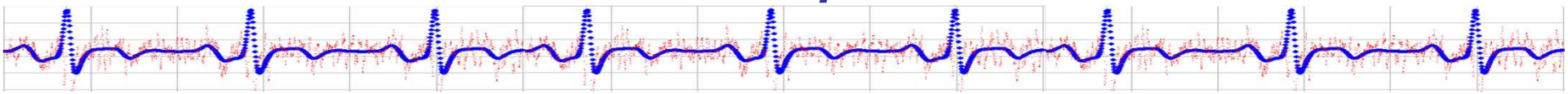
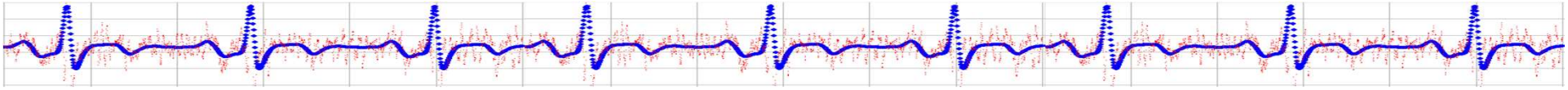
Empirical Research Methods in Information Science

IS 4800 / CS6350



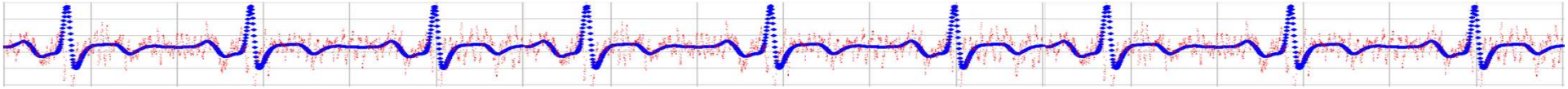
Lecture 12

Outline



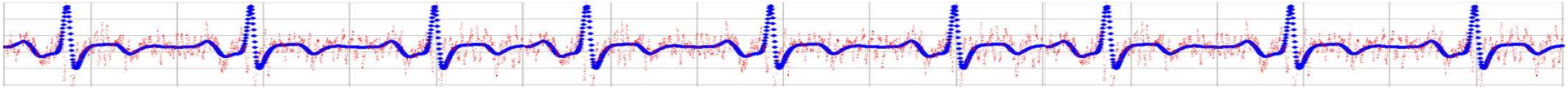
- Reading assessment
- Homework I5 (getting started)
- Chi-square
- Between subjects studies
- Start correlation…

I5 Part 1



- You are a manager at BigBucks, Inc., but you have noticed that your employees are becoming increasingly sedentary. You believe this is impacting productivity, and so you decide to conduct an experiment to test out two new fancy wearable fitness devices, the FlitBlit and the mappleBlotch, to see if they can motivate more physical activity after eight weeks of use with the staff.
- Be sure to include the following in your plan:
  - Hypotheses
  - Research model (the boxes and arrows diagram) and description of variables/measures
  - Human subjects issues, including eligibility criteria, recruitment procedures, and the number of potential subjects you need to reach with your recruitment
  - Detailed protocol , including recruitment, sampling and randomization methods
  - Analysis plan
- Your complete plan should be about 2-3 pages long, single spaced. Refer to the sample research plan for inspiration.

A common scenario



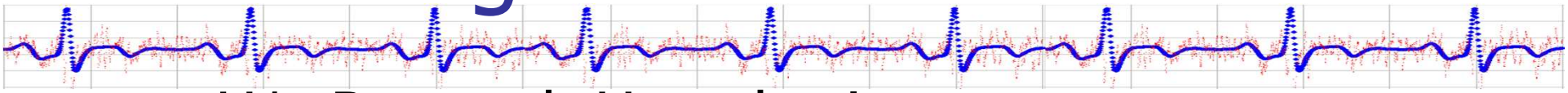
IV **DV**

Software A vs. Software B → **Efficiency (time to complete)**

Sampling Error
Measurement Error
Demand Effects
etc.

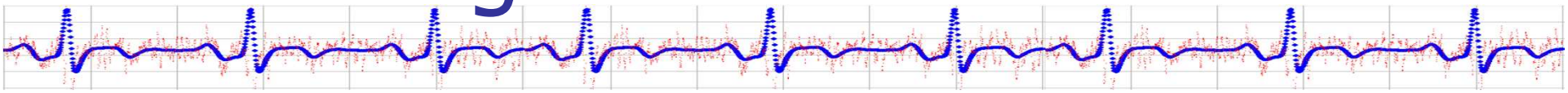
- **We measure difference A-B**
- **Is this real? Or just chance?**
- **We have very good models of randomness (noise).**
- **We can compute**
  p(observed difference A-B is due only to chance variation)
- **When should we conclude A-B is real (H1)?**
- **What can we conclude otherwise (H0)?**

Basic Process of Hypothesis Testing



- H1: Research Hypothesis:
  - Population 1 is different than Population 2
- H0: Null Hypothesis:
  - No difference between Pop 1 and Pop 2
  - *The difference is "null"*
- Compute *p(observed difference|H0)*
  - 'p' = probability observed difference is due to random variation
- If p<*threshold* then reject H0 => accept H1
  - p typically set to 0.05 for most work
  - p is called the "level of significance"

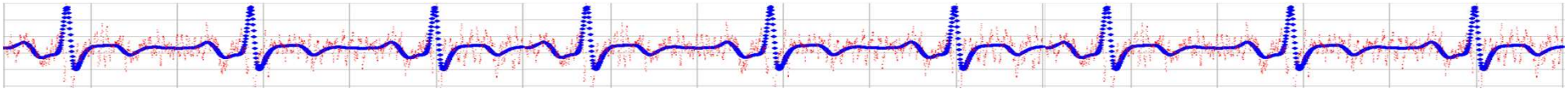
Type of Errors in Hypothesis Testing



	"The Truth"	
Your conclusion	H1 False	H1 True
Accept H1	Type I Error	Correct Decision
Reject H1	Correct Decision	Type II Error

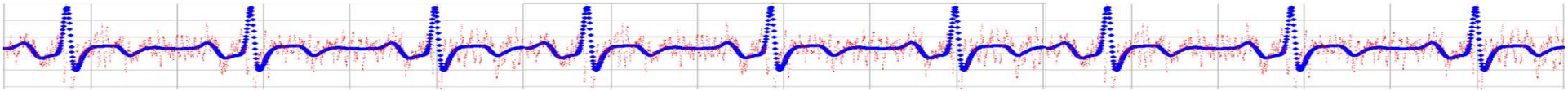
'p' = Probability of Type I Error

Chi-Square for Goodness of Fit



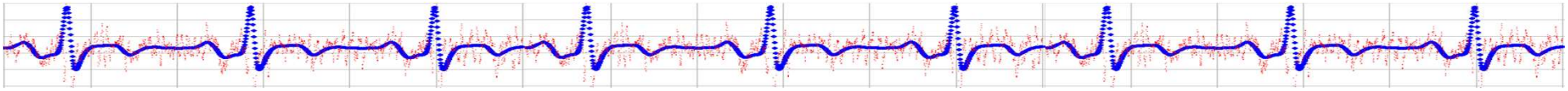
- Form of null hypothesis H0?
  - Observed frequency = Expected frequency
  - Populations (expected, observed) are actually the same on the nominal measure of interest
- Form of hypothesis H1?
  - Observed frequency ≠ Expected frequency
  - Populations (expected, observed) are different

