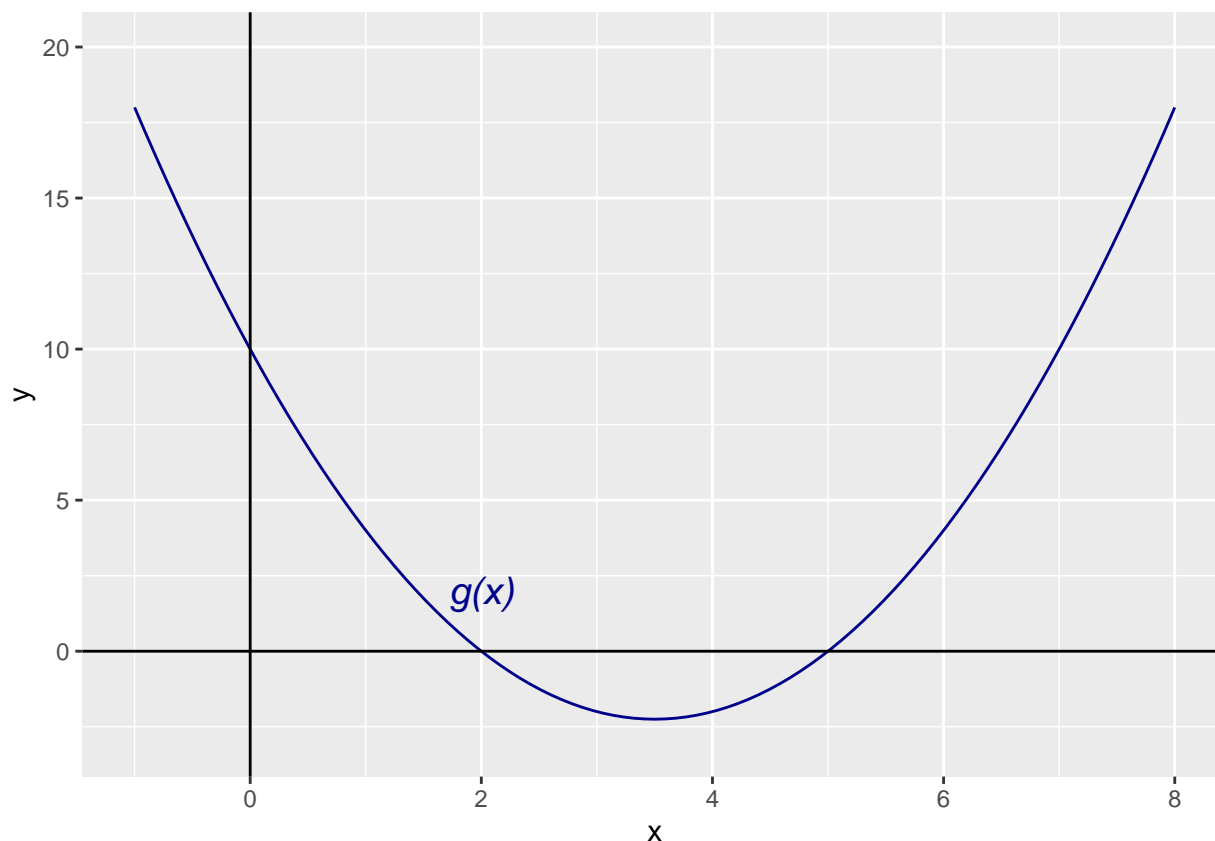
This is the plot of $g(x) = (x-2)(x-5) = x^2 - 7x + 10$.

The pedantic instructor might insist that the plot and the axes should have arrows on them, else it might imply that the function is actually an arc or a segment. Let's agree to forgive each other for the lack of arrows.

You should be able to show on this graph why your answer to the previous question makes sense. Can you?

Note that $g(x)$ is a second-order polynomial, because the x^2 term is of degree 2, the highest of the degrees of the three terms x^2, $-7x$, and 10.

e. Zero-degree polynomial.

OK, quick, what does a zero-degree polynomial function look like?

(This is the part where you think, again. What's the graph look like? Draw a coordinate plane and sketch it in your notebook, or draw it with your finger in the palm of your hand - before you go to the next page, please.)
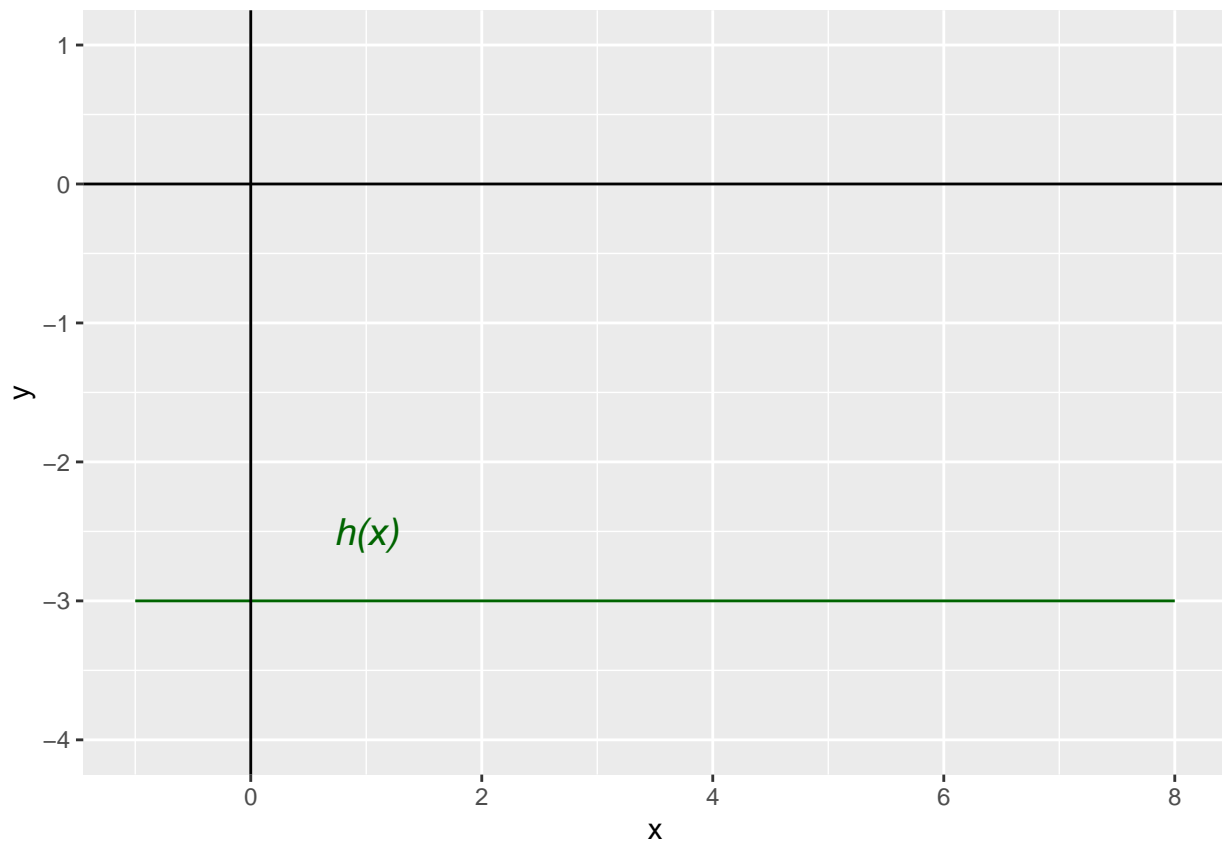
Above is a plot of $h(x) = -3$, a zero-order function. How does $h(x)$ vary with x?

