Empirical Research Methods in Information Science

IS 4800 / CS 6350

Lecture 17

Data Preparation

1

Data Preparation

4

Data Analysis

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

5

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 <u>roughly</u> normal.
  5. “Fix” measures that are not normal.
- <u>Before</u> running any statistics
- Policies should be defined in your proposal.

6

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)

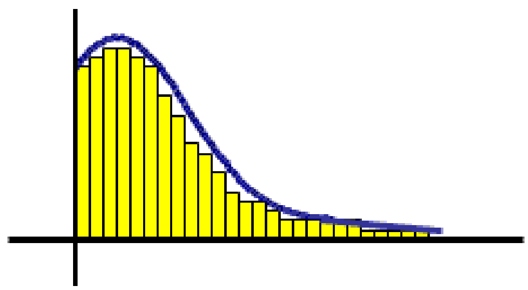
R: tests for normality

```
#Plot density function.
>plot(density(DV))

#Plot Q-Q line
>qqline(DV)

#Shapiro-Wilk Normality test
>shapiro.test(DV)
```

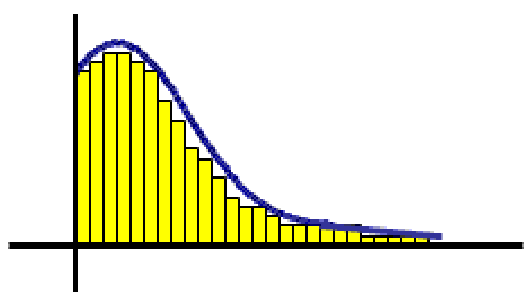
OK, so you just spent six months running your study and your primary outcome measure looks like this...



- What do you do?

9

What do you do?



- Data transformations
- Non-parametric tests

10

Data Transformations

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

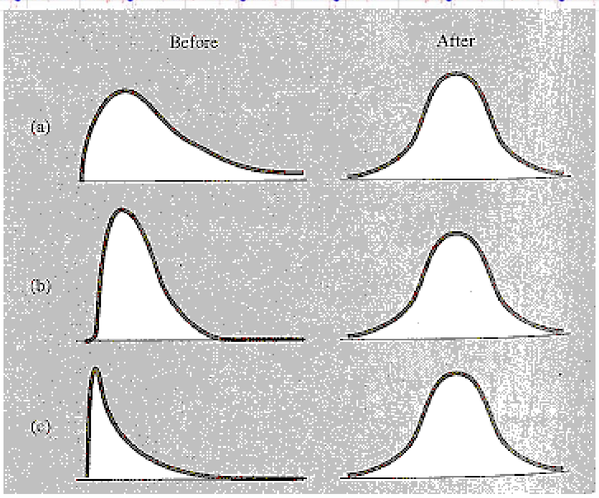
11

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

- Right Skew
  - Moderate – square root
  - Strong – log
  - Extreme – inverse, then reflect result
- Left Skew
  - Moderate – reflect then square root
  - Strong – reflect then log
  - Extreme - inverse

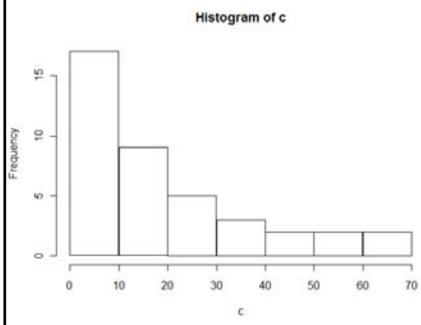
13

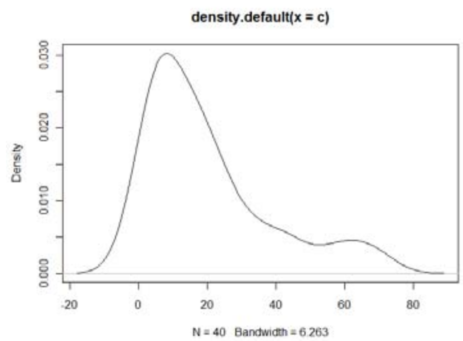
Data Transformations When Justified?

