

POS 3713: Homework Assignment #2
 Spring 2001
 Due on Friday, February 16th

The purpose of this assignment is to introduce you to various descriptive statistics, measures of central tendency, and measures of dispersion. You must **type** your answers to the following questions.

Every two years, the National Election Study is conducted at the University of Michigan. This is a national in-person survey that asks numerous questions related to political attitudes and political behavior. Most of the data that you will be analyzing in the assignments for this class will come from the 1996 National Election Study (NES).

Part A: Frequency and Contingency Tables

One of the most useful pieces of information to obtain in a data set is a frequency distribution of the variables. Using the 1996 NES, I have produced a frequency distribution of the variables for Region (identifying which region a respondent lives in) and Prayofn (how frequently does a respondent pray) . Use this information to answer the questions below.

Region

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1. Northeast (CT, ME, MA, NH, NJ, NY, PA	260	15.2	15.2	15.2
2. North Central (IL, IN, IA, KS, MI, MN	458	26.7	26.7	41.9
3. South (AL, AR, DE, DC, FL, GA, KY, LA	642	37.5	37.5	79.3
4. West (AK, AZ, CA, CO, HI, ID, MT, NM,	354	20.7	20.7	100.0
Total	1714	100.0	100.0	

How often R prays

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1. Several times a day	519	30.3	30.5	30.5
2. Once a day	391	22.8	23.0	53.4
3. A few times a week	345	20.1	20.3	73.7
4. Once a week or less	287	16.7	16.9	90.5
5. Never	161	9.4	9.5	100.0
Total	1703	99.4	100.0	
Missing System	11	.6		
Total	1714	100.0		

Question A1:

What is the level of measurement for these variables, Region and Prayofn (nominal, ordinal, or interval)? Explain your answer.

As Political Scientists, we are often interested in assessing the relationship between two or more variables, such as party identification and vote, or education and political knowledge. For variables that are measured at the nominal or ordinal level, it is useful to examine the relationship between two or more variables in the form of a contingency table or crosstabulation. One variable is displayed in the rows of the table, while the other variable is displayed in the columns of the table.

For example, in his 1996 article "Charting a Course to Conflict: Territorial Issues and Interstate Conflict, 1816-1992", Dr. Paul Hensel examines the relationship between the issues at stake in militarized disputes and the escalation level of these disputes. He argues that disputes over territorial stakes will be much more likely to escalate to war than disputes over other issues. He tests this relationship with the following data on issues at stake in militarized disputes. Both variables are measured at the nominal level.

Table 2
Militarized Dispute Escalation to War

Territorial Issues at Stake?	Non-War Disputes	Interstate War	Total
No	1420	36	1456
Yes	543	43	586
Total	1963	79	2042

You can see that 2.5% (36/1456) of disputes not involving territorial issues escalate to war, whereas 7.3% (43/586) of disputes over territory escalate to war. He concludes that disputes involving territorial stakes are 3 times more likely to escalate to the level of full-scale war.

Using the 1996 NES, I have created a contingency table of the two variables presented above in the frequency tables (Region, Prayofn). In the table below, the first value in each cell is the frequency. This is followed by the row percentage (frequency/row total), column percentage (frequency/column total), and total percentage (frequency/N for entire table). Use this information to answer the questions below.

How often R prays * Region Crosstabulation

			Region				Total
			1. Northeast (CT, ME, MA, NH, NJ, NY, PA)	2. North Central (IL, IN, IA, KS, MI, MN)	3. South (AL, AR, DE, DC, FL, GA, KY, LA)	4. West (AK, AZ, CA, CO, HI, ID, MT, NM,	
How often R prays	1. Several times a day	Count	60	126	237	96	519
		% within How often R prays	11.6%	24.3%	45.7%	18.5%	100.0%
		% within Region	23.2%	27.7%	37.1%	27.4%	30.5%
		% of Total	3.5%	7.4%	13.9%	5.6%	30.5%
	2. Once a day	Count	68	109	152	62	391
		% within How often R prays	17.4%	27.9%	38.9%	15.9%	100.0%
		% within Region	26.3%	24.0%	23.8%	17.7%	23.0%
		% of Total	4.0%	6.4%	8.9%	3.6%	23.0%
	3. A few times a week	Count	47	103	123	72	345
		% within How often R prays	13.6%	29.9%	35.7%	20.9%	100.0%
		% within Region	18.1%	22.6%	19.3%	20.5%	20.3%
		% of Total	2.8%	6.0%	7.2%	4.2%	20.3%
	4. Once a week or less	Count	53	74	91	69	287
		% within How often R prays	18.5%	25.8%	31.7%	24.0%	100.0%
		% within Region	20.5%	16.3%	14.3%	19.7%	16.9%
% of Total		3.1%	4.3%	5.3%	4.1%	16.9%	
5. Never	Count	31	43	35	52	161	
	% within How often R prays	19.3%	26.7%	21.7%	32.3%	100.0%	
	% within Region	12.0%	9.5%	5.5%	14.8%	9.5%	
	% of Total	1.8%	2.5%	2.1%	3.1%	9.5%	
Total	Count	259	455	638	351	1703	
	% within How often R prays	15.2%	26.7%	37.5%	20.6%	100.0%	
	% within Region	100.0%	100.0%	100.0%	100.0%	100.0%	
	% of Total	15.2%	26.7%	37.5%	20.6%	100.0%	

Question A2: How many respondents in the sample are from the North Central region and pray once a day?

