STAT 135
11. Hypothesis testing for categorical data

Spring 2022

Lecturer: Dr Rebecca Barter (*she/her*)
Office hours: Tu 9:30-10:30 (in person), Th 4-5pm (virtual)
Office: Evans 339

Email: rebeccabarter@berkeley.edu
Twitter: @rlbarter
GitHub: rlbarter

Pearson χ^2 “Goodness-of-fit” test

The χ^2-Goodness-of-fit test

Imagine that you have some **discrete** data and you want to fit a **Poisson** distribution to it

(i.e., estimated λ from the mean of the data, and fit a Poisson($\hat{\lambda}$))

You want to test whether the Poisson fit you have generated is a good fit for the data.

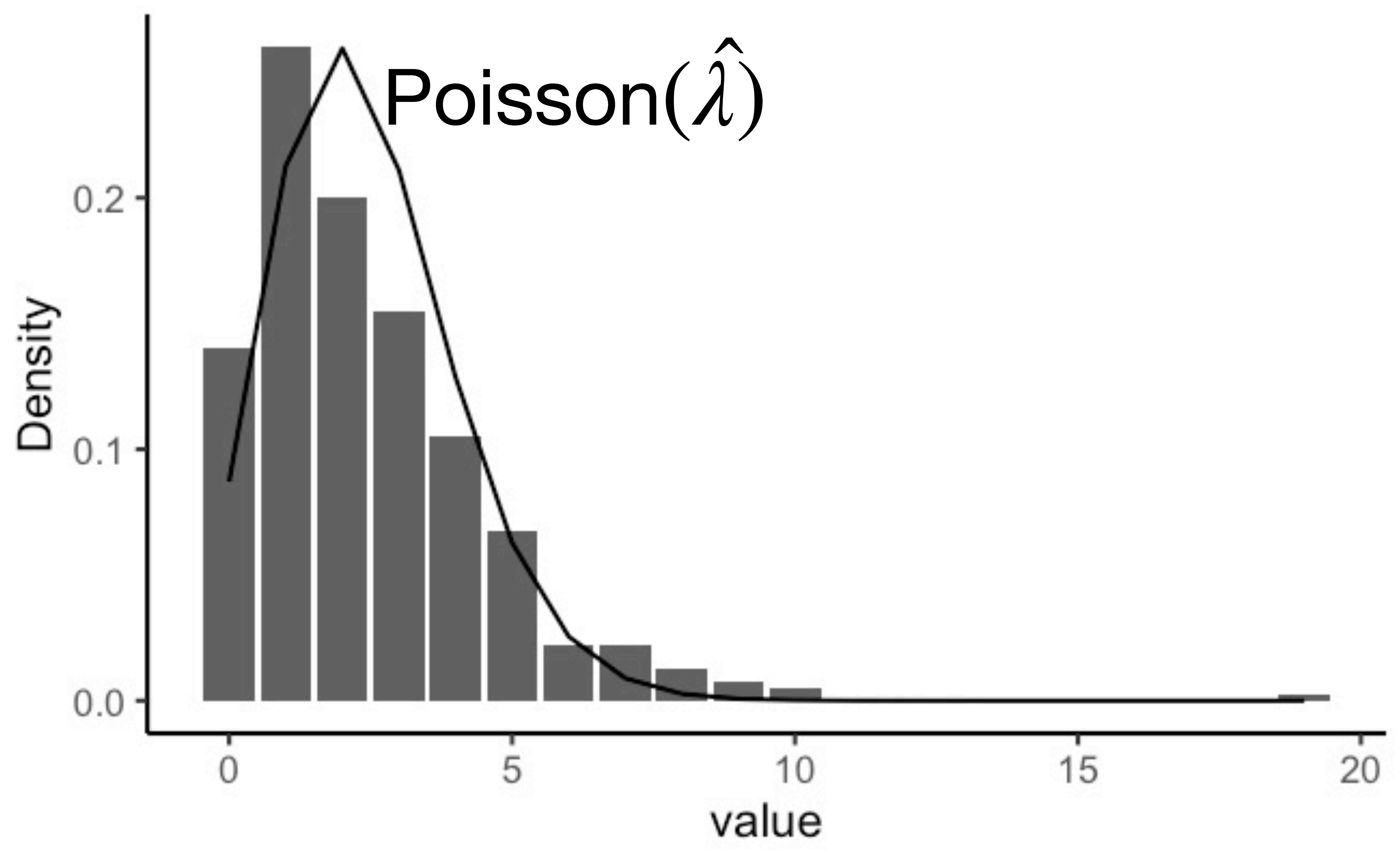
H_0 : the data come from the specified distribution

