Quantitative Human Physiology

Jeffrey A. Walker

2019-10-04

Contents

Directions

1. Create a Google Sheets file that you will share with me. Name the sheet "last name BIO 22x", for example "Walker BIO 221" or "Smith BIO 223"
2. Create one sheet for each assignment. These can be created as you do each
3. Name each sheet using the assignment name. For example "1. Diffusion"

Do not create separate Google Sheets files for each assignment. Do not spread a single assignment over multiple sheets.

BIO 221

Chapter 1

Diffusion

Goals: approximation, standard curve, scale, spreadsheet best practices, google search

The main goal of this exercise is to use *approximation* to demonstrate that diffusion is too "slow" to be an effective transport process at any scale larger than microscopic (microns). **Scale** roughly means size but typically size means something fairly precise (132.5 cm) while scale is more general. "At the scale of individual molecules" means something with a size of 1 nm or 10 nm or even 100 nm but certainly no 10^5 nm. When using scale, our precision is generally only at the **order of magnitude**, which are multiples of 10 (The range 1 nm to 100 nm is over two orders of magntidue).

1.1 Approximation

At some point, scientists are confronted with the recognition that some math problems do not have exact answers. This is very different than the math we learn in high school and college, which typically represents math as a construction that only has exact answers (2 + 2 = 4 and not any other number).

What is the expected time of diffusion of biologically important molecules at different scales? There is no exact answer to a question like "how long does it take for oxygen to diffuse across the respiratory membrane". Instead there are better answers and worse answers. The reason there isn't an exact answer is because

1. the question is about a system of particles that are in diffusion. The motion of individual particles is random and the time an individual particle travels a certain net distance is a function of its individual history of elastic collisions with other particles. But we can compute *expected* (or the average) time for diffusion over a specified distance using a simple diffusion model (see below)
2. The rate of diffusion is a function of the medium such as pure water, air, cytosol, or lipid bilayer. The diffusion coefficient in the table below is measured using water as the medium but in the questions below the media might include cytosol, extracellular fluid, and phospholipid membrane, so the water-based coefficient will only give us a close (or approximate) answer.
3. The final reason is the one that makes students the most uncomfortable. The goal of this exercise is to compare time of transport by diffusion over distances that differ by many orders of magnitude–from a few nm to a few meters (so one billion times or 9 orders of magnitude). While it is true that skeletal muscle cells vary in diameter and the epidermis varies in thickness, the question is looking for a answer only at a reasonable order of magnitude. Two meters is a reasonable answer to the question, what is the height of humans? 0.2 or 20 meters is not.

1.2 Problems

A simple formula to calculate the approximate time of diffusion (t) over a distance (x) is

$$t \approx \frac{x^2}{2D} \tag{1.1}$$

where D is a diffusion coefficient. The symbol ≈ means "approximately equal to".

1.2.1 Compute diffusion times for chemicals with known diffusion coefficients

Create a table in your sheet that looks like this:

Event	Distance (original units)	Original units	Distance (cm)	Diffusion coefficient	time in seconds	time in minutes	time in hours	time in days
Na+ through a Na+ channel								
O2 fom edge of type I muscle fiber to a mitochondria in the center of the cell								
O2 across the respiratory membrane (alveoli to pulmonary capilary)								
O2 across skin epidermis								
O2 from lung to big toe								
O2 from front to back of classroom								
Actelycholine across a synaptic cleft								

The first column lists diffusion "events", for example, the diffusion of a sodium ion (Na^+) through a Na^+ channel. Use Google search to find the distance traveled by the chemical in each event. Use this distance, the molecular weight of the molecule (see the table below), and the function above to compute the approximate time for transport of this molecule across this distance via diffusion. Compute this time in seconds, minutes, hours, and days. See Problem 2 below for actylcholine.

Spreadsheet best practices that must be followed to receive full points;

1. **Use functions for all math**; that is, don't do the math in your head or on a calculator and input the result. Do the math in the spreadsheet. The reason for doing the math in this worksheet is that this is good practice. If we do math on a calculator and then insert the result here, we have no record of the computation. Having a record of all computations is a best practice for reproducible science.
2. **Keep cells with numeric values (such as the diffusion distance) numeric**. For example, do not add units to the diffusion distance in the cell containing the numeric value. If the diffusion distance is 1 meter, enter "1" into the cell and not "1 meter". The reason is, you can't do math on words (that is, the value in the cell won't work in simple equations). But, it is good to record the units so add these to the adjacent column (Original units).

1.2.2 Compute diffusion time for a chemical with an unknown diffusion coefficient

Diffusion data for acetylcholine is not given in the table above. You might be able to find the coefficient with a google search but I want you to use the table of above to 1) generate a standard curve and 2) use the generated curve to estimate the diffusion coefficient for acetylcholine. Remember that diffusion is a function

Molecule	Molecular Weight (g/mol)	Diffusion Coefficient (cm^2/s)
H^+	1.008	9.31×10^{-5}
Na^+	22.99	1.33×10^{-5}
K^+	39.098	1.96×10^{-5}
Ca^{2+}	40.078	0.79×10^{-5}
Cl^-	35.453	2.03×10^{-5}
Ammonia (NH_3)	17.031	1.51×10^{-5}
Oxygen (O_2)	31.999	2.10×10^{-5}
Carbon dioxide (CO_2)	44.01	1.97×10^{-5}
Urea	60.055	1.38×10^{-5}
Glucose	180.156	5×10^{-6}
Sucrose	342.296	5.23×10^{-6}
Hemoglobin	68,000	6.9×10^{-7}
DNA	6,000,000	1.3×10^{-8}

Figure 1.1: Molecular weights of common biological molecules

of the *size of the molecule* – bigger molecules diffuse more slowly. In other words, the diffusion coefficient gets smaller as the size of the molecule gets bigger. Use the table above to create a *standard curve*, which shows the relationship between known X and known Y. Standard curves are everywhere in science research and recognizing that you can solve a problem, without asking for help, by generating a standard curve will make your future boss very happy. So, instead of trying to look up the diffusion coefficient of acetylcholine, look up its molecular weight with a quick google search (or test your chemical skills and compute its weight using acetylcholine's chemical formula) and compute its expected (or predicted value) given the standard curve.

In a traditional standard curve done on graph paper, one can predict the expected value of Y (which has not been directly measured) from the known value of X. Here, I want you to do this by generating the mapping function (a function that maps a value of X to a value of Y)

$$D = b_0 + b_1 MW \tag{1.2}$$

which is the equation for a line (recognize how it is simply Y = mX + b with b and mX re-arranged?) so b_0 is the intercept and b_1 is the slope. Use the data in the table to estimate the slope and intercept of the mapping function and then use this slope and intercept to compute the estimated diffusion coefficient for acetylcholine. When computing the slope and intercept, temember that we want to predict D from MW.

1. log transform the weights and the diffusion coefficients– this will make the relationship between the two linear.
2. plot log(D) on the Y-axis aganst log(MW) on the X-axis. Show the "trend line", also called the regression line and formula, which gives the slope and intercept
3. In seperate cells, use spreadsheet functions (google these) to compute the slope and intercept
4. use the equation for the regression line to compute the predicted log(diffusion coefficient) given a the log(molecular weight of acetylcholine). Use the cells with the slope and intercept in this function. ** Do not hardcode the slope and intercept in your formula **.
5. back transform the answer to get the right units (not log units) - the back-transformation is the inverse or anti-log. Make sure you know the base for the log function that you are using!

Chapter 2

Nernst Potential

Goals: quantitative thinking, spreadsheet best practices

2.1 Electrical potentials

Diffusion of an ion across a membrane differs from that of an electrically neutral chemical such as glucose because the dynamics of ion diffusion is driven by both a **chemical gradient** and an **electrical gradient**. The chemical gradient is due to a difference in the concentration of the chemical on either side of the membrane. The electrical gradient is a due to a difference in charge across the membrane. This difference in charge is largely due to differences in the concentration of electrolytes (especially Na+, K+, Cl-, Ca++) on either side of the membrane. The difference in charge gives rise to an **electrical potential** across the membrane, known as the **membrane potential**. The units of the membrane potential is milivolts (mV). If the inside of a membrane has more negative and fewer positive charges relative to the outside of the membrane we say the membrane potential is negative (so the sign of the potential is arbitrarily assigned as the inside relative to outside).

A cell can regulate its membrane potential by making the membrane more or less permeable to ions, which changes the distribution of charges on either side of a membrane. The membrane potential in a resting neuron (when it is not conducting a current) is called the **resting membrane potential** or simply "resting potential". The resting potential is an equilibrium potential in the sense that there is no net transport of charge. A neuron has a **resting membrane potential** of around -70 mV, meaning that the inside of the membrane has more negative and fewer positive charges relative to the outside.

