Reading Statistical Tables

Basic principles for understanding what the researcher is trying to tell you (that is, questions you should ask yourself when reading a table):

- What is the source of this table?
- How many variables are presented? What are their names?
- What is represented by the numbers presented in the first column? In the second column?

Table 1. Prevalence of Each Category of Adverse Childhood Experiences and ACE Score by Sex*

	No. (%)		
	Women (n = 9367)	Men (n = 7970)	Total (N = 17 337)
Adverse childhood experiences			
Emotional abuse	1227 (13.1)	602 (7.6)	1829 (10.5)
Physical abuse	2530 (27.0)	2382 (29.9)	4912 (28.3)
Sexual abuse	2310 (24.7)	1276 (16.0)	3586 (20.7)
Battered mother	1281 (13.7)	920 (11.5)	2201 (12.7)
Household alcohol/drug abuse	2759 (29.5)	1896 (23.8)	4655 (26.9)
Mental illness in household	1937 (20.7)	1058 (13.3)	2995 (17.3)
Parental separation or divorce	2293 (24.5)	1738 (21.8)	4031 (23.3)
Incarcerated household member	485 (5.2)	324 (4.1)	809 (4.7)
ACE score			
0	3271 (34.9)	3044 (38.2)	6315 (36.4)
1	2299 (24.5)	2237 (28.1)	4536 (26.2)
2	1443 (15.4)	1297 (16.3)	2740 (15.8)
3	969 (10.3)	685 (8.6)	1654 (9.5)
4	665 (7.1)	382 (4.8)	1047 (6.0)
5	390 (4.2)	212 (2.7)	602 (3.5)
6	210 (2.2)	74 (0.9)	284 (1.6)
≥7	120 (1.3)	39 (0.5)	159 (0.9)

*For ACE Study waves 1 and 2 combined. ACE indicates adverse childhood experiences. See "Methods" for definitions of each type of adverse event.

Source: Dube, S. R., Anda, R. F., Felitti, V. J., Chapman, D. P., Williamson, D. F., & Giles, W. H. (2001). Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the Adverse Childhood Experiences Study. *Jama*, *286*(24), 3089-3096.

Table 1. Demographic data for girls and women aged 13–24 years

	Participants (n=1244)*
Age-group	
13–17 years	575 (46·4%)
18–24 years	669 (53·6%)
Community setting	
Urban	201 (14·9%)
Rural	1043 (85·1%)
Orphan status†	
Biological mother died	125 (9·6%)
Biological father died	241 (18·4%)
Death of both biological parents	83 (7·6%)
Death of at least one biological parent‡	449 (36·0%)
Marital status	
Married	127 (9·7%)
Not married	1112 (90·3%)

Source : Reza, A., Breiding, M. J., Gulaid, J., Mercy, J. A., Blanton, C., Mthethwa, Z., ... & Anderson, M. (2009). Sexual violence and its health consequences for female children in Swaziland: a cluster survey study. *The Lancet*, *373*(9679), 1966-1972.

Table. 1 Unregistered births (1000s) in 2003 by region and level of development

	Births	Unregistered children
World	133 028	48 276 (36%)
Sub-Saharan Africa	26 879	14 751 (55%)
Middle East and north Africa	9790	1543 (16%)
South Asia	37 099	23 395 (63%)
East Asia and Pacific	31 616	5901 (19%)
Latin America and Caribbean	11 567	1787 (15%)
CEE<comma> CIS<comma> and Baltic states	5250	1218 (23%)
Industrialised countries	10 827	218 (2%)
Developing countries	119 973	48 147 (40%)
Least developed countries	27 819	16 682 (71%)

Source : Marmot, M., Friel, S., Bell, R., Houweling, T. A., & Taylor, S. (2008). Closing the gap in a generation: health equity through action on the social determinants of health. *The Lancet*, *372*(9650), 1661-1669.

Chapter 4: What is a measure of Central Tendency?

Numbers that describe what is typical of the distribution

•You can think of this value as where the middle of a distribution lies (the median).

or

•The value within a distribution of values that has the most cases (mode)

or

•The mathematical average (mean)

Measure of Central Tendency:

The Mode

- The category with the largest frequency (or percentage) in the distribution.

