CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture M5c: Goodness of Fit Testing; Chi-square

Jimi Oke

UMass Amherst

College of Engineering

November 13, 2025

Outline

Today's objectives

- Get introduced to the χ^2 distribution
- Conduct χ^2 tests
- Conduct goodness-of-fit tests using the χ^2 distribution

Reading: Section 6.3, Open Intro Statistics

Scenario 1:
Recall the assumptions made for samples on which we conduct inference:

1. Independence
2. Normality (i.e. if the Central Limit Theorem applies)

How do we test for these?

Scenario 2:
You want to create a focus group for a study which is representative of demographic groups of interest (e.g. income, race/ethnicity, education level, major, etc.) How would you evaluate whether your sample is representative/balanced or biased toward a certain group?

Group	A	B	C	D	Total
Observed counts	$\hat{n}_A$	$\hat{n}_B$	$\hat{n}_C$	$\hat{n}_D$	
Expected counts	n_A	n_B	n_C	n_D	

Scenario 2 (cont.)

If we define a test statistic χ^2 as:

$$\chi^2 = Z_A^2 + Z_B^2 + Z_C^2 + Z_D^2 \tag{1}$$

where:

$$Z_A^2 = \frac{(\hat{n}_A - n_A)^2}{n_A} = \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} \tag{2}$$

and so on, then it turns out that χ^2 follows a special distribution called the χ^2 distribution

Chi-square testing

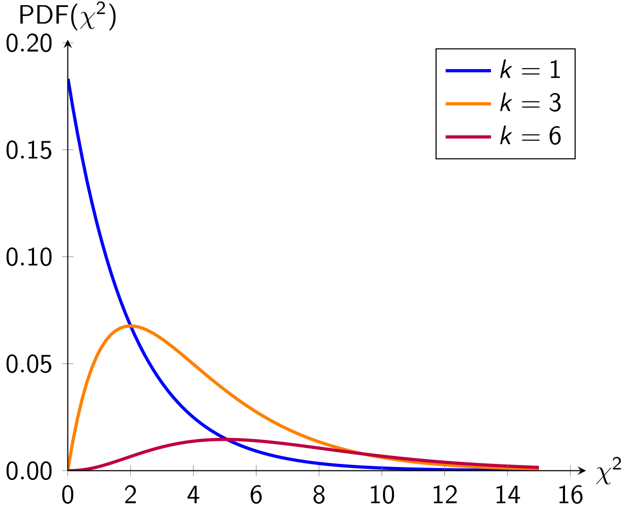
The **chi-square** (χ^2) test provides a framework to judge:

1. if categories of two factors occur **independently** of each other in a population
2. proportions in different categories are the same for all populations (homogeneity)
3. fitness of the frequency of observations from a multinomial experiment (i.e. multiple outcomes) to specified probabilities
4. more generally, the fitness of a theoretical probability distribution model to an experimental dataset

To conduct chi-square tests, we use the χ^2 distribution, whose properties we can exploit for statistical inference (hypothesis testing, and so on)

The chi-square distribution

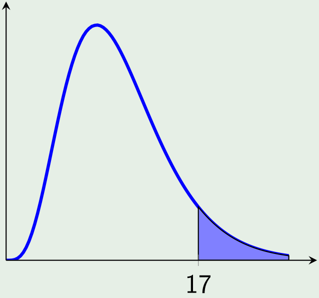
The chi-square (χ^2) distribution is specifed as $\chi^2(k)$ or χ^2_k, where k is the single parameter called **degrees of freedom** (or "df").



Chi-square distribution (cont.)

Example 1: CDF of chi-square distribution

Estimate the area under the χ^2 curve where $k = 9$.



(a) 0.05

(b) 0.02

(c) between 0.02 and 0.05

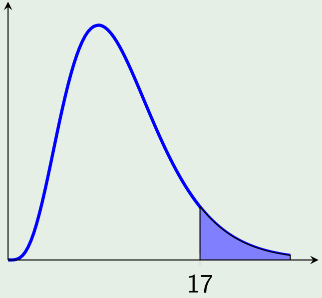
(d) between 0.05 and 0.1

(e) between 0.01 and 0.02

Chi-square distribution (cont.)

