Tests of Homogeneity and Independence

Lecture 52
Sections 14.5

Robb T. Koether

Hampden-Sydney College

Mon, Apr 30, 2012

Outline

Outline

1. Degrees of Freedom
2. Goodness-of-Fit Test vs. One-Sample Proportion Test
3. χ^2 Test vs. Two-Sample Proportion Test
4. Assignment

Degrees of Freedom

- Consider the political affiliation data data.

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20	10	10	40
Moderate	20	30	30	80
Col. Total	50	90	60	200

- What if we had only the row and column totals?
- How many values could we fill in arbitrarily?

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative				80
Liberal				40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10			80
Liberal				40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50		80
Liberal				40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal				40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20			40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20	10		40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20	10	10	40
Moderate				80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20	10	10	40
Moderate	20			80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20	10	10	40
Moderate	20	30		80
Col. Total	50	90	60	200

Degrees of Freedom

	Approve	Disapprove	Neutral	Row Total
Conservative	10	50	20	80
Liberal	20	10	10	40
Moderate	20	30	30	80
Col. Total	50	90	60	200

Outline

Goodness-of-Fit Test vs. One-Sample Proportion Test

- Recall the example of tossing the coin.
- This problem could be recast as a goodness-of-fit test with 2 categories (heads, tails).
- Let p_1 be the proportion of heads and p_2 the proportion of tails.
- Then we have

$$\begin{aligned} H_0 &: \quad p_1 = 0.50, p_2 = 0.50 \\ H_1 &: \quad H_0 \text{ is false.} \end{aligned}$$

Goodness-of-Fit Test vs. One-Sample Proportion Test

- Our sample contained 1000 coin tosses, with 525 heads and 475 tails.
- Our observed and expected counts are

	Male	Female
Observed	525	475
Expected	(500)	(500)

Goodness-of-Fit Test vs. One-Sample Proportion Test

- Compute χ^2.

$$\begin{aligned}\chi^2 &= \frac{(525-500)^2}{500} + \frac{(475-500)^2}{500} \\ &= \frac{625}{500} + \frac{625}{500} \\ &= 2.5.\end{aligned}$$

- p-value $= \chi^2\texttt{cdf(2.5,E99,1)} = 0.1138$.

Goodness-of-Fit Test vs. One-Sample Proportion Test

- Re-work this example as a one-sample proportion problem, as we did before.

Goodness-of-Fit Test vs. One-Sample Proportion Test

- Re-work this example as a one-sample proportion problem, as we did before.
- We get $z = 1.581$ and p-value $= 0.1138$.

Goodness-of-Fit Test vs. One-Sample Proportion Test

- Re-work this example as a one-sample proportion problem, as we did before.
- We get $z = 1.581$ and p-value $= 0.1138$.
- Furthermore, $z^2 = 2.5$.

Outline

χ^2 Test vs. Two-Sample Proportion Test

- We worked a two-sample proportion problem a while back.
- 3500 likely voters on Feb 6, 34% approved of the president's preformance.
- 3500 likely voters on Mar 26, 29% approved of the president's preformance.
- That is, 1190 out of 3500 approved on Feb 6 and 1015 out of 3500 approved on Mar 26.

χ^2 Test vs. Two-Sample Proportion Test

- We can recast this as a test of homogeneity
- Are the two populations (likely voters on Feb 6 and likely voters on Mar 26) homogeneous?

	Approve	Disapprove	Row Total
Feb 6	1190	2310	
Mar 26	1015	2485	
Col Tot			

χ^2 Test vs. Two-Sample Proportion Test

- We can recast this as a test of homogeneity
- Are the two populations (likely voters on Feb 6 and likely voters on Mar 26) homogeneous?

	Approve	Disapprove	Row Total
Feb 6	1190	2310	3500
Mar 26	1015	2485	3500
Col Tot			

χ^2 Test vs. Two-Sample Proportion Test

- We can recast this as a test of homogeneity
- Are the two populations (likely voters on Feb 6 and likely voters on Mar 26) homogeneous?

	Approve	Disapprove	Row Total
Feb 6	1190	2310	3500
Mar 26	1015	2485	3500
Col Tot	2205	4795	

χ^2 Test vs. Two-Sample Proportion Test

- We can recast this as a test of homogeneity
- Are the two populations (likely voters on Feb 6 and likely voters on Mar 26) homogeneous?

	Approve	Disapprove	Row Total
Feb 6	1190	2310	3500
Mar 26	1015	2485	3500
Col Tot	2205	4795	7000

χ^2 Test vs. Two-Sample Proportion Test

- We can recast this as a test of homogeneity
- Are the two populations (likely voters on Feb 6 and likely voters on Mar 26) homogeneous?

	Approve	Disapprove	Row Total
Feb 6	1190 (1102.5)	2310 (2397.5)	3500
Mar 26	1015 (1102.5)	2485 (2397.5)	3500
Col Tot	2205	4795	7000

χ^2 Test vs. Two-Sample Proportion Test

- Compute χ^2.

$$\chi^2 = \frac{87.5^2}{1102.5} + \frac{87.5^2}{2397.5} + \frac{87.5^2}{1102.5} + \frac{87.5^2}{2397.5}$$
$$= 6.9444 + 3.1934 + 6.9444 + 3.1934$$
$$= 20.2756.$$

- p-value $= \chi^2\texttt{cdf(20.2756,E99,1)} = 6.705 \times 10^{-6}$.

χ^2 Test vs. Two-Sample Proportion Test

- Re-work this example as a two-sample proportion problem, as we did before.

χ^2 Test vs. Two-Sample Proportion Test

- Re-work this example as a two-sample proportion problem, as we did before.
- We get $z = 4.5029$ and p-value $= 6.711 \times 10^{-6}$.

χ^2 Test vs. Two-Sample Proportion Test

- Re-work this example as a two-sample proportion problem, as we did before.
- We get $z = 4.5029$ and p-value $= 6.711 \times 10^{-6}$.
- Furthermore, $z^2 = 20.276$.

χ^2 Test vs. Two-Sample Proportion Test

- Re-work this example as a two-sample proportion problem, as we did before.
- We get $z = 4.5029$ and p-value $= 6.711 \times 10^{-6}$.
- Furthermore, $z^2 = 20.276$.
- χ^2 with 1 degree of freedom is z^2.

Outline

Assignment

Homework

- Read Section 14.5, pages 951 - 959.
- Let's Do It! 14.6, 14.7. (14.6 is an example of Simpson's paradox.)
- Exercises 23 - 27, 31, 34, page 959. (Ex. 31 and 34 are examples of Simpson's paradox.)
- Chapter Review 35 - 40, 42, 47, page 966.