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

Chapter 6.3

Question for Today

Say we have a population where the racial breakdown of the juror pool (registered voters) is:

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%

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 breakdown. Fairly obvious bias in juror selection!

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	0	0	100	0	$n = 100$

Question for Today

But what about the following? We expected 72 whites, but observe 75. 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

- Expected frequencies
- Observed frequencies

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 earlier to compute the expected counts, say we then observe:

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

Hypothesis Test

H_0 : the jurors are a random sample
i.e. there is no racial bias in who serves on a jury and the observed counts reflect natural sampling fluctuation

vs H_A : the jurors are not randomly sampled
i.e. there is racial bias in juror selection

Hypothesis Test:

We compute a test statistic and use a null distribution (the distribution of the test statistic if H_0 is true) to compute p-values:

1. means/proportions:
   - test statistic: z-score of $\overline{x}/\widehat{p}$
   - null distribution: normal distribution (z-table)
2. t-test:
   - test statistic: t-statistic
   - null distribution: t-distribution with $df = n - 1$ (t-table)
3. AVOVA:
   - test statistic: F-statistic
   - null distribution: F-distribution with $df_1 = k - 1$ and $df_2 = n - k$ (F-table)
4. Now: Goodness-of-fit:
   - test statistic: χ^2-statistic
   - null distribution: χ^2 distribution with $df = k - 1$

Test Statistic

For previous tests, we constructed a test statistic of the following form:

$$\frac{\text{point estimate} - \text{null value}}{\text{SE of point estimate}}$$

In our case, it's similar. For each of the k groups (in this case racial group) compute

$$Z = \frac{\text{observed count} - \text{expected count}}{\sqrt{\text{expected count}}}$$

Test Statistic

$$Z = \frac{\text{observed count} - \text{expected count}}{\sqrt{\text{expected count}}}$$

So when

- observed $=$ expected $\Rightarrow Z = 0$
- observed $>$ expected $\Rightarrow Z > 0$
- observed $<$ expected $\Rightarrow Z < 0$

Test Statistic

Now treat $+$'ve and $-$'ve differences as the same:

$$\begin{aligned} Z^2 &= \left(\frac{\text{observed count} - \text{expected count}}{\sqrt{\text{expected count}}}\right)^2 \\ &= \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} \end{aligned}$$

Why square it and not absolute value it? It's easier to do calculus on x^2 than $|x|$.

Test Statistic

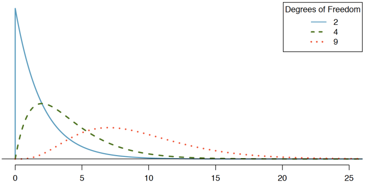
In the case of our trial data, we have 4 groups: white, black, hispanic, and other:

$$\begin{aligned}\chi^2 &= Z_w^2 + Z_b^2 + Z_h^2 + Z_o^2 \\ &= \frac{(205-198)^2}{198} + \frac{(26-19.25)^2}{19.25} + \frac{(25-33)^2}{33} + \frac{(19-24.75)^2}{24.75} \\ &= 5.89\end{aligned}$$

This is our test statistic.

p-values

To compute the p-value, 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-value

Use table on page 412:

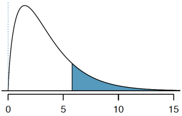


Figure B.2: Areas in the chi-square table always refer to the right tail.

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

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

Say:

- For all 4 groups, we observed = expected. Then

$$Z_i^2 = \frac{(\text{observed count}_i - \text{expected count}_i)^2}{\text{expected count}_i} = 0$$

for $i = 1, \ldots, 4$, and $\chi^2 = 0$. p-value $= 1$

- Say we observed 0 whites, blacks, hispanics and 275 others, then $\chi^2 = 15648.25$. p-value=0

Chi-Square Test for One-Way Tables

This is also called a chi-square test for one-way tables.

$$\chi^2 = \sum_{i=1}^{k} Z_i^2 = \sum_{i=1}^{k} \frac{(\text{observed count}_i - \text{expected count}_i)^2}{\text{expected count}_i}$$

Assumptions for Chi-Square Test

1. Independence: Each case is independent of the other
2. Sample size/distribution: Similarly like with proportions, we need at least 5 cases in each scenario (each cell in the table)
3. Degrees of freedom: We need at least $df = 2$, i.e. $k \geq 3$

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?