SACRAMENTO STATE
Leadership begins here.

Descriptive Statistics

Stephen E. Brock, Ph.D., NCSP
California State University, Sacramento

1

Descriptive Statistics

- Describes (or summarizes) data.
- Describes quantitatively (with numbers or graphs) how a particular characteristic is distributed among one or more groups of people.
- No generalizations beyond the sample represented by the data are made by descriptive statistics.
  - However, if your data reflects an ***entire population***, then the data are considered to be population parameters.
  - On the other hand, if your data represents a population ***sample***, then the data is considered a statistic that describes a sample. Inferential statistics are required to determine if the samples statistics can be generalized back to the population.

2

Group Activity: Prepare to teach the following concepts.

- What is the “mean” of a data set?
- What is the “standard deviation” of a data set?
- What are derived or “standard scores?”
- What is the “bell shaped curve?”

3

Preface: Preparing Data for Analysis

- Scoring standardized test.
  - Follow manual instructions
  - Have someone else double check 25% of the protocols
- Scoring self-developed measures.
  - Establish reliability

4

Preface: Conducting Data for Analysis

- Be careful and take your time.
- Work with a partner who can check your work.
- Break-up data entry sessions to make sure you avoid the errors caused by fatigue.
  - Don't try to get it all done in one sitting
  - Double check the work of others (e.g., research assistants).
  - Assuming, accurate data entry, there is no such a thing as "bad data."
- Develop a coding system and set up a database.
- The structure of the database will depend upon the type of data being used.

5

Preface: Identify Scale of Measurement

Determines what descriptive statistic will be used

Scale	Properties	E.G.
Nominal (to name)	Data represents qualitative or equivalent categories (not numerical).	Eye color, Gender, Race or ethnicity (could be a word in database, but…). Mode
Ordinal (to order)	Numerically ranked, but has no implication about how far apart ranks are.	Grades (always a number in the database). Mode, Median
Interval (equal)	Numerical value indicates rank and meaningfully reflects relative distance between points on a scale	Temperature (always a number in the database). Mode, Median, Mean
Ratio (equal)	Has all the properties of an interval scale, and in addition has a true zero point.	Length, Weight (always a number in the database). Mode, Median, Mean

6

Preface: Data Entry

- Give each participant a subject #
- Data is generally placed in columns
- Codes for categorical and nominal data are determined.
  - Including group membership (usually coded as a group number).

7

Preface Coding Descriptive Data

- Develop a way to code each of the following variables. Remember only nominal data can be coded with words or letters, all other data must be quantified.
  - Eye color
  - Grades
  - IQ scores
  - Weight
  - Gender
  - Art skill level
  - Temperature
  - Length
  - Ethnicity
  - Race results

8

A Sample Data Base

S#	EC	Gd	IQ	Wt	Sx	Art	Eth	RR
1	Bn	4	100	97	1	10	1	1
2	Bn	3	105	65	1	3	1	2
3	Bn	4	130	200	2	5	3	3
4	Bl	4	111	99	2	7	4	4
5	H	1	90	43	1	9	2	5
6	Bn	0	65	55	1	2	5	6
7	G	2	117	67	1	4	6	7
8	H	2	100	87	2	6	7	8
9	Bl	2	89	96	1	8	3	9
10	Bn	4	85	45	2	1	3	10

9

Preface: Coding Experimental Data

- Develop a way to code each of the following variables. Remember only nominal data can be coded with words **or** letters, all other data must be quantified.
  - Group membership (ADHD Int, v ADHD Hyp v Bipolar Type 1)
  - Hyperactivity

10

A Sample Data Base

S#	Group	T-Score
1	1	60
2	1	50
3	1	55
4	2	71
5	2	78
6	1	65
7	2	88
8	2	70
9	2	89
10	2	85

S#	Group	T-Score
11	1	50
12	1	55
13	2	79
14	2	65
15	1	50
16	1	61
17	1	58
18	2	65
19	2	88
20	1	50

S#	Group	T-Score
21	3	79
22	3	65
23	3	80
24	3	71
25	3	78
26	3	65
27	3	88
28	3	70
29	3	61
30	3	90

11

Types of Descriptive Statistics

- Univariate (single variable data set summaries)
  - Measures of Central Tendency (location)
  - Measures of Variability (dispersion)
  - Measures of Shape (symmetry of the normal curve)
  - Measures of Relative Position (rank, standard score)
- Bivariate (two data sets)
  - Measures of Relationship

12

Measures of Central Tendency

- Mode
  - Determined by looking at a set of scores and seeing which occurs most frequently.
  - Not typically used, however, it is the only appropriate statistic for nominal data.