Chi-Square for Goodness of Fit



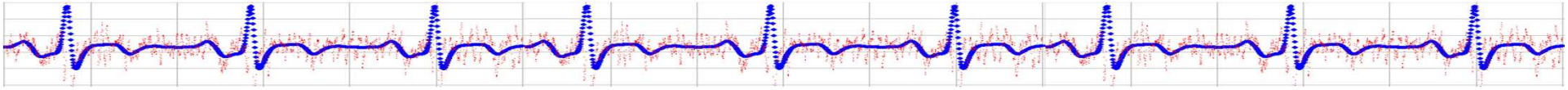
Is an observed frequency distribution significantly different from an expected distribution?

Chi-Square for Goodness of Fit



- Assumes
  1. You have a nominal variable
     - Values are exhaustive & mutually-exclusive
  2. You have an *Expected Frequency* table for the nominal variable
  3. None of the expected frequencies are "too small" (≥5)
  4. Random sampling

Formula for Chi-square statistic

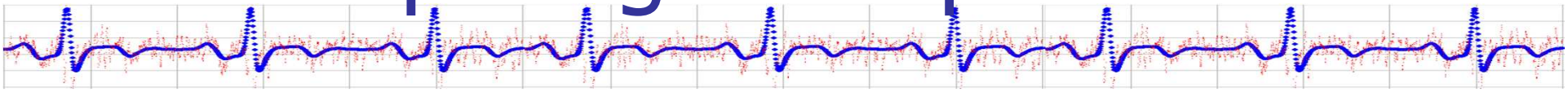


$$X^2 = \sum \frac{(O-E)^2}{E}$$

- O = Observed frequency for a given category
- E = Expected frequency for a given category

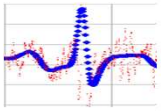
- Note: "statistic" is a function you apply to a set of data (in a statistical analysis)

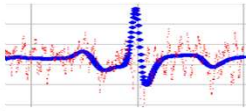
Computing Chi-square

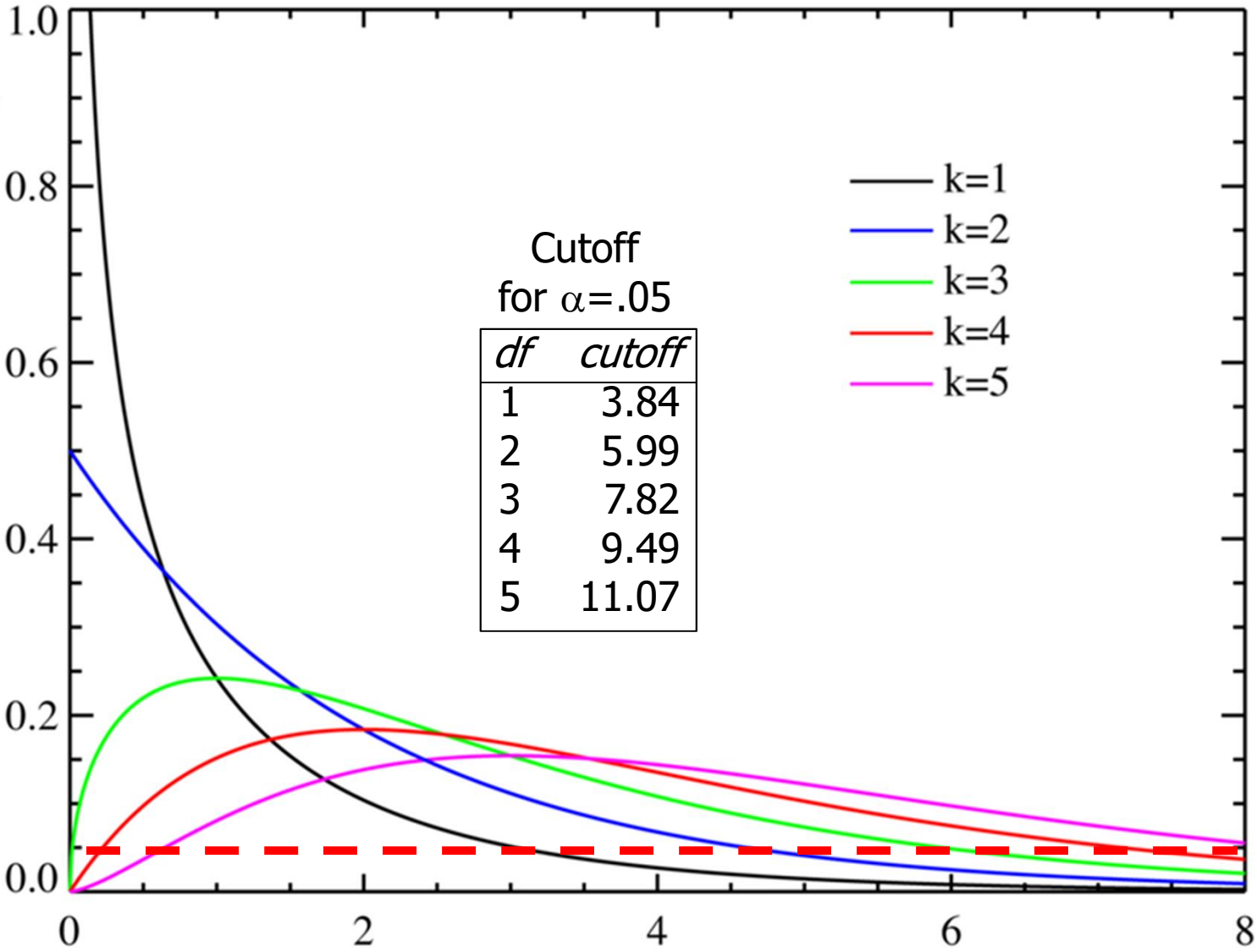


- Manually:
  - Determine df (= num categories – 1)
  - Compute Chi-square using formula
  - Lookup to see if statistic>table entry for significance,df
    - If yes => reject H0
    - If no => inconclusive

