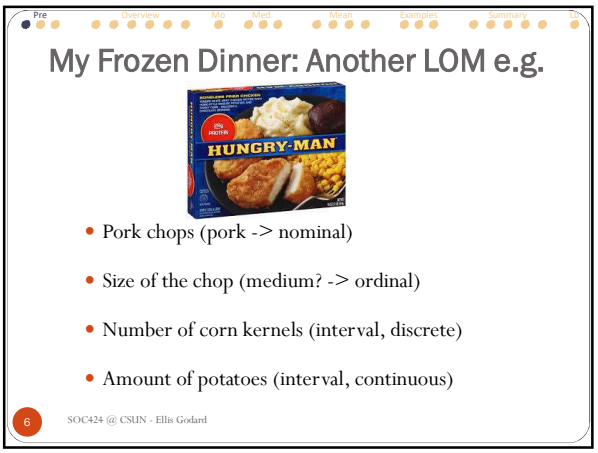
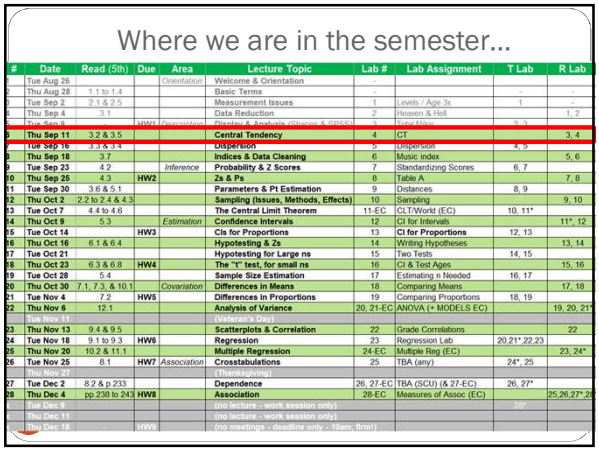
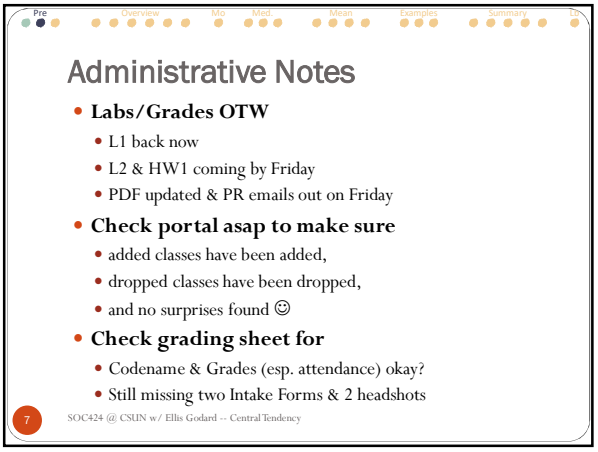
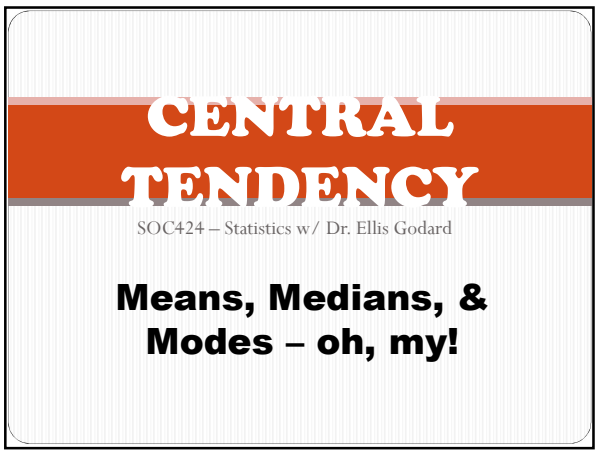
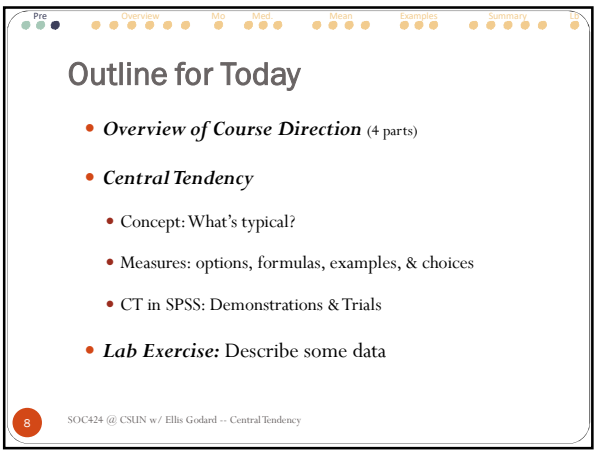

