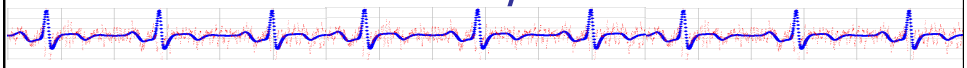
Empirical Research Methods in Information Science

IS 4800 / CS6350

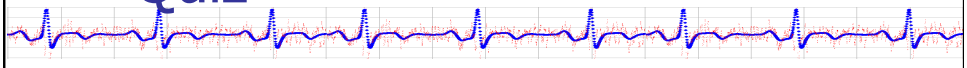


Lecture 10

Hypothesis Testing Using Chi^2

1

Quiz



https://tinyurl.com/IS4800NUM5

2

Homework Status?

- Surveys?
- Analysis Plan?

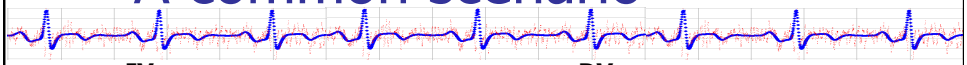
3

Hypothesis Testing

- How can we tell if our hunches about the world are right?
  - *Step 3 of the scientific method*
- How can we do this in the face of sampling error, measurement error, and extraneous variables?

4

A common scenario



IV **DV**

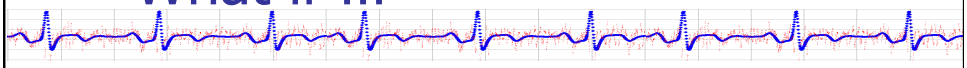
Software A vs. Software B → **Efficiency (time to complete)**

Sampling Error
Measurement Error
Covariates
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)?**

5

What if …

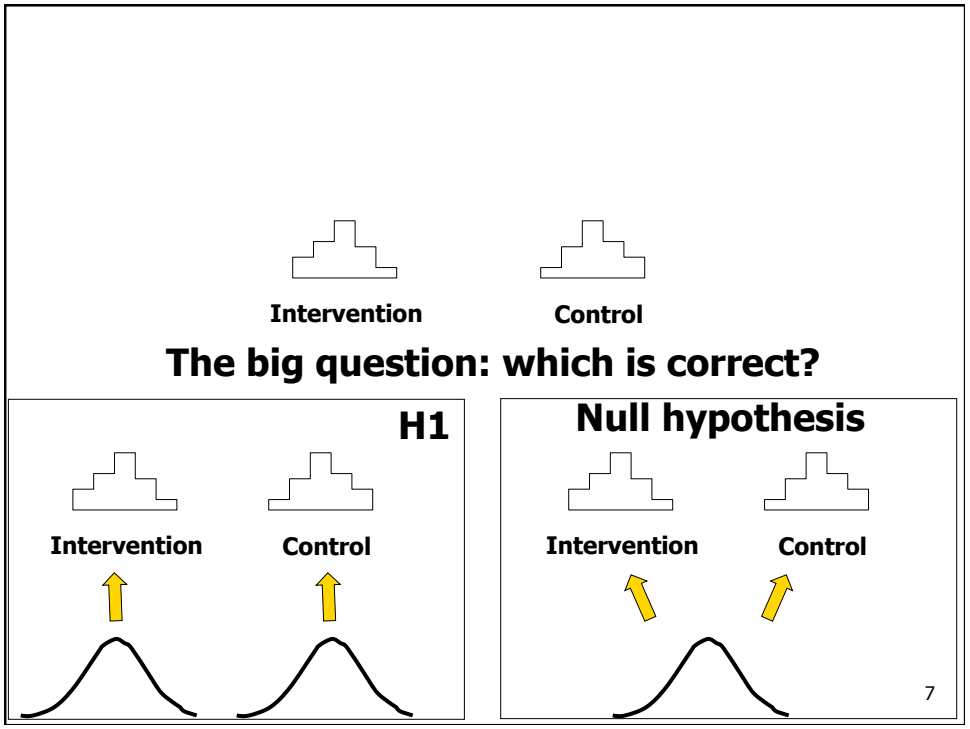


- You can test every subject in your population, AND
- There is no measurement error? AND
- There are no covariates or other noise?