f. Functions derived from other functions.

Let:

$$j(x) = f(x) + g(x)$$

and

$$k(x) = f(x) + g(x) + h(x)$$

$$
\begin{aligned}
j(x) &= f(x) + g(x) \\
&= (2x - 4) + (x^2 - 7x + 10) \\
&= x^2 - 5x + 6 \\
&= (x - 2)(x - 3)
\end{aligned}
$$

Which is a nice form for graphing this equation, as we get the zero-crossings, 2 and 3.
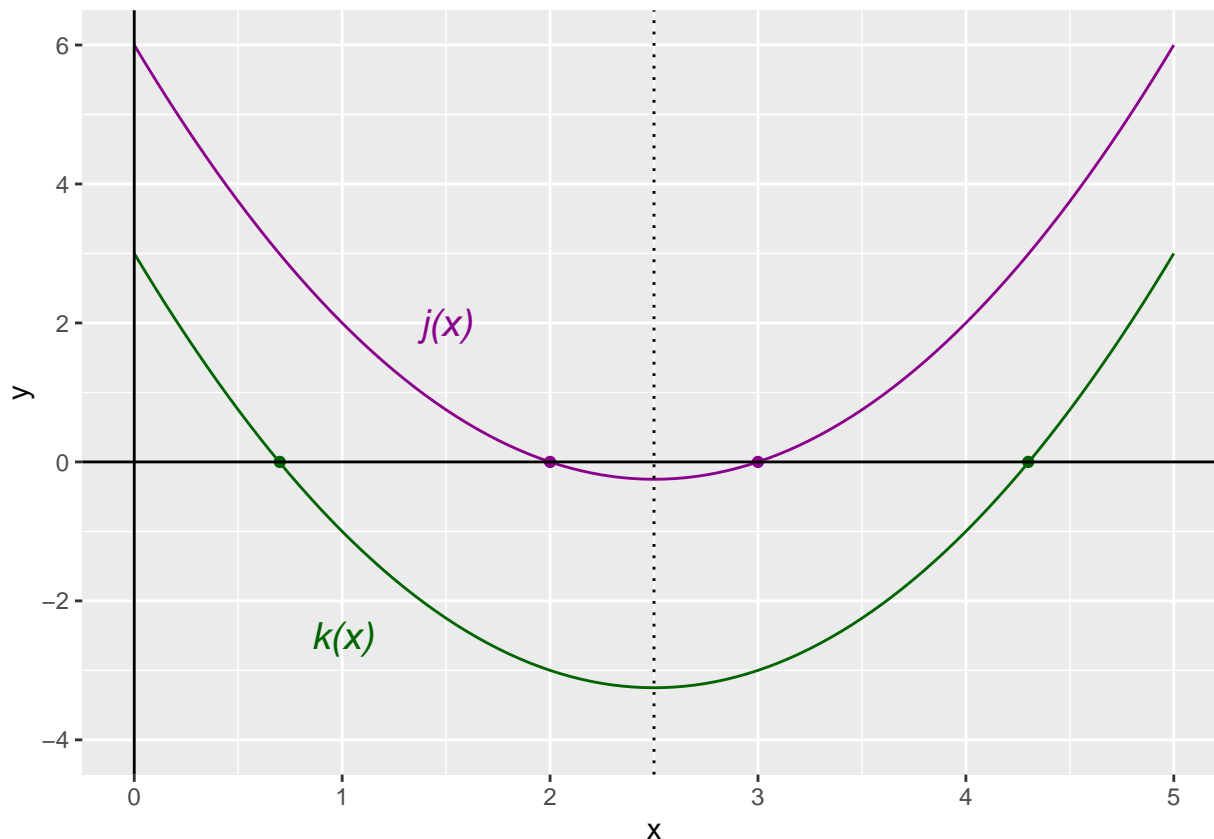
$$k(x) = f(x) + g(x) + h(x)$$
$$= (2x - 4) + (x^2 - 7x + 10) - 3$$
$$= x^2 - 5x + 6 - 3$$
$$= (x - 3)(x - 2) - 3$$

Why did I not write this as $k(x) = x^2 - 5x + 3$?

Let's look at $j(x)$ and $k(x)$ on the same plot...

Note that $k(x)$ is just $j(x) - 3$. In other words, $k(x)$ is shifted down by 3 on the y axis. The two functions have their minima at the same value of x, and they have the same slope for every value of x. Do you see that?

Also, note that for $k(x)$, at $y = -3$, x has values of 2 and 3. We can write this as $k(2) = k(3) = -3$.

Can you write the equation for the values at which $k(x)$ crosses the x-axis?

