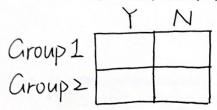
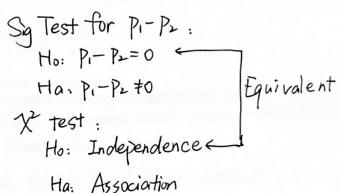
2×2 Contingency Tables and Test for 2 Independent Proportions

For a 2×2 contingency table, the χ^2 test of independence gives the <u>same</u> conclusion as the significance test for comparing two independent proportions. Why? How are these related?





Example: Suppose we want to check if Gender is independent of Vegetarianism. We collect the following data on 679 randomly selected individuals.

Vegetarian?					
Gender	Yes	No	Total		
Male	5	154	159		
Female	38	482	520		
Total	43	636	679		

Assumptions and Hypotheses for x2 test of independence:

SRS - 679 ppl random + representative of poplation of interest

· Min 5 expected wunts under Ho in every cell

Assumptions and Hypotheses for 2 independent proportions test:

· SRS - same

· 10 Succ + fail expected under Ho for each group

Example: The Physician's Health Study I was a randomized, double-blind, placebo-controlled trial whose participants were all male doctors in the USA. It studied the relationship between heart attacks (Y/N) and taking aspirin (Y/N). One treatment group took Aspirin, while another took a placebo daily, for several years. https://phs.bwh.harvard.edu/phs1.htm

The results are summarized in the contingency table below:

	Heart Attack?					
	Yes	No	Total			
Placebo	189	10,845	11,034			
Aspirin	104	10,933	11,037			
Total	293	21,778	22,071			

Have the assumptions been met to conduct a χ^2 test for independence? ·SRS of participants (Not representative of all American adults, at hest to males, doctors (or maybe well-educated trich, higher interest in health) min 5 expected per cell -> Check obs.
 State the hypotheses for this problem.

Ho: No association between Agririn *and Heart attack Ha, Association

Conduct the statistical analysis and state the conclusion.

Pearson's TS: $\chi^2 = 25.014$

p-value 20 from X2 distribution with 1 df.

Rej Ho at all usual a's -> Very strong evidence of association

Compute Conditional Probabilities to describe the association. between Aspirin and Heart attack. % of Aspirin had heart attacks: 104/11037 = . 0094 = 0.94% % of Placebo had heart attacks: 129/11034= .0171=1.71%

Compute the Relative Risk of heart attack for the placebo group vs aspirin.

 $RR = \frac{P_1}{\hat{P}_2} = \frac{1.71\%}{0.94\%} = 1.82$

Interpretation: Ppl who take placebo are almost twice as likely (as ppl who take Aspirin of having a heart attack Not all ppl though!

Identify the test statistics, p-values, and draw conclusions based on the Minitab Output below:

Chi-Square Test: Veg, Not

Expected counts are printed below observed counts Chi-Square contributions are printed below expected counts

		Veg	Not	Total
M-	→ ¹	5 10.07 2.552	154 148.93 0.173	159
F.	→ ²	38 32.93 0.780	482 487.07 0.053	520
	Total	43	636	679

Chi-Sq = 3.558, DF = 1, P-Value = 0.059

Test and CI for Two Proportions

M Sample X N Sample p
$$= 5 \div 159$$
 $= 5 \div 159$ $= 38$ $= 520$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $= 38$ $=$

Difference = p (1) - p (2)
Estimate for difference: -0.0416304

95% CI for difference: (-0.0767909, -0.00646989)

exp: $\frac{43 \times 159}{679} = 10.07$ $\frac{636 \times 159}{679} = 148.93$ $\frac{43 \times 520}{679} = 32.93$ $\frac{636 \times 520}{679} = 487.07$

contribution to TS:

$$\frac{(5-10.07)^2}{(0.07)^2} = 2.552 \qquad \frac{(154-148.93)^2}{(48.93)^2} = 0.173$$

$$\frac{(38-32.93)^{2}}{32.93} = 0.780 \quad \frac{(482-487.07)^{2}}{487.07} = 0.053$$

X= 2.552+0.173+0.780+0.053>3.558

$$DF = (2-1) \times (2-1) = 1$$
 use software P -value = $P(X_1^2 > 3.558) = 0.059$

-0.00646989) Z = -1.89 P-Value = 0.059

District a sees tost : P-Value = 0 063

 $\hat{p} - \hat{p} = \frac{5}{159} - \frac{38}{520} = -0.0416304$

95% CI for p- P2:

$$\hat{p}_1 - \hat{p}_2 \pm Z_{0.025} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

= (-0.0767909, -0.00646989)

Test Statistic:

$$TS \in Z = \frac{\vec{p}_1 - \vec{p}_2 - 0}{\vec{p}_1 - \vec{p}_1 \cdot (\vec{p}_1 + \vec{p}_2)} = \frac{\vec{5}_9 - \frac{38}{520}}{\sqrt{\frac{47}{679} \times (1 - \frac{47}{679}) \times (\frac{1}{159} + \frac{1}{520})}}$$

$$= -1.89$$

P-value = 2P(2<-1.89) = 0.059

a Fandom variable that follows x distribution with degrees of freedom 1.

$$\angle Z^2 = \chi_i^2$$

$$t_n = \frac{X \rightarrow N(0,1)}{\sqrt{Y/n} \times 1}$$

Extra Credit (0.5 pts)

Example: Collect data in class, or use the General Social Survey to answer a question of interest involving two categorical variables

Variables: Sexuality & Dating apps · Assumptions: Pating apps · Ho; Hetero 80 40 120 .TS:

Not 7 17 24 .DF:

87 57 144 .Pvalue:

· Conditional Prob: % who use Dating Apps for / not:

· Marginal distributions for sexuality and Dating apps usage:

· PR of using dating apps for the heterosexuals v.s. not heterosexuals: Interpretation of the RR.

· Joint distribution:

Deadline: March 91th 11:59 pm (No email; No paper)
Upload a polf file on Canvas "Extra Credit 2"

Nonparametric Statistics (Ch. 15)

Overview of Nonparametric Methods

- Normal-based procedures (Z and t tests) assume that the statistic has an approximately Normal distribution, either because the sample sizes are large enough or because the original distribution is
- When utilizing Nonparametric methods, we still have a parameter but we have no distribution assumptions for the response variable. In order to use these methods, our response variable Y must must be quantitative.
- Nonparametric methods make inferences about: Populations But not about means, instead inferencens about median or entire distribution.
- Nonparametric methods are useful when:
 - small sample Size (with outliers)
 - -several groups with very different variances
 - -data is quartificative but NOT continuous particularly

 Subjective ratings
- Nonparametric methods generally only need to assume:

-SRS from population of interest

Remarks:

Almost All the nonparametric methods we see in this class are based on RANKS

- · If assumptions of Normality and Equal Variances are met, then both parametric (normal-based) and nonparametric (distribution-free) methods can be used.
- · However, parametric methods are statistically more powerful and are able to find true significant differences (assuming all assumptions are met), while nonparametric methods sometimes cannot.