

# 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  $\chi^2$  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

$H_0$  : The data are consistent with the specified distribution.

vs  $H_A$  : The data are not consistent with the specified distribution.

$H_0$  can also be stated: the data are a random sample from the distribution and any differences of observed vs expected reflect natural sampling variation.

# Hypothesis Test in Our Case

$H_0$  : the jurors are randomly sampled i.e. there is no racial bias

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



# Null Distributions

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

## 1. means/proportions:

- ▶ test statistic: z-score of  $\bar{x}/\hat{p}$
- ▶ null distribution: normal distribution

## 2. t-test:

- ▶ test statistic:  $t$ -statistic
- ▶ null distribution:  $t$ -distribution with  $df = n - 1$

## 3. ANOVA:

- ▶ test statistic:  $F$ -statistic
- ▶ null distribution:  $F$ -distribution with  $df_1 = k - 1$  and  $df_2 = n - k$

## 4. Goodness-of-fit:

- ▶ test statistic:  $\chi^2$ -statistic
- ▶ null distribution:  $\chi^2$  distribution with  $df = k - 1$

# Deviations

Previously, many test statistics had the following form:

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

For goodness-of-fit, it's similar. For each of the  $k$  groups compute

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

Note when:

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

The  $Z$ 's measure deviations.

# Deviations

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

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

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

# Chi-Square Test Statistic

Finally sum all values of  $Z^2$ . This is the **chi-square test statistic for one-way tables**.

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

## Chi-Square Test Statistic

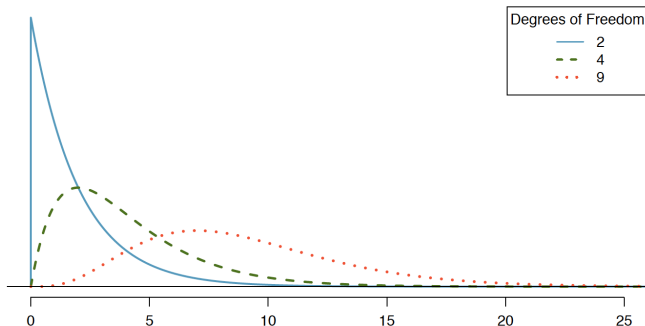
In the case of the jury data, we have 4 groups: white, black, hispanic, and other:

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

## p-values

We compare the test statistic to a  $\chi^2$  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.412:

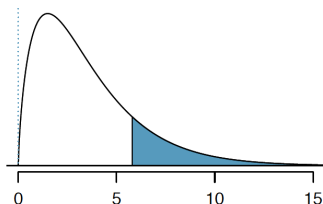


Figure B.2: Areas in the chi-square table always refer to the right tail.

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

Say we have two hypothetical scenarios of observed counts:

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

- For all 4 groups, say observed = expected, then

$$\chi^2 = 0 + 0 + 0 + 0 = 0$$

hence p-value = 1.

- Say we observed 275 others and 0 for the rest, then

$$\chi^2 = 2786.11 + 84.46 + 189.57 + 12588.11 = 15648.25$$

hence p-value = 0.



# Assumptions for Chi-Square Test

1. **Independence:** Each case is independent of the other
2. **Sample size:** 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?