14.E: Chi-square (Exercises)

1. What does a frequency table display? What does a contingency table display?

Answer:

Frequency tables display observed category frequencies and (sometimes) expected category frequencies for a single categorical variable. Contingency tables display the frequency of observing people in crossed category levels for two categorical variables, and (sometimes) the marginal totals of each variable level.

2. What does a goodness-of-fit test assess?
3. How do expected frequencies relate to the null hypothesis?

Answer:

Expected values are what we would observe if the proportion of categories was completely random (i.e. no consistent difference other than chance), which is the same was what the null hypothesis predicts to be true.

4. What does a test-for-independence assess?
5. Compute the expected frequencies for the following contingency table:

	Category A	Category B
Category C	22	38
Category D	16	14

Answer:

Observed	Category A	Category B	Total
Category C	22	38	60
Category D	16	14	30
Total	38	52	90

Expected	Category A	Category B	Total
Category C	$((60*38)/90) = 25.33$	$((60^*52)/90) = 34.67$	60
Category D	$((30*38)/90) = 12.67$	$((30*52)/90) = 17.33$	30
Total	38	52	90

6. Test significance and find effect sizes (if significant) for the following tests:
   1. $N = 19, R = 3, C = 2, \chi 2(2) = 7.89, \alpha = .05$
   2. $N = 12, R = 2, C = 2, \chi 2(1) = 3.12, \alpha = .05$
   3. $N = 74, R = 3, C = 3, \chi 2(4) = 28.41, \alpha = .01$
7. You hear a lot of people claim that The Empire Strikes Back is the best movie in the original Star Wars trilogy, and you decide to collect some data to demonstrate this empirically (pun intended). You ask 48 people which of the original movies they liked best; 8 said A New Hope was their favorite, 23 said The Empire Strikes Back was their favorite, and 17 said Return of the Jedi was their favorite. Perform a chi-square test on these data at the .05 level of significance.

Answer:

Step 1: H_0: "There is no difference in preference for one movie", H_A: "There is a difference in how many people prefer one movie over the others."

Step 2: 3 categories (columns) gives $df = 2$, $\chi^2_{\text{crit}} = 5.991$.

Step 3: Based on the given frequencies:

	New Hope	Empire	Jedi	Total
Observed	8	23	17	48
Expected	16	16	16	

$\chi^2 = 7.13$.

Step 4: Our obtained statistic is greater than our critical value, reject H_0. Based on our sample of 48 people, there is a statistically significant difference in the proportion of people who prefer one Star Wars movie over the others, $\chi^2(2) = 7.13$, $p < .05$. Since this is a statistically significant result, we should calculate an effect size: Cramer's $V = \sqrt{\frac{7.13}{48(3-1)}} = 0.27$, which is a moderate effect size.

8. A pizza company wants to know if people order the same number of different toppings. They look at how many pepperoni, sausage, and cheese pizzas were ordered in the last week; fill out the rest of the frequency table and test for a difference.

	Pepperoni	Sausage	Cheese	Total
Observed	320	275	251	
Expected				

9. A university administrator wants to know if there is a difference in proportions of students who go on to grad school across different majors. Use the data below to test whether there is a relation between college major and going to grad school.

		Major		
		Psychology	Business	Math
Graduate School	Yes	32	8	36
	No	15	41	12

Answer:

Step 1: H_0: "There is no relation between college major and going to grad school", H_A: "Going to grad school is related to college major."

Step 2: $df = 2$, $\chi^2_{\text{crit}} = 5.991$.

Step 3: Based on the given frequencies:

Expected Values		**Major**		
		Psychology	Business	Math
Graduate School	Yes	24.81	25.86	25.33
	No	22.19	23.14	22.67

$\chi^2 = 2.09 + 12.34 + 4.49 + 2.33 + 13.79 + 5.02 = 40.05$

Step 4: Obtained statistic is greater than the critical value, reject H_0. Based on our data, there is a statistically significant relation between college major and going to grad school, $\chi^2(2) = 40.05$, p < .05, Cramer's $V = 0.53$, which is a large effect

10. A company you work for wants to make sure that they are not discriminating against anyone in their promotion process. You have been asked to look across gender to see if there are differences in promotion rate (i.e. if gender and promotion rate are independent or not). The following data should be assessed at the normal level of significance:

		Promoted in last two years?	
		Yes	No
Gender	Women	8	5
	Men	9	7