# Hypothesis Testing

November 2, 2022

# Today

- Finish discussion of hypothesis testing, by focusing on chi-square
- Discussion of one application article

# Analysis of Two-Way Tables

| nes$relig_attend3 | nes$pid_3 Dem | Ind | Rep | Row Total |
|---|---|---|---|---|
| Low | 921 | 1008 | 444 | 2373 |
| | 0.388 | 0.425 | 0.187 | 0.405 |
| Mid | 670 | 524 | 293 | 1487 |
| | 0.451 | 0.352 | 0.197 | 0.254 |
| High | 754 | 603 | 644 | 2001 |
| | 0.377 | 0.301 | 0.322 | 0.341 |
| Column Total | 2345 | 2135 | 1381 | 5861 |

# Analysis of Two-Way Tables

- Difference of means and difference of proportions share a similar approach:
  - Compare two groups
  - Null hypothesis is no difference between groups
  - Examine the difference between groups in the sample and examinethe likelihood that the null hypothesis is true given the observed difference in the sample
- What if we have categorical variables with more than two outcomes?
  - More than two groups—ex., Democrats, Republicans, Independents
  - More than two outcomes—i.e., vote for candidate in a three-party race

# Analysis of Two-Way Tables

- Can make a two-way table (i.e. cross-tab) comparing categorical variables
  - *R* x *C* Table:
    - *R* rows
    - *C* columns
- Independent variable is the column variable
- Dependent variable is the row variable
- We are interested in the conditional distribution of the DV across categories of the IV
- Remember-To determine if the IV is related to the DV, then interested in whether the distribution of the DV is different in different categories of the IV

# Testing the Null Hypothesis

- Null hypothesis: No association between the row variable and the column variable
- If true, then the table would have the same distribution across all columns
- Can determine what the table would look like if the null hypothesis were true
- Need expected cell counts:
  - Expected cell count $= \frac{(row\ Total)(column\ Total)}{N}$
- Compare hypothetical table (with expected cell counts) to the table with observed cell counts from sample

# Example

| | nes$pid_3 | | | |
|---|---|---|---|---|
| nes$relig_attend3 | Dem | Ind | Rep | Row Total |
| Low | 921 | 1008 | 444 | 2373 |
| | 0.388 | 0.425 | 0.187 | 0.405 |
| Mid | 670 | 524 | 293 | 1487 |
| | 0.451 | 0.352 | 0.197 | 0.254 |
| High | 754 | 603 | 644 | 2001 |
| | 0.377 | 0.301 | 0.322 | 0.341 |
| Column Total | 2345 | 2135 | 1381 | 5861 |

# Chi-Square

- Can use chi-square statistic to determine the likelihood that we would get our table if $H_0$ were true
- Steps:
  - Take the difference between each observed count and its corresponding expected count
  - Square all these values (so all are either 0 or positive)
  - Divide each squared difference by the expected count
  - Sum over all cells
- This gives chi-square statistic:
  - $x^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$
- $X^2$ distribution is based on degrees of freedom:
  - Degrees of freedom= $(r\text{-}1)(c\text{-}1)$

# Statistical Significance with Chi-square

- We can determine the p-value using the chi-square distribution
- If $H_0$ is true, $\chi^2$ has approximately a chi-square distribution with (r-1)(c-1) degrees of freedom
- Can use p-value to test for statistical significance
  - Not directional, always only one-tail (chi-square distribution always positive)
  - To determine direction, need to compare conditional distributions, i.e., must assess it yourself

# Chi-Square Example

| nes$relig_attend3 | nes$pid_3 Dem | Ind | Rep | Row Total |
|---|---|---|---|---|
| Low | 921 | 1008 | 444 | 2373 |
| | 0.388 | 0.425 | 0.187 | 0.405 |
| Mid | 670 | 524 | 293 | 1487 |
| | 0.451 | 0.352 | 0.197 | 0.254 |
| High | 754 | 603 | 644 | 2001 |
| | 0.377 | 0.301 | 0.322 | 0.341 |
| Column Total | 2345 | 2135 | 1381 | 5861 |

