1.2 Tables and graphs

Lung cancer, chronic bronchitis, bad breath, stains on your clothes, and the expense. These are just a few of the consequences of smoking cigarettes. With what we know now about the dangers of smoking, are people smoking more or less than they were ten years ago, fifty years ago, or even one hundred years ago?

Reality is, we don't have information on each individual person's smoking rate, so we can't answer this question exactly. We do have information on the total number of cigarettes sold each year. So maybe we should look at that total. Uh oh, that isn't going to work. There are way more people now than there were fifty or a hundred years ago. So, even if the same percentage of people smoke, and even if they each smoke the same amount as their predecessors did, we would have a much bigger number of cigarettes smoked now just because there are more people now.

Turns out a reasonable measure is to compare the number of cigarettes smoked per year *per person*. By taking into account the number of people we will be able to see whether people are smoking more or less, on average. That's what we want.

Here are some representative years from the Center for Disease Control for the United States. The smoking rate is the average cigarettes per year per person. (Here "person" only includes adults.)

Year	1900	1915	1930	1940	1950	1965	1975	1990	2000	2006
Smoking rate	54	285	1,485	1,976	3,552	4,258	4,122	2,834	2,049	1,619

To make sense of these numbers, suppose there are five friends. Three don't smoke at all, so that is 0 cigarettes in a year. Another smokes only occasionally, maybe 100 cigarettes a year. The fifth smokes "a pack a day," which adds up to 7,300 cigarettes in a year because

$$\frac{1 \text{ pack}}{\text{day}} * \frac{20 \text{ cigarettes}}{\text{pack}} * \frac{365 \text{ days}}{\text{year}} = 20 \times 365 = \frac{7{,}300 \text{ cigarettes}}{\text{year}}$$

(Not sure about this calculation? Not to worry. More about unit conversions in §1.4.) These five people smoke a total of

$$0 + 0 + 0 + 100 + 7{,}300 = 7{,}400 \text{ cigarettes per year}$$

so when we divide by the number of people we get

$$\frac{7{,}400 \text{ cigarettes per year}}{5 \text{ people}} = 7{,}400 \div 5 = 1{,}480 \text{ cigarettes per year per person}$$

This group is fairly typical for the United States in 2012. They smoke less than the average of 1,619 cigarettes per year per person for 2006 (the last year the CDC published the data).

We can tell a lot of information from this table. For example, what was the smoking rate in 1964, and how does that compare to 2006? The answers appears in the table, a whopping 3,552 cigarettes per person in 1964 and 1,619 cigarettes per person in 2006.

When did the consumption first pass 3,000? That answer does not appear in the table, but we can use the information in the table to make a good guess. In 1940, there were an average of 1,976 cigarettes per person per year and by 1950, there were 3,552. Somewhere between 1940 and 1950 the number first climbed above 3,000. More specifically, the number we're looking for (3,000) is a lot closer to the 1950 figure (3,552) than to the 1940 figure (1,976). So, it would be reasonable to guess close to 1950. I'd say 1947. Of course, you might guess 1946 or 1948, or even 1949 and those would be good guesses too.

When did the consumption drop below 3,000 again? This answer also does not appear in the table, but falls somewhere between 1975 when consumption was 4,122 and 1990 when consumption was 2,834 . Here I'd guess just before 1990, say in 1989.

What's changing are the number of cigarettes smoked per person per year and the year. Those are our variables. The smoking rate is a function of year, and it's what we care about, so it's the dependent variable. Time, as measured in years, is the independent variable.

$$S = \text{smoking rate (cigarettes per year per person)} \sim \text{dep}$$
$$Y = \text{year (years since 1900)} \sim \text{indep}$$

Quick note on how we deal with actual years. Since the year 0 doesn't make sense in this problem, it is convenient to measure time in years since 1900. Officially we should rewrite our table as:

Y	0	15	30	40	50	65	75	90	100	106
S	54	285	1,485	1,976	3,552	4,258	4,122	2,834	2,049	1,619

Notice where the variable names are listed in the table. In a horizontal format like this table, the independent variable (Y) is in the top row, with the dependent variable (S) is in the bottom row. If you want to write your table in a vertical format, that's okay too. Just put the independent variable in the left column, with the dependent variable in the right column. It might help to remember that the independent variable goes first (either top or left) and the dependent variable follows (either bottom or right).