Chi-square probability distribution



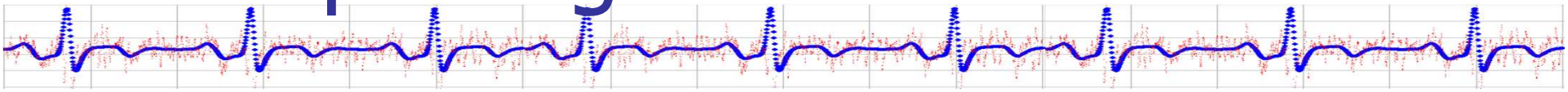




Cutoff for α=.05

df	*cutoff*
1	3.84
2	5.99
3	7.82
4	9.49
5	11.07

Reporting result



$$X^2(df) = chisq, p < sigthresh$$

Where,

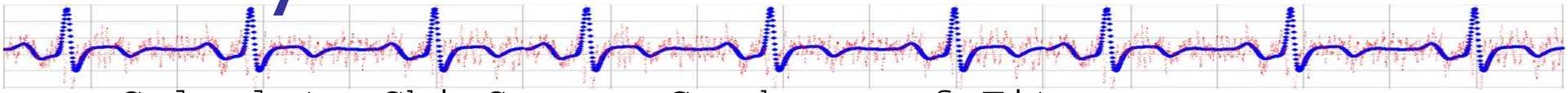
- df = degrees of freedom
- sigthresh = pre-defined significance threshold
  - Note: if p<<sigthresh, can report that as well, e.g., "p<.01", "p=.001"

For example: $X^2(2) = 11.89, p < 0.05$

If not significant, than use "n.s." instead of "p<…".

Usually also report expected and actual frequencies, or at a minimum, the total number of cases considered (aka "n").

Computing chi-square in Python



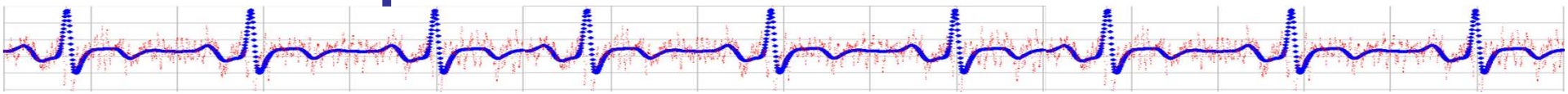
Calculate Chi Square Goodness of Fit

- The results returned are the Chi Square statistic, and the p value. Optionally enter the degrees of freedom, ddof. from scipy.stats
- import chisquare
- observed= [16, 18, 16, 14, 12, 12]
- expected= [16, 16, 14, 15, 13, 12]
- ddof = 2
- chisquare(observed, expected, ddof)

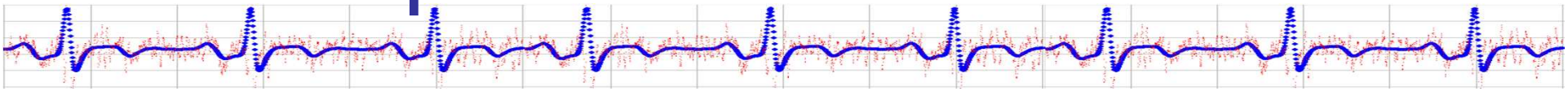
```
>>> (2.0, 0.8491450360846095 6)
```

Note: If you don't supply an expected distribution, it will use a default equal distribution…

Chi-Square Test for Independence



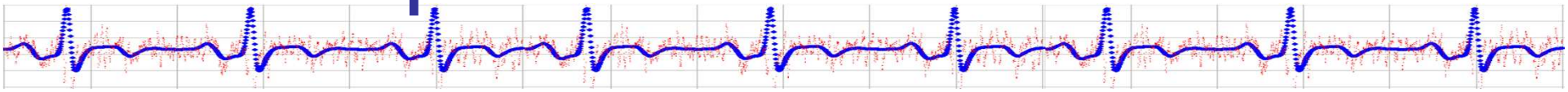
Chi-Square Test for Independence



- Are two nominal variables related (H1), or are they independent (H0)?

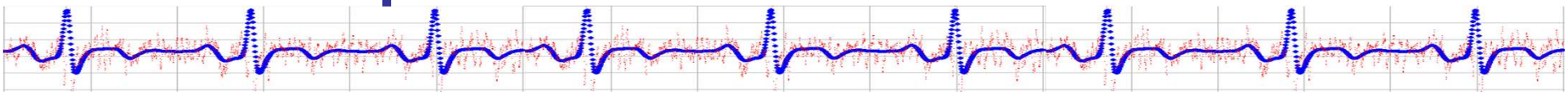
- Assumptions
  - Both variables must be nominal.
  - Cannot be related in a 'special' way (i.e., repeated measures)
  - Random sampling assumed

Chi-square Test for Independence

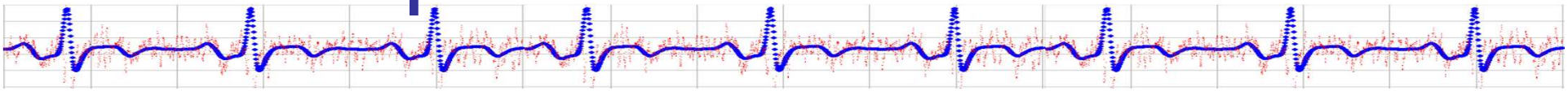


- Which of the following is it appropriate for?
  - Descriptive study designs
  - Demonstration study designs
  - Correlational study designs
  - Experimental study designs

Chi-Square Test for Independence



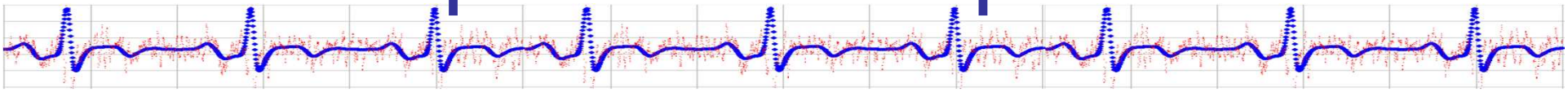
Chi-Square Test for Independence



- Are two variables related (H1), or are they independent (H0)?

- Assumptons
  - Both variables must be nominal.
  - Cannot be related in a 'special' way (i.e., repeated measures)
  - Random sampling assumed

Example from chapter



- Morning & night people using different modes of transportation.
- What kind of table is this?

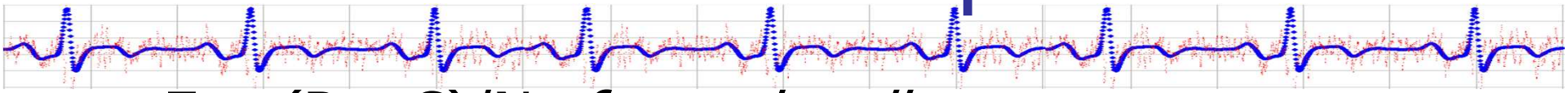
Contingency table

	Bus	Carpool	Own Car
Morning	60	30	30
Night	20	20	40

- What kind of study is this?

