Lecture 6 practice

Stats 7 Summer Session II 2022

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(a) The chi-square distribution, just like the normal distribution, has two parameters, mean and standard deviation.

False. The chi-square distribution has one parameter called degrees of freedom

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(b) The chi-square distribution is always right skewed, regardless of the value of the degrees of freedom parameter.

True

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(c) The chi-square statistic is always positive

True

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(d) As the degrees of freedom increases, the shape of the chi-square distribution becomes more skewed.

False. As the degrees of freedom increases, the shape of the chi-square distribution becomes more symmetric.

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(e) As the degrees of freedom increases, the mean of the chi-square distribution increases.

True

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(f) When finding the p-value of a chi-square test, we always shade the tail areas in both tails.

False, the test statistic is always positive, and a higher test statistic means a stronger deviation from the null hypothesis.

Checking our knowledge

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

(g) As the degrees of freedom increases, the variability of the chi-square distribution decreases.

False, as the degrees of freedom increases, the variability of the chi-square distribution increases.

Open source textbook

A professor using an open source introductory statistics book predicts that 60% of the students will purchase a hard copy of the book, 25% will print it out from the web, and 15% will read it online. At the end of the semester he asks his students to complete a survey where they indicate what format of the book they used. Of the 126 students, 71 said they bought a hard copy of the book, 30 said they printed it out from the web, and 25 said they read it online.

(a) State the hypotheses for testing if the professor's predictions were inaccurate

H_0: The distribution of the format of the book used by the students follows the professor's predictions.

H_A: The distribution of the format of the book used by the students does not follow the professor's predictions.

Open source textbook

(b) How many students did the professor expect to buy the book, print the book, and read the book exclusively online?

Total = 126	Observed	Expected %	Expected count
Hard copy	71	60%	126 × 0.60 = 75.6
Web	30	25%	126 × 0.25 = 31.5
Online	25	15%	126 × 0.15 = 18.9

Open source textbook

(c) This is an appropriate setting for a chi-square test. List the conditions required for a test and verify they are satisfied.

Total = 126	Observed	Expected %	Expected count
Hard copy	71	60%	75.6
Web	30	25%	31.5
Online	25	15%	18.9

Assuming students was not influenced by others, we have independence.

Each cell has at least 5 expected observations.

Open source textbook

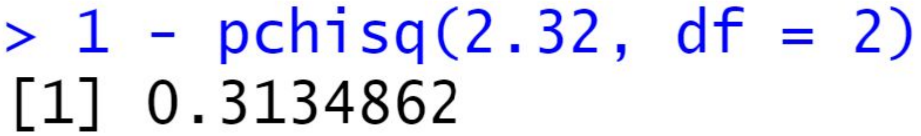
(d) Calculate the chi-squared statistic, the degrees of freedom associated with it, and the p-value.

Total = 126	Observed	Expected count	$(O - E)^2 / E$
Hard copy	71	75.6	$(71 - 75.6)^2 / 75.6 = 0.2799$
Web	30	31.5	$(30 - 31.5)^2 / 31.5 = 0.0714$
Online	25	18.9	$(25 - 18.9)^2 / 18.9 = 1.9688$

$\chi^2 = 0.2799 + 0.0714 + 1.9688 = 2.32$

df = 3 - 1 = 2

P-value = area above 2.32



```
> 1 - pchisq(2.32, df = 2)
[1] 0.3134862
```

Open source textbook

(e) Based on the p-value calculated in part (d), what is the conclusion of the hypothesis test? Interpret your conclusion in this context.

Our p-value was 0.3135, so there was a 31.35% probability of obtaining data as or more different from the expected percentages, if the data follow the expected distribution of textbook format.

P-value is large so we fail to reject H_0, we did not find evidence to support H_A.

We did not find evidence (p-value = 0.3135) that the student's preferred textbook format differs from the teacher's expectations.

Quitters

Does being part of a support group affect the ability of people to quit smoking? A county health department enrolled 300 smokers in a randomized experiment. 150 participants were assigned to a group that used a nicotine patch and met weekly with a support group; the other 150 received the patch and did not meet with a support group. At the end of the study, 40 of the participants in the patch plus support group had quit smoking while only 30 smokers had quit in the other group.

(a) Create a two-way table presenting the results of this study.

Quitters

	Quit	Didn’t quit	Total
Patch + support	40	110	150
Patch only	30	120	150
Total	70	230	300

(b) Answer each of the following questions under the null hypothesis that being part of a support group does not affect the ability of people to quit smoking, and indicate whether the expected values are higher or lower than the observed values

Quitters

	Quit	Didn’t quit	Total
Patch + support	40	110	150
Patch only	30	120	150
Total	70	230	300

i. How many subjects in the “patch + support” group would you expect to quit?