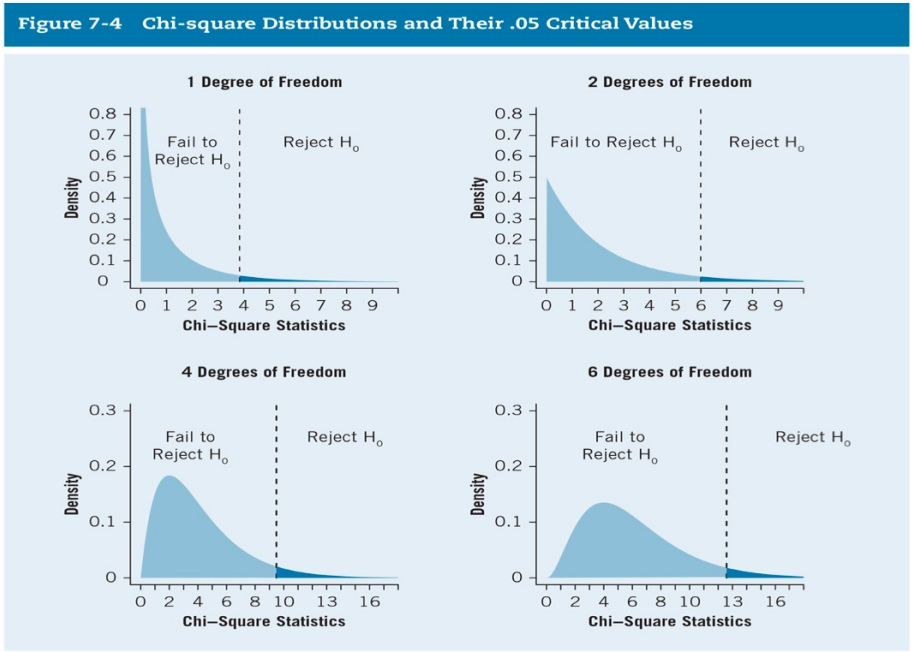
# Chi-square table—critical values

**Percentage Points of the Chi-Square Distribution**

| Degrees of Freedom | Probability of a larger value of $x^2$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.99 | 0.95 | 0.90 | 0.75 | 0.50 | 0.25 | 0.10 | 0.05 | 0.01 |
| 1 | 0.000 | 0.004 | 0.016 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 6.63 |
| 2 | 0.020 | 0.103 | 0.211 | 0.575 | 1.386 | 2.77 | 4.61 | 5.99 | 9.21 |
| 3 | 0.115 | 0.352 | 0.584 | 1.212 | 2.366 | 4.11 | 6.25 | 7.81 | 11.34 |
| 4 | 0.297 | 0.711 | 1.064 | 1.923 | 3.357 | 5.39 | 7.78 | 9.49 | 13.28 |
| 5 | 0.554 | 1.145 | 1.610 | 2.675 | 4.351 | 6.63 | 9.24 | 11.07 | 15.09 |
| 6 | 0.872 | 1.635 | 2.204 | 3.455 | 5.348 | 7.84 | 10.64 | 12.59 | 16.81 |
| 7 | 1.239 | 2.167 | 2.833 | 4.255 | 6.346 | 9.04 | 12.02 | 14.07 | 18.48 |
| 8 | 1.647 | 2.733 | 3.490 | 5.071 | 7.344 | 10.22 | 13.36 | 15.51 | 20.09 |
| 9 | 2.088 | 3.325 | 4.168 | 5.899 | 8.343 | 11.39 | 14.68 | 16.92 | 21.67 |
| 10 | 2.558 | 3.940 | 4.865 | 6.737 | 9.342 | 12.55 | 15.99 | 18.31 | 23.21 |
| 11 | 3.053 | 4.575 | 5.578 | 7.584 | 10.341 | 13.70 | 17.28 | 19.68 | 24.72 |
| 12 | 3.571 | 5.226 | 6.304 | 8.438 | 11.340 | 14.85 | 18.55 | 21.03 | 26.22 |
| 13 | 4.107 | 5.892 | 7.042 | 9.299 | 12.340 | 15.98 | 19.81 | 22.36 | 27.69 |
| 14 | 4.660 | 6.571 | 7.790 | 10.165 | 13.339 | 17.12 | 21.06 | 23.68 | 29.14 |
| 15 | 5.229 | 7.261 | 8.547 | 11.037 | 14.339 | 18.25 | 22.31 | 25.00 | 30.58 |
| 16 | 5.812 | 7.962 | 9.312 | 11.912 | 15.338 | 19.37 | 23.54 | 26.30 | 32.00 |
| 17 | 6.408 | 8.672 | 10.085 | 12.792 | 16.338 | 20.49 | 24.77 | 27.59 | 33.41 |
| 18 | 7.015 | 9.390 | 10.865 | 13.675 | 17.338 | 21.60 | 25.99 | 28.87 | 34.80 |
| 19 | 7.633 | 10.117 | 11.651 | 14.562 | 18.338 | 22.72 | 27.20 | 30.14 | 36.19 |
| 20 | 8.260 | 10.851 | 12.443 | 15.452 | 19.337 | 23.83 | 28.41 | 31.41 | 37.57 |
| 22 | 9.542 | 12.338 | 14.041 | 17.240 | 21.337 | 26.04 | 30.81 | 33.92 | 40.29 |
