

**Social Studies 201**  
**Answers to Problem Set 3**

1. (a) The calculations for determining the mean and standard deviation of the percentage support for each of the Conservative, Liberal, and New Democratic parties are contained in Tables 1 to 3. There are two formulas that can be used to obtain the variance and standard deviation. For the Conservative and Liberal party, I use the differences about the mean formula; for the NDP, I use the alternative formula with the square of the  $X$  values. Either formula can be used for any of these, and the results should be identical.

**Table 1: Calculations for mean and standard deviation, Percentage support for Conservative Party**

$X$	$X - \bar{X}$	$(X - \bar{X})^2$
19	7.3	53.29
15	3.3	10.89
13	1.3	1.69
10	-1.7	2.89
8	-3.7	13.69
9	-2.7	7.29
8	-3.7	13.69
82	0.1	103.43

There are seven values here so  $n = 7$ . Mean percentage support for the Conservative party is  $\Sigma X/n = 82/7 = 11.714$  or 11.7. The variance is

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{103.43}{7 - 1} = \frac{103.43}{6} = 17.238.$$

The standard deviation is the square root of the variance, that is,

$$s = \sqrt{s^2} = \sqrt{17.238} = 4.152$$

or 4.2 percentage points. The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{4.152}{11.714} \times 100 = 0.354 \times 100 = 35.4.$$

Table 2: Calculations for mean and standard deviation, Percentage support for Liberal Party

$X$	$X - \bar{X}$	$(X - \bar{X})^2$
38	-8	64
49	3	9
49	3	9
45	-1	1
45	-1	1
44	-2	4
52	6	36
322	0	124

Mean percentage support for the Liberal party is  $\Sigma X/n = 322/7 = 46$ . The variance is

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{124}{7 - 1} = \frac{124}{6} = 20.667.$$

The standard deviation is the square root of the variance, that is,

$$s = \sqrt{s^2} = \sqrt{20.667} = 4.546$$

or 4.5 percentage points. The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{4.546}{46} \times 100 = 0.0988 \times 100 = 9.9.$$

For the NDP,

$$\bar{X} = \frac{\Sigma X}{n} = \frac{71}{7} = 10.143$$

Table 3: Calculations for mean and standard deviation, Percentage support for NDP – using alternative formula

$X$	$X^2$
11	121
11	121
12	144
11	121
9	81
9	81
8	64
71	733

or 10.1 per cent. The variance is

$$\begin{aligned}
 s^2 &= \frac{1}{n-1} \left( \sum X^2 - \frac{(\sum X)^2}{n} \right) \\
 &= \frac{1}{7-1} \left( 733 - \frac{71^2}{7} \right) \\
 &= \frac{1}{7-1} \left( 733 - \frac{5041}{7} \right) \\
 &= \frac{733 - 720.143}{6} \\
 &= \frac{12.857}{6} \\
 &= 2.143
 \end{aligned}$$

and the standard deviation is

$$s = \sqrt{s^2} = \sqrt{2.143} = 1.464$$

or 1.5 percentage points.

The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{1.464}{10.143} \times 100 = 0.1443 \times 100 = 14.4.$$

Table 4: Summary Statistics for percentage support for three political parties

Measure	Conservative	Liberal	NDP
$\bar{X}$	11.7	46.0	10.1
$s^2$	17.2	20.7	2.1
$s$	4.2	4.5	1.5
CRV	35.4	9.9	14.4

- (b) In order to compare the variation in percentage support, the above measures are summarized in Table 4. From the standard deviations, the Liberal party had the most variable level of support (4.5), followed by the Conservative Party (4.2), and then the NDP (1.5). However, the variability in percentage support is very similar for the Conservative and Liberal party, with variability for the NDP less than one-half that of either of the two other parties.

The coefficient of relative variation gives a different picture, with the Conservative Party most variable, relative to mean support (CRV of 35.4). This is followed by the NDP (CRV of 14.4), and Liberals have the lowest relative variability (9.9). The reason for this difference in ranking is that the Liberal Party had a much higher level of support than the Conservative Party, with a mean about four times as large. As a result, since there are similar standard deviations for the Liberals and Conservatives, relative to the lower mean the Conservative Party is much more varied than are the Liberals to their higher mean support. While the standard deviation for the NDP is lowest of the three parties, the low mean level of NDP support leads to a larger relative variation for the NDP than for the Liberals.

While the analyst is incorrect, there is no single answer to which party has greatest variability – in absolute terms Liberal support is most variable but in relative terms Conservative support is most variable.

2. Tables 5 and 6, and the calculations following, provide the standard deviations.

Table 5: Calculations for standard deviation of opinion for Edmonton

Opinion about same-sex marriage	$X$	$f$	$fX$	$fX^2$
Strongly support	1	56	56	56
Somewhat support	2	56	112	224
Somewhat oppose	3	38	114	342
Srongly oppose	4	117	468	1,872
Total		267	750	2,494

For Edmonton, the mean opinion is

$$\bar{X} = \frac{\Sigma fX}{n} = \frac{750}{267} = 2.809$$

or 2.8. The variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left( \Sigma fX^2 - \frac{(\Sigma fX)^2}{n} \right) \\ &= \frac{1}{267-1} \left( 2,494 - \frac{750^2}{267} \right) \\ &= \frac{1}{266} \left( 2,494 - \frac{562,500}{267} \right) \\ &= \frac{2,494 - 2,106.742}{266} \\ &= \frac{387.258}{266} \\ &= 1.456 \end{aligned}$$

and the standard deviation is  $s = \sqrt{s^2} = \sqrt{1.456} = 1.207$  or 1.2.

