GENERAL INSTRUCTIONS - This is due by 11:59 PM on Saturday Nov. 2, 2024. When you are finished making your final edits, simply walk away from this document. If you are completing both data sets, this will satisfy your Midterm Exam 2 requirements (...you can ignore any late/missing notifications from Canvas)

Chi-Square Analysis - Version A.1

A researcher would like to perform a preliminary study regarding which factors are most important to people who are pursuing a new relationship. A sample of 200 "interested" singles ("looking") between the ages of 30 and 40 are asked to identify the most important factor in the decision process (positive affect—"happy" vs. reliable vs. physically attractive). To start, the researcher would like to know whether there is a difference between the factors identified by cis-gender women compared to those identified by men. The data are as follows:

	"Happy"	**Reliable**	**Attractive**
Male	**24**	**28**	**45**
Female	**23**	**61**	**19**

Using a chi-square analysis, with $\alpha = 0.05$, what can you argue about the data (note: stating *statistically significant* is not enough)? Regardless of whether the results are significant or not, describe the trends in the data and supply the necessary summary statistics for the Chi-Square analysis as you would in a formal write-up (i.e., in APA style). Click here for Chi-Square critical values.

SHOW ALL WORK - Upload a clear snapshot of all your base calculations to this doc. Include a summary analysis with APA notation. (Append your work below, along with your APA summary and notation, typed)

THIS SPACE FOR YOUR WORK

Calculation:

Firstly, we find row totals, column totals, and recall the sample size (N) as follows.
$R_{Male} = 24 + 28 + 45 = 97$
$R_{Female} = 23 + 61 + 19 = 103$
$C_{Happy} = 24 + 23 = 47$
$C_{Reliable} = 28 + 61 = 89$
$C_{Attractive} = 45 + 19 = 64$
$N = R_{Male} + R_{Female} = 97+103 = 200$
$N = C_{Happy} + C_{Reliable} + C_{Attractive} = 47 + 89 + 64 = 200$

With the row totals, column totals, and sample size in hand, we can now calculate the expected frequency values.

GENERAL INSTRUCTIONS - This is due by 11:59 PM on Saturday Nov. 2, 2024. When you are finished making your final edits, simply walk away from this document. If you are completing both data sets, this will satisfy your Midterm Exam 2 requirements (...you can ignore any late/missing notifications from Canvas)

Because the expectation value for a given row and column is simply the product of the respective row and column totals divided by the sample size

$$E_{i,j} = \frac{R_i C_j}{N}$$

we have

$E_{Male,Happy} = \frac{97*47}{200} = 22.795$;$E_{Male,Reliable} = \frac{97*89}{200} = 43.165$; $E_{Male,Attractive} = \frac{97*64}{200} = 31.04$

$E_{Female,Happy} = \frac{103*47}{200} = 24.205$; $E_{Female,Reliable} = \frac{103*89}{200} = 45.835$;$E_{Female,Attractive} = \frac{103*64}{200} = 32.96$

Now with the expectations in hand, we're ready to calculate the respective contributions to the chi-square statistic. We call each contribution $C_{i,j} = \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$ where

$$\chi 2 = \sum_{i,j} C_{i,j}$$

$$C_{Male,Happy} = \frac{(O_{Male,Happy} - E_{Male,Happy})^2}{E_{Male,Happy}} = \frac{(24-22.795)^2}{22.795} = 0.0636$$

$$C_{Male,Reliable} = \frac{(O_{Male,Reliable} - E_{Male,Reliable})^2}{E_{Male,Reliable}} = \frac{(28-43.165)^2}{43.165} = 5.328$$

$$C_{Male,Attractive} = \frac{(O_{Male,Attractive} - E_{Male,Attractive})^2}{E_{Male,Attractive}} = \frac{(45-31.04)^2}{31.04} = 6.278$$

$$C_{Female,Happy} = \frac{(O_{Female,Happy} - E_{Female,Happy})^2}{E_{Female,Happy}} = \frac{(23-24.205)^2}{24.205} = 0.0600$$

$$C_{Female,Reliable} = \frac{(O_{Female,Reliable} - E_{Female,Reliable})^2}{E_{Female,Reliable}} = \frac{(61-45.835)^2}{45.835} = 5.018$$

$$C_{Female,Attractive} = \frac{(O_{Female,Attractive} - E_{Female,Attractive})^2}{E_{Female,Attractive}} = \frac{(19-32.96)^2}{32.96} = 5.913$$

