Example: Construct a 95% confidence interval for the difference in the cure rates of the two drugs.

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$= \left(\frac{51}{120} - \frac{88}{150}\right) \pm 1.96 \sqrt{\frac{\left(\frac{51}{120}\right)\left(\frac{69}{120}\right)}{120} + \frac{\left(\frac{88}{150}\right)\left(\frac{62}{150}\right)}{150}}$$

$$= -0.16167 \pm 1.96(0.0604)$$

$$= (-0.28, -0.04)$$

$\therefore$ we are 95% confident that $p_1 - p_2 \in (-0.28, -0.04)$.

Since $0 \notin (-0.28, -0.04)$, we are 95% confident that $p_1 < p_2$, that is, we conclude with 95% confidence that the proportion of patients cured by Drug X is between 4% and 28% lower than the proportion of patients cured by Drug Z.

Chapter 22: Comparing Counts (Chi-Squared Tests)

Chi-squared tests are used to make comparisons of several proportions corresponding to one or more categorical variables.

We will consider three chi-squared tests:

- **Chi-Squared Goodness-of-Fit Test:** compares the observed distribution of a single categorical variable to an expected distribution based on a model.
- **Chi-Squared Test of Homogeneity:** compares the distribution of a single categorical variable in several groups.
- **Chi-Squared Test of Independence:** used to determine if there is an association between two categorical variables measured on the same population.

These tests are hypothesis tests that require the **counts** or number of observations from a random sample that fall into each category of a categorical variable. We display these observed counts in the **cells** of a frequency table. For each cell, there is also an expected value calculated from some hypothesized proportion. H_0

All three tests use the **chi-squared statistic**:

χ^2 large means observed vales far from expected values (evidence to reject H_0)

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

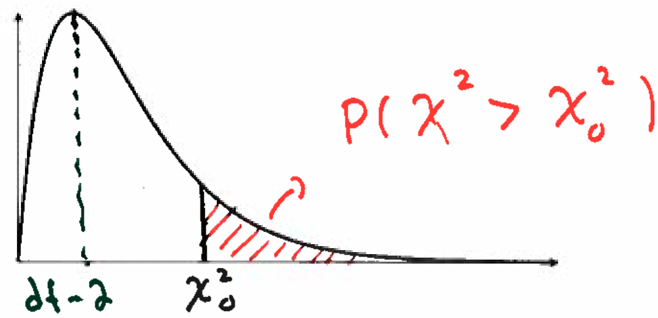
Note: We use the chi-squared statistic for hypothesis tests only. It is not used for constructing confidence intervals.

Note: Although observed values must be integers, expected values may be decimals.

counts

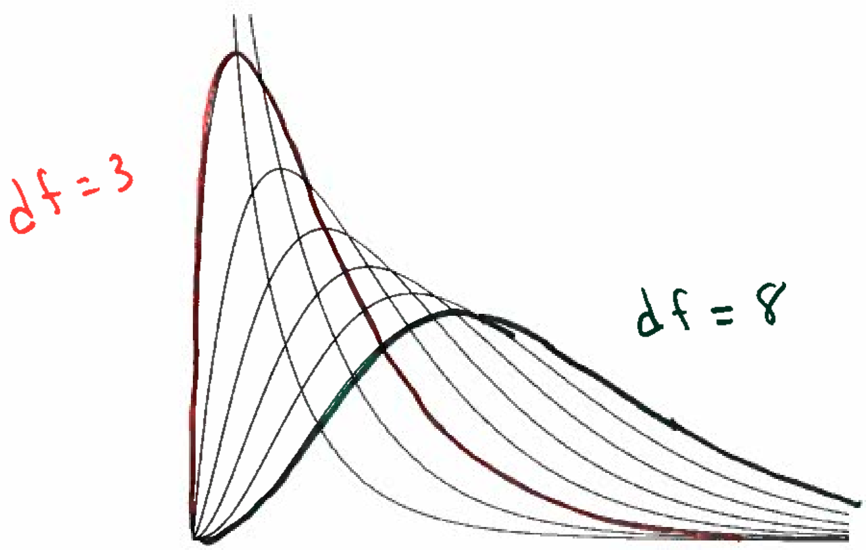
Chi-Squared Distribution

When H_0 is correct and the sample size is large enough, χ^2 can be described by a **chi-squared distribution**. In all three cases, we use this distribution to perform an **upper-tailed** hypothesis test.