Here's a hint (the quadratic formula):
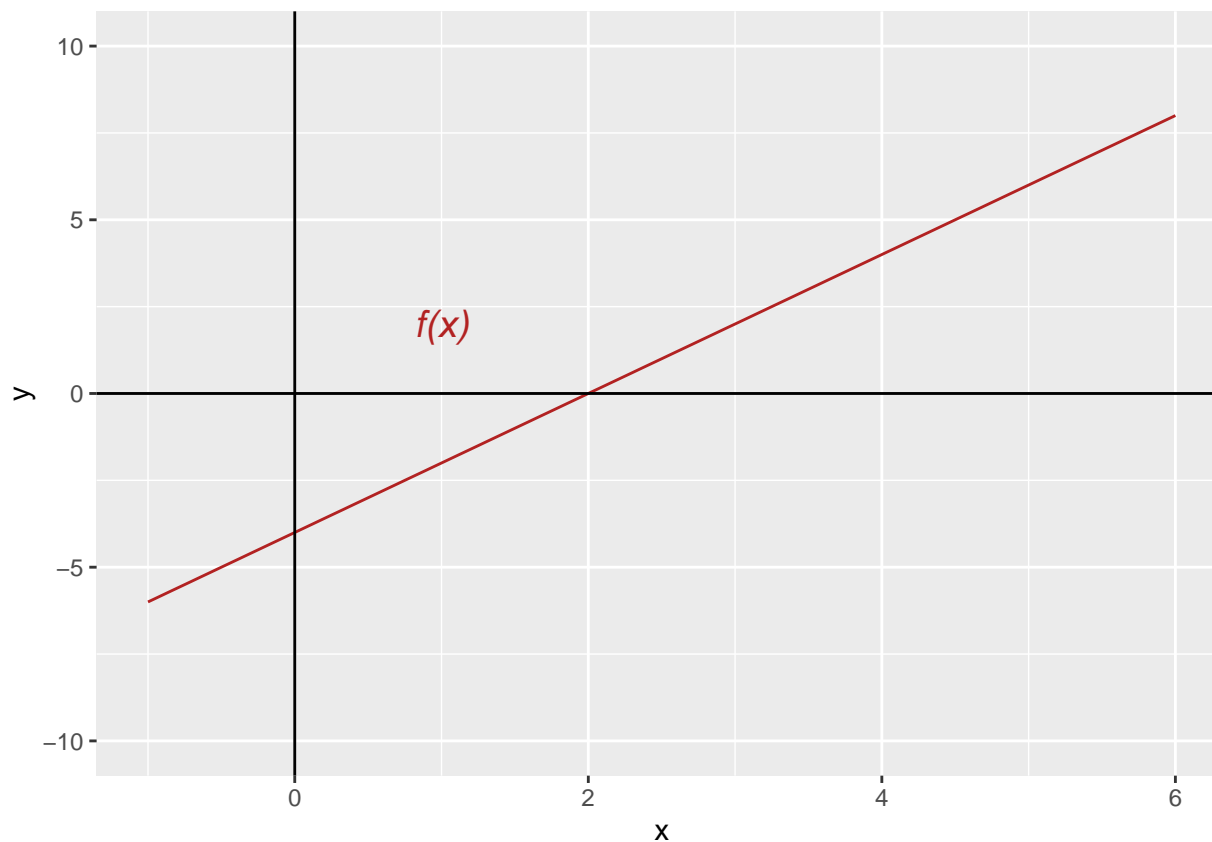
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

which is the solution for the values of x which make $y = 0$ in an equation of form: $y = ax^2 + bx + c$.

What are a, b, and c?

So, we get $\frac{5 \pm \sqrt{13}}{2} \approx 2.5 \pm 1.8 \approx 0.7, 4.3$ for the zero-crossings.

5. Differential Calculus

a. How does $f(x)$ change as x changes? $f(x) = 2x - 4$

For every increase of x by 1, $f(x)$ increases by 2. How does this fact reveal itself in the graph of $f(x)$?
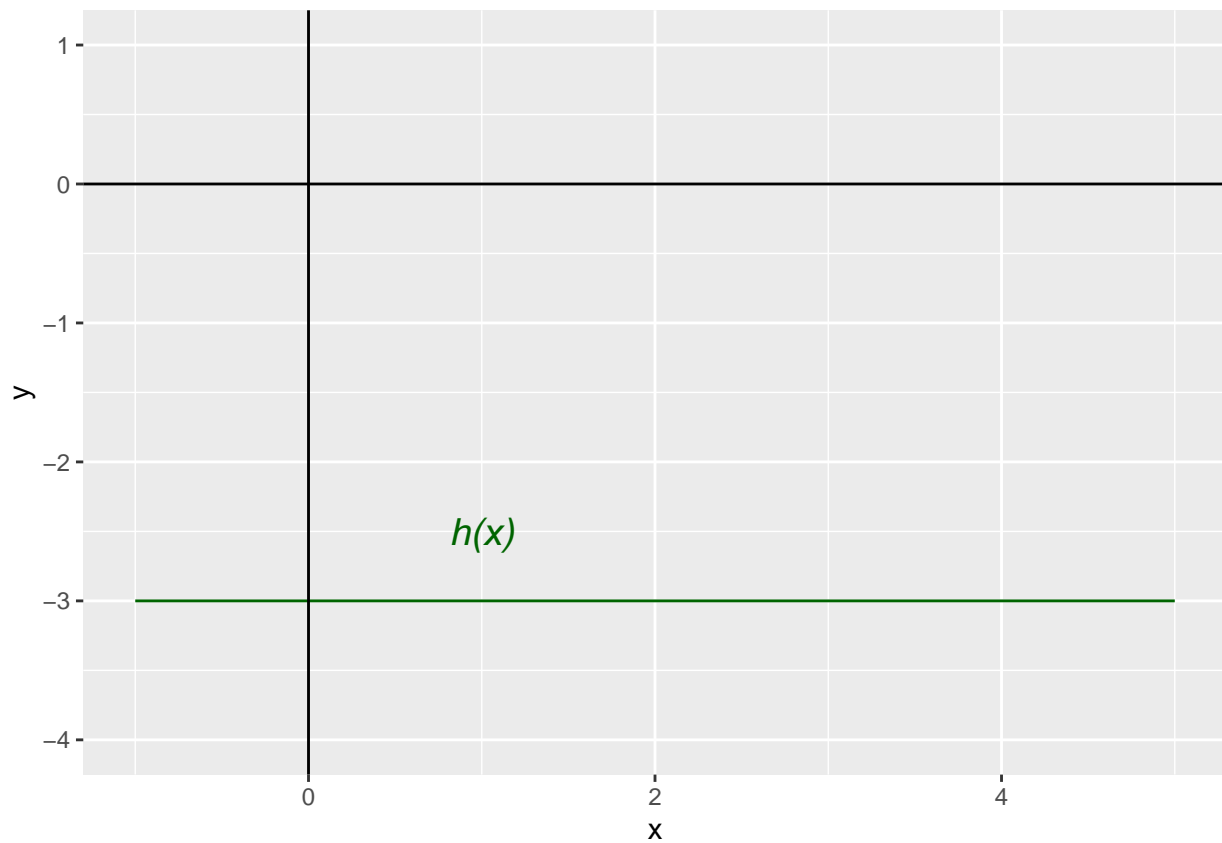
We can write this in several ways:

- The derivative of $f(x)$ with respect to x is 2.
- $\frac{d}{dx} f(x) = 2$ This is Liebniz' Notation
- $f'(x) = 2$ This is Lagrange's Notation
- $y' = 2$

These are all equivalent. Which do you like best?

b. The derivative of $h(x)$ wrt x.

