

# Warm-Up 8

Prof. Jordan C. Hanson

July 19, 2020

## 1 Formula Area

- $N$ : Usually  $N$  refers to the size of a data set.
- *Frequency*: The number of times a particular value occurs in a data set.
- *Relative frequency*: The frequency divided by  $N$ .
- *Qualitative or categorical data*: Data properties that are not countable, *e.g.* hottest chile species.
- *Quantitative discrete data*: Data properties that are countable but not continuous (rational) numbers.
- *Quantitative continuous data*: Data properties that are countable and continuous (rational).
- A *histogram* is a bar-chart of frequencies or relative frequencies on the y-axis, with the categories or bins on the x-axis.
- The mean  $\bar{x}$  of a sample is

$$\bar{x} = N^{-1} \sum_{i=1}^N x_i = \sum_{i=1}^M f_{r,i} x_i \quad (1)$$

where  $N$  is the size of the data set,  $M$  is the number of bins, and  $f_{r,i}$  are the relative frequencies.

- *Median*: A value that separates half of the data set above it, and half below it.
- *Quartiles*:  $Q1$ ,  $Q2$ , and  $Q3$  are values that separate the data into four equal frequencies.
- The standard deviation  $s$  of a sample is given by

$$s^2 = (N - 1)^{-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \bar{x^2} - \bar{x}^2 \quad (2)$$

where  $N$  is the size of the data set, and  $M$  is the number of bins.

## 2 Unit 0 Review and Warm-up

1. What type of data is **number of hours of sleep per night**?
2. Suppose a sample of hours per night of sleep is 4, 4, 5, 5, 6, 7, 8, 8 for a person. Answer the following:
  - (a) What is the mean and standard deviation of the data set? How many standard deviations below the mean is the first data point?
  - (b) What is the median or  $Q2$  of the data set?
  - (c) What are the quartiles of the data set?
  - (d) Draw a time-series of the data set.