## Slide 1

You won't see one of these in your output. Ever. Trust me.

**NORMAL DISTRIBUTION**

Yeah, probably not this one either, but... definitely not the first one.

**PARANORMAL DISTRIBUTION**

3

## Slide 2

### Best place for movies...

1. Amazon Prime — 15%
2. Disney+ — 15%
3. Hulu — 5%
4. Netflix — 17%
5. YouTube — 7%
6. An actual *theater* — 41%

SOC497 @ CSUN w/ Ellis Godard

28

6

## Slide 3

### Where we are...

| DEADLINES | | | LECTURE | | LABS | | | |
|---|---|---|---|---|---|---|---|---|
| Date | Things Due | Reading (15th) | Category | TOPIC | Lab # | Lab Assignment | Type | T | R |
| Tue Jan 21 | | | Orientation | Orientation | | | | | |
| Thu Jan 23 | | chapt. 1 | Orientation | Inquiry & Research | 1 | Empirical Claims | Topical | | 1 |
| Tue Jan 28 | | chapt. 4 | Orientation | Research Design | (2) | Mini-Design | Solo | 1, (2) | |
| Thu Jan 30 | | | Orientation | SPSS Intro | 3 | Computing (miles) | Data | | (2), 3 |
| Tue Feb 4 | HW1 Article Review | chapt. 2 | Orientation | Explanation & Theory | 4 | Writing Hypo | Topical | 3,4 | |
| Thu Feb 6 | | ch 17 & p 43 | Orientation | Researching Research (gsssub) | | | | | |
| | | | Measurement | Recoding (ir-ish) | 5 | Recoding (colleges) | Data | | 4, 5 |
| Tue Feb 11 | HW2 Annotations | pp. 124-137 | Orientation | Conceptualization | (6) | Measuring Love | Solo | 5, (6) | |
| Thu Feb 13 | | chapt 5 | | Operationalization | | Internal Validity | | | (6), 7 |
| Tue Feb 18 | | pp. 421-426 | Measurement | Elementary Stats | 8 | Univariates (gsssub) | Data | 7, 8 | |
| Thu Feb 20 | | | | Measures of Research | | Levels (EC0 code) | Solo | 7 | |
| Tue Feb 25 | HW3 Dataset Basics | p. 415-420, 434-438 | Measurement | Quantitative Analysis | 9 | Coding Heroes | Topical | EC1, 9 | 8, EC1 |
| Thu Feb 27 | | ch.6 & 427-428 | Measurement | Indices & Scales | 10 | Lab (abort) | Data | | 9, 10 |
| Tue Mar 4 | | pp. 256-258 | Measurement | Surveys: Wording | 11 | Draft Survey Questions | Topical | 10, 11 | |
| Thu Mar 6 | HW4 Welfare Analysis | chapt. 7 | Measurement | Sampling & Errors | (12) | Apartments | Solo | | 11, (12) |
| Tue Mar 11 | | p. 426-427, 429-433 | Description | Crosstabs | (13) | Crosstabs (gsssub) | Data | (12), 13 | |
| Thu Mar 13 | MIDTERM | | | Project Overview (mandatory lecture) | - | - | - | | 13 |
| Tue Mar 18 | | | | | | | | | |
| Tue Mar 25 | HW5 Indices | pp. 256-263 | Measurement | Surveys: Format | 14 | Draft Survey Format | Topical | 14 | |
| Thu Mar 27 | | - | Description | Testing (Ordinal) Association | 15 | Ordinal Tests (USDecline) | Data | | 14, 15 |
| Tue Apr 1 | | chapt. 8 | Observation | Experimental Research | 16 | Review Survey Draft | Topical | 15, 16 | |
| Thu Apr 3 | HW 6 Proposal | - | Description | Three Interval Tests | 17 | 3 Tests | Data | | 16, 17 |
| Tue Apr 8 | | pp. 263-288 | Observation | Surveys: Mode | (18) | Conduct Survey | Solo | 17, (18) | |
| Thu Apr 10 | | pp. 433-434 | Description | Multivariate & Modeling | 19 | Multiple R (multireg) | Data | | (18), 19 |
| Tue Apr 15 | HW7 Three Tests | ch 13 & 428-429 | Orientation | Qualitative Analysis | (20) | Jaywalking Study | Solo | 19, (20) | |
| Thu Apr 17 | | pp. 420-421 | Observation | Data Entry | (21) | Data Entry | Solo | | (20), (21) |
| Tue Apr 22 | | chapt. 10 | Description | Fieldwork & Ethnography | (22) | Grounded Theory | Solo | (21), (22) | |
| Thu Apr 24 | HW8 Experiment | chapt. 15 | Description | Control Variables & Elab | EC2 | Elaboration (extra credit) | Data | | (22), EC2 |
| Tue Apr 29 | | chapt. 11 | Observation | Unobtrusive Research | 23 | Content Analysis | Topical | EC2, 23 | |
| Thu May 1 | HW9 Survey Analysis | pp. 316-318 | Observation | Focus Groups | 24 | Evaluation Design | Topical | | 23, 24 |
| Tue May 6 | | chapt. 12 | Description | Oral Presentations | (25) | Oral Presentations Eval | Solo | 24, (25) | (25) |
| Thu May 8 | | | | | | | - | - | |
| Thu May 15 | Final Exam | | | (no class meeting - due 5pm, firm!) | | | - | - | |