Horizontal table format:

indep				
dep				

Vertical table format:

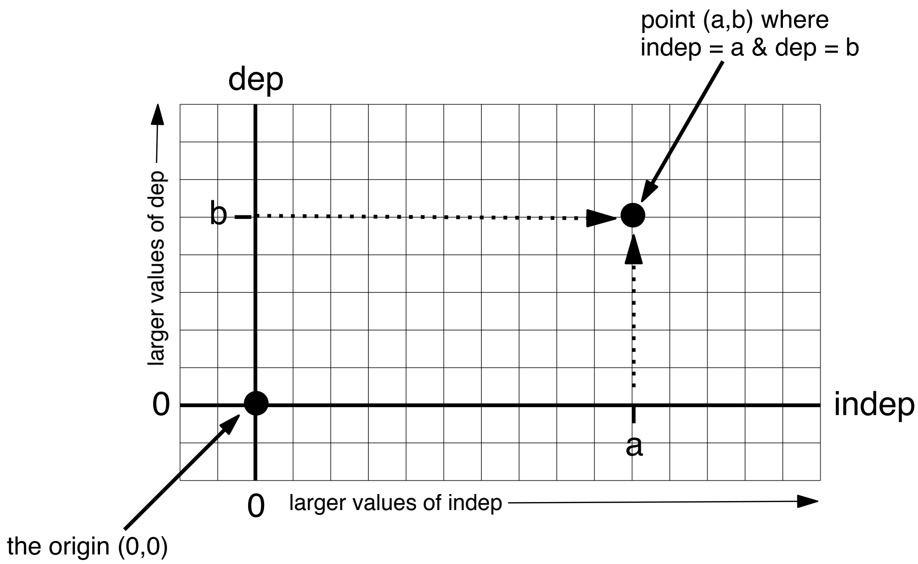
indep	dep

Where the variables go in a table is not something you can figure out. It's a **convention** – a custom, practice, or standard used within the mathematical community. Though based on reason, it often involves some arbitrary choice, which is why we can't figure it out. So, whenever some practice is introduced to you as a "convention", you need to memorize it.

Tables are useful because they contain specific numbers, but it can be difficult to guess or see general trends. For that, a picture is worth a thousand words – or numbers, in this case. By picture we mean the graph of the function.

Throughout this text, we draw graphs by hand. On graph paper. Seriously. You might wonder why we do that when graphing calculators, spreadsheet programs, graphing "apps," or computer algebra systems all can draw graphs for us. The answer is we want to understand graphs better. I promise – drawing them by hand will help you do that. Different folks have different opinions on the importance of graphing by hand, so be sure to ask your instructor what you are expected to do. Even if you're allowed to use some type of graphing technology, I strongly encourage you to practice drawing graphs by hand as well.

There is a standard set up for a graph.



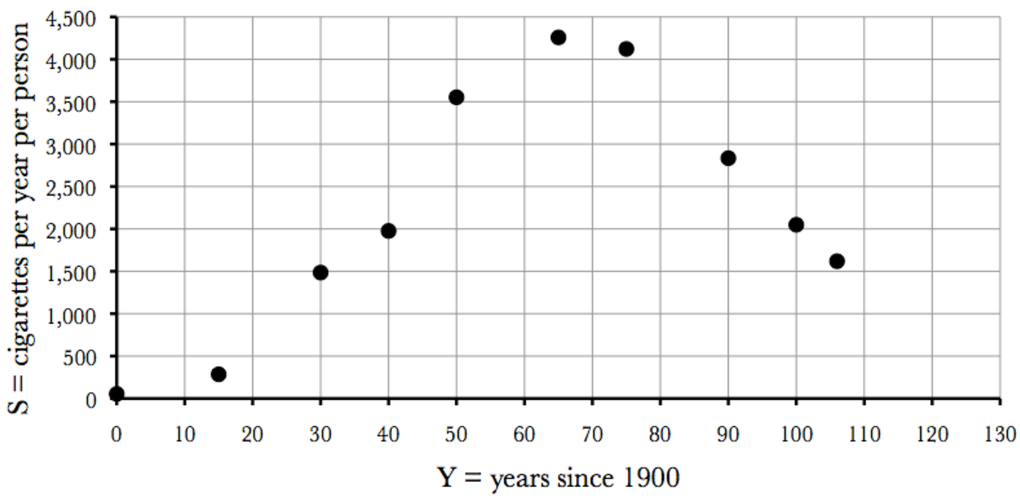
The graph is based on a horizontal line and vertical line, called the **axes**. Where they cross is a point called the **origin**. It represents where each variable is 0. By convention, the independent variable is measured along the horizontal axis, with larger values progressing to the right of the origin, and negatives to the left. Similarly, the dependent variable is measured along the vertical axis, with larger values progressing up from the origin, and negatives down. Each gridline counts the same number, called the **scale**, but the scale for the vertical may be different from the scale for the horizontal. Each pair of values of the independent and dependent variable from our table correspond to a point on our graph.