When the membrane potential is negative, the electrical gradient for a cation is directed "into" (this is the "downhill" direction). This makes sense, the positively charged cations are attracted to the inside (notice that there is no sense of attraction and repulsion in diffusion of electrically neutral chemicals). By contrast, the electrical gradient for an ion is directed "out of" the cell – anions are attracted to the more positively charged outside of the membrane. The magnitude of the electrical gradient is proportional to the membrane potential.

Why "potential"? A potential is a difference in potential energy over some space. A chemical gradient is a **chemical potential**. An electrical gradient is an electrical potential, and an electrochemica gradient is an **electrochemical potential**. A potential can be used to power (provide the energy for) something such as rotating a turbine or lighting a light bulb. The chemical potential of Na+ is used to transport molecules such as glucose up their chemical gradient. The membrane potential is used for numerous cell functions such as generating a current that runs down the membrane of neuron or muscle cell.

Table 2.1: Concentrations of major ions in a neuron

Ion	Intracellular concentration	Extracellular concentration
Sodium (Na+)	15 mM	145 mM
Potassium (K+)	150 mM	4 mM
Calcium (Ca2+)	70 nM	2 mM
Hydrogen ion (proton, H+)	63 nM (pH 7.2)	40 nM (pH 7.4)
Magnesium (Mg2+)	0.5 mM	1 mM
Chloride (Cl−)	10 mM	110 mM
Bicarbonate (HCO3−)	15 mM	24 mM

2.2 The Nernst potential

For an ion, the chemical and electrical gradients add together to create an electrochemical gradient, or **electrochemical potential**. The rate of diffusion of an ion is proportional to its electrochemical gradient and this gradient is the difference between the membrane potential and the ion's **equilibrium potential**. The equilibrium potential of an ion is the membrane potential at which there is no net transport (diffusion) of the ion across the membrane (despite the membrane being permeable to the ion). There is no net transport - that is, the ion is at equilibrium, because the ion's electrical and chemical gradients are equal in magnitude but opposite in direction. The equilibrium potential for an ion is the ion's **Nernst potential**.

The Equilibrium (Nernst) potential of an ion is

$$E_{ion} = \frac{62\text{mV}}{Z} \log_{10} \frac{[\text{ion}]_{\text{out}}}{[\text{ion}]_{\text{in}}} \tag{2.1}$$

where Z is the valence of the ion.

2.3 Electrochemical potential

If the membrane potential is not at an ion's equilibrium potential, there will be diffusion of the ion down it's electrochemical gradient if the membrane is permeable to the ion. This electrochemical potential of an ion is

$$\Delta E_{ion} = E_{membrane} - E_{ion} \tag{2.2}$$

The greek letter Δ ("delta") is often used in math and science for "difference". For a cell membrane, the direction of a negative ΔE_{ion} is into the cell and the direction of a positive ΔE_{ion} is out of the cell. Sometimes ΔE_{ion} is called the "driving force" but a phrase like the electrochemical potential "drives" Na+ into the cell can lead to the misconception that all Na+ ions are moving into the cell. This is diffusion and diffusion is the net transport.

2.4 Problem Set

1. Use the table of ion concentrations above to compute the Nernst potential for the ions Na+, K+, Ca++, and Cl-. Set up the table in google sheets to look like this:

Ion	[ion]_in	[ion]_out	Z	E (ion)
Na+				
K+				
Ca++				
Cl-				

2. Use your computations of equilibrium coefficients for Na+ and K+ to compute the electrochemical difference ("driving force") of each ion at different membrane potentials. Describe the consequence on the initial rate and direction of diffusion of the ion using something like "fast and into the cell". Set up the table in google sheets to look like this

K+			
E (membrane)	E (ion)	Delta E	Qualitative Description of direction and magnitude
-70			
30			
-150			
150			
Na+			
E (membrane)	E (ion)	Delta E	Qualitative Description of direction and magnitude
-70			
30			
-150			
150			

Chapter 3

Muscle

Chapter 4

Skeleton

BIO 223

Chapter 5

Blood

Goals: back-of-the-envelope estimation, scale, google search

Many complex problems in biology can be broken down into a series of smaller problems and a common smaller problem is the **estimation** of some number, such as the number of bacteria per cell. Estimation problems range from **back-of-the-envelope estimations** that are imprecise but useful in that they give one a general sense of the **magnitude** of a phenomenon to more precisely modeled estimates that are used for making **decisions under uncertainty**. Back-of-the-envelope estimations are called that because most can, literally, be done with a pencil and the back of an envelope. They can be done with pencil because the computations uses rounded instead of exact numbers like 10 or 300 that are easily multiplied/divided. In this module, you will compute some back-of-the-envelope estimations.

A problem like "how many bacteria can colonize a cell" depends on the distribution of the sizes of the bacteria, the size of the cell, and how packed the cell is with its own molecules and organelles. Here, I simply want to get you started on addressing a problem like this with very simple models of the problem. Along the way, solving the problem should give you a sense of scale of what it is like to be a bacterium or a virus living in a cellular world.

5.1 How to solve an estimation problem

I'll solve an analogous problem: How many beach balls can fit in a barn? If a barn has Volume V_{barn} and a beach ball has volume V_{ball} then the number of balls that could fit into the barn would be approximately, $N_{balls} = \frac{V_{barn}}{V_{ball}}$. To solve this, I need to **paramaterize the model** by assigning numbers to these variables. And, the answer is dependent on what numbers I choose for the size of the beach balls and the shape and size of the barn and how filled the barn is with hay (or furniture or horses or whatever) – that is how much of the volume of the barn is available for beach balls. A back-of-the-envelope calculation simply uses a *reasonable* value for the parameters. So, here are my numbers.

1. the barn is a typical vermont barn. I have a sense of what "typical" is because I live in New England and see barns every day. But what if I were a martian, and had never seen a barn? Then I would need to find this information from a reliable source. So, to find "typical", I used a google search and found what looks like a reputable source that says a typical hay barn is 30 feet wide by 40 feet long. I estimated wall height from the figure as half the width and I used a 12/12 pitch for the roof, so the peak is centered and 15 feet (half of the width) high.

Again – if you don't have to look up information to parameterize your model, don't!

2. I used a big beach ball of 2 feet in diameter (because big beach balls are fun). I didn't need to look this up!

	A	B	C	D	E	F	G	H	I	J	K
1		W (ft)	L (ft)	H (ft)	volume (ft^3)	ball diameter (ft)	volume of cube enclosing ball (ft^3)	max number of balls	volume fraction already filled	number of balls	
2	main room	30	40	20	24000	2	8	3000	0.1	2700	
3	attic	30	40	15	9000	2	8	1125	0.1	1012.5	
4									Total	3712.5	
5											
6	Source for barn size: "The most often utilized size is 30 feet wide and 40 feet long." http://www.thebarnpeople.com/the-barn-people-barn-inventory.html										

Figure 5.1: Estimation of maximum number of beach balls that could colonize a barn. Cells in red are computed.

3. 10% of the barn is filled with hay.

I use Google Sheets to compute the number of balls for each room (the main room and the attic) and then add these. Here is my sheet

My column labels include the units of the measure. **Do not add units to the measure itself** because this makes the format of the cell "text" instead of "number" and you cannot refer to the cell in an equation. I also cite the source of the parameterization below the table (I cite the source for the size of the barn. The size of the beach ball I just made up).

5.2 Problem set

Do these on the same sheet. Name the sheet "1. Blood".

1. How many red blood cells in a drop of blood? Note, I don't want you to look up how many RBCs are in a drop, I want you to estimate it using a back-of-the-envelope estimation. You don't need to look up the volume of a drop of water if you are able to use available information in your head to derive a reasonable volume for a drop of water.

For the next three questions, assume the host cell is "empty", that is, it contains no organelles or molecules that take up space.

2. How many bacteria could colonize a red blood cell?
3. How many bacteria could colonize a macrophage?
4. How many virus particles could colonize a red blood cell?

– These should all be on the same google sheet.

– **Do not hardcode parameters**, that is, if a virus is 30 feet wide do not put "30" in an equation but instead make your equations reference the cell with this information.

– You may need to google search information to parameterize the model, such as, how big a virus is. Part of the goal of this is for you to develop your skills finding reliable information using a google search. There is variation in virus size or cell size so use something in the middle or "typical". Again – these are back-of-the-envelope estimates so you don't need to be very precise – in fact all of these problems could be computed by most working biologists without looking up any information. We all have a pretty good sense for how big a virus, a bacterium, a blood cell, and a drop of water is. But you can look up this information because you are at the beginning of your biology career.