| 24 | 10.856 | 13.848 | 15.659 | 19.037 | 23.337 | 28.24 | 33.20 | 36.42 | 42.98 |
| 26 | 12.198 | 15.379 | 17.292 | 20.843 | 25.336 | 30.43 | 35.56 | 38.89 | 45.64 |
| 28 | 13.565 | 16.928 | 18.939 | 22.657 | 27.336 | 32.62 | 37.92 | 41.34 | 48.28 |
| 30 | 14.953 | 18.493 | 20.599 | 24.478 | 29.336 | 34.80 | 40.26 | 43.77 | 50.89 |
| 40 | 22.164 | 26.509 | 29.051 | 33.660 | 39.335 | 45.62 | 51.80 | 55.76 | 63.69 |
| 50 | 27.707 | 34.764 | 37.689 | 42.942 | 49.335 | 56.33 | 63.17 | 67.50 | 76.15 |
| 60 | 37.485 | 43.188 | 46.459 | 52.294 | 59.335 | 66.98 | 74.40 | 79.08 | 88.38 |

# Chi-square Distribution

**Figure 7-4 Chi-square Distributions and Their .05 Critical Values**



# Chi-square test example

**Table 7-7 Chi-square Calculation for Opinions on Diplomacy versus Military Force, by Gender**

| | Gender | | |
|---|---|---|---|
| **Favor diplomacy or force?** | ***Female*** | ***Male*** | **Total** |
| Diplomacy | $f_o = 245.0$<br>$f_e = 214.3$<br>$f_o - f_e = 30.7$<br>$(f_o - f_e)^2 = 942.5$<br>$(f_o - f_e)^2/f_e = \mathbf{4.4}$ | $f_o = 176.0$<br>$f_e = 206.7$<br>$f_o - f_e = -30.7$<br>$(f_o - f_e)^2 = 942.5$<br>$(f_o - f_e)^2/f_e = \mathbf{4.6}$ | 40.4%<br>(421) |
| Middle | $f_o = 125.0$<br>$f_e = 135.9$<br>$f_o - f_e = -10.9$<br>$(f_o - f_e)^2 = 118.8$<br>$(f_o - f_e)^2/f_e = \mathbf{.9}$ | $f_o = 142.0$<br>$f_e = 131.1$<br>$f_o - f_e = 10.9$<br>$(f_o - f_e)^2 = 118.8$<br>$(f_o - f_e)^2/f_e = \mathbf{.9}$ | 25.6%<br>(267) |
| Force | $f_o = 160.0$<br>$f_e = 179.7$<br>$f_o - f_e = -19.7$<br>$(f_o - f_e)^2 = 388.1$<br>$(f_o - f_e)^2/f_e = \mathbf{2.2}$ | $f_o = 193.0$<br>$f_e = 173.3$<br>$f_o - f_e = 19.7$<br>$(f_o - f_e)^2 = 388.1$<br>$(f_o - f_e)^2/f_e = \mathbf{2.2}$ | 33.9%<br>(353) |
| Total | 100.0%<br>(530) | 100.0%<br>(511) | 99.9%[a]<br>(1,041) |

*Note:* Shaded cells aren't used directly in the chi-square test statistic formula but are included for reference.