In the graph of smoking rates, the independent variable is Y, the year, so that goes on the horizontal axis for our graph. Our dependent variable is S, the smoking rate, so that goes on the vertical axis. For the scale, it works nicely to count by 10 years and count by 500s for the smoking rate.

There's a certain amount of guess and check involved in figuring out a good scale for each axis. As a general rule of thumb we would like the graph to be as large as possible so we can see all of its features clearly. But, not so big that it runs off the graph paper. What matters is that the gridlines are evenly scaled and that they can handle large enough numbers. Speaking of which, it's a good idea to leave a little room to extend the graph a little further than the information we have in the table, in case we get curious about values beyond what we have already.

With realistic numbers it's normal to have numbers in the table that are not exactly where the gridlines are. It is very helpful to count by round numbers (2s, 5s, 10s, etc.)

because it makes guessing in between easier. Easier for you drawing the graph. Easier for someone reading your graph.



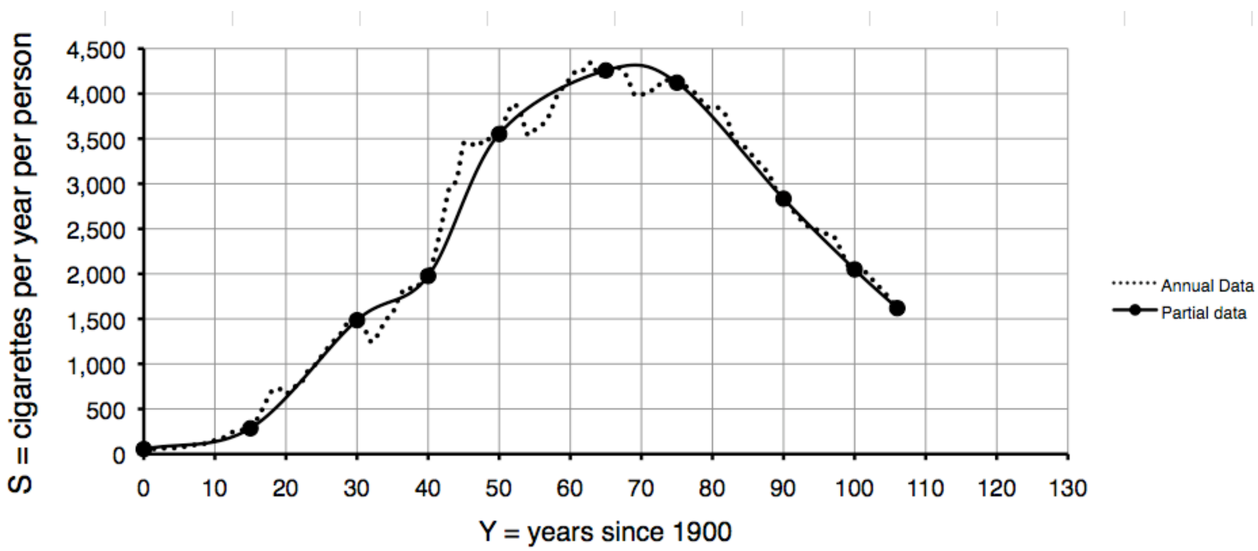
To plot each point, we start at the origin and move right to that Y-value, and then up to that S-value. When a value doesn't land exactly on a grid mark, we have to guess in between. For example, in 1900, when $Y = 0$ so we don't move right at all, just up to $S = 54$. The first labeled gridline on our graph is 500. Where's 54? It's between 0 and 500, very close to 0. Our point is just a tiny bit above the origin. In 1915 we have $Y = 15$. Our labeled gridlines are for 10 and 20, so 15 must land halfway in between. The smoking rate to 285, which is around halfway between 0 and 500. Etc.