Example 1: CDF of chi-square distribution (cont.)

Estimate the area under the χ^2 curve where $k = 9$.

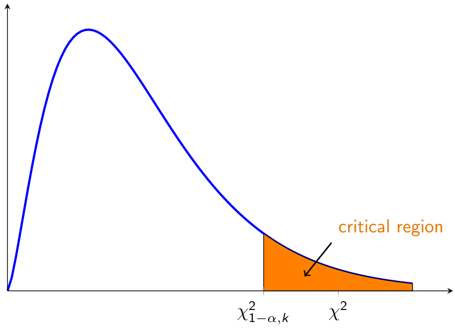


In Python, you can find the area using `1 - chi2.cdf(17, 9)` $= 0.0487$.
Answer: (c) between 0.02 and 0.05

Hypothesis testing

When conducting hypothesis tests using the chi-square distribution, the test is always upper-tailed.

- Test statistic: χ^2
- Null hypothesis: χ^2 follows a chi-square distribution
- Alternative hypothesis: χ^2 does not follow a chi-square distribution

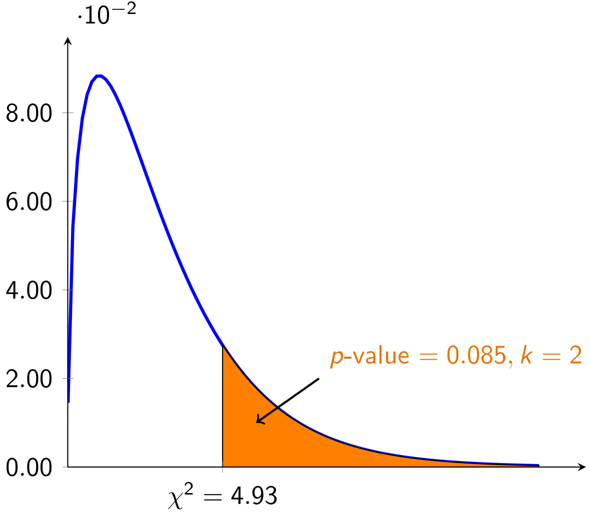


The critical value is $\chi^2_{1-\alpha,k}$ (`chi2inv(1 - alpha, nu)`).

- If $\chi^2 > \chi^2_{1-\alpha,k}$, **reject null hypothesis**
- If $\chi^2 \leq \chi^2_{1-\alpha,k}$, **fail to reject null hypothesis**

p-values for chi-square tests

As our focus is on upper-tailed tests, the p-value is given as the area to the right of the test statistic under the chi-square curve (`chi2.sf(test, df)`)



Working with the chi-square distribution

Example 2: Hypothesis testing

What conclusion would be appropriate for an upper-tailed chi-square test given the following:

$$\begin{aligned} \alpha &= 0.05 \\ \chi^2 &= 12.25 \quad \text{(Test statistic)} \\ k &= 4 \quad (\chi^2 \text{ parameter: degrees of freedom}) \end{aligned}$$

Working with the chi-square distribution

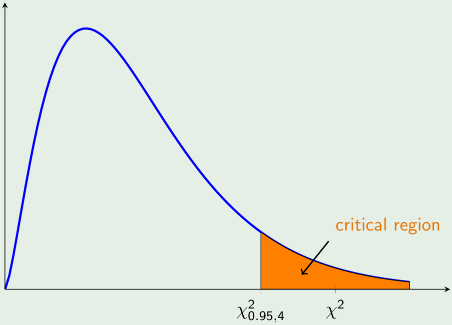
Example 2: Hypothesis testing (cont.)