H_1 : the data *do not* come from the specified distribution

The χ^2-Goodness-of-fit test

H_0 : the data come from the specified distribution

H_1 : the data *do not* come from the specified distribution



The χ^2-Goodness-of-fit test

H_0 : the data come from the specified distribution

H_1 : the data *do not* come from the specified distribution

	Value (x)	0	1	2	3	4	5	6	7	8	9	10	19
X_i	Value (x)	0	1	2	3	4	5	6	7	8	9	10	19
O_i	Observed frequency	56	104	80	62	42	27	9	9	5	3	2	1
E_i	**Poisson Expected frequency**	**34.9**	**85.1**	**103.8**	**84.4**	**51.5**	**25.1**	**10.2**	**3.6**	**1.1**	**0.3**	**0.1**	**0**

$n = 12$ (#cells)

$N = 400 \quad (\sum_{i=1}^{n} O_i)$

$$\hat{\lambda} = \bar{X} = \frac{1}{N}\sum_i X_i O_i$$

Model the data using $X \sim Pois(2.44)$

$E = 400 \times P(X = 0) = 12.8$

$E = 400 \times P(X = 6) = 10.2$

The χ^2-Goodness-of-fit test

H_0 : the data come from the specified distribution

H_1 : the data *do not* come from the specified distribution

The χ^2 test statistic is given by

The sum is over the "cells"/ "counts"

(Under H_0)

$$\boxed{X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-c-1}}$$

Number of independent "cells"/"counts"

Number of parameters estimated from the data

Where

O is the ***observed*** value of your data

E is the ***expected*** value of your data under your fit

The χ^2-Goodness-of-fit test example

We want to fit a poisson distribution to counts of bacterial clumps in milk samples. At first a Poisson distribution seems to be a good fit, but we aren't sure. The following table, taken from Bliss and Fisher (1953) summarizes and counts the number of bacterial clumps in 400 milk samples.

X_i	Clumps per sample	0	1	2	3	4	5	6	7	8	9	10	19
O_i	Observed frequency	56	104	80	62	42	27	9	9	5	3	2	1

To fit a Poisson distribution, we will use the the MLE estimate for λ, which corresponds to the sample mean.

$$\hat{\lambda} = \frac{1}{400}(0 \times 56 + 1 \times 104 + \ldots + 19 \times 1) = 2.44$$

The χ^2-Goodness-of-fit test example

Generate the expected value under the Poisson(2.44) distribution:

$$E = NP(X = x) \qquad \text{where } X \sim Pois(2.44), \text{ and } N \text{ is total \#obs}$$

Clumps per sample (x)	0	1	2	3	4	5	6	7	8	9	10	19
Observed frequency	56	104	80	62	42	27	9	9	5	3	2	1
Expected frequency	**34.9**	**85.1**	**103.8**	**84.4**	**51.5**	**25.1**	**10.2**	**3.6**	**1.1**	**0.3**	**0.1**	**0**

$$E = 400 \times P(X = 0) = 34.9$$

$$E = 400 \times P(X = 6) = 10.2$$

The χ^2-Goodness-of-fit test example

There is a rule that says you need to aggregate cells with expected counts less than 5

Clumps per sample (x)	0	1	2	3	4	5	6	7	8	9	10	19
Observed frequency	56	104	80	62	42	27	9	9	5	3	2	1
Expected frequency	**34.9**	**85.1**	**103.8**	**84.4**	**51.5**	**25.1**	**10.2**	**3.6**	**1.1**	**0.3**	**0.1**	**0**

