POLS 5377 Scope & Method of Political Science

Week 4 Introduction to Statistics

Basic Descriptive Statistics

Healey. (2016) *Statistics: A Tool for Social Research,* Chapter 2

Key Question:

- How to compute and interpret percentages, proportions, ratios, rates, and percentage change?
- How to construct and analyze frequency distributions for variables at each level of measurement?

Outline

- Percentage & Proportion
- Ratio, rates, & percentage change
- Frequency distribution

Percentage & Proportion

- Consider the easiness of processing the information in readers' brain

Religious Group	Frequency	Percentages
Protestant	116,203,000	53.71%
Catholic	57,199,000	26.44%
Jewish	2,680,000	1.24%
Muslim	1,349,000	0.62%
Buddhist	1,189,000	0.55%
Unitarian	586,000	0.27%
Other	2,992,000	1.38%
None	34,169,000	15.79%
	N = 216,367,000	100.00%

Difficult to process

Easier to interpret

Percentage & Proportion

- Percentage and proportion report **relative** size
- Compare a **part** (specific category) to a **whole** (all categories)
- Example: in a room of 10 people, 6 of them are women, and 4 are men. Express the distribution in terms of proportion and percentage:
  - The proportion of women in the room is 0.4 (4/10).
  - The percentage of women in the room is 40% (4/10*100%).

Percentage & Proportion

- Formula

$$\text{Percentage: } \% = \left(\frac{f}{N}\right) \times 100$$

$$\text{Proportion } = \left(\frac{f}{N}\right)$$

where f= frequency, or the number of cases in any category

N= the number of cases in all categories

Percentage & Proportion

- With small number of cases (less than 20), report actual frequencies
- Always report number of observations along with proportions and percentages
- Can calculate percentages and proportions for variables at all levels of measurement

Ratio

- Compare the relative sizes of categories
- Compare **parts** to **parts**
- Ratio = $\frac{f_1}{f_2}$
  - f_1 is the number of cases in first category
  - f_2 is the number of cases in second category
- Example: In a class of 23 females and 19 males, the ratio of males to females is:
  - 19/23 = 0.83
  - For every female, there are 0.83 males

Rates

- Express the number of actual occurrences of an event (births, deaths, homicides) vs. the number of possible occurrences per some unit of time

$$\text{Crude death rate} = \frac{\text{Number of deaths}}{\text{Totoal population}} \times 1000$$

- If a town of 2300 had 17 deaths last year, the death rate is:
  - (17/2300) * 1000 = (.00739) * 1000 = 7.39
  - The town had 7.39 death for every 1000 residents
- This computation allows you to compare the deaths across countries without the impact of the size of population.

Rates

- Steps:
  1. Determine the number of occurrences (*f*)
  2. Determine the possible number of occurrences (n)
  3. Divide the actual occurrences (*f*) by the possible occurrences (n)
  4. Multiply by a power of 10 (often 1,000 or 100,000)

Percentage Change

- Measures the relative increase or decrease in a variable over time

$$\text{Percentage change} = \left(\frac{f_2 - f_1}{f_1}\right) \times 100$$

where f_1 = first score, frequency, or value

f_2 = second score, frequency, or value

- Percentage change can also be calculated with percentages, rates, or other values
- Percentage change can be positive (indicating increase from time 1 to time 2) or negative (indicating decrease from time 1 to time 2)

Percentage Change

- Example 1

In 1990, a state had a murder rate of 7.3. By 2000, the rate had increased to 10.7. What was the relative change?

- (10.7 – 7.3 / 7.3) * 100 = (3.4 / 7.3) * 100 = 46.58%
- The rate increased by 46.58%.

- Example 2

Projected Population Growth for Six Nations, 2012–2050

Nation	Population, 2012 (f_1)	Population, 2050 (f_2)	Increase/Decrease ($f_2 - f_1$)	Percentage Change
China	1,350,400,000	1,310,700,000	−39,700,000	−2.94
U.S.	313,900,000	422,600,000	108,700,000	34.63
Nigeria	170,100,000	402,400,000	232,300,000	136.57
Mexico	116,100,000	143,900,000	27,800,000	23.94
U.K.	63,200,000	79,600,000	16,400,000	25.95
Canada	34,900,000	48,600,000	13,700,000	39.26

Frequency Distribution

- Data can appear as individual listings (raw scores) where each unit is shown with its value.
- Example, in a statistic class of 5, each student receive the following scores on an examination:
  - Abby 85, Cathy 90, Jason 81, Rob 95, Joe 90
- With only 5 students (number of case N = 5), it's easy to read the values and know that the students did well. However, if there were 50, 100, or 200 students, evaluating the exam performance would be much difficult.

