Introduction to Statistics and Data Science using *eStat*

Chapter 11 Testing Hypothesis for Categorical Data

11.1.2 Goodness of Fit Test for Continuous 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

11.1.2 Goodness of Fit Test for Continuous Data

[Example 11.1.2] The age of 30 people who visited a library in the morning is as follows. Test the hypothesis that the population is normally distributed at the significance level of 5%.

28 55 26 35 43 47 47 17 35 36 48 47 34 28 43
20 30 53 27 32 34 43 18 38 29 44 67 48 45 43

11.1 Goodness of Fit Test

< Answer of Example 11.1.2>

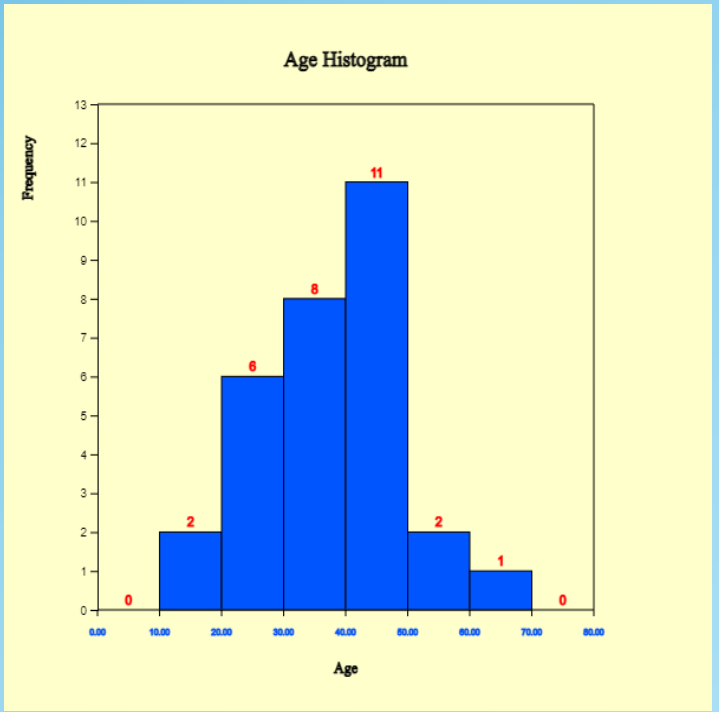
File EX110102_AgeOfLibraryVisitor.cs

Analysis Var: 1: Age — by Group: ---

(Selected data: Raw Data)

SelectedVar: V1

	Age	V2	V3	V4	V5
1	28				
2	55				
3	26				
4	35				
5	43				
6	47				
7	47				
8	17				
9	35				
10	36				
11	48				
12	47				
13	34				
14	28				
15	43				
16	20				
17	30				
18	53				
19	27				
20	32				
21	34				
22	43				
23	18				
24	38				
25	29				



Histogram Frequency Table	Group Name	()
Interval (Age)		Total
1 [10.00, 20.00)	2 (6.7%)	2 (6.7%)
2 [20.00, 30.00)	6 (20.0%)	6 (20.0%)
3 [30.00, 40.00)	8 (26.7%)	8 (26.7%)
4 [40.00, 50.00)	11 (36.7%)	11 (36.7%)
5 [50.00, 60.00)	2 (6.7%)	2 (6.7%)
6 [60.00, 70.00)	1 (3.3%)	1 (3.3%)
Total	30 (100%)	30 (100%)

11.1 Goodness of Fit Test

< Answer of Example 11.1.2>

Interval id	Interval	Observed frequency
1	$X < 20$	2
2	$20 \leq X < 30$	6
3	$30 \leq X < 40$	8
4	$40 \leq X < 50$	11
5	$50 \leq X < 60$	2
6	$X \geq 60$	1

- **Hypothesis**

  H_0 : Sample data follow a normal distribution.

