

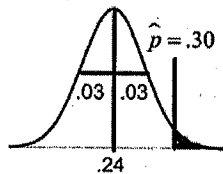
AP Statistics – Lesson Notes - Chapter 26: Comparing Counts

Comparing one Count against an expected percentage

If we know that M&M bags typically have 24% blue M&Ms, how unusual would it be to find a bag of 200 M&Ms with 60 Blue M&Ms?

We could answer this question using a 1-Proportion Z Test:

$$\begin{aligned} H_0: p = .24 \\ H_A: p > .24 \end{aligned} \quad \hat{p} = \frac{60}{200} = .30 \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.24)(.76)}{200}} = .03$$



$$z = \frac{.30 - .24}{.03} = 2$$

$$p\text{-value} = \text{normalcdf}(.30, 999, .24, .03)$$

$$p\text{-value} = .023$$

With low p-value, we reject H_0 . This bag's percentage is higher than 24%.
 $z = 2$ puts this bag in the top 2.5% of natural variation for percentage of blue.

Comparing a Row of Counts against an expected percentages

But what if we knew the expected percentages of all the colors, and we knew the counts of all the colors in a particular bag. How could we come up with a single measure of how well all the proportions, taken together, 'fit' the expected proportions?

M&M bag expected color distributions are:

13% Brown, 14% Yellow, 13% Red, 24% Blue, 20% Orange, 16% Green.

A particular bag of 200 M&Ms has the following counts:

29 Brown, 26 Yellow, 29 Red, 52 Blue, 36 Orange, 28 Green.

Is this bag's color distribution unusual?

The test which allows us to answer this question is called a

Chi-squared Goodness-Of-Fit (χ^2 GOF) test.

Chi-Squared Goodness-of-Fit Test

Conditions:

- **Counts:** This test can only be performed on data which represent *counts* for the categories of a categorical variable.
- **SRS:** The data have to be a random (or representative) sample from a population.
- **Each expected count is at least 5:** Some wiggle room: you can proceed if 80% of the expected counts are >5 and none are <1 .

Chi-Squared Goodness-of-Fit Test

Expected count = the expected number in each category if you multiply n by the expected percentages for each category.

M&M bag expected color distributions are:
 13% Brown, 14% Yellow, 13% Red, 24% Blue, 20% Orange, 16% Green.
 A particular bag of 200 M&Ms has the following counts:
 29 Brown, 26 Yellow, 29 Red, 52 Blue, 36 Orange, 28 Green.

	Observed	Expected	Calculation
Brown	29	26	(13%) * 200
Yellow	26	28	(14%) * 200
Red	29	26	(13%) * 200
Blue	52	48	(24%) * 200
Orange	36	40	(20%) * 200
Green	28	32	(16%) * 200
200			

Conditions:

Counts? *yes* ✓

SRS? *no, but representative* ✓

Expected Counts at least 5? *yes* ✓

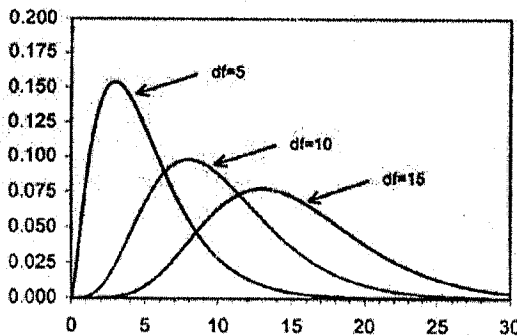
Compute the Chi-Squared statistic: $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$

*test statistic
 (no longer mean
 # std dev, though)*

	Observed	Observed - Expected	(Observed - Expected) ²	(Observed - Expected) ² / Expected
Brown	29/26	3	9	9/26 = .3462
Yellow	26/28	-2	4	4/28 = .1429
Red	29/26	3	9	9/26 = .3462
Blue	52/48	4	16	16/48 = .3333
Orange	36/40	-4	16	16/40 = .4000
Green	28/32	-4	16	16/32 = .5000

$\chi^2 = 2.0686$

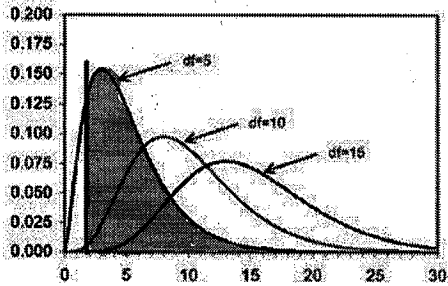
To find p-value, we have to use the appropriate distribution for χ^2 values. Different from Normal and t-distribution, and is called a χ^2 - distribution



Like t-distributions, χ^2 - distribution depends upon degrees-of-freedom. For goodness of fit:

$df = (\# \text{categories}) - 1$

Here, $df = 6 - 1 = 5$



...and our $\chi^2 = 2.0686$

$$p\text{-value} = \chi^2 cdf(2.0686, 999, 5)$$

LB UB df

$$p\text{-value} = .840$$

With this very high p-value, we would fail to reject the null hypothesis.
(But what are the appropriate hypotheses?)

