JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section C

The Chi-Square Test: Mechanics

Hypothesis Testing Problem

- Testing equality of two population proportions using data from two samples
  - H_o: $p_1 = p_2$ $\quad$ H_o: $p_1 - p_2 = 0$
  - H_a: $p_1 \neq p_2$ $\quad$ H_A: $p_1 - p_2 \neq 0$

  - In the context of the 2x2 table, this is testing whether there is a relationship between the rows (HIV status) and columns (treatment type)

Statistical Test Procedures

- (Pearson's) Chi-Square Test (x^2)
  - Calculation is easy (can be done by hand)
- Works well for "big" sample sizes
- Gives (essentially) the same p-value as z-test for comparing two proportions
- Unlike z-test, can be extended to compare proportions between more than two independent groups in one test

The Chi-Square Approximate Method

- Looks at discrepancies between observed and expected cell counts in a 2x2 table
  - 0 = observed
  - E = expected = $\frac{\text{row total} \times \text{column tot al}}{\text{grand total}}$

- Expected refers to the values for the cell counts that would be expected if the null hypothesis is true
  - The expected cell proportions if the underlying population proportions are equal

Chi Square Test to Compare Proportions

- Recall, general "recipe" for hypothesis testing . . .
  1. Start by assuming H_o true
  2. Measure distance of sample result from H_o
  3. Compare test statistic (distance) to appropriate distribution to get p-value

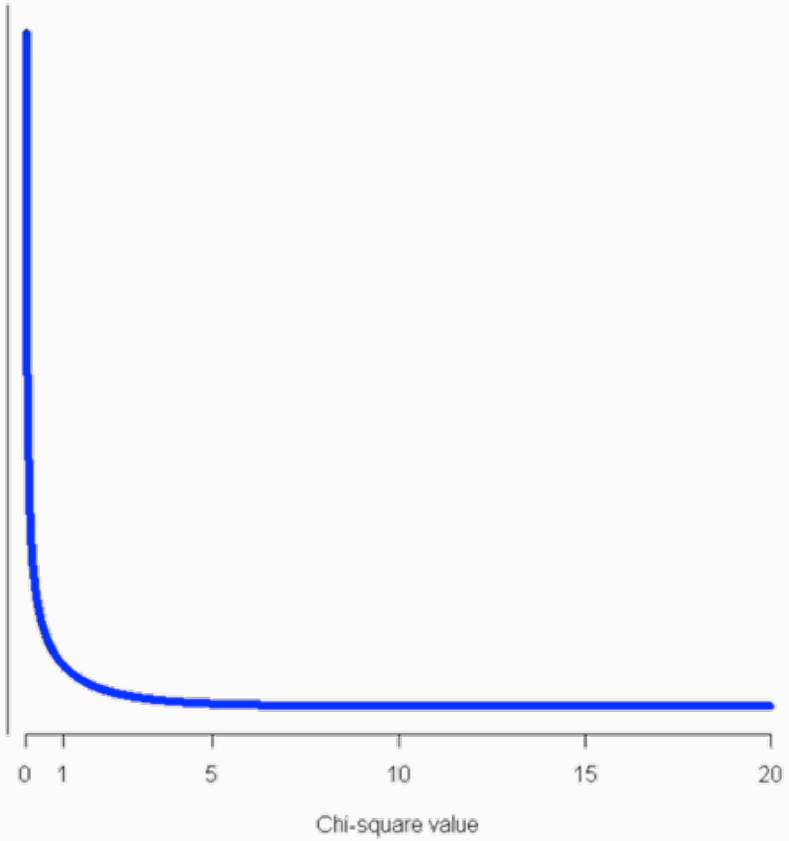
$$\chi^2 = \sum_{4\ cells} \frac{(0 - E)^2}{E}$$

The Chi-Square Approximate Method

- The sampling distribution of this statistic when the null is a chi-square distribution with one degree of freedom