  H_1 : Sample data do not follow a normal distribution

  ⇨

  H_0 : Sample data follow $N(38.000, 11.519^2)$

  H_1 : Sample data do not follow $N(38.000, 11.519^2)$

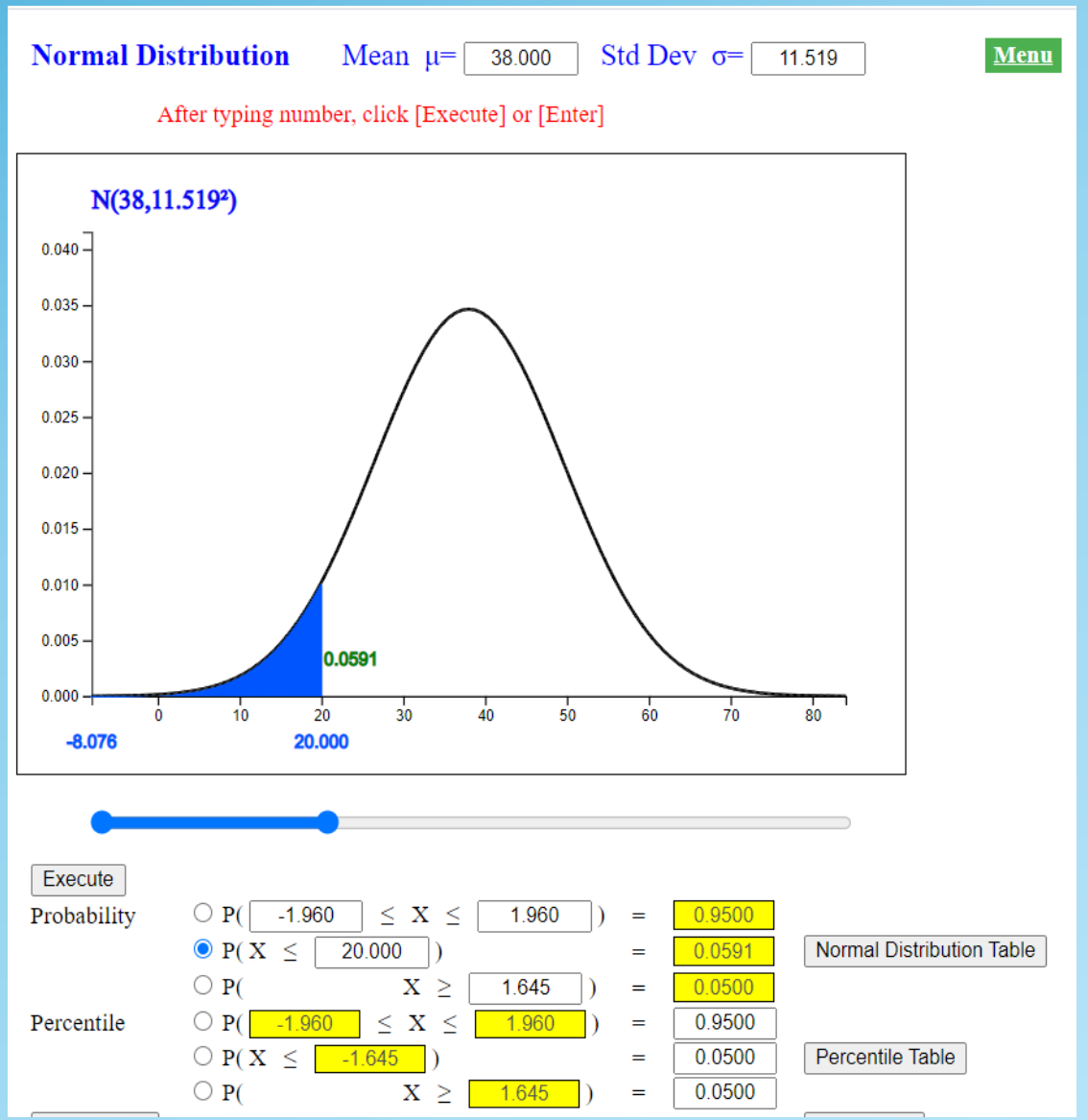
Descriptive Statistics	Analysis Var (Age)
Observation	30
Missing Observations	0
Mean	38.000
Variance (n)	128.267
Variance (n-1)	132.690
Std Dev (n)	11.325
Std Dev (n-1)	11.519
Minimum	17.000
1st Quartile	29.250
Median	37.000
3rd Quartile	46.500
Maximum	67.000
Range	50.000
Interquartile Range	17.250
Coefficient of Variation (n)	29.80 %
Coefficient of Variation (n-1)	30.31 %

11.1 Goodness of Fit Test

< Answer of Example 11.1.2>

- Probability of $N(38.000, 11.519^2)$

$$P(X < 20) = 0.059$$
$$P(20 \leq X < 30) = 0.185$$
$$P(30 \leq X < 40) = 0.325$$
$$P(40 \leq X < 50) = 0.282$$
$$P(50 \leq X < 60) = 0.121$$
$$P(X \geq 60) = 0.028$$



11.1 Goodness of Fit Test

< Answer of Example 11.1.2>

Interval id	Interval	Observed frequency	Expected probability	Expected frequency
1	$X < 20$	2	0.059	1.77
2	$20 \leq X < 30$	6	0.185	5.55
3	$30 \leq X < 40$	8	0.325	9.75
4	$40 \leq X < 50$	11	0.282	8.46
5	$50 \leq X < 60$	2	0.121	3.63
6	$X \geq 60$	1	0.028	0.84

