POLITICAL SCIENCE 205
INTRODUCTION TO RESEARCH METHODOLOGY

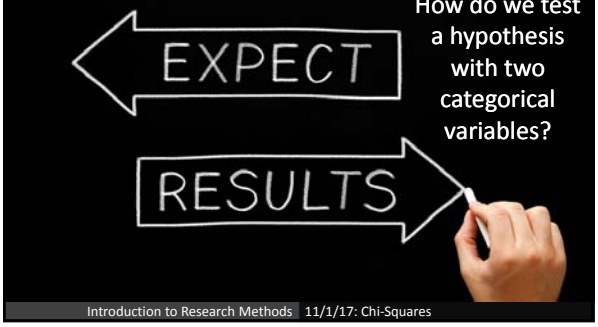
Agenda

I. Quick Review
II. One-Sample Chi-Square
III. Bivariate Chi-Square

Chi-Square
Wednesday, November 1

Week 6, Day 2

Lecture Question



How do we test a hypothesis with two categorical variables?

Quick Review

SPSS Tool/Test	When do you use it?
Descriptive statistics	To describe single variables
Bar graphs & histograms	To visualize single variables
Z-scores	To determine if one score on one variable is average.
T-test for independent samples	To determine if the outcome of a variable is different between two categories
Recoding variables	To collapse the categories of our variable
One-way ANOVA	To determine if the outcome of a variable is different amongst multiple categories

Quick Review

SPSS Tool/Test	Menu Command
Descriptive statistics	Analyze > Descriptive Statistics > Frequencies
Bar graphs & histograms	Analyze > Descriptive Statistics > Frequencies
Z-scores	Analyze > Descriptive Statistics > Descriptives
T-test for independent samples	Analyze > Compare Means > Independent Samples T-Test
Recoding variables	Transform > Recode into Different Variable
One-way ANOVA	Analyze > Compare Means > One-Way ANOVA

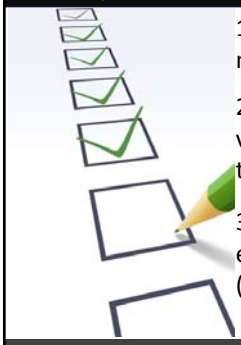
T-Test & ANOVA Reviewed

	T-Test	ANOVA
Definition	The difference in the average scores of one variable (dependent variable) between two groups (two categories of one independent variable).	Examines the difference in the average scores on one variable (dependent variable) between multiple groups (multiple categories of one independent variable).
Test statistic	T-value	F-statistic
Degrees of Freedom	$(n_1\text{-}1) + (n_2\text{-}1)$	(# of groups -1) (# of observations - # of groups)
Significance	Under .05 = The probability the difference is due to chance is under 5%	Under .05 = The probability the difference is due to chance is under 5%
SPSS Command	Analyze > Compare Means > Independent Samples T-Test	Analyze > Compare Means > One-Way ANOVA

One-Sample Chi-Square

Chi-Square: Tests our expected distributions on a single **categorical (nominal or ordinal)** variable.

Assumptions of One-Sample Chi-Square



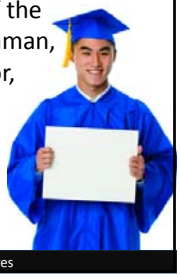
1. We only use this test for nominal or ordinal variables.

2.The test highlights how the values we **observe** are different than what we **expect**.

3. This test is to examine our expectations of a single variable (one-sample).

Hypothesis

Hypothesis: There is a nonrandom difference in the frequencies we observe in the categories of the variable *class standing*: Freshman, Sophomore, Junior, Senior, Graduate Student.



Data on Class Standing

Standing
1
4
5
1
3
4
2
1
1
2

Freshmen (1)	Sophomore (2)	Junior (3)	Senior (4)	Graduate (5)	Total
4	2	1	2	1	10

But is this difference between the observed values and the expected values meaningful?

Freshmen (1)	Sophomore (2)	Junior (3)	Senior (4)	Graduate (5)	Total
2	2	2	2	2	10

The One-Sample Chi-Square

A test statistic that allows us to determine if the values we observe are meaningful OR if they are randomly distributed.

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

X^2 = Chi-square

O = Observed values

E = Expected values

Calculating the One-Sample Chi-Square

Category	Observed Frequency	Expected Frequency	Difference	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Freshman	4	2	2	4	2
Sophomore	2	2	0	0	0
Junior	1	2	-1	1	0.5
Senior	2	2	0	0	0
Graduate	1	2	-1	1	0.5
Total	10	10			**3**

$X^2 = 3$ Always a positive number, as X^2 increases, our observed values are more extreme.

The Chi-Square for our Hypothesis

$X^2 = 3$

Use Table B5 on page 412 to find the chi-square needed to reject the null hypothesis

We need to calculate the degrees of freedom:

df = number of categories – 1
= 5 – 1
= 4

Hypothesis: There is a nonrandom difference in the frequencies we observe in the categories of the variable *class standing*: Freshman, Sophomore, Junior, Senior, Graduate Student.