Hypotheses for Chi-square Goodness of Fit test:

The Chi-Square statistic is **always positive** (because any negative difference is squared) and the statistic is measuring the overall difference from the 'model' which would be an exact fit to the expected percentages.

In this sense, a Chi-Square test is **always one-sided** because the best fit would be no difference between observed and expected for all categories and χ^2 would equal zero.

That means the Null Hypothesis is that all category proportions match expected values. The Alternative Hypothesis would be that at least one proportion does not match the expected (population) proportion:

$$H_0 : p_{brown} = .13, p_{yellow} = .14, p_{red} = .13, p_{blue} = .24, p_{orange} = .20, p_{green} = .16$$

$$H_A : \text{At least one } p \text{ is different than expected}$$

So with $p\text{-value} = .843$, we fail to reject that the proportions are as expected. The evidence does not support that the proportions of this bag of M&Ms are different than expected.

Chi-Squared Distribution Shape / Sketching

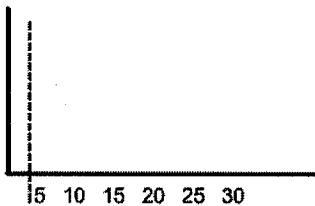
How can we sketch the χ^2 distribution when it is skewed?

The **mode** (peak) is at $df - 2$.

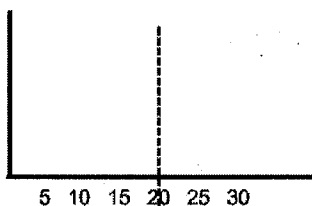
The **mean** is at df .

The lower df is, the higher the peak is.

$$\chi^2(df = 5)$$



$$\chi^2(df = 20)$$



Chi-Squared Goodness-of-Fit Test on a TI-84

- 1) Observed counts into L1.
- 2) Expected counts into L2.

by hand

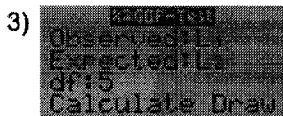
$$3) \text{ At top of L3: } L3 = \frac{(L1 - L2)^2}{L2}$$

$$4) \chi^2 = 2\text{nd STAT (LIST), MATH, sum(L3)}$$

$$5) p\text{-value} = \chi^2 cdf(\chi^2, 999, df)$$

$$\text{Remember: } df = (\# \text{ categories}) - 1$$

using test function



CNTRB = the contibuted values for each category.
$$\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

Chi-Squared Test for Homogeneity

The Chi-square Goodness-of-Fit test is like comparing one sample proportion to a known 'expected' proportion (1-PropZTest), except for a number of categories, at the same time. When we did proportion tests, we also sometimes compared two proportions to each other, rather than to a 'known' proportion: a 2-PropZTest. Can we do comparisons of multiple distributions, but for a number of categories?

Yes, and we can actually compare more than two distributions at a time (any number of distributions) and compare any number of categories. This is called a **Chi-square Test for Homogeneity**.

It looks similar to the Chi-square GOF, but instead of being given the 'expected' proportions for the categories, we determine them from the data itself by using the proportions in the right-hand 'totals' column.

The following data show the number of undergraduate engineering degrees awarded at ASU for the previous three academic years. Has the distribution of different degree fields changed significantly over these years?

	2012-2013	2013-2014	2014-2015
Electrical/Computer/SW (ECS):	288	325	425
Mechanical/Aerospace/Industrial (MAI):	278	299	308
Chemical/Biomedical (CB):	174	200	217
Civil/Construction (CC):	130	116	113

The analysis is similar to a GOF, but we find the 'expected' values by using the expected value for a category computing from the right-hand totals for all the data:

	2012-2013	2013-2014	2014-2015	Totals	<i>expected percentages</i>
(ECS):	288 / $\frac{288}{314.326}$	325 / $\frac{325}{339.617}$	425 / $\frac{425}{384.056}$	1038	$1038/2873 = .361...$
(MAI):	278 / $\frac{278}{267.995}$	299 / $\frac{299}{289.559}$	308 / $\frac{308}{327.447}$	885	$885/2873 = .308...$
(CB):	174 / $\frac{174}{178.966}$	200 / $\frac{200}{193.366}$	217 / $\frac{217}{218.668}$	591	$591/2873 = .2057...$
(CC):	130 / $\frac{130}{108.712}$	116 / $\frac{116}{117.459}$	113 / $\frac{113}{132.829}$	359	$359/2873 = .1249...$
Totals	870	940	1063	2873	

χ^2 contributions

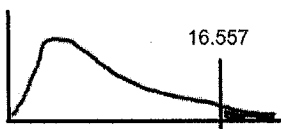
	2012-2013	2013-2014	2014-2015
(ECS):	2.205	.629	4.365
(MAI):	.374	.308	1.155
(CB):	.138	.228	.013
(CC):	4.169	.018	2.960