What we have so far is a **scatter plot** of points. Can you see why it's called that? Anyway, our whole goal here was to be able to understand smoking rates better by having a graph. You may already begin to see a curve suggested by the points. Time to draw it in. I don't mean drawing a line between each pair of points, like you do in the children's game "connect the dots." That isn't quite right. It was probably more of a continuous trend and so the graph should be smoother.

When we draw in this smooth curve for the graph, what we are really doing is making a whole lot of guesses all at once. For example, from the table we guessed that the smoking rate passed 3,000 in around 1947, and dropped back to that level in around 1989. What does the graph show? If we look where the horizontal gridline for 3,000 crosses our graph, it crosses in two places. First, between the vertical gridlines for 40 and 50, and perhaps slightly closer to 50. I'd say $Y \approx 47$, in the year 1947. Sure. The second time is between the gridlines for 80 and 90, much closer to 90. Looks like $Y \approx 88$, in the year 1988. We guessed 1989. Close enough.

Don't forget that when we drew in that curve it was really just a guess. We're sure about the points we plotted, but we're only guessing about where to draw the curve in. That means we're not sure about the other points. If we knew a lot more points we could have a more accurate graph.

Turns out more data is available from the CDC. The full table of data from the CDC shows that consumption first topped 3,000 as early as 1944. Here's an example where the history tells you more than the mathematics as cigarette consumption rose sharply during World War II. Our guess about 1988 or 1989 was spot on. Look at how the graph from the full data compares to our guess.



5.3 Growth factors

Obesity among children ages 6-11 continues to increase. From 1994 to 2010, the proportion of children classified as obese rose from an average of 1.1 out of every ten children in 1994 to around 2 out of every ten children in 2010. Assuming that the prevalence of childhood obesity increases exponentially, what is the annual percent increase and what does the equation project for the year 2020? Well, unless we are able to make drastic improvements in how children eat and how much they exercise.

Because we are told obesity is increasing exponentially we can use the template for an exponential equation.

$$\text{dep} \;=\; \text{start} \;*\, \text{growth factor}^{\text{indep}}$$

The variables are

$$\begin{aligned} C &= \text{obese children (out of every ten)} \sim \text{dep} \\ Y &= \text{year (years since 1994)} \sim \text{indep} \end{aligned}$$

The starting amount is 1.1 children out of every ten in 1994 so our equation is of the form

$$C = 1.1 * g^Y$$

Trouble is we don't actually know what the growth factor g is. Yet.

We do know that in 2010 we have $Y = 2010 - 1994 = 16$ years and $C = 2$. We can put those values into our equation to get

$$1.1 * g^{16} = 2$$

No good reason for switching sides, just wanted to have the variable on the left. That's supposed to be true but we don't know what number g is so we can't check. Argh.

Oh, wait a minute. The only unknown in that equation is the growth factor g. What if we solve for g? First, divide each side by 1.1 to get

$$\frac{\cancel{1.1} * g^{16}}{\cancel{1.1}} = \frac{2}{1.1}$$

which simplifies to

$$g^{16} = \frac{2}{1.1} = 2 \div 1.1 = 1.818181818\ldots$$

Since we want to solve for the base (not the exponent), we have a power equation. We use the ROOT FORMULA with power $n = 16$ and value $v = 1.818181818$ to get

$$g = \sqrt[n]{v} = \sqrt[16]{1.818181818} = 16 \sqrt[x]{}\ 1.818181818 = 1.038071653 \approx 1.0381$$

Want a quicker way to find the growth factor? Forget the entire calculation we just did. It all boils down to two steps:

$$\frac{2}{1.1} = 2 \div 1.1 = 1.818181818\ldots$$

and then

$$g = \sqrt[16]{1.818181818} = 16 \sqrt[x]{}\ 1.818181818 = 1.038071653 \approx 1.0381$$

We can do this calculation all at once as

$$g = \sqrt[16]{\frac{2}{1.1}} = 16 \sqrt[x]{} (2 \div 1.1) = 1.038071653 \approx 1.0381$$

Notice we added parentheses because the normal order of operations would do the root first and division second. We wanted the division calculated before the root.

