Considering Fractions

Introduction

Whether we are interested in the opioid crisis, road safety, low birthweight babies, suicide, or vaccinations, we often work with fractions to measure the phenomenon. A fraction might be a proportion, like the proportion of babies born who weigh less than 250 grams; it might be a rate, such as the number of deaths per years of exposure to an environmental hazard; or a ratio of two fractions (or two rates) so we can compare, say, the vaccine uptake of two groups.

In this discussion, we examine fractions to understand how they are calculated, and we ask questions about where the numbers come from, how they are measured, who or what is being counted and isn't being included, and whether different numbers might be more informative.
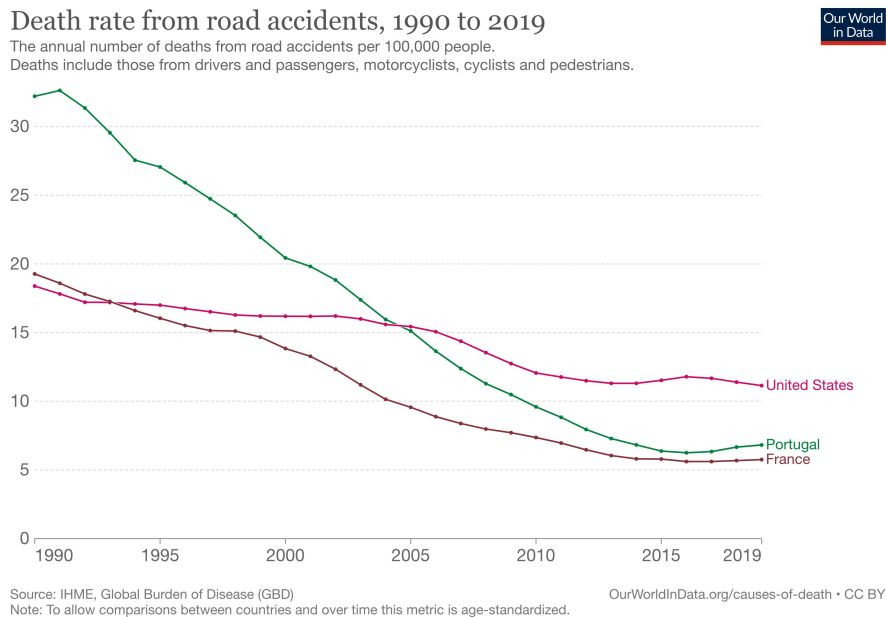

Figure 1: Screenshot from Krugman's tweet.

An Informative Fraction

Let's begin by looking at road safety. Nobel prize-winning economist Paul Krugman

tweeted a chart (see Figure 1) as he was winding up a long trip in Europe, including a visit to Portugal. The chart compared the road safety in Portugal to the U.S. and France over the past 50 years.

Let's focus on 2019 and the United States. Take a look at the scale on the y-axis to see what measurement is being plotted. It's a fraction, where the numerator is the number of deaths from road accidents in 2019 in the U.S. and the denominator is the population of the US in 100,000s.

$$\frac{\#\text{deaths from road accidents in the US in 2019}}{\#\text{ of people (100,000s) in the U.S. in 2019}}$$

1. If there are roughly 323,800,000 people in the US in 2019, how many 100,000s of people are there?

2. Why doesn't the chart simply report the number of deaths in road accidents that year? For the U.S. that would be 36,096.

3. Why not simply use a fraction, like

$$\frac{36096}{323800000} = 0.0001114762$$

Or a percentage such as:

$$100 \times \frac{36096}{323800000} = 0.01114762\%$$

Digging Deeper: Where do the numbers come from?

Let's dig a bit deeper into the road safety fraction and answer the following.

- Who is doing the measuring?
- What is being measured (be as precise as possible)?
- Who is being measured? Who might be missed?
- What might be missed by this measurement?