The critical value is given by:

$$\chi^2_{1-\alpha,k} = \chi^2_{0.95,4} = \texttt{chi2.ppf(.95, 4)} = 9.4877$$

We compare the test statistic:

$$\chi^2 = 12.25 > \chi^2_{0.95,4} = 9.4877$$



The test statistic is in the critical region. We therefore **reject the null hypothesis**.

Working with the chi-square distribution (cont.)

Example 3: *p*-value

Give as much information as you can about the *p*-value for an upper-tailed chi-square test in this situation:

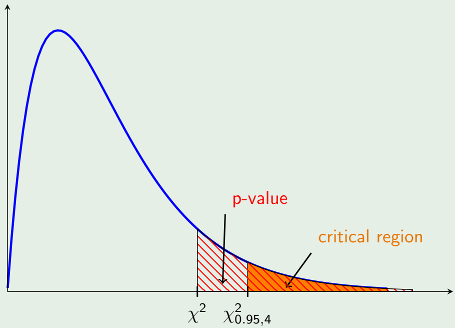
$$\begin{aligned} k &= 4 \\ \chi^2 &= 7.5 \end{aligned}$$

Working with the chi-square distribution (cont.)

Example 3: *p*-value (cont.)

First, we can look up the critical value at a given significance level.
Let us choose $\alpha = 0.05$. The critical value is given by:

$$\chi^2_{0.95,4} = \texttt{chi2inv}(.95, 4) = 9.4877$$



Since the test statistic, $\chi^2 = 7.5 < \chi^2_{0.95,4} = 9.4877$, then the test statistic is in the region of non-rejection.

We can therefore say that the *p*-value > 0.05.

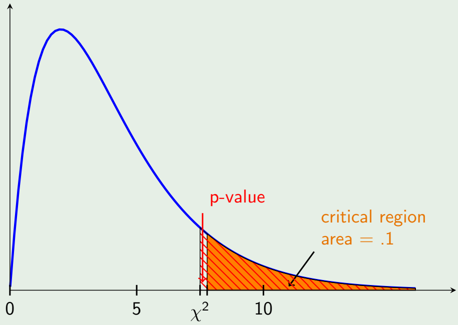
But how much greater?

Working with the chi-square distribution (cont.)

Example 3: *p*-value (cont.)

We can increase α further and try $\alpha = 0.10$. Thus:

$$\chi^2_{0.90,4} = \texttt{chi2inv}(.9, 4) = 7.7794$$



We see that the critical value is still greater than χ^2, but the difference is much smaller.

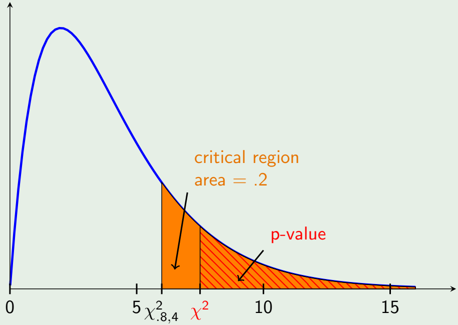
$$\chi^2 = 7.5 < \chi^2_{0.90,4} = 7.7794$$

Thus, we can say that the *p*-value is quite close to 0.10.

Working with the chi-square distribution (cont.)

Example 3: *p*-value (cont.)

We can also see that $\chi^2_{.8,4} = \texttt{chi2inv}(.8, 4) = 5.9886$. This is evidence that the *p*-value is not greater than 0.20.



From MATLAB, we find that the *p*-value is given by `chi2cdf(7.5,4,'upper')`:

$$F(\chi^2, k) = F(7.5, 4) = \boxed{0.1117}$$

Note that F here represents the CDF of the χ^2 distribution (not the F distribution).

Goodness-of-fit testing

- Chi-square tests can be used to examine whether data from a sample fit expected values from a theoretical model/distribution
- If the null hypothesis that the observed frequencies are equal to their theoretical counterparts, then the test statistic follows a chi-square distribution, and thus we use it to conduct the hypothesis test
- Generally, we can write the **test statistic** as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{3}$$

  where o_i is the number of observations in each group or interval i, and e_i is the expected/theoretical frequency for each group/interval.
- The k (degrees of freedom) parameter is given by $k = k - 1$.

