Lecture 22: Chi-Square Tests for Goodness-of-Fit

Chapter 6.3

Question for Today

Say we had $n = 100$ people picked as jurors, we expect the breakdown to be:

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	72	7	12	9	$n = 100$

Question for Today

Say we observe the following. Is there a bias? i.e. a non-random mechanism?

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	75	6	11	8	$n = 100$

Chi-Square Tests

Chi-square χ^2 tests allow us to compare

- Observed counts
- Expected counts

Chi-Square Tests

Chi-square χ^2 tests allow us to compare

- Observed counts
- Expected counts

i.e. What is the "goodness" of the fit of the observed counts to the expected counts?

The Data

Let's use $n = 275$ people. Assuming the same proportions as above, we compute the expected counts. Ex: $198 = 275 \times 0.72$.

Race	White	Black	Hispanic	Other	Total
Expected Counts	198	19.25	33	24.75	275

The Data

Let's use $n = 275$ people. Assuming the same proportions as above, we compute the expected counts. Ex: $198 = 275 \times 0.72$.

Race	White	Black	Hispanic	Other	Total
Expected Counts	198	19.25	33	24.75	275
Observed Counts	205	26	25	19	275

Hypothesis Test in General

Hypothesis Test in Our Case

Null Distributions

To compute p-values we compare a test statistic to a null distribution: the distribution of the test statistic under H_0.

Null Distributions

To compute p-values we compare a test statistic to a null distribution: the distribution of the test statistic under H_0.

Ex: For tests on means/proportions

- test statistic: z-score of $\overline{x}/\widehat{p}$
- null distribution: standard normal distribution

Null Distributions

To compute p-values we compare a test statistic to a null distribution: the distribution of the test statistic under H_0.

Ex: For tests on means/proportions

- test statistic: z-score of $\overline{x}/\widehat{p}$
- null distribution: standard normal distribution

Now

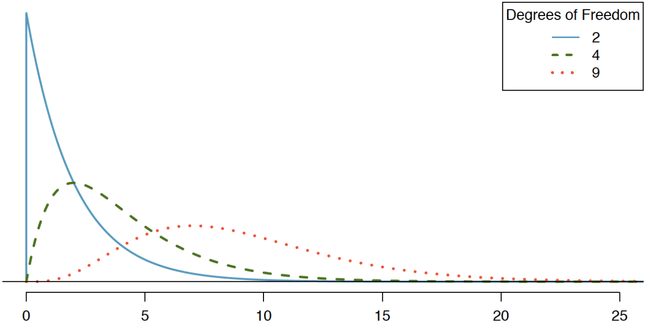
- test statistic: χ^2-statistic
- null distribution: χ^2 distribution with $df = k - 1$

Deviations

Chi-Square Test Statistic

p-values

We compare the test statistic to a χ^2 distribution with $df = k - 1$ degrees of freedom.
Note: not $df = n - 1$ like with t-test.



p-values

The p-value is the area to the right of the test statistic. Use p.432:

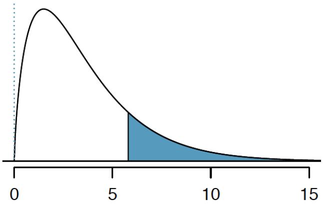


Figure B.2: Areas in the chi-square table always refer to the right tail.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52

p-values

The p-value is the area to the right of the test statistic. Use p.432:

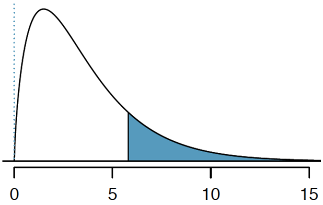


Figure B.2: Areas in the chi-square table always refer to the right tail.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52

In our case, $df = k - 1 = 3$, and $\chi^2 = 5.89$, which is in between $(4.64, 6.25)$, so p-value is in between $(0.1, 0.2)$. Not overwhelming evidence against H_0.

Hypothetical Scenarios

Using the same expected counts as earlier...

Race	White	Black	Hispanic	Other	Total
Expected Counts	198	19.25	33	24.75	275
Observed Counts					275

say we have two sets of hypothetical observed counts...

Assumptions for Chi-Square Test

Next Time

We look at chi-square tests for two-way tables to test for independence. i.e. are two variables independent from each other?