Empirical Research Methods in Information Science

IS 4800 / CS6350

Lecture 10

Hypothesis Testing Using Chi^2

1

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"

4

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

5

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?

6

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 $p(H_1|H_0)$?

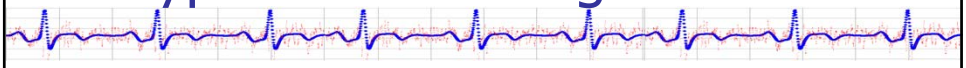
7

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….

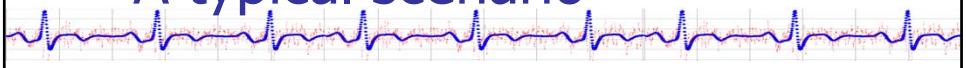
8

Hypothesis Testing



9

A typical scenario



- Between-subject design
- Let every subject try both Wizziword & Word
- Measure performance
- Research Hypothesis:
  - Wizziword is better

10

What if …

- You can test every subject in your population, AND
- There is no measurement error?

→

- Compute descriptives for two treatments
- If Wizziword perf > Word perf conclude H1 is supported
- No uncertainty, No 'p' value

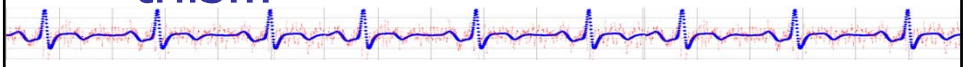
11

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"

12

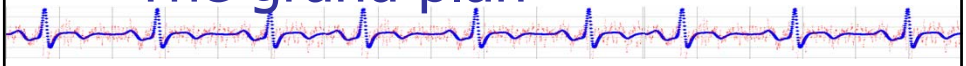
Other ways of thinking about this…



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

13

The grand plan



- X^2 tests
  - For nominal measures
  - Can apply to a single measure
- Correlation tests
  - For two numeric measures
- t-test for independent means
  - For categorical IV, numeric DV

14

Terminology Soup

- Descriptive statistics = describing state of the world
- Inferential statistics = Hypothesis testing
  - When you only have access to a sample
- Univariate analysis/statistics = statistics on only one variable at a time
- Multivariate analysis/statistics = statistics on more than one variable at a time

15

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"

16

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?

17

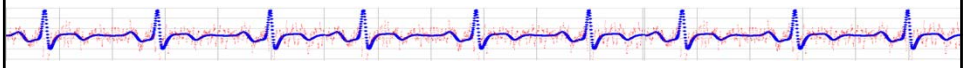
Type of Errors in Hypothesis Testing

		"The Truth"
	H0 True	**H0 False**
Decide to Reject H0	**Type I Error**	**Correct Decision**
Do not Reject H0	**Correct Decision**	**Type II Error**

