

**Social Studies 201****January 28, 2005****Measures of Variation****Overview**

Measures of variation (range, interquartile range, standard deviation, variance, and coefficient of relative variation) are presented in the notes in this module. These measures are less commonly used than are averages, and are not as familiar as are measures of centrality (mode, median, or mean). Measures of variation indicate the extent of dispersion or variation of a distribution.

In these notes, each measure is defined and examples are presented. The following quick examples demonstrate ways that measures of variation form an essential aspect of social science analysis.

**Quick examples**

- The range of grades in Social Studies 201 last semester was 42 percentage points, from a minimum of 50% to a maximum of 92%.
- Literacy skills vary both across and within provinces. A Statistics Canada study demonstrated that while “a student’s socio-economic background is a key factor, it accounted for less than half the variation in provincial literacy scores.”

From <http://www.statcan.ca/Daily/English/040714/d040714b.htm>

- Class grades vary more in an introductory than in upper level classes – the standard deviation of grades in a 100-level class was 8 percentage points but was only 3 percentage points in a 300-level class.

- “Most countries today are culturally diverse. ... In very few countries can the citizens be said to share the same language, or belong to the same ethnonational group.”

Will Kymlicka, *Multicultural Citizenship*, Oxford, Clarendon Press, 1995, p. 1.

- “Income inequality among families remained stable.”

From <http://www.statcan.ca/Daily/English/040520/d040520b.htm>

- “Between 1981 and 1996, the second, third, and fourth quintiles lost 2.8 per cent of their before-tax income, a total of \$14 billion, to the upper quintile .... These figures support the idea that there is increasing polarization of income in Canada.”

From L. Tepperman and J. Curtis, *Sociology: A Canadian Perspective*, Don Mills, Oxford University Press, 2004, p. 371.

Each of these quick examples refers to how members of a sample or population differ from each other – variation in grades, literacy skills, culture, and income. While the measures of centrality for each of these variables describes the average or typical member, just as important is how members differ from each other. It is this variation that forms the topic for this module.

## Topics

- Introduction
- Range
- Interquartile range
- Variance and standard deviation
  - Ungrouped data
  - Grouped data
- Coefficient of relative variation
- Conclusion

## Readings

Chapter 5, Sections 5.7 - 5.11

### Introduction

Measures of variation are statistics describing the extent of variation among members of a sample or population. These measures are not as well understood or as commonly used as are measures of centrality (mode, median, and mean). But measures of variation are just as important as are averages to understanding distributions and populations. Not everyone in a population is at the average, and measures of variation describe how members of a population differ.

Measures of variation are single numbers, or statistics, that summarize how varied or how dispersed members of a population are. The measures of variation that we discuss in this class are as follows. Formal definitions and examples of each are contained in later parts of Module 4.

- **Range.** The range is the maximum value of a variable minus its minimum value.
- **Interquartile range.** The interquartile range is the value of the 75th percentile minus the value of the 25th percentile. This is the range for the “middle one-half” of a distribution.
- **Standard deviation.** The standard deviation is a measure of the “average” or “standard” amount by which individual cases differ from the mean of a distribution. The exact form of this average, or standard, is provided later in these notes.
- **Variance.** The variance is the square of the standard deviation. Alternatively stated, the variance is an “average” of squares of differences of individual values from the mean. This will become apparent when you study the formulae and examples later in this module.
- **Coefficient of relative variation.** The coefficient of relative variation is the value of the standard deviation divided by the mean. This measure of variation provides an indication of how much variation there is in a distribution, relative to the mean of the distribution.

Most of these measures of variation are not used in ordinary conversations or in newspapers and magazines – or they are much less commonly used than averages. One key to understanding measures of variation is to use them to compare two distributions. In the examples and exercises in this module, two distributions are often compared. When you work through these examples and exercises, carefully examine the two distributions and the associated measures of variation. This will help you develop an understanding of how measures of variation can be used to compare variability.

Consider the hypothetical distributions A and B in Figures 1 and 2. Measures of variation for distribution A have smaller values than the corresponding measures of variation for distribution B. These indicate that distribution A is less varied and distribution B is more varied. Each distribution is symmetrical and has the same mean, at the centre of the distribution. But distribution B in Figure 2 is twice as spread out as distribution A in Figure 1. Most measures of variation for the distribution in Figure 2 will be double the size of the corresponding measure for the distribution in Figure 1.

