SE

The chi-squared test

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}$$

To compare predictions and observations

Comparing frequencies

Consider a situation in which we have count data and want to know if it matches our predictions.

We can compare numbers observed to the numbers expected (i.e., predicted).

If they're similar, prediction method seems good.
If they're different, prediction method seems bad.

But sampling error and chance will cause a mismatch, how different is too different?

Obs	Exp
4	4
6	5
5	3
5	6
2	3

Obs	Exp
1	4
9	5
7	3
2	6
4	2

Standard Statistical Test Procedure

Ask question about population count data

H_0 observed = expected (prediction good)
H_A observed ≠ expected (prediction bad)

Calculate a **test statistic** from data

Q: If H_0 true, what is prob. of **test statistic**?
A: probability = p

If p small, then H_0 probably not true → reject H_0
If p not small, then H_0 may be true → fail to reject H_0

We need a good test statistic, call it χ^2

Obs	Exp
4	4
6	5
5	3
5	6
2	3

small χ^2

Obs	Exp
1	4
9	5
7	3
2	6
4	2

large χ^2

The Chi-Squared test statistic

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i}\right] = \sum \frac{(obs - exp)^2}{exp}$$

The chi-squared statistic is the sum of the squared deviations between the observed and expected values, relative to the expected values.

This value is larger for bigger mismatches.

e.g., for the two data sets shown: χ^2 = 2.03 χ^2 = 15.45

Obs	Exp
4	4
6	5
5	3
5	6
2	3

small χ^2

Obs	Exp
1	4
9	5
7	3
2	6
4	2

large χ^2

Goodness of fit χ^2 tests

Any mathematical model can be used to predict the number of observations.

Observed	Model predicts
4	5
6	5
5	5
...	...
...	...

The df = the number of categories (k) - 1,
- (the # parameters that the data is used to help estimate).

Examples:

Uniform distribution: df=k-1 (no parameters estimated)
Binomial distribution: df=k-3 (2 parameters come from data, p & k)
Normal distribution: df=k-3 (2 parameters come from data, μ & σ^2)
Poisson distribution: df=k-2 (1 parameter comes from data, $\mu=\sigma^2$)

Types of Chi-Squared tests

A chi-squared test compares observed to predicted using a *model*. There are 3 general types of tests.

Observed	Model predicts
4	5
6	5
5	5
...	...
...	...

1. "Goodness of Fit" tests: the method is a mathematical model. Any model can be used.
2. Tests of "independence" or "homogeneity": the independence of two factors is assumed and used to predict the number of observations in categories representing combinations of factors.

	1/2	1/3	1/6	
1/4	7	3	2	
3/4	17	13	6	Total=48

→

1/4 of 1/2 of 64 = 6	1/4 of 1/3 of 64 = 4	1/4 of 1/6 of 64 = 2
3/4 of 1/2 of 64 = 18	3/4 of 1/3 of 64 = 12	3/4 of 1/6 of 64 = 6

Using the Chi-Squared test statistic

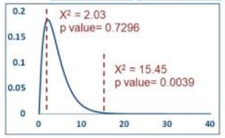
1. Line up the observed and expected values
2. Calculate the chi-squared statistic $X^2 = \sum_{i=1}^{5}\left[\frac{(O_i - E_i)^2}{E_i}\right]$
3. Compare this X^2 to the χ^2 distribution.

Area under curve to the right is the probability of getting an X^2 value that large by chance alone, the **p value**.

4. Interpret magnitude of the p value to "*reject*" or "*fail to reject*" the null hypothesis that *the method* used to make predictions is good.

Obs	Exp
4	4
6	5
5	3
5	6
2	3

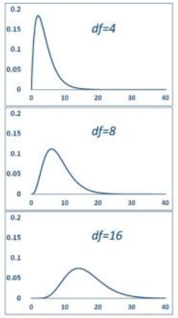
Obs	Exp
1	4
9	5
7	3
2	6
4	2



The Chi-Squared distribution

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i}\right] = \sum \frac{(obs - exp)^2}{exp}$$

- Always positive
- Larger as mismatches increase in magnitude.
- Larger as number of categories (k) increases.
- Skewed right (becomes normal for large df).

