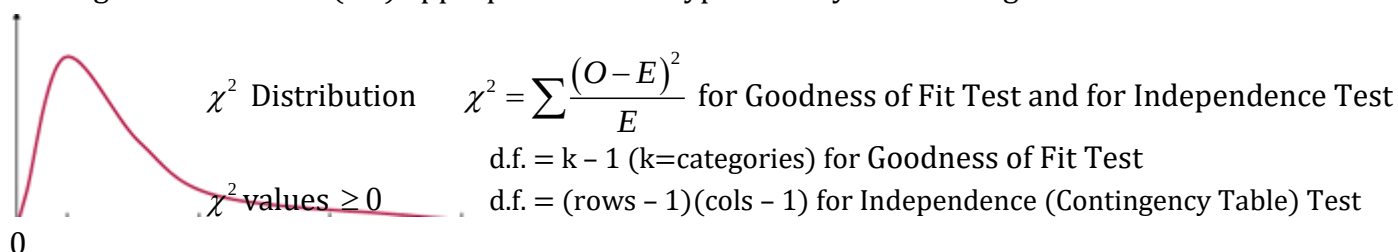


MAT1080 Spring 2019 Worksheet #9 Examples
(Actual problems in separate document)

χ^2 (Chi-Square) Distribution

A chi-square distribution is a type of distribution similar to a z (normal) distribution or a t distribution. Unlike those other two distributions it does not have symmetrical left and right tails, in fact you cannot have a value less than zero. It is used in this course for either a Goodness of Fit test, or for an Independence (Contingency Table) Test. Similar to using a t table, to use a χ^2 table you need to know the degrees of freedom (d.f.) appropriate for the type of test you are doing.



All tests are right-tailed.

Chi-Square Goodness of Fit Example #1:

A tax preparation company wants to determine the proportions of Normandale statistics students who used different methods to prepare their taxes. It wants to test a previous survey's claim concerning the distribution of proportions of Normandale statistics students who used the methods to prepare their taxes and claim the previous survey is inaccurate. Use a P-Value test at the 1% significance level.

Previous survey of tax preparation methods.		New sample poll of tax preparation methods.	
Method	percentage of students	Method	# of students
Accountant	11%	Accountant	5
By Hand	8%	By Hand	3
Computer Software	45%	Computer Software	28
Friend/Family	27%	Friend/Family	23
Tax Preparation Software	9%	Tax Preparation Software	7

One requirement for this test is that the expected frequency of each class is at least five. The first step is to derive the expected frequencies from the expected percentages. An expected percentage could be considered relative frequency, but we want absolute frequency (a count). The expected frequency is the sample size times the expected probability: $E_i = np_i$

		Observed Frequency	Expected frequency		
Tax Preparation Method	% of People	O	E	$O - E$	$\frac{(O - E)^2}{E}$
Accountant	11%	5			
By Hand	8%	3			
Computer Software	45%	28			
Friend/Family	27%	23			
Tax Preparation Software	9%	7			

		n=66			$\chi^2 = \sum \frac{(O - E)^2}{E} =$

Example #1 Solution:

Remember for a Chi-Square Goodness of Fit Test:

- All expected frequencies must be ≥ 5 , akin to the number of expected successes and failures in some binomial tests having this criteria. Otherwise, you cannot complete the test.
- The test is ALWAYS right-tailed.
- The null hypothesis states that the frequency distribution fits the specified distribution.
- The alternative hypothesis states that the frequency distribution does not fit the specified distribution
- If all observed=expected (frequencies are as expected), then $\chi^2 = 0$.
- The greater the differences between observed and expected, the greater χ^2 is.
- The null hypothesis is usually stated in words, but it is the same as $\chi^2 = 0$.
- The alternative hypothesis is usually stated in words, but it is the same as $\chi^2 > 0$.

Since there are extra steps to calculate the test statistic, we will do that first.

		Observed Frequency	Expected frequency	$O - E$	$\frac{(O - E)^2}{E}$
Tax Preparation Method	p% of People	O	E=np		
Accountant	11%	5	7.26		
By Hand	8%	3	5.28		
Computer Software	45%	28	29.7		
Friend/Family	27%	23	17.82		
Tax Preparation Software	9%	7	5.94		
		-----			-----
		n=66			$\chi^2 = \sum \frac{(O - E)^2}{E} =$

Optional: As a check, the sum of the expected frequency column should be within rounding error of n=66.

The rest of the calculation for χ^2 and P can be done with technology.

For StatCrunch, read "STATCRUNCH Tutorial for Chi-square Goodness-of-fit test (Chapter 110(a).pdf" in the D2L content area. The quick summary is that you would fill in two columns for the observed and expected data, then go to Stat/Goodness-of-fit/Chi-Square Test, choose the observed and expected columns, and press "Compute!".

You obtain:

$$\chi^2 = 3.480 \quad P = 0.4809$$

StatCrunch would have automatically calculated the degrees of freedom, d.f. = k - 1 = 5 - 1 = 4 before finding P.

Example 1 Test of Significance

Step 1: The claim is that the actual distribution is not equal to the claimed distribution.

Step 2: The complement of the claim is that the actual distribution is equal to the claimed distribution

Step 3:

H_0 : The distribution is 11% by accountant, 8% by hand, 45% by computer software, 27% by friend/family, and 9% by tax preparation software.

H_1 : The distribution of tax preparation methods differs from the expected distribution (claim)

Step 4: $\alpha=0.01$

Step 5: χ^2 Goodness of Fit Test, right-tailed

Step 6: Test statistic $\chi^2 = \sum \frac{(O-E)^2}{E} = 3.480$

d.f. = $k - 1 = 5 - 1 = 4$ $P = 0.4809$

(we will relax the requirement for a drawing when using the P-Value method)

Step 7: $P > \alpha$, outside of rejection zone, fail to reject H_0

Step 8:

At the 1% level of significance, there is not sufficient evidence to support the claim that the distribution of tax preparation methods differs from the expected distribution

Example2 Chi-Square Independence Test (Contingency Table Test)

An Independence Test tests whether two variables are dependent or independent.