Table 1: Variation of distributions of A and B – Figures 4.1 and 4.2

Figure 1	Figure 2
less varied	more varied
less dispersed	more dispersed
more concentrated	less concentrated

Differences between variation of distributions A and B are summarized in Table 1. Each of the words **variation**, **dispersion**, and **concentration** can be used to contrast differences in variation, or spread of values, in the two distributions.

Figure 1: Less varied distribution A

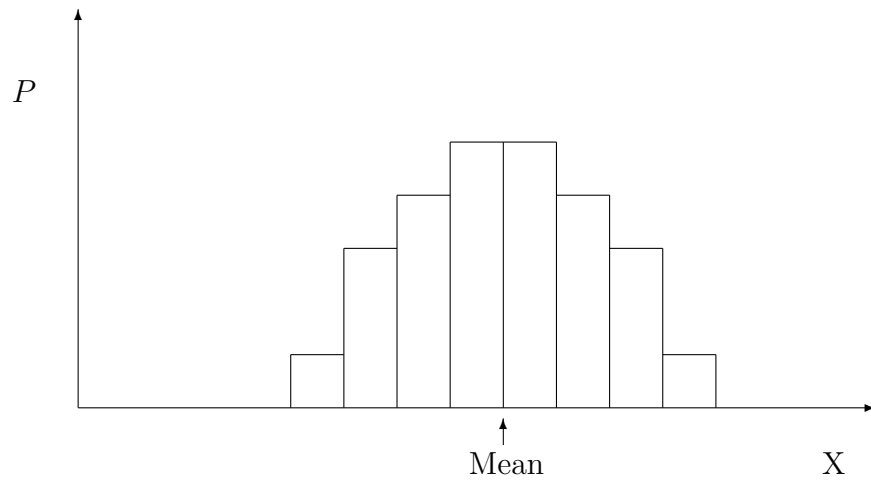
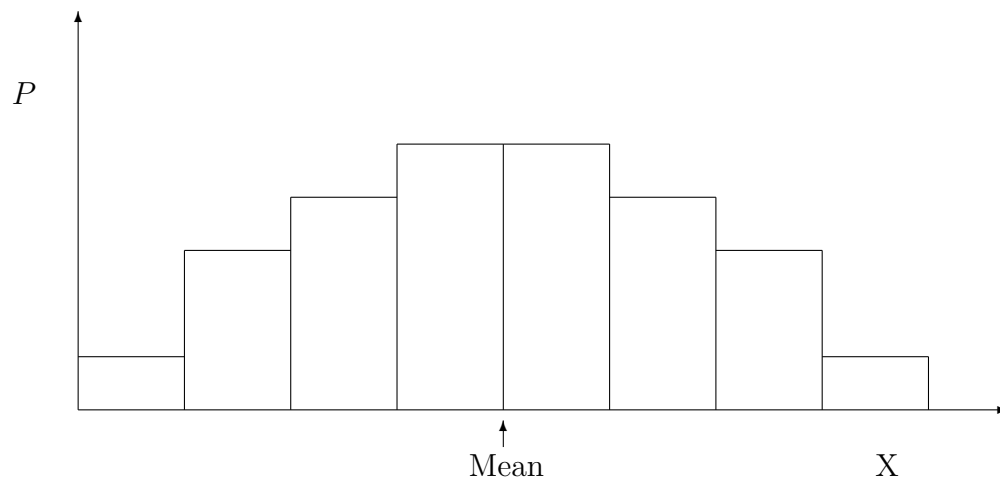


Figure 2: More varied distribution B



## Range

See text, section 5.8.1, pp. 216-218.

“Provinces clearly vary in their reading performance, ranging from 501 in New Brunswick to 550 in Alberta.”

From J. Douglas Willms, *Variation in Literacy Skills Among Canadian Provinces: Finding from the OECD PISA*, Statistics Canada catalogue number 81-595-MIE, No. 0012, Ottawa, 2004, p. 14. Available at the Statistics Canada web site:  
<http://www.statcan.ca/Daily/English/040714/d040714b.htm>

The first measure of variation is the range, the set of values over which the variable varies.

