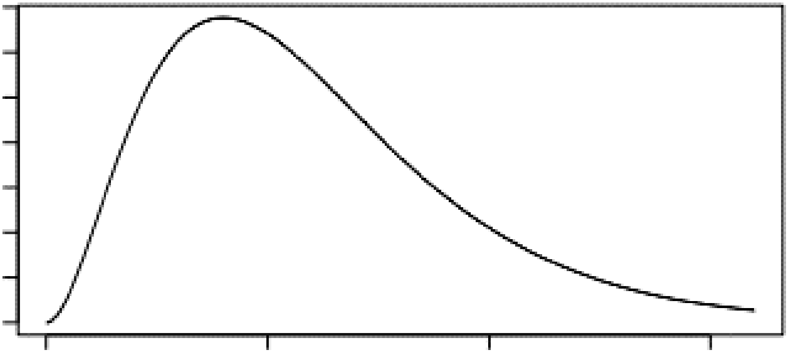
χ^2 Tests

Math 122 - Introduction to Statistics and Probability

The symbol

$$\chi$$

is the Greek letter "chi"



A χ^2 distribution.

Observed Frequencies

In this section we consider observed frequencies (counts) arranged in rows or columns and how they match some pre-described distribution.

Goodness of Fit

A GOODNESS OF FIT test will test claims that observed frequencies either do or do not fit a claimed distribution.

H_0: The observed frequencies match the claimed distribution.
H_1: The observed frequencies do not match the claimed distribution.

Expected Frequencies

For each category or class or possible outcome, we can calculate an EXPECTED FREQUENCY.

Expected Frequencies

Suppose that all pieces of a certain type of candy are red, blue, green, or orange.

Calculate the expected frequencies among 200 pieces if all colors are equally likely.

Expected number of pieces of each color is $\frac{200}{4} = 50$

If all categories are equally likely, the expected frequency of each category is

$$\frac{\text{number of trials}}{\text{number of categories}}$$

Expected Frequencies

Suppose that all pieces of a certain type of candy are red, blue, green, or orange.

Suppose the claimed distribution is

red	0.1
blue	0.4
green	0.4
orange	0.1

Calculate the expected frequencies among 200

red	$0.1 \times 200 = 20$
blue	$0.4 \times 200 = 80$
green	$0.4 \times 200 = 80$
orange	$0.1 \times 200 = 20$

Notation

$$O = \text{OBSERVED FREQUENCY}$$
$$E = \text{EXPECTED FREQUENCY}$$

Test Statistic for Goodness of Fit

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

Interpretation of χ^2

If the observed frequencies are VERY different from the expected, then χ^2 will be large.
In this case, we reject H_0. The observations do not match the claimed distribution.

If the observed frequencies are not very different from the expected, then χ^2 will be small.
In this case, we fail to reject H_0. The observations seem to match the claimed distribution.

This is a right tail test.

FIND χ^2

1. Enter the observed frequencies in $L1$.
2. Enter the expected frequencies in $L2$.
3. Enter $(L1 - L2)^2/L2$ in $L3$.
4. Add up the entries in $L3$ with “sum(”
    - Found under
      2ND, LIST (STAT), MATH, 5: sum(

$$P = \chi^2 cdf(\text{test statistic}, 9999999999, \text{categories-1})$$

Last digits in reported weights

We suspect that the weights reported by a group of people are not actual measurements. If they were, the final digits in the weights would be uniformly distributed (all 10 digits equally likely). Use the frequency counts below to test the claim that the final digits in these weights are not uniformly distributed.

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	7	14	6	10	8	4	5	6	12	8

Number of Games Played in World Series

If teams in the World Series are equally matched and if each team is equally likely to win each game, then we can calculate the probability of a team winning the Series in a fixed number of games.

Number of Games Played in World Series

Below is a table of frequencies of the number of games needed to win the Series in years past along with the expected frequencies of each number.

Games	Actual	Expected
4	19	$2/16$
5	21	$4/16$
6	22	$5/16$
7	37	$5/16$

Use this data to test the claim that the actual frequencies match the distribution expected if the teams are equally matched.

M&M's

35 women and 40 men were asked if they prefer yellow, orange, or red M&M's.

	Yellow	Orange	Red	**Row Totals**
Female	14	10	11	35
Male	16	14	10	40
Column Total	30	24	21	Total: 75

This is a CONTINGENCY TABLE

M&M's

We can use a contingency table such as the one below to test a claim such as

"A person's M&M preference is independent of that person's gender."

	Yellow	Orange	Red	**Row Totals**
Female	14	10	11	35
Male	16	14	10	40
Column Total	30	24	21	Total: 75

Expected Frequencies

Assume that the rows and columns are independent and calculate the probability that a randomly chosen person falls into one of the categories in the table.

There are a total of 75 people in the table.

30 are in the first column.
35 are in the first row.

The probability that a randomly chosen person is in the first column is $\frac{30}{75}$
The probability that a randomly chosen person is in the first row is $\frac{35}{75}$
The probability that a randomly chosen person is in the first row AND the first column is $\frac{30}{75} \cdot \frac{35}{75}$

Expected Frequencies

The expected frequency for an entry in the table if rows and columns are independent is

$$\frac{\text{number in column}}{\text{total}} \cdot \frac{\text{number in row}}{\text{total}} \cdot \text{total}$$

(You will only need this if you decide to use a TI instead of the online calculator.)

Observed and Expected Frequencies

Observed Frequencies:

	Yellow	Orange	Red	**Row Totals**
Female	14	10	11	35
Male	16	14	10	40
Column Total	30	24	21	Total: 75

Expected Frequencies:

	Yellow	Orange	Red	**Row Totals**
Female	14	11.2	9.8	35
Male	16	12.8	11.2	40
Column Total	30	24	21	Total: 75

We can now do a χ^2 GOF test.

M&M's

"A person's M&M preference is independent of that person's gender."

	Yellow	Orange	Red
Female	14	10	11
Male	16	14	10

Clinical Trials

Chantix is a drug used as an aid for those who want to stop smoking. The adverse reaction of nausea has been studied in clinical trials. A contingency table summarizes the study below.

Use a 0.01 significance level to test the claim that nausea is independent of whether the subject took a placebo or Chantix.

	Placebo	Chantix
Nausea	10	30
No Nausea	795	791