Becomes

Clumps per sample (x)	0	1	2	3	4	5	6	7
Observed frequency	56	104	80	62	42	27	9	20
Expected frequency	**34.9**	**85.1**	**103.8**	**84.4**	**51.5**	**25.1**	**10.2**	**5**

The χ^2-Goodness-of-fit test example

Clumps per sample (x)	0	1	2	3	4	5	6	7
Observed frequency	56	104	80	62	42	27	9	20
Expected frequency	34.9	85.1	103.8	84.4	51.5	25.1	10.2	5
$\frac{(O_i - E_i)^2}{E_i}$	12.8	4.2	5.5	5.9	1.8	0.14	0.14	45

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 12.8 + 4.2 + \ldots + 0.14 + 45 = 75.4$$

$$X^2 \sim \chi^2_{\underset{\text{8 “cells”}}{8}-1-\underset{\text{1 parameter estimated }(\lambda)}{1}} = \chi^2_6$$

The χ^2-Goodness-of-fit test example

H_0 : the data come from the specified distribution

H_1 : the data *do not* come from the specified distribution

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 75.4$$

$$X^2 \sim \chi^2_6$$

P-value = $p(X^2 \geq 75.4) \approx 0$

The p-value is always a "*greater than*" expression!

So we reject the null that our data came from the fitted poisson distribution!

The χ^2-Goodness-of-fit test

1. Compute the expected counts under H_0
2. Aggregate any cells with expected counts less than 5
3. Compute $\frac{(O_i - E_i)^2}{E_i}$ for each cell
4. Compute the test statistic $X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
5. Identify the degrees of freedom, $n - c - 1$ (number of cells minus number of parameters estimated from the data minus 1)
6. Compute the p-value based on whether the test statistic is *larger* than the test statistic using the χ^2_{n-c-1} distribution

The χ^2-Goodness-of-fit test example

See <u>chisq_goodness_of_fit.R</u>

The χ^2-Goodness-of-fit test example

In a famous experiment, Mendel crossed 556 smooth, yellow male peas with wrinkled green female peas.

According to now established genetic theory, the relative frequencies of the progeny should be

Type	Frequency
Smooth yellow	9/16
Smooth green	3/16
Wrinkled yellow	3/16
Wrinkled green	1/16

And the data that Mendel claimed to observe is as follows

Type	Observed count
Smooth yellow	315
Smooth green	108
Wrinkled yellow	102
Wrinkled green	31

Is there any evidence that Mendel just made up his data?

The χ^2-Goodness-of-fit test example

In a famous experiment, Mendel crossed 556 smooth, yellow male peas with wrinkled green female peas.

Is there any evidence that Mendel just made up his data?

Type	Frequency
Smooth yellow	9/16
Smooth green	3/16
Wrinkled yellow	3/16
Wrinkled green	1/16

$$E = 556 \times 9/16 = 312.75$$

$$E = 556 \times 3/16 = 104.25$$

$$E = 556 \times 3/16 = 104.25$$

$$E = 556 \times 1/16 = 34.75$$

The χ^2-Goodness-of-fit test example

In a famous experiment, Mendel crossed 556 smooth, yellow male peas with wrinkled green female peas.

Is there any evidence that Mendel just made up his data?

Mendel's data:

Type	Observed count (O)	Expected count (E)	$\frac{(O_i - E_i)^2}{E_i}$
Smooth yellow	315	312.75	0.016
Smooth green	108	104.25	0.135
Wrinkled yellow	102	104.25	0.049
Wrinkled green	31	34.75	0.405

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

$$= 0.604$$

$$X^2 \sim \chi^2_{n-c-1} = \chi^2_{4-0-1} = \chi^2_3$$

$$p = P(X^2 \geq 0.604)$$
$$= 0.895$$

Mendel's data is consistent with our modern genetic distributions!