– Cite a webpage giving the source of the information, as I've done for the barns. There is no "right" or "wrong" place to get this information, only more or less reliable. I'm not grading you on where you get it, but I want to see where you get it! And all I want for a citation is the webpage, this is not a formal citation that you put in a scientific paper.

Chapter 6

Immune

Goals: combinations

How many kinds of antibody can a human make by V(D)J recombination?

An individual human produces many different antibody proteins, where "different" is amino acid sequence. How is this possible given that there are only a few "antibody" genes? Part of the answer is V(D)J recombination. An antibody is constructed from two pairs of polypeptides. Each pair consists of a light chain and a heavy chain. Each chain has a "variable" region and a "constant" region. The heavy chain is constructed from one gene (located on chromosome 14) while the light chain is constructed from two genes: the light chain locus λ ("lambda") located on chromosome 22 and the light chain locus κ ("kappa") located on chromosome 2. The variable region of both light chain loci is composed of a V part and a J part. The variable region of the heavy chain locus is composed of V, D, and J parts. A V, D, or J part consists of multiple copies of the exon that will be spliced into the mRNA but each of these copies has a slightly different nucleotide sequence and some of the copies do not produce functional mRNA.

To make the heavy chain mRNA for the antibody

1. Choose one of the copies of the V region of the heavy chain locus.
2. Choose one of the copies of the D region of the heavy chain locus.
3. Choose one of the copies of the J region of the heavy chain locus.

combine with the C (constant) region to make the heavy chain mRNA

To make the light chain mRNA for the antibody

1. Choose one of the copies of the V region of one (either λ or κ) light chain locus.
2. Choose one of the copies of the J region of the same light chain locus.

combine with the C region to make the light chain mRNA.

Finally, combine the light and heavy chains (these are actually translated independently and then joined into the protein but the math is the same).

So an antibody is a combination of combinations. It is a combination (light + heavy combined) of combinations (V, J, and D combined)

6.1 Combinations

If there are n1 elements in set 1 and n2 elements in set 2, how many combinations of 1 element of each set are there? Answer: n1 x n2

	A	B	C	D
1				
2		lambda	kappa	heavy
3	V	3	4	5
4	D			2
5	J	2	2	3
6				
7	combinations	6	8	30
8				
9	light chain combinations		14	
10	heavy chain combinations		30	
11	possible antibody mRNA types		420	

Figure 6.1: How many kinds of antibodies

In the table below, I use this math to compute the number of antibodies that could be made using only V(D)J recombination. The

6.2 Problem set

Do these on the same sheet. Name the sheet "2. Immune"

1. (Ken, Jeff, David, and Doug) is the set of male faculty members in the Biology department. (Chris, Terry, Rachel1, Rachel2, and Rachel3) is the set of female faculty members in the Biology department. If the biology department has a square dance, how many combinations of male-female partners could there be? Write all of these out to confirm (write this in a column of your google sheet)
2. Figure 4.3 in this online textbook is a table containing the number of copies of each of the gene segments. Use this table to compute the number of different antibodies that can be synthesized using V(D)J recombination alone.

Chapter 7

Cardiovascular

7.1 Background

This exercise explores equations 12-1 and 12-2 from Vander's Physiology.

Regulation of blood flow is critical to increase or decrease delivery of blood to organs as they need more or less blood. Blood flow can be modeled with the equation for fluid flow used in almost any system (rivers, wind, etc)

$$F = \frac{\Delta P}{R} \quad (12.1) \tag{7.1}$$

1. F is flow
2. P is pressure. Here, and almost everywhere you'll see it, Δ (the greek letter "delta") means "change", so ΔP ("delta p") is a *difference* in pressure between two points in space. Here this is two points along the length of a blood vessel. It is the difference in pressure that is driving the blood to flow.
3. R is the resistance to flow due to friction. Friction sucks kinetic energy from moving objects (the lost kinetic energy is transformed to heat).

Resistance is an important concept in understanding human physiology. Resistance can be modeled using the Poiseuille equation

$$R = \frac{8L\eta}{\pi r^4} \quad (12.2) \tag{7.2}$$

1. L is the length of section of blood vessel
2. η (the greek letter "eta") is the viscosity of the fluid (or more specifically, the dynamic viscosity)
3. r is the radius of the lumen of the vessel.

7.2 Problems

Create a sheet named "3. cardiovascular"

1. Create a table like that below. Write down the units of each of the terms. There is no "right" answer, because units can be written different ways. For example I could write the units of volume as L ("liter") or gallon or L^3 (here "L" is length). Write down a definition of each term. Write a formal definition and then add your own interpretation of that definition. For example, Wikipedia defines density as the "mass per unit volume" which I'll interpret as "the amount of matter in given amount of space",

which doesn't quite capture the nuances but is helpful for understanding. Also notice that wikipedia's definition of density here is an equation expressed as words, this can help with thinking about units.

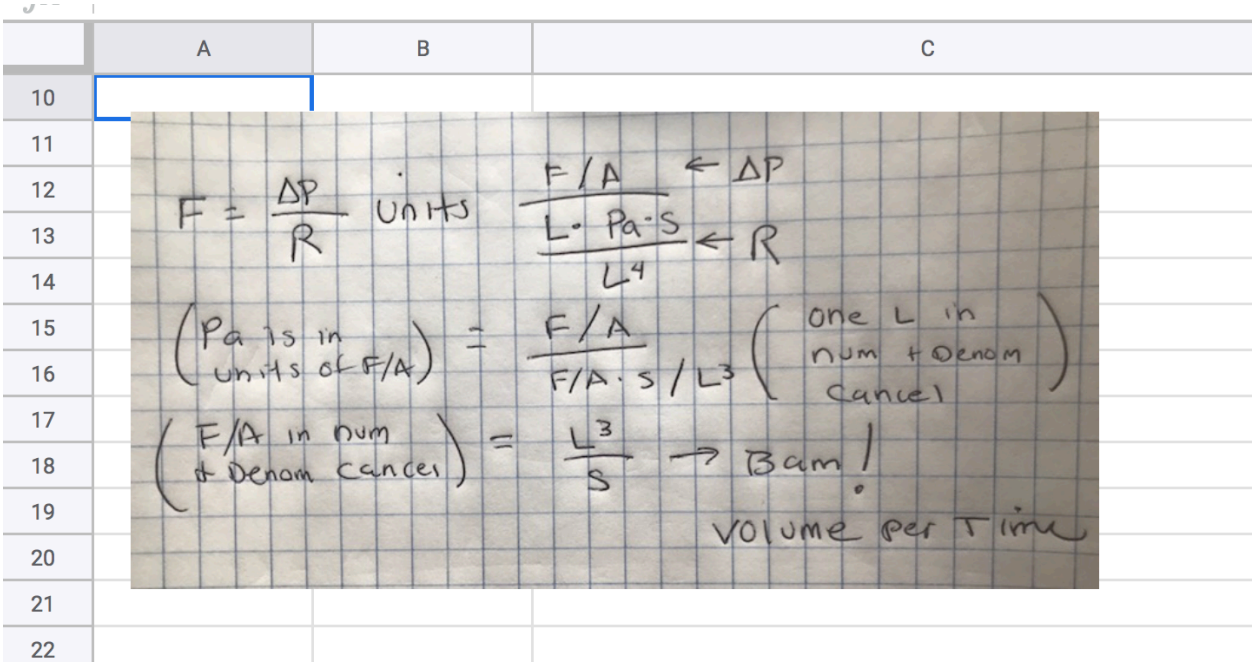
	A	B	C
1		units	definition
2	Flow		
3	pressure		
4	resistance		
5	length		
6	viscosity		
7	lumen radius		
8	stress		

2. The typical units of viscosity, $\text{Pa} \cdot \text{s}$ is not very intuitive. Wikipedia gives a nice way to think about viscosity:

   Viscosity is the material property which relates the viscous stresses in a material to the rate of change of a deformation (the strain rate)

Using your knowledge of stress and strain from last semester, how would you express this in units? To answer this 1) write "Viscosity…relates the viscous stresses in a material to the rate of change of a deformation" as an equation, and then 2) determine the units from this equation. Show how the units expressed this way equals $\text{Pa} \cdot \text{s}$. Do this with pencil and paper, snap a photo, and insert it below the table in your Google sheet.

Here is an example to follow: While the equation for flow is useful for understanding how variation in pressure and resistance cause variation in flow, if I use the equation to define flow, I would get something like "the change in pressure of the fluid per unit resistance", which isn't very helpful in thinking about flow. Flow is "the volume of fluid that moves past a point per unit time". So how do I get from the equation to this definition? I worked this out, snapped a photo, re-sized the image to 800 pixels wide, then inserted the image in my google sheet.