Recall the graph of $h(x)$:

How does $h(x)$ change as x changes?

This equation captures that relationship: $\frac{d}{dx}h(x) = 0$

In other words, $h(x)$ does not vary with x; it has constant value of -3 regardless of the value of x. The slope of this line is zero, which is another way of saying the same thing.

Recall that:

$$k(x) = j(x) + h(x)$$

We can differentiate both sides of this equation, with respect to x to get:

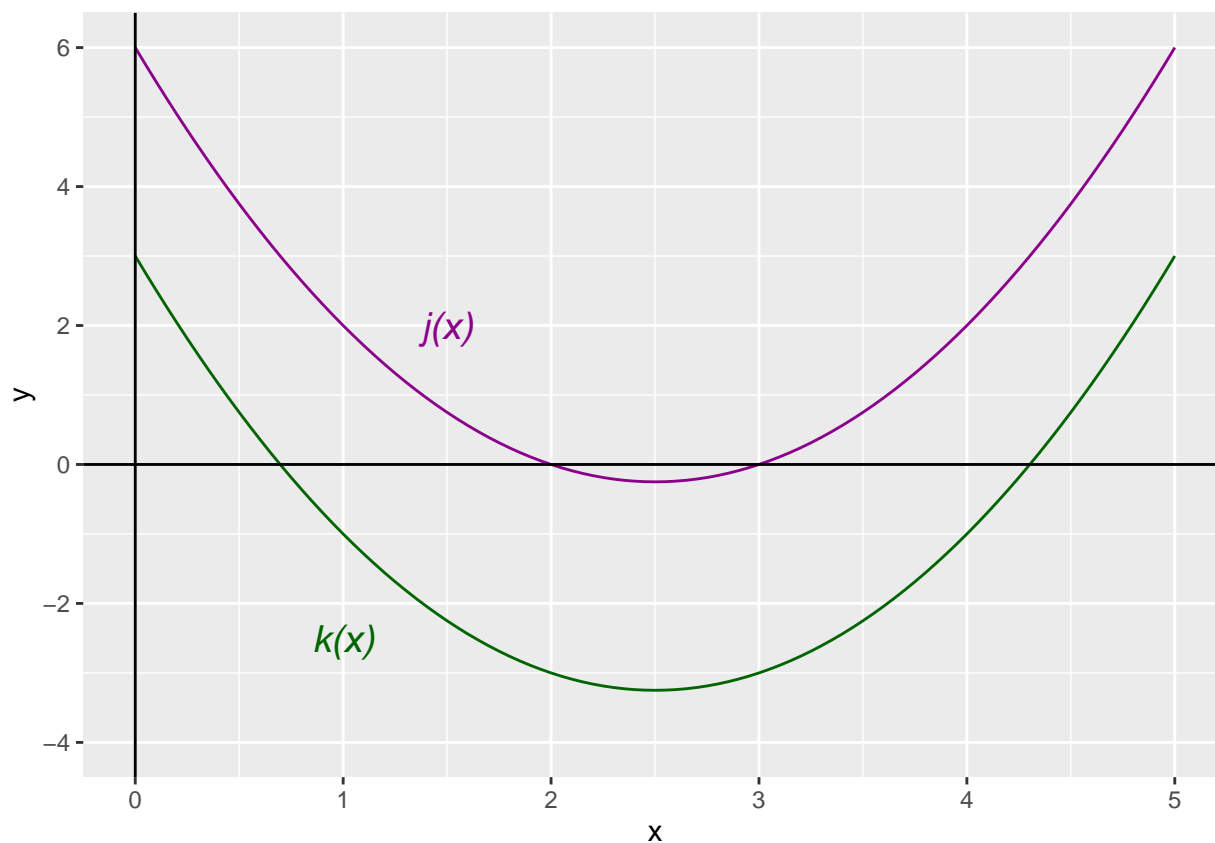$$\frac{d}{dx}k(x) = \frac{d}{dx}j(x) + \frac{d}{dx}h(x)$$

Since

$$\frac{d}{dx}h(x) = 0$$

we have:

$$\frac{d}{dx}k(x) = \frac{d}{dx}j(x)$$

Given your understanding that the derivative of a function is the slope of that function, the equation above should be clear from the graph below:

See how the functions always have the same slope for all values of x?

c. Taking the derivative of a polynomial function.

This is where we see the benefit of writing $g(x)$ in this form. . .