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$\chi^2 = 16.557$$

$$df = (\# \text{rows} - 1)(\# \text{columns} - 1)$$

$$df = 3 * 2 = 6$$



$$p\text{-value} = \chi^2 \text{cdf}(16.557, 999, 6) = .011$$

For a Chi-square Test for Homogeneity, the hypotheses are:

H_0 : *The proportions are homogeneous across the distributions*

H_A : *At least one set of proportions are not homogeneous*

Here, with: $p\text{-value} = \chi^2 \text{cdf}(16.557, 999, 6) = .011$

...we would reject H_0 .

“ There has been a change in the distribution of engineering degree fields over the last 3 years. “

That is useful to know, but doesn't provide much information. In what way have the fields changed?

To analyze this, we need to consider each cell count compared to the expected value for that cell. The difference is a difference between an observed value and a 'model' or 'expected' value. As with linear regression, these difference represents **residuals**.

For each cell, you can calculate the **standardized residual**:

$$\text{standardized cell residual} = c = \frac{(\text{Observed} - \text{Expected})}{\sqrt{\text{Expected}}}$$

(This is just the square root of the cell's contribution to χ^2 but preserving sign)

Cell residuals:

	2012-2013	2013-2014	2014-2015
(ECS):	-1.484	-.793	2.042
(MAI):	.611	.555	-1.075
(CB):	-.371	.477	-.113
(CC):	2.042	-.135	-1.72

It can be shown that these residuals are really z-scores, so values above 2 are 'unusual' because they are 2 standard deviation above expected value.

These residuals indicate that in 2012-2013 there was an unusually high number of degrees awarded in Civil/Construction engineering, and in 2014-2015 there was an unusually high number of degrees awarded in Electrical/Computer/Software engineering. All the other fields are within expected values for each year.

Chi-Squared Test for Homogeneity on a Ti-84

- 1) Enter the entire table into matrix A.
- 2) STAT, TESTS, χ^2 -Test
- 3) Calculate. It will put each cell's expected value into matrix B.



This is **much** faster than doing the test by hand, and we will always use the calculator.

Unfortunately, the calculator does not provide a convenient way to look at the residuals. If you find that the data is not homogeneous, you need to compute each cell's residual by hand (using the original data and the expected values from the B matrix).

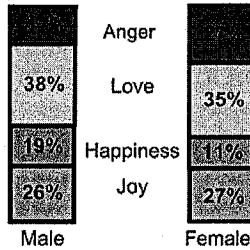
Chi-Squared Test for Independence

There is one final use for Chi-square analysis. What if we have a table, but the columns don't represent different populations, they instead represent a 2nd variable?

A researcher asked children which of four different emotions they associated with the color red. The results were recorded by gender. Is there an association between gender and emotion attached to red? (Are the variables, 'gender' and 'emotion for red' independent?)

	Joy	Happiness	Love	Anger
Male:	28	20	40	18
Female:	61	25	80	60

Until now, the only way we had to analyze this was to make a segmented bar graph and visually look for differences in the percentages:



But a Chi-square analysis will give a p-value for how significant the variations are from each other.



H_0 : Gender and Emotion for Red are independent (differences are not significant)

H_A : Gender and Emotion for Red are not independent (differences are significant)

With high p-value, we cannot reject H_0 . Gender and Emotion for Red are independent.

Homogeneity vs. Independence

Both are analysis of a two-way table. It really depends upon how the data were collected:

1) Ask a **SRS** of voters two questions: Do you support building a border fence? What is your party affiliation?

	Democrat	Republican	Other	
Support border fence	18	35	20	Independence
Oppose border fence	38	12	26	

2) Divide voters into groups by party affiliation. Ask a **SRS of each group** one question: do you support building a border fence?

	Democrat	Republican	Other	
Support border fence	18	35	20	Homogeneity
Oppose border fence	38	12	26	

Summary: How do you know which test to use?

χ^2 Goodness-Of-Fit (GOF)

When you have **1** population, and you are examining **1** categorical variable.

Data in L1, Expected counts in L2, χ^2 GOF-Test.

H_0 : All proportions match expected proportions

H_A : At least one proportion doesn't match expected proportions

χ^2 Test for Homogeneity

When you have **multiple** populations, and you are examining **1** categorical variable with data from each of the populations.

Data in matrix A, χ^2 -Test (matrix B contains expected counts - residuals by hand)

H_0 : The proportions are homogeneous across the distributions

H_A : At least one set of proportions are not homogeneous

χ^2 Test for Independence

When you have **1** population, and you are examining **2** categorical variables.

Data in matrix A, χ^2 -Test (matrix B contains expected counts - residuals by hand)

H_0 : Variables are independent

H_A : Variables are not independent