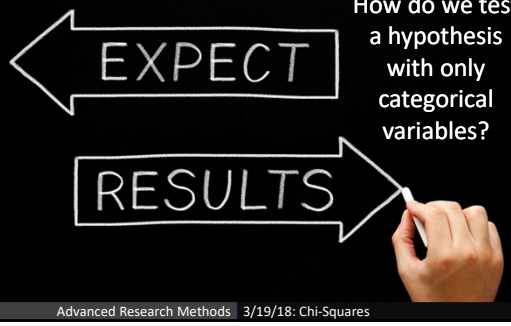
POLITICAL SCIENCE 3055
ADVANCED RESEARCH METHODS

Chi-squares
Tuesday, March 19

Agenda
1. Deciding to use Chi-square
2. One-Sample Chi-Square
3. Bivariate Chi-Square

Week 9, Day 1

Lecture Question



How do we test a hypothesis with only categorical variables?

Deciding to use Chi-square

A **test statistic** refers to the different calculations we use in order to test hypotheses.

1	T-Test
2	ANOVAs
3	Chi-squares
4	Correlations
5	Regressions

Chi-Squares

Chi-Square: Tests our *expectations* of **categorical (nominal or ordinal)** variables.

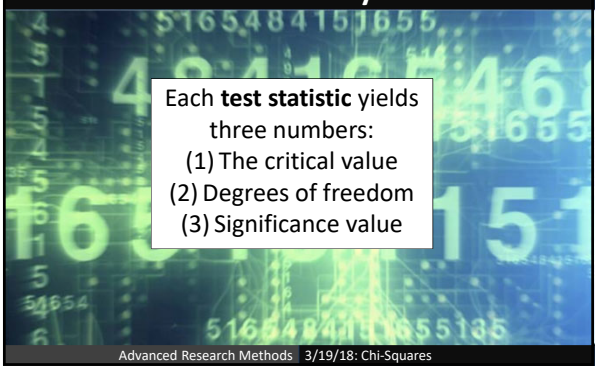
Types of hypotheses

		What type of hypothesis?
1	T-Test	Non-directional
2	ANOVA	Non-directional
3	Chi-square	Non-directional
4	Correlation	Directional
5	Regression	Directional

Levels of Measurement

		Independent variable	Dependent variable
1	T-Test	Nominal or Ordinal (Binary)	Interval-ratio
2	ANOVA	Nominal or Ordinal	Interval-ratio
3	Chi-square	Nominal or Ordinal	Nominal or Ordinal
4	Correlation	Interval-ratio	Interval-ratio
5	Regression	Interval-ratio	Interval-ratio

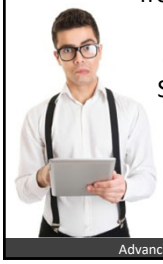
What does a test statistic yield?

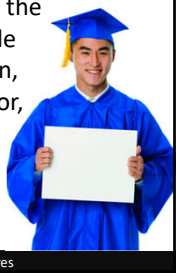


Each **test statistic** yields three numbers:
(1) The critical value
(2) Degrees of freedom
(3) Significance value

One-Sample Chi-Square

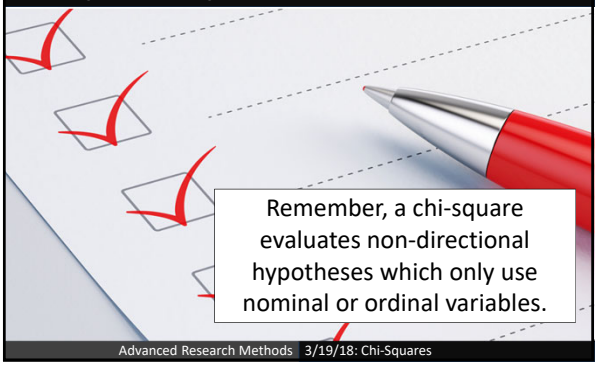
Hypothesis: There is a nonrandom difference in the frequencies we observe in the categories of the variable *class standing*: Freshman, Sophomore, Junior, Senior, Graduate Student.





