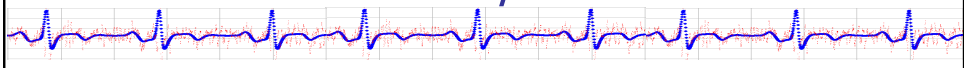
Empirical Research Methods in Information Science

IS 4800 / CS 6350



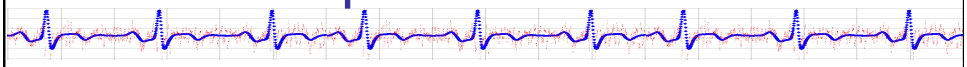
Lecture 16

Data Preparation

Reporting Results

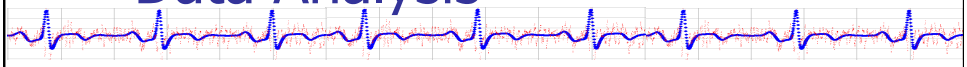
1

Data Preparation



2

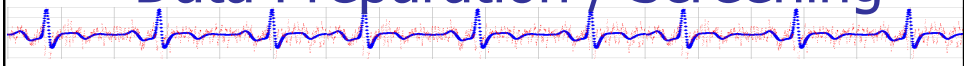
Data Analysis



1. Data Preparation
2. Baseline Analysis
3. Primary Outcome Analysis
4. Secondary Outcome Analysis
5. Subgroup Analysis (and other "fishing")

3

Data Preparation / Screening



- After collecting data
  1. Check that all data has been correctly recorded, no erroneous datasets, etc.
  2. Deal with missing data.
  3. Deal with outliers (review, transform, drop)
  4. Check each interval/ratio sample (measure) to see if normal.
  5. "Fix" measures that are not normal.
- Before running any statistics
- Policies should be defined in advance (proposal).

4

R: tests for normality

```
#Plot Q-Q line
>qqline(DV)

#Shapiro-Wilk Normality test
#Probability my data came from a
# normal distribution
>shapiro.test(DV)
```

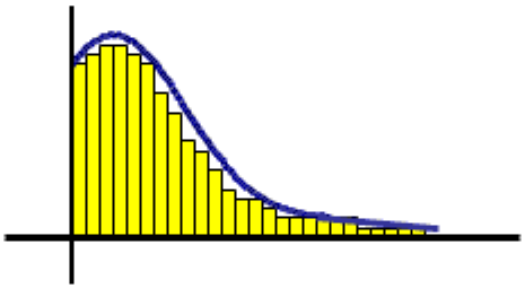
5

When to do transformations

- When underlying population does not meet assumptions for the tests you are doing
  - For parametric tests:
    - Populations are normal
    - Have equal variances (for multiple group comparisons)
- Causes (some)
  - Ceiling/Floor effects
  - Outliers
- Have to assess this from samples
  - Looking at histograms (unimodal, symmetric)
  - Statistics (e.g., kurtosis)

6

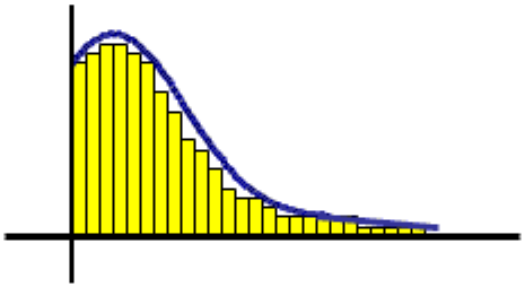
OK, so you just spent six months running your study and your primary outcome measure looks like this…



- What do you do?

7

What do you do?



- Data transformations
- Non-parametric tests

8

Data Transformations

- Applying a function to <u>every</u> data point (for a given measure) in your samples.