crispin lover
@hog_mild
Mean Girls implies the existence of Median Girls, Mode Girls, and Range Girls
11:44 AM · 2021-02-09 · Twitter for iPhone


Pre Overview Mo Med Mean Examples Summary To
My Frozen Dinner: Another LOM e.g.
HUNGRY-MAN
Pork chops (pork -> nominal)
Size of the chop (medium? -> ordinal)
Number of corn kernels (interval, discrete)
Amount of potatoes (interval, continuous)
6 SOC424 @ CSUN - Ellis Godard


Where we are in the semester...
Date Read (5th) Due Area Lecture Topic Lab # Lab Assignment T Lab R Lab
1 Tue Aug 26 Orientation Welcome & Orientation - - - -
2 Thu Aug 28 1.1 to 1.4 Basic Terms - - - -
3 Tue Sep 2 2.1 & 2.5 Measurement Issues 1 Levels / Age 3x 1 -
4 Thu Sep 4 3.1 Data Reduction 2 Heaven & Hell - 1, 2
5 Tue Sep 9 HW1 Description Displays & Analysis (Shapes & SPSS) 3 Data Maps 2, 3 -
6 Thu Sep 11 3.2 & 3.5 Central Tendency 4 CT - 3, 4
7 Tue Sep 16 3.3 & 3.4 Dispersion 5 Dispersion 4, 5 -
8 Thu Sep 18 3.7 Indices & Data Cleaning 6 Music index - 5, 6
9 Tue Sep 23 4.2 Inference Probability & Z Scores 7 Standardizing Scores 6, 7 -
10 Thu Sep 25 4.3 HW2 Zs & Ps 8 Table A - 7, 8
11 Tue Sep 30 3.6 & 5.1 Parameters & Pt Estimation 9 Distances 8, 9 -
12 Thu Oct 2 2.2 to 2.4 & 4.3 Sampling (Issues, Methods, Effects) 10 Sampling - 9, 10
13 Tue Oct 7 4.4 to 4.6 The Central Limit Theorem 11-EC CLT/World (EC) 10, 11* -
14 Thu Oct 9 5.3 Estimation Confidence Intervals 12 CI for Intervals - 11*, 12
15 Tue Oct 14 HW3 CIs for Proportions 13 CI for Proportions 12, 13 -
16 Thu Oct 16 6.1 & 6.4 Hypotesting & Zs 14 Writing Hypotheses - 13, 14
17 Tue Oct 21 Hypotesting for Large ns 15 Two Tests 14, 15 -
18 Thu Oct 23 6.3 & 6.8 HW4 The "t" test, for small ns 16 CI & Test Ages - 15, 16
19 Tue Oct 28 5.4 Sample Size Estimation 17 Estimating n Needed 16, 17 -
20 Thu Oct 30 7.1, 7.3, & 10.1 Covariation Differences in Means 18 Comparing Means - 17, 18
21 Tue Nov 4 7.2 HW5 Differences in Proportions 19 Comparing Proportions 18, 19 -
22 Thu Nov 6 12.1 Analysis of Variance 20, 21-EC ANOVA (+ MODELS EC) - 19, 20, 21*
Tue Nov 11 Veterans Day (no class)
23 Thu Nov 13 9.4 & 9.5 Scatterplots & Correlation 22 Grade Correlations - 22
24 Tue Nov 18 9.1 to 9.3 HW6 Regression 23 Regression Lab 20,21* 22,23 -
25 Thu Nov 20 10.2 & 11.1 Multiple Regression 24-EC Multiple Reg (EC) - 23, 24*
26 Tue Nov 25 8.1 HW7 Association Crosstabulations 25 TBA (any) 24*, 25 -
Thu Nov 27 Thanksgiving (no class)
27 Tue Dec 2 8.2 & p 233 Dependence 26, 27-EC TBA (SCU) (& 27-EC) 26, 27* -
28 Thu Dec 4 pp 238 to 243 HW8 Association 28-EC Measures of Assoc (EC) - 25,26,27*,28*
Tue Dec 9 (no lecture - work session only) TBA -
Thu Dec 11 (no lecture - work session only)
Tue Dec 16 HW9 (no meetings scheduled/only if/then, if/m!)