3. The radius of the lumen of an arteriole leading into a capilary increases 50%. What is the change in blood flow to the capillary? Use the google sheet to show your work, including *all* calculations.

Chapter 8

Resipiratory – Why we need Hemoglobin

We need hemoglobin because our blood cannot carry enough dissolved O2 to support our cell activity. That's the short answer. Let's explore a quantitative answer.

8.1 How much O2 can dissolve into blood?

We can model what is going on in the alveoli using a beaker filled with water. Diffusion of gas molecules from the air into the water ("going into solution") or the reverse ("coming out of solution") is not simply a function of the concentration gradient of the gas between the air and the water because *gasses have different solubilities in solution*.

Equilibrium (when an equal amount of gas is going into and out of solution) is modeled by the following equation

$$c_{O_2} = h_{O_2} P_{O_2} \tag{8.1}$$

- c_{O_2} is the concentration of O2 in the water
- P_{O_2} is the partial pressure of O2 in air (so a measure of concentration)
- h_{O_2} is Henry's solubility coefficient (or "constant of proportionality").

Scientists in different fields have different ways of expressing the relationship between c_{O_2} and P_{O_2} so you may land on a web page or a textbook that expresses the constant in something like $\frac{1}{c_{O_2}}$ or even a dimensionless constant. I like this way for physiology because it pumps our intuition about how P_{O_2} controls our dissolved O2 levels.

This equation tells us how much O2 will dissolve in the water at different partial pressures of O2 in the air, or, switching back to the lung, how much O2 will dissolve in the blood plasma given the partial pressure of O2 in the alveolar air.

Below is a table of P_{O_2} of alveolar air and the resulting concentration of dissolved O2 at equilibrium.

P_O2 (mmHg)	c_O2 (mL O2/dL blood)
20	0.066
30	0.091
40	0.137
50	0.156
60	0.195
70	0.220
80	0.257
90	0.281
100	0.307
110	0.349

1. Transfer the data into your spreadsheet.
2. Plot c_{O_2} (y-axis) against P_{O_2} (x-axis)
3. Compute the slope and and in the cell next to the computation, write the units.
4. What is this slope? (the concept not the value)
5. What is c_{O_2} in healthy arterial blood entering an organ (use $P_{O_2} = 97$ mmHg)?
6. How much dissolved O2 is ejected from our left ventricle each minute? Again, use $P_{O_2} = 97$ mmHg
7. How much O2 do our tissues need each minute? For this, you need to look up resting O2 consumption, which is usually in units of mL O2 per min per Kg. From this, you use the mass of a person to compute their O2 consumption per minute.
8. Compare the dissolved O2 sent by the left ventricle to the O2 required at rest? Do we send enough dissolvd O2 to our tissues?

Chapter 9

Renal

9.1 A back of the envelope calculation of GFR

1. Estimate GFR using a back-of-the-envelope calculation. The calculation is

GFR = cardiac output X renal fraction X plasma fraction X filtration fraction

Look up reasonable values for the four variables to parameterize this equation. Do the computation in your Google Sheet. Insert the units of GFR in the adjacent cell.

9.2 Using renal clearance to measure GFR in an indivudal

$$C_s = \frac{\dot{M}_s}{P_s} \tag{9.1}$$

where C_s is the clearance of solute s, M_s ("m dot") is the mass of s excreted in the urine per unit time, and P_s is the plasma concentration of s.

2. What are the units of C_s? These are the units of what kind of measure (for example Force per Area are the units of a pressure)?

3. Remember that a dot over a variable is a first derivative; here we assume that this is constant and so $\dot{M}_s = \frac{\Delta Mass}{\Delta Time}$. What are the units of $\dot{M}_s$? This kind of measure is "kinda like" the kind of measure in #2. Google around to see what we call $\dot{M}_s$.

4. The clearance of a solute is useful in pharmacology but we can also use the concept to measure the GFR in a person. This is done using a solute s that is filtered but no amount is either 1) secreted into the nephron, or 2) is not reabsorbed from the nephron). Inulin is an example. We could give a person some inulin and then measure the urine concentration of inulin (U_{in}), the volume of urine generated per time ($\dot{V}$), and the plasma concentration of inulin (P_{in}) to compute the GFR

$$GFR = \frac{U_{in}\dot{V}}{P_{in}} \tag{9.2}$$

(Note that I use $\dot{V}$ and not V to make it crystal clear that this is a measure of the volume of urine produced per time not simply a volume).

Using this information, compute the GFR for a person in which 1) inulin was given continuously to generate a constant plasma concentration of 1.0 mg/dL. 1.6 L of urine was collected over a 10 hour period. The urinary concentration of inulin was 462 mg/L.

Chapter 10

Digestion, Nutrition, and Metabolism

10.1 Estimating Causal Effects

Think about headlines in human health, performance and disease: red wine decreases colon cancer, or coffee increases dementia, or oxygenated water increases marathon performance. These "conclusions" constantly seem to be flipping. To understand why, it is important to think about the problem with math. The *mathematical* way to think about these assertions is X -> Y, or "X causes Y". Importantly, if a scientist says something like "X causes Y", this does not mean that X is the only cause of Y – other things may also cause Y. For example: vegetarian diet -> low blood cholesterol AND running -> low blood cholesterol AND statins -> low blood cholesterol.

Most importantly "cause" is not binary (causes v. doesn't cause) but has some magnitude (trivially small, or small, or big, or huge). Here, we use the greek letter β ("beta") to indicate effect size.

We are going to use Google Sheets to create fake data that were generated by a known causal process (known β), and then use a **statistical model** to estimate the causal process (estimate β) from the fake data. The statistical model is **regression**, which is the principle statistical method used in the biological sciences to estimate causal effects. We are purposefully using abstract notation (X and Y) instead of meaningful variables (dietary cholesterol and atherosclerotic plaque development) because it is good to be able to think abstractly.

	A	B	C	D	E	F	G	H
1		True Value	Estimate			True Value	Estimate	
2	beta_0	0			beta_0	0		
3	beta_1	0.5	0.472		beta_1	0.5	-0.004	
4	E[b1]	0.5			beta_2	-0.7		
5					r	0.7	0.701	
6					E[b_1]	0.01		
7								
8	sd	0.9828666119	0.9735327391			1.0066371	0.9910731407	1.032272965
9	ID	X	Y		Z	X1	X2	Y
10	1	1.441519473	0.7105884214		-2.957506448	-1.450754737	-2.42224558	0.1065029495
11	2	-1.067223794	-0.3866287056		-1.024169967	-1.095180579	-1.44544354	0.1770615012
12	3	1.690301322	0.8807498814		1.655578157	1.308331782	1.883125131	-1.010955082
13	4	-0.2457007644	0.2254569788		0.2663960621	-1.002803094	-0.5660069822	0.8461727216
14	5	-1.306222918	-0.8542635306		0.2086526157	-0.2589033115	0.7362888197	0.1166237296
15	6	-0.2527736106	-0.5132589204		0.9056618838	-0.1066822945	1.550229258	-1.604065218

10.2 Simulation 1

Open your Google spreadsheet and

10.2.1 Step 1. Set up the parameters

1. In column A, cells 2-4, insert "beta_0", "beta_1", "E[b1]" (see figure above)
2. In row 1, columns B and C, insert "True Value", "Estimate"
3. In B2, insert a number (it doesn't matter)
4. In B3, insert 0.5 (this is the true generating effect of X on Y)
5. In B4, insert =B3 (this is the expected value of the generating effect of X on Y given the statistical model)

10.2.2 Step 2. Generate fake data

1. In row 9, coumns A-C, insert "ID", "X", "Y"
2. In A10 insert "1"
3. In B10 insert `=normsinv(rand())`
4. In C10 insert `=$B$2 + $B$3*B10 + sqrt(1-$B$3^2)*normsinv(rand())`
5. In A11 insert `=A10 + 1`
6. Highlight cells B10 and C10. Click on the handle on the lower right corner of the box and drag down 1 row. Your formulas from row 10 should now be in row 11.
7. Highlight cells A11, B11, C11. Click on the handle on the lower right corner of the box and drag down and down and down until you get to row 1000. You should have copied all three formuilas all the way down.

What is step 2 doing? It is creating fake data. The value is caused by three things, the value in Cell B2, the product of B3 and X, and a random number. The value in B3 is the contribution of X to Y or how "X causes Y" or the "causal effect of X on Y". If B3 is 0 then there is no causal effect. If B3 is 1 or -1, then the random component is zero.