9

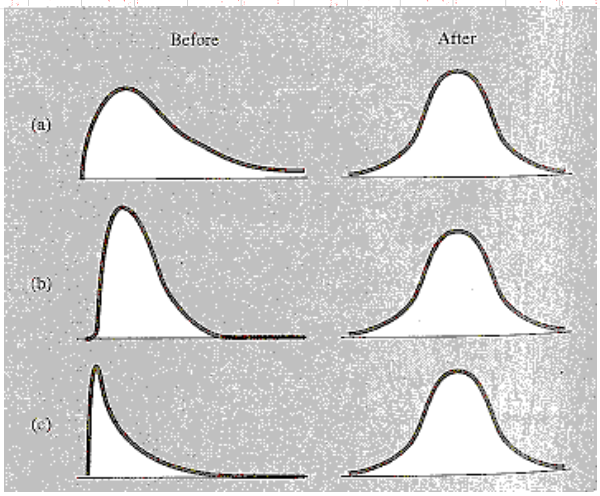
Data Transformations

To fix unimodal but skewed distribution

Moderate right skew –
Use square-root transform

Strong right skew –
Use log transform

Extreme right skew –
Use inverse transform
and reflect results

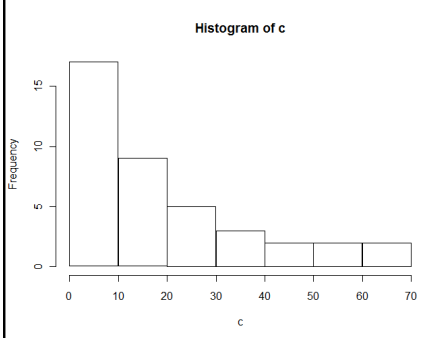


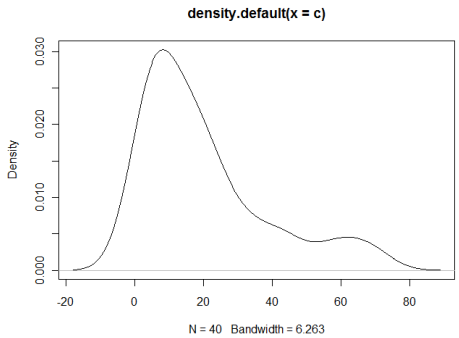
Data Transformations When Justified?

- When absolute values do not matter
  - Measure is dimensionless
    - E.g. scores on composite Likert questionnaire
  - Measure indirectly assesses underlying factor of interest
- If this is not true, you can still do parametric inferential stats on transformed data, but be careful in reporting descriptive stats.
- Must apply transform to all data points equally

12

Example: mydata

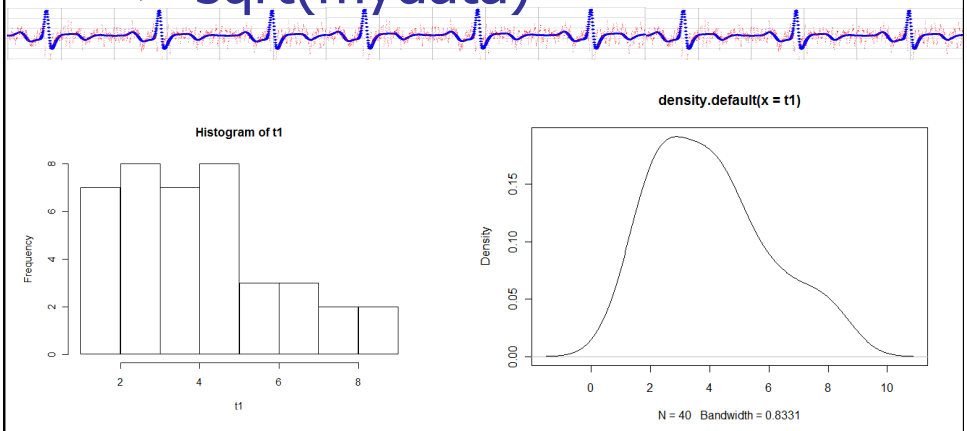




Shapiro-Wilk normality test
W = 0.8417, p-value = 5.693e-05

13

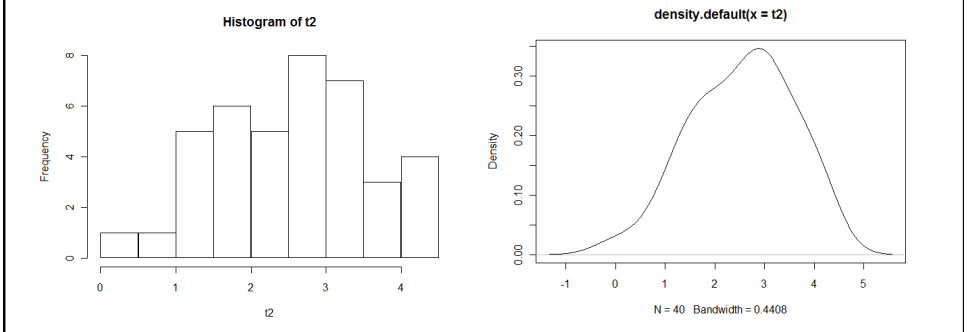
> sqrt(mydata)



Shapiro-Wilk normality test
W = 0.948, p-value = 0.06461

14

> log(mydata)



Shapiro-Wilk normality test
W = 0.9782, p-value = 0.6241

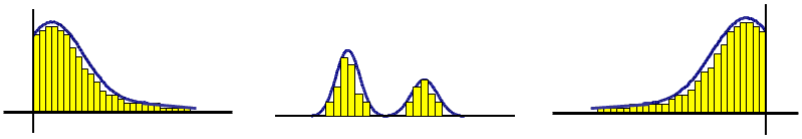
15

Non-parametric statistics

- Make no assumption about underlying population distribution.
- If your data does meet assumptions for parametric statistics (t-test etc), then non-parametric stats provide less power.
- Generally operate on rank order-transformed data.