Now summing up the contributions:

$$\chi 2 = \sum_{i,j} C_{i,j} = 0.0636 + 5.328 + 6.278 + 0.0600 + 5.018 + 5.913 = 22.660$$

Now recall that the degrees of freedom is calculated as

$$df = (row\ dimension - 1)(column\ dimension - 1) = (2 - 1)(3 - 1) = 2$$

With a significance level of $\alpha = 0.05$, $df = 2$ we find from a $\chi 2$ critical value tabulation that the $\chi 2$ critical value is 5.991.

As our $\chi 2$ test statistic (22.660) is greater than the critical value (5.991), we reject the null hypothesis under consideration, namely that there is no significant interaction between gender (male, female) and other factors (happy, reliable, attractive). The observed interaction appears to be unlikely to have resulted from random chance alone.

GENERAL INSTRUCTIONS - This is due by 11:59 PM on Saturday Nov. 2, 2024. When you are finished making your final edits, simply walk away from this document. If you are completing both data sets, this will satisfy your Midterm Exam 2 requirements (...you can ignore any late/missing notifications from Canvas)

Before we move on to the summary, note the significance of the contributions of different rows and columns to the chi-square test statistic. Namely we see that the "happy" column produced contributions remarkably smaller than the other factor columns. We can make this claim more rigorous by examining the standardized pearson residuals as calculated in some linked python code

Pearson Standardized Residuals
[[0.40209893 -4.31744124 4.23415245]
[-0.40209893 4.31744124 -4.23415245]]

and observing that values from the first column are less than 3, and values from the right column are greater than 3, indicating that the "happy" column doesn't contribute significantly to a lack of fit, while the other factors do. Here's the row percentages too, we'll use those.

Observed Row Percentages:
[[24.74226804 28.86597938 46.39175258]
[22.33009709 59.22330097 18.44660194]]

Expected Row Percentages:
[[23.5 44.5 32.]
[23.5 44.5 32.]]

The Pearson standard residuals apparently reflect the disparity between expected and observed row percentages.

Summary:

Across the three attraction factors we see that, assuming no influence of gender on preference, expected preferences are 23.5%, 44.5% and 32% across factors Happy, Reliable, and Physically Attractive respectively. In looking at the observed relationship between gender (male/female) and attraction factors, we find remarkable disparity in the preferences for reliability and physical attractiveness between genders. Namely males disproportionately favor physical attractiveness (46.4% observed vs 32% expected) over reliability (28.9% observed vs 44.5% expected), and females disproportionately favor reliability (59.2% observed vs 44.5% expected) over physical attractiveness (18.4% observed vs 32% expected). We find that between males and females observed preference for happiness is close to what was expected and shows a minimal relationship between gender and preference for happiness in a partner: 24.7% and 22.3% observed respectively to a 23.5% expected.
We found that there is sufficient evidence to reject the null hypothesis that there is no relationship between gender and attraction factor preference. As such we claim that there is evidence to suggest that the association between gender and attraction factor preference is statistically significant, with $\alpha = 0.05$, $\chi^2(2, 200) = 22.66$; $p < 0.05$.