'p' = Probability of Type I Error

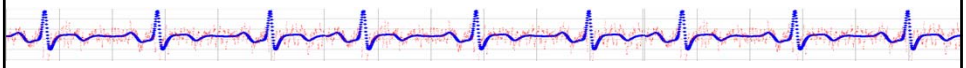
18

Chi-Square for Goodness of Fit



19

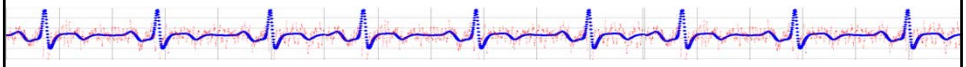
Chi-Square for Goodness of Fit



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

20

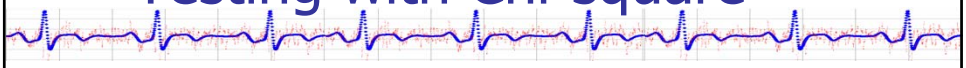
Chi-Square for Goodness of Fit



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

21

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

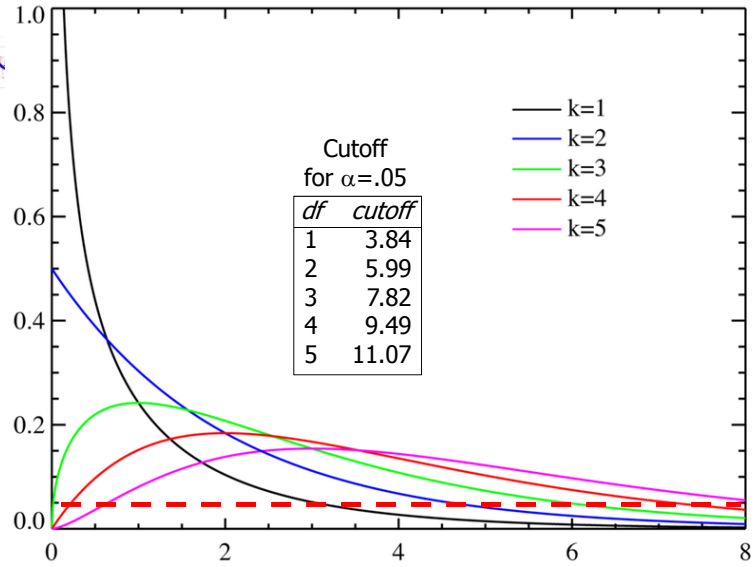
22

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

23

Chi-square probability distribution



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

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").

26

Survey Feb 5, 2013
Guns in Congress

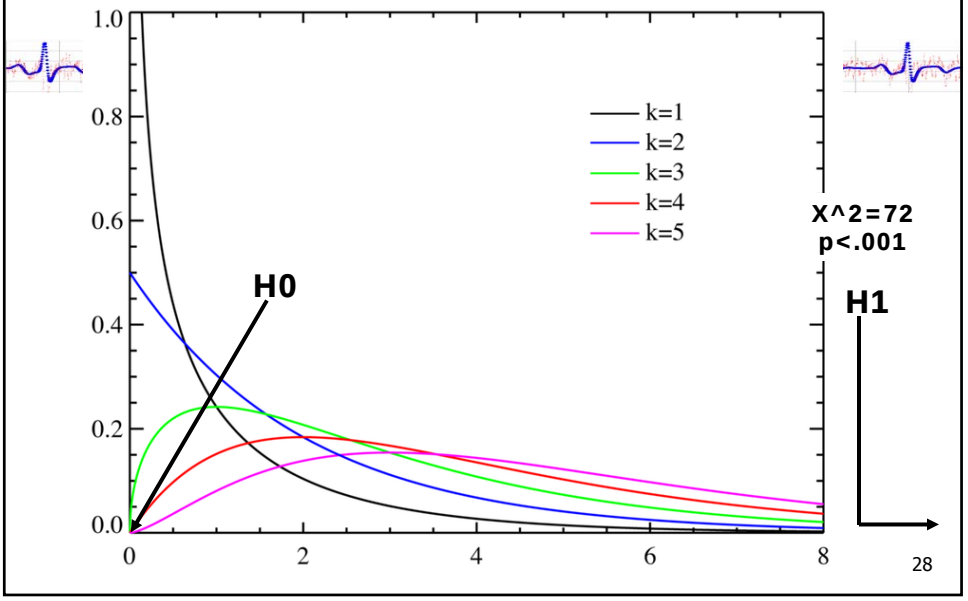
- US: 32% Congress: 55% N=296

Own Gun?	Expected	Observed
Yes	95	163
No	201	133

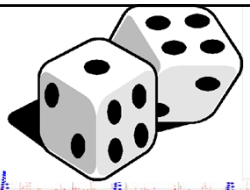
$$X^2 = \frac{(163-95)^2}{95} + \frac{(133-201)^2}{201}$$

$$= 72$$

Chi-square probability distribution



Example



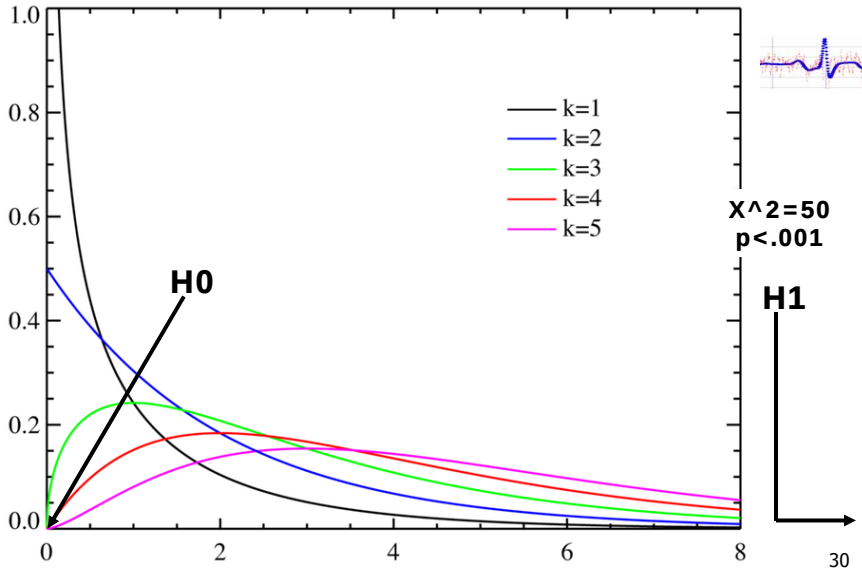
- You go gambling in a shady casino, and suspect that the games are rigged.
- You focus your attention on one 6-sided die being used in a game and keep track of 60 rolls:

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

- Is the die "loaded"?