[a]Does not total to 100% due to rounding.

# Chi-square example

- I predict that the more conservative someone is, the more likely they are to drink caffeinated beverages
- I conduct a survey based on an SRS of adults in the United States
- Independent variable—liberal/conservative scale (liberal, moderate, conservative)
- Dependent variable—amount of caffeinated beverages a person consumes (none, some, many)
- I give you the data, make a cross-tab and find the chi-square

# Data for chi-square example

- R1: Cons, Some
- R2: Cons, Some
- R3: Lib, Many
- R4: Lib, Many
- R5: Mod, None
- R6: Cons, None
- R7: Mod, Many
- R8: Lib, Some
- R9: Mod, None
- R10: Mod, Some
- R11: Lib, Many
- R12: Cons, Some
- R13: Lib, None
- R14: Cons, None
- R15: Mod, Some
- R16: Mod, Many
- R17: Cons, Many
- R18: Cons, None
- R19: Lib, Some
- R20: Cons, Many
- R21: Mod, Some
- R22: Mod, None
- R23: Mod, Some
- R24: Mod, Many

# Application—Emotion & Gender Stereotypes in Candidate Evaluation

- Brooks, Deborah Jordan. 2011. "Testing the Double Standard for Candidate Emotionality: Voter Reactions to the Tears and Anger of Male and Female Politicians." The Journal of Politics 73: 597-615.
- Some conventional wisdom suggests voters hold female candidates to a higher standard than male candidates
  - Displays of emotion — e.g., anger and crying
  - Gender stereotypes generate negative reactions by voters
- Do voters hold double standards that disadvantage female candidates?
- No systematic empirical examinations & little theoretical development on the subject

# Overview—Competing Theoretical Perspectives

- (1) Descriptive stereotypes
- Confirmation of a stereotype (i.e., women tend to get more "emotional") is attributed (by voters) as reflecting one's disposition
  - Will be punished more than stereotype-inconsistent behavior
- Stereotypes attribute tearful displays more to women, leading to greater perceptions of weakness among them (compared to men)
- Emotion by men is less likely to be attributed to their disposition or some "personality deficit"
- (2) Prescriptive stereotypes
  - Men are less expected to be emotional, and thus showing emotion is more harmful to them (compared to women)
- (3) Political leaders are largely immune to emotional stereotypes
  - Men and women are penalized equally (or not at all)

# Analytical & Data Overview

- Internet experiment (via Polimetrix): April 17-29, 2009
  - Nationally representative sample
  - 1,120 total respondents, with random assignment
- Two treatment groups & a control group
  - “Crying” condition — 384 respondents
  - “Anger” condition — 371 respondents

# Analytical & Data Overview—the Treatments

- Respondents randomly assigned one of two newspaper articles about a fictional candidate
  - "Congresswoman Karen Bailey" or "Congressman Kevin Bailey"
  - 1-2 paragraphs describe an incident of selected behavior (otherwise same as control group story)
    - I.e., Candidate displaying "crying" or "anger"
- Article describes announcement that he/she is seeking a U.S. Senate seat
  - Describes legislative successes & challenges over previous two years
  - Described as "a moderate and oriented towards bipartisan solutions"

# Treatment Text—Crying Treatment

- Article headline/title & subtitle:
  - "Congresswoman (Congressman) Karen (Kevin) Bailey has a teary week: Cries over campaign rigors, legislative challenges"
- Treatment text:
  - "The announcement has been overshadowed by other events, however. In an incident earlier this week, the congresswoman cried during a speech she was giving while discussing the rigors of campaigning in her last election. During another recent press conference, Representative Bailey cried while being questioned by reporters who were pressing her repeatedly to comment on the failure of a recent bill to which she had devoted a great deal of time. With tears in her eyes, Representative Bailey cut the press conference short and said only 'I really can't talk about this right now.' The congresswoman's office declined to comment on the incidents."

