Inferential Statistics

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

1

Portfolio Activity #9

- ***Identify data analysis resources.***
  - 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.
  - Discuss in small groups and be prepared to share with the rest of the class.

2

Descriptive Statistics

- Describes data.
- Describes quantitatively 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 represents an entire population, then the data are considered to be ***population parameters***.

3

Inferential Statistics

- If study data represents a population ***sample***, then we will need to make "inferences" about the likelihood the sample data can be generalized to the population.
- Inferential statistics allow the researcher to make a probability statement regarding how likely it is that the sample data is generalizable back to the population.
  - For example…. Is the difference between means real or the result of sampling error?
  - "Inferential statistics are the data analysis techniques for determining how likely it is that results obtained from a sample or samples are the same results that would have been obtained for the entire population" (p. 337)

4

Inferential Statistics

- *"… whereas descriptive statistics show how often or how frequent an event or score occurred, inferential statistics help researchers to know whether they can generalize to a population of individuals based on information obtained from a limited number or research participants"*
  - Gay et al., (2012, p. 341)

5

Inferential Statistics

- **Do not** "prove" beyond any doubt that sample results are a reflection of what is happening in the population.
- **Do** allow for a probability statement regarding whether or not the difference is real or the result of sampling error.

6

Activity: Inferential Statistics

- To make my discussion more concrete, in small groups …
  - Identify a population
  - Discuss how to select a sample
  - Determine how to divide the sample into 2 groups
  - Identify an IV and a DV
  - Indicate what the use of inferential statistics will allow you to do
  - We will use these designs throughout class

7

Basic Concepts Underlying Inferential Statistics

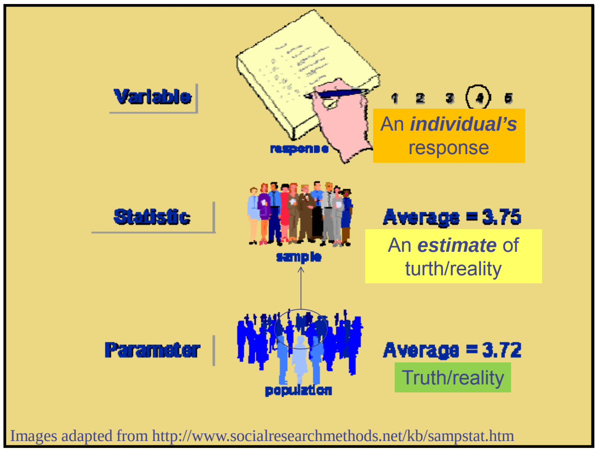
- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

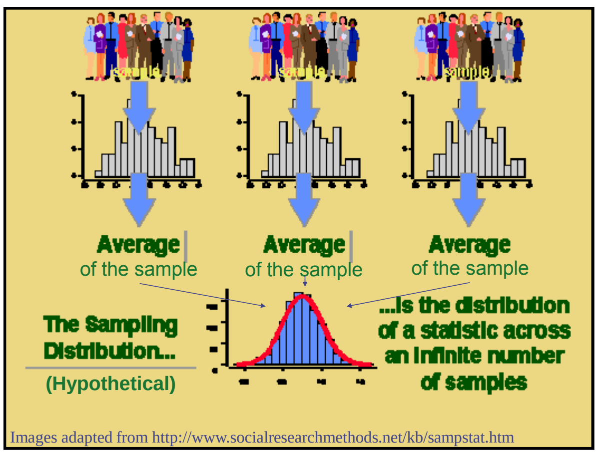
8

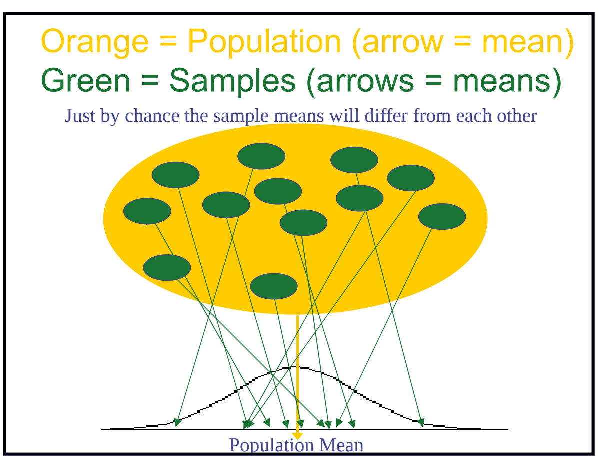
Basic Concepts Underlying Inferential Statistics