16

Rank-Order Transform



17

Rank-Order Transform

- Another kind of data transformation.
  - For non-uni-modal distributions or distributions with extreme outliers, or for ordinal measures
- Transform
  - Take all data from a given measure (from all groups), sort and assign ranking (ignore ties).
  - Then, replace the original values with the corresponding rank.

18

Rank-Order Transform Example

- DV Measure: Satisfaction
  - With sign in food court: 5.5, 7.1, 6.2, 6.7
  - Without sign: 2.3, 3.4, 1.7, 5.2
- Rankings:

1.7	2.3	3.4	5.2	5.5	6.2	6.7	7.1
1	2	3	4	5	6	7	8

- Resulting data for comparison:
  - With sign in food court: 5, 8, 6, 7
  - Without sign: 2, 3, 1, 4

19

Rank-Order Tests

Parametric Test	Rank-Order Test
Pearson correlation	Spearman's rho *Just compute Pearson on the ranks*
t test for independent means	Wilcoxon rank-sum test -aka- Mann-Whitney U test
t test for dependent means	Wilcoxon signed rank test

Also use these tests on non-transformed data when measure is naturally rank ordered to begin with

20

Mann-Whitney U test aka **Wilcoxon rank-sum test**

- H0: two samples are drawn from a single population, and therefore their probability distributions are equal
- For interval or ordinal data
- "U" statistic
  - "sample 1" = the sample for which the ranks seem to be smaller (other sample is "sample 2")
  - For each observation in sample 1, count the number of observations in sample 2 that are smaller than it (count a half for any that are equal to it).
  - The total of these counts is U.

21

Critical values for U test

Nondirectional α=.05

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12
1	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1
3	-	-	-	-	0	1	1	2	2	3	3	4
4	-	-	-	0	1	2	3	4	4	5	6	7
5	-	-	0	1	2	3	5	6	7	8	9	11
6	-	-	1	2	3	5	6	8	10	11	13	14
7	-	-	1	3	5	6	8	10	12	14	16	18
8	-	0	2	4	6	8	10	13	15	17	19	22
9	-	0	2	4	7	10	12	15	17	21	23	26
10	-	0	3	5	8	11	14	17	20	23	26	29
11	-	0	3	6	9	13	16	19	23	26	30	33
12	-	1	4	7	11	14	18	22	26	29	33	37

22

Example, continued

- Resulting data for comparison:
  - With sign in food court: 5, 8, 6, 7
  - Without sign: 2, 3, 1, 4
- U = 0 + 0 + 0 + 0 = 0
- Lookup n1=4, n2=4
- Or use R…

23

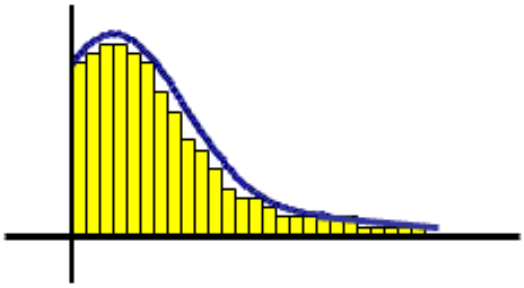
Mann-Whitney U test

```
> wilcox.test(nonsmokers,smokers)
Wilcoxon rank sum test with continuity
correction
data: nonsmokers and smokers
W = 76.5, p-value = 0.04715
alternative hypothesis: true location
shift is not equal to 0
```

Also…

> wilcox.test(d$outcome~d$condition)

What do you do?
Review



- Data transformations
- Non-parametric tests

25

Data Analysis

1. Data Preparation
2. Baseline Analysis
3. Primary Outcome Analysis
4. Secondary Outcome Analysis
5. Subgroup Analysis (and other "fishing")

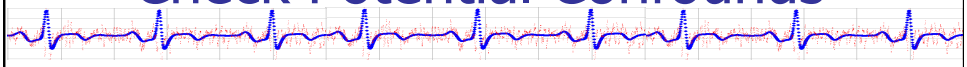
26

Baseline Analysis
aka Check Potential Confounds

- For between-subjects studies –
  - Primary purpose: show that randomization worked
  - Goal: no significant differences between randomized groups on non-outcome measures