Correlational

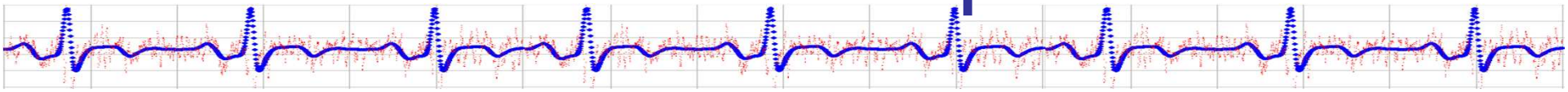
Expected frequencies if variables are independent



- E = (R x C)/N *for each cell*
  - R = row count
  - C = column count
  - N = total number in all cells

	Bus	Carpool	Own Car
Morning	60	30	30
Night	20	20	40

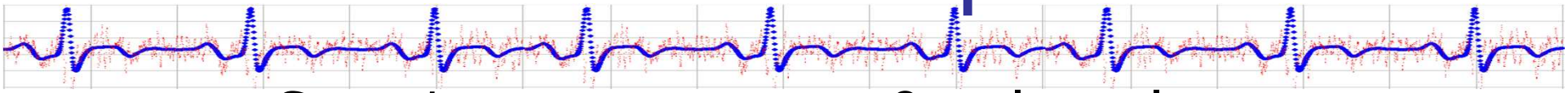
Expected frequencies if variables are independent



- Step 1 – compute row & col totals

	Bus	Carpool	Own Car	
Morning	60	30	30	**120**
Night	20	20	40	**80**
	80	**50**	**70**	

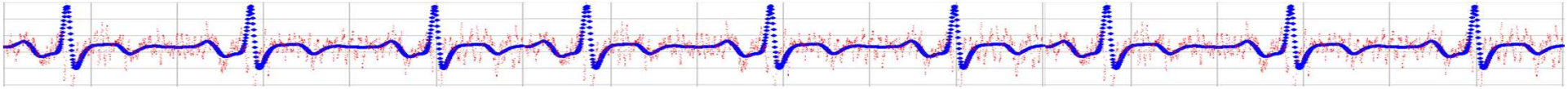
Expected frequencies if variables are independent



- Step 1 – compute row & col totals
- Step 2 – ea cell = (R x C)/N

	Bus	Carpool	Own Car	
Morning	**(48)** 60	**(30)** 30	**(42)** 30	**120**
Night	**(32)** 20	**(20)** 20	**(28)** 40	**80**
	80	**50**	**70**	

Formula

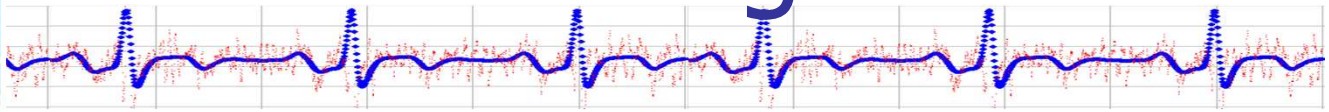


- Same as goodness-of-fit test.

$$X^2 = \sum \frac{(O-E)^2}{E}$$

- df = (NumRows-1)x(NumColumns-1)

Survey Feb 5, 2013 Guns in Congress



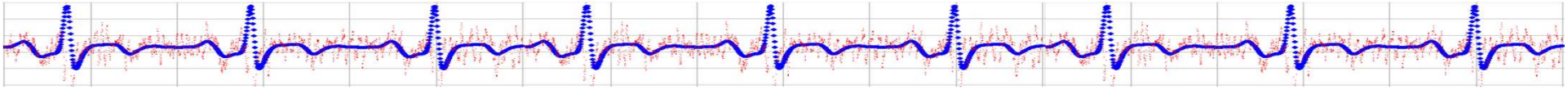
- Q: Is Gun Ownership related to NRAGrade?

	No	Yes
A	14	131
B	5	5
C	7	8
D	4	2
F	103	17

df=?

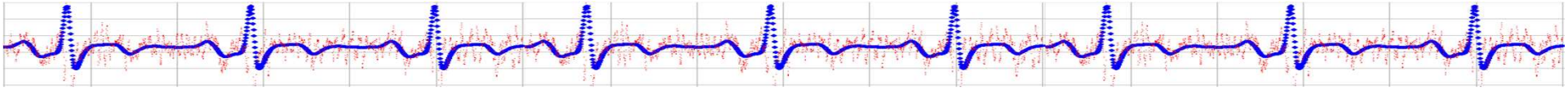
X^2 = ?

Fill in this table…



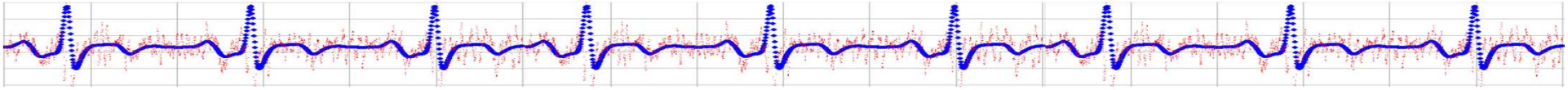
	No	Yes	
A	14 ()	131 ()	? (?)
B	5 ()	5 ()	? (?)
C	7 ()	8 ()	? (?)
D	4 ()	2 ()	? (?)
E	103 ()	17 ()	? (?)
	?	?	? (?)

Fill in this table…



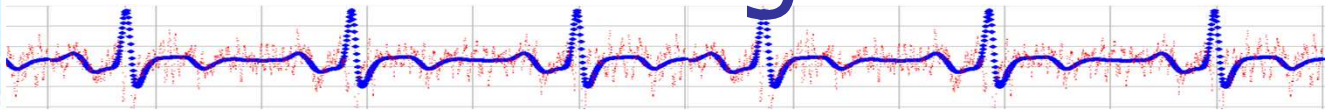
	No	Yes	
A	14 ()	131 ()	145 (49%)
B	5 ()	5 ()	10 (3%)
C	7 ()	8 ()	15 (5%)
D	4 ()	2 ()	6 (2%)
E	103 ()	17 ()	120 (41%)
	133	163	296 (100%)

Fill in this table…



	No	Yes	
A	14 (65)	131 (80)	145 (49%)
B	5 (4)	5 (5)	10 (3%)
C	7 (7)	8 (8)	15 (5%)
D	4 (3)	2 (3)	6 (2%)
E	103 (54)	17 (67)	120 (41%)
	133	163	296 (100%)

Survey Feb 5, 2013 Guns in Congress



- Q: Is Gun Ownership related to NRAGrade?