The χ^2-Goodness-of-fit test example

See chisq_goodness_of_fit.R

Two-sample test for proportions Using a normal approximation (Revisited)

Two-sample tests of proportions

Remember the two-sample test of proportions?

$$X_1, \ldots, X_n \sim Bernoulli(p_1),$$
$$Y_1, \ldots, Y_m \sim Bernoulli(p_2)$$

$$H_0 : p_1 = p_2 \qquad H_1 : p_1 \neq p_2$$

$$\boxed{z = \frac{\text{estimate} - \text{null value}}{SD_{H_0}(\text{estimate})} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \overset{\text{Under } H_0}{\sim} N(0,1)}$$

(Estimated from combined dataset)

Two-sample tests of proportions: example

You want to test whether the proportion of right-handedness is significantly different between Australians and Americans

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

	Right-handed	**Left-handed**	**Total**
Australians	43	9	52
Americans	44	4	48
Total	87	13	100

Using entire data, we estimate

$$\hat{p} = 87/100 = 0.87$$

$$\hat{p}_1 = \frac{43}{52} = 0.83 \qquad \hat{p}_2 = \frac{44}{48} = 0.92$$

Two-sample tests of proportions: example

	Right-	Left-	Total
Australian	43	9	52
Americans	44	4	48
Total	87	13	100

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

Australians	Americans
$n = 52$	$m = 48$
$\hat{p}_1 = \frac{43}{52} = 0.83$	$\hat{p}_2 = \frac{44}{48} = 0.92$

$$\hat{p} = 0.87$$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.83 - 0.92}{\sqrt{0.87 \times 0.13 \left(\frac{1}{52} + \frac{1}{48}\right)}} = -1.33$$

P-value $= 2(1 - \Phi(1.33)) = 0.18$

Reformulating the hypothesis test as independence

	Right-handed	Left-handed	Total
Australians	43	9	52
Americans	44	4	48
Total	87	13	100

You want to test whether the proportion of right-handedness (p) is significantly different between Australians and Americans

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

Reformulating the hypothesis test as independence

	Right-handed	Left-handed	Total
Australians	43	9	52
Americans	44	4	48
Total	87	13	100

Instead, let's test whether the proportion of people suffering from Australianism (as opposed to Americanism) is significantly different between left-handed people and right-handed people

$$H_0 : \boxed{p_1} = \boxed{p_2}$$
$$H_1 : \boxed{p_1} \neq \boxed{p_2}$$

We should get the same p-value

Two-sample tests of proportions: example

	Right-	Left-	Total
Australian	43	9	52
Americans	44	4	48
Total	87	13	100

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

Right-handed	Left-handed
$n = 87$	$m = 13$
$\hat{p}_1 = \frac{43}{87} = 0.49$	$\hat{p}_2 = \frac{9}{13} = 0.69$

$$\hat{p} = 52/100 = 0.52$$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.49 - 0.69}{\sqrt{0.52 \times 0.48 \left(\frac{1}{87} + \frac{1}{13}\right)}} = -1.33$$

P-value $= 2(1 - \Phi(1.33)) = 0.18$

Equivalent hypothesis tests

The test was originally formatted as:

H_0 : The proportion of **right-handedness** is *the same* **across Australians and Americans**

H_1 : The proportion of **right-handedness** is *different* **across Australians and Americans**

Equivalent hypothesis tests

And we showed that this was equivalent to:

H_0 : The proportion of people who are **Australian** is *the same* **across left-handed and right-handed**

H_1 : The proportion of people who are **Australian** is *different* **across left-handed and right-handed people**

Reformulating the hypothesis test as independence

These tests can be reformatted as a test for independence of two categorical variables:

H_0 : Australianism/Americanism and hand-dominance are **independent**

H_1 : Australianism/Americanism and hand-dominance are not **independent**

Fisher’s exact test

(an <u>exact</u> two sample test for proportions)

Fisher's exact test

	Right-handed	Left-handed	Total
Australians	43	9	52
Americans	44	4	48
Total	87	13	100