Here's the easy version in a formula.

Growth Factor Formula

If a quantity is growing (or decaying) exponentially, then the growth (or decay) factor is

$$g = \sqrt[t]{\frac{a}{s}}$$

where s is the starting amount and a is the amount after t time periods.

Let's check. In our example, the number of obese children grew from the starting amount of $s = 1.1$ to the amount of $a = 2$ after $t = 16$ years. By the Growth Factor Formula we have

$$g = \sqrt[t]{\frac{a}{s}} = \sqrt[16]{\frac{2}{1.1}} = 16 \sqrt[x]{} (2 \div 1.1) = 1.038071653 \approx 1.0381 \quad \checkmark$$

Either way, we knew from the beginning that our equation was in the form $C = 1.1 * g^Y$. Now that we found the growth factor $g \approx 1.0381$ we get our final equation

$$C = 1.1 * 1.0381^Y$$

For example, we can check that in 2010, we have $Y = 16$ still and so

$$C = 1.1 * 1.0381^{16} = 1.1 \times 1.0381 \wedge \underline{16} = 2.000874004 \approx 2 \quad \checkmark$$

You might wonder why we didn't just round off and use the equation

$$C = 1.1 * 1.04^Y$$

Look what happens when we evaluate at $Y = 16$ then. We would get

$$C = 1.1 * 1.04^{16} = 1.1 \times 1.04 \wedge \underline{16} = 2.06027937 \approx 2.1$$

Not a big difference (2.1 vs. 2.0) but enough to encourage us to keep extra digits in the growth factor in our equation. Lesson here is: don't round off the growth factor too much.

Back to the more reliable equation

$$C = 1.1 * 1.0381^Y$$

We can now answer the two questions. First, in 2020 we have $Y = 2020 - 1994 = 26$ and so

$$C = 1.1 * 1.0381^{26} = 1.1 \times 1.0381 \wedge \underline{26} = 2.908115507 \approx 2.9$$

According to our equation, by 2020 there would be approximately 2.9 obese children for every ten children.

The other question was what the annual percent increase is. Think back to an earlier example. Remember that Jocelyn was analyzing health care costs? They began at \$2.26 million and grew 6.7% per year. She had the equation

$$H = 2.26 * 1.067^{Y}$$

So the growth factor $g = 1.067$ in the equation came from the growth rate $r = 6.7\% = .067$. Our equation modeling childhood obesity is

$$C = 1.1 * 1.0381^{Y}$$

The growth factor of $g = 1.0381$ in our equation must come must come from the growth rate $r = .0381 = 3.81\%$. Think of it as converting to percent $1.0381 = 103.81\%$ and then ignoring the 100% to see the 3.81% increase. Childhood obesity has increased around 3.81% each year. Well, on average.

Here's the general formula relating the growth rate and growth factor.

PERCENT CHANGE FORMULA: (updated version)

- If a quantity changes by a percentage corresponding to growth rate r, then the growth factor is

$$g = 1 + r$$

- If the growth factor is g, then the growth rate is

$$r = g - 1$$

Let's check. We have $g = 1.0381$ and so the growth rate is

$$r = g - 1 = 1.0381 - 1 = .0381 = 3.81\%$$

Not sure we really need these formulas, but there you have it.

By the way, formula works just fine if a quantity decreases by a fixed percent. One example we saw was Joe, who drank too much coffee. The growth (or should I say decay) factor was $g = .87$. That corresponds to a growth (decay) rate of

$$r = g - 1 = .87 - 1 = -.13 = -13\%$$

Again, the negative means that we have a percent decrease.

5.4 Linear vs. exponential models

Sofía bought her car new for \$22,500. Now the car is fairly old and just passed 109,000 miles. Sofía looked online and estimates the car is still worth \$5,700. She wonders when the car would be practically worthless, meaning under \$500.

We can describe the variables in this story.

$$\begin{aligned} M &= \text{mileage (thousand miles)} \sim \text{indep} \\ C &= \text{value of car (\$)} \sim \text{dep} \end{aligned}$$

Notice we are measuring the mileage in thousands. The information we are given is

M	0	109
C	22,500	5,700

But what's the equation? Hmm. Don't know for sure what type of equation might work here. Tell you what, let's compare what a linear and exponential model would tell us about the value of the car.

