The ***Chi-Square*** Test

Not all research involves interval-level data

For example, responses might be:

yes, *no*, *undecided*

strongly agree, *agree*, *disagree*, *strongly disagree*

People might be categorized as

Republican, Democrat, or Independent

School might be classified as

1A, 2A, 3A, etc

Often use the ***Chi-Square Test*** to analyze these type of responses

We look at *association*, rather than means

For example,

Is political affiliation associated with attitude toward an issue?

Is gender associated with selection of an academic major?

Is place of residence associated with attitude toward an issue?

We want to determine if there's any association between a person's political affiliation (*Republican, Democrat, Independent*) and their attitude towards a downtown renovation project (*for, against, undecided*).

Randomly survey 180 people asking them for the political affiliation and how they feel about the proposal.

Create a ***contingency table***

		Political Party Affiliation			
		Republican	Democrat	Independent	Total
View of Downtown Redevelopment Proposal	For	20	20	20	60
	Against	20	20	20	60
	Undecided	20	20	20	60
	Total	60	60	60	180

Marginal Totals

Marginal Totals

View of Downtown Redevelopment Proposal	Political Party Affiliation			
	Republican	Democrat	Independent	Total
For	20	20	20	60
Against	20	20	20	60
Undecided	20	20	20	60
Total	60	60	60	180

Fairly even distribution of cases over all the cells, hence

probably little, if any, association between the two variables

View of Downtown Redevelopment Proposal	Political Party Affiliation			
	Republican	Democrat	Independent	Total
For	40	10	10	60
Against	10	40	10	60
Undecided	10	10	40	60
Total	60	60	60	180

If a greater concentration in a few cells, then

a greater chance that there's some sort of association

If we determine there is an association between political affiliation and viewpoint, then if we know one we can predict the other.

CAUTION: Just because there's association ***does not*** mean there's causation

We may find an association between receiving high grades in math and high grades in science, *BUT* we would not say that receiving high grades in math *CAUSES* high grades in science.

NULL HYPOTHESIS: *There is no association between political affiliation and viewpoint about downtown renovation*.

ANOTHER EXAMPLE:

You randomly sample 98 people at the mall asking them what type of community they live in (*urban, suburban, or rural*) and whether they intend to vote (yes, no, undecided).

Your contingency table is as follows:

Observed Frequencies

Voter Intention	Urban	Suburban	Rural	
Yes	8	17	7	32
No	6	8	15	29
Undecided	19	7	11	37
	33	32	33	

Calculate the marginal totals for each row and each column

OBSERVED FREQUENCIES

EXPECTED Frequencies

Cell = (Row Total *times* Column Total) */* n

Observed Frequencies

Voter Intention	Type of Community: Urban	Suburban	Rural	
Yes	8	17	7	32
No	6	8	15	29
Undecided	19	7	11	37
	33	32	33	

Urban, Yes =

(32 x 33)/98 = 10.78

Suburban, Yes =

(32 x 32)/98 = 10.45

Rural, Yes =

(32 x 33)/98 = 10.78

Urban, No =

(29 x 33)/98 = 9.77

Suburban, No =

(29 x 32)/98 = 9.47

etc

Observed Frequencies

Voter Intention	Type of Community: Urban	Suburban	Rural
Yes	8	17	7
No	6	8	15
Undecided	19	7	11

Expected Frequencies

Voter Intention	Type of Community: Urban	Suburban	Rural
Yes	10.78	10.45	10.78
No	9.77	9.47	9.77
Undecided	12.46	12.08	12.46

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

		Observed	Expected			
		f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
Urban	Yes	8	10.78	-2.78	7.73	**0.72**
Suburban	Yes	17	10.45	6.55	42.90	**4.11**
Rural	Yes	7	10.78	-3.78	14.29	**1.33**
Urban	No	6	9.77	-3.77	14.21	**1.45**
Suburban	No	8	9.47	-1.47	2.16	**0.23**
Rural	No	15	9.77	5.23	27.35	**2.80**
Urban	Undecided	19	12.46	6.54	42.77	**3.43**
Suburban	Undecided	7	12.08	-5.08	25.81	**2.14**
Rural	Undecided	11	12.46	-1.46	2.13	**0.17**

16.38 $= \chi^2$

Degrees of Freedom =

number of rows – 1

times

number of columns – 1

df = (r-1)(c-1)

= (3-1)(3-1)

= 2 x 2

= **4**

Observed Frequencies

		Type of Community		
		Urban	*Suburban*	*Rural*
Voter Intention	*Yes*	8	17	7
	No	6	8	15
	Undecided	19	7	11

Critical Values for Chi-Square (χ^2)

Degrees of Freedom (*df*)	LEVEL OF SIGNIFICANCE		
	.10	.05	.01
1	2.706	3.841	6.635
2	4.605	5.991	9.21[illegible]
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
11	17.275	19.675	24.725
12	18.549	21.026	26.217
13	19.812	22.362	27.688
14	21.064	23.685	29.141
15	22.307	24.996	30.578
16	23.542	26.296	32.000
17	24.769	27.587	33.409
18	25.989	28.869	34.805
19	27.204	30.144	36.191
20	28.412	31.410	37.566
21	29.615	32.671	38.932
22	30.813	33.924	40.289
23	32.007	35.172	41.638
24	33.196	36.415	42.980
25	34.382	37.652	44.314

Critical Value for α=.05
Is 9.488 or **9.49**

$\chi^2 = 16.38$

Hence, we reject the null hypothesis that there is no association between type of community and intention to vote.

POWER

Effect Size: .10 – small

.25 – medium

.40 – large

E.S. $= \sqrt{(\chi^2/(\chi^2+n))}$

$= \sqrt{(16.38/(16.38+98))}$

$= \sqrt{(16.38/115.68)}$

$= \sqrt{0.14}$

= .38

Large effect size, hence we are confident there is an association between type of community and intention to vote