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

14

Example

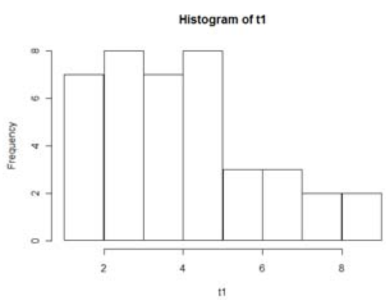


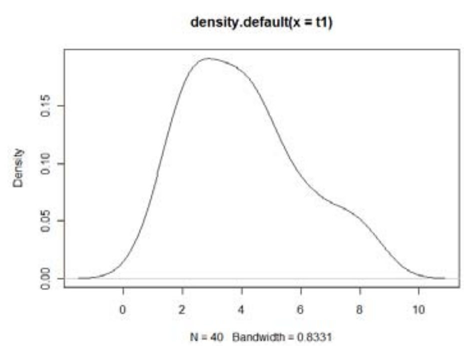


Shapiro-Wilk normality test
W = 0.8417, p-value = 5.693e-05

15

> sqrt(mydata)

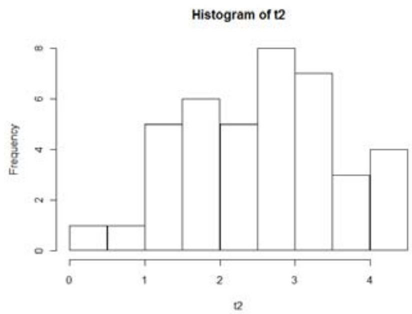


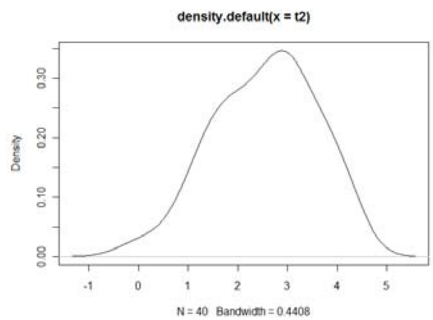


Shapiro-Wilk normality test
W = 0.948, p-value = 0.06461

16

> log(mydata)





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

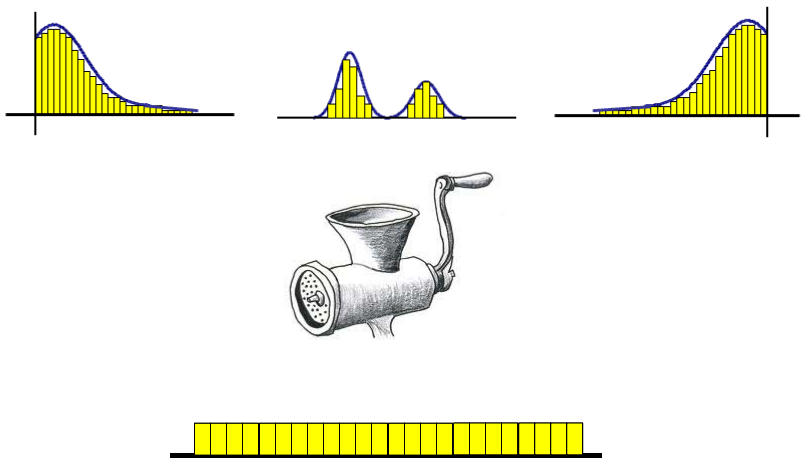
17

Non-parametric statistics

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

18

Rank-Order Transform



19

Rank-Order Transform

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

20

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

21

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

22

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

Critical values for U test

Nondirectional α=.05

	n_2											
n_1	**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

Example, continued

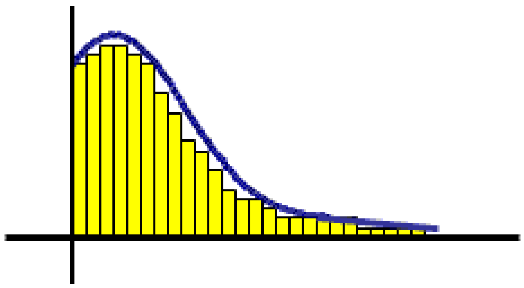
- Resulting data for comparison:
  - With sign in food court: 5, 8, 6, 7
  - Without sign: 2, 3, 1, 4
- U = 0 + 0 + 0 + 0 = 0
- Lookup n1=4, n2=4
- Or use R...