Test statistic

Given n observations with expected probabilities p_{i0}, the statistic:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \sum_{i=1}^{k} \frac{(o_i - np_{i0})^2}{np_{i0}} \tag{4}$$

is approximately **chi-square** distributed with $k = k - 1$, **provided that** $np_i \geq 5$ **for all** i.

Steps to perform chi-square goodness-of-fit tests

Step 1. Order/sort the observations (if necessary)

Step 2. Group the observations (or aggregate into reasonable intervals)

Step 3. Find the frequencies for each group

Step 4. Compute the theoretical frequencies for each group:

$$n \times P(a_i < X \leq b_i) \tag{5}$$

Step 5. Compute the chi-square statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{6}$$

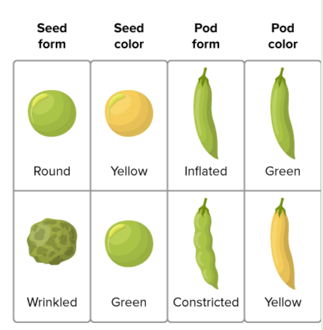
Step 6. Perform the hypothesis test and conclude

Example 4: Peas in a pod

The distribution of the dominant alleles (forms of a gene) in peas (Y = yellow color, R = round shape) is binomial with $p = \frac{9}{16}$. An experiment was performed to test the hypothesis that all the observed frequencies are equal to the theoretical ones (null hypothesis) versus the alternative that at least one is not. In this experiment, n four-seed pods were examined. In a randomly selected pod, possible X values were 0, 1, 2, 3 and 4.

The following data are given (**one-way table**):

Cell *i*	**1**	**2**	**3**	**4**	**5**
YR peas/pods	**0**	**1**	**2**	**3**	**4**
Observed	16	45	100	82	26
Expected					
$(o_i - e_i)^2/e_i$					



Goodness-of-fit testing

Example 4: Peas in a pod (cont.)

Given $n = 16 + 45 + 100 + 82 + 26 = 269$, we compute the theoretical frequencies:

$$
\begin{aligned}
e_i &= n \times \binom{n}{i-1} p^{i-1} (1-p)^{4-(i-1)} \\
e_1 &= 269 \times \binom{4}{0} \left(\frac{9}{16}\right)^0 \left(\frac{7}{16}\right)^4 = 9.86 \\
e_2 &= 269 \times \binom{4}{1} \left(\frac{9}{16}\right)^1 \left(\frac{7}{16}\right)^3 = 50.38 \\
e_3 &= 269 \times \binom{4}{2} \left(\frac{9}{16}\right)^2 \left(\frac{7}{16}\right)^2 = 97.75 \\
e_4 &= 269 \times \binom{4}{3} \left(\frac{9}{16}\right)^3 \left(\frac{7}{16}\right)^1 = 83.78 \\
e_5 &= 269 \times \binom{4}{4} \left(\frac{9}{16}\right)^4 \left(\frac{7}{16}\right)^0 = 26.93
\end{aligned}
$$

Goodness-of-fit testing

Example 4: Peas in a pod (cont.)

Our updated table is now:

Cell *i*	**1**	**2**	**3**	**4**	**5**
YR peas/pods	**0**	**1**	**2**	**3**	**4**
Observed	16	45	100	82	26
Expected	9.86	50.38	97.75	83.78	26.93
$(o_i - e_i)^2/e_i$					

We can then compute the normalized squared deviations

Goodness-of-fit testing

Example 4: Peas in a pod (cont.)

$$
\begin{aligned}
(n_1 - e_1)^2/e_1 &= (16 - 9.86)^2/9.86 = 3.823 \\
(n_2 - e_2)^2/e_2 &= (45 - 50.68)^2/50.68 = 0.637 \\
(n_3 - e_3)^2/e_3 &= (100 - 97.75)^2/97.75 = 0.052 \\
(n_4 - e_4)^2/e_4 &= (82 - 83.78)^2/83.78 = 0.038 \\
(n_5 - e_5)^2/e_5 &= (26 - 26.93)^2/26.93 = 0.032
\end{aligned}
$$

Goodness-of-fit testing

Example 4: Peas in a pod (cont.)