H_0 : Australianism/Americanism and hand-dominance are **independent**

H_1 : Australianism/Americanism and hand-dominance are **not independent**

We can use Fisher's exact test to compute an **exact** p-value

Fisher's exact test

	A_1	A_2	Total
B_1	$\boxed{N_{11}}$	N_{12}	$n_{1\cdot}$
B_2	N_{21}	N_{22}	$n_{2\cdot}$
Totals	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

Under H_0: $p_1 = p_2$ and $N_{11} \sim$ Hypergeometric$(n_{\cdot\cdot}, n_{1\cdot}, n_{\cdot 1})$

If $X \sim Hypergeometric(N, K, n)$ then $P(X = k)$ corresponds to:

Prob of drawing *k* objects with a particular feature in *n* draws (w/o replacement) from a finite population of size *N* that contains exactly *K* objects with that feature

Fisher's exact test

	A_1	A_2	Total
B_1	$\boxed{N_{11}}$	N_{12}	$n_{1\cdot}$
B_2	N_{21}	N_{22}	$n_{2\cdot}$
Totals	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

Under H_0: $p_1 = p_2$,
$N_{11} \sim$ Hypergeometric$(n_{\cdot\cdot}, n_{1\cdot}, n_{\cdot 1})$

So

$$P(N_{11} = n_{11}) = \frac{\binom{n_{1\cdot}}{n_{11}}\binom{n_{2\cdot}}{n_{21}}}{\binom{n_{\cdot\cdot}}{n_{\cdot 1}}}$$

$$E(N_{11}) = \frac{n_{\cdot 1} n_{1\cdot}}{n_{\cdot\cdot}}$$

Let N_{11} be our test statistic. Reject H_0 for extreme values of N_{11}

Our p-value for the test of independence is

$$P\left(\,|N_{11} - E(N_{11})| \geq |n_{11} - E(N_{11})|\,\right)$$

Fisher's exact test

	Right-handed	Left-handed	Total
Australians	43	9	52
Americans	44	4	48
Total	87	13	100

$$N_{11} \sim Hypergeoetric(100, 87, 52)$$

$$n_{11} = 43 \qquad E(N_{11}) = \frac{n_{.1}n_{1.}}{n_{..}} = \frac{52 \times 87}{100} = 45.24$$

So our p-value is

$$\begin{aligned} P(|N_{11} - E(N_{11})| \geq |n_{11} - E(N_{11})|) &= P(|N_{11} - 45.24| \geq 2.24) \\ &= P(N_{11} - 45.24 \geq 2.24) + P(N_{11} - 45.24 \leq -2.24) \\ &= P(N_{11} \geq 47.48) + P(N_{11} \leq 43) \quad = 0.24 \end{aligned}$$

Fisher's exact test example

See *fisher_exact_test.R*

Fisher’s exact test example

- A psychological experiment was done to investigate the effect of anxiety on a person’s desire to be alone or in company.
- A group of 30 subjects was randomly divided into two groups of sizes 13 and 17.
- The subjects were told that they would be subject to electric shocks.
  - The “high anxiety” group was told that the shocks would be quite painful
  - The “low-anxiety” group was told that they would be mild and painless.
- Both groups were told that there would be a 10-min wait before the experiment began and each subject was given the choice of waiting alone or with other subjects.

Fisher's exact test example

	Wait Together	Wait Alone	Total
High-Anxiety	12	5	17
Low-Anxiety	4	9	13
Total	16	14	30

Test whether there is a significant difference between the high- and low-anxiety groups.

Fisher's exact test example

	Wait Together	Wait Alone	Total
High-Anxiety	12	5	17
Low-Anxiety	4	9	13
Total	16	14	30

If $X \sim Hypergeometric(N, K, n)$ then $P(X = k)$ corresponds to:

Prob of drawing *k* objects with a particular feature in *n* draws (w/o replacement) from a finite population of size *N* that contains exactly *K* objects with that feature

$k = n_{11}$ is the number of subjects who chose to "wait together" from the
n = 17 subjects in the high-anxiety group
which are a part of a population of total sample size N = 30,
that contains a total of K = 16 "wait together" subjects