- We can use this to determine how likely it was to get such a big discrepancy between the observed and expected by chance alone

Chi-Square with One Degree of Freedom



Display Data in a 2x2 Table

- 2x2 table setup

		Drug Group		
		AZT	Placebo	
HIV Transmission	Yes	13	40	53
	No	167	143	310
		180	183	363

- The observed value for cell one is 13
- Let's calculate its expected value

Display Data in a 2x2 Table

- Expected value computation: $\frac{\text{row total} \times \text{column total}}{\text{grand total}}$

		Drug Group: AZT	Drug Group: Placebo	
HIV Transmission	Yes	13	40	53
	No	167	143	310
		180	183	363

$$\frac{53 * 180}{363} = 26.3$$

Expected Values

- We could do the same for the other three cells; the below table has expected counts

		Drug Group		
		AZT	Placebo	
HIV Transmission	Yes	26.3	26.7	53
	No	153.7	156.3	310
		180	183	363

Example of Calculations: Chi-Square

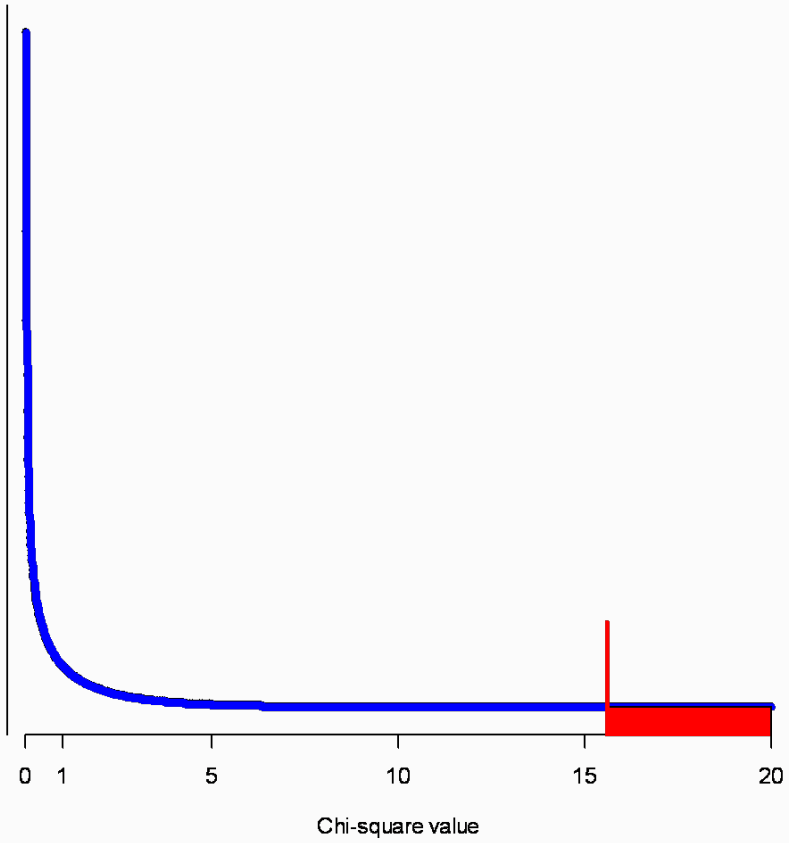
- Test statistic ("distance")

$$\chi^2 = \sum_{4\ cells} \frac{(0 - E)^2}{E}$$

- In our example

$$\chi^2 = 15.6$$

Sampling Dist'n: Chi-Square w/1 Degree of Freedom



Using Stata: AZT/HIV Example

- Results from csi command

```
.   csi 13 40 167 143

                 |   Exposed   Unexposed  |     Total
-----------------+------------------------+----------
           Cases |        13          40  |        53
        Noncases |       167         143  |       310
-----------------+------------------------+----------
           Total |       180         183  |       363
                 |                        |
            Risk |  .0722222    .2185792  |  .1460055
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
 Risk difference |         -.146357       | -.2171766   -.0755374
      Risk ratio |         .3304167       |  .1829884    .5966235
 Prev. frac. ex. |         .6695833       |  .4033765    .8170116
 Prev. frac. pop |         .3320248       |
                 +-----------------------------------------------
                           chi2(1) =    15.59  Pr>chi2 = 0.0001
```