25

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
```

What do you do?
Review



- Data transformations
- Non-parametric tests

27

Data Analysis

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

28

Baseline Analysis
aka Check Potential Confounds

Primary purpose: show that randomization worked

Goal: no significant differences between randomized groups on non-outcome measures

29

Check Potential Confounds

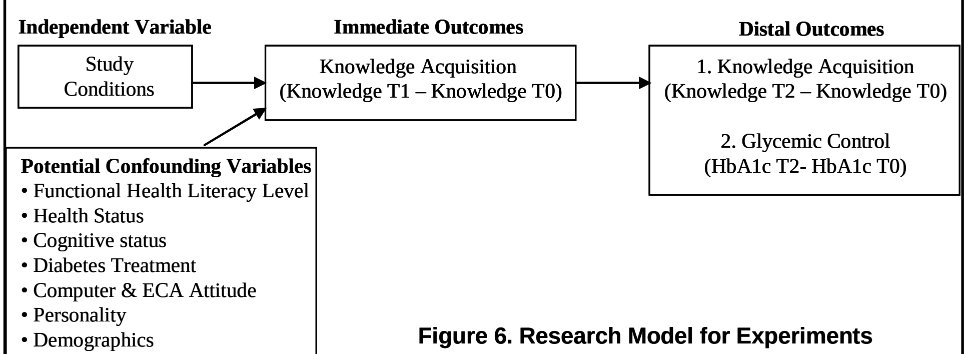


Figure 6. Research Model for Experiments

30

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

31

Sample Baseline Analysis Report

4 treatment between-subjects design

Demographics	All	CONTROL	ACT	DIET	ACT+DIET	P
Age, means (sd)	33.0(12.6)	32.0 (14.5)	33.5 (12.8)	32.9(11.1)	32.4 (12.3)	.813
Gender						.349
Male	48 (39.3%)	12 (38.7%)	15 (48.3%)	13(43.3%)	8 (26.6%)	
Female	74 (60.7%)	19 (61.3%)	16 (51.6%)	17(56.6%)	22 (73.3%)	
Smoking	9 (8.1%)	3 (9.6%)	2 (6.45%)	1 (3.33%)	3 (10.0%)	.801
Race						.835
Am. Indian	1 (0.8%)	1 (3.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Asian	40 (32.7%)	10 (32.2%)	7 (22.5%)	12 (40.0%)	11 (35.0%)	
Black	10 (8.0%)	1 (3.2%)	3 (9.6%)	3 (10.0%)	3 (10.0%)	
White	63 (51.6%)	17 (54.8%)	18 (58.0%)	14 (46.6%)	14 (46.6%)	
Hispanic	8 (6.5%)	2 (6.4%)	3 (9.6%)	1 (3.3%)	2 (6.45%0	
Education						.769
High Sc	2 (1.6%)	1 (3.2%)	1 (3.2%)	0 (0.0%)	0 (0.0%)	
Tech/Voc	3 (2.4%)	0(0.0%)	1 (3.2%)	1 (3.3%)	1 (3.3%)	
Some College	27 (22.1%)	7 (22.5%)	10 (32.2%)	4 (13.3%)	6 (20.0%)	
College Grad	66 (54.0%)	19 (61.2%)	13 (42.9%)	17 (56.6%)	17 (56.6%)	
Adv Degree	24 (19.6%)	4 (12.9%)	6 (19.3%)	8 (26.6%)	6 (20.0%)	

32

Data Analysis

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

33

Subgroup Analysis

34

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

35

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.
- More practically:
  - Plan your power analysis for your one primary hypothesis
  - Treat all other comparisons as tentative results