So $N_{11} \sim Hypergeometric(30,16,17)$

Fisher's exact test example

	Wait Together	Wait Alone	Total
High-Anxiety	12	5	17
Low-Anxiety	4	9	13
Total	16	14	30

$N_{11} \sim Hypergeometric(30,16,17)$ $(Hypergeometric(N, K, n))$

If anxiety had no effect on choice, we would expect to see:

$$EN_{11} = n\frac{K}{N} = \frac{n_{\cdot 1} n_{1\cdot}}{n_{\cdot\cdot}} = \frac{16 \times 17}{30} = 9.07$$

High-anxiety people choosing to wait together

Fisher's exact test example

	Wait Together	Wait Alone	Total
High-Anxiety	12	5	17
Low-Anxiety	4	9	13
Total	16	14	30

Our p-value is thus

$$P(|N_{11} - 9.07| \geq |12 - 9.07|)$$

$$= P(N_{11} - 9.07 \geq 2.93) + P(N_{11} - 9.07 \leq -2.93)$$

$$= P(N_{11} \geq 12) + P(N_{11} \leq 6.14)$$

$$= 0.063$$

So we don't *quite* have enough evidence to reject the null

Fisher’s exact test example

See *fisher_exact_test.R*

Fisher’s exact test example

	Wait Together	Wait Alone	Total
High-Anxiety	12	5	17
Low-Anxiety	4	9	13
Total	16	14	30

Fisher’s exact test and the two-sample test are appropriate for proportions are for **2x2 contingency tables**

Pearson χ^2 test for homogeneity

The χ^2 test for homogeneity

Suppose now that instead of a 2×2 table of data, we have an arbitrary $I \times J$ table, and we want to see if the proportions of each category differ across the different groups.

When Jane Austin died, she left the novel Sanditon unfinished.

An impersonator finished the novel, attempting to emulate Austin's style.

Are the word distributions the same across the following books?

Word	Sense & Sensibility	Emma	Sanditon I (Austin)	Sanditon II (Impersonator)
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4
Total	375	440	202	196

The χ^2 test for homogeneity

The six word counts for each book can be modeled as a multinomial random variable with unknown cell probabilities and the total given in the “Total” row.

Word	Sense & Sensibility	Emma	Sanditon I (Austin)	Sanditon II (Impersonator)
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4
Total	375	440	202	196

$$Multinomial(375, \pi_{a,1}, \pi_{an,1}, \pi_{this,1}, \pi_{that,1}, \pi_{with,1}, \pi_{without,1})$$

(The multinomial dist is like the binomial dist but with I different categories.)

$$H_0 : \pi_{i1} = \pi_{i2} = \ldots = \pi_{iJ}, \qquad i = 1,\ldots,I$$

(i.e., within each row, the proportions are equal)

The χ^2 test for homogeneity

The six word counts for each book can be modeled as a multinomial random variable with unknown cell probabilities and the total given in the “Total” row.

$$Multinomial(375, \pi_{a,sense}, \pi_{an,sense}, \pi_{this,sense}, \pi_{that,sense}, \pi_{with,sense}, \pi_{without,sense})$$

(The multinomial dist is like the binomial dist but with I different categories.)

$$H_0 : \pi_{i1} = \pi_{i2} = \ldots = \pi_{iJ}, \qquad i = 1,\ldots,I$$

(i.e., within each row, the proportions are equal)

Word	Sense & Sensibility	Emma	Sanditon I (Austin)	Sanditon II (Impersonator)
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4
Total	375	440	202	196

The χ^2 test for homogeneity

$$H_0 : \pi_{i1} = \pi_{i2} = \ldots = \pi_{iJ}, \qquad i = 1,\ldots,I$$

For the Jane Austin data, the null is that **each** of the following are true:

$$H_0 : \pi_{a,sense} = \pi_{a,emma} = \pi_{a,sand1} = \pi_{a,sand2}$$

$$\pi_{an,sense} = \pi_{an,emma} = \pi_{an,sand1} = \pi_{an,sand2}$$

$$\pi_{this,sense} = \pi_{this,emma} = \pi_{this,sand1} = \pi_{this,sand2}$$

$$\pi_{that,sense} = \pi_{that,emma} = \pi_{that,sand1} = \pi_{that,sand2}$$

$$\pi_{with,sense} = \pi_{with,emma} = \pi_{with,sand1} = \pi_{with,sand2}$$

$$\pi_{without,sense} = \pi_{without,emma} = \pi_{without,sand1} = \pi_{without,sand2}$$