First, linear. The template is

$$\textsc{Linear Equation Template:} \quad \text{dep} = \text{start} + \text{slope} * \text{indep}$$

The starting value of Sofía's car is \$22,500 so we just need to find the slope. We expect the slope to be negative because her car is worth less the more she drives it.

$$\begin{aligned} \text{slope} &= \text{rate of change} = \frac{\text{change dep}}{\text{change indep}} = \frac{\$5{,}700 - \$22{,}500}{109 \text{ thousand miles}} = \frac{-\$16{,}800}{109 \text{ thousand miles}} \\ &= \text{(-)}16{,}800 \div 109 = -154.1284404\ldots \approx -\$154/\text{thousand miles} \end{aligned}$$

Her car loses value at a rate of around \$154 for each thousand miles she drives.

We are ready to write the linear equation.

$$\textbf{linear:} \quad C = 22{,}500 - 154M$$

As usual we wrote subtract, instead of adding the negative slope. Quick check – when $M = 109$ we get

$$C = 22{,}500 - 154 * 109 = 22{,}500 - 154 \times \underline{109} = 5{,}714 \approx \$5{,}700 \quad \checkmark$$

Remember we don't expect the exact answer because we rounded off the slope.

When will Sofía's car be worth under \$500 according to this linear equation? Let's solve the equation to find out. When $C = 500$, use our linear equation to get

$$22{,}500 - 154M = 500$$

Subtract 22,500 from each side and simplify to get

$$-154M = -22{,}000$$

Now divide each side by -154 and simplify to get

$$M = \frac{-22{,}000}{-154} = \text{(-) } 22{,}000 \div \text{(-)}154 = 142.738095\ldots \approx 143$$

According to the linear equation, Sofía's car will be worth under \$500 at about 143,000 miles. Since her car already has 109,000 miles on it, that means in another 143,000 − 109,000 = 34,000 miles. For a typical driver that's two or three more years.

Next, let's take a look at the exponential model. Here goes. The template is

$$\text{EXPONENTIAL EQUATION TEMPLATE:} \quad \text{dep} \; = \; \text{start} \; * \text{growth factor}^{\text{indep}}$$

We know everything except the growth factor. We expect it to be less than 1 because her car is worth less the more she drives it. Perhaps we should say "decay" factor here since the function is decreasing. The starting amount is $s = 22{,}500$ and the ending amount is $a = 5{,}700$ after $t = 109$ thousand miles. By the GROWTH FACTOR FORMULA we have

$$g = \sqrt[t]{\frac{a}{s}} = \sqrt[109]{\frac{5{,}700}{22{,}500}} = 109 \sqrt[x]{} (5{,}700 \div 22{,}500) = 0.98748222\ldots \approx .9875$$

We are ready to write the exponential equation.

$$\textbf{exponential:} \quad C = 22{,}500 * .9875^M$$

Quick check – when $M = 109$ we get

$$C = 22{,}500 * .9875^{109} = 22{,}500 \times .9875 \wedge \underline{109} = 5711.19365\ldots \approx \$5{,}700 \quad \checkmark$$

Again, we don't expect the exact answer because we rounded off the decay factor.

When will Sofía's car be worth under \$500 according to this exponential equation? Let's solve the equation to find out. When $C = 500$, use our exponential equation to get

$$22{,}500 * .9875^M = 500$$

Divide each side by 22,500 and simplify to get

$$.9875^M = \frac{500}{22{,}500} = 500 \div 22{,}500 = .02222222\ldots$$

By the THE LOG-DIVIDES FORMULA with growth factor $g = .9875$ and the value $v = .02222222$ we get

$$M = \frac{\log(v)}{\log(g)} = \frac{\log(.022222222)}{\log(.9875)} = \log(.02222222) \div \log(.9875) = 302.6256856 \approx 300$$

According to the exponential equation, Sofía's car will be worth under \$500 at about 300,000 miles. Hard to imagine the car would last that long. Essentially the exponential model says the car will always be worth at least \$500, if only for parts, I guess. Quite different from our answer from the linear equation.

