Chi-Square Tests

PSYC 300B - Lecture 7
Dr. J. Nicol

Chi-Square Tests

- Used to analyze data measured on a ***nominal scale*** (i.e., categorical data)
- The data are ***frequencies*** of observations that fall into each category of the independent variable(s)

Assumptions of Chi-Square Tests

- Chi-square tests are non-parametric, but they do make two important assumptions:
  - Each observation comes from a different individual
  - The expected frequency of a cell is 5 or more
  - When the expected frequencies are too small the probability of making a Type I error is distorted and chi-square may not be a valid test (Howell, 2014)

Observed and Expected Frequency

- Chi-square tests compare observed and expected frequencies
- ***Observed frequency***: the number of individuals from the sample who are classified in a particular category
- ***Expected frequency:*** the value that is predicted from the proportions in the H_0 and the sample size (N) (i.e., they are the frequencies you would expect if H_0 were true)
- Expected frequencies define the shape of a distribution that would be obtained if the sample proportions were in perfect agreement with the proportions specified in the H_0

The Goodness-of-Fit Test

- The test uses a sample frequency distribution to test hypotheses about the proportions (or shape) of a population frequency distribution
- The test asks whether the deviations from the proportions stated in H_0 are large enough to permit the conclusion that they are not a chance occurrence
- Determines whether the frequency distribution across the categories of a nominal variable is significantly different than the proportions specified by H_0

Formula for the Goodness-of-Fit Test

$$\chi^2 = \sum \frac{(FREQ_{OBSERVED} - FREQ_{EXPECTED})^2}{FREQ_{EXPECTED}}$$

$$FREQ_{EXPECTED} = p(N)$$

$$df = \#cells - 1$$

Effect Size

$$w = \sqrt{\frac{\chi^2}{N}}$$

- Interpreting the magnitude of *w* (Cohen, 1992):
  - 0.10 is a small effect
  - 0.30 is a medium effect
  - 0.50 is a large effect

The Chair of the Psychology Department suspects that some of her faculty are more popular than others, so she asks a sample of N = 185 psychology students to report their favourite professor

Professor A	Professor B	Professor C	Professor D
47	45	62	31

Conduct a hypothesis test (α = 0.05) to determine if the observed data support the Chair's suspicion

A developmental psychologist would like to determine if infants have colour preferences. In a preferential looking task she records the amount of time that a sample of N = 80 infants spent looking at each of four displayed colours. The one that was looked at the longest was identified as the infant's preferred colour

Red	Green	Blue	Yellow
25	18	23	14

Conduct a hypothesis test (α = 0.05) to determine if the data indicate that infants have a colour preference

A psychology professor would like to determine whether there has been a significant change in distribution of grades in his department over the years. It is known that the overall grade distribution for the department in 2010 was: 14% A's, 26% B's, 31% C's, 19% D's and 10% F's

A	B	C	D	F
42	64	61	21	12

A sample of N = 200 psychology students from last semester produced the above grade distribution

Conduct a hypothesis test (α = 0.01) to determine if the data indicate a significant change in the grade distribution

The Test of Independence

- Used when we are interested in asking whether the frequency distribution of one variable is contingent (i.e., related to) on a second variable
- Each person in the sample is classified on both of the two variables, creating a two-dimensional contingency table
- Two variables are independent when there is no consistent, predictable relationship between them (i.e., the frequency distribution for one variable is not related to, or dependent on, the categories of the second variable)

The Test of Independence

- As a result, when two variables are independent, the frequency distribution for one variable will have the same proportions for all categories of the second variable
- Look at row and column percentages to interpret effects: these percentages reflect the patterns of data better than the frequencies themselves (because the frequencies will be dependent on the sample sizes in different categories)

Formula for The Test of Independence

$$\chi^2 = \sum \frac{(FREQ_{OBSERVED} - FREQ_{EXPECTED})^2}{FREQ_{EXPECTED}}$$

$$FREQ_{EXPECTED} = \frac{(FREQ_{COLUMN})(FREQ_{ROW})}{N}$$

$$df = (\#columns - 1)(\#rows - 1)$$

$$Effect\ Size: \ w = \sqrt{\frac{\chi^2}{N}}$$

Researchers collected data from older (N = 60) and younger (N = 60) adults concerning their preferred cell phone model after using each one for a month

	Cell Phone		
Age Group	Model A	Model B	Model C
Younger	27	20	13
Older	21	34	5

Conduct a hypothesis test (α = 0.05) to determine if the data indicate that cell phone model preference is related to age?

Conduct a hypothesis test (α = .05) to determine if there is a relationship between the type of music the females were exposed to and their willingness to give out their number

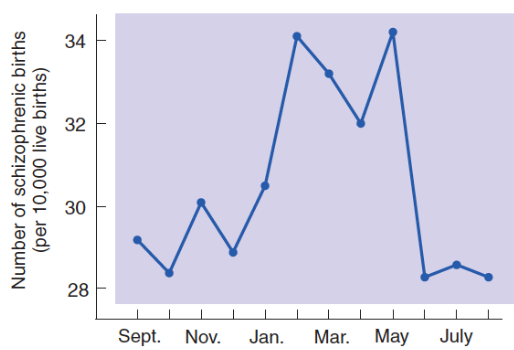
	Gave Phone Number?	
Music	Yes	No
Romantic	23	21
Neutral	12	31
Total	1,303	1,320

Calculate and interpret the odds ratio comparing the likelihood females gave out their number after listening to romantic music to the likelihood they did if they had listened to neutral music

Odds Ratio

- An odds ratio is an elegant and easily understood metric for expressing effect size (Field, 2013, p.744-745)
- They are most interpretable in 2 × 2 contingency tables are not useful for larger contingency tables
- $\text{Odds}_{\text{GAVE NUMBER\_ROMANTIC}} = 23/21 = 1.1$
- $\text{Odds}_{\text{GAVE NUMBER\_NEUTRAL}} = 12/31 = 0.39$
- Odds Ratio= 1.1/0.39 = 2.82
- The odds are 2.82 times higher that a female gave her number after listening to romantic than neutral music

A disproportionate amount of people with schizophrenia were born in late winter and early spring



Kendell & Adams (1991)

Researchers collected date-of-birth information from a sample of N = 100 healthy adults and N = 50 adults diagnosed with schizophrenia

	Season of Birth			
Schizophrenia	Summer	Fall	Winter	Spring
No Disorder	26	24	22	28
Disorder	9	11	18	12

Conduct a hypothesis test (α = 0.05) to determine if the data indicate that schizophrenia is related to season of birth?