27

Check Potential Confounds



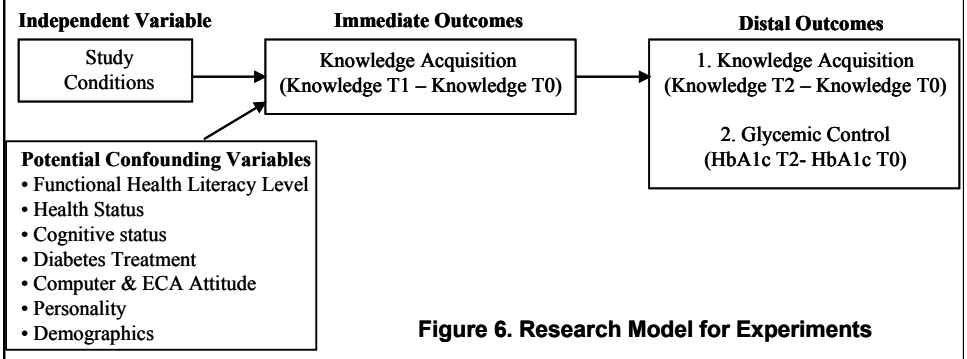
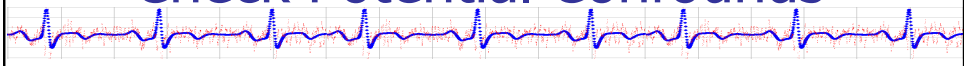


Figure 6. Research Model for Experiments

28

Check Potential Confounds



- Ensure potential confounds are balanced across study conditions
  - Shows quality of randomization
  - t-test or chi-sq TFI to see if significantly different
- If significantly different…
  - May want to "correct" the data prior to analysis
    - Add as covariate in ANCOVA, or consult statistician
  - Contentious whether adjustments should even be made

29

Sample Baseline Analysis Report

2 treatment between-subjects design

Table 1. Baseline Measures

Variable	All, N = 263	Control, n = 131	Intervention, n = 132
Female, n (%)	161 (61.2)	72 (55.0)	89 (67.4)
Age at enrollment, mean ± SD	71.3 ± 5.4	70.8 ± 5.2	71.7 ± 5.6
Race, n (%)			
Black	165 (62.7)	86 (65.6)	79 (59.8)
White	75 (28.5)	35 (26.7)	40 (30.3)
Other	23 (8.7)	10 (7.6)	13 (9.8)
Hispanic, n (%)	20 (7.6)	10 (7.6)	10 (7.6)
Education			
<High school	56 (21.3)	30 (22.9)	26 (19.7)
High school	79 (30.0)	34 (26.0)	45 (34.1)
>High school	128 (48.7)	67 (51.1)	61 (46.2)
Married, n (%)	87 (33.1)	43 (32.8)	44 (33.3)

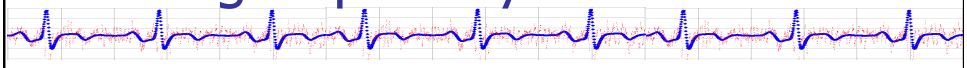
30

Data Analysis

1. Data Preparation
2. Baseline Analysis
3. Primary Outcome Analysis
4. Secondary Outcome Analysis
5. Subgroup Analysis (and other "fishing")

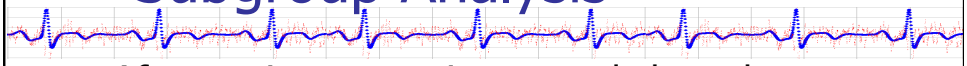
31

Subgroup Analysis



32

Subgroup Analysis



- After testing your primary study hypotheses…
- And testing your explicit secondary hypotheses..
- Do subgroup analyses
  - "Among which group of participants does the intervention work the best?"
  - How did it work for men vs. women ("gender effects")?
  - Divide the data into smaller groups (demographic, etc) and re-run descriptive & inferential statistics
  - Exploratory data analysis, aka "fishing"
- Note: you will have lower power

33

Multiple Outcome Analyses

- At the .05 significance level, 1-in-20 tests will be significant by chance
- To be conservative, if you are doing *k* tests at α=.05, you should do your power analysis for a α/k significance level (Bonferroni correction).
- More practically:
  - Plan your power analysis for your one primary hypothesis
  - Treat all other comparisons as tentative results

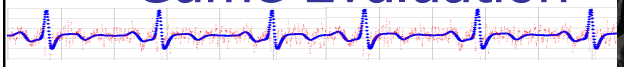
34

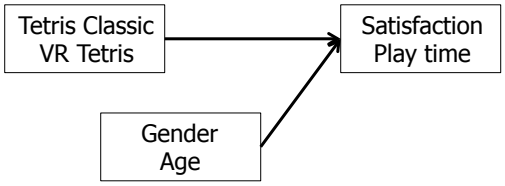
Data Analysis Summary

1. Check data for missing values, outliers & fix.
2. Check distributions of each measure for normality.
   - If outliers, determine if they can be dropped.
   - Apply transforms if necessary and justified.
3. Baseline analysis
   1. Check for possible confounds (aka sample bias, for all demographics and anything else that seems plausible).
4. Conduct primary analyses (descriptives and hypothesis tests).
5. Conduct secondary analyses
6. Conduct sub-group analyses and other secondary tests.