→

- Compute descriptives for two treatments
- If Software A Efficiency > Software B Efficiency conclude H1 is supported
- No uncertainty, p=0.0

6



Basic Process of Hypothesis Testing

- H1: Research Hypothesis:
  - Population 1 is different than Population 2 on some measure
- 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"

8

The plan for hypothesis testing

- X^2 tests
  - For 1 or 2 nominal measures
- Correlation tests
  - For two numeric measures
- t-test for independent means
  - For categorical IV, numeric DV

9

Survey Feb 5, 2013 Guns in Congress

- How many members of congress own guns?
- Dataset:

Title Member Party State GunOwner NRAGrade

- 531 seated members
- 165 Yes, 147 No, 219 “no comment”

10

Survey Feb 5, 2013
Guns in Congress

- Question: Is congressional gun ownership greater than US national gun ownership?
  - 32% of households have guns
    - 2010 data, 2011 study University of Chicago's National Opinion Research Center
    - Note: rate of individual ownership is less
  - Congress: 165 Yes, 147 No

11

Survey Feb 5, 2013
Guns in Congress

- Question: Is the rate of congressional gun ownership greater than the US national gun ownership rate?
- US: 32% Congress: 55%
  - Based on percentages, yes.
- Are congressional members representative of US households wrt gun ownership?
- If we were to take a random sample of 312 US households, how likely would we be to see this outcome?

12

Survey Feb 5, 2013 Guns in Congress

- Null Hypothesis (H_0): distribution of gun ownership in congress = distribution of gun ownership in US households.
- Research Hypothesis (H_1): distribution of gun ownership in congress ≠ distribution of gun ownership in US households.
- Inferential statistics: What is the probability of drawing a sample from US households that has the rate of gun ownership observed in congress?
  - What is $p(H_1|H_0)$?

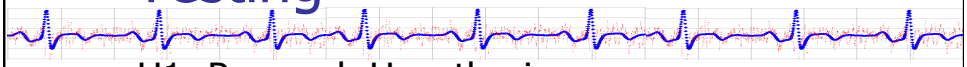
13

Survey Feb 5, 2013 Guns in Congress

- Significance:
  - How small does $p(H_1|H_0)$ need to be before we say that congress is really different?
  - Convention: $p(H_1|H_0) <= 0.05$
  - "Level of significance" $\alpha = 0.05$
- How to compute $p(H_1|H_0)$?
  - Chi square distribution….

14

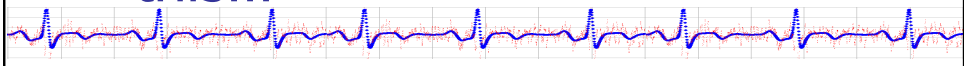
Basic Process of Hypothesis Testing



- 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"

15

Other ways of thinking about this…



- "Innocent until proven guilty."
- How surprising would this result be if there really were no difference?
- Why do things this way???

16

Terminology Soup

- Descriptive statistics = describing state of the world
- Inferential statistics = Hypothesis testing
  - When you only have access to a sample, or want to account for other kinds of noise
- Univariate analysis/statistics = statistics on only one variable (DV) at a time
- Multivariate analysis/statistics = statistics on more than one variable (DV) at a time

17

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"

18

Examples

- Research Question:
  - Which is better, Guitar Hero or Rock Band?
- Research Question:
  - Is the ownership of Wii vs. Xbox consoles significantly different among NU students compared to ownership in the general US population?
- Research Question:
  - Are Wii owners more likely to own Guitar Hero vs. Rock Band, compared to Xbox owners?

