Data 6, Summer 2023 **Weichert and Ferrari**

Considering Fractions Solutions

Discussion 1 *Summer 2023*

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.

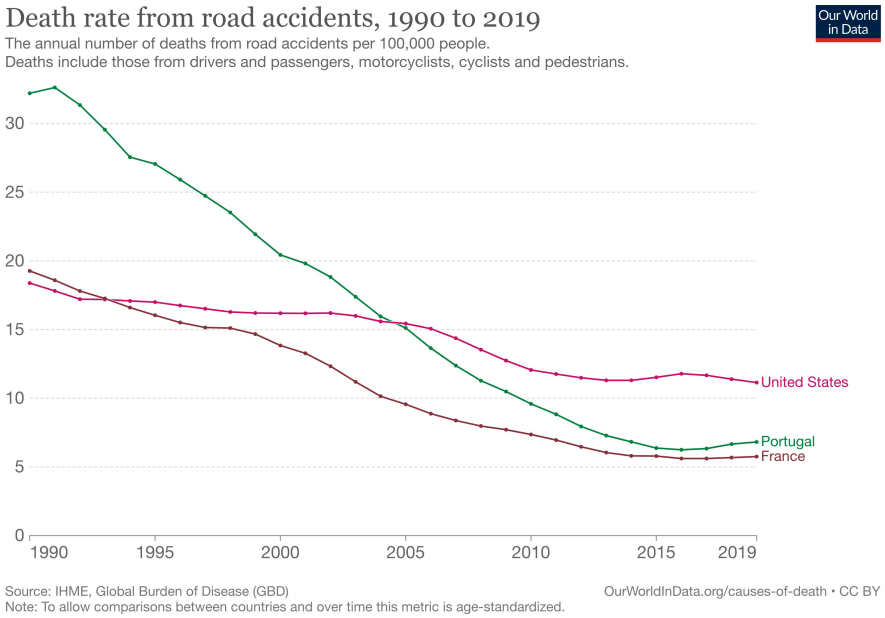


Figure 1: Screenshot from Krugman's tweet.

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?

   Solution: There are 3238 100,000s of people.

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.

   Solution: While 36,000 is a large number, we want to compare it to Portugal and France's numbers. Since the US is the third largest country in the world, and France and Portugal are much smaller with about 65 and 10 million people, respectively. Comparing the absolute number is not a fair comparison. We need to take into account the population size.

3. Why not simply use a fraction, like

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

   Or a percentage such as:

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

   Solution: The fraction is harder to make sense of when it is so tiny. The same goes for the percent. For this reason, the rate per 100,000 people is reported instead. The rate per 100,000 people is technically still a fraction, but it's more digestible because it is a number larger than one.

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.

Solution: Who is doing the measuring?
According to the caption on the plot, The data are reported by OurWorldInData.org, which in turn is provided by the Institute for Health Metrics and Evaluation; https://www.healthdata.org/2019/ and their Global Burden of Disease (GBD) database. Country profiles in the GBD are "based on over 80,000 different data sources used by researchers to produce the most scientifically rigorous estimates possible." We would need to dig deeper to find out where each country's (U.S., France, Portugal) data is coming from. Presumably the data are from each country's government.

What is being measured (be as precise as possible)?
The numerator is the number of road accident deaths in a year, where a road accident death is defined as the death of a driver, passenger, cyclist, or pedestrian.

The denominator is provided by the census.

Who is being measured? Who might be missed? An alternative is to measure the drivers only.

What might be missed by this measurement? People who drive longer distances are more exposed to traffic accidents.

Countries where there is more mass transit might be safer for traveling than those where driving is the primary mode of transportation.

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?

Solution: If say one country had a larger fraction of elderly people than another country, then the traffic accidents for the "older" country could be inflated because they had more elderly people who were more accident prone. A simple example. In Country A, half the people are over 75 and half are below, and the accident rate for those over 75 is 20 and for those under 75 it is 5 accidents per 100,000. In Country B, the accident rate for those over and under 75 is the same as in Country A, but 90% of the population is under 75. The accident rate for Country A is $0.5 \times 20 + 0.5 \times 5 = 12.5$. But in Country B it is $0.1 \times 20 + 0.9 \times 5 = 6.5$.

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?

