Social Studies 201
Answers to Problem Set 3
March 5, 2004

1. Tables 1 and 2 and the calculations following provide the means (required for the CRV, even though not directly requested in this question), variances, standard deviations, and coefficients of relative variation.

   Table 1: Calculations for measures of variation for smoking

Rate of smoking	X	P	PX	PX^2
Did not smoke	0	76	0.0	0.00
1-2	1.5	9	13.5	20.25
3-9	6	8	48.0	288.00
10-24	17	6	102.0	1,734.00
25+	30	1	30.0	900.00
Total		100	193.5	2,942.25

   For smoking, the variance is

$$\begin{aligned} s^2 &= \frac{1}{100}\left(\Sigma PX^2 - \frac{(\Sigma PX)^2}{100}\right) \\ &= \frac{1}{100}\left(2,942.25 - \frac{193.5^2}{100}\right) \\ &= \frac{1}{100}\left(2,942.25 - \frac{37,442.25}{100}\right) \\ &= \frac{2,942.25 - 374.4225}{100} \\ &= \frac{2,567.8275}{100} \\ &= 25.678275 \end{aligned}$$

   and the standard deviation is $s = \sqrt{s^2} = \sqrt{25.67825} = 5.0674$ or 5.1 cigarettes per day.

For computing the CRV, it is necessary to calculate the mean, and this is

$$\bar{X} = \frac{\Sigma PX}{100} = \frac{193.5}{100} = 1.935$$

or 1.9 cigarettes per day.

The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{5.0674}{1.935} \times 100 = 2.619 \times 100 = 261.9$$

Table 2: Calculations for measures of variation of drinking

Rate of drinking	X	P	PX	PX^2
Did not drink	0	23	0	0
Only special events	1	17	17	17
1-3	2	29	58	116
4-6	5	17	85	425
7+	10	14	140	1,400
Total		100	300	1,958

For drinking, the variance is

$$\begin{aligned} s^2 &= \frac{1}{100}\left(\Sigma PX^2 - \frac{(\Sigma PX)^2}{100}\right) \\ &= \frac{1}{100}\left(1,958 - \frac{300^2}{100}\right) \\ &= \frac{1}{100}\left(1,958 - \frac{90,000}{100}\right) \\ &= \frac{1,958 - 900}{100} \\ &= \frac{1,058}{100} \\ &= 10.58 \end{aligned}$$

and the standard deviation is $s = \sqrt{s^2} = \sqrt{10.58} = 3.2527$ or 3.3 time per week.

Table 3: Summary statistics for variation of smoking and drinking

Measure	Smoking	Drinking
$\bar{X}$	1.94	3.00
s^2	25.68	10.58
s	5.07	3.25
CRV	261.9	108.42

For computing the CRV, it is necessary to calculate the mean, and this is

$$\bar{X} = \frac{\Sigma PX}{100} = \frac{300}{100} = 3.00$$

or 3.0 times per week.

The coefficient of relative variation (CRV) is

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{3.2527}{3.0} \times 100 = 1.0842 \times 100 = 108.42$$

The statistics are summarized in Table 3. Using the measures of variation, the rate of smoking is a more varied distribution than is the rate of drinking. Note though that the variance and standard deviation for drinking are number of times, rather than actual consumption, as in the case of smoking. Also, drinking is measured on a monthly basis while smoking is measured on a daily basis. From these considerations, the different values of the standard deviation and variance may not be good measures of variation for purposes of comparison of these two distributions.

The CRV provides a measure that is independent of units, so a comparison between distributions is more legitimate. The CRV for smoking is over double that for drinking so, on this basis, smoking is more variable than is drinking.

The one problem with the CRV in this case is that the mean for smoking is very low (only 1.9 cigarettes per day), so when calculating the CRV, the standard deviation is divided by a very small mean, thus

elevating the value of the CRV. Looking at the frequency distributions makes it appear as if smoking is less varied, since over three-quarters of respondents do not smoke at all. For drinking, there appears to be more variation. In this case, these measures of variation may not give a reliable guide concerning the relative variation for the two distributions.

2. Variation and patterns

The tables from T:\Students\PUBLIC\201\ssae98.sav and the calculations and comments follow.

a. *Descriptives*

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
DEBT1 Debt before Fall 1998	606	0	40000	3134.95	6635.048
PAY Hourly pay in dollars	384	2.00	45.00	9.7200	5.42857
INC income in thousands of dollars	617	15	250	65.46	43.888
Valid N (listwise)	328				

The range is the maximum minus the minimum value. The coefficient of relative variation (CRV) is

CRV = (standard deviation / mean) x 100

The measures of variation are summarized in the following table.