The One-Sample Chi-Square is intended to demonstrate how this test statistic works by only using one variable.

Chi-square Requirements



Remember, a chi-square evaluates non-directional hypotheses which only use nominal or ordinal variables.

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 of *a single variable* are meaningful OR if the values 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

$X^2 = 3$

We look up the critical value to find where the tails begin and end for this data.

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

$X^2 = 3$

We look up the critical value to find where the tails begin and end for this data.

We need to calculate the degrees of freedom:

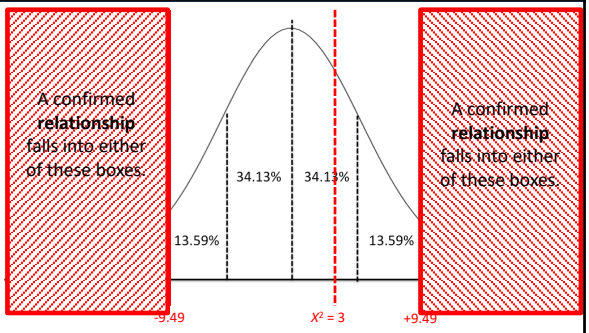
df = number of categories – 1
= 5 – 1
= 4

$X^2 = 9.49$

This is our value needed to reject the null hypothesis

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.

Interpreting the Critical Value



Bivariate Chi-Square

Bivariate Chi-Square: Finally, this test statistic examines how two **categorical (nominal or ordinal)** variables are related to one another in a hypothesis.

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

$X^2 = 1.269$

We look up the critical value to find where the tails begin and end for this data

We need to calculate the degrees of freedom:

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

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

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

$X^2 = 1.269$

We look up the critical value to find where the tails begin and end for this data

We need to calculate the degrees of freedom:

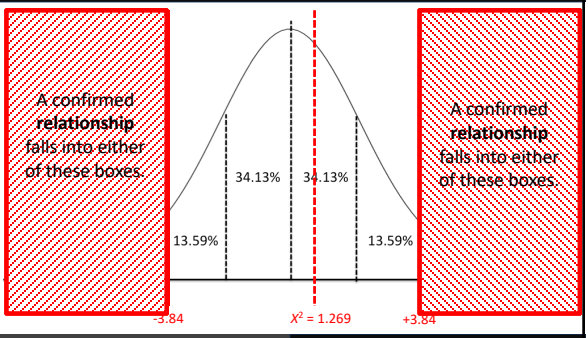
df = (number of rows - 1)(number of columns - 1)
= (2 – 1)(2 – 1)
= (1)(1)
= 1

$X^2 = 3.84$

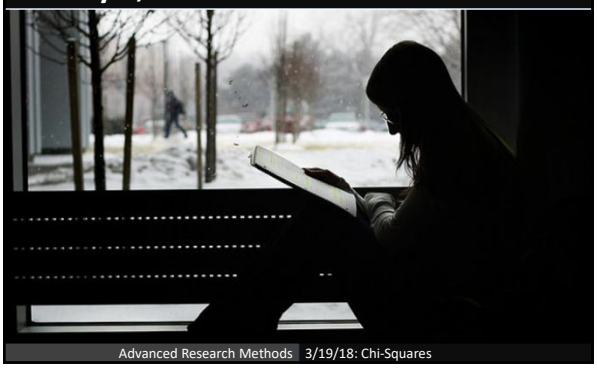
This is our value needed to reject the null hypothesis

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

Interpreting the Critical Value



Activity 4, Part D



What does a Chi-Square give us?

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 the **critical value** to find whether we accept or reject the null hypothesis.

- Significance value

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

The language of the Chi-Square

Chi-square (Critical value)	The chi-square for this hypothesis is 1.269.
The degrees of freedom	The degrees of freedom for this hypothesis is 1.
Statistical Significance	The probability the relationship is due to chance is above 5%. We reject the hypothesis.