Comparing Proportions between Two Populations

- To create a 95% confidence interval for the difference in two proportions

$$\hat{p}_1 - \hat{p}_2 \pm 2 \times \hat{SE}(\hat{p}_1 - \hat{p}_2)$$

- To get a p-value for testing:
  - H_o: $p_1 = p_2$ $\quad\quad$ H_o: $p_1 - p_2 = 0$
  - H_a: $p_1 \neq p_2$ $\quad\quad$ H_A: $p_1 - p_2 \neq 0$

- Two sample z-test or chi-squared test (give same p-value)

Extendibility of Chi-Squared

- Chi-squared test can be extended to test for differences in proportions across more than two independent populations
  - Analogue to ANOVA with binary outcomes

Extendibility of Chi-Squared

- Example: health care indicators by immigrant status*

TABLE 2—Health Status, Health Care Access, and Health Care Use, by Immigrant Status: National Survey of America's Families, 1999

	US Born, % (SE)		Foreign Born, % (SE)	
	Citizen Parents	Noncitizen Parents	Citizen (Naturalized)	Noncitizen
Health and well being				
Fair/poor current health status	3.91(0.22)	13.52 (1.85)	4.05 (2.65)	12.22 (1.95)
Negative behavior at ages 6–11 y	6.46 (0.56)	6.91 (2.00)	2.74 (1.76)	2.19 (1.05)
Negative behavior at ages 12–17 y	7.50 (0.42)	3.58 (1.94)	1.49 (0.91)	8.46 (2.67)
No involvement in activities at ages 6–17 y	16.75 (0.53)	42.64 (3.29)	15.21 (3.67)	29.12 (2.50)
Health insurance coverage and health care use and access				
Lack of medical insurance at any time in past 12 mo	15.34 (0.55)	34.37 (2.62)	12.86 (3.68)	52.3 (2.77)
No usual source of care other than ER	5.78 (0.27)	18.21(1.93)	12.19 (4.18)	27.93 (2.63)
At least one doctor visit in past year	77.03 (0.54)	65.43 (2.28)	77.04 (5.38)	51.75 (2.48)
ER visit in past year	25.43 (0.47)	23.47 (1.96)	11.59 (3.62)	12.45 (1.72)
At least one visit to dentist in past year (≥3 y old)	80.47 (0.44)	62.73 (2.81)	84.65 (3.42)	55.59 (2.81)
Visit to mental health specialist in past year (≥3y old)	7.17 (0.32)	2.83 (0.89)	5.55 (1.86)	1.77 (0.46)
Subset of items targeted specifically to families with incomes at or below 200% of FPL				
Lack of medical insurance at any time in past 12 mo	26.68 (1.12)	38.76 (3.20)	37.19 (10.58)	68.58 (3.58)
Current Medicaid/SCHIP/state coverage	40.66 (0.95)	48.83 (3.30)	18.28 (5.48)	19.57 (2.50)
Aware of separate SCHIP program	50.25 (1.20)	48.51 (3.14)	46.6 (11.74)	39.71 (4.16)
Aware of Medicaid program	89.85 (0.70)	90.83 (1.27)	89.95 (4.26)	80.79 (3.10)

Note. ER = emergency room; FPL = federal poverty level; SCHIP = State Child Health Insurance Program. All χ^2 *P*s were less than .05.

Notes: *Huang, Z., et al. (2006). Health status and health service access and use among children in U.S. immigrant families, *American Journal of Public Health* 96: 4.