Variable	Range	S	CRV
Debt1 (dollars)	40,000	6,635	212
Pay (dollars/hour)	43	5.43	55.8
Income (thousands of dollars annually)	235	43.9	67.0

Pay has the smallest variation of the three variables, whether measured in the actual units (range and standard deviation) or as CRV. This is at least partly because the unit for pay is dollars per hour, as opposed to the larger units of dollars per year (income) or total dollars of debt. It may also be that pay levels are fairly similar for students, since they work in a relatively narrow set of types of jobs, at least for the most part. If incomes are converted to dollars, income has the largest variation, $235,000 for the range and $43,900 for standard deviation. In relative terms though, debt is more variable than income. While incomes of the households or families of students are quite varied, in relative terms it appears that students are even more varied by level of debt. It may be that debt is quite high for some students and very low or zero for many others.

b. *Means* procedure

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
V5 Increase Corporate Taxes * FUTURE Economic Future?	694	98.2%	13	1.8%	707	100.0%
GPAR grade point average * FUTURE Economic Future?	569	80.5%	138	19.5%	707	100.0%
INC income in thousands of dollars * FUTURE Economic Future?	611	86.4%	96	13.6%	707	100.0%

Report

FUTURE Economic Future?		V5 Increase Corporate Taxes	GPAR grade point average	INC income in thousands of dollars
1 Better Off	Mean	3.68	74.33	57.96
	N	297	249	272
	Std. Deviation	1.109	7.025	37.268
2 About Same	Mean	3.79	73.59	69.14
	N	254	201	216
	Std. Deviation	.978	6.986	42.605
3 Worse Off	Mean	4.06	73.80	74.07
	N	143	119	123
	Std. Deviation	.977	7.369	53.597
Total	Mean	3.80	73.95	65.16
	N	694	569	611
	Std. Deviation	1.044	7.080	43.314

Comments

For grade point average, there appears to be no particular pattern and the mean grades are fairly similar for each group. That is, the grade point average is largest for those who think they will be better off (74.3), next highest for those who think they will be worse off (73.80), and in between for those who think they will be about at the same economic level in the future (73.6). But these differences in grades do not seem very large, so there is little to say about this. The standard deviations are also fairly similar, each around 7 grade points.

For V5, attitude about corporate taxes, the mean is lowest (3.68) for those who think they will be better off, then greater (3.79) for those who think they will be about the same, and

highest (4.06) for those who think they will be worse off. A larger mean denotes greater agreement, and there is a clear pattern of greater agreement with increasing corporate taxes for those who think they will be not as well off. Standard deviations are greatest for those who think they will be better off, but again the standard deviation does not differ greatly across the three groups.

Finally, the pattern of different means and standard deviations is clearest for income. Mean income is over ten thousand dollars lower for those who think they will be better off, as compared with those who think they will be about the same. And moving from the latter group to those who think they will be worse off, there is again a big jump in mean income. It appears that those who come from higher income backgrounds think they will be worse off in future while those who come from lower income backgrounds are more optimistic. The same pattern of relative values observed for the means is also apparent for the standard deviations, those reporting the highest mean income also have the largest variation in incomes. For those who think they will be better off, both the mean and standard deviation of the three groups is smallest.

3. Tables for calculating probabilities

V5 Increase Corporate Taxes * FUTURE Economic Future? Crosstabulation

Count

		FUTURE Economic Future?			
		1 Better Off	2 About Same	3 Worse Off	Total
V5 Increase Corporate Taxes	1 Strongly Disagree	10	5	3	18
	2	35	15	6	56
	3	79	78	28	185
	4	88	86	49	223
	5 Strongly Agree	85	70	57	212
Total		297	254	143	694

GPAR grade point average * FUTURE Economic Future? Crosstabulation

Count

		FUTURE Economic Future?			
		1 Better Off	2 About Same	3 Worse Off	Total
GPAR grade point average	55 less than 60	8	3	3	14
	62 60-64	12	17	11	40
	67 65-69	32	38	21	91
	72 70-74	73	50	23	146
	77 75-79	60	49	29	138
	82 80-84	53	35	29	117
	87 85+	11	9	3	23
Total		249	201	119	569

3. Part A. The calculations for the standard deviation are in Table 4 and following.

Table 4: Calculations for standard deviation of *V5* for those who said they would be *Worse off*