Interval id	Interval	Observed frequency	Expected probability	Expected frequency
1	$X < 30$	8	0.244	7.32
2	$30 \leq X < 40$	8	0.325	9.75
3	$40 \leq X < 50$	11	0.282	8.46
4	$X \geq 50$	3	0.149	4.47
	Total	30	1.000	30.00

11.1 Goodness of Fit Test

<Answer of Example 11.1.2>

- Test Statistic

$$\chi^2_{obs} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$
$$= \frac{(8 - 7.32)^2}{7.32} + \frac{(8 - 9.75)^2}{9.75} + \frac{(11 - 8.46)^2}{8.46} + \frac{(3 - 4.47)^2}{4.47} = 1.623$$

- Decision Rule

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

Since $\chi^2_{4-2-1;\,0.05} = 3.841$, H_0 cannot be rejected.

11.1 Goodness of Fit Test

<Answer of Example 11.1.2>

Categorical : Goodness of Fit Test

Menu

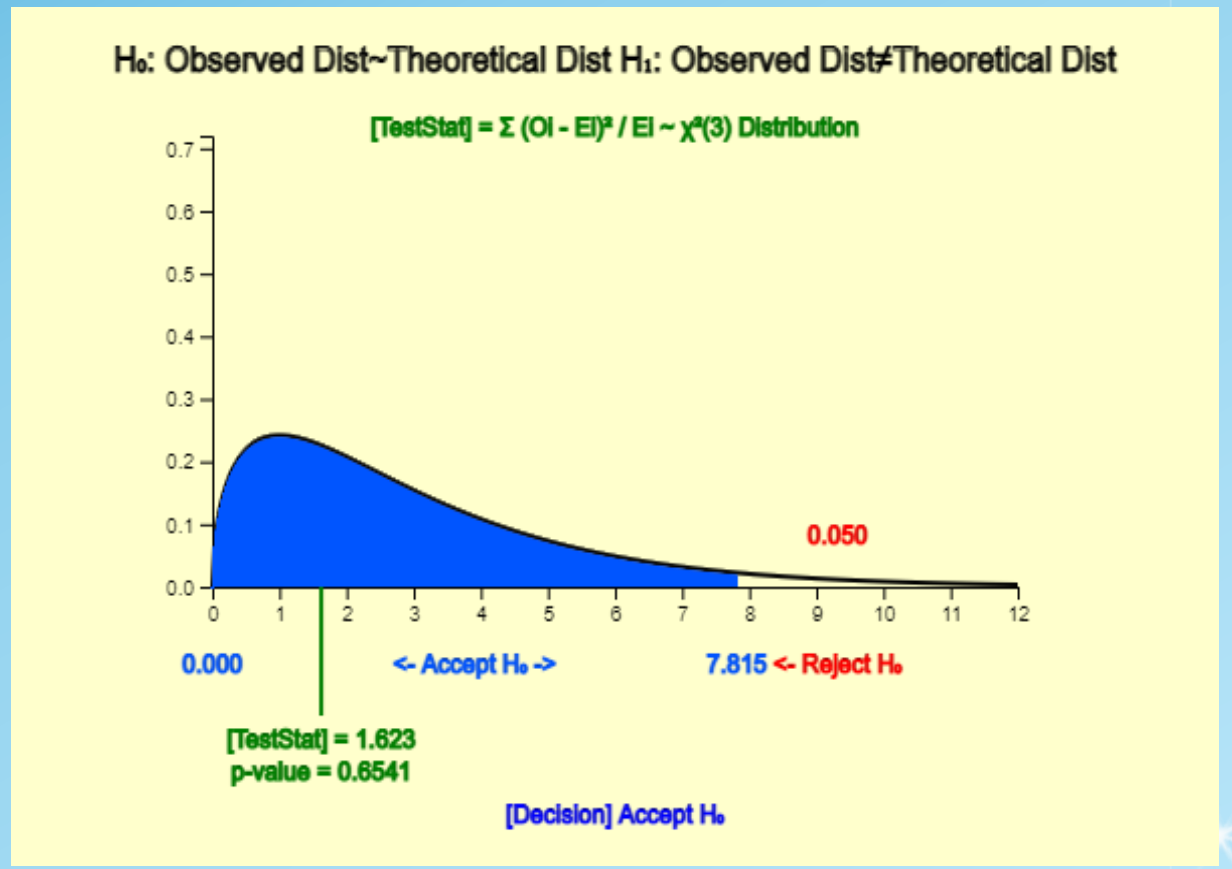
[Hypothesis] H_0 : *Observed & theoretical Dist. are the same*

H_1 : *Observed & theoretical Dist. 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	8	0.244	7.32
Row 2	8	0.325	9.75
Row 3	11	0.282	8.46
Row 4	3	0.149	4.47
Row 5			
Row 6			
Row 7			
Row 8			
Row 9			
		Total	30.00



Thank you