H_1 : at least one proportion is not equal to the others for at least one row

The χ^2 test for homogeneity

$$H_0 : \pi_{i1} = \pi_{i2} = \ldots = \pi_{iJ}, \qquad i = 1,\ldots,I$$

(i.e., within every row, the proportions are equal)

H_1 : at least one proportion is not equal to the others for at least one row

We can view this as a **goodness of fit test**

Does the model prescribed by the null hypothesis fit the data?

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot})^2}{n_{i\cdot}n_{j\cdot}/n_{\cdot\cdot}}$$

Total for jth category (book)

Total for ith observation (word)

$$E_{ij} = \frac{n_{\cdot j}n_{i\cdot}}{n_{\cdot\cdot}}$$

Total for the entire table

The χ^2 test for homogeneity

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot})^2}{n_{i\cdot}n_{j\cdot}/n_{\cdot\cdot}}$$

(Under H_0)

$$X^2 \sim \chi^2_{(I-1)(J-1)}$$

$J = 4$

$I = 6$

Word	Sense & Sensibility	Emma	Sanditon I (Austin)	Sanditon II (Impersonator)
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4
Total	375	440	202	196

df = # independent counts - # of independent parameters

$= \quad J(I-1) \qquad - \qquad (I-1)$

(Since each of the J multinomials has $I-1$ independent counts)

(Since the totals from each multinomial is fixed)

The χ^2 test for homogeneity summary

Test statistic:

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot})^2}{n_{i\cdot}n_{j\cdot}/n_{\cdot\cdot}}$$

(Under H_0)

$$X^2 \sim \chi^2_{(I-1)(J-1)}$$

P-value = $P(X^2 \geq x^2)$, where $X^2 \sim \chi^2_{(I-1)(J-1)}$

The χ^2 test for homogeneity example

Let's fill in the table with the expected counts

$$(E_{ij}) = \frac{n_{\cdot j} n_{i \cdot}}{n_{\cdot \cdot}}$$

$$(160) = \frac{517 \times 375}{1213}$$

$$(22.9) = \frac{91 \times 375}{1213}$$

Word	Sense & Sensibility	Emma	Sanditon I (Austin)	Sanditon II (Impersonator)	Total
a	147 (160)	186 (187.8)	101 (86.2)	83 (83.5)	517
an	25 (22.9)	26 (26.8)	11 (12.3)	29 (14.7)	91
this	32 (31.7)	39 (37.2)	15 (17.1)	15 (16.3)	101
that	94 (87)	105 (102.1)	37 (46.9)	22 (41.6)	258
with	59 (59.4)	74 (69.7)	28 (32)	43 (33)	204
without	18 (14)	10 (16.4)	10 (7.5)	4 (6.8)	42
Total	375	440	202	196	1,213

The χ^2 test for homogeneity example

Let's fill in the table with the expected counts

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-n_{i}.n_{.j}/n_{..})^2}{n_{i}.n_{j}./n_{..}}$$

$$= \frac{(147-160)^2}{160} + \frac{(25-22.9)^2}{22.9} + \ldots$$

$$= 45.58$$

Word	Sense & Sensibility	Emma	Sanditon I (Austin)	Sanditon II (Impersonator)
a	147 **(160)**	186 **(187.8)**	101 **(86.2)**	83 **(83.5)**
an	25 **(22.9)**	26 **(26.8)**	11 **(12.3)**	29 **(14.7)**
this	32 **(31.7)**	39 **(37.2)**	15 **(17.1)**	15 **(16.3)**
that	94 **(87)**	105 **(102.1)**	37 **(46.9)**	22 **(41.6)**
with	59 **(59.4)**	74 **(69.7)**	28 **(32)**	43 **(33)**
without	18 **(14)**	10 **(16.4)**	10 **(7.5)**	4 **(6.8)**