You have just created fake data with a known generating mechanism! But it is imperative to check the the equations you entered don't have bugs. If the equations were entered correctly, the standard deviation of the X and Y columns should both be one. Check this

10.2.3 Step 3. Fake data check

1. In A8, insert "sd"
2. In B8, insert `=stdev(B10:B1000)`
3. Copy B8 and paste in C8.

These numbers should be close to 1.0 (something is probably wrong if it is less than 0.95 or more than 1.05). Refresh the spread sheet by typing command-R (Mac) or control-R (Windows)

10.2.4 Step 4. Does a statistical model recover the known effect?

1. In C3, insert `=slope(C10:C1000, B10:B1000)`
2. In C3, round to three places after the decimal

This is the slope of the regression (the statistical model) of Y on X. It is the **estimate** of the causal effect. The number should be very close to the true value.

This slope is the **regression coefficient** b1. The cell labeled "E[b1]" is the "expectation of b1" or the expected value of b1. Your estimate of beta_1 should also be very close to E(b1) since E(b1) is equal to the true generating effect (beta_1).

10.2.5 What you did

10.2.5.1 ... in a nutshell

you generated Y using a "data generating" mechanism and then using the available data (X and Y), you used a statistical analysis to see if you could recover this data generating mechanism. The data generating mechanism is the set of two coefficients β_0 and β_1.

10.2.5.2 the data generating mechaniusm in a little more detail

The fake data are two variables, X and Y. Y is caused by three things:

$$y_i = \beta_0 + \beta_1 x_i + \sigma_i \tag{10.1}$$

the subscript is the "ith" individual (if ID=7 then i=7). The three components generating y_i are

1. β_0 is "the intercept"; it is common to all i
2. $\beta_1 x_i$ is the product of the effect (β_1) and an individuals value of x. β_1 is the same for all i but the product is unique to each i.
3. σ_i is "the error"; this is the random variation due to other factors that "cause" Y but are unique to each i. That is, these factors are **not correlated** with X.

10.2.6 The model you fit is

$$y_i = b_0 + b_1 x_i + e_i \tag{10.2}$$

1. b_0 is the intercept
2. b_1 is the slope
3. e_i is the residual (the difference between the modeled value and the actual value)

Notice that the statistical model is the same as the generating model. It is not at all surprising that the statistical model "recovers" the data generating mechanism (or the "true values"). **The problem in science is, we don't know the data generating model so we don't know the correct statistical model**. This will hopefully make more sense in the next exercise.

10.3 Simulation 2

10.3.1 Step 5. Set up the parameters

1. In column E, rows 2-6, insert the labels "beta_0", "beta_1", "beta_2", "r", "E(b_1)"
2. In row 1, columns F and G, insert the labels "True Value", "Estimate"
3. In F2, insert a number (it doesn't matter) (this is the baseline value of generating model)
4. In F3, insert 0.5 (this is the true generating effect of X_1 on Y)
5. In F4, insert -0.7 (this is the true generating effect of X_2 on Y)
6. In F5, insert 0.7 (this is the true correlation between X_1 and X_2)

10.3.2 Step 6. Generate fake data

1. In row 9, coumns E-H, insert "Z", "X1", "X2", "Y"
2. In E10 insert `=normsinv(rand())`
3. In F10 insert `=sqrt($F$5)*$E10 + sqrt(1-$F$5)*normsinv(rand())`
4. In G10, copy the equation from F10 and insert into G10
5. In H10, insert `=$F$2 + $F$3*F10 + $F$4*G10 + sqrt(1-$F$3^2 - $F$4^2 - 2*$F$3*$F$4*$F$5)*normsinv(rand(`
6. Highlight cells E10 through H10. Click on the handle on the lower right corner of the box and drag down and down and down until you get to row 1000. You should have copied all four formuilas all the way down.

What is step 6 doing? Like Step 2 in Simulation 1 above, it is creating fake data. But here the Y value is caused by five things:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i \tag{10.3}$$

1. β_0 is "the intercept"; it is common to all i
2. $\beta_1 x_{1i}$ is the product of the effect (β_1) and an individuals value of x_1. β_1 is the same for all i but the product is unique to each i. This is the causal or generating effect of X_1 on Y
3. $\beta_2 x_{2i}$ is the product of the effect (β_2) and an individuals value of x_2. β_2 is the same for all i but the product is unique to each i. This is the causal or generating effect of X_2 on Y
4. σ_i is "the error"; this is the random variation due to other factors that "cause" Y but are unique to each i. That is, these factors are **not correlated** with X.

what is the 5th cause of Y?

5. r – the correlation between X_1 and X_2. A correlation is a measures of association and is always between -1 and 1

10.3.3 Step 7. Fake data check

1. Check the standard deviation of X_1, X_2, and Y as in Step 3 above. All of these should be close to 1.0
2. insert `=correl(F10:F1000, G10:G1000)` in G5. This should be close to the true correlation in F5 (The starting correlation is 0.7, so the estimate should be 0.67-0.73)

10.3.4 Step 8. Does a statistical model recover the known effect?

1. In G3, insert `=slope(H10:H1000, F10:F1000)`
2. In G3, round to three places after the decimal
3. In F6, insert `=F3 + F4*F5`

As in Step 4 in Siumulation 1 above, this is the slope of the regression (the statistical model) of Y on X. It is the **estimate** of the causal effect. The number will not be very close to the true generating value (β_1), at least using the default values specified in Step 5. But it should be close to E(b1) (the expected value of b1), given the statistical model. But, unlike simulation 1, E(b1) is not similar to β_1, the true generating effect. Huh?

1. E(b1) should not equal the true value of beta_1 (at least using default values in Step 5), unlike in Simulation 1.
2. Your estimate of beta_1 should be very close to E(b1) but not to beta_1

10.3.5 What's going on is the whole point of this exercise

You have measured Y and X_1 but have not measured X_2. Because you haven't measured X_2, it is *not* in your statistical model, so your statistical model is just like that in Simulation 1.

$$y_i = b_0 + b_1 x1_i + e_i \tag{10.4}$$

But the generating model for Y is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i \tag{10.5}$$

That is, your statistical model has an **omitted causal variable** (X_2) and your estimate of the effect of X_1 is **biased**. This kind of bias is called **omitted variable bias**. The true effect of X_1 on Y is β_1 but you are actually estimating E(b1) with the regression coefficient! Researchers often think their result will get closer to the truth as the sample size increases but if an causal effect is missing from a statistical model, the estimate of the effects of the factors in the model will never ever get closer to the truth - instead it gets closer to the wrong thing (the biased expectation of the effect given the model).

10.4 Questions

10.4.1 Simulation 1

1. Given the default parameters specified above, what is the estimated effect of X_1 on Y in Simulation 1?
2. What is the true effect of X_1 on Y in Simulation 1?
3. If you increase your sample size, will the estimated effect of X_1 on Y move toward the true effect of X_1 on Y in Simulation 1?

10.4.2 Simulation 2

4. Given the default parameters specified above, what is the estimated effect of X_1 on Y in Simulation 2?
5. What is the true effect of X_1 on Y in Simulation 2?
6. If you increase your sample size, will the estimated effect of X_1 on Y move toward the true effect of X_1 on Y in Simulation 2?
7. If you did a study of X_1 on Y and the true generating model of Y is that in Simulation 2, what would you conclude about the effect of X_1 on Y?

10.4.3 Simulation 2 with new parameters

Redo Simulation 2 with the parameters: $\beta_1 = 0.0$. Leave $\beta_2 = -0.7$ and $r = 0.7$.

8. What is the estimated effect of X_1 on Y?
9. What is the true effect of X_1 on Y in Simulation 2?
10. If you did a study of X_1 on Y and the true generating model of Y is these new parameters in Simulation 2, what would you conclude about the effect of X_1 on Y?