19

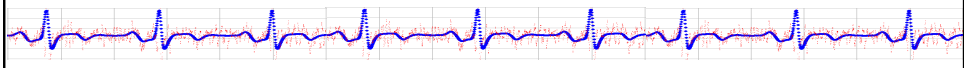
Type of Errors in Hypothesis Testing

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

'p' = Probability of Type I Error

20

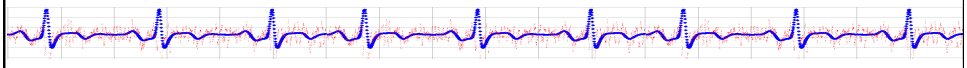
Chi-Square for Goodness of Fit



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

21

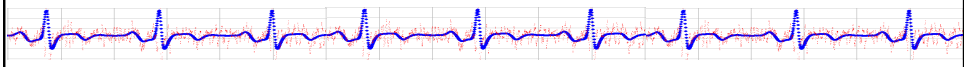
Chi-Square for Goodness of Fit



- 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

22

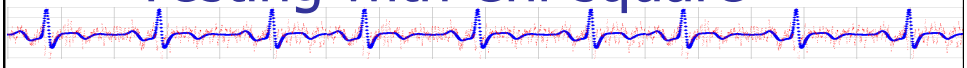
Chi-Square for Goodness of Fit



- 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

23

Procedure for Hypothesis Testing with Chi-square



1. Formulate your research hypothesis (including statement of expected frequencies)
2. Determine hypothesis test parameters
   - significance threshold
3. Collect your data
4. Compute Chi-Square statistic and draw conclusion

24

Formula for Chi-square statistic

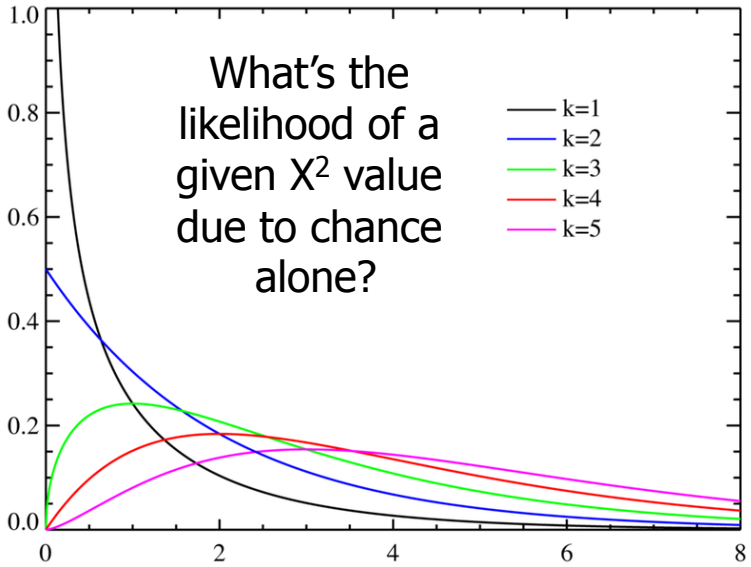
$$X^2 = \sum \frac{(O-E)^2}{E}$$

- 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)

25

Chi-square probability distribution



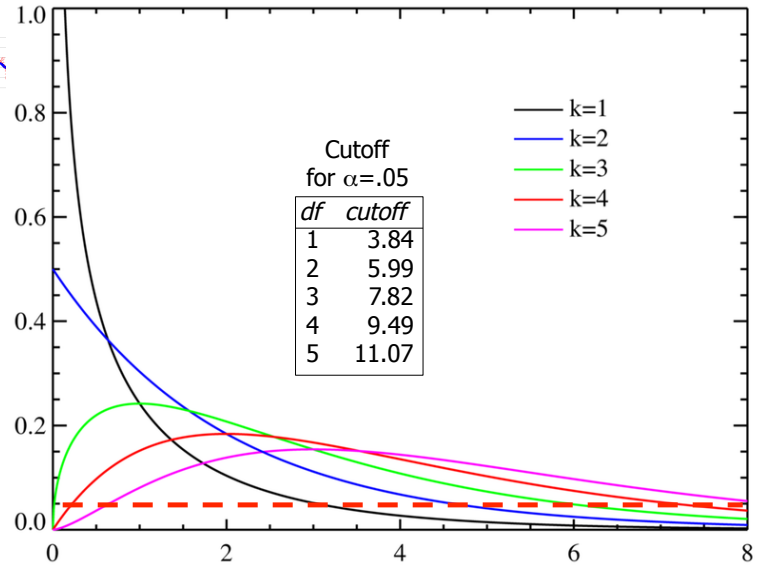
26

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

27

Chi-square probability distribution



28

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").