$$g(x) = x^2 - 7x + 10$$

$$\frac{d}{dx}g(x) = 2x - 7$$
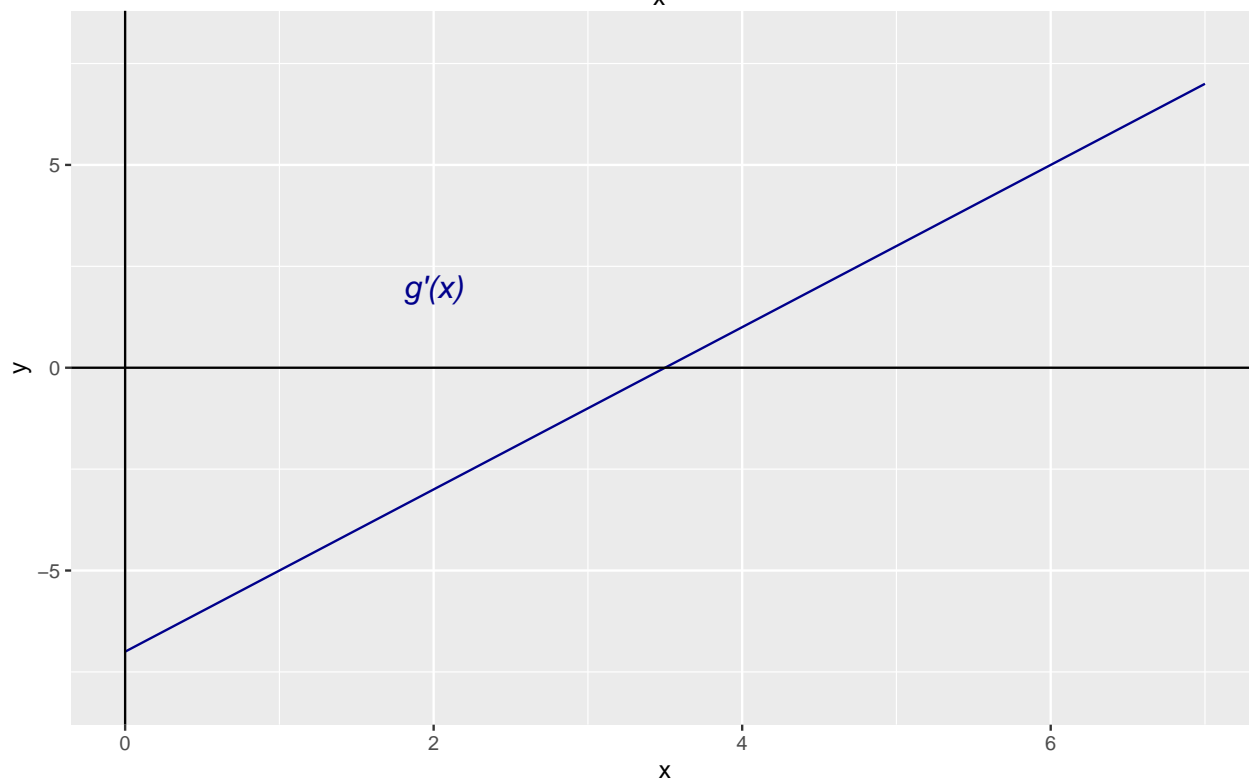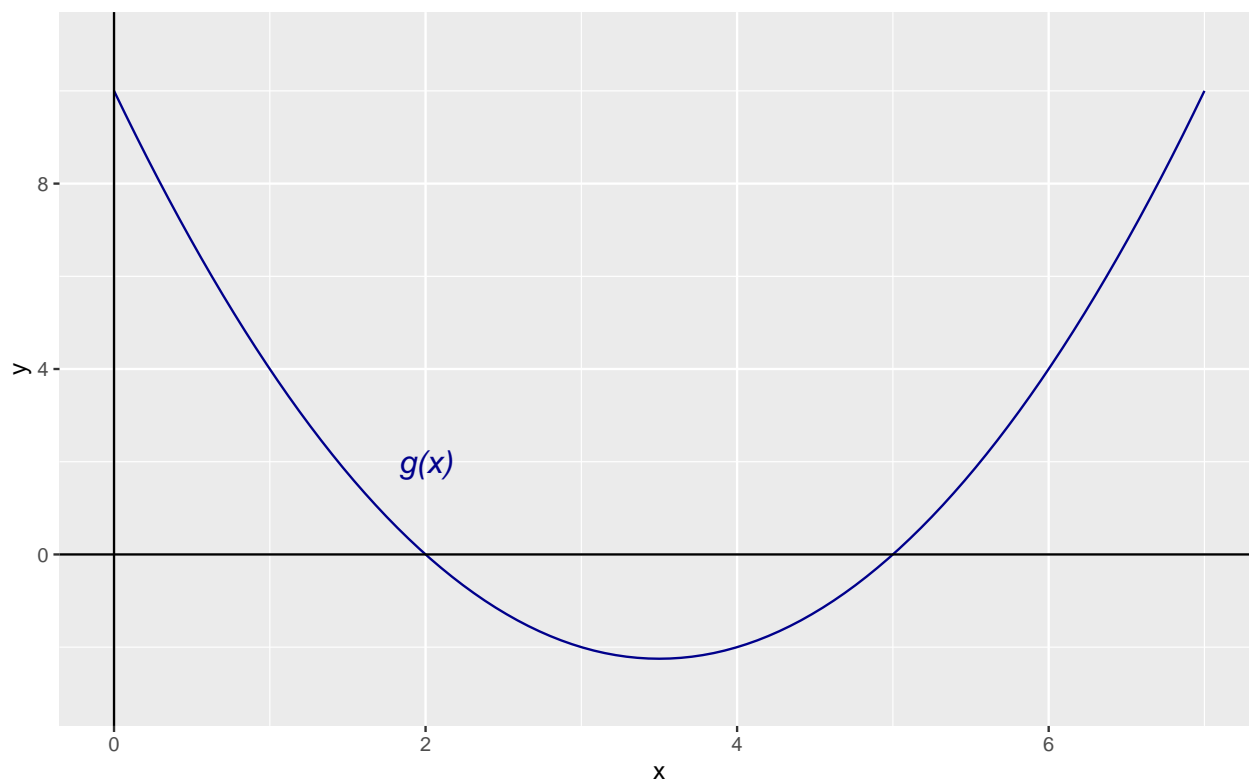
The formula for the derivative of a function,