- Standard Error
  - Samples are virtually never a perfect match with the population (i.e., identical to population parameters).
  - The variation among the sample means drawn from a given population, relative to the population mean, is referred to as ***sampling error***.
  - The variation among an infinite number of sample means, relative to the population mean, typically forms a normal curve.
  - The standard deviation of the distribution of sample means is usually called the standard error of the mean.
  - Smaller standard error scores indicates less sampling error.

9


Variable
1 2 3 4 5
An individual's response
response
Statistic
Average = 3.75
An estimate of turth/reality
sample
Parameter
Average = 3.72
Truth/reality
population
Images adapted from http://www.socialresearchmethods.net/kb/sampstat.htm


Average
of the sample
Average
of the sample
Average
of the sample
The Sampling Distribution...
(Hypothetical)
...is the distribution of a statistic across an infinite number of samples
Images adapted from http://www.socialresearchmethods.net/kb/sampstat.htm


Orange = Population (arrow = mean)
Green = Samples (arrows = means)
Just by chance the sample means will differ from each other
Population Mean

100 Sample Means

64	82	87	94	98	100	103	108	113	121
67	83	88	95	98	100	103	108	114	111
68	83	88	96	98	100	104	109	115	123
70	84	89	96	98	101	104	109	116	124
71	84	90	96	98	101	105	110	117	125
72	84	90	97	99	101	105	110	117	127
74	84	91	97	99	102	106	111	118	130
75	85	92	97	99	102	106	111	119	131
75	86	93	97	99	102	107	112	119	136
78	86	94	97	99	103	107	112	120	142

100 samples of 20 7th grade CA students on the WJIII Broad Reading Cluster yielded the following means

13

100 Sample Means

- Median, 99.5
- Mode, 97
- Mean, 100.04
- Standard Deviation, 15.6
  - AKA Standard Error of the Mean
  - 68% of the time sample means will be ?
  - 95% of the time sample means will be?

14

100 Sample Means

- The standard error of the mean can be ***estimated*** from the standard deviation of a single sample using this formula

$$SE_{\bar{x}} = \frac{SD}{\sqrt{N - 1}}$$

As sample size goes up, sampling error goes down. WHY???

15

Basic Concepts Underlying Inferential Statistics

- Standard Error
  - Small group discussion
    - How might sampling error have affected the conclusions drawn from your study?

16

Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

17

Basic Concepts Underlying Inferential Statistics

- Null Hypothesis (H_o)
  - A statement that the obtained differences (or observed relationships) being investigated are not significant (e.g., the observed sample mean differences are in fact just a chance occurrence).
  - In other words, the findings are not indicative of what is really going on within the population (the differences are due to sampling error)
  - Stating: "***The null hypothesis was rejected.***"
    - Is synonymous with: ***"The differences among sample means are big enough to suggest they are likely real and not chance occurrences."***

18

Basic Concepts Underlying Inferential Statistics

- Null Hypothesis (H_o)
  - **Small group discussion:**
    - What is the Null Hypothesis for the studies you just constructed?
    - If you conclude that the Null Hypothesis should be rejected what does it mean?
    - To test a null hypothesis you will need a test of significance (and a selected probability value).

19

Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

20

Basic Concepts Underlying Inferential Statistics

- Tests of Significance
  - What does this mean?
  - $t = 7.3$, $df = 105$, $p = .03$

21

Basic Concepts Underlying Inferential Statistics

- Tests of Significance
  - The inferential statistic that allows the researcher to conclude if the null hypothesis should or should not be rejected.
  - A test of significance is usually carried out using a pre-selected significance level (or alpha value) reflecting the chance the researcher is willing to accept when making a decision about the null hypothesis
    - Typically no greater than 5 out of 100.
  - Is a "significant" difference always an "important" difference????