Response to *V5*	X	f	fX	fX^2
Strongly disagree	1	3	3	3
	2	6	12	24
	3	28	84	252
	4	49	196	784
Srongly agree	5	57	285	1,425
Total		143	580	2,488

The variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\Sigma fX^2 - \frac{(\Sigma fX)^2}{n}\right) \\ &= \frac{1}{143-1}\left(2,488 - \frac{580^2}{143}\right) \\ &= \frac{1}{143}\left(2,488 - \frac{336,400}{143}\right) \\ &= \frac{2,488 - 2,352.4476}{143} \\ &= \frac{135.5524}{143} \\ &= 0.9479 \end{aligned}$$

and the standard deviation is $s = \sqrt{s^2} = \sqrt{0.9479} = 0.9736$ or 0.974.

This value for the standard deviation is slightly different from the value of 0.977 given in 2.b. This difference is likely due to differences in rounding procedures.

Part B. From the table of question 3. (a).

i. The probability of selecting someone who believes they will be worse off is

$$\frac{\text{number worse off}}{\text{total}} = \frac{143}{694} = 0.206.$$

ii. The chance of selecting someone who disagrees with increasing corporate taxes

$$P(1) + P(2) - P(1 \text{ and } 2) = \frac{18}{694} + \frac{56}{694} - \frac{0}{694} = \frac{74}{694} = 0.107$$

.

iii. The probability of selecting better off and agrees is

$$P(\text{better off and agree}) = \frac{88 + 85}{694} = \frac{173}{694} = 0.249.$$

iv. The probability of neutral or the same in the future is

$$P(\text{neutral}) + P(\text{same}) - P(\text{neutral and same}) = \frac{185}{694} + \frac{254}{694} - \frac{78}{694} = \frac{361}{694} = 0.520.$$

v. The probability of strongly agreeing, given better off, is

$$P(\text{strongly agree/better off}) = \frac{85}{297} = 0.286.$$

The probability that the individual selected strongly agrees, given worse off is

$$P(\text{strongly agree/worse off}) = \frac{57}{143} = 0.399.$$

vi. The probability of better off, given strongly disagreeing, is

$$P(\text{better off/strongly disagree}) = \frac{10}{18} = 0.556.$$

The probability of better off, given strongly agreeing, is

$$P(\text{better off/strongly agree}) = \frac{85}{212} = 0.401.$$

vii. If A is the event of strongly agreeing and B is the event of believing they will be worse off, one way of checking to see whether the two events are independent or not is to see whether $P(A/B)$ and $P(A)$ are equal. These probabilities are

$$P(A/B) = \frac{57}{143} = 0.399.$$

$$P(A) = \frac{212}{694} = 0.305.$$

These two probabilities differ by almost 0.1, so are not real close to each other. As a result, these two events can be considered dependent events.

Alternatively, a check to see whether $P(B/A)$ and $P(B)$ are equal is another way to examine this. These probabilities are

$$P(B/A) = \frac{57}{212} = 0.269.$$

$$P(B) = \frac{143}{694} = 0.206.$$

and again there is a considerable difference between these two probabilities so the events of strongly agreeing with increasing corporate taxes and believing they will be worse off are dependent events.

viii. If A is the event of having a neutral response and B is the event of believing they will be better off, these two events can be checked for independence by examining the probabilities of $P(A/B)$ and $P(A)$. These probabilities are

$$P(A/B) = \frac{79}{297} = 0.266.$$

$$P(A) = \frac{185}{694} = 0.267.$$

These two probabilities are very close to equal, so A and B can be considered to be independent of each other in practice.

Alternatively, a check to see whether $P(B/A)$ and $P(B)$ are equal is another way to examine this. These probabilities are

$$P(B/A) = \frac{79}{185} = 0.427.$$

$$P(B) = \frac{297}{694} = 0.428.$$

Again, these two probabilities are very close to equal, so A and B are practically independent. That is, the neutral response about increasing corporate taxes and a positive evaluation of the future are independent events.