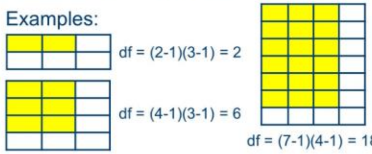


Independence/homogeneity χ^2 tests

Independence of rows and columns used to predict the number of observations in each combination.

The df = (rows -1) x (columns -1).

Examples:



Obs

	1/2	1/3	1/6	
1/4	7	3	2	
3/4	17	13	6	Total=48

Exp

	1/2	1/3	1/6	
1/4	6	4	2	
3/4	18	12	6	Total=48

Obs

	1/4	1/2	1/4	
1/2	7	16	9	
1/2	9	16	7	Total=64

Exp

	1/4	1/2	1/4	
1/2	8	16	8	
1/2	8	16	8	Total=64

Independence vs homogeneity tests

The math is identical, but the experimental design differs

Independence: samples randomly from *one population* and measures numbers observed in two sets of categories.

Homogeneity: samples randomly from *multiple populations* and measures numbers in one set of categories (the population is one of the sets of categories).

Chi-Squared test formal procedure

▶ Create null and alternative hypotheses:
H_0: Observation counts match predicted counts.
H_A: Observation counts don't match predicted counts.

▶ Calculate X^2 test statistic: $X^2 = \sum_{i=1}^{5}\left[\frac{(O_i - E_i)^2}{E_i}\right]$

▶ Compare X^2 to various χ^2 values (α=0.05).

▶ Determine probability, **p value**, of seeing X^2 as large as we do.

▶ Decide to "reject H_0" or "fail to reject H_0" based on the p value.
H_0: p > 0.05, observation counts roughly match predicted counts.
H_A: p < 0.05, observation counts don't match predicted counts.

Chi-Squared test conclusions

What does it mean to **reject** the null hypothesis.

Goodness of fit test:
- 1+ assumptions of the math model are violated.

Independence test:
- There is a non-random association between some of the factors.

Homogeneity test:
- The frequencies of the categories differ between the populations.

Chi-Squared Test

The chi-squared test is perhaps the most versatile test there is.

Observations can be compared to predictions from **ANY** mathematical model.

- as long as the data is # of observations (i.e., count data).

This allows for an almost unlimited number of types of tests.

Chi-Squared 2x2 special cases: Fisher's Exact Test

2x2 tables may not be the best for χ^2 analyses, especially if:

- The overall total # observations is less than 20
- The overall total # observations is 20-40, with any predictions less than 5.

In these cases, Fisher's Exact Test should be used instead.

The math is complicated, but this test can give exact p values for these cases when the χ^2 test would be inaccurate.

Chi-Squared 2x2 special cases: the Yates correction

The X^2 distribution is discrete, but the χ^2 distribution is continuous. This can cause problems, especially for 2x2 tables with small predictions.

Therefore, the Yates Correction was proposed to improve the match and give more conservative X^2 values (i.e., lower risk of type I error).

$$X^2_{Yates} = \sum_{i=1}^{4}\left[\frac{\left(|O_i - E_i| - \frac{1}{2}\right)^2}{E_i}\right]$$

Most modern statisticians consider this correction to be *too* conservative (i.e., it increases the risk of type II error too much).

Chi-Squared test minimum predicted value caution

Problem: extremely small predictions for categories can create large X^2 values from just one or two chance observations.

Rules of thumb:
- no prediction less than 1.0 allowed.
- maximum of 20% of predictions < 5.

Solution: combine adjacent categories to create larger predicted values.

Fruit	Predicted
Apples	7
Bananas	12
Oranges	2
Grapes	4
Plums	9

→

Fruit	Predicted
Apples	7
Bananas	12
Oranges or grapes	6
Plums	9

StatsExamples.com