Pre Overview Mo Med Mean Examples Summary To
Administrative Notes
Labs/Grades OTW
L1 back now
L2 & HW1 coming by Friday
PDF updated & PR emails out on Friday
Check portal asap to make sure
added classes have been added,
dropped classes have been dropped,
and no surprises found ☺
Check grading sheet for
Codename & Grades (esp. attendance) okay?
Still missing two Intake Forms & 2 headshots
7 SOC424 @ CSUN w/ Ellis Godard -- Central Tendency

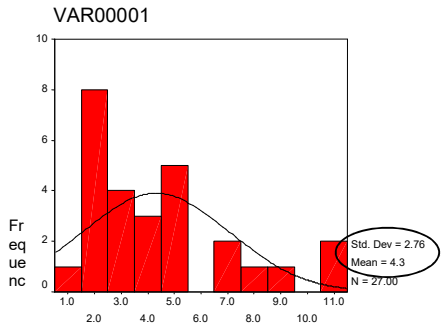

CENTRAL TENDENCY
SOC424 – Statistics w/ Dr. Ellis Godard
Means, Medians, & Modes – oh, my!


Pre Overview Mo Med Mean Examples Summary To
Outline for Today
Overview of Course Direction (4 parts)
Central Tendency
Concept: What's typical?
Measures: options, formulas, examples, & choices
CT in SPSS: Demonstrations & Trials
Lab Exercise: Describe some data
8 SOC424 @ CSUN w/ Ellis Godard -- Central Tendency

Toward Bivariate Inference…

- Want to make inferences about a population based on data from sample
  - Estimate the parameters from the statistics
  - For example, population mean from sample mean
    - "Mu" from "Ybar"
- Start w/ sample (i.e. descriptive) statistics
  - How to summarize your data
- Build towards discussion of parameters
  - Then will be ready for *bivariate* statistics

You've already seen 1 of each…



Course Direction

We'll look at statistics in this order:

- **Descriptive** statistics for **univariate** data
- **Inferential** statistics for **univariate** data

then ….

- **Descriptions** of **bivariate** relationships
- **Inferences** from **bivariate** relationships

3 Measures of Central Tendency

- **Mode**
  - most frequently occurring value
- **Median**
  - middle value, when cases ordered by value
- **Mean:**
  - arithmetic average, "center of gravity"

Univariate Descriptive Statistics

- Central tendency
  - Where is the data centered?
  - What's the typical value?
  - Do different kinds of centers differ?
- Dispersion
  - How are the data spread out?
    - Widely? Narrowly?
    - Highly concentrated or very dispersed?
- You've seen 1 of each already…

What's New & What Isn't

- **You learned those stats in 5th grade** (maybe 4th? 6th)
- But I recognize that **there *are* new things** about them now
  - **Opening a dataset** (FILE>OPEN>DATA)
  - **Getting those stats from SPSS**
    - Always check 6: mean, median, mode, quartiles, std dev, & range
  - **Choosing variables** (nominal, ordinal, & interval)
    - Don't use the Measure column – look @ the Values column
  - **Choosing which statistic to use for which variable**
    - Guidance throughout (and again @ the end of) this lecture
  - **Putting them in a meaningful sentence**
    - Not for a mathematician – for a relative, boss, neighbor, etc.
    - Reporting & Interpreting them – not just the number; what's it *mean*?
    - Describe the sample / Summarize the data

About Modes...