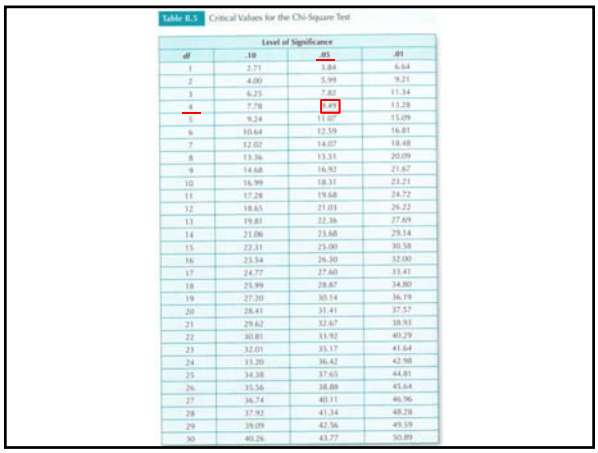


Table B.5 Critical Values for the Chi-Square Test

	Level of Significance		
df	.10	.05	.01
1	2.71	3.84	6.64
2	4.00	5.99	9.21
3	6.25	7.82	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	16.99	18.31	23.21
11	17.28	19.68	24.72
12	18.65	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.60	33.41
18	25.99	28.87	34.80
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.42	42.98
25	34.38	37.65	44.81
26	35.56	38.88	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

The Chi-Square for our Hypothesis

$X^2 = 3$

Use Table B5 on page 412 to find the chi-square needed to reject the null hypothesis

Hypothesis: There is a nonrandom difference in the frequencies we observe in the categories of the variable *class standing*: Freshman, Sophomore, Junior, Senior, Graduate Student.

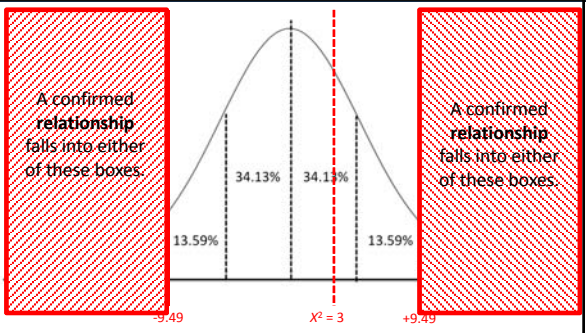
We need to calculate the degrees of freedom:

df = number of categories – 1
= 5 – 1
= 4

$X^2 = 9.49$

This is our value needed to reject the null hypothesis

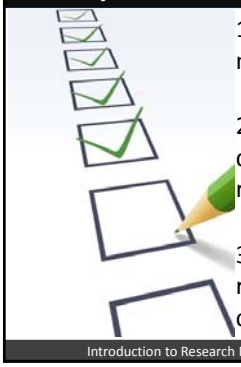
Interpreting the Critical Value



Bivariate Chi-Square

Bivariate Chi-Square: Examines the expected distributions on two **categorical (nominal or ordinal)** variables and whether or not their distributions are related to one another.

Assumptions of the Bivariate Chi-Square



1. We only use this test for nominal or ordinal variables.

2.The test examines how the distribution of one variable is related to another.

3. This test is for examining the relationship between two categorical variables.

Hypothesis

Hypothesis: The categories of **gender** are significantly related to the categories of **party identification.**

Crosstabs

Gender	PartyID
1	2
1	2
2	1
1	1
1	2
2	2
1	1
1	2
2	1
1	2

Let's examine this data in a different format: in a **crosstab.**

A **crosstab** compares the distributions of one variable against another.

	Male	Female	Total
Democrat	2	2	4
Republican	5	1	6
Total	7	3	10

Observed & Expected Values

	Male	Female	Total
Democrat	2	2	4
Republican	5	1	6
Total	7	3	10

Observed Values

These are the observed values of the crosstab. How do we obtain the expected values?

	Male	Female	Total
Democrat	2.8	1.2	4
Republican	4.2	1.8	6
Total	7	3	10

Expected Values

$$\frac{Total_{ROW} \times Total_{COLUMN}}{n}$$

Chi-Square

$$X^2 = \sum \frac{(O-E)^2}{E} \qquad X^2 = 1.269$$

Category	Observed Frequency	Expected Frequency	Difference	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Male Democrat	2	2.8	-.8	.64	0.229
Male Republican	5	4.2	.8	.64	0.533
Female Democrat	2	1.2	.8	.64	0.152
Female Republican	1	1.8	-.8	.64	0.355
Total	10	10			**1.269**

The Chi-Square for our Hypothesis

$X^2 = 1.269$

Use Table B5 on page 412 to find the chi-square needed to reject the null hypothesis