# Treatment Text—Anger Treatment

- Article headline/title & subtitle:
  - "Congresswoman (Congressman) Karen (Kevin) Bailey erupts at colleague, reporter: Calls colleague 'obstinate S.O.B.'"
- Treatment text:
  - "The announcement has been overshadowed by other events, however. In an incident earlier this week, the congresswoman yelled at a committee chair who was trying to block her bill from coming to the House floor, calling him an 'obstinate S.O.B.' Journalists reporting on the committee hearings caught the representative's outburst on tape. In a different incident last month, reporters cornered the congresswoman outside of a Washington, D.C. restaurant to ask for her comments on a recent bill she had sponsored. When reporters persisted in their questioning and continued to follow her down the street, the congresswoman shoved one of the reporters out of the way and yelled, 'I said I was through taking questions tonight, and I meant it.' 'She could have made the same point without pushing me and yelling," commented the reporter after Representative Bailey left the scene. The congresswoman's office declined to comment on the incidents..."

# Treatment Text-Control Group

- Article headline/title & subtitle:
  - “Congresswoman (Congressman) Karen (Kevin) Bailey announces Senate bid”
- No treatment text — just the base text of the hypothetical article (which every group reads)

# Analysis Overview

- Three primary dependent variables (measuring voter perceptions) on a 7-point scale:
  - Overall favorability
  - Likely effectiveness in the Senate
  - Likely effectiveness as U.S. President in 10 years

- Other questions measuring various traits were explored — partly to establish validity of treatments
  - E.g., strong vs. weak; caring vs. uncaring; emotional vs. unemotional; angry vs. calm, etc.
- No controls necessary — Randomization!
- Baseline results = no a priori evidence of emotionality stereotypes

# Empirical Results

TABLE 1. **Summary Results for Crying Experiment**

| | ANOVA RESULTS | | | | | | MEANS | | MEANS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Candidate Gender | | Cry (vs. the control group) | | Candidate Gender * Cry | | Overall | | | | | | |
| | *F* | *p-val* | *F* | *p-val* | *F* | *p-val* | CONTROL (male + female candidates) | CRY (male + female candidates) | CONTROL Male cand | CRY Male cand | CONTROL Female cand | CRY Female cand | *N* |
| **Favorability** | 1.38 | *0.24* | 4.30** | *0.04* | 0.27 | *0.60* | 4.5 | 4.3 | 4.5 | 4.4 | 4.5 | 4.2 | 746 |
| **Senate Effectiveness** | 0.00 | *0.99* | 8.55*** | *0.00* | 0.16 | *0.69* | 4.5 | 4.3 | 4.5 | 4.3 | 4.6 | 4.2 | 746 |
| **Presidential Effectiveness** | 3.63* | *0.06* | 17.41*** | *0.00* | 0.13 | *0.72* | 3.8 | 3.4 | 3.9 | 3.5 | 3.7 | 3.2 | 745 |
| Strong | 0.11 | *0.74* | 68.41*** | *0.00* | 0.50 | *0.48* | 4.4 | 3.6 | 4.4 | 3.6 | 4.4 | 3.5 | 747 |
| Strong Leadership | 0.04 | *0.85* | 35.75*** | *0.00* | 1.08 | *0.30* | 4.5 | 3.8 | 4.4 | 3.9 | 4.5 | 3.8 | 743 |
| Unemotional | 1.32 | *0.25* | 588.33*** | *0.00* | 3.30* | *0.07* | 3.9 | 2.0 | 3.9 | 2.1 | 4.0 | 1.9 | 746 |
| Appropriate | 0.15 | *0.70* | 52.41*** | *0.00* | 0.01 | *0.92* | 4.5 | 3.8 | 4.5 | 3.7 | 4.5 | 3.8 | 747 |
| Situational Attribution | 0.86 | *0.35* | *NA* | | *NA* | | *NA* | | NA | 3.5 | NA | 3.6 | 383 |