For Southern Alberta, the mean opinion is

$$\bar{X} = \frac{\Sigma fX}{n} = \frac{263}{85} = 3.094$$

Table 6: Calculations for standard deviation of opinion for Southern Alberta

Opinion about same-sex marriage	$X$	$f$	$fX$	$fX^2$
Strongly support	1	10	10	10
Somewhat support	2	18	36	72
Somewhat oppose	3	11	33	99
Strongly oppose	4	46	184	736
Total		85	263	917

or 3.1. The variance is

$$\begin{aligned}
 s^2 &= \frac{1}{n-1} \left( \sum fX^2 - \frac{(\sum fX)^2}{n} \right) \\
 &= \frac{1}{85-1} \left( 917 - \frac{263^2}{85} \right) \\
 &= \frac{1}{84} \left( 917 - \frac{69,169}{85} \right) \\
 &= \frac{917 - 813.753}{84} \\
 &= \frac{103.247}{84} \\
 &= 1.229
 \end{aligned}$$

and the standard deviation is  $s = \sqrt{s^2} = \sqrt{1.229} = 1.109$  or 1.1.

The standard deviation for the two distributions is very similar, with slightly greater variation for Edmonton ( $s = 1.2$ ) than for Southern Alberta ( $s = 1.1$ ). Opinions for Southern Alberta residents tends toward strong opposition to same-sex marriage, with over one-half of respondents indicating strongly oppose. While Edmonton residents also tend toward opposition, there is somewhat less concentration of opinions at strongly oppose, with more respondents spread across the other categories.

Relative variation, as measured by the CRV, gives the same picture, with the Edmonton CRV being 43.0 and the Southern Alberta CRV being 35.8.

- Tables 7 and 8 and the calculations following provide the means, standard deviations, and coefficient of variation.

Table 7: Calculations for mean and standard deviation of education for connected

Years of school completed	$X$	$P$	$PX$	$PX^2$
None through seven	3.5	0.7	2.45	8.575
Eight	8	0.7	5.60	44.800
Nine through eleven	10	16.3	163.00	1,630.000
Twelve	12	82.3	987.60	11,851.200
Total		100.0	1,158.65	13,534.575

For those connected to the internet, the mean years of schooling is

$$\bar{X} = \frac{\Sigma PX}{100} = \frac{1,158.65}{100} = 11.5865$$

or 11.6 years of schooling. The variance is

$$\begin{aligned} s^2 &= \frac{1}{100} \left( \Sigma PX^2 - \frac{(\Sigma PX)^2}{100} \right) \\ &= \frac{1}{100} \left( 13,534.575 - \frac{1158.65^2}{100} \right) \\ &= \frac{1}{100} \left( 13,534.575 - \frac{1,342,469.823}{100} \right) \\ &= \frac{13,534.575 - 13,424.698}{100} \\ &= \frac{109.877}{100} \\ &= 1.099 \end{aligned}$$

and the standard deviation is  $s = \sqrt{s^2} = \sqrt{1.099} = 1.048$  or 1.0 years.

The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{1.048}{11.586} \times 100 = 0.0905 \times 100 = 9.05.$$

Table 8: Calculations for mean and standard deviation of education for not connected

Years of school completed	$X$	$P$	$PX$	$PX^2$
None through seven	3.5	5.5	19.25	67.375
Eight	8	11.2	89.60	716.800
Nine through eleven	10	30.0	300.00	3,000.000
Twelve	12	53.3	639.60	7,675.200
Total		100.0	1,048.45	11,459.375

For those connected to the internet, the mean years of schooling is

$$\bar{X} = \frac{\Sigma PX}{100} = \frac{1,048.45}{100} = 10.4845$$

or 10.5 years of schooling. The variance is

$$\begin{aligned} s^2 &= \frac{1}{100} \left( \Sigma PX^2 - \frac{(\Sigma PX)^2}{100} \right) \\ &= \frac{1}{100} \left( 11,459.375 - \frac{1,048.45^2}{100} \right) \\ &= \frac{1}{100} \left( 11,459.375 - \frac{1,099,247.403}{100} \right) \\ &= \frac{11,459.375 - 10,992.474}{100} \\ &= \frac{466.901}{100} \\ &= 4.669 \end{aligned}$$



Table 9: Summary Statistics for years of primary and secondary education of connected and not connected households

Measure	Connected	Not connected
$\bar{X}$	11.6	10.5
$s^2$	1.10	4.67
$s$	1.05	2.16
CRV	9.05	20.61

and the standard deviation is  $s = \sqrt{s^2} = \sqrt{4.669} = 2.161$  or 2.2 years.

The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{2.161}{10.484} \times 100 = 0.2061 \times 100 = 20.61.$$

The summary statistics are presented in Table 9. The two distributions are similar in that over one-half of all respondents report having twelve years of primary or secondary schooling completed. But those not connected to the internet tend to have somewhat less years of school completed – there are more of these respondents at each of the lower levels of schooling than in the case of the connected households. As a result, the mean years of schooling would be expected to be lower for the not connected than the connected households. The higher concentration of households at the twelve years of schooling level for the connected households leads to one expecting a smaller variation for the connected than the not connected households.

The latter is what the data in Table 9 show. The mean years of school completed is approximately one year less for the not connected households, as compared with the connected. The standard deviation for the connected households (1.05) is less than one-half that for the not connected households (2.16). Given similar mean years of schooling for each of the two groups, the CRV for the connected group is also approximately one-half the CRV for the not connected.