	No	Yes
A	14	131
B	5	5
C	7	8
D	4	2
F	103	17

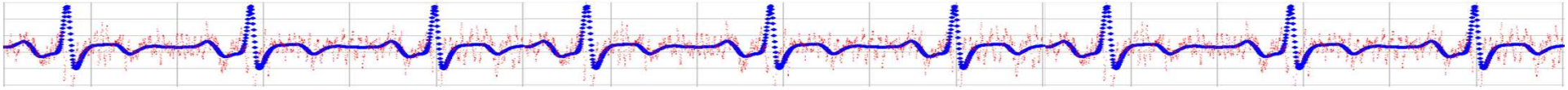
Cutoff for α=.05

df	*cutoff*
1	3.84
2	5.99
3	7.82
4	9.49
5	11.07

df=(2-1) x (5–1)

X^2 = 155

Exercise

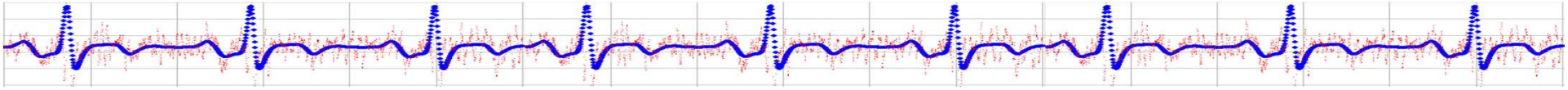


- For each problem, write
  1. What kind of study design is it?
  2. Two populations being compared
  3. Research hypothesis
  4. Null hypothesis
  5. Test criteria
  6. Expected frequencies
  7. Observed frequencies
  8. Test results
     - publication format and
     - English

Cutoff for α=.05

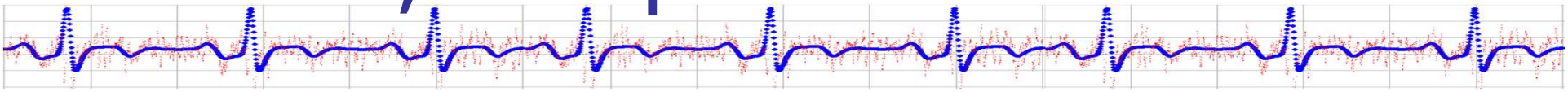
df	*cutoff*
1	3.84
2	5.99
3	7.82
4	9.49
5	11.07

A Brief Note About Power



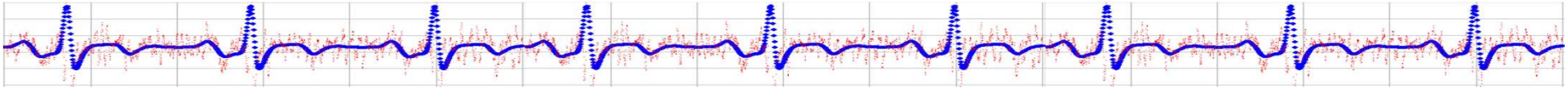
- The "power" of a statistical test is its ability to detect differences in data that are inconsistent with the null hypothesis.
  - p(rejecting H0|H1)
  - aka Concluding H1, given that H1 is actually true.

Relationship between alpha, beta, and power.



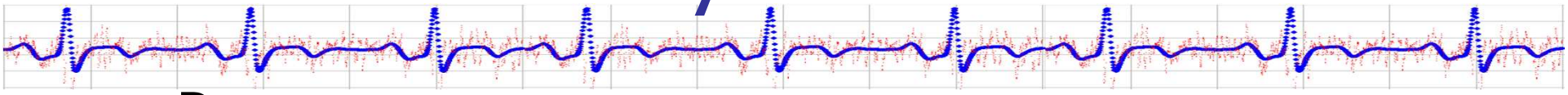
	"The Truth"	
	H1 True	**H1 False**
Decide to Reject H0 & accept H1	Correct p = power	Type I err p = alpha
Do not Reject H0 & do not accept H1	Type II err p = beta	Correct p = 1-alpha

Effect size



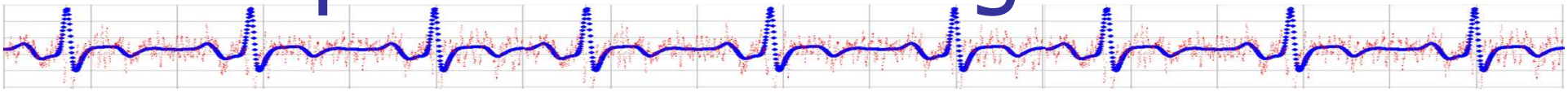
- The *amount* of change in the DVs seen.
- Can have statistically significant test but small effect size.

Power Analysis



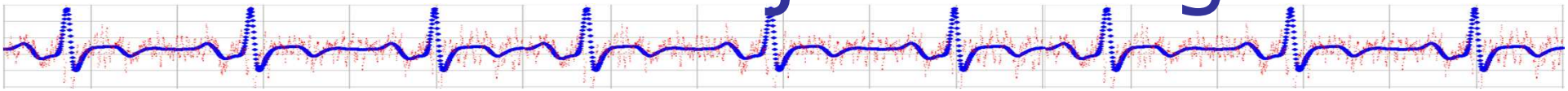
- Power
  - Increases with effect size
  - Increases with sample size
  - Decreases with decreasing (more stringent) alpha
- Should determine number of subjects you need ahead of time by doing a 'power analysis'
- Standard procedure:
  - Fix alpha and beta (power)
  - Estimate effect size from prior studies
    - Categorize based on Table 13-8 in Aron (sm/med/lg)
  - Determine number of subjects you need
  - For Chi-square, see Table 13-10 in Aron Ch 13, or pwr.chisq.test in "pwr" package.

Two Group Between-Subjects Experimental Design



- B&A: "Randomized Two Group Design"
- Have two experimental conditions (treatments, levels, groups)
- Randomly assign subjects to conditions
  - Each subjects sees one condition
- Measure (numeric) outcome in each group