*$p \leq .1$, ** $p \leq .05$, *** $p \leq .01$

*Note:* **Favorability/Senate Effectiveness/Presidential Effectiveness** (Most favorable/most effective response=7, Least favorable/least effective response=1), **Strong** (Strong=7, Weak=1), **Strong Leadership** (Describes extremely well=7, Does not describe well at all=1); **Unemotional** (Unemotional=7, Emotional=1), **Appropriate** (Appropriate Behavior=7, Inappropriate Behavior=1), **Situational Attribution** (Situation=7, Personality/Disposition=1).

# Empirical Results

TABLE 2 **Summary Results for Anger Experiment**

| | ANOVA RESULTS | | | | | | MEANS | | MEANS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Candidate Gender | | Angry (vs. the control group) | | Candidate Gender * Angry | | Overall | | | | | | |
| | *F* | *p-val* | *F* | *p-val* | *F* | *p-val* | CONTROL (male + female candidates) | ANGER (male + female candidates) | CONTROL Male cand | ANGRY Male cand | CONTROL Female cand | ANGRY Female cand | *N* |
| **Favorability** | 1.38 | *0.24* | 63.42*** | *0.00* | 0.37 | *0.54* | 4.5 | 3.7 | 4.5 | 3.8 | 4.5 | 3.6 | *735* |
| **Senate Effectiveness** | 0.00 | *0.96* | 39.78*** | *0.00* | 0.10 | *0.75* | 4.5 | 3.9 | 4.5 | 3.9 | 4.6 | 3.9 | *734* |
| **Presidential Effectiveness** | 2.93* | *0.09* | 53.83*** | *0.00* | 0.06 | *0.81* | 3.8 | 3.0 | 3.9 | 3.1 | 3.7 | 2.9 | *735* |
| Calm | 0.27 | *0.61* | 369.93*** | *0.00* | 1.38 | *0.24* | 4.4 | 2.6 | 4.4 | 2.7 | 4.4 | 2.5 | *735* |
| Unemotional | 2.42 | *0.12* | 173.07*** | *0.00* | 4.49** | *0.03* | 3.9 | 2.7 | 3.9 | 2.9 | 4.0 | 2.5 | 734 |
| Appropriate | 2.66* | *0.10* | 216.79*** | *0.00* | 3.59* | *0.06* | 4.5 | 2.9 | 4.5 | 3.1 | 4.5 | 2.7 | *735* |
| Situational Attribution | 1.83 | *0.18* | *NA* | | *NA* | | *NA* | | *NA* | 3.5 | *NA* | 3.3 | *371* |

*$p \leq .1$, **$p \leq .05$, ***$p \leq .01$

*Note:* **Favorability/Senate Effectiveness/Presidential Effectiveness** (Most favorable/most effective response=7, Least favorable/least effective response=1), **Calm** (Calm=7, Angry=1), **Unemotional** (Unemotional=7, Emotional=1), **Appropriate** (Appropriate Behavior=7, Inappropriate Behavior=1), **Situational Attribution** (Situation=7, Personality/Disposition=1).

# Treatment Results

- Crying leads to a small drop in favorability, and moderate penalties in terms of future effectiveness
- But, no evidence that women are penalized more than men
- Similar results for the anger condition
- Later in the study — Brooks considers the survey respondent's gender as an additional modifier
  - I.e., A conditional hypothesis — too complex for us at the moment
  - Basic results show that women respondents/voters penalize female candidates for crying more than they penalize men

# Next steps

- Next week, will wrap up and review these approaches
  - We will focus on another application article that uses these approaches

- Midterm exam #2—week of November 9th
  - Everything since first midterm—e.g., probability theory, sampling distributions, difference of means & proportions, statistical significance & hypothesis testing, etc.