35

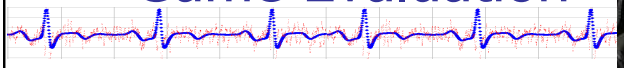
Example Game Evaluation

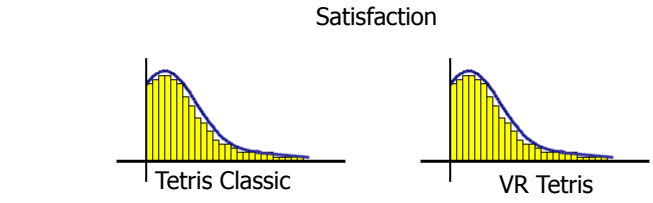


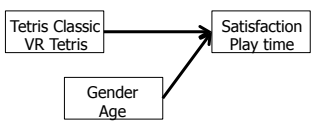


36

Example Game Evaluation

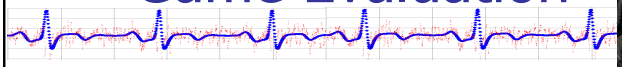




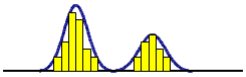


37

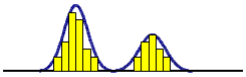
Example Game Evaluation



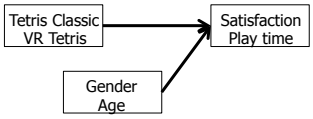
Play Time



Tetris Classic

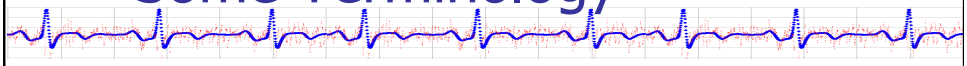


VR Tetris



38

Some Terminology



- Adherence/Compliance – the degree to which subjects follow your instructions
- Dose-Response – relationship between amount of intervention (e.g., days of software training) and outcome (e.g., proficiency)
- Effectiveness – how well your intervention works if subjects follow your instructions exactly
- Efficacy – how well your intervention works in the real world, taking nonadherence into account
- Intent-to-treat: include all Ss in outcome, regardless

39

Mo Terminology

- Longitudinal study – takes place over time, with measurements taken at multiple time points
- Cross-sectional study – takes place at one point in time, but associate outcome measure with a chronological measure (e.g., age, years experience)
- Retention – fraction of subjects who remain in a longitudinal study
- Attrition – opposite of retention

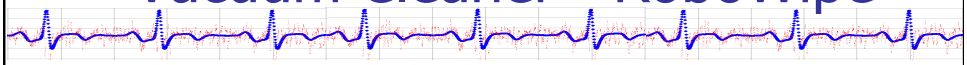
40

Example: Tomorrow's Technology Tonight

- You are head of QA.
- Engineers have just designed a new PC input device that lets users move their cursor using nasal sounds.
- You are put in charge of evaluating their claim that this is a faster way of controlling a PC than other conventional methods. You run a series of standardized tests with random users measuring user time to click on a target (in ms) comparing NasalPoint to a mouse.
- You also assess the degree to which subjects' feel silly using the product using a composite index survey.
- You suspect that people with large noses may have difficulty using the product, so you also record the size of subjects' noses (mm base to tip).
- Your final dataset for the silliness measure indicates a significant floor effect.

42

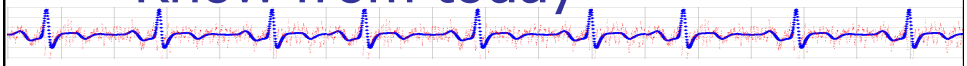
Example: RobotsRUs Robotic Vacuum Cleaner - RoboWipe



- You are head usability engineer and want to conduct an assessment of the effectiveness of your new robotic toilet cleaner.
- You provide a random sample of households with RoboWipe for a month and enlist another random sample of households as a control group.
- You conduct a chemical analysis of their toilet water, measuring bacteria in parts-per-million, at the end of the month, for all households.
- You are particularly interested in whether younger adults are better able to use the robot, compared to older adults, so you record the average age of adults in each household and use this to split the datasets into "older households" and "younger households".
- At the end of the study you discover that three of the households that had been given the robots never turned them on.
- After collecting all of your data you find that the bacteria measure for the households with the robot has a clear bimodal distribution.