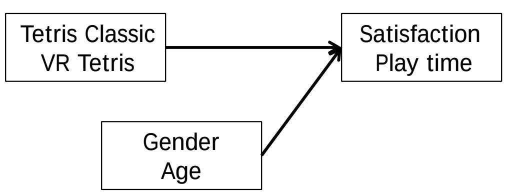
37

Data Analysis Summary

1. Check data for missing values & fix.
2. Check distributions of each measure for approximate 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.

38

Example
Game Evaluation

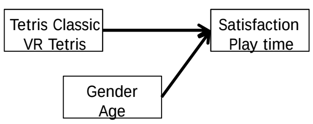


39

Example
Game Evaluation

Satisfaction

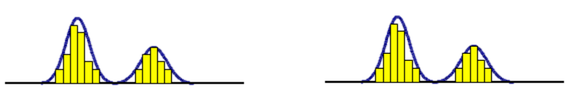
Tetris Classic VR Tetris



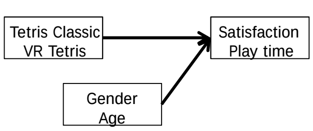
40

Example
Game Evaluation

Play Time



Tetris Classic VR Tetris



41

Adverse Events

43

Adverse Events

- If you are working under the oversight of an IRB – you must report any "adverse events" in a timely manner
  - "any untoward medical occurrence that may present itself during treatment or administration with a pharmaceutical product, and which may or may not have a causal relationship with the treatment."
- For large study, you may have a formal "Data & Safety Monitoring Board" (DSMB)

44

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

45

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

46

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.
- One of the subjects in your study gets a serious nosebleed and passes out during the test, but she indicates that this is common for her.
- 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.

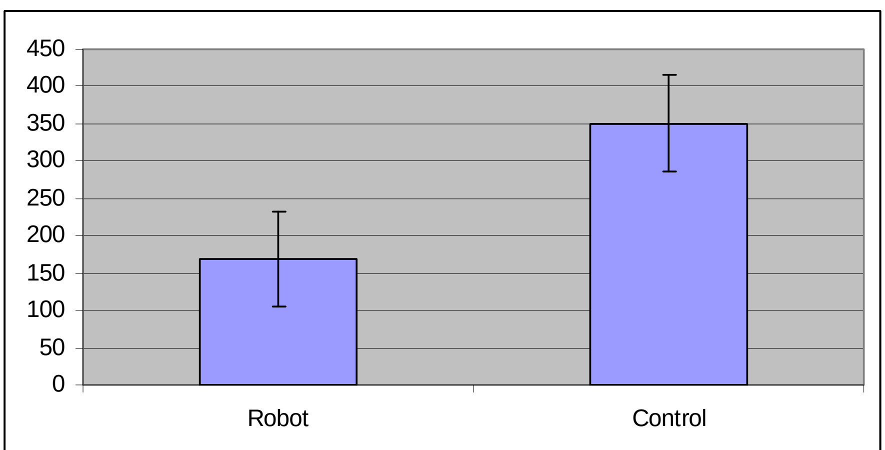
Example: RobotsRUs Robotic Vacuum Cleaner - RoboWipe

- 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 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.
- During the study three subjects in the intervention group report becoming nauseous, they think because of the cleaning fluid mist from the robot.
- 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.

Confidence Intervals

50

Example



51

Confidence Interval

- 95% CI = you are 95% sure that the true population mean falls within the interval.
- Based on a sample and the distribution of means.

52

Confidence Interval for Means of Samples from a Normal Distribution

