Economics 250
Grouped Data

Sometimes data are grouped into classes, perhaps to preserve the anonymity of respondents to a survey. For example, we might be told that 10% of households have annual consumption spending of between \$30,000 and \$50,000, and so on for other ranges.

How can we calculate statistics with such data? Imagine that there are K classes, indexed by a label i. The midpoints of the classes are m_i and the *counts* (numbers) of observations are c_i.

The sample mean is:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{K} c_i m_i.$$

Notice that the total number of observations is:

$$n = \sum_{i=1}^{K} c_i.$$

We can write the *proportion* of observations in class i by

$$w_i = \frac{c_i}{n},$$

so that the mean is:

$$\overline{x} = \sum_{i=1}^{K} w_i m_i.$$

Finally, the sample variance is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{K} c_i (m_i - \overline{x})^2.$$