29

Survey Feb 5, 2013 Guns in Congress

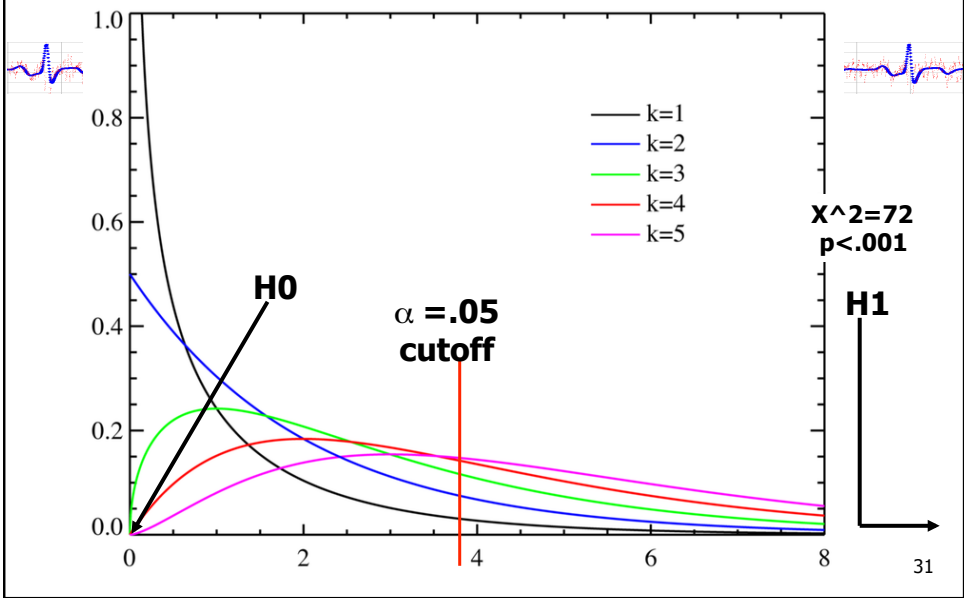
- US: 32% Congress: 55% N=296

Own Gun?	Expected	Observed
Yes	95	163
No	201	133

$$X^2 = \frac{(163-95)^2}{95} + \frac{(133-201)^2}{201}$$

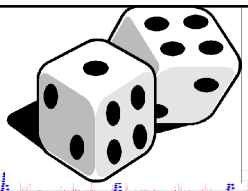
$$= 72$$

Chi-square probability distribution



31

Example



- You go gambling in a shady casino, and suspect that the games are rigged.
- You focus your attention on one 6-sided die being used in a game and keep track of 60 rolls:

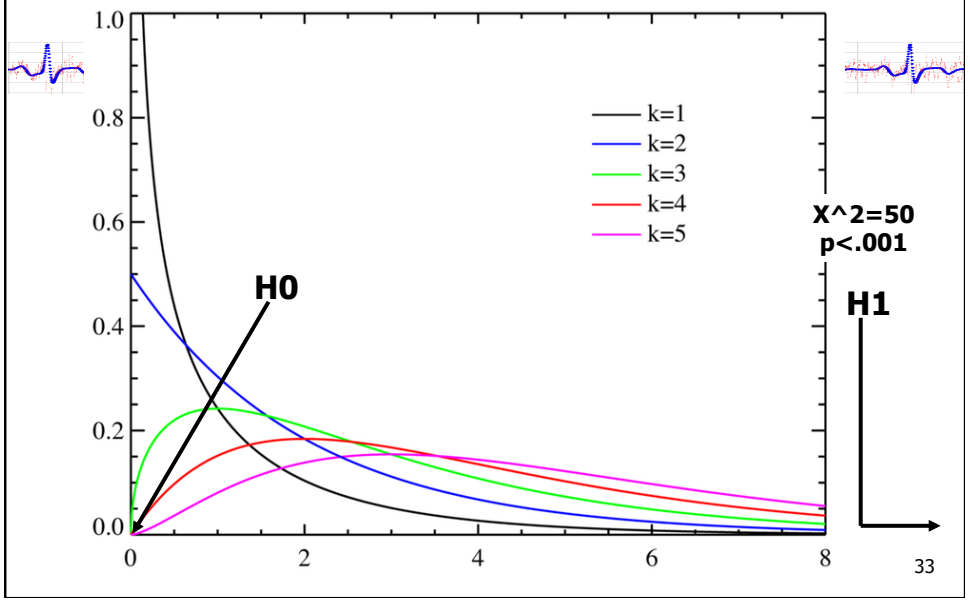
Roll	1	2	3	4	5	6
Count	6	5	7	9	3	30

- Is the die "loaded"?