13

Measures of Central Tendency

- Median
  - The point above and below which 50% of the scores are found.
  - Unlike the mode, may not be one of the obtained results.
    - e.g., if there are an even number of scores the median is the point halfway between the two middle scores
    - i.e., in "13, 25, 27, 45" the median = 26
  - When an extreme score is a part of the data set the median will not be the best estimate of the group's performance.
  - Appropriate for use when the data is ordinal.

14

Group Activity: Prepared to teach the following concepts.

- What is the "mean" of a data set?

15

Measures of Central Tendency

- Mean
  - Most frequently used measure of central tendency
  - The arithmetic average of the scores
  - Appropriate for use when the data is interval or ratio.

16

Measures of Central Tendency

- For example…
  - Compute measures of central tendency for the following data set of math standard scores
  - 96, 96, 97, 99, 100, 101, 102, 104, 195
    - Mode = 96
    - Median = 100
    - Mean = 110.6
  - What does each measure of central tendency tell you about the data set?
    - Mode = **most frequently** obtained score
    - Median = **middle point** of obtained range of scores
    - Mean = when a data set includes one or more extreme scores the mean will reflect the average performance of the group as a whole, but not the most typical result.

17

Measures of Variability

- Range
  - The difference between the highest and the lowest score.
  - A quick estimate of variability.

18

Measures of Variability

- Variance
  - The amount of spread among the scores
  - In a data set (35, 25, 30, 40, 30) with a mean of 32 the variance is obtained by doing the following:
    - 35-32 = 3
    - 25-32 = -7
    - 30-32 = -2
    - 40-32 = 8
    - 30-32 = -2
    - Because the sum of these scores is 0, to estimate the variance each number is squared (9+49+4+64+4 = 130)
    - 130/5 (the number of cases) = 26
    - Mathematically, to say the variance is 26 is not a problem, but do we typically deal with squared units (do we ask a clerk if 100 squared $ is enough)?

19

Group Activity: Prepared to teach the following concepts.

- What is the "standard deviation" of a data set?

20

Measures of Variability

- Standard Deviation (SD)
  - The square root of the variance returns the variance to the metric of the obtained score.
  - The most practical estimate of variability.
  - Small SD indicates the scores are close together (little variability)
    - What will this distribution "look" like?
  - Large SD indicates the scores are far apart (large variability)
    - What will this distribution "look" like?
  - If the distribution is normal, over 99% of the obtained scores will fall with in + or – 3 standard deviations from the mean.

21

Once upon a time . . .

One sunny Saturday morning, down by the banks, of the Hankie Pankie. A group of Woodchucks congregated with the intent of competing in the international Woodchuck, wood-chucking competition. These are the results of the pounds of wood-chucked in one 24 hour period.

22

Compute Mean, Median, Mode, Range, Variance, Standard Deviation

- Larry - 95 lbs
- Charles - 100lbs
- Vic - 125 lbs
- Bertha - 85 lbs
- Bunny - 90 lbs
- Chauncy - 95 lbs

23

Measures of Central Tendency and Variability

- Mode: Most frequently occurring
- Median: Point above/below which 50% of scores occur
- Mean: The average of the scores
- Range: The difference between the highest and lowest scores
- Variance: Amount of spread among the scores.
- Standard Deviation: Measure of Variability of a distribution of test scores.

24

Group Activity: Prepared to teach the following concepts.

- What is the “bell shaped curve?”

25

Measures of Shape

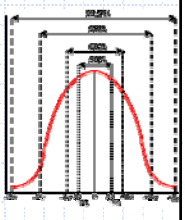
- When population or sample scores on a particular characteristic are graphed, the shape of the “normal curve” resembles a bell.
- The majority of scores fall in the middle (near the mean), and a few scores fall at the extreme ends of the curve.
- The height of a “normal curve” will be determined by the variability of the scores.