Is one's favorite way to eat ice cream dependent on gender? A survey was taken to determine this and the results shown below. At the 1% level of significance, test the claim the favorite way to eat ice cream is dependent on gender.

		Favorite way to eat ice cream					
Gender		Cup	Cone	Sundae	Sandwich	Other	Total
Male	O	600	288	204	24	84	
	E						
Female	O	410	340	180	20	50	
	E						
Total							

The table has left space for further calculation of the expected frequencies. We will quick show how to do this manually, but in general technology can calculate the expected values when it calculates χ^2 and P.

NOTE: The manual calculation of the expected frequencies is being shown here, but is not required on the worksheet.

The first step is to calculate the marginal totals.

		Favorite way to eat ice cream					
Gender		Cup	Cone	Sundae	Sandwich	Other	Total
Male	O	600	288	204	24	84	1200
	E						
Female	O	410	340	180	20	50	1000
	E						
Total		1010	628	384	44	134	2200

(Calculating the grand total both across and down is a good double-check the totals were done correctly)

The expected value in each cell is given by the formula $E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$

		Favorite way to eat ice cream					
Gender		Cup	Cone	Sundae	Sandwich	Other	Total
Male	O	600	288	204	24	84	1200
	E	550.91	342.55	209.45	24	73.09	
Female	O	410	340	180	20	50	1000
	E	459.09	285.45	174.55	20	60.91	
Total		1010	628	384	44	134	2200

So, for instance the first value of E=550.91 was obtained by

$$E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})} = \frac{(1200)(1010)}{(2200)} \sim 550.91$$

Then χ^2 could be calculated by using $\chi^2 = \sum \frac{(O-E)^2}{E}$, where the sum is being taken over all the cells in all rows and columns.

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E} = \frac{(600-550.91)^2}{550.91} + \frac{(288-342.55)^2}{342.55} + \frac{(204-209.45)^2}{209.45} + \frac{(24-24)^2}{24} + \frac{(84-73.09)^2}{73.09} \\ &+ \frac{(410-459.09)^2}{459.09} + \frac{(340-285.45)^2}{285.45} + \frac{(180-174.55)^2}{174.55} + \frac{(20-20)^2}{20} + \frac{(50-60.91)^2}{60.91} \approx 32.630\end{aligned}$$

The calculation for the expected frequencies, as well as χ^2 and P can be done with technology.

For StatCrunch, read "STATCRUNCH Tutorial for Chi-square test of independence (Chapter 110(b).pdf)" in the D2L content area. The quick summary is that you would first copy the data into StatCrunch:

var1	Cup	Cone	Sundae	Sandwich	Other
Male	600	288	204	24	84
Female	410	340	180	20	50

↑

put the male/female row labels as data

Enter the row labels under the first column (var1)

Change the column headers starting from column 2 to the column labels (for example, cup)

Enter the number for each cell of the contingency table

Having row and column labels, while optional, will aid in analyzing the results.

Go to Stat/Tables/Contingency/With Summary (With Data is actually something different).

Choose multiple **columns with data under Select columns** by **holding the Ctrl key** (for example, cup and cone columns).

Choose the **column with the row labels under Row labels** (for example, var1).

Choose **Expected count under Display** (in order to get the expected counts).

Choose **Chi-square test for independence under Hypothesis tests**.

Press **Compute!**

You should see the expected counts in parantheses after each observed count. They should be the same as what as shown above. If these values are asked for on the worksheet they can copied over.

$$\chi^2 = 32.63 \quad P = 0.00000142 \text{ (StatCrunch will say } P < .0001)$$

StatCrunch would have automatically calculated the degrees of freedom:

$$\text{d.f.} = (\text{rows} - 1)(\text{cols} - 1) = (1)(4) = 4 \text{ before finding } P.$$

Remember for a Chi-Square Independence Test (Contingency Table Test):

- The test is ALWAYS right-tailed.
- The null hypothesis states that the two variables are independent.
- The alternative hypothesis that the two variables are dependent.
- If all observed=expected (frequencies are as expected, one variable is not dependent on the other in some way, they are totally independent), then $\chi^2 = 0$.
- The greater the differences between observed and expected (the more the variables have a dependency), the greater χ^2 is.
- The null hypothesis is usually stated in words, but it is the same as $\chi^2 = 0$.
- The alternative hypothesis is usually stated in words, but it is the same as $\chi^2 > 0$.

Example 2 Test of Significance

Step 1: The claim is that the favorite way to eat ice cream is dependent of gender.

Step 2: The complement of the claim is that the favorite way to eat ice cream is independent of gender.

Step 3:

H_0 : The favorite way to eat ice cream is independent of gender

H_1 : The favorite way to eat ice cream is independent of gender (claim)

Step 4: $\alpha=0.01$

Step 5: χ^2 Chi-Square Independence Test, right-tailed

Step 6: Test statistic $\chi^2 = \sum \frac{(O-E)^2}{E} = 32.63$

d.f. = (rows - 1)(cols - 1) = (1)(4) = 4 P = 0.00000142 (StatCrunch will say P < .0001)
(we will relax the requirement for a drawing when using the P-Value method)

Step 7: $P \leq \alpha$, inside of rejection zone, reject H_0

Step 8:

At the 1% level of significance, there is sufficient evidence to support the claim that the favorite way to eat ice cream is dependent of gender.