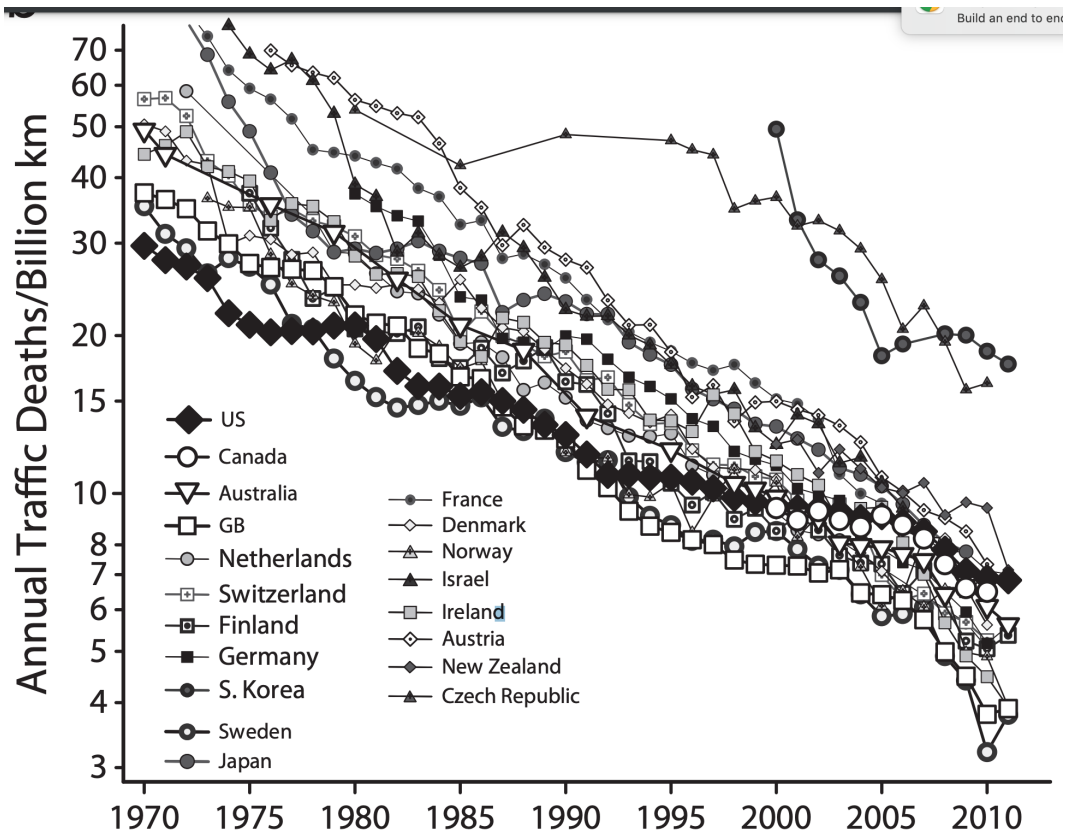
   "Seems like deaths per 100,000 Vehicle Miles Traveled, rather than per 100k population, might be a better comparison, given the need to drive everywhere in most of the US."

   "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."

Solution: The first comment refers to exposure. People who drive longer distances are more exposed to the danger of driving. The plot below looks at accidents normalized to hours on the road.



The second comment refers to the change in road systems in Portugal over 50 years. Making historical comparisons can be problematic when many things are changing. In this case, road quality, added safety features in cars, and speed limits might all reduce traffic accidents.

The third comment refers to the confounding factor of regulations on driving under the influence, which differs by country. This might be used to support lowering the allowable level of blood alcohol to make roads safer.

The last comment pertains to the focus only on automobile accidents. A country can impact "travel" safety by offering more opportunities for mass transit.

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.

Solution: In every age category the accident rate for females is lower than for males. This may in part be due to females driving fewer miles (lower exposure).

The accident rate for those over 65 is the highest among all age groups.

The accident rate for those under 15 must be for children who are passengers or pedestrians because the driving age is typically at least 16.

The greatest number of accidents occur in the 25-64 age group. This is undoubtedly because there are more people in this age group.

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?

Solution: Here are three alternatives:

- Change the ratio to deaths per million miles traveled in a car. This would take into account the exposure to the dangers of driving. The graph in the previous solution shows what might happen.
- Change the ratio to deaths per million miles traveled by any form of ground transportation. This would take into account the alternative, potentially safer modes of travel. I would expect that France would be even safer. I don't know enough about transportation in Portugal to venture a guess.
- If interested in the impact of driving under the influence, a different fraction might be to look at the number of deaths where the driver had a blood alcohol level exceeding Sweden's level of 0.02. I would expect to see the US to be much larger than Portugal, which in turn, is larger than Sweden's rate.

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?

Solution: We chose to look at suicides in the United States. According to the Centers for Disease Control and Prevention (CDC) WISQARS Leading Causes of Death Reports, in 2019: suicide was the tenth leading cause of death overall in the United States, claiming the lives of over 47,511 people.

We can create a rate per 100,000 for this number: in 2019, there were 14.7 suicides per 100,000 people in the U.S.:

$$100000 \times \frac{47511}{323800000} = 14.7$$

The data for the numerator is from The denominator is provided by the US census.

Disaggregating by race shows Native American and White populations have markedly higher suicide rates. Disaggregating by age and sex shows that males are more likely to commit suicide and two age groups — teen and fifty-something — have the highest rates.

Why might we want to compare suicide rates among races using an age-adjusted rate? What if we want to focus on one race in particular; how might adjusting for age be problematic?

What might we do with the insights gleaned from this disaggregated data?