$$l(x) = ax^b$$

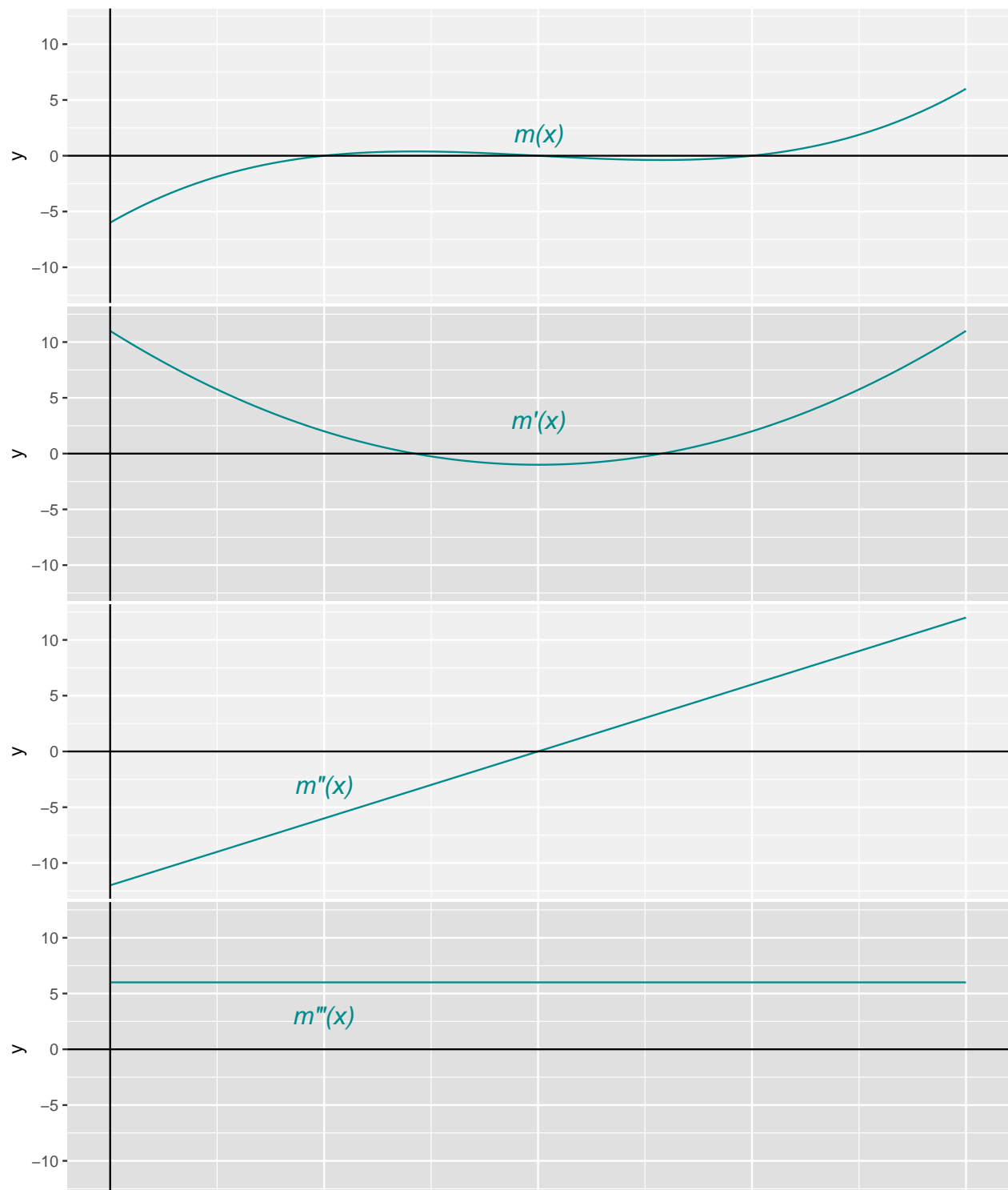with respect to x where a and b are constants is

$$l'(x) = abx^{b-1}$$

For polynomials, simply repeat this for each term of the function.

d. Let's look at $g(x)$ and $g'(x)$ together.

g(x)

g'(x)

Note the value of x where $g'(x) = 0$ and the behavior of $g(x)$ at that value of x. How would you describe this?

e. Plots for the derivatives of a 3rd-degree polynomial.

f. Notes on derivatives and notation

If

$$n(y) = 3y^2 + 4y + 7$$

and y is not a function of x, then

$$\frac{d}{dx}n(y) = 0$$

but

$$\frac{d}{dy}n(y) = 6y + 4$$

Which is just a way of encouraging you to make sure you pay attention to the independent variable that you are using when taking a derivative. This is a reason I do not like the $y' = 2$ notation: it can be ambiguous as to the independent variable (although usually not, in context).

Finally, let's review the notation for higher derivatives showing Leibniz and Lagrange notation:

$$\frac{dy}{dx} = f'(x)$$

$$\frac{d^2y}{dx^2} = f''(x)$$

$$\frac{d^3y}{dx^3} = f'''(x)$$

g. Partial Derivatives

Functions may have many input variables. For such a function, the derivative of that function with respect to a single input variable is referred to as the partial derivative of that function with respect to that single variable. It is interpreted as the change in the output variable based on a change in a given input variable, *holding all other variables constant.* For partial derivatives, it is standard to use the greek letter delta, δ, instead of d. For example, consider:

$$f(x, y) = x^2 + xy + 4$$

$$\frac{\delta}{\delta x}f(x, y) = 2x + y$$

$$\frac{\delta}{\delta y}f(x, y) = x$$

Take a moment to consider what these statements mean, in the context of *holding all other variables constant.*

h. Important Derivatives

(From Tallarida's "Pocket Book of Integrals and Mathematical Functions, 2nd Edition")

Note: $u = f(x)$, $v = f(x)$, a and n are constants, and e is the base of the natural logrithm.

$$\frac{d}{dx}(au) = a\frac{du}{dx}$$

$$\frac{d}{dx}(u + v) = \frac{du}{dx} + \frac{dv}{dx}$$

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

$$\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$$

$$\frac{d}{dx}(e^u) = e^u\frac{du}{dx}$$

$$\frac{d}{dx}ln(u) = \frac{1}{u}\frac{du}{dx}$$

i. An application - The Wage Equation

Let's look at one example you will see this semester, the wage equation, where wage is a function of education. Note that Woolridge uses *log* to denote the natural logarithm with a base of e, differing from log_{10}, which has a base of 10.

$$log(wage) = \beta_0 + \beta_1 educ + u$$

For reasons that you may already understand, and which you will certainly understand in a few weeks, *wage* and other monetary variables are often modeled as natural log functions.

To take the derivative of the Wage Equation *wrt education*, we apply the rule:

$$\frac{d}{dx}log(u) = \frac{1}{u}\frac{du}{dx}$$

Differentiating both sides of the Wage Equation *wrt education*, we get:

$$\frac{d}{d\,educ}log(wage) = \frac{d}{d\,educ}(\beta_0 + \beta_1\,educ + u)$$

$$\frac{1}{wage}\frac{d\,wage}{d\,educ} = \beta_1$$

Rearranging, we get:

$$\beta_1 = \frac{\frac{d\,wage}{wage}}{d\,educ}$$

which is interpted as β_1 being the *percent change* in *wage* that results from an additional unit (typically a year) of education. Note that this *percent change* outcome is a direct result of $\frac{d}{dx}e^x = e^x$, and it is for this reason that we use the *natural* log for variables such as *wage*. Does intuitively make sense to you that people people, regardless of their salary level would view a 10% raise similarly? Lots more on this later in the semester.

The perceptive student will question if β_0 and β_1 are truly constants *wrt educ*. We will cover these important assumptions during the class, and learn the meaning of *exogeneity* and *endogeneity*.

6. Integral Calculus

a. Integrals - Basic Idea and Intuition

The integral of a function represents how that function accumulates as a function of some input variable. Graphically, the integral of a function is the area under the curve, either for the entire range of the function (the indefinite integral), or for some specific range (the definite integral).