29

Chi-square probability distribution

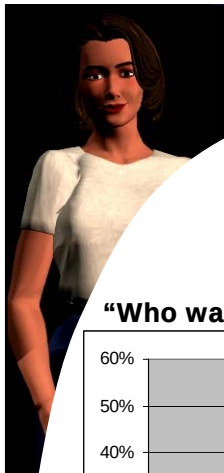


30

Example

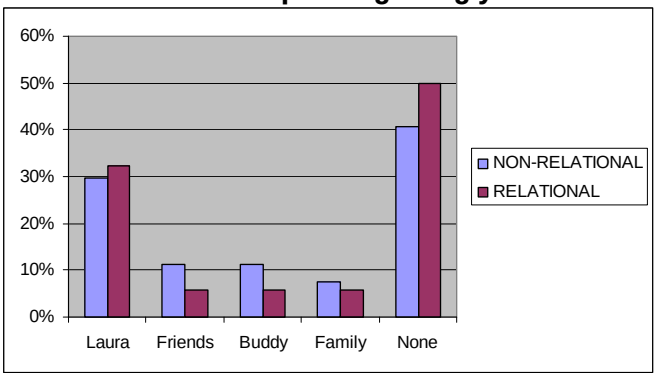
- OS market share in US is:
  - 96% PC; 3% Macintosh; 1% Linux
- You do a survey in your company and find the following user breakdown:
  - 440 PC; 9 Mac; 3 Linux
- Is your company weird?

31



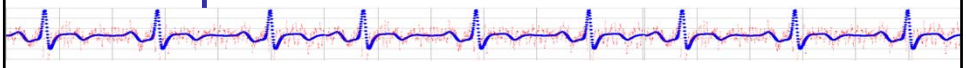
Relational Results All Subjects

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



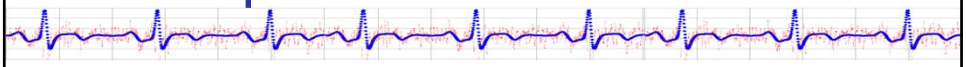
X^2 $p<.05$

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



34

Group Exercise



35

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

36

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

37

Computing Chi-square in R

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

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

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

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

38

Chi-Square Test for Independence

39

Chi-Square Test for Independence

- Are two variables related, or are they independent?
- Assumptons
  - Both variables must be nominal.
  - Cannot be related in a 'special' way (i.e., repeated measures)
  - Random sampling assumed

40

Chi-square Test for Independence

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

41

Example from chapter

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

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

42

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

43

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**	

44

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**	

46

Formula

- Same as goodness-of-fit test.

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

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

47

Survey Feb 5, 2013 Guns in Congress

- Given the data, what questions could we ask about the relatedness of nominal measures?

Title Member Party State GunOwner NRAGrade

- Q: Is Gun Ownership related to party?
- Q: Is Gun Ownership related to NRAGrade?

48

Survey Feb 5, 2013
Guns in Congress

- Q: Is Gun Ownership related to party?

	No	Yes
D	115	44
R	18	119

df=(2-1) x (2–1)
X^2 = 102

49

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=(2-1) x (5–1)
X^2 = 155

50

Group 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

51

X^2 Test for Independence in R

```
#use contingency table
chisq.test(table(x,y))

#e.g., data$favcolor, data$ownmac
chisq.test(table(data$favcolor,
                 data$ownmac))
```

52

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.

54

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

55

Effect size

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

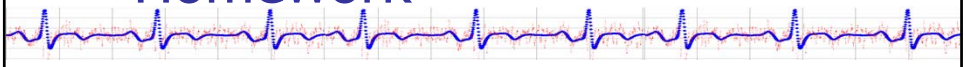
56

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

57

Homework



- Read B&A Ch 13 from 418
- Read Aron Ch 11 (or review stats Pearson correlation)
- *You don't need to memorize formulas.*
- Finish Homework I9 – Designing a Composite Measure for a Questionnaire

58