**Definition.** The range is the maximum value of the variable minus the minimum value of the variable.

$$\text{Range} = \text{maximum} - \text{minimum}.$$

That is, the range is the largest minus the smallest value of the cases in a data set.

Alternatively, the range is sometimes defined as a listing of the minimum and maximum values.

Since the range requires determining the smallest and largest values, the a variable must have at least an ordinal level of measurement in order to determine the range.

The unit of measure for the range is the same as the unit of measure for the variable  $X$ . For example, if  $X$  represents income of individuals in dollars, and the minimum and maximum incomes are measured in dollars, then the range is measured in dollars.

**Example 4.1 – Range of grades.** A student takes five classes during her first semester at the University of Regina. The grades obtained are 64, 74, 68, 79, and 85.

What is the range of grades of the student?

**Answer.** The variable is grade, presumably in the unit of one percentage point, so this variable has at least an interval level of measurement and the range can be meaningfully determined.

The lowest grade was 64 and the highest was 85, so the range is 21.

$$\text{Range} = \text{maximum} - \text{minimum} = 85 - 64 = 21$$

The range of grades for this student is 21 percentage points.

**Example 4.2 – Range of ages.** For a sample of eleven University of Regina undergraduates, the ages are 27, 19, 18, 18, 18, 29, 20, 36, 24, 18, and 19 years.

What is the range of ages for these eleven students?

**Answer.** The variable is age in years, a variable with a ratio scale of measurement so the range can be meaningfully determined.

The lowest age is 18 and the highest age is 36, so the range is 18.

$$\text{Range} = \text{maximum} - \text{minimum} = 36 - 18 = 18$$

The range of ages for this sample of students is 18 years.

**Example 4.3 – Attitudes to treating gay and lesbian couples as married.** A sample of just under seven hundred undergraduate students was asked to state their view about the statement, “Tax laws and job benefits should recognize gay and lesbian couples as married.” Respondents stated their degree of agreement or disagreement with this statement on a five-point scale, from 1, indicating strongly disagree, to 5, indicating strongly agree. The distribution of responses is reported in Table 2.

What is the range of views on this topic?

**Answer.** For this statement, responses are measured on a scale from 1 (strongly disagree) to 5 (strongly agree). This measurement is at the ordinal level, so the range is meaningful.

The range of opinions is  $5 - 1 = 4$ . Alternatively stated, in words, the range is from 1 to 5, or from strongly disagree to strongly agree.

Table 2: Frequency, percentage, and cumulative percentage distributions of views of undergraduate student responses to treating gay and lesbian couples as married

Response ( $X$ )	Frequency	%	Cumulative %
1 (strongly disagree)	141	20.3	20.3
2 (somewhat disagree)	86	12.4	32.7
3 (neutral)	183	26.3	59.0
4 (somewhat agree)	161	23.2	82.2
5 (strongly agree)	124	17.8	100.0
Total	695	100.0	

**Note.** The distribution of cases between these lower and upper limits has no effect on the determination of the range. All that matters for the range is what the smallest and largest values are.

### Recap – some characteristics of the range

The range provides a useful first indication of the dispersion or concentration of values in a distribution. At the same time, as demonstrated by Example 4.3, the range is limited in that it does not indicate how varied the cases are between the minimum and maximum values.

Imagine two rooms with ten people in each room – in room A the ages of the people have a range from 18 to 22, a range of 4 years; in room B the ages of the people have a range from 15 to 55, a range of 40 years. A glance around each room will quickly indicate that the people in room A have a similar set of ages (more concentrated or less dispersed), while the people in room B are more varied in their ages (less concentrated or more dispersed).

One of the first questions a researcher asks about a variable is what is the range of the variable across the data set. This does not tell the researcher a great deal about the values of the variable, but provides a useful first indication of the spread or variation of the values. The researcher is likely to follow up by calculating some of the other measures of variation in the following notes.



Social Studies 201 – January 28, 2005. Introduction to measures of variation9

Last edited January 28, 2005.