(Row total) x (Column total) / (Table total) = 150 x 70 / 300 = 35.

This is lower than the observed value of 40.

Quitters

	Quit	Didn’t quit	Total
Patch + support	40	110	150
Patch only	30	120	150
Total	70	230	300

ii. How many subjects in the “patch only” group would you expect to not quit?

(Row total) x (Column total) / (Table total) = 150 x 230 / 300 = 115

This is lower than the observed value of 120.

Offshore drilling

The table below summarizes a data set we first encountered in a previous exercise that examines the responses of a random sample of college graduates and non-graduates on the topic of oil drilling. Complete a chi-square test for these data to check whether there is a statistically significant difference in responses from college graduates and non-graduates.

	College grad	Not a College grad	Total
Support	154	132	286
Oppose	180	126	306
Do not know	104	131	235
Total	438	389	827

Offshore drilling

	College grad	Not a College grad	Total
Support	154	132	286
Oppose	180	126	306
Do not know	104	131	235
Total	438	389	827

First we need to identify the hypotheses:

H_0: The opinion of college grads and non-grads is not different on the topic of drilling for oil and natural gas off the coast of California.

H_A: Opinions regarding the drilling for oil and natural gas off the coast of California has an association with earning a college degree.

Offshore drilling

Now we need to calculate the expected counts:

	College grad	Not a College grad	Total
Support	154 286 x 438 / 827 = 151.5	132 286 x 389 / 827 = 134.5	286
Oppose	180 306 x 438 / 827 = 162.1	126 306 x 389 / 827 = 143.9	306
Do not know	104 235 x 438 / 827 = 124.5	131 235 x 389 / 827 = 110.5	235
Total	438	389	827

Offshore drilling

	College grad	Not a College grad	Total
Support	154 151.5	132 134.5	286
Oppose	180 162.1	126 143.9	306
Do not know	104 124.5	131 110.5	235
Total	438	389	827

Check the conditions:

- Independence: The samples are both random and unrelated, so independence between observations is reasonable.
- Sample size: All expected counts are at least 5.

Offshore drilling

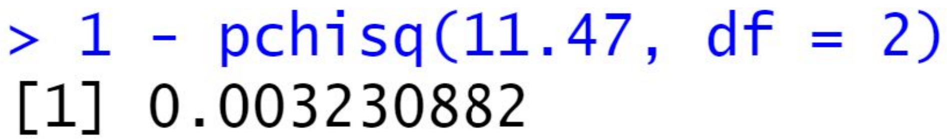
	College grad	Not a College grad	Total
Support	154 151.5	132 134.5	286
Oppose	180 162.1	126 143.9	306
Do not know	104 124.5	131 110.5	235
Total	438	389	827

Calculate the test statistic:

$$\chi^2 = (154 - 151.5)^2 / 151.5 + (132 - 134.5)^2 / 134.5 + \ldots + (131 - 110.5)^2 / 110.5 = 11.47$$

df = (# rows - 1) x (# of columns - 1) = (3 - 1) x (2 - 1) = 2

p-value = area above 11.47 with 2 df



```
> 1 - pchisq(11.47, df = 2)
[1] 0.003230882
```

Offshore drilling

P-value = 0.0032 < α = 0.05 so reject null, evidence to support alternative

There is strong evidence (P-value = 0.0032 < α = 0.05) that there is an association between support for offshore drilling and having a college degree.

Active learning

A teacher wanting to increase the active learning component of her course is concerned about student reactions to changes she is planning to make. She conducts a survey in her class, asking students whether they believe more active learning in the classroom (hands on exercises) instead of traditional lecture will helps improve their learning. She does this at the beginning and end of the semester and wants to evaluate whether students' opinions have changed over the semester. Can she used the methods we learned in this chapter for this analysis? Explain your reasoning.

No. The samples at the beginning and at the end of the semester are not independent since the survey is conducted on the same students.

So far...

We have learned inference for categorical variables

	1 Proportion	2 Proportions	Chi square goodness of fit	Chi square independence
Data scenario	1 sample of a categorical variable with 2 levels	2 samples of a categorical variable with 2 levels	1 sample of a categorical variable with 3+ levels	1 sample of 2 categorical variables with 2+ levels each
Distribution used	Normal	Normal	Chi square	Chi square
Inference	Hypothesis test and Confidence interval	Hypothesis test and Confidence interval	Hypothesis test	Hypothesis test

Now we will learn inference for numerical variables, starting with a mean for a single sample.

Credits

Examples adapted from OpenIntro Statistics (4th edition) by David Diez, Mine Cetinkaya-Rundel, and Christopher D Barr https://www.openintro.org/book/os/ protected under the Creative Commons License