Hypothesis: The categories of **gender** are significantly related to the categories of **party identification.**

We need to calculate the degrees of freedom:

df = (number of rows - 1)(number of columns - 1)
= (2 – 1)(2 – 1)
= (1)(1)
= 1

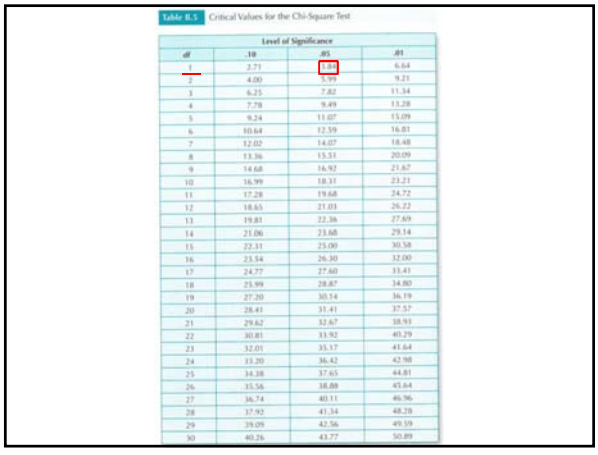


Table B.5 Critical Values for the Chi-Square Test

	Level of Significance		
df	.10	.05	.01
1	2.71	3.84	6.64
2	4.00	5.99	9.21
3	6.25	7.82	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	16.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.60	33.41
18	25.99	28.87	34.80
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.42	42.98
25	34.38	37.65	44.31
26	35.56	38.88	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

The Chi-Square for our Hypothesis

$X^2 = 1.269$

Use Table B5 on page 412 to find the chi-square needed to reject the null hypothesis

Hypothesis: The categories of **gender** are significantly related to the categories of **party identification.**

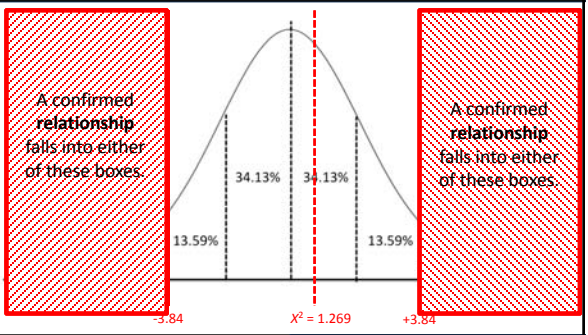
We need to calculate the degrees of freedom:

df = (number of rows - 1)(number of columns - 1)
= (2 – 1)(2 – 1)
= (1)(1)
= 1

$X^2 = 3.84$

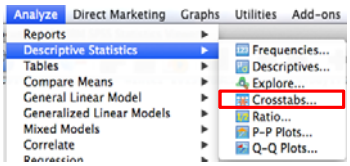
This is our value needed to reject the null hypothesis

Interpreting the Critical Value



Activity 4, Part B

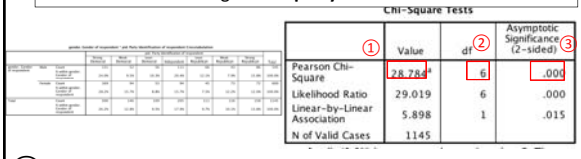
Using SPSS to calculate a Chi-Square



Analyze > Descriptive Statistics > Crosstabs

Interpreting the SPSS Chi-Square

Hypothesis: The categories of **gender** are significantly related to the categories of **party identification.**



Chi-Square Tests

	① Value	② df	③ Asymptotic Significance (2-sided)
Pearson Chi-Square	28.784[a]	6	.000
Likelihood Ratio	29.019	6	.000
Linear-by-Linear Association	5.898	1	.015
N of Valid Cases	1145		

①Chi-square When examining the relationship between **gender** and **party ID**, my chi-square value is **28.784**.

②Degrees of freedom The degrees of freedom for this statistic is **6**.

③Significance value There is a **0.00%** chance that the difference in these frequencies we observe on **gender and party ID** is random. This relationship is **significant** and we accept the hypothesis.

To Review: Chi-Square

Definition: Examines the expected distributions on two **categorical** variables and whether or not their distributions are related to one another.

This is the **Bivariate Chi-Square.** This is more commonly used than the . Remember, you need:

1. Only two variables
2. Nominal or ordinal measurement

- Chi-square

Our **test statistic** that we calculated from the equation. The equation allows us to determine how different our observed values are from what we expect

- Degrees of freedom

The degrees of freedom helps us find our **critical value** to interpret our chi-square. We calculate this by subtracting 1 from both the number of rows and categories; then multiplying those together. We use Table B5 (p. 412) to find our **critical value** needed to reject the null hypothesis.

- Significance value

Only calculated by SPSS, the significance value tells us if our f-statistic is higher or lower than our **critical value**, without having to calculate the chi-square or degrees of freedom on your own.

Activity 4, Part B