**14 days** · **23 days**

SOC497 @ CSUN w/ Ellis Godard

4

## Slide 4

Admin · Choices · Levels · Statistics · Measures · Choosing · SPSS

### Missing Info

◆ 1 Intakes Form

◆ 3 Headshots

◆ 2 said "no" to PDF
  ▪ email if you change your mind

SOC497 @ CSUN w/ Ellis Godard

8

## Slide 5
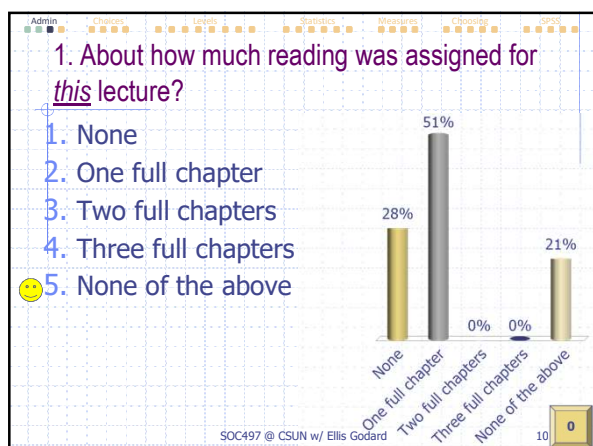
SOC497/L: SOCIOLOGY RESEARCH METHODS

# Elementary Statistics:

## Choices & Implications

Ellis Godard

## Slide 6

Admin · Choices · Levels · Statistics · Measures · Choosing · SPSS

### Outline for Today

◆ Operationalization Choices
  ▪ Evolution, Components
  ▪ Levels of Measurement

◆ Review of Elementary Statistics
  ▪ Central Tendency & Dispersion
  ▪ Choosing Statistics
  ▪ Distributions & Shapes

◆ Example & Lab

SOC497 @ CSUN w/ Ellis Godard

9

## Slide 1

1. About how much reading was assigned for *this* lecture?

1. None
2. One full chapter
3. Two full chapters
4. Three full chapters
5. None of the above

Bar chart:
- None: 28%
- One full chapter: 51%
- Two full chapters: 0%
- Three full chapters: 0%
- None of the above: 21%

SOC497 @ CSUN w/ Ellis Godard

10    0

## Slide 2 — Historic Shifts in Operations

*Historic* Shifts in Operations

- Sex vs. Gender, and dichotomy>complexity
- Measurements of "race" in the U.S. Census
  - None in 1790, 1800, 1810
  - Color in 1820, '30, '40; Mulatto added 1850
  - Chinese & Indian added 1860, Japanese in 1870
  - All to "Black" and "White" in 1920
  - Mexican added 1930, dropped in 1940
  - 1960 shift from physical appearance to self-identity
  - 1980 – five categories
  - 1990 & 2000 others added
  - 2010 – biracial and other
  - 2020 – citizenship? (almost)