```
> mean <- 5
> stdev <- 2
> samplesize <- 20
> error <- qnorm(0.975)*stdev/sqrt(samplesize)
> left <- mean-error
> right <- mean+error
> left
[1] 4.123477
> right
[1] 5.876523
>
```

53

Confidence Interval for t-test

```
t.test(sample1,sample2,var.equal=T)

Two Sample t-test

data:  c(1, 2, 3, 4, 5) and c(3, 4, 5, 6, 7)
t = -2, df = 8, p-value = 0.08052
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
 -4.3060041  0.3060041
sample estimates:
mean of x mean of y
        3         5
```

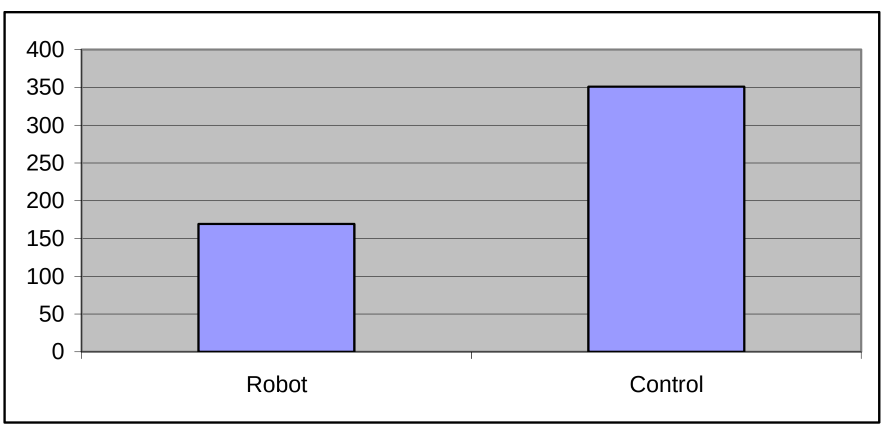
54

Step #2 – Excel Data & Graph

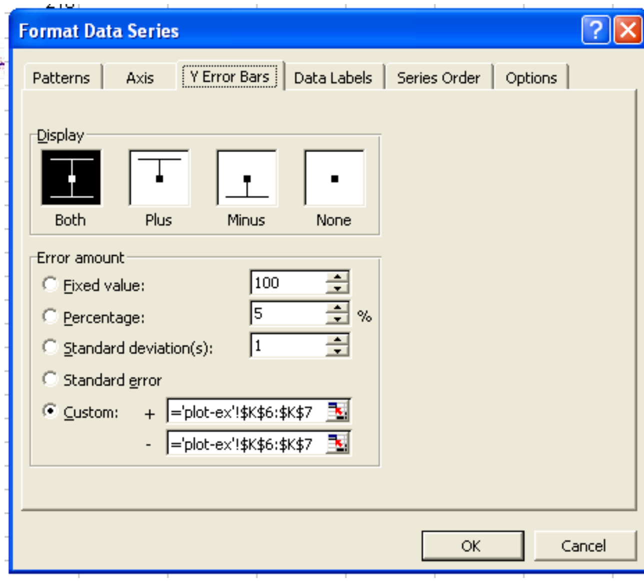
Means	Outcome
Robot	169
Control	351

95% CI

Low	High	Delta
105	233	64
232	414	63

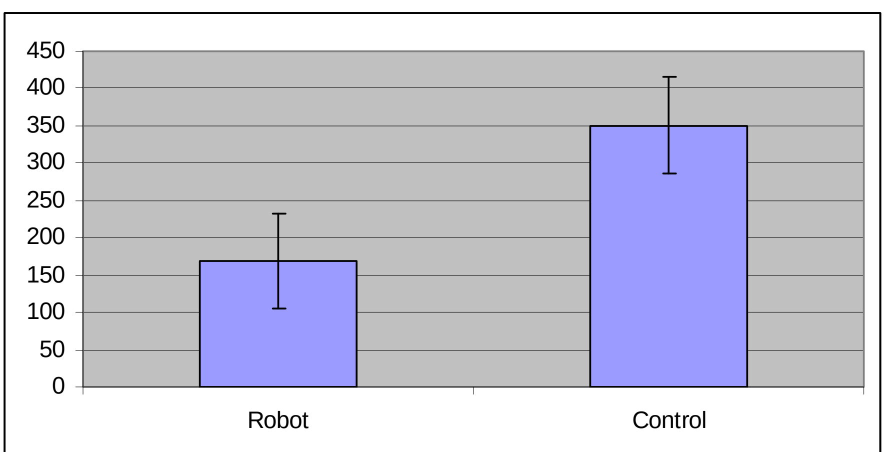


Step #3 – Confidence Interval



56

Result



57

Know from today

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

58

Sample Project Presentations

59

Homework

- Complete T1
  - Write T1 report
  - Prepare T1 presentation

60