X^2 = 50

32

Chi-square probability distribution

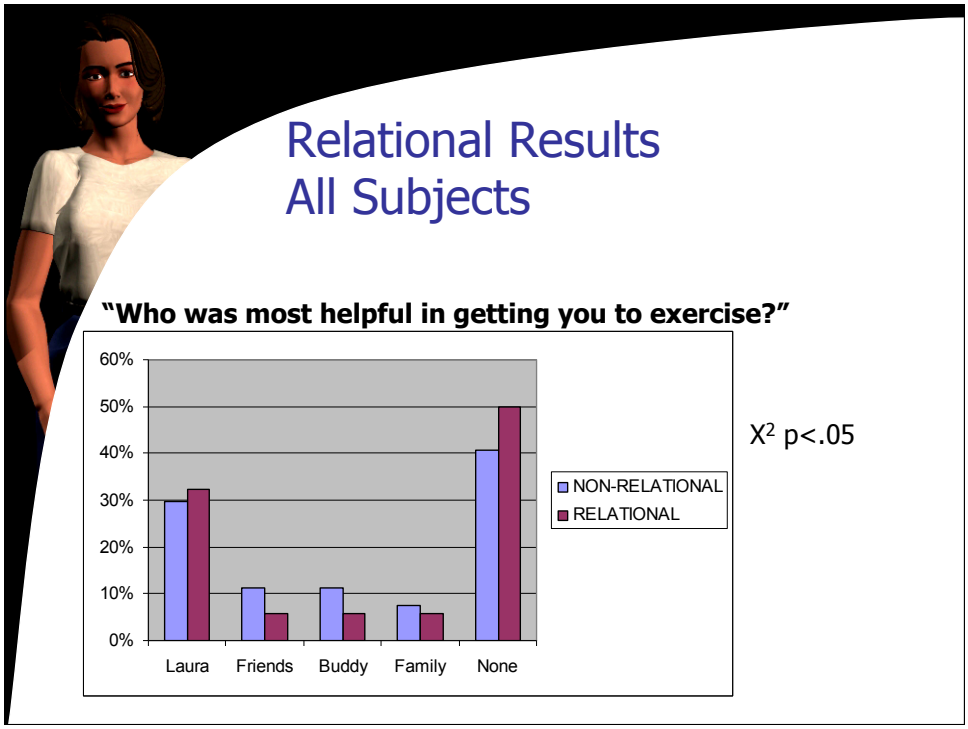


33

Example

- OS market share in US is:
  - 96% PC; 3% Macintosh; 1% Linux
- You do a survey in your company and find the following user breakdown:
  - 440 PC; 9 Mac; 3 Linux
- Is your company weird?

34

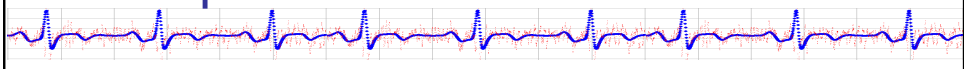


Relational Results
All Subjects

"Who was most helpful in getting you to exercise?"

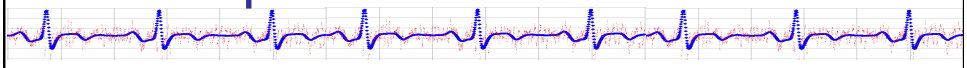
X^2 $p<.05$

What do you do if you don't have any information to base Expected frequencies on?

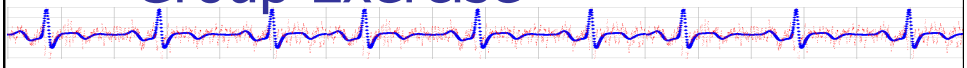


37

Group Exercise



Group Exercise



- For each problem, write
  1. Two populations being compared
  2. Research hypothesis
  3. Null hypothesis
  4. Test criteria
  5. Expected frequencies
  6. Observed frequencies
  7. 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

Chi-square Goodness of Fit test

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

40

Computing Chi-square in R

```
#For x=vector of observed frequency counts…
#assume equiprobable
chisq.test(x)

#If 'frame$y' is a factor or strings, then use
chisq.test(table(frame$y))

#for probs = expected probabilities vector
chisq.test(x, p=probs)
```

- See if significance<threshold
  - Yes => reject H0
  - No => inconclusive

41