Answering these questions helps us figure out whether the numbers are: trustworthy; measuring what we think they are measuring; and comprehensive.

4. Answer these questions for the road safety example, and consider the numerator and denominator, separately. You may need to do some Internet research to get the answers.

5. The footnote for the figure mentions that "to allow comparisons between countries and over time, this metric is age-standardized." Age standardization means that rates are calculated separately for each age group in a country and then combined according to a specific weighting, where the weighting is the same for each country. Can you think of a scenario where the comparison between countries would be misleading without age-standardization?

6. The tweets in response to Krugman raise several questions about who is missing and whether or not it is the right metric. Below is a sample of four responses to Krugman's plot. Explain what each tweet is recommending: does it make a case for adjusting the ratio or using a different numerator or denominator?

> "Seems like deaths per 100,000 Vehicle Miles Traveled, rather than per 100k population, might be a better comparison, given the need to drive everywhere in most of the US."

> "In the 70's and 80's there were almost no highways in Portugal. In the 90's and 2000's a lot were built, and fatalities decreased significantly."

> "DUI limit is 0.05 [in Portugal] vs 0.08 in US. Makes a huge difference. Sweden is 0.02 which is effectively do not drink at all if you're driving."

> "But that is also a policy decision. Build safer forms of travel (rail, tram, segregated bike lanes) and fatalities will drop. Best metric would be deaths per km traveled across all modes not just cars."

Comparisons

We often want to disentangle a rate to see if the rate is roughly the same across different groups of people. These groups might be disaggregated according to age, income, race, sex, state, etc. In considering the road accident death rates, we might want to group the measurements by age to see if there's a case for raising the driving age or strictly enforcing tests for older drivers. Or, we might want to separate out the accidents that involved alcohol from those that did not. Below is a table from a report by the CDC that looks at fatal road accidents by age group.

This table breaks the population into 8 subgroups according to age and sex. The idea of looking at metrics for subgroups defined by two features (age and sex in this case) is called intersectionality.

	0-14		15-24		25-64		65+	
	Count	per 100k	Count	per 100k	Count	per 100k	Count	per 100k
Male	605	2.0	4,149	19.0	16,674	19.7	5,243	21.8
Female	505	1.7	1,718	8.2	5,909	6.9	2,790	9.3

7. Make three observations about these data.

8. Reflect on the tweeted critiques of the US - Portugal comparison and suggest an alternative. This may be a different rate, a feature to disentangle and compare the rates at a finer scale, or a group to include or exclude from the measurements. Justify your choice. What would you expect to see?

Your Turn

Choose a health outcome that you want to learn more about. Some ideas are: Low birthweight, Opioid Overdose, Abortion, Suicide, Flu Vaccination, HIV Mortality, Divorce.

9. Measurement: How would you best measure this health outcome as a fraction?

   (a) What is the numerator?

   (b) What is the denominator?

   (c) Find these values online and compute the fraction. Make sure your source is reliable, and include your source in your response.

10. Is there an alternative measurement of exposure that might provide a different way to assess the impact of the outcome? How does this alternative measure the phenomenon differently?

11. Disentangling: Identify subgroups that you think would make informative comparisons. Why did you choose the grouping that you chose?

Reading Assignment #1 in class portion

All reading assignments are meant to be done individually, and all responses should be in your own words. If including thoughts from the reading or from your peers in discussion make sure to cite and/or quote appropriately. See full assignment online for response length/grading rubric.

4. [In-Discussion] On page 8 of the pdf of the Introduction there is a discussion of feedback loops starting with "This is one example of a WMD feedback loop …". In pairs, read the paragraph about feedback loops and credit scores. Please describe an example of a feedback loop from you or your discussion partner's own life.

5. [In-Discussion] Quick Write: Having completed the reading and discussion please feel free to re-answer questions 1 or 2 (from pre-discussion) adding any new insights you gained during discussion or describe how the reading and discussion impacted your understanding of data science?