Extendibility of Chi-Squared

- Zoom in

TABLE 2—Health Status, Health Care Access, and Health Care Use, by Immigrant Status: National Survey of America's Families, 1999

	US Born, % (SE)		Foreign Born, % (SE)	
	Citizen Parents	Noncitizen Parents	Citizen (Naturalized)	Noncitizen
Lack of medical insurance at any time in past 12 mo	15.34 (0.55)	34.37 (2.62)	12.86 (3.68)	52.3 (2.77)
No usual source of care other than ER	5.78 (0.27)	18.21(1.93)	12.19 (4.18)	27.93 (2.63)
At least one doctor visit in past year	77.03 (0.54)	65.43 (2.28)	77.04 (5.38)	51.75 (2.48)
ER visit in past year	25.43 (0.47)	23.47 (1.96)	11.59 (3.62)	12.45 (1.72)
At least one visit to dentist in past year (≥3 y old)	80.47 (0.44)	62.73 (2.81)	84.65 (3.42)	55.59 (2.81)
Visit to mental health specialist in past year (≥3y old)	7.17 (0.32)	2.83 (0.89)	5.55 (1.86)	1.77 (0.46)

Note. ER = emergency room; FPL = federal poverty level; SCHIP = State Child Health Insurance Program. All χ^2 *P*s were less than .05.

Extendibility of Chi-Squared

- Zoom in

TABLE 2—Health Status, Health Care Access, and Health Care Use, by Immigrant Status: National Survey of America's Families, 1999

	US Born, % (SE)		Foreign Born, % (SE)	
	Citizen Parents	Noncitizen Parents	Citizen (Naturalized)	Noncitizen
Lack of medical insurance at any time in past 12 mo	15.34 (0.55)	34.37 (2.62)	12.86 (3.68)	52.3 (2.77)
No usual source of care other than ER	5.78 (0.27)	18.21(1.93)	12.19 (4.18)	27.93 (2.63)
At least one doctor visit in past year	77.03 (0.54)	65.43 (2.28)	77.04 (5.38)	51.75 (2.48)
ER visit in past year	25.43 (0.47)	23.47 (1.96)	11.59 (3.62)	12.45 (1.72)
At least one visit to dentist in past year (≥3 y old)	80.47 (0.44)	62.73 (2.81)	84.65 (3.42)	55.59 (2.81)
Visit to mental health specialist in past year (≥3y old)	7.17 (0.32)	2.83 (0.89)	5.55 (1.86)	1.77 (0.46)

Note. ER = emergency room; FPL = federal poverty level; SCHIP = State Child Health Insurance Program. All χ^2 *P*s were less than .05.

Extendibility of Chi-Squared

- Zoom in

TABLE 2—Health Status, Health Care Access, and Health Care Use, by Immigrant Status: National Survey of America's Families, 1999

	US Born, % (SE)		Foreign Born, % (SE)	
	Citizen Parents	Noncitizen Parents	Citizen (Naturalized)	Noncitizen
Lack of medical insurance at any time in past 12 mo	15.34 (0.55)	34.37 (2.62)	12.86 (3.68)	52.3 (2.77)
No usual source of care other than ER	5.78 (0.27)	18.21(1.93)	12.19 (4.18)	27.93 (2.63)
At least one doctor visit in past year	77.03 (0.54)	65.43 (2.28)	77.04 (5.38)	51.75 (2.48)
ER visit in past year	25.43 (0.47)	23.47 (1.96)	11.59 (3.62)	12.45 (1.72)
At least one visit to dentist in past year (≥3 y old)	80.47 (0.44)	62.73 (2.81)	84.65 (3.42)	55.59 (2.81)
Visit to mental health specialist in past year (≥3y old)	7.17 (0.32)	2.83 (0.89)	5.55 (1.86)	1.77 (0.46)

Note. ER = emergency room; FPL = federal poverty level; SCHIP = State Child Health Insurance Program. All χ^2 *P*s were less than .05.