HW 4

18| Table 2.13 shows data from the 2002 General Social Survey cross classifying a person's perceived happiness with their family income. The table displays the observed and expected cell counts and the standardized residuals for testing independence.

a) Show how to obtain the expected cell count of 35.8 for the first cell.

This is done using the formula $\mu_{11} = (n_{1\cdot} \cdot n_{\cdot 1})/n$

$\mu_{11} = [(21+159+110)\cdot(21+53+94)]/(21+159+110+53+372+221+94+249+83$

$= 35.77093 \approx 35.8$

b) For testing independence, $X^2 = 73.4$. Report the df value and the P-value and interpret.

$df = (I-1)(J-1) = ((3)-1)((3)-1) \boxed{= 4}$

P-value: $\boxed{<.001}$

A P-value of <.001 suggests that it is very unlikely that family income and reported happiness are independent.

c) Interpret the standardized residuals in the corner cells having counts 21 and 83.

These residuals are both $|\text{residual}| > 2$, suggesting that these results are likely too high for us to consider family income and reported happiness are not truly independent.

d) Interpret the standardized residuals in the corner cells having counts 110 and 94.

These residuals are well above $|\text{residual}| > 2$, again suggesting that family income and reported happiness are truly independent.

19| Table 2.14 was taken from the 2002 General Social Survey.

a) Test the null hypothesis of independence between party identification and race. Interpret.

$n_{1\cdot} = 871+444+873 = 2188$ $\quad n_{2\cdot} = 302+80+43 = 425$ $\quad n_{\cdot 1} = 871+302 = 1173$ $\quad n_{\cdot 2} = 444+80 = 524$

$n_{\cdot 3} = 873+43 = 916$ $\quad n = 2188+425 = 2613$ $\quad \mu_{11} = \frac{(2188)(1173)}{(2613)} = 982.21$

$\mu_{12} = \frac{(2188)(524)}{(2613)} = 438.77$ $\quad \mu_{13} = \frac{(2188)(916)}{(2613)} = 767.01$ $\quad \mu_{21} = \frac{(425)(1173)}{(2613)} = 190.79$

$\mu_{22} = \frac{(425)(524)}{(2613)} = 85.23$ $\quad \mu_{23} = \frac{(425)(916)}{(2613)} = 148.99$

$$X^2 = \frac{(871-982.21)^2}{982.21} + \frac{(444-438.77)^2}{438.77} + \frac{(873-767.01)^2}{767.01} + \frac{(302-190.79)^2}{190.79} + \frac{(80-85.23)^2}{85.23} + \frac{(43-148.99)^2}{148.99}$$

continued on next page →

$\chi^2 = 12.59 + 0.06 + 14.65 + 64.82 + 0.32 + 75.40 = 167.84$

$df = ((3)-1)((2)-1) = 2$

P-value $\boxed{= < .001}$

Based on a χ^2 test for independence, we can reject the null hypothesis that race and party affiliation are independent, suggesting that there is some correlation between race and party affiliation.

21| Each subject in a sample of 100 men and 100 women is asked to indicate which of the following factors (one or more) are responsible for increases in teenage crime:

	A An increasing income gap	B An increase in single parent families	C Insufficient parent-child time
Male	60	81	75
Female	75	87	86

a) Is it valid to apply the χ^2 test of independence to this 2x3 table?

No, because responders were allowed to select more than one response we know that some of the cell responses are dependent on others and therefore not independent.

b) Explain how this table actually provides information needed to cross classify gender with each of the three variables. Construct the contingency table relating gender to opinion about whether factor A is responsible for increases in teenage crime.

By splitting the table into partial tables, we can evaluate how each gender feels on each factor one at a time.

Response to Factor A

	Yes	No
Male	60	40
Female	75	25

22| Table 2.15 classifies psychiatric patients by their diagnosis and whether their treatment prescribed drugs.

a) Construct a test of independence and interpret the P-Value.

Table 2.15	Drugs	No Drugs	
Schizophrenia	105	8	113
Affective Disorder	12	2	14
Neurosis	18	19	37
Personality Disorder	47	52	99
Special Symptoms	0	13	13
	182	94	276

$df = ((2)-1)((5)-1) = 4$

* The sample size is small, and with $\mu_{22} = 4.77$ and $\mu_{25} = 4.43$ we risk the use of a χ^2 test for independence being invalid. However, I will use the χ^2 test anyway because these μs are close to 5 and the rest are safely above 5, and I am unsure of any better methot to use (not a 2x2, no Fisher Exact test).