The integral of a function is often written as a capital letter-version of the function. For example:

$$F(x) = \int f(x)dx$$

The area under the curve of the function $f(x)$ from a to b is the integral of $f(x)$ wrt x evauated at b minus that same function evaluated at a. This is known as the Fundamental Theorem of Calculus.
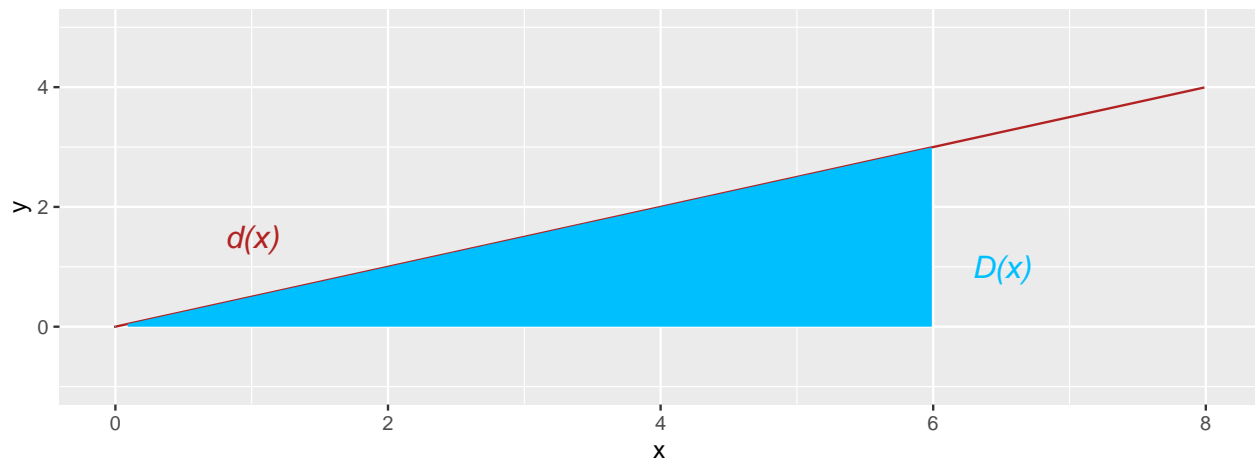
$$F(b) - F(a) = \int_a^b f(x)dx$$

b. Integrating a linear function.

Consider

$$d(x) = \frac{1}{2}x$$

which has the indefinite integral:

$$D(x) = \int \frac{1}{2}x\,dx$$

We can define a function $D(a)$, the area of the triangle shown in blue, as follows:

$$D(a) = \int_0^a \frac{1}{2}x\,dx$$

$$D(a) = \frac{x^2}{4}\bigg|_0^a = [\frac{a^2}{4} - \frac{0^2}{4}] = \frac{a^2}{4}$$

Computing, the area shown in blue: $D(6) = 9$. To check, we can compute the area of the triangle under $d(x)$ from $x = 0$ to $x = 6$.

The area of a triangle is $\frac{1}{2}base \times height$

The height is: $d(6) = \frac{1}{2} \times 6 = 3$

The base is 6, so the area is $\frac{1}{2} \times 6 \times 3 = 9$.

c. Cumulative Distribution Functions (CDFs)

Pay close allention to the indices on the integral and the variables of the function being integrated. See p. 148 in Devore. $F(x)$ is a function of x, which is the upper index on the integral. $F(x)$ is the cumulative area under the curve from $-\infty$ to x of the probability density function $f(y)$.

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\, dy$$

d. Double Integrals

$$D(a,b) = \int_{x=0}^{a} \int_{y=0}^{b} f(x,y)\, dy\, dx = \int_{y=0}^{b} \int_{x=0}^{a} f(x,y)\, dx\, dy$$

Intuitively, if you are going to compute the volume of a solid, it does not matter whether you compute (or measure) along the x-axis or y-axis first, you will get the same answer. That's what this equation says. On the left side, you first integrate over y, then over x, working your way out. On the right side, you integrate over x first, and over y.

Here's an example:

$$f(x,y) = 2xy$$

$$F(a,b) = \int_{y=0}^{a} \int_{x=0}^{b} 2xy\, dx\, dy$$

$$= \int_{y=0}^{a} 2y \int_{x=0}^{b} x\, dx\, dy$$

$$= \int_{y=0}^{a} 2y \frac{x^2}{2}\Big|_{0}^{b} dy$$

$$= \int_{y=0}^{a} 2y \left[\frac{b^2}{2} - \frac{0^2}{2}\right] dy$$

$$= \int_{y=0}^{a} y\, b^2 dy$$

$$= b^2 \int_{y=0}^{a} y\, dy$$

$$= b^2 \frac{y^2}{2}\Big|_{0}^{a}$$

$$= \frac{a^2 b^2}{2}$$

e. Double Summations

Note that for a discrete function, $p(x, y)$, the equivalent functional representation would be:

$$D(a, b) = \sum_{x=0}^{a} \sum_{y=0}^{b} p(x, y) = \sum_{y=0}^{b} \sum_{x=0}^{a} p(x, y)$$

7. Histograms and Density/Mass Functions

Consider the set of Data Points (7, 8, 5, 10, 5, 6, 8, 7, 8, 7, 4, 9, 6, 7, 6, 9, 6, 7, 12, 9).