Between-Subjects Design

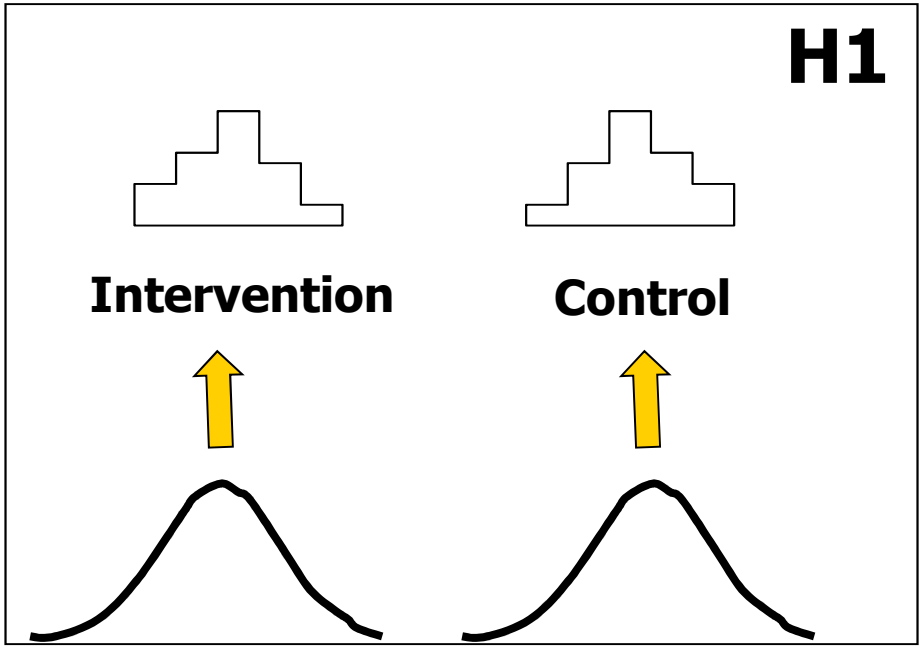


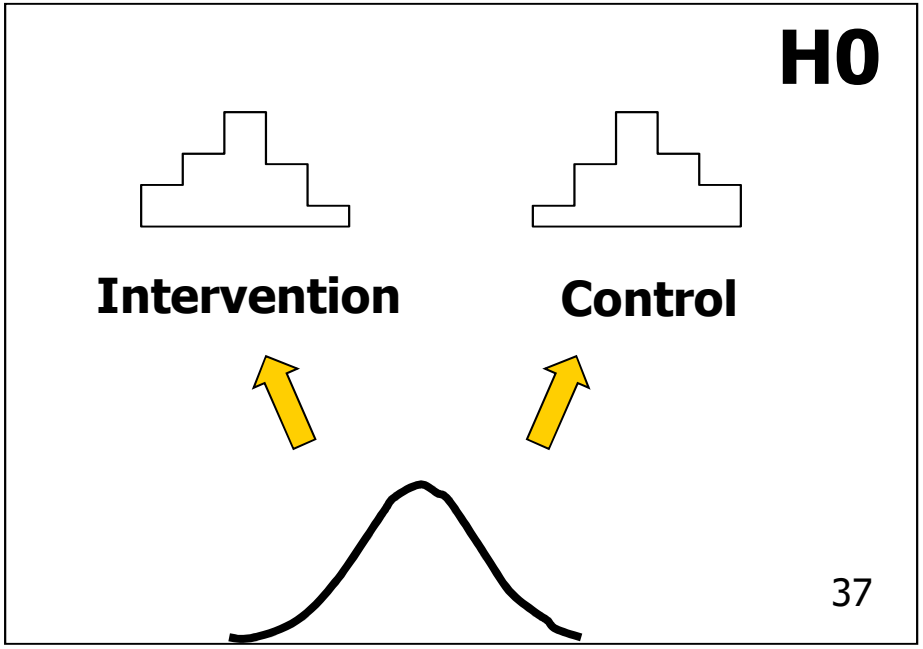
- Each group is a **sample** from a population
- Big question: are the populations the same (null hypothesis) or are they significantly different?

Intervention

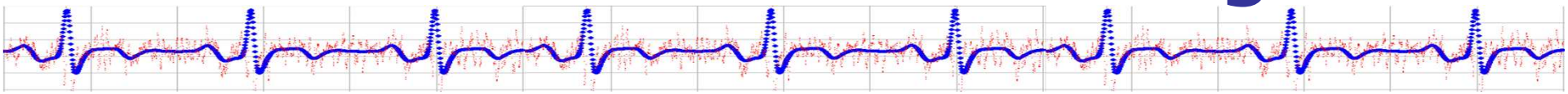
Control

The big question: which is correct?

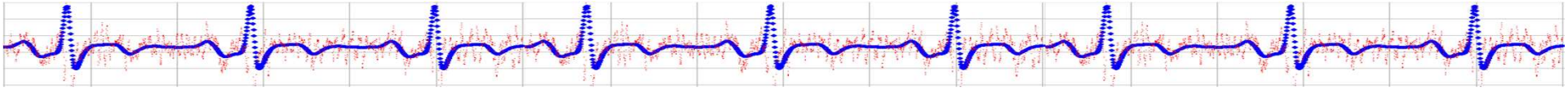




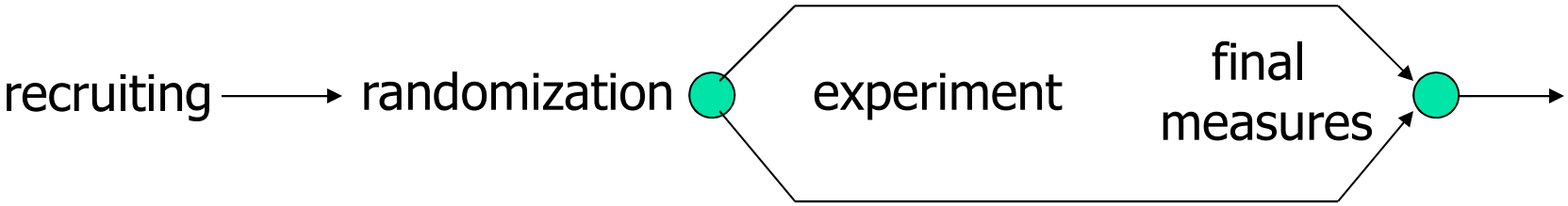
Hold that thought…
More next time on testing this!