22

Basic Concepts Underlying Inferential Statistics

- Tests of Significance
- Small group discussion:
  - What are the stakes involved in your study?
    - In other words, what will happen if you are wrong (i.e., you conclude your IV has an effect when it really does not)?
    - Does it out weigh the benefits of being right?

23

Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

24

Basic Concepts Underlying Inferential Statistics

- Type I Error
  - Incorrectly concluding that the null hypothesis should be rejected (i.e., concluding that a finding is significant or not likely a chance occurrence) **when in fact it reflects a chance sampling error**.
- Type II Error
  - Incorrectly concluding that the null hypothesis should be accepted (i.e., concluding that the finding is a chance sampling error, or not significant), **when in fact it reflects a real difference within the population** being sampled

25

Basic Concepts Underlying Inferential Statistics (H_o = Null Hypothesis)

		Decision made by the researcher	
		Accept H_o	Reject H_o
Truth/ Reality	Difference is not chance (it is "significant")	Type II Error Saying there is no relationship, difference, gain, when there in fact is such.	Correct Decision
	Difference is a chance occurrence resulting from sampling error (not "significant")	Correct Decision	Type I Error Saying there is a relationship, difference, gain, when in fact such does not exist.

We should keep the risk of Type 1 Error small if we cannot afford the risk of wrongly concluding that the IV has an effect within in the population.

26

Basic Concepts Underlying Inferential Statistics

- In small groups discussion:
  - In your study what concerns you the most: making a Type I or a Type II error?
  - Why (should be connected to the prior discussion)?

27

Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

28

Basic Concepts Underlying Inferential Statistics

- Levels of Significance
  - Reflects the chance the researcher is willing to take of making an incorrect decision about the obtained result (i.e., that the result was due to sampling error).
  - There are a variety of tests of significance (e.g., *t*-test, *F*-test, chi-square).
  - As a rule the larger the score on a given test, the greater the likelihood that the result is significant (i.e., not a chance occurrence, not a reflection of sampling error, or an indication that the null hypothesis should be rejected).

29

Basic Concepts Underlying Inferential Statistics

- Level of Significance
  - For every test, the researcher must select a minimum value that the statistical test must exceed to be regarded as significant.
  - Generally, the larger the sample size the smaller the test score must be to reach statistical significance. [Why is this the case?]
  - A level of significance (or *alpha* ["α"]) value of .05 ($p<.05$) means that the researcher is willing to accept a 5% chance of making a Type I error.
    - In other words, the researcher would be 95% sure that the difference or relationship observed is not a chance occurrence and can reasonably be generalized to the population (i.e., it is not due to "sampling error").

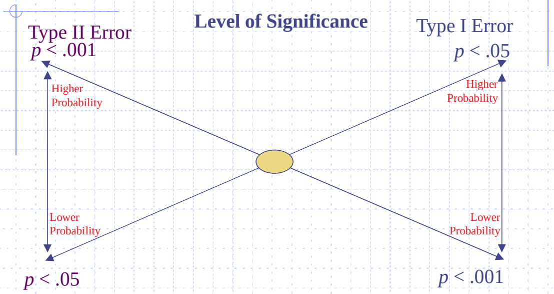
30

Basic Concepts Underlying Inferential Statistics

- Levels of Significance
  - Reducing the probability of making a ***Type I*** error, by increasing the level of significance required to reject the null hypothesis (e.g., from .05 to .01) , increases the probability of making a Type II error.
  - Reducing the probability of making a ***Type II*** error, by decreasing the level of significance required to reject the null hypothesis (e.g. from .05 to .10), increases the probability of making a Type I error.

31

Basic Concepts Underlying Inferential Statistics



Type II Error: Accepting an invalid null hypothesis (i.e., rejecting real differences)

Type I Error: Rejecting a valid null hypothesis (i.e., accepting differences due to sampling error)

32

Basic Concepts Underlying Inferential Statistics

- Levels of Significance
  - Small group discussion:
    - What is the level of significance you are going to select in your study and why?
    - p = .10
    - p = .05
    - p = .01
    - p = .001

33

Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

34

Basic Concepts Underlying Inferential Statistics

- Practical Significance
  - Reflects the possibility that a statistically significant finding may be unimportant
    - A generalizable but very small difference
    - A difference so small it is practically **in**significant.
  - *Effect size* (ES) reflects how many standard deviation scores the obtained findings are apart.
  - An ES of .33 or more is typically used to determine if the difference is meaningful.
    - In other words, one third of a SD difference is typically considered important or practically significant.