a. Stem-and-leaf plot - a horizontal histogram

Let's order and bin the data:

```
4
5 5
6 6 6 6
7 7 7 7 7
8 8 8
9 9 9
10
12
```

b. Histogram vs. Density Function

Note the differences in the following two plots, and look at the R code to see how similar they are. Can you spot the difference? Also note how you can include code as part of of your reporting. This will be a basic skill to master.

```r
x <- c(7,8,5,10,5,6,8,7,8,7,4,9,6,7,6,9,6,7,12,9)
hist(x, col="lightblue", main="Histogram of x", breaks = seq(3.5, 12.5, by=1))
```
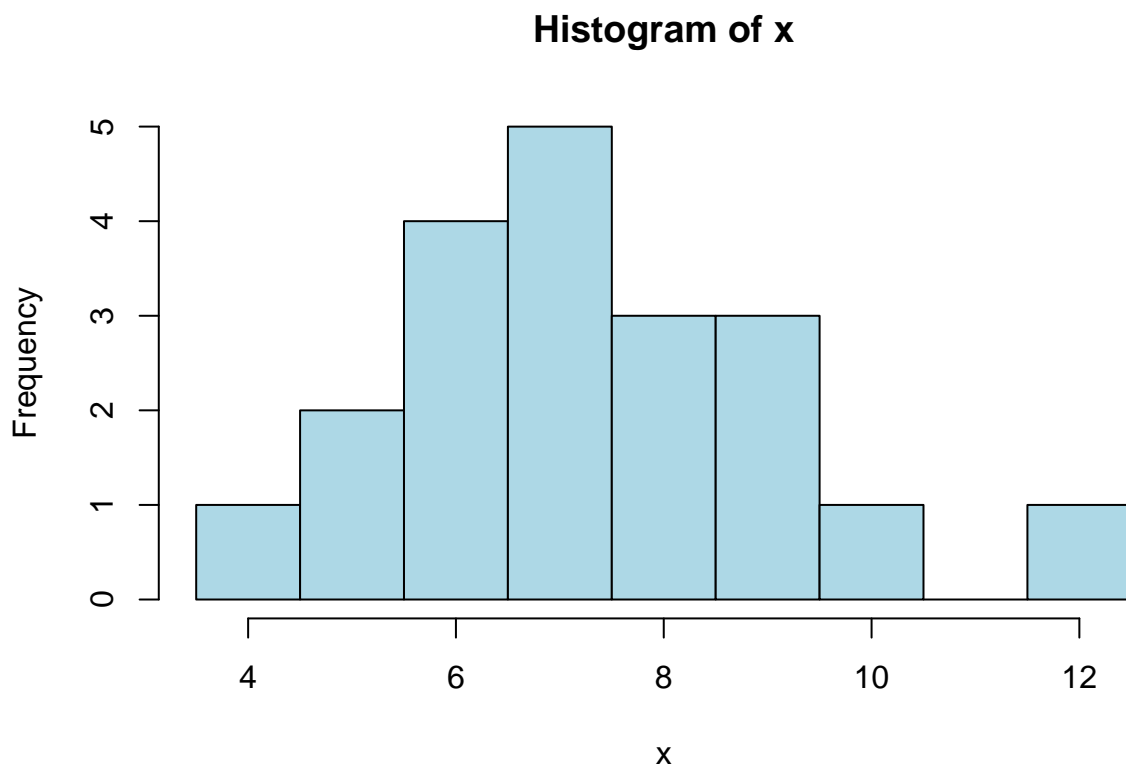

Why did I include this statement "breaks = seq(3.5, 12.5, by=1)" in the code? What is its function? Please use this pattern, where appropriate, when you generate histrograms.

```
#
# Note: this is the place to comment your code, NOT the place to write analysis that is
#  part of your report.
#
# As a data scientist, write your report for readers who may not be able to read
#  your code.  Putting prose in the code comments like this is a TERRIBLE idea.
#  Please don't do it. However, commenting your code so that others can understand
#  it, and so that you can understand it later, is a fabulous idea.
#
# The seq(3.5, 12.5, by=1) comment offsets the histogram by 0.5 so that the labels appear
# directly under the center of the bars - an example of a useful code comment
# and the answer to the question on the previous page.
#
hist(x, prob=TRUE,  col="lightblue", main="PMF of x", breaks = seq(3.5, 12.5, by=1))
```

PMF of x
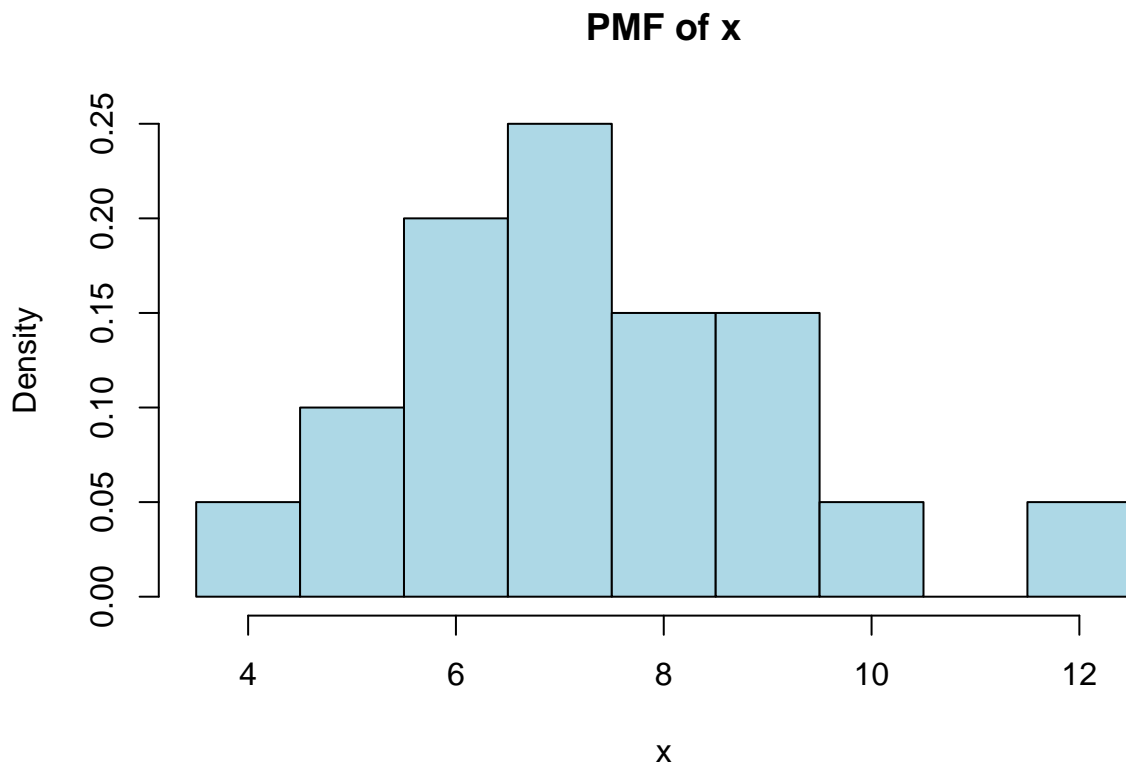
Probability Density Functions are for continuous input variables; *Probability Mass Functions* are for discrete input variables. Devore makes this distinction (pp 100, 143). However, Woolridge does not make this clear distinction, nor does R (as you can see in the graph above). If you want to get this perfect, override the default y-axis name when generating histograms using hist in R.