Time to compare models. Which one makes more sense? First things first, the car already has a lot of miles on it. Don't know what make or model the car is, but another couple of years seems a reasonable time until is worth under \$500. That's what the linear equation projects. On the other hand, the exponential model project it will hold that value for a long time, essentially for parts. That makes sense too.

Wait a minute. Does a car lose the same value for each thousand miles it's driven? That's what it means to be linear. Every thousand miles, same decrease. Nah, that's not right. Once the car is old, another 1,000 miles or so probably won't affect the value at all. Also, when a car is new, once you drive it off the lot and then that strange vinyl smell wears off and it's officially "used," the car is worth a lot less. Even if it hasn't been driven much at all. What would each model say the car was worth soon after Sofía bought it, say with 10,000 miles on it? With $M = 10$, the estimates are

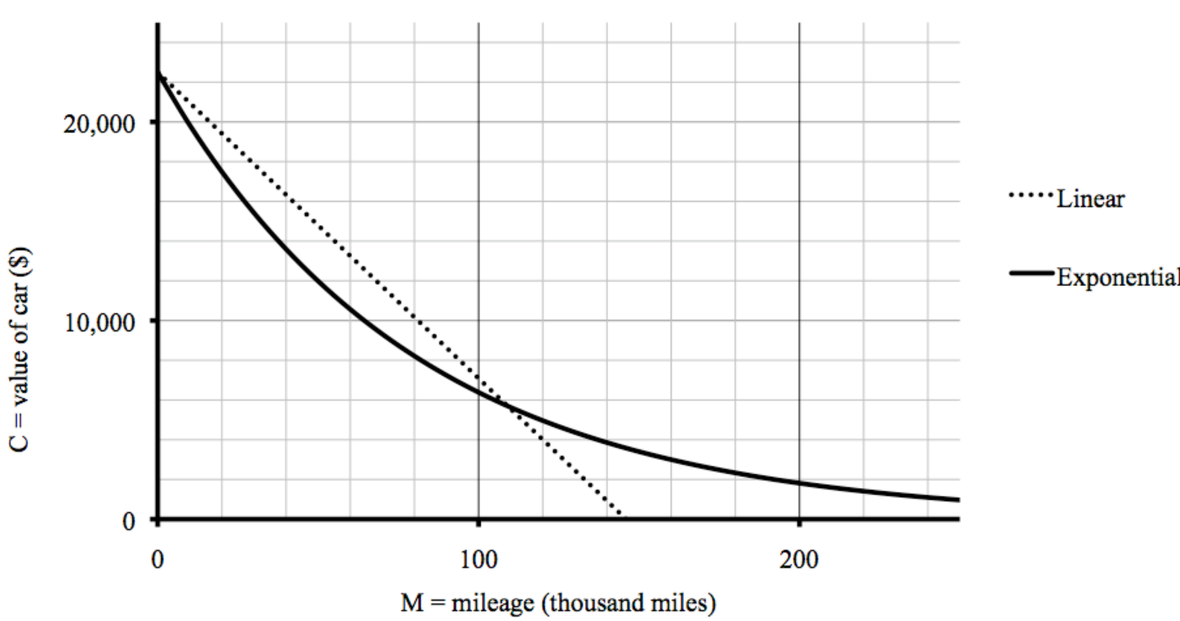
$$\textbf{linear:}\quad C = 22{,}500 - 154 * 10 = 22{,}500 - 154 \times \underline{10} = \$20{,}960$$

$$\textbf{exponential:}\quad C = 22{,}500 * .9875^{10} = 22{,}500 \times .9875 \wedge \underline{10} \approx \$19{,}841$$

The lower value, from the exponential equation, seems more reasonable.

Here are a few more values and the graph. The graph is shows both the line and exponential curve have intercept just over \$22,000, which should be \$22,500. The line and curve intersect again between 100,000 and 120,000 miles (close to the exact mileage of 109,000) at right under \$6,000 (close to the exact value of \$5,700).

M	0	10	50	80	109	200	250
C (if linear)	22,500	20,960	14,800	10,180	5,714	~~- 8,300~~	~~-16,000~~
C (if exponential)	22,500	19,841	11,996	8,225	5,711	1,818	969



There's no way of knowing whether the function is linear or exponential. It is probably not exactly either one. But if we have to pick, the exponential model seems closer to reality.