43

Know from today



- Major steps in data analysis
- When to transform, types of transforms
  - Including rank-order
- Non-parametric tests, when to use

44

Exercise:
Do a baseline analysis!

Table 1. Baseline Measures

Variable	All, N = 263	Control, n = 131	Intervention, n = 132	P
Female, n (%)	161 (61.2)	72 (55.0)	89 (67.4)	n.s.
Age at enrollment, mean ± SD	71.3 ± 5.4	70.8 ± 5.2	71.7 ± 5.6	<.05
Race, n (%)				n.s.
Black	165 (62.7)	86 (65.6)	79 (59.8)	
White	75 (28.5)	35 (26.7)	40 (30.3)	
Other	23 (8.7)	10 (7.6)	13 (9.8)	
Hispanic, n (%)	20 (7.6)	10 (7.6)	10 (7.6)	n.s.
Education				n.s.
<High school	56 (21.3)	30 (22.9)	26 (19.7)	
High school	79 (30.0)	34 (26.0)	45 (34.1)	
>High school	128 (48.7)	67 (51.1)	61 (46.2)	n.s.
Married, n (%)	87 (33.1)	43 (32.8)	44 (33.3)	

45

Presenting Research Results

46

Written Study Reports

- Ignore most of the stuff on APA style (fonts, etc.)
- Information on paper structure very important & relevant

47

Written Study Reports

- Objectives (also critiques)
  - Describe what your study is about
  - Motivate your study
  - Assure reader you have conducted a sound study
    - Research Methods – often presented in small font
  - Present results in an objective manner
  - Discuss implications
  - Discuss future work
  - Enable replication

48

Typical Study vs. IS/CS/HCI Paper Structure

- Abstract
- Introduction
  - Motivation
  - Related work
  - Hypotheses
- Method
- Results
- Discussion
  - Limitations
  - Implications
  - Future work
- References

- Abstract
- Introduction
  - Motivation
- Related work
- System design
- Evaluation
  - Hypotheses
  - Method
  - Results
  - Discussion – summary, limitations
- Conclusion
  - Implications
  - Future work
- References

49

The Abstract

- Concise summary
- Abstract for an empirical study should include
  - Information on the problem under study
  - The nature of the subject sample
  - A description of methods, equipment, and procedures
  - A statement of the results
  - A statement of the conclusions drawn
- Often the last thing you write

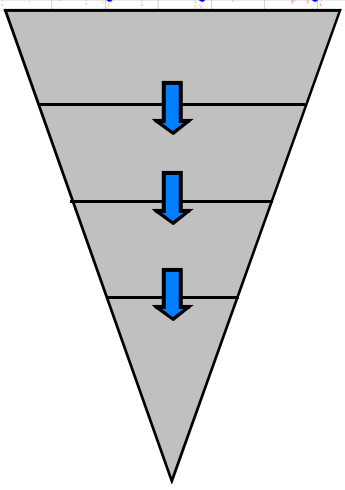
50

The Introduction

- Part of paper giving justification for study
- Usually has the following information
  - Introduction to the topic under study
  - Brief review of research and theory related to the topic
  - A statement of the problem to be addressed
  - A statement of the purpose of the research
  - A brief description of the research strategy
  - A description of predictions and hypotheses
- CS/IS papers often put Related Work as a separate section after Introduction
  - For each, describe how your work is different

51

Organization of the Introduction: General to Specific



Present a general introduction to your topic

Review relevant literature

Link literature review to your hypotheses

State your hypotheses

52

The Method Section

- Includes information on exactly how a study was carried out
- Subsections
  - Participants or subjects
    - Describe in detail the participant or subject sample
    - Human participants go in a *Participants* subsection, and animal subjects in a *Subjects* subsection
  - Apparatus or materials
    - Describe in detail any equipment or materials used
    - Equipment is usually described in an *Apparatus* subsection and written materials in a *Materials* subsection

53

The Method Section

- Procedure
  - Describe
    - Exactly how the study was carried out
    - The conditions to which subjects were exposed or under which observed
    - The behaviors measured and how they were scored
    - When and where observations were made
    - Debriefing procedures
- Enough detail should be included in all sections so that the study could be replicated

54

The Results Section

- Objective, dry, boring – *just the facts*
- All relevant data and analyses are reported in the results section
- Do not present raw data
- Data should be reported in summary form
  - Descriptive statistics
  - Inferential statistics
- Results of descriptive and inferential statistics must be presented in narrative format
- Describe the source of any unconventional statistical tests