SOC497 @ CSUN w/ Ellis Godard    13

## Slide 3 — Last Lab Assigned

- Threats to Internal Validity
- Some you plan for; some you can't
  - Choices in research design have implications for what happens during data collection
- This lecture & lab:
  - Choices in <u>measurement</u> have implications for what happens (or can happen) during analysis

SOC497 @ CSUN w/ Ellis Godard    11

## Slide 4 — Design & Meanings

- Data *results* depend on Data *definitions*
- Consider Data Sources – e.g. Violent Crime
  - Different measurements > Diff results
  - FBI's UCR vs Criminal Victimization Survey
    - UCR under-reports; CVS (self-reports) may over-report
- Operational Focus – e.g. IPV
  - Measures of *incidence* – incidents
  - Measures of *prevalence* – rate, victims

SOC497 @ CSUN w/ Ellis Godard    14

## Slide 5 — Evolution w/i any Research Project

- Ways measures change *during* a study
  - Recoding & computing (labs 3 & 5)
  - Indices & Scales (lab 10)
  - Select cases (fiprbs in 424; some projects in 497)
  - Bivariate intersections (labs 13, 15, 17, 19)
- Timing of that varies
  - Ideally, would have spelled out in advance
  - Often, arise during data analysis
  - Also, they change over time…

SOC497 @ CSUN w/ Ellis Godard    12

## Slide 6 — Measurement & Meaning

- Consequence of Changes
  - Changing the measurement changes the meaning!
  - Different operations, measure different things – or, at least, measure the same things differently
- Everything matters
  - Seemingly minor changes in wording, aren't minor
  - Things taken for granted, are now hotly contested
  - Nothing "has to" be measured in any one way

SOC497 @ CSUN w/ Ellis Godard    15

## Basic Elements of Variation

- ◆ **Attribute:** Characteristic or quality of something
  - ■ Examples:
    - ◆ young, female, Armenian, queer, wealthy, plumber
  - ■ Single measurement's value (as if) for 1 case
    - ◆ *One* of those e.g.'s - *Could* apply to *many* cases
- ◆ <u>Variable:</u> a logical set of related attributes
  - ■ Examples:
    - ◆ age, gender, ethnicity, sexual orientation, SES, occupation
  - ■ Two primary characteristics/requirements:
    - ◆ Exhaustiveness: Able to classify every observation
    - ◆ Mutually exclusivity: each case fits 1&*only*1 value

SOC497 @ CSUN w/ Ellis Godard          16

## Interval Variables

- ◆ Distance btwn attributes is measured & uniform
  - ■ Meaningful standard intervals
    - ◆ Distance between successive values is clear
  - ■ Meaningful spectrum of values
    - ◆ Distance between *any* two values can be *calculated*
    - ◆ Difference between any two cases can be *calculated*
- ◆ No "true zero"
  - ■ Ratio measures originate/end at zero
    - ◆ e.g. age, Kelvin
  - ■ Values can be compared
    - ◆ Jill has twice as much as Joe
  - ■ No other differences for our purposes
    - ◆ just use NOI

SOC497 @ CSUN w/ Ellis Godard          19

## Levels of Measurement

- ◆ Nominal variables:
  - ■ <u>Only</u> exhaustive & mutually exclusive – just names
  - ■ Values cannot be ordered/ranked – apples/oranges
  - ■ Examples: gender, race, religion, department
- ◆ Ordinal variables:
  - ■ <u>Also</u> rank-ordered – more/less, higher/lower
    - ◆ RanksRelative, not absolute
    - ◆ Difference between 2 values or cases is unclear
    - ◆ Range covered by each value *may* be unclear too
  - ■ Examples: short/medium/tall; <HS/HS/BA/+

SOC497 @ CSUN w/ Ellis Godard          17

## 3. Which value is from an interval variable?

1. <8 years of education
2. 12-14 years old
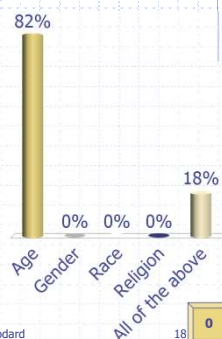3. >5 sexual partners
4. $18K-$20K in income
5. None of the above