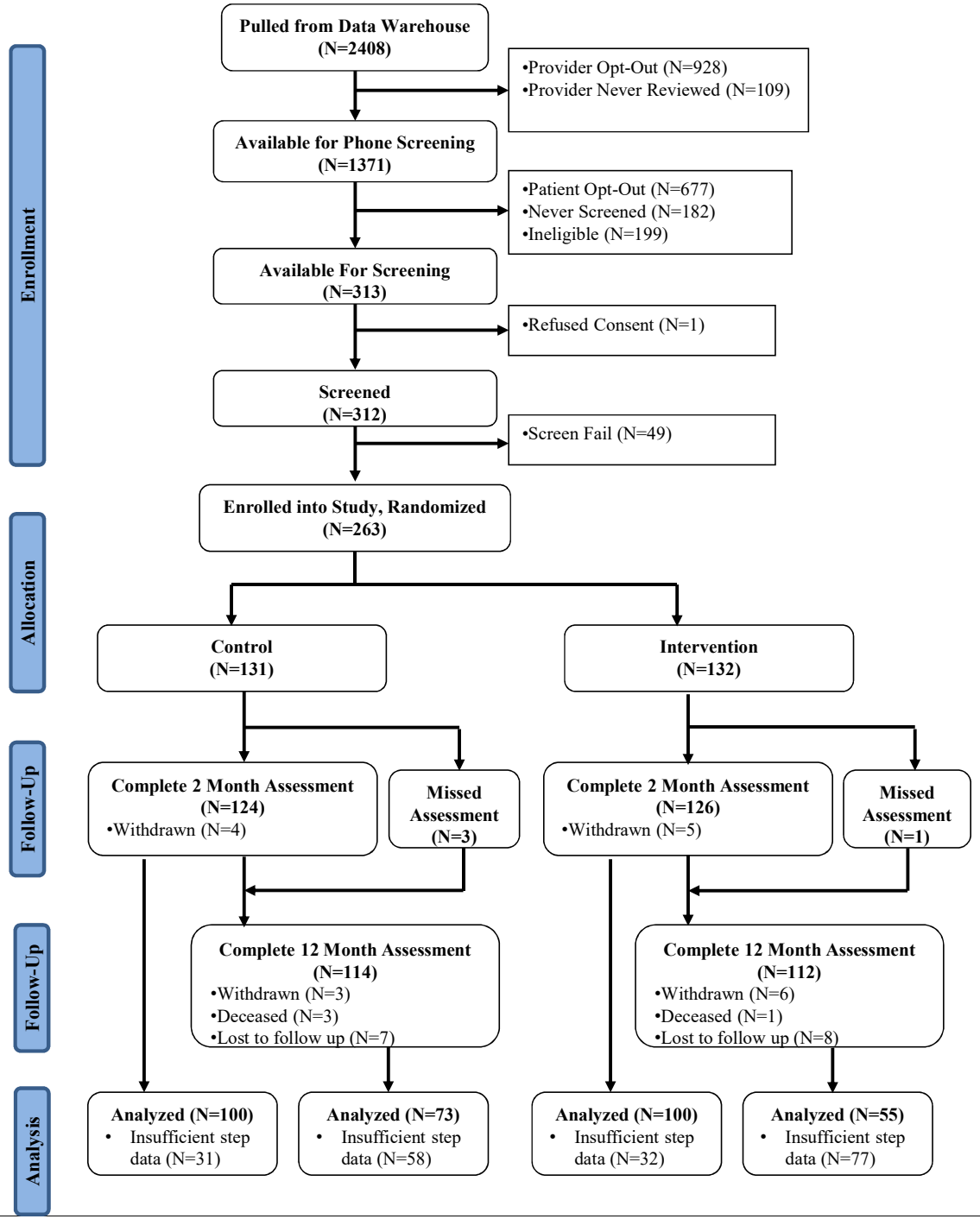


Sidebar: Randomization

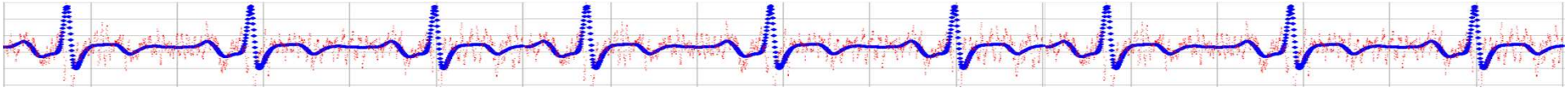


- Crucial: method must not be applied subjectively
- Point in time at which randomization occurs is important



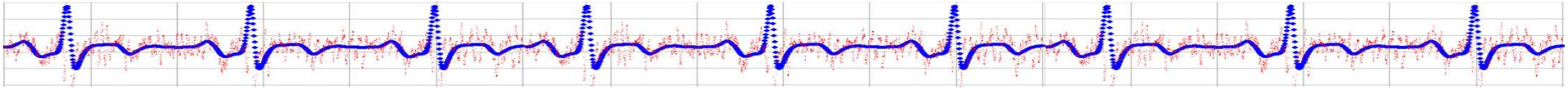

Enrollment
Allocation
Follow-Up
Follow-Up
Analysis
Pulled from Data Warehouse (N=2408)
•Provider Opt-Out (N=928)
•Provider Never Reviewed (N=109)
Available for Phone Screening (N=1371)
•Patient Opt-Out (N=677)
•Never Screened (N=182)
•Ineligible (N=199)
Available For Screening (N=313)
•Refused Consent (N=1)
Screened (N=312)
•Screen Fail (N=49)
Enrolled into Study, Randomized (N=263)
Control (N=131)
Intervention (N=132)
Complete 2 Month Assessment (N=124)
•Withdrawn (N=4)
Missed Assessment (N=3)
Complete 2 Month Assessment (N=126)
•Withdrawn (N=5)
Missed Assessment (N=1)
Complete 12 Month Assessment (N=114)
•Withdrawn (N=3)
•Deceased (N=3)
•Lost to follow up (N=7)
Complete 12 Month Assessment (N=112)
•Withdrawn (N=6)
•Deceased (N=1)
•Lost to follow up (N=8)
Analyzed (N=100)
• Insufficient step data (N=31)
Analyzed (N=73)
• Insufficient step data (N=58)
Analyzed (N=100)
• Insufficient step data (N=32)
Analyzed (N=55)
• Insufficient step data (N=77)

Intent-to-Treat



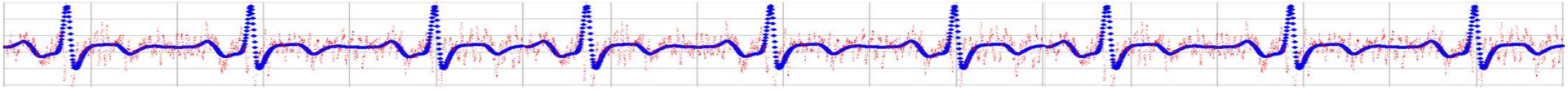
- You want to test a new support line ticket system.
- You randomize 20 support employees to use the new system, 20 to use the old one, then collect satisfaction and performance measures after one month.
- You discover that 6 of the employees using the new system stopped using it after a week.
- What do you do?