55

Commonly Used Statistical Citations

Statistical Test	**Format**
Analysis of variance	$F(1,85) = 5.96, p < .01$
Chi-square	$\chi^2(3) = 11.34, p < .01$
t test	$t(56) = 4.78, p < .01$
Pearson correlation coefficient	$r = -.87, p<.05$

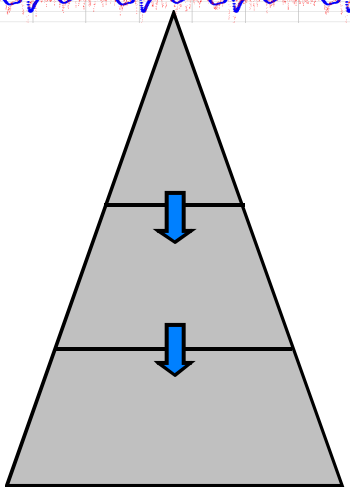
56

The Discussion Section

- This is where you can take some liberties with describing what the results *mean*
- Results are interpreted, conclusions drawn, and findings are related to previous research
- Section begins with a brief restatement of hypotheses
- Next, indicate if hypotheses were confirmed
- The rest of the section is dedicated to integrating findings with previous research
- It is fine to speculate, but speculations should not stray far from the data

58

Organization of Discussion: Specific to General



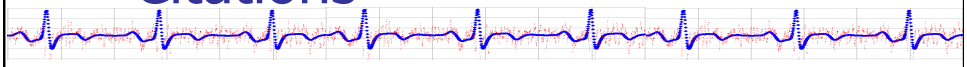
Restate your hypotheses or major finding

Tie your results with previous research and theory

State broad implications of your results, methodological implications, directions for future research

59

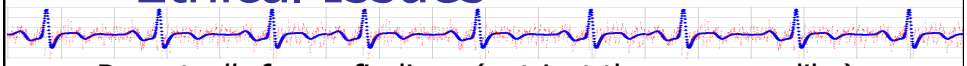
Citations



- Liberally cite previous & related work.
- If you copy passages you *must* cite and, depending on length, format to indicate it is copied.
- Suggest using EndNote, BibTex or similar.

73

Ethical Issues



- Report *all* of your findings (not just the ones you like)
- Adhere to your original plan
  - Report any deviations and why
  - Power analysis, statistics, measures
- Do not drop subjects or data points without rigorous justification
- If your hypothesis test was not significant *you cannot say anything about difference in means (example).*
- If you did not do an experiment, attempting to control for extraneous variables, *you cannot mention or imply causality.*

74

Oral Presentation of Study Results

Oral Presentation

- Main concepts and ideas
- Do *not* go into great detail on experimental methods – just enough so people understand roughly what you did
- Focus on motivation, results, implications
  - If listener wants details they can read the paper or ask questions