35

Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

36

Basic Concepts Underlying Inferential Statistics

- Two- & One-tailed test
  - Tests of significance can be either one- or two-tailed (Two-tailed is most common).
  - If it is hypothesized that the difference or relationship will only occur in one direction (you have a specific directional hypothesis) then use a one-tailed test.
    - A smaller difference (exactly half) will be required to be considered significant if you use a one-tailed test.
  - However, if it is possible for the difference or relationship to go either way, then use a two-tailed test.
    - A bigger difference will be required to be considered significant if you use a two-tailed test.

37

Basic Concepts Underlying Inferential Statistics

- Two- & One-tailed test
  - Examples:
    - One-tailed test [in each case, the null hypothesis (indirectly) predicts the direction of the difference]
      - Females will not score significantly higher than males on an IQ test.
      - Blue collar workers are will not buy significantly more product than white collar workers.
      - Superman is not significantly stronger than the average person.
    - Two-tailed test (the two-tailed probability is exactly double the value of the two-tailed probability)
      - There will be no significant difference in IQ scores between males and females.
      - There will be no significant difference in the amount of product purchased between blue collar and white collar workers.
      - There is no significant difference in strength between Superman and the average person

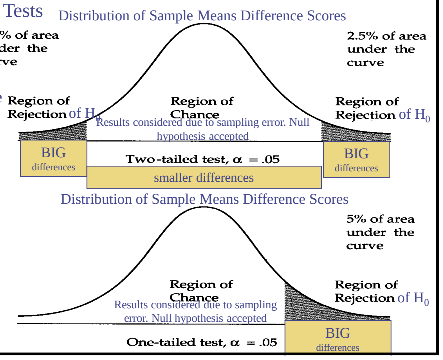
from http://www.statpac.com/surveys/statistical-significance.htm

38

Basic Concepts Underlying Inferential Statistics

Two & One Tailed Tests

1. All possible difference or relationship scores
2. Extreme scores are considered significant
3. Differences could be + or -
4. If the difference (or relationship) could only be positive …
5. If on the other hand your hypothesis is non-directional …



Basic Concepts Underlying Inferential Statistics

- Standard error of the mean
- Null Hypothesis (H_o)
- Tests of Significance
- Type I and Type II Errors
- Levels of Significance
- Practical Significance
- Two- & One-tailed Tests
- Degrees of Freedom

40

Basic Concepts Underlying Inferential Statistics

- Degrees of Freedom
  - "Suppose we ask you to name any five numbers. You agree and say '1, 2, 3, 4, 5.' In this case *N* is equal to 5, you had 5 choices and you could select any number for each choice. In other words, each number was 'free to vary;' it could have been any number you wanted. Thus, you had 5 degrees of freedom for your selection (*df* = N). Now suppose we tell you to name 5 numbers and you begin with '1, 2, 3, 4, …' and we say 'Wait! The mean of the five numbers you choose must be 4.' Now you have no choice - your last number must be 10 to achieve the required mean of 4 (i.e., 1 + 2 + 3 + 4 + 10 = 20/5 – 4). That final number is not free to vary; in the language of statistics, you lost one degree of freedom because of the restriction that the mean must be 4. In this situation, you only had 4 degrees of freedom (*df* = N – 1)."
    - Gay et al. (2012, p. 350)

41

Types of Tests of Significance

- Parametric Tests
  - Requires that…
    - the DV data represent an interval or ratio scale.
    - the participants be independently selected (random sampling).
    - the variable measured not be extremely skewed.
- Nonparametric Tests
  - Make no assumptions about the shape of the distribution
  - Can be used when the DV data represents a nominal or ordinal scale.
    - When used it is more difficult to reject the null hypothesis, thus if appropriate researchers typically use parametric statistics.