P-value = $P(X^2 \geq 45.58) = 0.00006,$ where $X^2 \sim \chi^2_{(I-1)(J-1)} = \chi^2_{15}$

The χ^2 test for homogeneity

See chisq_homogeneity.R

Pearson χ^2 test for independence

(This is essentially **exactly the same** as the test for homogeneity)

The χ^2 test for independence

The χ^2 test for independence is eerily similar to (read: *exactly the same as*) the χ^2 test for homogeneity, but is aimed at answering a slightly different question

H_0 : the row variable and column variable are independent

H_1 : the row variable and column variable are ***not*** independent

We can view this as a **goodness of fit test**

Do the counts match what you would expect if the cols/rows were indep?

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot})^2}{n_{i\cdot}n_{j\cdot}/n_{\cdot\cdot}}$$

$$E_{ij} = \frac{n_{\cdot j}n_{i\cdot}}{n_{\cdot\cdot}}$$

The χ^2 test for independence summary

Test statistic:

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - n_{i.}n_{.j}/n_{..})^2}{n_{i.}n_{j.}/n_{..}}$$

(Under H_0)

$$X^2 \sim \chi^2_{(I-1)(J-1)}$$

P-value = $P(X^2 \geq x^2)$, where $X^2 \sim \chi^2_{(I-1)(J-1)}$

The χ^2 test for independence example

Suppose that a psychologist wants to test whether there is a relationship between personality and color preference.

H_0 : color preference and personality are independent

Let's fill in the expected counts

Total for jth category

Total for ith observation

$$E_{ij} = \frac{n_{\cdot j} n_{i \cdot}}{n_{\cdot\cdot}}$$

Total for the entire table

	Blue	Red	Yellow	Total
Extroverted	5 (**9**)	20 (**15**)	5 (**6**)	30
Introverted	10 (**6**)	5 (**10**)	5 (**4**)	20
Total	15	25	10	50

$$(9) = \frac{30 \times 15}{50}$$

$$(6) = \frac{20 \times 15}{50}$$

The χ^2 test for independence example

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-n_{i.}n_{.j}/n_{..})^2}{n_{i.}n_{j.}/n_{..}}$$

$I = 2$

$J = 3$

	Blue	Red	Yellow	Total
Extroverted	5 **(9)**	20 **(15)**	5 **(6)**	30
Introverted	10 **(6)**	5 **(10)**	5 **(4)**	20
Total	15	25	10	50

$$= \frac{(5-9)^2}{9} + \frac{(10-6)^2}{6} + \ldots$$

$$= 9.02$$

P-value = $P(X^2 \geq 9.02) = 0.011,$ where $X^2 \sim \chi^2_{(I-1)(J-1)} = \chi^2_2$

The χ^2 test for independence

See <u>chisq_indep.R</u>

Comparison of all the χ^2 tests

The χ^2 goodness-of-fit test

H_0 : the data comes from the specified distribution

Value (n total)	Count
a	56
b	27
c	3
d	15

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-c-1}$$

P-value = $P(X^2 \geq x^2)$

The χ^2 test for homogeneity

$$H_0 : \pi_{i1} = \pi_{i2} = \ldots = \pi_{iJ}, \qquad i = 1,\ldots,I$$

Group (J total)

Value (I total)

	A	B	C	D
a	56	104	80	62
b	12	1	15	12
c	32	32	45	15

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(I-1)(J-1)}$$

P-value = $P(X^2 \geq x^2)$

The χ^2 test for independence

H_0 : the row variable and column variable are independent

Value (I total)	Group (J total)			
	A	B	C	D
a	56	104	80	62
b	12	1	15	12
c	32	32	45	15

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(I-1)(J-1)}$$

P-value = $P(X^2 \geq x^2)$