


I pledge my honor that I have abided by the Stevens Honor System.

Team 2 - Kush Parmar, Jo-Anne Rivera, Simerjeet Mudhar

9.25 Sexual harassment in middle and high schools

A nationally representative survey of students in grades 7 to 12 asked about the experience of these students with respect to sexual harassment.¹¹ One question asked how many times the student had witnessed sexual harassment in school. Here are the data categorized by gender:  HARAS1

Gender	Times witnessed		
	Never	Once	More than once
Girls	140	192	671
Boys	106	125	732

Find the expected counts for this 2×3 table.

Expected counts = (row total * column total) / n

Girls	140	192	671	1003
Boys	106	125	732	963
	246	317	1403	1966

Girls, Never Cell Expected Count: $1003 * (246 / 1966) = 125.5$

Boys, Never Cell Expected Count: $963 * (246 / 1966) = 120.5$


Girls, Once Cell Expected Count: $1003 * (317 / 1966) = 161.72$

Boys, Once Cell Expected Count: $963 * (317 / 1966) = 155.28$

Girls, More than Once Cell Expected Count: $1003 * (1403 / 1966) = 715.77$

Boys, More than Once Cell Expected Count: $963 * (1403 / 1966) = 687.23$

9.26 Do the significance test

Refer to the previous exercise. Compute the chi-square statistic and the P -value. Write a short summary of your conclusions from the analysis of these data.  HARASI

H_0 : Harassment/gender are not independent

H_a : Harassment/gender are independent

Observed Count(O)	Expected Count (E)	$\frac{(O-E)^2}{E}$
140	125.5	1.68
192	161.72	5.67
671	715.77	2.8
106	120.5	1.75
125	155.28	5.90
732	687.23	2.91

$$\chi^2 = 20.71$$


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

> 1-pchisq(20.71,2)
[1] 3.183323e-05

When we look at the Chi squared table we can see that the critical value is 5.991 when the significance level is 0.05 and the degrees of freedom are 2. When we test the chi-squared value to the critical value we can see that it is not greater than the

significance level. Hence, we reject the null hypothesis.

9.27 Sexual harassment online or in person

In the study described in Exercise 9.25, the students were also asked whether or not they were harassed in person and whether or not they were harassed online. Here are the data for the girls: 
HARASG

Harassed in person	Harassed online	
	Yes	No
Yes	321	200
No	40	441

- (a) Analyze these data using the method presented in Chapter 8 for comparing two proportions (page 508).
- (b) Analyze these data using the method presented in this chapter for examining a relationship between two categorical variables in a 2×2 table.
- (c) Use this example to explain the relationship between the chi-square test and the z test for comparing two proportions.
- (d) The number of girls reported in this exercise is not the same as the number reported for Exercise 9.25. Suggest a possible reason for this difference.

a. P1: girls who get harassed online

P1.1: girls who get harassed online and in person

P1.2: girls who get harassed online but not in person

P2: girls who don't get harassed online

P2.1 girls who don't get harassed online but in person

P2.2 girls who don't get harassed online and not in person

$$H_0: \hat{p}_1 = \hat{p}_2, \quad H_{alt}: \hat{p}_1 \neq \hat{p}_2$$

$$\hat{p}_1 = \frac{321}{361} = 0.889, \hat{p}_2 = \frac{200}{641} = 0.312 \rightarrow \hat{p} = \frac{321+200}{361+641} = \frac{521}{1002} = 0.52$$

$$SEp = \sqrt{(0.52)(1 - 0.52)\left(\frac{1}{361} + \frac{1}{641}\right)} = 0.033$$

$$D = 0.8882 - 0.3120 = 0.5772$$

$$z = \frac{0.5772}{0.033} = 17.5$$

P value of 17.5 from the z test is about 0.0002 which is in the the range of error, hence we reject the null hypothesis.

b. $321 + 200 = 521 \rightarrow$ yes to in person harassment

$40 + 441 = 481 \rightarrow$ no to in person harassment

$321 + 40 = 361 \rightarrow$ yes to online harassment

$200 + 441 = 641 \rightarrow$ no to online harassment

$$\frac{521 \times 361}{1002} = 187.71$$

$$\frac{521 \times 641}{1002} = 333.29$$

$$\frac{481 \times 361}{1002} = 173.29$$

$$\frac{481 \times 641}{1002} = 307.71$$

$$\frac{(321-187.71)^2}{187.71} = 94.64$$

$$\frac{(200-333.29)^2}{333.29} = 53.30$$

$$\frac{(40-173.29)^2}{173.29} = 102.52$$

$$\frac{(441-307.71)^2}{307.71} = 57.33$$


$$X^2 = 94.6 + 53.3 + 102.52 + 57.3 = 308.2$$

$$(2-1)(2-1) = 1 \text{ degree of freedom}$$

Reject null hypothesis.