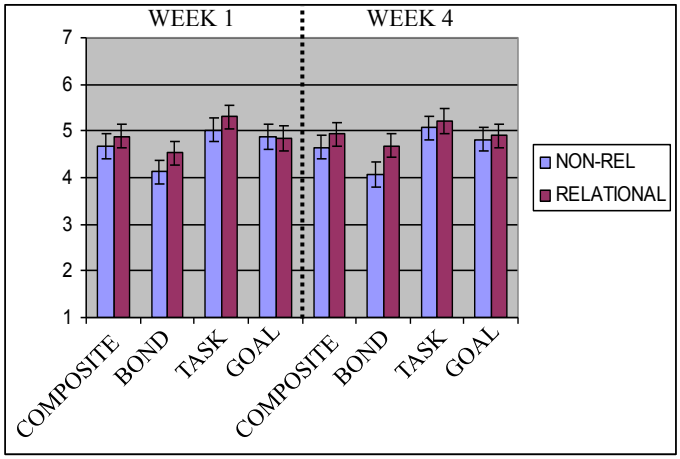
76

Oral Presentation
Don't do this…

Measure	Change From Day1	To Day2	ALL CONDS df	t	p	CONTROL df	t	p	NON-REL df	t	p	RELATIONL df	t	p
WAI/COMP	7	27	54	0.205	0.838				24	0.014	0.989	29	0.361	0.720
WAI/BOND	7	27	54	0.519	0.606				24	0.376	0.710	29	1.489	0.147
WAI/TASK	7	27	54	0.134	0.894				24	0.409	0.686	29	0.661	0.514
WAI/GOAL	7	27	54	0.155	0.877				24	0.081	0.936	29	0.329	0.745
CONTINUE LAURA	30	44	54	0.868	0.389				24	0.625	0.538	29	0.619	0.541
MIN/DAY	-6-0	22-30	81	1.470	0.145	26	1.274	0.214	24	0.124	0.903	29	1.104	0.279
	1-7	22-30	81	0.691	0.492	26	0.758	0.456	24	0.109	0.914	29	0.358	0.723
	22-30	38-44	81	3.626	**0.001**	26	2.480	**0.020**	24	1.959	0.062	29	1.804	0.082
DAY/WK>30MIN	-6-0	22-30	81	6.653	**0.000**	26	2.323	**0.028**	24	5.284	**0.000**	29	4.347	**0.000**
	1-7	22-30	81	6.272	**0.000**	26	2.401	**0.024**	24	3.818	**0.001**	29	4.597	**0.000**
	22-30	38-44	81	8.990	**0.000**	26	4.043	**0.000**	24	5.322	**0.000**	29	6.530	**0.000**
STEP/DAY	1-7	22-30	81	1.778	0.079	26	1.197	0.242	24	2.366	**0.026**	29	0.236	0.815
DAY/WK>10KSTEP	1-7	22-30	77	3.986	**0.000**	25	1.355	0.188	23	3.591	**0.002**	27	2.055	**0.050**
STAGE	Intake	30	81	6.988	**0.000**	26	3.403	**0.002**	24	4.000	**0.001**	29	4.738	**0.000**
	30	44	81	2.019	**0.047**	26	1.185	0.247	24	1.000	0.327	29	1.409	0.169
SELF-EFFICACY	1	29	81	4.782	**0.000**	26	0.872	0.391	24	3.314	**0.003**	29	4.750	**0.000**
	29	44	81	2.770	**0.007**	26	1.525	0.139	24	4.550	**0.000**	29	0.085	0.933
PROS	1	29	81	1.998	**0.049**	26	1.418	0.168	24	0.456	0.653	29	1.540	0.134
	29	44	81	0.393	0.695	26	1.147	0.262	24	0.225	0.824	29	0.308	0.760
CONS	1	29	81	0.902	0.370	26	1.124	0.271	24	0.499	0.622	29	0.823	0.417
	29	44	81	0.740	0.462	26	0.386	0.703	24	0.611	0.547	29	0.339	0.737
CONTINUE FT	30	44	81	1.520	0.133	26	1.442	0.161	24	1.163	0.256	29	0.000	1.000

77

Oral Presentation
Do use as many figures as possible



78

Oral Presentation Guide for Visuals

- Visuals should be *exhibits* that you talk about
  - Do not put lots of text on charts
  - Do not read your charts for your presentation
- Use interactivity, video, images to keep your audience awake

79

Common Questions

- How did you evaluate that?
- How did you measure that?
- How did you control for *extraneous variable X*?
- Why didn't you use statistic Y?
- Isn't that a biased sample?
- What was your control group?
- How did you do *study procedure Z?*

80

Outline for Descriptive Study Oral Presentation

- Motivation
- Research Questions
  - Not hypotheses, unless meaningful Chi^2 GoF
- Method
  - Including exact text of any surveys
- Results
  - Demographics
  - Outcomes
  - Lots of visuals!
- Conclusions / Discussion

81

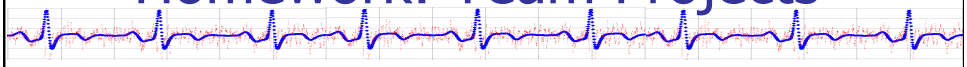
Team Project Presentation Grading

- Worth ~5% of your grade.

- Timeliness
- Oral presentation
- Visuals
- Ability to answer questions
- Use of appropriate stats & charts

83

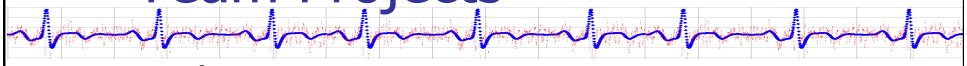
Homework: Team Projects



- Team Project Writeup
  - At least two pages long
  - Outline – as discussed today
  - Raw data and R runs as appendices
  - You must have *some* literature citations

84

Team Projects



- Oral presentation
  - 10 minutes (hard upper bound)
  - 5 minutes critique
  - Main idea, hypotheses, study design, results, conclusions
  - Visualization of data
  - Either
    - Email ppt to me by noon, or
    - Bring memory stick, laptop, put online

85

Tips

- Describe your sample
  - Minimal demographics – number of subjects, broken down by gender
  - Better: age, occupation, major, year
- Minimize text on your charts
- If you use a novel measure (e.g., new survey) you must give details on the measure
  - Actual questions asked
  - Any reliability/validity/psychometrics done
- If you do interviews, include actual quotes
- Build from data to conclusions
- Practice your timing/delivery with your project team

86

Homework

- T1 –
  - Report and Presentation Next Class (3/20)
- Next time:
  - Single-subject designs (B&A Ch 12).

89