Chapter 11

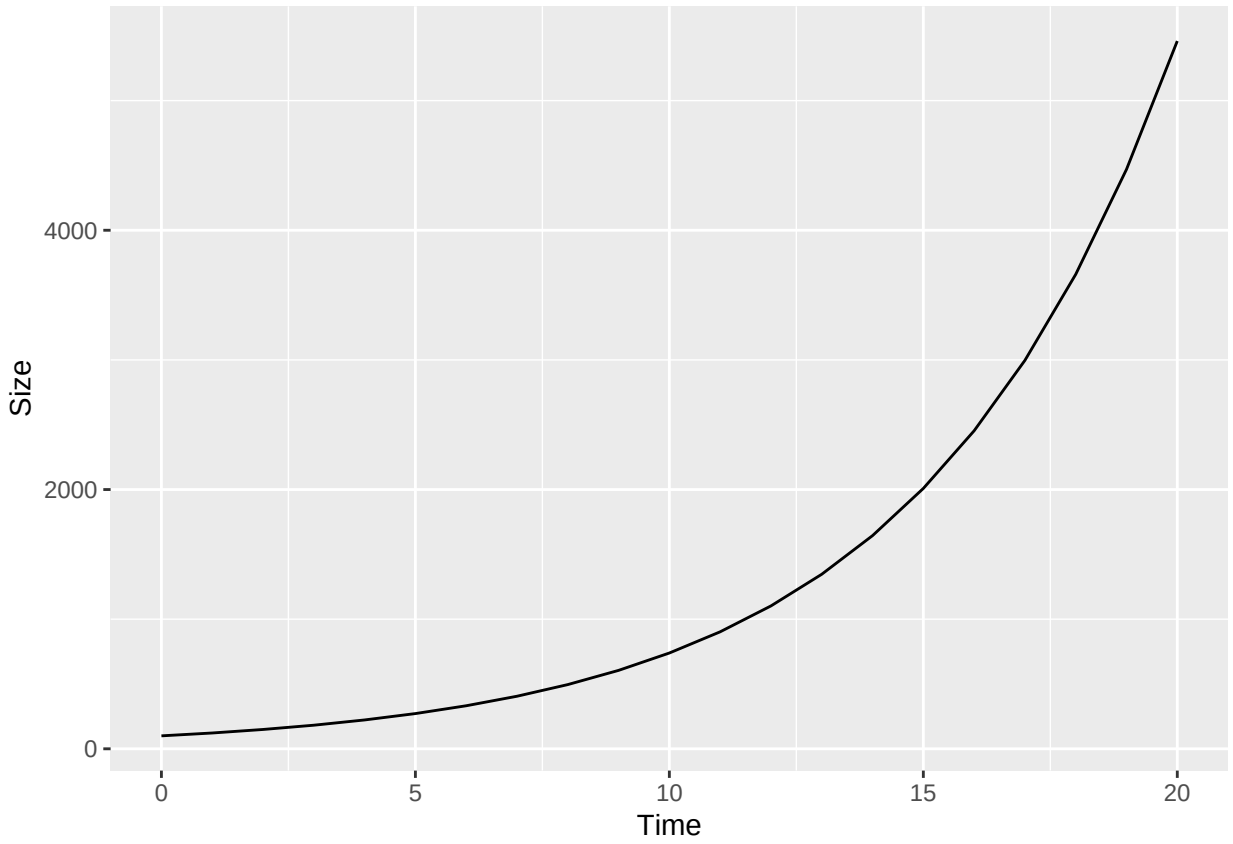
Cancer

11.1 Numerical Self-Discovery

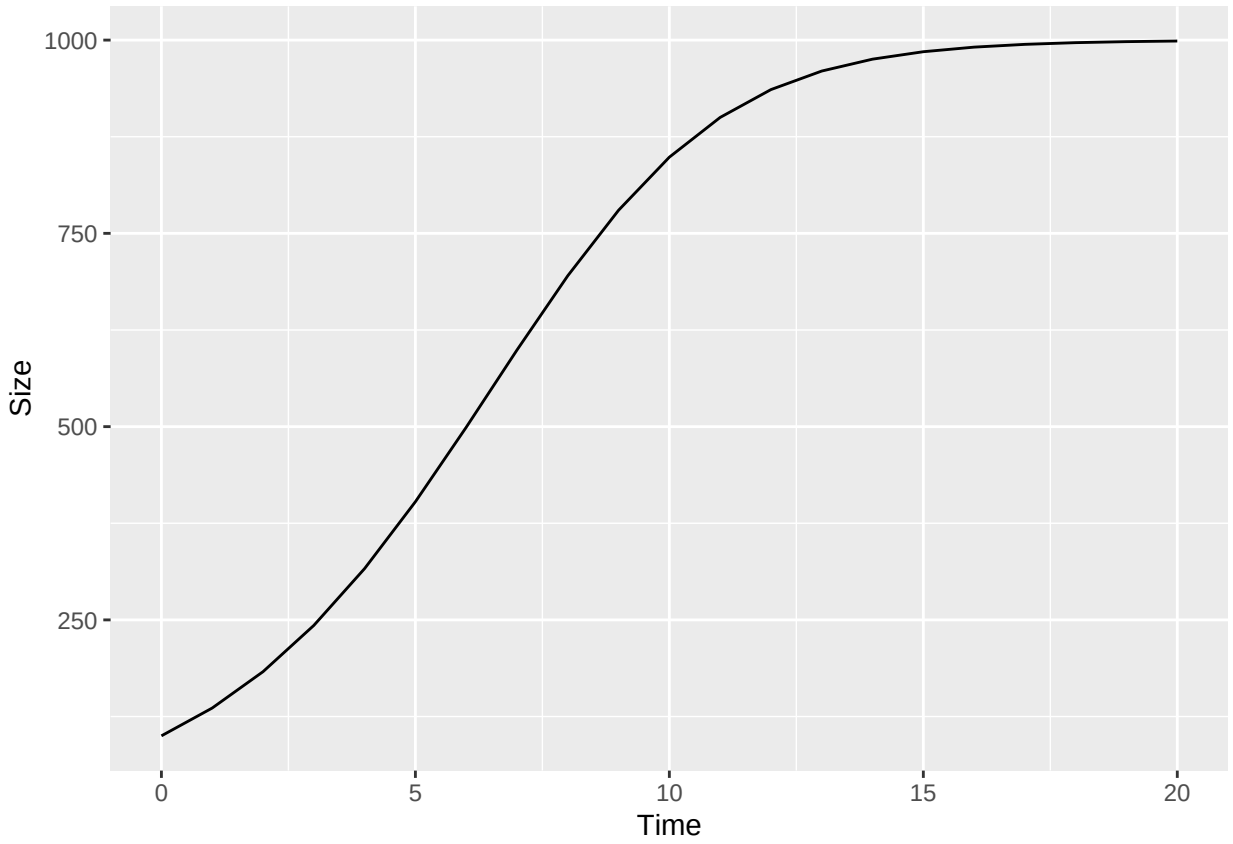
Students in Biology are inundated with equations both from Biology classes and from classes such as Physics and Chemistry. For most of us, these equations are pretty hard to really understand, even if we can do the math. By "understand", I mean, to treat the equation as a function, with inputs (the parameters) and outputs (what the function equates to given the parameters) and to have a picture in one's head of how this function behaves. By "behave", I mean what the function looks like and how "what it looks like" depends on the parameters. A spreadsheet is a terrific way to self-discover the behavior of a function (A scripting language like R is even better). This worksheet is a step-by-step guide on how to do this. The skills learned here should be transferable to any function, although a spreadsheet might not be the perfect tool for more complex functions.

The major feature of a tumor is the abnormal growth of the tumorogenic cell population. Physiologists are interested in the **dynamics** of tumor cell growth, which means *how* the cell population (or a proxy such as tumor volume) increases in size over time. How the growth occurs can give insights into the mechanisms driving the growth.

The beginning of tumor growth, when there are few cells, will probably be approximately exponential, which looks like this.



Exponential growth cannot continue for long. A growth model that might better describe the longer term growth of a tumor is the logistic model, which looks something like this



The logistic model is

$$\frac{\Delta N}{\Delta T} = rN(\frac{K-N}{K}) \tag{11.1}$$

Here we aren't concerned with deriving the logistic equation or with the assumptions of using this function to model tumor growth. We are concerned only with **how to explore the behavior** of the function itself. I'll leave derivation and assumptions to other worksheets.

You see this function in your textbook or on the blackboard. How do you use a spreadsheet to self-discover? That is, how do you get the equation into a spreadsheet? First, what is all the stuff in the equation? The parameters are r (the "intrinsic" growth rate), N (population size, that is the number of cells), and K (the maximum size of the tumor that can be sustained. This is sometimes called the "carrying capacity"). r, N, and k are on the right hand side (rhs) of the equal sign and variables on the rhs are the **inputs** to the function. The left hand side (lhs) of the equation is the output. The output is just a single number but there are three things on the lhs, Δ, N (population size, again), and T (time). The Δ is not a variable it is a symbol meaning "change in" and used to specify how one variable changes given change in another variable. So $\frac{\Delta N}{\Delta T}$, which is read as "delta en over delta tee" is the **change in population size over time**. The unit of time (minutes, days, generations) isn't specified and doesn't matter here.