53%   33%   5%   0%   9%
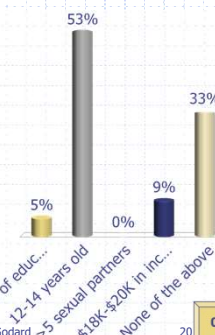
SOC497 @ CSUN w/ Ellis Godard          20

## 2. Which variable has rankable values?

✓ 1. Age
2. Gender
3. Race
4. Religion
5. All of the above

82%   0%   0%   0%   18%

SOC497 @ CSUN w/ Ellis Godard          18

## Question 2 from 1st Day Quiz

- ◆ 2. which LOM appropriate for...
  - ■ College major – nominal
  - ■ Socioeconomic status (low medium high) – ordinal
  - ■ Average GPA – interval
  - ■ Occupation (plumber, accountant, teacher, etc.) – Nominal
  - ■ Able to compose web pages in HTML (yes or no) – nominal (b/c 2)
  - ■ Verbal complexity (on a 100-point continuum) – interval

SOC497 @ CSUN w/ Ellis Godard          22

## Levels are a Measurement *Choice*

- ◆ Behavior – e.g. class enrollment
  - Full-time/part-time
  - 0, 1-2, 3-4, 5+
  - {specific number of classes or units}
- ◆ Characteristics – e.g. gender
  - Male / Female
  - Masculine, Mixed, Feminine
  - Of 100 binary aspects (body? behavior? clothes?), # male?
- ◆ Orientations – e.g. sexual
  - Straight / Gay  (Heteronormative / Non-normative?)
  - Heterosexual, Bisexual, Homosexual
  - Kinsey Scale (10 point)

SOC424 w/ Dr. Ellis Godard          23

## Course Progress

**Typically learn statistics in this order:**

- ◆ Descriptive **statistics for** univariate **data**
- ◆ Inferential **statistics for** univariate **data**
- ◆ Descriptions **of** bivariate **relationships**
- ◆ Inferences **from** bivariate **relationships**

SOC497 @ CSUN w/ Ellis Godard          27

## Why Do Levels Matter?

*Associated w/ Different Statistics, 2 ways…*

- ◆ Each level is described differently
  - Each requires different *univariate* procedures
    - Can't compute "average religion"
    - *Can* compute *most frequent* (modal) religion
  - Combinations require different *bivariate* procedures
- ◆ Stat. techniques requires (min.) level
  - Each has a set of assumptions about data
  - Inc. mathematical manipulation of the values
    - Addition, subtraction, multiplication, division
    - Require at least interval level of measurement

SOC564 w/ Dr. Ellis Godard -- Slide          24

## Univariate Analysis

- ◆ Describing a sample in terms of a single variable
  - Gender is a good example. How many of the respondents were women, as compared to men.
- ◆ Including discussion of Distributions
  - How answers R distributed across possible responses
  - What is the shape of the distribution? (…)
  - Where is this distribution centered? (typical value)
  - How spread out is the distribution? (dispersion)

SOC497 @ CSUN w/ Ellis Godard          28

## Introduction to Statistics

- ◆ Meaning
  - Numbers or Procedures?
  - Parameters (about populations) vs. Statistics (about samples) !
- ◆ Purposes
  - Descriptive statistics – numerically summarize observations
  - Inferential statistics – generalize beyond a sample
- ◆ Complexity
  - Univariate analysis – single variables, *not* relationships, causes, etc.
  - Bivariate analysis – two variables; test relationships, causes, etc.
  - Multivariate Analysis – more than two variables

SOC497 @ CSUN w/ Ellis Godard          26

## Shapes of Distributions

- ◆ Basic Targets
  - Normal – not same as "bell-shaped"
  - Skewed – left/right? heavy/slight?
- ◆ Oddities
  - U-shaped (polarized), log, etc.
  - Flat/even/uniform
  - Logarhythmic
- ◆ Warnings
  - May be nothing distinctive
  - Don't exaggerate – almost certainly *not* normal!

SOC497 @ CSUN w/ Ellis Godard          29

## Slide 30

# Histograms @ Amazon

◆ Which shorts are better?