- c. When we look at the z test, it is trying to explain that the proportion of girls who are harassed online differs from those who have not. This suggests that there is, in fact, a link between being harassed online and in person. When we look at the chi-squared test, similarly to the z-test, we also reach this conclusion; therefore, there is a relationship between these factors.
- d. There could have just been a different subject pool, or because the questions themselves are phrased a bit differently, subjects responded differently.

9.28 Data for the boys

Refer to the previous exercise. Here are the corresponding data for boys:  HARASB

Harassed in person	Harassed online	
	Yes	No
Yes	183	154
No	48	578

Using these data, repeat the analyses that you performed for the girls in Exercise 9.27. How do the results for the boys differ from those that you found for girls?

- a. p_1 - boys harassed online, p_2 - boys harassed in person

$\hat{p}_1 = \frac{183}{183+48} = \frac{183}{231} = 0.792$; Population of boys who were harassed in person within the population of boys who were harassed online.

$\hat{p}_2 = \frac{154}{154+578} = \frac{154}{732} = 0.210$; Population of boys who were harassed in person within the population of boys who were not harassed online.

$$H_0: p_1 = p_2, \quad H_{alt}: p_1 \neq p_2$$

$$\hat{p} = \frac{183+154}{231+732} = \frac{337}{963} = 0.350$$

$$SE_p = \sqrt{0.350(1 - 0.350)\left(\frac{1}{231} + \frac{1}{732}\right)} = 0.036$$

$$z = \frac{0.792-0.210}{0.036} = 16.167$$

$$P(Z < 16.16) > 0.9998, P(z > 17.54) < 0.0002$$

P value is less than 0.0004, hence we must reject the null hypothesis.

Row Totals:

$183 + 154 = 337 \rightarrow$ Yes to in person harassment

$48 + 578 = 626 \rightarrow$ No to in person harassment

Column Totals:

$183 + 48 = 231 \rightarrow$ Yes to online harassment

$154 + 578 = 732 \rightarrow$ No to online harassment

Grand Total: $231 + 732 = 963$

Expected count:

$$\frac{337(231)}{963} = 80.84 \rightarrow \text{for row Yes, column Yes}$$

$$\frac{337(732)}{963} = 256.16 \rightarrow \text{for row Yes, column No}$$

$$\frac{626(231)}{963} = 150.16 \rightarrow \text{for row No, column Yes}$$

$$\frac{626(732)}{963} = 475.84 \rightarrow \text{for row No, column No}$$

$$X^2 = \sum \frac{(O.C. - E.C.)^2}{E.C.}$$

$$\frac{(183 - 80.84)^2}{80.84} = 129.103$$

$$\frac{(154 - 256.16)^2}{256.16} = 40.473$$

$$\frac{(48 - 150.16)^2}{150.16} = 69.504$$

$$\frac{(578 - 475.84)^2}{475.84} = 21.933$$

$$X^2 = 129.103 + 40.473 + 69.504 + 21.933 = 261.013$$

$$df = (row - 1)(col - 1) = (2 - 1)(2 - 1) = 1$$

Since X^2 has a P-value less than the minimum probability, we must reject the null hypothesis. Accept the alternative hypothesis.

P-value < 0.0005

When comparing the results for the boys versus the girls, we can see that there is a greater probability that these variables can be independent for the boys' result rather than for the girls' result. This can be explained through how the z statistic for boys (16.16) is a lot lower than the girls' (17.54). An opposite relationship can be seen with the P-value, where the range for the boy's P-value seems to be greater than the range of the girl's P-value. The chi-square statistic of the boys' is smaller than the girls' ($261.013 < 308.2$). Thus, we can see through the chi-square test, that while the chi-square statistic gets larger, the P-value decreases.

