Introduction to Statistics and Data Science using *eStat*

Chapter 11 Testing Hypothesis for Categorical Data

11.1.1 Goodness of Fit Test for Categorical Data

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

11.1 Goodness of Fit Test

11.1.1 Goodness of Fit Test for Categorical Data
11.1.2 Goodness of Fit Test for Continuous Data

11.2 Testing Hypothesis for Contingency Table

11.2.1 Independence Test
11.2.2 Homogeneity Test

11.1 Goodness of Fit Test

11.1.1 Goodness of Fit Test for Categorical Data

[Example 11.1.1] The result of a survey of 150 people before a local election to find out the approval ratings of three candidates is as follows.

- Looking at this frequency table alone, it seems that A candidate has a 40 percent approval rating, higher than the other candidates.
- Based on this sample survey, perform the goodness of fit test whether three candidates have the same approval rating or not. Use『eStatU』with the 5% significance level.

Candidate	Number of Supporters	Percent
A	60	40.0%
B	50	33.3%
C	40	25.7%
Total	150	100%

11.1 Goodness of Fit Test

<Answer of Example 11.1.1>

- **Hypothesis**

H_0 : Three candidates have the same approval rating. $(p_1 = p_2 = p_3 = \frac{1}{3})$

H_1: Three candidates have different approval rating.

- **Observed and Expected Frequency**

Candidate	Observed frequency (denoted as O_i)	Expected frequency (denoted as E_i)
A B C	O_1 = 60 O_2 = 50 O_3 = 40	E_1 = 50 E_2 = 50 E_3 = 50
Total	150	150

11.1 Goodness of Fit Test

<Answer of Example 11.1.1>

- **Test Statistic**

$$\chi^2_{obs} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3}$$
$$= \frac{(60 - 50)^2}{50} + \frac{(50 - 50)^2}{50} + \frac{(40 - 50)^2}{50} = 4$$

- **Decision Rule**

‘If $\chi^2_{obs} > \chi^2_{k-1;\,\alpha}$, reject H_0‘

Since $\chi^2_{3-1;\ 0.05} = 5.991$, H_0 cannot be rejected.