64% 18% 13% 5% 0%   67% 15% 7% 4% 6%

4.4 out of 5
81 global ratings
See customer reviews ›
5 star  4 star  3 star  2 star  1 star

4.3 out of 5
735 global ratings
See customer reviews ›
5 star  4 star  3 star  2 star  1 star

SOC497 @ CSUN w/ Ellis Godard        30

## Slide 33

# Measures of Dispersion

|  | Formula | Appropriate if… |
|---|---|---|
| **Variation Ratio** | The percent of cases not in the modal category | **Nominal** (because nothing else works) |
| **Range** | Simple subtraction of the lowest value (the "minimum") from the highest (the "maximum") | **Ordinal** (maybe?) **interval** (if sample range differs from population's) |
| **Inner-quartile range (IQR)** | Same as *the range*, but only *of the middle half* of the cases when ordered – i.e. from the 25th percentile to the 75th | **Interval** (if skewed) **Ordinal** (esp. if range is full spectrum of variable) |

SOC497 @ CSUN w/ Ellis Godard        33

## Slide 31

# Univariate Descriptive Statistics

◆ Central Tendency (what is typical?)
- Mean: arithmetic average ("center of gravity")
- Median: value of middle case (when sample ordered)
- Mode: most frequently occurring value (*not* majority)

◆ Dispersion (how far are they spread out?)
- Standard Deviation (NOT the variance*)
- Range (Max – Min) & IQR (middle 50%)
- Variation* ratio: percent that isn't the mode

  * Note that these are NOT the same word – *variance* vs. *variation ratio*

SOC497 @ CSUN w/ Ellis Godard        31

## Slide 34

# More Measures of Dispersion

|  | Formula | Appropriate if… |
|---|---|---|
| **Variance** | The average squared difference between each value and the mean $$s = \frac{\sum (Y_i - \overline{Y})^2}{n}$$ (for population: $\sigma^2$) | *NEVER* USE IT FOR DISPERSION! *NEVER* ASK SPSS FOR IT! YOU WON'T INTERPRET IT ON ITS OWN! |
| **Standard Deviation** | Square root of variance $$s = \sqrt{s^2} = \sqrt{\frac{\sum (Y_i - \overline{Y})^2}{n}}$$ (for population: $\sigma$) | **Interval** (unless skewed; then use IQR, or range?) |

SOC497 @ CSUN w/ Ellis Godard        34

## Slide 32

# Measures of Central Tendancy

|  | Formula | Appropriate if… |
|---|---|---|
| **Mean** | Simple average (the sum of all the values, divided by the number of values) $$\overline{Y} = \frac{\sum Y_i}{n}$$ | **Interval** (unless heavily skewed) |
| **Median** | If even number of cases, the median *case* (not value) is the (n/2)th case. Otherwise, it is the [(n+1)/2]th case | **interval** (if skewed), **ordinal** (and/or mode?) |
| **Mode** | Highest (relative) frequency | any, but esp. **nominal** (because = only choice) |

SOC497 @ CSUN w/ Ellis Godard        32

## Slide 35

# Standard Deviations & Standard Errors

◆ Deviations across *sample* distributions:
- About 68% of the values in a normal distribution will fall w/i 1 standard deviation of the mean, 95% within 1.96, and 99.9 within 3.

◆ Deviations across *sampling* distributions:
- Same idea, same distribution, same %'s
- But the standard deviation of a *sampling* distribution is called the standard error
- Don't confuse that with *sampling* error

SOC497 @ CSUN w/ Ellis Godard        35

---

### Slide 36

## Criteria for Selection

- What else makes a measure of central tendency or dispersion "appropriate"? (in order of importance)
  1. **Scale of measurement –** no medians or means for nominal data; don't trust means for ordinal data;
  2. **Shape of distribution –** e.g. for skewed interval data, use both – the mean will differ from the median in the direction of skew i.e. higher if skewed right, lower if skewed left)
  3. **Robustness –** Here, a statistic is "robust" if it resists sampling deviations. The mean is fairly robust, but the median is less misleading if there are scraggly tales
  4. **Efficiency –** use the highest level of precision available (modes are least precise)
  5. **When in doubt, use more than one**