Frequency Distribution

- A frequency distribution summarizes all the values into an easy-to-read format.
- To construct a frequency table, order the data by value and list the number of cases recorded for each value.
- In a statistics class with 50 students, the following grades were recorded:

100, 98, 95, 94, 91, 90, 87, 86, 85, 84, 80, 80, 77, 70, 65, 64, 60, 100, 95, 94, 90, 87, 86, 85, 84, 80, 70, 95, 90, 87, 86, 85, 80, 95, 90, 87, 86, 85, 95, 90, 87, 86, 85, 90, 87, 85, 90, 87,87, 87

Frequency Distribution

- We need an easier way to organize the data. So, we can list the scores by their frequency.
- I arranged all the score in order, and count the frequency (*f*) of each score.
- For example, there are 2 students received a score of 100. The frequency of 100 is 2. The table on the next page shows the frequency distribution of all the 50 scores.

Score	Frequency (*f*)
100	2
98	1
95	5
94	2
91	1
90	7
87	9
86	5
85	6
84	2
80	2
77	3
70	2
65	1
64	1
60	1
	N = 50

Frequency Distribution

- As you can see, the table is long and it's hard to get useful information. So, I group the scores into categories.
- The categories should be of equal size and mutually exclusive. Because there is no scores lower than 60, so I keep the last category open.
- The grouping changes the data from ratio into ordinal level measurement.

Score	Frequency (*f*)
100 -90	18
89 – 80	26
79 – 70	3
69 and below	3
	N = 50

Frequency Distribution

- Frequency distributions are tables that report the number of cases in each category of a variable
- Frequency distributions summarize distribution of a variable by reporting the number of times each score of a variable occurred
- General Rule for categories of frequency distribution:
  - Exhaustive
  - Mutually exclusive
  - Each case counted in one and only one category

Frequency Distribution – Nominal Variables

- Frequency distributions for nominal variables

Gender	Frequency
Males	53
Females	60
	$N = 113$

Religious Group	Frequency
Protestant	116,203,000
Catholic	57,199,000
Jewish	2,680,000
Other	6,116,000
None	34,169,000
	$N = 216,367,000$

Frequency Distribution – Ordinal Variables

- Support for birth control on a university campus

Do you strongly agree, agree, disagree, or strongly disagree that the University Health Center should provide condoms and other "safe sex" items on demand and at no additional cost to students?

Response	Frequency	Percentage
Strongly agree	350	25.55%
Agree	462	33.72%
Disagree	348	25.40%
Strongly disagree	210	15.33%
	1370	100.00%

Do you strongly agree, agree, disagree, or strongly disagree that the University Health Center should provide condoms and other "safe sex" items on demand and at no additional cost to students?

Response	Frequency	Percentage
Strongly agree or Agree	812	59.27%
Disagree or Strongly disagree	558	40.73%
	1370	100.00%

Frequency Distribution – Interval/Ratio Variables

- Basic consideration
  - Complexity
  - Large number of categories
  - Requires collapsing or grouping of categories
  - Decide the number of categories and the width of those categories
  - **Class intervals** refer to the categories used in the frequency distribution

Frequency Distribution – Interval/Ratio Variables

- Grouping

Age of Students in a College Class (fictitious data)

Interval width = 1 year of age	
Ages	Frequency
18	5
19	6
20	3
21	2
22	1
23	1
24	1
25	0
26	1
	$N = 20$

Age of Students in a College Class (fictitious data)

Interval width = 2 years of age		
Age	Frequency	Percentage
18–19	11	55.0%
20–21	5	25.0%
22–23	2	10.0%
24–25	1	5.0%
26–27	1	5.0%
	$N = 20$	100.0%

Cumulative Frequency & Cumulative Percentage

- Cumulative frequencies and percentages refer to how many cases fall below a given score or class interval

Age of Students in a College Class

Age	Frequency	Cumulative Frequency	Percentage	Cumulative Percentage
18–19	11	11	55.0%	55.0%
20–21	5	16	25.0%	80.0%
22–23	2	18	10.0%	90.0%
24–25	1	19	5.0%	95.0%
26–27	1	20	5.0%	100.0%
	$N = 20$		100.0%	

The End