To explore how the rate of population growth changes over time, We want to create a plot with N on the y-axis and $Time$ on the x-axis. The rate is then the tangent to the curve for any time, T. In order to do this, we need to **find the solution to the equation** by solving for N. First, remembering that

$$\frac{\Delta N}{\Delta T} = N_{t+1} - N_t \tag{11.2}$$

so substitute this in on the lhs

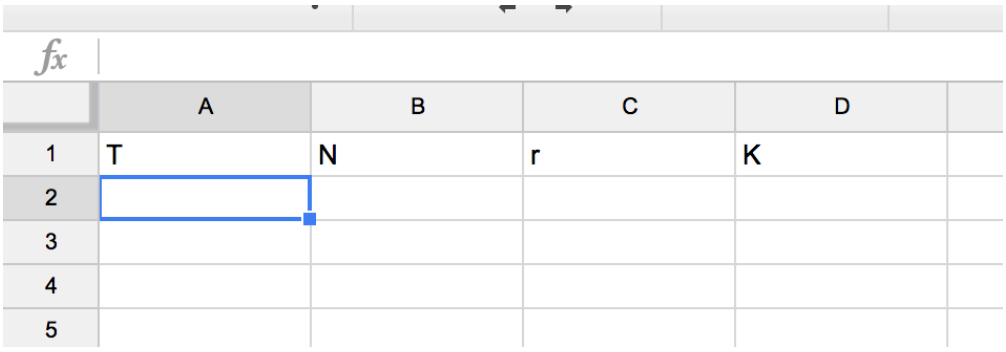


Figure 11.1: Step 1: Set up the header row of parameter labels

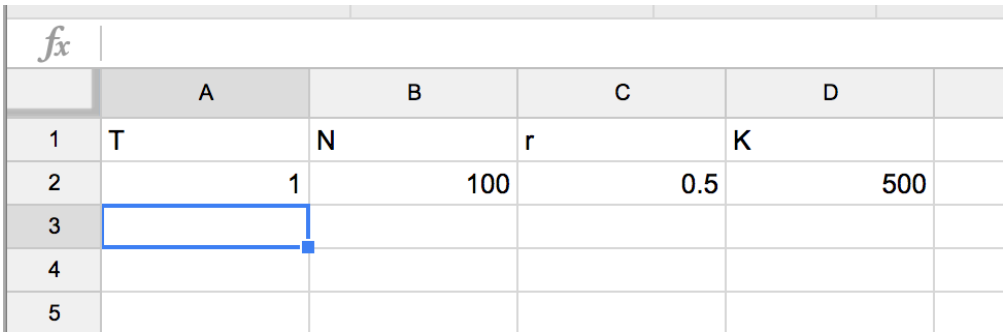


Figure 11.2: Step 2: Initialize T and N and parameterize r and K

$$N_{t+1} - N_t = rN_t(\frac{K - N_t}{K}) \tag{11.3}$$

and re-arrange ...

$$N_{t+1} = N_t + rN_t(\frac{K - N_t}{K}) \tag{11.4}$$

Equation (11.4) is function we can work with in a spreadsheet! So open your Google Sheet project and...

11.1.1 Step 1

First set up a row of column labels containing all the variables that we care about (Fig 11.1). This includes a column T for $Time$ that is not explicitly on the rhs but is there implicitly!

11.1.2 Step 2

Now add an initial value for T and N and some value for the parameters r and K (Fig 11.2). I inserted "1" for T, "100" for N, ".5" for r and "500" for K. Don't worry about the values you insert because the beauty of using a spreadsheet for self-discovery is that you *will* change these later!

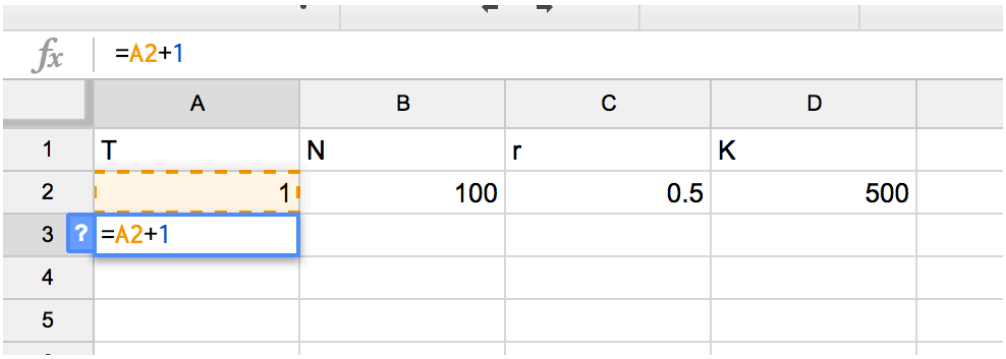


Figure 11.3: Step 3: Add a formula to generate a column of T values

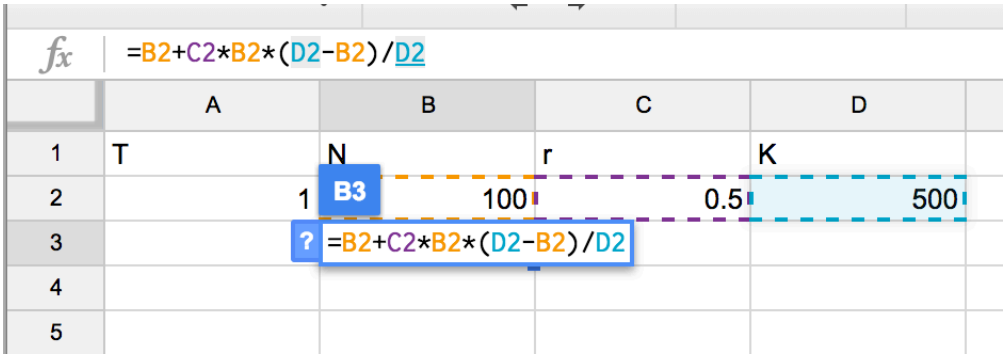


Figure 11.4: Step 4: Add the formula for N. This is step is the key step and the whole focus of the activity

11.1.3 Step 3

Now you need to add some formula to the spreadsheet. Google sheets or MS Excel recognize a formula of the text you enter starts with "=". The first formula you want to enter is for T and you want to enter this in the cell immediately below the initial value for T (Fig 11.3). If you've followed how I've input the header row and intial value row, then the formula goes into cell A3. As you type, the formula appears in the formula editor. Type "=" then either type "A2" (referring to the cell containing $T-1$ or simply click in cell A2, which inserts "A2" into the formula editor. Then click back into the formula editor and added "+1" then hit "return".

11.1.4 Step 4

Now you need to enter a formula for "N" in cell B3 (Fig 11.4). This formula is of course the function that we're focusing on (Equation (11.4)). The key is knowing how to translate a formula in the form you see it written in a textbook to a spreadsheet formula. In a textbook, you see rN for the product of r and N but if you type "=rN" into a spreadsheet you'll get an error because it doesn't know what r, N or rN is. So instead of entering "r" you need to type in the cell that contains r, and instead of "N" you need to type in the cell that contains N and finally you need to tell the spreadsheet to multiply the values of the two cells (r and N) using the product symbol "*". *So the rN part of the function is entered into the spreadsheet as* "=*C2*B2". Hopefully the rest of the formula makes sense.

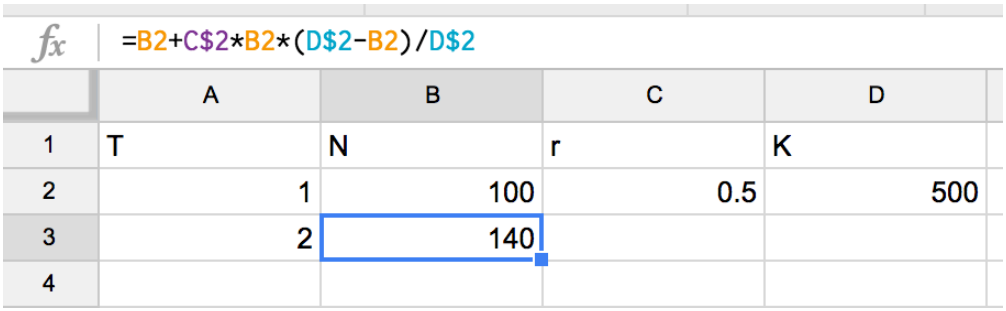


Figure 11.5: Step 5: Make the reference to the cells containing the values for r and K constant using the dollar sign