$\mu_{11} = \frac{(182)(113)}{276} = 74.51$ $\mu_{21} = \frac{(182)(14)}{276} = 9.23$ $\mu_{31} = \frac{(182)(37)}{276} = 24.40$ $\mu_{41} = \frac{(182)(99)}{276} = 65.28$ $\mu_{51} = \frac{(182)(13)}{276} = 8.57$

$\mu_{21} = \frac{(94)(113)}{276} = 38.49$ $\mu_{22} = \frac{(94)(14)}{276} = 4.77$ $\mu_{23} = \frac{(94)(37)}{276} = 12.60$ $\mu_{24} = \frac{(94)(99)}{276} = 33.72$ $\mu_{25} = \frac{(94)(13)}{276} = 4.43$

$$\chi^2 = \frac{(105-74.51)^2}{74.51} + \frac{(12-9.23)^2}{9.23} + \frac{(18-24.40)^2}{24.40} + \frac{(47-65.28)^2}{65.28} + \frac{(0-8.57)^2}{8.57} + \frac{(8-38.49)^2}{38.49} + \frac{(2-4.77)^2}{4.77}$$

$$+ \frac{(19-12.60)^2}{12.60} + \frac{(52-33.72)^2}{33.72} + \frac{(13-4.43)^2}{4.43}$$

$$= 12.48 + .83 + 1.68 + 5.12 + 8.57 + 24.15 + 1.6 + 3.25 + 9.91 + 16.58 = 84.17$$

$\boxed{\text{P-value: } < .001}$

Based on the χ^2 test, so the null hypothesis that disorder and use of drugs in treatment are dependent can be rejected. This would suggest that there is not a firm treatment protocol for psychiatric disorders according to this data.

$$\frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\mu_{ij}(1-p_{i\cdot})(1-p_{\cdot j})}}$$

b) Obtain standardized residuals and interpret.

$p_{1\cdot} = \frac{n_{1\cdot}}{n} = \frac{(113)}{276} = .409$ $p_{2\cdot} = \frac{(14)}{276} = .051$ $p_{3\cdot} = \frac{(37)}{276} = .134$ $p_{4\cdot} = \frac{(99)}{276} = .359$

$p_{5\cdot} = \frac{(13)}{276} = .047$ $p_{\cdot 1} = \frac{(182)}{276} = .660$ $p_{\cdot 2} = \frac{(94)}{276} = .341$

$$r_{11} = \frac{(105)-(74.51)}{\sqrt{(74.51)(1-(.409))(1-(.660))}} = 7.88 \qquad r_{12} = \frac{(8)-(38.49)}{\sqrt{(38.49)(1-(.409))(1-(.341))}} = -7.87$$

$$r_{21} = \frac{(12)-(9.23)}{\sqrt{(9.23)(1-(.051))(1-(.660))}} = 0.94 \qquad r_{22} = \frac{(2)-(4.77)}{\sqrt{(4.77)(1-(.051))(1-(.341))}} = -1.60$$

$$r_{31} = \frac{(18)-(24.40)}{\sqrt{(24.40)(1-(.134))(1-(.660))}} = -1.39 \qquad r_{32} = \frac{(19)-(12.60)}{\sqrt{(12.60)(1-(.134))(1-(.341))}} = 2.39$$

$$r_{41} = \frac{(47)-(65.28)}{\sqrt{(65.28)(1-(.359))(1-(.660))}} = -2.83 \qquad r_{42} = \frac{(52)-(33.72)}{\sqrt{(33.72)(1-(.359))(1-(.341))}} = 4.84$$

$$r_{51} = \frac{(0)-(8.57)}{\sqrt{(8.57)(1-(.047))(1-(.660))}} = -4.47 \qquad r_{52} = \frac{(13)-(4.43)}{\sqrt{(4.43)(1-(.047))(1-(.341))}} = 5.14$$

Most of the residuals have an absolute value greater than 2, suggesting the data is not independent. In fact, for Schizophrenia and Affective disorder there appears to be a trend of prescribing drugs more often than expected, and less often for the other mental disorders.

