Module 13 – Two Categorical Variables: Tables and the Chi-Square Test

Review Against All Odds: Unit 29 (Inference for Two-Way Tables)

Stat Procedure Diagram – Where we are

		Descriptive Statistics (Describing Pops or Samples)		**Inferential Statistics (from Samples)**
	Variable Types	Display	Describe	Estimation
Univariate	categorical (nominal or ordinal)*	Bar Graph/Pie Chart	Counts/Percentages	*When binary/dichotomous:* Confidence interval for proportions
	quantitative/continuous	Histogram/Stem & Leaf Box Plot	Mean/St Dev (normal) Median/Min, Q1, Q3, Max (skewed)	Confidence interval for means
		Display	Describe	Significance Tests/Hypothesis Tests
Bivariate	2 categorical	Tables or Bar Graphs	Two-way tables/Crosstabulation	Chi-square test (for goodness of fit)
	1 categorical, 1 quant.	Bar Graphs	Comparison of means/averages	T-test (one sample/group, two samples/groups) ANOVA (two or more samples/groups)
	2 quant.	Scatterplot	Correlation Coef. (Coef. of determ)/ Regression Line	T-test for correlation
		Display	Describe	Significance Tests/Hypothesis Tests
Multivariate	Response Variable is Quant.	-	Ordinary Least Squares Regression (OLS)	F-test for overall model T-tests for each explanatory variable
	Response Variable is categorical (dichotomous)	-	Logistic Regression	Chi-square tests of significance

NOTE: Items highlighted in yellow are covered in this course.

**When a categorical variable has two categories, it is called dichotomous.*

Shifting focus to …***comparing two categorical variables (with any number of categories)***

- ***Presenting tables and hypothesis testing for 2 cat. variables***
  - Part 1: Rules for presenting tables
  - Part 2: Hypothesis testing (Chi-Square test)

Part 1: Rules for presenting tables

- 1. Put the Explanatory variable (IV) in the columns and the Response variable (DV) in the rows.
- 2. Display percentages (have the columns add to 100 percent).
- 3. Interpretation sentence A: Compare categories ACROSS the rows
- 4. Interpretation sentence B: Summarize the relationship

Part 1: Rules for presenting tables

- At college graduation, how does the average Mountain State student's quantitative skill level compare to the average Valley State student…Prairie State student?

- Take a random selection of 50 graduating seniors from each school and give them a quantitative skills test with a pass/fail result.

- We will present with a 2 by 3 table
  (2 cat by 3 cat)

Part 1: Rules for presenting tables

	Mountain State	Valley State	Prairie State
Pass (>50)	29	25	21
Fail (<50)	21	25	29
Total	50	50	50

	Mountain State	Valley State	Prairie State
Pass (>50)	58%	50%	42%
Fail (<50)	42%	50%	58%
Total	100%	100%	100%

- 1. Put the Explanatory variable (IV) in the columns and the Response variable (DV) in the rows.
- 2. Display percentages (have the columns add to 100 percent).
- 3. Interpretation sentence A: Compare categories ACROSS the rows
- 4. Interpretation sentence B: Summarize the relationship

Part 1: Rules for presenting tables

	Mountain State	Valley State	Prairie State
Pass (>50)	58%	50%	42%
Fail (<50)	42%	50%	58%
Total	100%	100%	100%

- 1. Put the Explanatory variable (IV) in the columns and the Response variable (DV) in the rows.
- 2. Display percentages (have the columns add to 100 percent).
- 3. Interpretation sentence A: Compare categories ACROSS the rows
- 4. Interpretation sentence B: Summarize the relationship

3. Example: Of all the Mountain State students, 58% passed, compared to 50% at Valley State, and 42% at Prairie State.

4. Example: Mountain state has the highest passing rate (over half passed), followed by Valley State, and then by Prairie State (less than half passed).
At 16%, the difference between the top and bottom passing rate is substantial.

Part 2: Hypothesis Testing

- Testing the relationship between two categorical variables using the Chi-square test. Can we generalize association reflected in the sample to the population?

- Null Hypothesis Statement H_0: There ***is no*** association/relationship between school affiliation and quantitative test result.

- Alternative Hypothesis Statement H_a: There ***is*** an association/relationship between school affiliation and quantitative test result.

- We will calculate a Chi-Square statistic.
- Degrees of Freedom = (r-1) * (c-1) where r= # of rows and c= # of columns
- We compare the Chi-Square statistic to the critical value at alpha=.05 to determine statistical significance

Part 2: Hypothesis Testing

- Calculating the Chi-Square statistic – We'll use observed vs expected results.

Observed Counts				
	Mountain State	Valley State	Prairie State	Total
Pass (>50)	29	25	21	75 (50%)
Fail (<50)	21	25	29	75 (50%)
Total	50	50	50	150 (100%)

Expected Counts				
	Mountain State	Valley State	Prairie State	Total
Pass (>50)	25	25	25	75
Fail (<50)	25	25	25	75
Total	50	50	50	150

$$X^2 = \sum \frac{(\text{Observed value - Expected value})^2}{\text{Expected value}}$$

$X^2 = [(29\text{-}25)^2 / 25] + [(25\text{-}25)^2 / 25] + [(21\text{-}25)^2 / 25] + [(21\text{-}25)^2 / 25] + [(25\text{-}25)^2 / 25] + [(29\text{-}25)^2 / 25]$

$X^2 = (0.64) + (0) + (0.64) + (0.64) + (0) + (0.64)$

$X^2 = 2.56$

Part 2: Hypothesis Testing

- To determine statistical significance we need to find out if our ***test statistic*** is greater than our ***critical value***
- Chi-Square test statistic = 2.56
- Chi-Square critical value with df = (r-1) * (c-1) at alpha = .05
  - (2-1) * (3-1) = df
  - 1 * 2 = 2 df

TABLE D CHI-SQUARE DISTRIBUTION CRITICAL VALUES

	p											
df	**.25**	**.20**	**.15**	**.10**	**.05**	**.025**	**.02**	**.01**	**.005**	**.0025**	**.001**	**.0005**
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42

Part 2: Hypothesis Testing

- Because the ***Chi-Square statistic (2.56)*** is less than the ***critical value (5.99)*** our result is not statistically significant and we ***fail to reject*** the null hypothesis

- The probability that we would get this result if there was no relationship between school affiliation and quantitative test score in the population is ***greater than 5%.*** We do not have strong evidence that the sample result can be generalized to the population.

TABLE D CHI-SQUARE DISTRIBUTION CRITICAL VALUES

	p											
df	**.25**	**.20**	**.15**	**.10**	**.05**	**.025**	**.02**	**.01**	**.005**	**.0025**	**.001**	**.0005**
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42