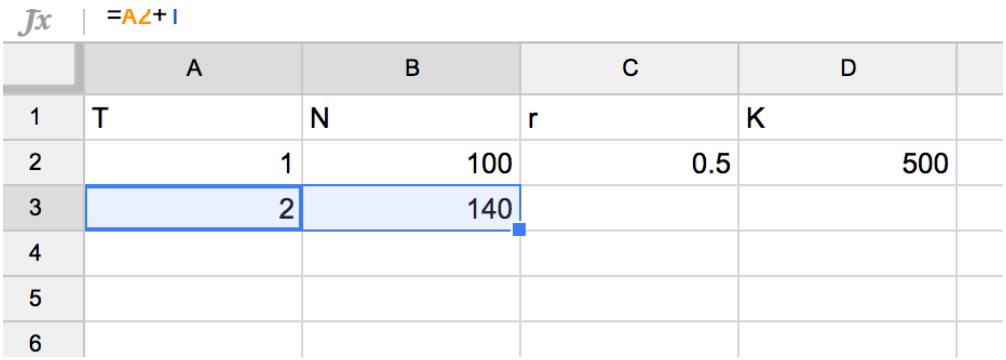


Figure 11.6: Step 6: Highlight then click-and-drag the little box in the bottom-right corner to copy the formula down

11.1.5 Step 5

The next step is a spreadsheet power tip. You want to compute N for many values of T. You could just keep re-typing in the formula but the more efficient method is to simply copy the formula down. The formula in row 3 refers to the values of cells in row 2. When a formula is copied from row 3 to row 4, all the row referents in the formula are increased by 1, so now the formula in row 4 refers to cells in row 3. This is beautiful. Except when it isn't. Some parts of the formula that you entered in cell B3 are "constant", that is we do not want these to "move down" with the formula. So to keep the row constant, simply add a dollar sign in front of the number (=row) part of the cell address (Fig 11.5). For example, r, in cell C2 is a constant, so you need to add a dollar sign in front of the 2 in "C2". The paramter K (cell D2) is also a constant so any part of the formula that refers to D2 should be kept constant by adding a dollar sign in front of the 2 in "D2". Sweet!

11.1.6 Step 6

Now, the formula is ready to copy "down". This can be done several ways. The way I usually do it is by highlighting the cells that I want to copy down and click-and-drag (down) the bold little square in the bottom right corner of the bounding box of the highlighted cells (Fig 11.6). Bam! You should have multiple rows with T increasing from 1 to however many rows you copied the formula and Values of N for each value of T.

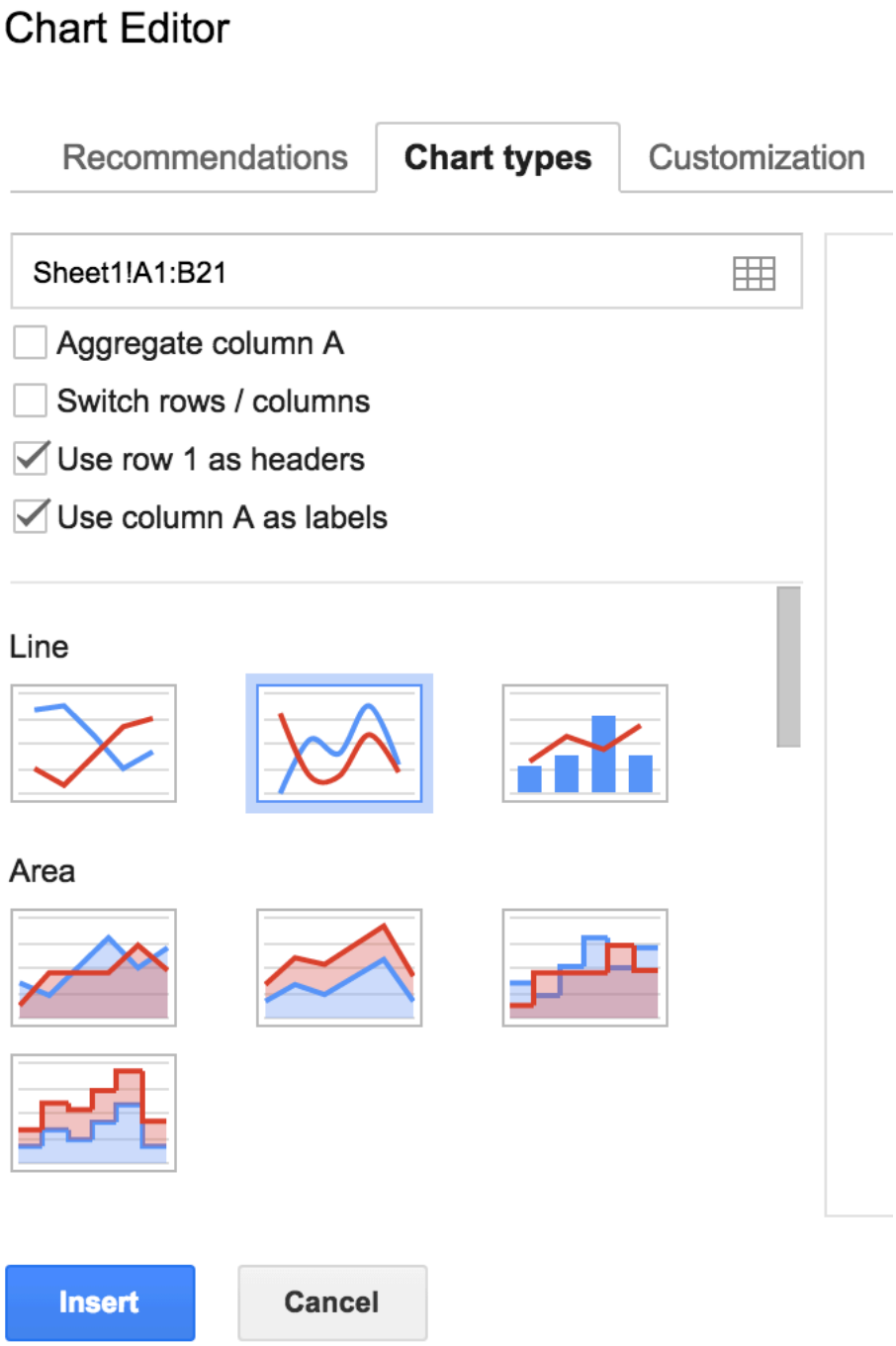


Figure 11.7: Step 7: Make a figure

11.1.7 Step 7

Now just plot the data. Click-and-drag to highlight the columns T and N including the labels "T" and "N". Now choose the menu "Insert" and the item "Chart". A default chart might appear. In the chart editor window, click on chart type and click on the **Line Chart** with the smooth lines (this is the chart icon highlighted in Fig 11.7 by the light blue line). Then click "Use column A as labels" which makes the values in the column labeled "T" the x-axis. The column labeled N will be the y-axis. Now you have a plot of the function. Now it's time to explore the function (Equation (11.4)) by changing r or changing K or changing the initial N. The plot will immediately re-draw in response to these changes. Explore around then come back to the sections below.

11.2 Stuff to explore

11.2.1 Getting to know the model

Set the initial T to 1, the initial N to 100, r to 0.6, and K to 500.

1. how does the function change if you set r to .8? What about setting r to -.5? Can r take the value -.5?
2. how does the function change if you set K to something different, say to 1000? What about to something less than the initial N, say 50. Does setting K to something less than the initial N make sense. Think about a scenario that it does. If you cannot, then maybe it doesn't make sense. Or maybe it does, but you just cannot think of a scenario! can K be a negative value?

11.2.2 A wee bit more to explore

Set the initial T to 1, the initial N to 1, r to 0.6, and K to 500.

3. r is the intrinsic rate of growth. Is r the rate of growth at $T = 10$? What about at $T = 1$
4. If not, what is the rate of growth at $T = 10$ or $T = 1$?

The answer is $r(\frac{K-N}{K})$ for the value of N associated with $T = 1$ or $T = 10$. Call this actual rate of growth R. Think of R as r penalized by density. Create a new column that computes R for each value of T. Make sure that you have about 30 rows of data (so T going from 1 to 30).

5. Create a plot with R on the y-axis and N on the x-axis.
6. How does R change as time goes by?
7. What is the y-intercept equal to? (hint, it's a parameter in Equation (11.4)
8. What is the x-intercept equal to? (hint, it's also a parameter in Equation (11.4)

11.2.3 Yet more stuff to explore

Set the initial T to 1, the initial N to 1, r to 0.6, and K to 500. Remember that we started with the function for $\frac{\Delta N}{\Delta T}$

9. Create a column that computes $\frac{\Delta N}{\Delta T}$ for each value of T. Hint: you cannot do this for the first row ($T = 1$), why?
10. Create a plot with $\frac{\Delta N}{\Delta T}$ on the y-axis and T on the x-axis.
11. How does $\frac{\Delta N}{\Delta T}$ change as time goes by (as T gets bigger)?
12. Why?