25| For tests of H_0: independence, $\{\hat{\mu}_{ij} = (n_{i\cdot} \cdot n_{\cdot j})/n\}$

a) Show that $\{\mu_{ij}\}$ have the same row and column totals as $\{n_{ij}\}$

Given that $\sum n_{i\cdot} = n$ and $\sum n_{\cdot j} = n$

For any given "i"

$\mu_{i\cdot} = \sum \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ $\rightarrow$ because we are keeping "i" constant $\rightarrow \mu_{i\cdot} = n_{i\cdot} \cdot \sum \frac{n_{\cdot j}}{n} = n_{i\cdot} \cdot \frac{(n)}{n} = n_{i\cdot} \cdot 1 = n_{i\cdot}$

Thus, $\{\mu_{i\cdot}\} = \{n_{i\cdot}\}$. This finding will be true for all rows, and the same logic will verify that $\{\mu_{j\cdot}\} = \{n_{\cdot j}\}$

b) For 2x2 tables, show that $\hat{\mu}_{11} \cdot \hat{\mu}_{22} / \hat{\mu}_{12} \cdot \hat{\mu}_{21} = 1.0$. Hence $\{\hat{\mu}_{ij}\}$ satisfy H_0

If: $\hat{\theta} = \frac{\hat{\mu}_{11} \cdot \hat{\mu}_{22}}{\hat{\mu}_{12} \cdot \hat{\mu}_{21}} = \hat{\mu}_{11} \cdot \hat{\mu}_{22} \cdot \hat{\mu}_{12}^{-1} \cdot \hat{\mu}_{21}^{-1}$

Then:

$$\hat{\theta} = \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n} \cdot \frac{n_{2\cdot} \cdot n_{\cdot 2}}{n} \cdot \frac{n}{n_{1\cdot} \cdot n_{\cdot 2}} \cdot \frac{n}{n_{2\cdot} \cdot n_{\cdot 1}} = \frac{n_{1\cdot}\, n_{\cdot 1}\, n_{2\cdot}\, n_{\cdot 2}}{n_{1\cdot}\, n_{\cdot 2}\, n_{2\cdot}\, n_{\cdot 1}} = 1$$

26) A χ^2 variate with degrees of freedom equal to df has representation: $Z_1^2 + \ldots + Z_{df}^2$, where $Z_1, \ldots, Z_{df}$ are independent standard normal variates.

a) If Z has a standard normal distribution, what distribution does Z^2 have?

$\boxed{\text{$Z^2$ has a χ^2 distribution with whatever number of degrees of freedom df was assigned to Z.}}$

b) Show that, if Y_1 and Y_2 are independent χ^2 variates with degrees of freedom df_1 and df_2, then $Y_1 + Y_2$ has a χ^2 distribution with $df = df_1 + df_2$.

Fisher's Exact test: $P(n_{11}) = \dfrac{\binom{n_{1\cdot}}{n_{11}}\binom{n_{2\cdot}}{n_{\cdot 1}-n_{11}}}{\binom{n}{n_{\cdot 1}}}$

30) The table contains the results of a study comparing radiation therapy and surgery for laryngeal cancer treatment. Use Fisher's exact test to test $H_0: \theta = 1$ against $H_a: \theta > 1$. Interpret the results.

	Cancer Controlled: Yes	No	
Surgery	21	2	23
Radiation Therapy	15	3	18
	36	5	41

$$P(21) = \frac{\binom{23}{21}\binom{18}{36-21}}{\binom{41}{36}} + \frac{\binom{23}{22}\binom{18}{36-22}}{\binom{41}{36}} + \frac{\binom{23}{23}\binom{18}{36-23}}{\binom{41}{36}}$$

$$= \frac{253 \cdot 816}{749398} + \frac{23 \cdot 3060}{749398} + \frac{1 \cdot 8568}{749398}$$

$$= .275 + .094 + .011 = 0.381$$

With a P-value = 0.381, it is not convincing that the outcome of cancer control is actually dependent on the treatment chosen. It is not possible to say, with reasonable confidence, that the type of treatment chosen actually effects whether or not laryngeal cancer is controlled.