- = the most frequently occurring score
- Any level of measurement *might* have one
  - *Could* be *no* mode (equal distribution)
  - *Could* be *more* than one (bimodal, tri-, etc)
    - But don't go overboard – looking for **concentrations**
- = the *least* you can say about center
  - Might not even be center – could be 1 end
  - Doesn't utilize differences among values
  - Ignores all non-modal values
- In a *symmetrical* (more soon) distribution:
  mean = median = mode

About Medians...

- Proportional measure
  - Score that falls in the middle *when they're ranked by magnitude*
  - 50th percentile, ½ above and ½ below
  - Cumulative percentage shows it – find the first value w/ cum % > 50
- It's a value, not a case
  - Of {0, 3, 7, 8, 8}, it's 7
  - If Larry had a value of 7, the median is 7, not Larry
  - The *third case* is in the middle, & the *value* of that case is 7, *not 3*
  - Of {3, 4, 7, 8, 8} it's still 7, the value of the 3rd case, *not 3*
- It's the value of the middle case, when they're in order
  - In {8, 7, 3, 0, 7}, it's still 7

About Medians...

- If n = odd, there is a unique median {0,3,7,8,8}
  - Rank scores, find *value* of the middle case (NOT the middle value)
  - If there are 5 cases, median *case* is (5+1)/2 = 3rd
    and the *value of that case* is 7 (not 3!)
- If even number of cases - e.g. {0,7,8,8}
  - Some say to average the middle two positions
    - Average of value at position (n/2) and the next one: (n/2)+((n/2)+1)/2
  - But that's weird… no case had a value of 7.5;
    how could there be a median that doesn't even exist?
  - Better to report both ("median of 7 to 8"?), though
    usually they'll be the same (because you'll have 100s or 1000s of cases)

Properties of Medians...

- can use for ordinal data
  - *must* be at *least* ordinal level data
  - positional measure
    - 50% above, 50% below – like a percentile ranking
- does not use all information from the data
  - generally not affected by outliers
- inference procedures are difficult
  - cannot be manipulated algebraically

About Means (Notation Intro.)

- μ = "Mu" = population mean
- $\hat{Y}$ = "Yhat" = estimate of a value
- $\overline{Y}$ = "Ybar" = sample mean
  - sum over each case (until you reach end) & divide by sample size
  - Just a simple "average"
    - "Average" sometimes used w/ non-interval data
    - Only interval data has a mathematical average

Basics of Statistical Notation

- N = population size; n = sample size
- Y_i = each particular case
  - Individual person, school, church, etc.
  - *Not* the sample, but a single element of it
- Y_1 = 1st case
- Y_n = last case
- Σ = summation

Properties of Means…

- **Uses all data/scores**
- **Manipulated algebraically easily**
  - so lends itself to statistical inference procedures
- **Must have interval data**
  - Can't add and divide meaningfully if not equidistant
- **Affected by outliers (extreme values)**
  - like a fulcrum – center of gravity of data
- **Best computed for bell-shaped distributions**
  - not as good if bimodal or heavily skewed

Weighted Means

- **Imagine separate means, for 2 groups**
  - e.g. avg hats owned by 5 girls and 50 boys
  - Can't just average averages
  - But girls' hats avg = total girls hats / 5, and boys' hats = total boys hats / 50
  - So if you multiply each group's avg, by # in that group, you can reverse engineer the separate totals
  - And adding the two separate totals, is like adding all of the individual scores w/o having them
- **So, variation: weighted average**
  - overall mean for combined set of samples
  - weight by # cases in each set

    (N1Ybar1 + N2Ybar2 …) / (n1 + n2 ..)
  - (5*girlsavg + 50*boysavg) / (5+50)

Non-Computer Examples

- Find the median of:

  {black, black, white, Asian, white)
- Find the mode and median of:

  {4, 6, 12, 7, 10, 3, 4}
- Find median of:

  {4, 6, 12, 10, 3, 4}
- Find mean, median, & mode for

  {2, 2, 7, 8, 3, 2, 11, 13, 9}
- More?
- Find the mean, median, & mode for

  {7, 6, 10, 7, 5, 9, 3, 7, 5, 13}

Comparisons w/ Descriptives

- Imagine these three variables:
  - Preferred network
    - That's nominal – no mean or median
  - Amount of time spent online
    - If the values are "some", "lots", "too much", that's ordinal – no mean
  - Number of friends
    - If that's an actual count, we can average them
- Using the appropriate measure of central tendency, we could compare two different samples, or subgroups…

Comparisons w/ Descriptives

Social Media Use by Men				
	Mean	Median	Mode	Dispersion
Preferred network	x	x	LinkedIn	53% (variation ratio)
Amount of time	x	"Some"	"None"	2 (range)
Number of friends	267.4	163	29	8.35 (std. dev) or 86 (IQR)

Social Media Use by Women				
	Mean	Median	Mode	Dispersion
Preferred network	x	x	Instagram	62% (variation ratio)
Amount of time	x	"Some"	"Tons"	2 (range)
Number of friends	345.2	119	42	11.72 (std. dev) or 112 (IQR)

3 Measures of Central Tendency

	Formula	Appropriate if…
Mean	**Simple average (the sum of all the values, divided by the number of values)** $\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$	**interval (unless heavily skewed); *is not meaningful for nominal, & arguably not for ordinals – just interval variables!***
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 (unless mode is a majority?) – *nominals don't have a middle***
Mode	**Highest (relative) frequency**	**Esp. nominal (no other choice) but any (?)**

Means or Medians for Interval?

- If a distribution is symmetric, mean = median
  - If distribution is normal, mean = median = mode
- If a distribution is skewed, the mean is in the direction of the skew from the median
  - If skewed to right, mean is higher than median
  - If skewed to left, mean is lower than median
- Distribution = like seesaw for the median
  - Add more on one end, it may slide in that direction
- But like magnets for the mean
  - Any addition anywhere pulls mean that direction

Criteria for Selection

What 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 –** for skewed interval data, mean + median say different things; which helps most?
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 highest level of precision appropriate for level of measurement (modes are least precise; means most)
5. **When in doubt, use more than one –** may need to *triangulate* center– pinpoint using multiple measures

Last point about estimates…

- **"Y-bar" is mean of a sample**
  - **vs "Mu" (the population mean)**
  - **Anything w/ a ^ is an estimate**
- **e.g. "Y-hat" is an estimate of "Mu"**
  - not the only way to estimate it, nor the best
  - just the easiest; it's a "point estimate"
- **Doesn't take into account dispersion**
  - Sampling theory allows us to consider disp.
  - Will delve into that next lecture… but first…

HW Hints / Clarifications

- **"Report"** includes giving the actual value
  - not just the number e.g. 1 for gender, but *which gender*
- **"Interpret"** means making sense of it
  - Not just 12.4 for # of years of educ, but "some college, on avg"
- Summarize clearly
  - State in terms of **the sample**
- Make full use of the CT measure
  - **Describe** the typical case
- Mode is not "most" – just most *common*
  - "most" implies well above a majority (>50%), but mode could be 8%

Lab Exercise

- Get GSS2018-trimmed.sav from Canvas - 745 variables
- Pick 3 variables, 1 @ each Level of Measurement (N, O, & I)
  - To determine LOM, use the Values column and/or frequency distributions and/or the actual survey – but do NOT use the "Measure" column in SPSS
  - In the Values column, any variable might have missings (97, 98, 99?) but *only* nominal and ordinal variables need other labels (intervals values *are* labels)
- Get *all* measures of central tendency for each, & print
  - When getting a frequency distribution, Click "Statistics" then click Mean, Median, and Mode
- Write at least a sentence about each variable:
  - Be clear about what each variable *measures*
    - It may look like a math class, but we're doing sociology
  - Report *appropriate* measure(s) of central tendency
    - Do not *report* (write about) all of them. Use criteria outlined above.
    - Do *not* simply report the choice (e.g. mean) or the value (e.g. 3); use the statistics to describe the sample (e.g. mostly female, highly educated, mostly happy with the class