11.1 Goodness of Fit Test

<Answer of Example 11.1.1>

- **Confidence Interval**

A: $0.40 \pm 1.96\sqrt{\frac{0.40(1-0.40)}{150}} \Leftrightarrow (0.322, 0.478)$

B: $0.33 \pm 1.96\sqrt{\frac{0.33(1-0.33)}{150}} \Leftrightarrow (0.255, 0.405)$

C: $0.27 \pm 1.96\sqrt{\frac{0.27(1-0.27)}{150}} \Leftrightarrow (0.190, 0.330)$

11.1 Goodness of Fit Test

<Answer of Example 11.1.1>

Goodness of Fit Test

Menu

[Hypothesis] H_0 : *Observed & theoretical Distributions are the same*

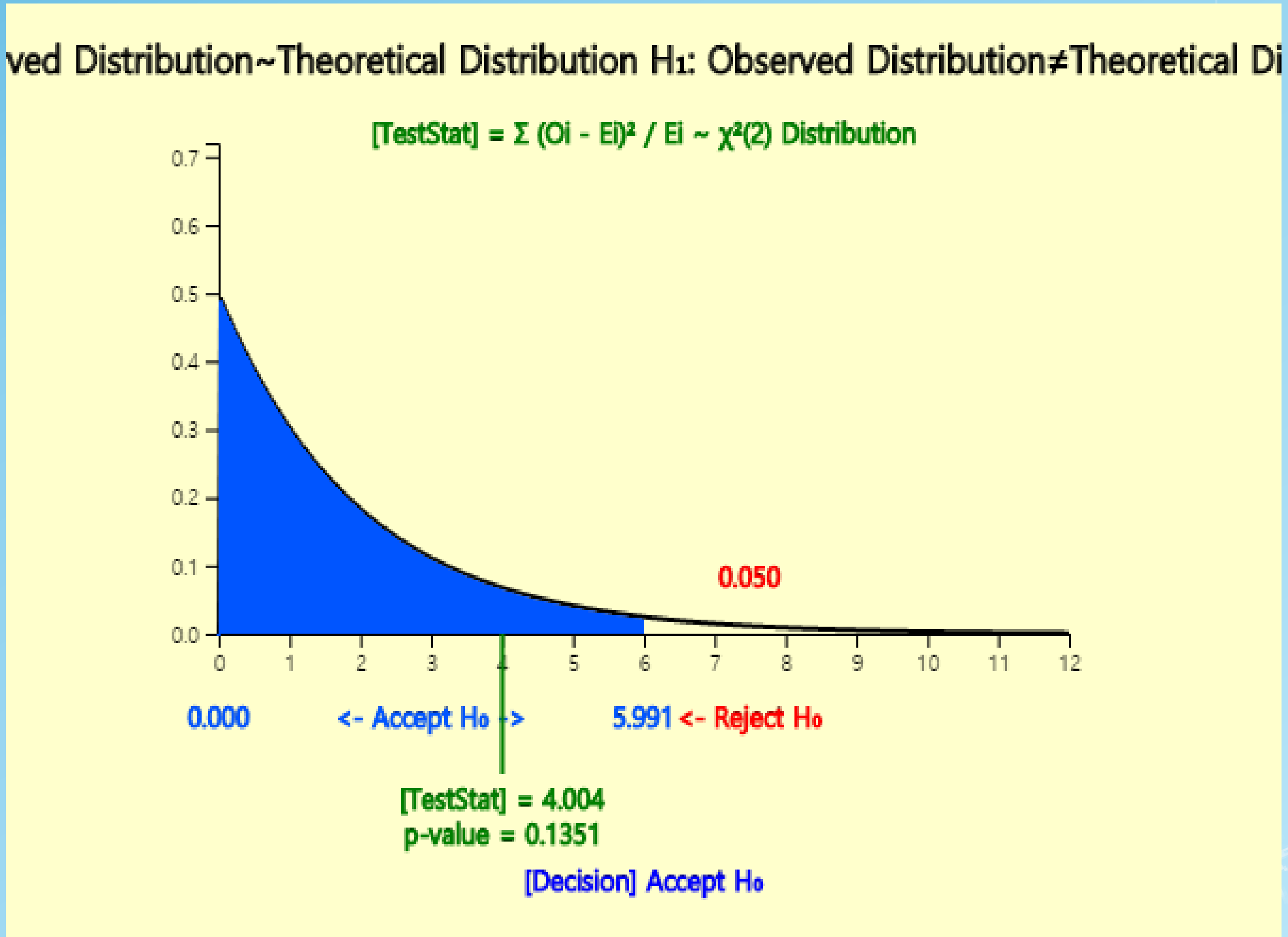
H_1 : *Observed & theoretical Distributions are different*

[Test Type] χ^2 *test* Significance Level α = ◉ 5% ○ 1%

[Sample Data] *Enter cell from upper left cell*

	Observed Frequency O	Expected Probability p	Expected Frequency E(>5)
Row 1	60	0.333	49.95
Row 2	50	0.333	49.95
Row 3	40	0.333	49.95
Row 4			
Row 5			
Row 6			
Row 7			
Row 8			
Row 9			
		합계	149.85

Execute



11.1 Goodness of Fit Test

[Goodness of Fit Test]

- A categorical variable X which has possible values $x_1, x_2, \ldots, x_k$ and their probabilities are $p_1, p_2, \ldots, p_k$ respectively.
- Observed frequencies from n samples are $(O_1, O_2, \ldots, O_k)$ and expected frequencies $(E_1, E_2, \ldots, E_k)$. The significance level is α.
- Hypothesis:

H_0: Distribution of $(O_1, O_2, \ldots, O_k)$ follows $(p_{10}, p_{20}, \ldots, p_{k0})$
H_1: Distribution of $(O_1, O_2, \ldots, O_k)$ do not follow $(p_{10}, p_{20}, \ldots, p_{k0})$

- Decision Rule:

'If $\chi^2_{obs} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} > \chi^2_{k-m-1;\,\alpha}$, reject H_0'

m is the number of population parameters estimated from the samples.
$*$ E_i should be greater than 5

Thank you