The Mode: An Example

- Which of the three candidates represents the "mode" for these candidates?
- Variable=Candidates
  Candidate A – 11,769 votes
  Candidate B – 39,443 votes
  Candidate C – 78,331 votes

Level of measurement = _____

The Mode=_____

The Mode: An Example

- **Which of the three candidates represents the "mode" for these candidates**
- **Variable=Candidates**
  Candidate A – 11,769 votes
  Candidate B – 39,443 votes
  Candidate C – 78,331 votes

Level of measurement = nominal (why?)

The Mode= Candidate C (why?)

The mode can be calculated for variables within all levels of measurement that are: nominal, ordinal, or interval-ratio.

Measure of Central Tendency: The Median

- The score that divides the distribution into two equal parts, so that half the units (cases) are above it and half below it.
- The median is the middle score in a distribution.
- The median is appropriate for ordinal or interval-ratio data.

Finding the Median for an Ordinal Variable

Job Satisfaction (I am very satisfied with my job)

Values	Freq	Cummulative Frequency
Agree Strongly	5	5
Agree	10	15
Undecided	3	18
Disagree	7	25
Dis. Strongly	3	28
Total Cases:	28	

Steps to Determine Median for Ordinal Var

1. divide total # of cases by 2: **28/2 = 14**
2. determine/calculate the cumulative frequencies
3. locate the value (category) that holds the middle case (unit): "**agree**" contains the 14th case

Finding the Median for an Interval/Ratio Variable

What is the interval/ratio variable below?

What is the median # of hate crimes?

What is the "unit of analysis"?

Number of Hate Crimes in State

NC = 39
Penn = 141
TX = 287
Ohio = 255
Fla = 240

Finding Median for Interval/Ratio Variable

# of hate crimes by state

Cases
NC = 39
Penn = 141
TX = 287
Ohio = 255
Fla = 240

States ordered low to high

NC = 39
Penn = 141
Fla = 240
Ohio = 255
TX = 287

# of cases (or units) = 5

Steps to Determine:

1. Order the variable (hate crimes) from highest to lowest or vice versa
2. Add 1 to the total # of units (states) if there is an odd # of units (e.g .,1+5=6)
3. divide resulting number by 2 (6/2 = 3)
4. Count down that many units (cases) to identify the middle or median (Fla)

Percentile

- A score at or below which a specific percentage of the distribution falls.
- For example, the 75th percentile is a score for which 75% of the cases are at or below it.

Percentile

Table 1: Satisfaction with Health

	Freq	Cum Freq	%	Cum %
Very Low	5	5	18	18
Low	7	12	25	43
Moderate	6	18	21	64
High	7	25	25	89
Very High	3	28	11	100
Total N:	28			

Steps to Determine Percentile: determine cumulative percentages and then locate the percentile of interest. The 75th percentile would be which category: _____

The Mean

The arithmetic average obtained by adding up all the scores and dividing by the total number of scores.

The mean is used with interval-ratio data.

Can be used with ordinal data but is not very accurate/precise.

Formula for the Mean

$$\overline{Y} = \frac{\sum Y}{N}$$

"Y bar" ($\overline{Y}$) equals the average or the sum of all the scores, Y, divided by the number of scores, N

(for example add up the # of hate crimes for the states and then divide by "N" or the number of states).

Calculating the mean with frequency distributions (ordinal variable):

Satisfaction with Health

	Freq	Category x Freq
1 - Very High	5	5
2 - High	7	14
3 - Moderate	6	18
4 - Low	7	28
5 - Very Low	3	15
Total N:	28	80

$$\overline{Y} = \frac{\sum fY}{N}$$

Steps to Determine:

1. multiply each category by its frequency (category x frequency)
2. sum all the "category x freq" scores to determine total (80)
3. divide total (80) by total number of cases (total N or 28) to get average score (2.86)

In-Class Exercise:

Calculate the mode, median, and mean for the grouped frequency below.