A χ^2-curve has several properties:

- The total area under a χ^2-curve is 1.
- A χ^2-curve is right-skewed. It starts at 0 on the horizontal axis and extends to the right indefinitely, getting very close to, but never touching, the horizontal axis.
- There are infinitely many χ^2-curves. Each χ^2-curve is identified by its number of **degrees of freedom** (df).
- The mode (peak) is at $df - 2$ and the mean is df.
- As the number of degrees of freedom increases, the χ^2-curves look increasingly like normal curves.



Section 22.1: Goodness-of-Fit Test

Situation: We have a sample of size n consisting of observations for **one categorical variable** that has k categories. We want to know if the distribution of the data matches the distribution predicted by a specified model.

We can display the **counts** in each category in a one-way frequency table. The table has k cells, one for each of the k categories. In this case, the number of **degrees of freedom** is

$$\begin{aligned} df &= k - 1 \\ &= \text{number of categories} - 1 \\ &= \text{number of cells} - 1 \end{aligned}$$

Example: A random sample of 240 assorted nuts were selected from a shipment and the types were recorded:

Variety	Walnut	Hazelnut	Almond	Pistachio
Observed	95	70	33	42

cell

4 categories
$\Rightarrow$ 4 cells

$df = 4 - 1 = 3$

For the null hypothesis of a goodness-of-fit test, we state

$$H_0 : \text{The data matches the specified model}$$

and so H_0 can be expressed in the form

proportion in each category according to model

$$H_0 : \; p_1 = \underline{p_{1_0}}, \;\; p_2 = \underline{p_{2_0}}, \;\; \ldots \;\; p_k = \underline{p_{k_0}}$$

The alternative hypothesis is

$$H_A : \text{The data does not match the model} \quad \text{or} \quad H_A : H_0 \text{ is not true}$$

which means that **at least one** of the hypothesized proportions is not correct, but we <u>don't know which</u> one(s).

if more than two proportions

If we reject the null hypothesis, we can only say that there is enough statistical evidence to conclude that the data does not match the model, but we can't say how.

Example: Returning to the previous example, suppose that the shipment is labeled as having 45% walnuts, 20% hazelnuts, 20% almonds, and 15% pistachios.

p_1 p_2 p_3 p_4

H_0: $p_1 = 0.45$, $p_2 = 0.2$, $p_3 = 0.2$, $p_4 = 0.15$

H_A: H_0 is not true. (at least one of these proportions is not correct)

To perform a goodness-of-fit hypothesis test, we need to find the count that is **expected** for each category when the null hypothesis is true:

$$\text{expected value of } i^{\text{th}} \text{ cell/category } = np_{i_0}$$

n = sample size

where p_{i_0} is the relative frequency (or probability) of the i^{th} category cell/category given in the null hypothesis. assume H_0 true

Example: $n = 240$

Variety	Walnut	Hazelnut	Almond	Pistachio
Observed	95	70	33	42
Labeled proportion	0.45	0.2	0.2	0.15
Expected	108	48	48	36

add to 240

240(0.45) 240(0.2) 240(0.15)

In order to approximate the test statistic χ^2 by a chi-squared distribution we must have a sufficiently large sample size.

Expected Cell Frequency Condition: The expected count (frequency) in every cell of the table should be at least 5.

Note: If any of the expected cell counts are less than 5, categories may be combined (in a sensible way) to create acceptable expected cell counts. If you combine categories, compute the *df* using the reduced number of categories.

Example: Suppose that a second shipment was labeled as containing 70% peanuts, 25% almonds, 4% walnuts, and 1% Brazil nuts. A random sample of size 240 was taken.

$df = 3$

Variety	Peanuts	Almonds	Walnuts	Brazil nuts
Observed	180	53	5	2
Labeled proportion	0.7	0.25	0.04	0.01
Expected	168	60	9.6	2.4

< 5

$240(0.7)$

Variety	Peanuts	Almonds	Walnuts and Brazil nuts
Observed	180	53	7
Labeled proportion	0.7	0.25	0.05
Expected	168	60	12

$df = 3 - 1 = 2$

A goodness-of-fit (hypothesis) test has five steps:

1. **Assumptions/Conditions:**
   - Data must be in **counts** and **not** in proportions or percentages.
   - Responses of individuals should be independent of each other.
   - Sample must be chosen randomly (to generalize to the population).
   - Sample size must be large enough: each cell should have an **expected count** of at least 5.