Part C. From the table of question 3. (b).

i. The probability of a grade of 75 or above is

$$P(75+) = \frac{\text{number } 75+}{\text{total}} = \frac{138 + 117 + 23}{569}\frac{278}{569} = 0.489.$$

ii. The chance of obtaining a grade of 75 or above and worse off is

$$P(75+ \text{ and worse off}) = \frac{29 + 29 + 3}{569} = \frac{61}{569} = 0.107$$

.

iii. The conditional probability of 70-74, given better off, is

$$P(70\text{-}74/\text{better off}) = \frac{73}{249} = 0.293.$$

The conditional probability of 70-74, given about the same, is

$$P(70\text{-}74/\text{about same}) = \frac{50}{201} = 0.249.$$

The conditional probability of 70-74 is

$$P(70\text{-}74/\text{worse off}) = \frac{23}{119} = 0.193.$$

The probability that the individual selected obtains a grade of 70-74 is worse off is

$$P(\text{70-74}) = \frac{146}{569} = 0.257.$$

From these four probabilities, evaluation of the future and obtaining a grade of 70-74 are connected. The events of 70-74 and better off are dependent events, since $0.293 > 0.257$. Again, the events of 70-74 and worse off are dependent events, since $0.193 < 0.257$. The events of 70-74 and same are less dependent since 0.249 and 0.257 are similar. This might be expected since worse off reduces probability of 70-74 and better off increases probability of 70-74, so about the same might be expected to produce a similar probability of 70-74, as compared with the overall probability.

iv. The probability of a grade of 80 plus is

$$P(\text{80 plus}) = \frac{117 + 23}{569} = \frac{140}{569} = 0.246$$

and

$$P(\text{80 plus/better off}) = \frac{53 + 11}{249} = \frac{64}{249} = 0.257.$$

These two probabilities are very close to each other so they are very close to being independent events.

Part D. **Comments on evaluation of future and connections with grade point average and view on corporate taxes**.

From question 2, there is no consistent relationship between grade point and evaluation of the future. However, there is a relatively clear cut relationship between attitude about increasing corporate taxes and evaluation of the future. Those with a less positive evaluation of their future generally expressed more support for increasing corporate taxes.

From part B, in v., the probability of strong agreement with increasing corporate taxes is greater for those who think they will be worse off than for those who think they will be better off. From vi, the probability of considering oneself better off is greater for those who strongly disagree with increased corporate taxes, as compared with those who

strongly agree. From vii., the chance of strongly agreeing with increasing corporate taxes is greater for those who think they will be worse off than the overall probability of strongly agreeing these two events are dependent. Together these support the results of question 2 agreement with increased corporate taxes is more associated with a negative evaluation of the future while disagreement with increased corporate taxes is more associated with positive evaluation of the future.

In the case of grade point average, from C. iii. there is some dependence of a grade of 70-74 and the different evaluations of the future. The chance of obtaining a grade of 70-74 is greater the more positive evaluation the respondent has of the future. From C. iv. though, there is little dependence between the grade of 80 plus and being better off.

In general, there appears to be less connection between grades and evaluation of the future than there is between attitude on corporate taxes and evaluation of the future.

4. **Adolescent self-perception of health**.

   (a) Frequency in I – this appears to be based on a study of adolescents and would require obtaining data in order to make this statement.

   Risk in II could be frequency if what is referred to is actual risks based on empirical estimates of probabilities, using studies of smoking and drinking. It could be subjective or judgmental if all that is meant is the adolescent judgment of risks – most individuals would not know the exact probabilities of risks, so there is a subjective aspect to estimating this risk. As a result, this would appear to be some combination of frequency and subjective probability.

   Risk in III is frequency. Here the probablities or risks are estimated using data from a study of health of adolescents.

   (b) The conditional probability of depressive episode, given 12- to 14-year old girl, is

$$P(\text{depression}/\text{12-14 girl}) = 0.06$$

since 6% of such girls have these episodes. For boys, the corresponding probability is

$$P(\text{depression/12-14 girl}) = 0.02$$

As a result, the risk of having a major depressive episode is considerably raised by being a girl, and lowered by being a boy. Since there are about one-half boys and one-half girls, the overall probability of having a major depressive episode is halfway between 6% and 2%, or 4% or 0.04. Since these probabilities each differ from 0.04, the events are dependent events.

(c) The quote states that 15- to 17-year old boys are at no more risk of depression than those aged 12-14. So the risk of depression for boys of these two ages is much the same. The events of being 15-17 or of being 12-14 are each independent of depression, at least across the teenage years for boys.

(d) What I had intended is for you to point out that from I, self-rating of health is dependent on whether the adolescent smoke, drank, or was obese. That is, these latter events increase the probability that these adolescents rated their health as poor, fair, or good, as oppposed to very good or excellent health. However, I forgot to point out that the scale for self-rating of health is a five point scale from poor to excellent, and this information is not given anywhere in the question. As a result, there seems to be no way of telling whether these events are dependent or independent, given the information provided in the question.

But in II, there appears to be independence of the events of smoking and drinking and knowing risks. That is, knowledge of risks does not appear to change the probability that the adolescent smokes or drinks.