GENERAL INSTRUCTIONS - This is due by 11:59 PM on Saturday Nov. 2, 2024. When you are finished making your final edits, simply walk away from this document. If you are completing both data sets, this will satisfy your Midterm Exam 2 requirements (...you can ignore any late/missing notifications from Canvas)

Spearman's rho Analysis - Version A.1

Do opposites attract? Does similarity or dissimilarity foster attraction? A social psychologist investigating this question asked 15 college students to fill out a questionnaire concerning their attitudes toward a variety of topics (e.g., politics; music). Some time later, they were shown the attitudes of a stranger to the same items and were asked to rate the stranger as to the probable liking for the stranger and probable enjoyment of working with him. If similarities attract, then there should be a direct relationship between the attraction of the stranger, and the proportion of similar attitudes. Variable X summarizes how similar each college student and each stranger were in their responses on a variety of topics (higher ratings imply that the college student and stranger were more similar). Variable Y summarizes how each college student rated their attraction (e.g., their desire to get to know) for the stranger. Using a Spearman's rho analysis, with $\alpha = 0.05$ determine the value of the Spearman's rho correlation coefficient for the ordinal scale data for Variables X and Y. Is the relationship significant? If so, describe the relationship. Click here for the critical values for correlations.

College Student	Rank of Variable X	Rank of Variable Y
1	7	6
2	7	8
3	9	10
4	1	1
5	8	6
6	3	5
7	9	9
8	5	3
9	12	13
10	15	14
11	14	14
12	2	3
13	12	9
14	11	12
15	4	4

SHOW ALL WORK - Upload a clear snapshot of all your base calculations to this document. Include a typed summary analysis with APA notation on your upload.

GENERAL INSTRUCTIONS - This is due by 11:59 PM on Saturday Nov. 2, 2024. When you are finished making your final edits, simply walk away from this document. If you are completing both data sets, this will satisfy your Midterm Exam 2 requirements (...you can ignore any late/missing notifications from Canvas)

THIS SPACE FOR YOUR WORK

Calculation:

We begin by calculating the differences between variables Similarity (X) and Attraction (Y)

$$d_i = R(X_i) - R(Y_i)$$

and then squaring them

College Student	Rank of Variable X	Rank of Variable Y	$d_i = R(X_i) - R(Y_i)$	d_i^2
1	7	6	$7 - 6 = 1$	1
2	7	8	$7 - 8 = -1$	1
3	9	10	$9 - 10 = -1$	1
4	1	1	$1 - 1 = 0$	0
5	8	6	$8 - 6 = 2$	4
6	3	5	$3 - 5 =- 2$	4
7	9	9	$9 - 9 = 0$	0
8	5	3	$5 - 3 = 2$	4
9	12	13	$12 - 13 =- 1$	1
10	15	14	$15 - 14 = 1$	1
11	14	14	$14 - 14 = 0$	0
12	2	3	$2 - 3 = -1$	1
13	12	9	$12 - 9 = 3$	9
14	11	12	$11 - 12 = -1$	1
15	4	4	$4 - 4 = 0$	0

Note the sample size

$$N = 15$$

Spearman's Rho is calculated as

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N}.$$

As such we calculate $\sum_{i=1}^{N} d_i^2 = 1 + 1 + 1 + 4 + 4 + 4 + 1 + 1 + 1 + 9 + 1 = 28$

so $6\sum_{i=1}^{N} d_i^2 = 6 * 28 = 168.$

GENERAL INSTRUCTIONS - This is due by 11:59 PM on Saturday Nov. 2, 2024. When you are finished making your final edits, simply walk away from this document. If you are completing both data sets, this will satisfy your Midterm Exam 2 requirements (...you can ignore any late/missing notifications from Canvas)

Next $N^3 - N = 15^3 - 15 = 3360$.
Putting this together we find

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} = 1 - \frac{168}{3360} = 1 - 0.05 = 0.95$$

Now recall for Spearman's Rank Correlation, the degrees of freedom are found as follows

$$df = N - 2$$

which for us is simply

$$df = N - 2 = 15 - 2 = 13$$

For a significance level $\alpha = 0.05$ and degrees of freedom $df = 13$, the Spearman's Rank Correlation critical value is found from a table to be 0.514.
As our calculated

$$r_s = 0.95 > 0.514$$

we reject the null hypothesis and say that we have statistical significance.

Here's some code that checks my work.

Summary:

We find that similarity and attraction rankings appear to be in strong agreement with one another. We find a statistically significant Spearman correlation between similarity and attraction rankings $r_s(13) = 0.95,\ p < 0.05$.