42

The *t* Test

- Assesses the significance of the difference observed between two means
  - Independent Samples *t* test is used when the two samples are randomly formed without any kind of matching
  - Nonindependent Samples *t* test is used to compare groups when they were formed using some type of matching procedure, or when you are looking at a single group's pre- and post-test results.
- Small group discussion: What kind to *t* test will you use in your study?

43

Simple Analysis of Variance

- Simple or "One-Way ANOVA" is used to determine whether there is a significant difference between two or more means.
- Why not just compute a number of different *t* tests?
- An *F* ratio is used to determine if significant differences exist among the means being compared.
- If the *F* ratio is significant, then multiple comparisons are used to determine which means are significantly different from which other means.

44

Factorial Analysis of Variance

- The statistic used when a study employs two or more independent variables.
- Also assesses the interaction observed among the variables.
- Example:
  - IVs = IQ & Reading inst. Method
  - DV = Reading test SS
  - Graph the following two data sets
  - Why is this considered a 2X2 design?

Reading inst. method

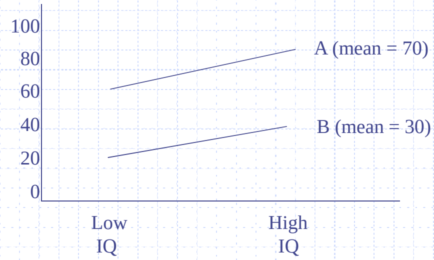
		A	B	
I	High	80	40	60
Q	Low	60	20	40
		70	30	

Reading inst. method

		A	B	
I	High	80	60	70
Q	Low	20	40	30
		50	50	

45

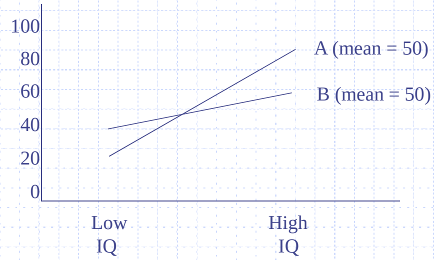
Factorial Analysis of Variance



NOTE: Graph method A and method B with two separate lines.

46

Factorial Analysis of Variance



NOTE: Graph method A and method B with two separate lines.

47

Analysis of Covariance (ANCOVA)

- An example of statistical (vs. experimental) control.
  - Matching is an example of experimental control.
- Because ANCOVA can reduce random sampling error by equating different groups, it increases the power of the significance test (the test's ability to reject the null hypothesis).

48

Chi Square

- A test of significance used when the data are in the form of frequency counts or percentages and proportions that can be converted into frequencies.
- Appropriate for use when using nominal both IV and DV data that is either a true category (male/female) or an artificial category (tall/short).

49

Digit Naming Speed Test

32533 13586 39292 64894 91665 68953 19645 15953 38311 28659

From A

68248 83542 99634 91826 61368 34113 65481 16544 35635 45318

Form B

50

Data Analysis: DNS ADHD vs. non-ADHD

	C1	C2	C3	C4	C5	C6
	S num	ADD DNT	C DNT	DNSDiff	DNSRank	signtest
1	1	19.50	18.25	1.25	6	+
2	2	32.95	24.08	8.87	15	+
3	3	24.35	21.65	2.70	8	+
4	4	20.70	16.50	4.20	12	+
5	5	15.50	19.53	-4.03	2	-
6	7	32.90	20.40	12.50	18	+
7	8	22.30	25.60	-3.30	3	-
8	9	26.60	17.80	8.80	14	+
9	10	20.75	25.15	-4.40	1	-
10	11	23.50	21.30	2.20	7	+
11	12	22.45	18.45	4.00	11	+
12	13	33.47	23.02	10.45	16	+
13	14	29.37	22.01	7.36	13	+
14	15	35.40	19.99	15.41	19	+
15	16	23.50	23.41	0.09	4	+
16	17	60.50	24.62	35.88	20	+
17	*	*	*	*	*	0
18	19	30.65	18.25	12.40	17	+
19	20	28.25	25.27	2.98	10	+
20	22	25.40	24.95	0.45	5	+
21	23	23.60	20.65	2.95	9	+

51

Next week

- Qualitative Research: Overview, Data Collection/Analysis, Narrative and Ethnographic Research
- Read Educational Research Chapters 13, 14 16, 20 & 21.

52