Satisfaction with Parking

Level of Satisfaction	Frequency
1 Very Satisfied	190
2 Satisfied	316
3 Somewhat Satisfied	54
4 Somewhat Dissatisfied	17
5 Dissatisfied	2
6 Very dissatisfied	2
TOTAL	581

Ordinal (Grouped) Data: Mode

Category with the most cases or "Satisfied (#2)"

Satisfaction with Parking

Level of Satisfaction	Frequency
1 Very Satisfied	190
2 Satisfied	316
3 Somewhat Satisfied	54
4 Somewhat Dissatisfied	17
5 Dissatisfied	2
6 Very dissatisfied	2
TOTAL	581

Ordinal (Grouped) Data: Median

- Make sure values are ordered
- Add one to total frequency (if an odd #): 581 + 1 = 582
- Divide by 2: 582/2 = 291
- Calculate cumulative frequency and determine which category contains the 291st person (answer is "Satisfied" or #2)

Level of Satisfaction	Frequency	Cumulative Freq
1 V. Satisfied	190	190
2 Satisfied	316	506
3 Somewhat Sat.	54	560
4 Somewhat Dis.	17	577
5 Dissatified	2	579
6 V. Dissatisfied	2	581
TOTAL	581	

Ordinal (Grouped) Data: Mean

- Multiply frequency (# of people) times category
- Sum the scores obtained; 1,074
- Divide by total frequency 1074/581 to obtain mean category (mean=1.85 people per household)

Level of Satisfaction	Frequency	Category x Frequency
1 Very Satisfied	190	190
2 Satisfied	316	632
3 Somewhat Satisfied	54	162
4 Somewhat Dissatisfied	17	68
5 Dissatisfied	2	10
6 Very Dissatisfied	2	12
TOTAL	581	1,074

Considerations for Choosing a Measure of Central Tendency

- For a nominal variable, the mode is the only measure that can be used.
- For ordinal variables, the mode and the median may be used. The median provides more information (taking into account the ranking of categories). Can also use interval/ratio but not precise.
- For interval-ratio variables, the mode, median, and mean may all be calculated. The mean provides the most information about the distribution, but the median is preferred if the distribution is skewed.

When choosing the appropriate measure of central tendency for a distribution, what should you consider?

the level of measurement of the variables

(e.g., mode for nominal level)

What is usually the appropriate measure of central tendency for interval-ratio level?

the mean

What is the primary "weakness" of the mean?

the mean is highly influenced by extreme scores in one direction

(e.g., the mean may not "represent" the true distribution of the cases very well)

Example of mean "unrepresentative" of sample

Sample 1: Score for Five Women	Sample 2: Score for Five Women
100	100
110	110
125	125
125	125
135	450

What is the mode: _____

What is the median: _____

What is the mean: _____

Example of mean "unrepresentative" of sample

Sample 1: Score for Five Women	Sample 2: Score for Five Women
100	100
110	110
125	125
125	125
135	450

What is the mode: 125 and 125

What is the median: _____

What is the mean: _____

Example of mean "unrepresentative" of sample

Sample 1: Score for Five Women	Sample 2: Score for Five Women
100	100
110	110
125	125
125	125
135	450

What is the mode: 125 and 125

What is the median: 125 and 125

What is the mean: _____

Example of mean "unrepresentative" of sample

Sample 1: Score for Five Women	Sample 2: Score for Five Women
100	100
110	110
125	125
125	125
135	450

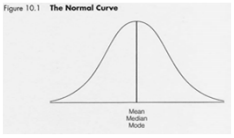
What is the mode: 125 and 125

What is the median: 125 and 125

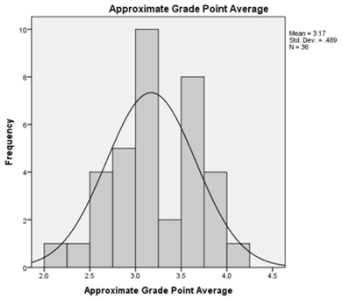
What is the mean: 119 and 182

Normal Distributions (also called normal curve)

- **Normal Distribution**
  - Used with **linear** variables
  - A bell-shaped and symmetrical theoretical distribution (a theoretical distribution of cases is not an actual distribution of cases),
  - with the mean, the median, and the mode all coinciding at its peak and
  - with frequencies gradually decreasing at both ends of the curve.



Normal Distributions



Normal Distributions

- **Normal Distribution**
  - What happens when we have a few cases that are far above or below the other cases?
  - Negatively Skewed: a few extremely low values
  - Positively Skewed: a few extremely high values