Now, we have all we need to compute the chi-square test statistic:

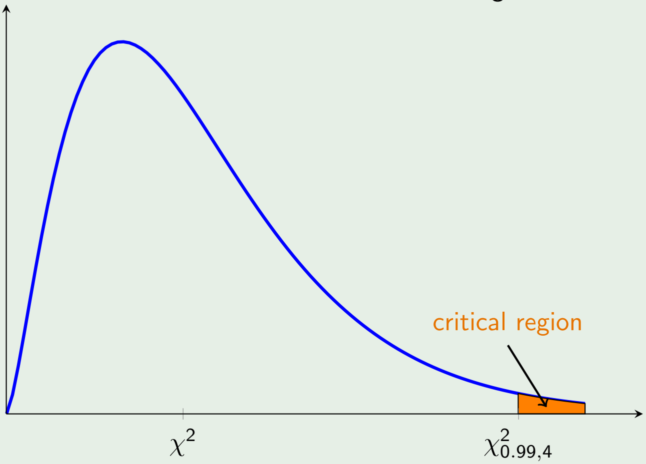
$$\begin{aligned}\chi^2 &= \sum_{i=1}^{G} (o_i - e_i)^2 / e_i \\ &= 3.823 + 0.637 + 0.052 + 0.038 + 0.032 \\ &= 4.582\end{aligned}$$

The degrees of freedom, $k = G - 1 = 5 - 1 = 4$.
Thus, the critical value is given by $\chi^2_{0.99,4} = 13.2767$.

Goodness-of-fit testing

Example 4: Peas in a pod (cont.)

Since $\chi^2 = 4.582 < \chi^2_{0.99,4} = 13.2767$, we **fail to reject** the null hypothesis. This indicates that the binomial model is a good fit for the data.



Example 5

We will revisit Scenario 2 as an in-class activity:

- Each of you will be assigned to a group
- In your group, count the following:
  - number of CEE majors ($\hat{n}_A$)
  - number of ME majors ($\hat{n}_B$)
  - number of IE majors ($\hat{n}_C$)
  - number of Other majors ($\hat{n}_D$)
- Goal: test whether the distribution of majors in your sample (group) is representative of the expected (true) counts (n_A, n_B, n_C, n_D) based on the proportions obtained from the class-wide survey you just completed.

Summary

- The χ^2 distribution has a single parameter k (degrees of freedom)
- It is the underlying the distribution for chi-square tests, which are used to evaluate whether observed data fit a given distribution (or that proportions/frequencies observed from a sample follow a theoretical model/assumption).
- Critical value: $\chi^2_{1-\alpha,k} \equiv$ `chi2.ppf(1—alpha, k)`
- p-value: `chi2.cdf(test, k)`
- When given a one-way table, the test statistic χ^2 is computed as

$$\chi^2 = \sum_{i=1}^{G} \frac{(o_i - e_i)^2}{e_i}, \quad k = G - 1 \tag{7}$$

  where G is the number of groups
- Reject null hypothesis: p-value $< \alpha$ OR $\chi^2 > \chi^2_{1-\alpha,k}$
- Fail to reject null hypothesis: p-value $\geq \alpha$ OR $\chi^2 \leq \chi^2_{1-\alpha,k}$
- Reading: Open Intro Stats, Section 6.3

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As the
unprocessed data that should have been added to the final page this e
has been added to receive it.
If you rerun the document (without altering it) this surplus page will g
because LaTeX now knows how many pages to expect for this documen