

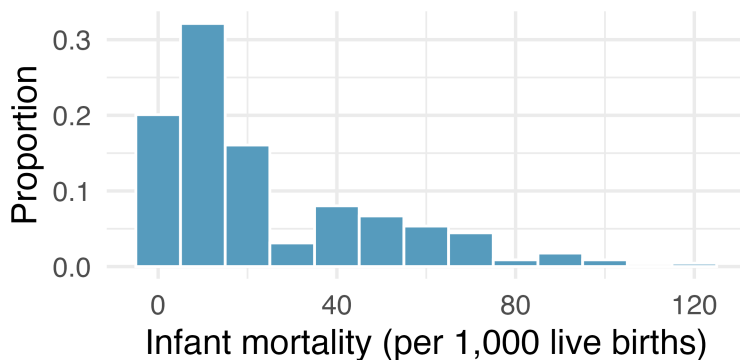
# Numerical data (part 2)

## Practice problems

9/18/24

Please work on the practice problems in your group. Problems with an asterisk \* will be assigned to the weekly problem set.

1. The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of health in a country. The histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014. In particular, this is a *relative frequency* histogram, which shows proportions instead of raw counts on the y-axis:



- a. Estimate  $Q_1$ , the median  $m$ , and  $Q_3$  from the histogram.
  - b. Would you expect the mean of this dataset to be smaller or larger than the median? Explain your reasoning.
2. Suppose that an exam has a total of 100 possible points, and the average score was an 85 with standard deviation of 15. Is the distribution of the scores on this exam symmetric? If not, what shape would you expect this distribution to have? Explain your reasoning.

3. (\*) The statistic  $\frac{\bar{x}}{m}$  can be used as a measure of skewness. Suppose we have a distribution where all observations are greater than 0 (i.e.  $x_i > 0$  for all observations  $i = 1, \dots, n$ ). What is the expected shape of the distribution under the following conditions? Explain your reasoning.
- a.  $\frac{\bar{x}}{m} = 1$
  - b.  $\frac{\bar{x}}{m} > 1$
  - c.  $\frac{\bar{x}}{m} < 1$
4. Returning to the sunflowers:
- a. Calculate the median number of healthy sunflowers in a grid cell for each of the sampling strategies. How do they compare to the sample means?
  - b. Pick one of the sampling schemes. Draw out the boxplot for the data you obtained under that sampling scheme.