2. **Hypotheses:**

$$H_0 : \text{The data matches the model.}$$
$$H_A : \text{The data does not match the model.}$$

3. **Test Statistic:** Use the hypothesized model (Ho) to compute the expected value for each cell. Then compute the test statistic

$$\chi_0^2 = \sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

4. ***P*-value:** The P-value $= P(\chi^2 > \chi_0^2)$ can be computed using software or using the χ^2-table with $df = k - 1$, where k is the number of cells.

5. **Conclusion:** Report and interpret the P-value in context. Given a significance level α,
   - if P - value $\leq \alpha$, we reject H_0 at level α.
   - if P- value $> \alpha$, we do not reject H_0 at level α.

Example: A shipment of assorted nuts is labeled as having 45% walnuts, 20% hazelnuts, 20% almonds, and 15% pistachios. An inspector randomly picks 240 nuts from the shipment and finds that there are 95 walnuts, 70 hazelnuts, 33 almonds, and 42 pistachios. Do these findings support an accusation of mislabeling? Use $\alpha = 0.05$.

Variety	Walnut	Hazelnut	Almond	Pistachio
Observed	95	70	33	42
Labeled proportion	0.45	0.2	0.2	0.15
Expected	108	48	48	36
	≥ 5	≥ 5	≥ 5	≥ 5

1. **Assumptions/Conditions:**
   - data in counts
   - independent, random sample
   - expected counts ≥ 5

2. **Hypotheses:**

   $H_0: p_1 = 0.45,\ p_2 = 0.2,\ p_3 = 0.2,\ p_4 = 0.15$

   H_A: H_0 is not true (at least one nut does not have the claimed proportion)

3. **Test Statistic:**

$$\chi_0^2 = \sum_{\text{all cells}} \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

$$= \frac{(95-108)^2}{108} + \frac{(70-48)^2}{48} + \frac{(33-48)^2}{48} + \frac{(42-36)^2}{36}$$

$$= 17.336$$

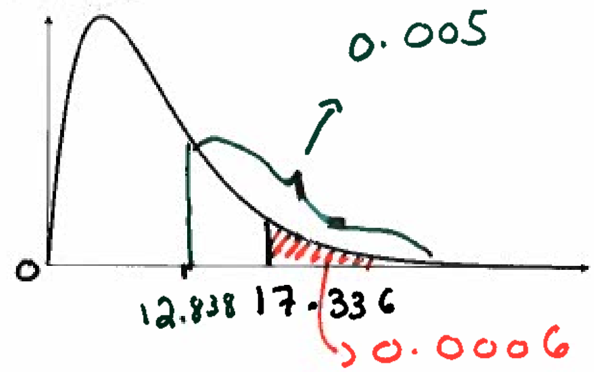
4. ***P*-value:** Use $df = 4 - 1 = 3$

By the χ^2- table,

P-value.

$= P(\chi^2 > 17.336) < 0.005$

≤ 0.05



Using StatCrunch: P-value $= 0.0006 \leq 0.05$

Chi-Square goodness-of-fit results:

Observed: Observed
Expected: Expected

N	DF	Chi-Square	P-value
240	3	17.335648	0.0006

5. **Conclusion:** Since P-value $\leq \alpha = 0.05$, we reject H_0 at the 0.05 Significance level, that is, there is enough statistical evidence to support an accusation of mislabeling.

Example: According to Statistics Canada, the percentages of birth months in Canada has the following distribution:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
National Birth %	8.2	7.4	8.3	8.1	8.8	8.5	8.9	9	8.6	8.5	7.9	7.8

In our Stat 151 class survey, 152 students responded and the following counts were observed:

	Jan	Feb	Mar	Apr	May	Jun
Observed	17	8	8	15	14	7
Expected	12.464	11.248	12.616	12.312	13.376	12.92

$152(0.082)$
$= 12.464$

	Jul	Aug	Sep	Oct	Nov	Dec
Observed	16	15	16	15	8	13
Expected	13.528	13.68	13.072	12.92	12.008	11.856

Is there enough evidence to conclude that the distribution of birth months of students in this class is not the same as the national distribution? Use $\alpha = 0.05$.

1. **Assumptions/Conditions:**
   - not randomly collected → no population inferences
   - independent
   - data in counts
   - expected counts ≥ 5
2. **Hypotheses:**

   H_0: The data matches the model.

   H_A: The data doesn't match the model.