26

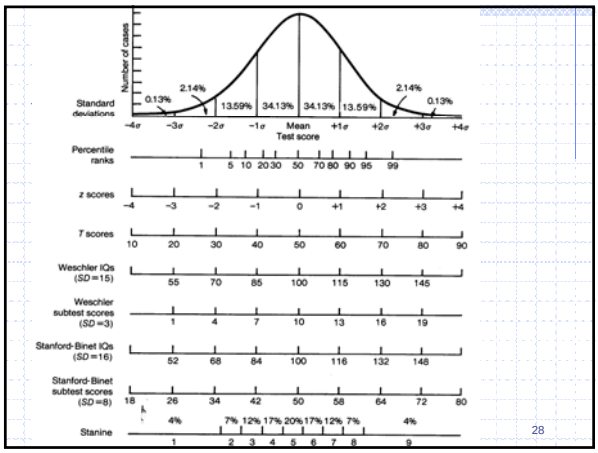
Measures of Shape

The Normal (or bell shaped) Curve

- If a variable is normally distributed it falls in a normal or bell shaped curve.
- Characteristics
  - 50% of scores are above/below the mean
  - Mean, median, mode have the same value (a reason for looking at all three)
  - Most scores are near the mean. Fewer scores are away from the mean.
  - The same number of scores are found + and - a standard deviation from the mean.



27



28

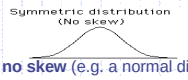
Measures of Shape

Skewed Distributions

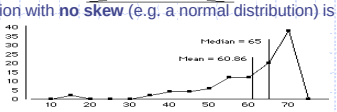
- Assumptions that form the basis for parametric inferential statistical analyses require a normal (or near normal) distributions.
  - Non-parametric statistics are used when the distribution is ***not*** "normal."
- Skewed distributions are asymmetrical (either positively or negatively).
- Examination of the mean, median, and mode will tell you if a distribution is skewed or not.

29

Symmetrical and Asymmetrical Distributions

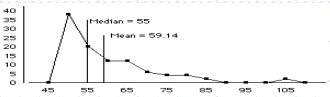


A distribution with **no skew** (e.g. a normal distribution) is symmetrical



Negatively Skewed Distribution

A **negatively skewed** distribution has a longer tail to the left



Positively Skewed Distribution

A **positively skewed** distribution has a longer tail to the right

30

Activity

- Interpret these descriptive statistics

Group	N	Mean Hyp. T-score	Median T-Score	SD	Range
ADHD Hyp	10	72	70	14.17	69 to 90
ADHD Int	10	55	50	22.79	45 to 61
Bipolar Typ 1	10	77	71	14.98	70 to 99

- Assuming these data are normally distributed, what do you think their respective bell shaped curves might look like?

31

Group Activity: Prepared to teach the following concepts.

- What are derived or "standard scores?"

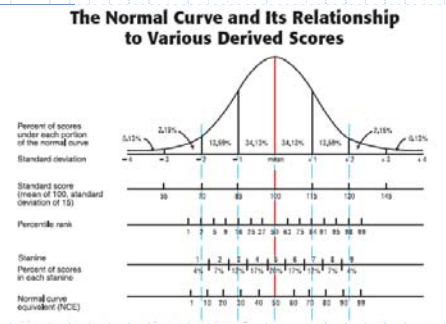
32

Measures of Relative Position

- Percentile Rank
  - Percent of scores that fall at or below a given score
  - Appropriate for ordinal data, typically computed for interval data.
  - Ranks are much closer together at the center of the distribution.
- Standard Scores
  - A derived score that reflects how far a score is from a reference point (typically the mean)
  - *Z*-Scores, # of SDs from the mean.
  - *T* Scores, a *z* score that has been transformed in some way
  - Difference between all scores are equal regardless of location on distribution.

33

Measures of Relative Position



34

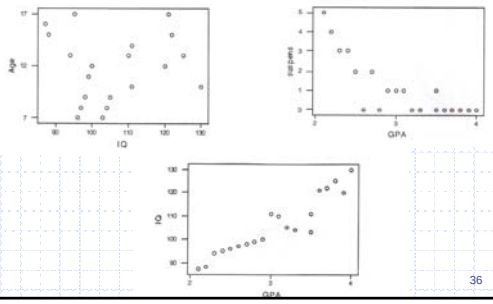
Measures of Relationship

- Correlation
  - Determine whether and to what degree a relationship exists between two or more quantifiable variables
  - Degree of relationship is expressed via the correlation coefficient.

35

Measures of Relationship

- Scatter Plots



36

Activity

- Teaching Descriptive Statistics
  - Mean
  - Standard Deviation
  - Standard Score
  - Bell Shaped Curve

37

April 23

- Data Analysis: Inferential Statistics
  - Read *Educational Research* Chapter 19.
  - Portfolio Element #10 Due:

    Identify resources that will assist you in analyzing data. These resources do not necessarily need to be CSUS resources. Portfolio entries could include student descriptions of the data analysis resources identified. Alternatively, any descriptive handout(s) describing how to locate/use a given resource may be included.

38