Question A3: What percentage of respondents pray a few times a week?

Question A4: What percentage of all respondents pray several times a day **and** live in the West region (hint: you are looking for a total percentage)?

Question A5: What can you conclude in general about the relationship between how often a person prays and the region they live in? Is there any tendency for people in one region of the US to pray more often than people in other regions? If so, what could account for this difference?

Part B: Descriptive Statistics for Interval Measures

For variables measured at the interval level, it is useful to examine measures of central tendency (mean, median, mode) and measures of dispersion (such as variance). Measures of central tendency give us an idea of the typical or average case in a distribution, while measures of dispersion allow us to compare how much heterogeneity there in a distribution of scores. For a survey question, one can interpret the measures of dispersion as tapping attitudes regarding the amount of agreement among survey respondents. Variables with higher levels of dispersion have more disagreement.

The 1996 NES survey asks each respondent to place Bill and Hillary Clinton on a *feeling thermometer* ranging from 0 to 100. Ratings between 50 degrees and 100 degrees mean that a

respondent feels favorable and warm toward Bill or Hillary. Ratings between 0 and 50 degrees mean that a respondent doesn't feel favorable toward Bill or Hillary, while a score of 50 indicates that the respondent does not feel particularly warm or cold towards Bill or Hillary. To answer the questions below, you will be comparing people's attitudes about Bill Clinton and Hillary Clinton, using these feeling thermometer scales. I have included two tables below that report the measures of central tendency and measures of dispersion for these feeling thermometer variables.

Statistics

		Clinton thermometer	Hillary Clinton Thermometer
N	Valid	1705	1685
	Missing	9	29
Mean		59.34	52.81
Median		70.00	60.00
Mode		85	70

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Clinton thermometer	1705	100	0	100	59.34	29.58	874.802
Hillary Clinton Thermometer	1685	100	0	100	52.81	29.85	891.005
Valid N (listwise)	1683						

Question B1: Compare the mean thermometer levels for Bill and Hillary Clinton. Who do the respondents have warmer feelings for (represented by higher average scores on the thermometer variables)? Do you agree with these averages, i.e., would you rank Bill and Hillary Clinton in the same order based on your feelings about them?

Question B2: Compare the median and mode thermometer levels for Bill and Hillary Clinton. Is your comparison of these two measures of central tendency similar to what you found in part (a)?

Question B3: According to the measures of dispersion (range, standard deviation, and variance), do people have more similar feelings for Bill or Hillary; in other words, do people agree more about Bill or Hillary?

Part C: Measures of Dispersion for Nominal Variables, IQV

Finally, we will examine a measure of dispersion for nominal variables, the Index of Qualitative Variation (IQV). The 1996 NES asks a series of questions about how the respondents feel about President Clinton on a variety of dimensions. For example, one question asks the respondents how intelligent they think Clinton is, while another asks how moral they think Clinton is. The possible responses range from extremely well (they think Clinton is very intelligent or moral) to not well at all. Use the information in the frequency tables below to answer the following questions.

INTELLIGENCE (CLINTON)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Extremely well	656	38.3	38.6	38.6
	2. Quite well	847	49.4	49.8	88.4
	3. Not too well	160	9.3	9.4	97.8
	4. Not well at all	38	2.2	2.2	100.0
	Total	1701	99.2	100.0	
Missing	System	13	.8		
Total		1714	100.0		

MORAL (CLINTON)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Extremely well	103	6.0	6.1	6.1
	2. Quite well	555	32.4	33.1	39.2
	3. Not too well	586	34.2	34.9	74.2
	4. Not well at all	433	25.3	25.8	100.0
	Total	1677	97.8	100.0	
Missing	System	37	2.2		
Total		1714	100.0		

Question C1: Calculate the Index of Qualitative Variation (IQV) using the formula on p. 92 of Healey for both variables (intelligent and moral).

Question C2: According to the IQV, do respondents think Clinton is more intelligent or more moral? Which of Clinton's personality traits do people agree upon the most (remember that a higher IQV score indicates *less* agreement among respondents)? Speculate on why agreement among respondents is different for each variable.