SOC497 @ CSUN w/ Ellis Godard                36

---

### Slide 39

## Question 1 from 1st Day Quiz

- Interval
  - Central tendency: mean (median if skewed)
  - Dispersion: standard deviation (mode " ")
- Ordinal
  - Central tendency: median (& mode?)
  - Dispersion: range (IQR?)
- Nominal
  - Central tendency: mode (only!)
  - Dispersion: variation ratio (Index of Qual. Variation)

SOC497 @ CSUN w/ Ellis Godard                39

---

### Slide 37

## 4. To measure nominal dispersion use…

1. Mode
   0%
2. Median
   5%
3. Variance
   2%
4. Variation Ratio
   86%
5. Standard Deviation
   7%

SOC497 @ CSUN w/ Ellis Godard                37    0

---

### Slide 40

## Questions 3-6 from 1st Day Quiz

- 3. Which level(s) appropriate for crosstabs?
  - Ord, Nom
- 4. which level(s) for t test or F test (ANOVA)?
  - Interval (t: 2 levels of Ord or Nom; F: >2 of O)
- 5. Which level(s) for regression?
  - Any (though most of you did interval, if that)
- 6. Formula for z (or t)
  - Math 140! ☹

SOC497 @ CSUN w/ Ellis Godard                40

---

### Slide 38

## Typical Choices

- **Nominal**
  - Mode and variation ratio (because there are *no alternatives*)
- **Ordinal**
  - Median & IQR (because you can at least put the values in order - & IQR better than Range)
- **Interval**
  - Mean & Standard deviation (taking advantage of equal increments between values) unless it's skewed (then median & IQR)

SOC497 @ CSUN w/ Ellis Godard                38

---

### Slide 41

## SPSS Tips for Today's Lab

- Two options for basic stats (used the 2nd)
  - ANALYZE – DESCRIPTIVES – DESCRIPTIVES
    - Mean, stdev, a few others – but *not* all you need!
  - ANALYZE – DESCRIPTIVES – FREQUENCIES
    - Stats & Choose which ones you want
    - More options, more output, tables, etc.
    - Use *this* one!!

SOC497 @ CSUN w/ Ellis Godard                41

---

## SPSS Tips for HW5 (Happy etc)

- ◆ Pearson's Chi-square statistic
  - ■ ANALYZE – DESCRIPTIVES – CROSSTABS
    - ◆ DV = row; IV = column
    - ◆ Stats – Chi-square
- ◆ Correlation coefficient
  - ■ ANALYZE – CORRELATE – BIVARIATE
- ◆ Just get those stats! **_BOTH_** of them!!
  - ■ Don't worry about crosstab or corr. Matrix
  - ■ If you don't recall these, do office hours!

SOC497 @ CSUN w/ Ellis Godard          42

## Team Scores

| Points | Team | Points | Team |
|--------|------|--------|------|
| 3.57 | Netflix | | |
| 3.57 | Disney+ | | |
| 3.33 | YouTube | | |
| 3.08 | An actual theater | | |
| 3 | Hulu | | |

SOC497 @ CSUN w/ Ellis Godard          47

## Lab Exercise

- ◆ From the given dataset (@ Canvas)
- ◆ Pick 3 variables
  - ■ One for each level of measurement (I, O, & N)
  - ■ **_Do NOT_** use the "measure" column in SPSS! EVER! ☺
    - ◆ Look at the values column and/or codebook and/or freq tables
- ◆ Submit
  - ■ Printout of frequency tables & histograms
  - ■ Description of the shape of each distribution
  - ■ Report & interpret cent. tendency & dispersion of each
    - ◆ Report the pieces *and* tell a story – a sentence or 2 each (x3)
    - ◆ *Use* the statistics to *describe the sample*

SOC497 @ CSUN w/ Ellis Godard          43

## 5. Which of these did I *NOT* "wear" virtually on my head today during lecture?

1. An orange cat, kneading
   2%
2. A cowboy hat w/ bandana
   5%
3. A fireman's hat w/ badge
   2%
4. A Viking helmet w/ horns
   0%
5. A skiing helmet in the snow
   90%

SOC497 @ CSUN w/ Ellis Godard          46          0