Intent-to-Treat



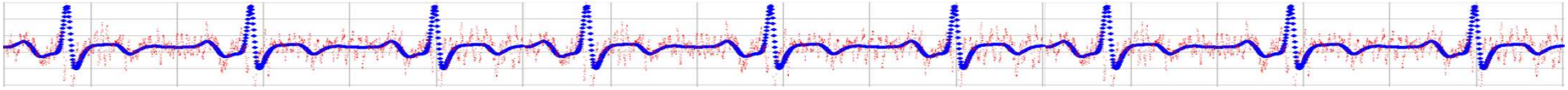
- Once a subject is randomized, every effort is made to include their outcome measures (DV) in the analysis
  - Even if they did not use the Intervention
  - Even if they went on vacation
  - Even if they died …
  - Assume worst case for lost data (e.g., intervention did not work)
- Efficacy = IV/DV effect under ideal conditions (e.g., lab study) = "method effectiveness"
- Effectiveness (aka "use effectiveness") = IV/DV effect under real world conditions
- Intent-to-treat assesses "effectiveness"

Sidebar: Randomization



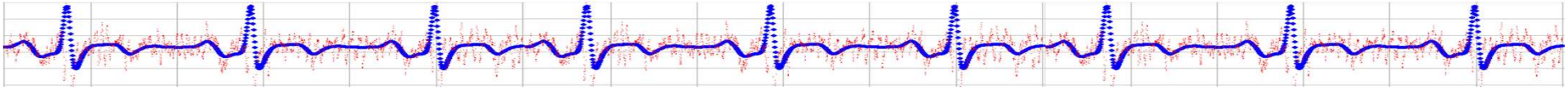
- Simple randomization
  - Flip a coin
  - Random number generator
  - Table of random numbers
  - Partition numeric range into number of conditions
- Problems?

Sidebar: Randomization



- Blocked randomization
  - Avoids serious imbalances in assignments of subjects to conditions
  - Guarantees that imbalance will never be larger than a specified amount
  - Example: want to ensure that every 4 subjects we have an equal number assigned to each of 2 conditions => "block size of 4"
  - Method: write all permutations of N conditions taken B at a time (for B = block size)
    - Example: 1122, 1212, 2112, 2121, 2211, 1221
  - At the start of each block, select one of the orderings at random
  - Should use block size > 2, block size = multiple of # arms

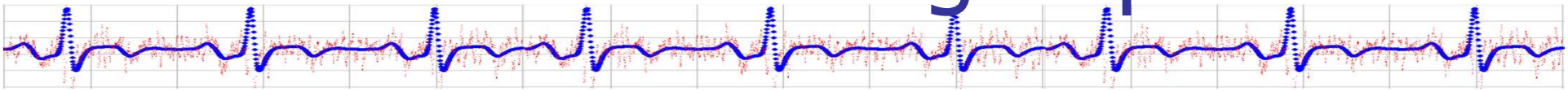
Sidebar: Randomization



- Stratified randomization
  - First stratify Ss based on measured factors (prior to randomization) (e.g., gender)
  - Within each strata, randomize
    - Either simple or blocked

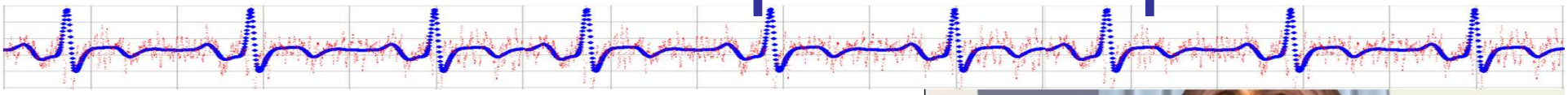
Strata	Sex	Condition assignment
1	M	ABBA BABA…
2	F	BABA BBAA…

Sidebar: Control groups

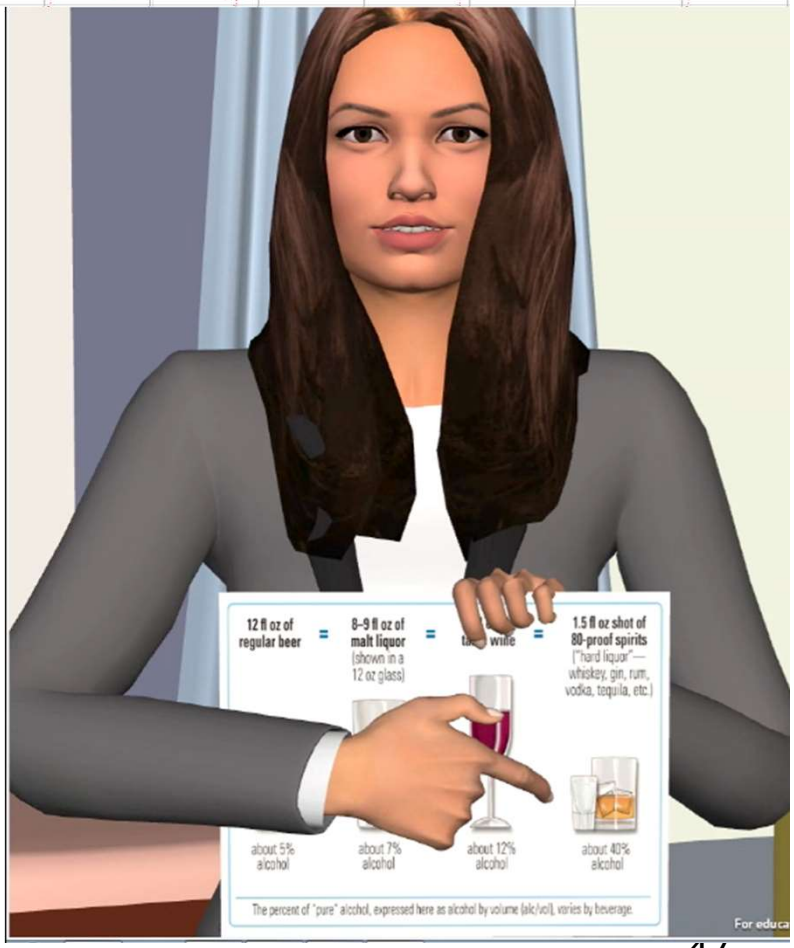


- A controlled experiment ("experimental design") generally compares the results obtained from an experimental sample against a control sample, which is identical to the experimental sample except for the one aspect whose effect is being tested.
- You must carefully select your control group in order to demonstrate that only the IV of interest is changing between groups.
- The control group must also comprise a reasonable comparison.

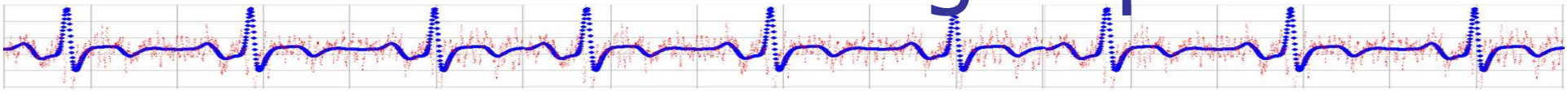
Control Groups: Example



- Say you are developing a conversational agent that counsels college students with depression (using CBT) and co-morbid binge drinking (using BMI).
- What is a good control group?



Sidebar: Control groups



- Standard-of-care control (new vs. old)
- Non-intervention control
- "A vs. B" design (shootout)
- "A vs. A+B" design

- Problem: the "intervention" may cause more than just the desired effect
  - Example: giving more attention to intervention Ss in educational intervention
- Some solutions:
  - Attention control
  - Placebo control
  - Wait list control (also addresses ethics)