Hypothesis Testing: Categorical Data 10

Additional Solved Problems

Problem (1)

A sample of 50 randomly selected men with high triglyceride levels consumed 2 tablespoons of oat bran daily for six weeks. After six weeks, 60% of the men had lowered their triglyceride level. A sample of 80 men consumed 2 tablespoons of wheat bran for six weeks. After six weeks, 25% had lower triglyceride levels. Is there a significance difference in the two proportions at $\alpha = 0.01$?

Solution

Since the statistics are given in percentages (proportions) then we have to use a two-sample problem comparing two binomial proportions as follows:

Let

p_1 = proportion of men consumed 2 tablespoons daily of oat bran who had lowered their triglyceride level after six weeks.

p_2 = proportion of men consumed 2 tablespoons daily of wheat bran who had lowered their triglyceride level after six weeks.

Step (1): Sample Proportions

- Sample proportion of men consumed 2 tablespoons daily of oat bran is:

$$\hat{p}_1 = 60\% = 0.60$$

- Sample proportion of men consumed 2 tablespoons daily of wheat bran is:

$$\hat{p}_2 = 25\% = 0.25$$

Step (2): In order to compute $\hat{p}$, we must find x_1 and x_2 as follows:

$$x_1 = n_1 * \hat{p}_1 = (50)(0.60) = 30$$
$$x_2 = n_2 * \hat{p}_2 = (80)(0.25) = 20$$

Step (3): Estimated common proportions $\hat{p}$ and $\hat{q}$ are obtained as follows:

$\hat{p} = (30 + 20) / (50 + 80) = 50/130 = 0.385$

$\hat{q} = 1 - \hat{p} = 1 - 0.385 = 0.615$

Step (4): Hypotheses to be tested are:

$$H_0: p_1 = p_2 \text{ vs. } H_1: p_1 \neq p_2$$

Step (5): Compute the Test Statistic (Z)

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = \frac{|0.60 - 0.25| - \left(\frac{1}{2(50)} + \frac{1}{2(80)}\right)}{\sqrt{(0.385)(0.615)(\frac{1}{50} + \frac{1}{80})}} = \frac{0.33375}{0.08772} = 3.80$$

Step (6): Critical Value

$$Z_{1-(\alpha/2)} = Z_{1-(0.01/2)} = Z_{0.995} = 2.575 \approx 2.58$$

Step (7): Decision

Now by using the critical value method, we get Z = 3.80 > $Z_{0.995} = 2.58$, then the decision will be reject H_0 and accept H_1 at level of significance $\alpha = 0.01$.

Conclusion

The results are highly significant. Therefore, we can conclude that there is enough evidence to support the claim that there is a difference in proportions.

Notations

- $n_1\hat{p}\hat{q}$ = (50)(0.385)(0.615) = 11.839 > 5
- $n_2\hat{p}\hat{q}$ = (80)(0.385)(0.615) = 18.942 > 5
- The p-value = 2 × [1 − Φ (3.80)]= 2 × [1 – 0.9999] = 0.0001 < 0.05

Problem (2)

A sample of 50 randomly selected men with high triglyceride levels consumed 2 tablespoons of oat bran daily for six weeks. After six weeks, 60% of the men had lowered their triglyceride level. A sample of 80 men consumed 2 tablespoons of wheat bran for six weeks. After six weeks, 25% had lower triglyceride levels. By using a 2 x 2 contingency-table approach can we conclude that there is a significance difference in the two proportions at $\alpha = 0.01$?

Solution

Step (1): First compute the observed and expected tables as given below respectively:

Observed Table

Triglyceride level	Type of consumed food for six weeks		Total
	Oat bran	Wheat bran	
Lowered	30	20	50
Non-Lowered	20	60	80
Total	50	80	130

Expected Table

Triglyceride level	Type of consumed food for six weeks		Total
	Oat bran	Wheat bran	
Lowered	19.231	30.769	50
Non-Lowered	30.769	49.231	80
Total	50	80	130

Note that the minimum expected value is 19.231, which is > 5.

Step (2): Use Table 6 (*Percentage points of the chi-square distribution*) page 880 in the Appendix to find the critical value $\chi^2_{(1,\ 1-\alpha)}$ as follows:

$$\chi^2_{(1,\ 1-\alpha)} = \chi^2_{(1,\ 1-0.01)} = \chi^2_{(1,\ 0.99)} = 6.63$$

Step (3): Thus, Equation 10.5, can be applied as follows:

$$X^2 = \frac{\left(|O_{ij} - E_{ij}| - \frac{1}{2}\right)^2}{E_{ij}}$$

$$X^2 = (|O_{11} - E_{11}| - .5)^2 / E_{11} + (|O_{12} - E_{12}| - .5)^2 / E_{12} + (|O_{21} - E_{21}| - .5)^2 / E_{21} + (|O_{22} - E_{22}| - .5)^2 / E_{22}$$

$$X^2 = \frac{(|30 - 19.231| - 0.5)^2}{19.231} + \frac{(|20 - 30.769| - 0.5)^2}{30.796} + \frac{(|20 - 30.769| - 0.5)^2}{30.769} + \frac{(|60 - 49.231| - 0.5)^2}{49.231}$$

$$= 5.483 + 3.427 + 3.427 + 2.142 = 14.299 \approx 14.30$$

Step (4): Decision and Conclusion

Because we get: $X^2 = 14.30 > \chi^2_{(1,\ 0.99)} = 6.63$ and $\rightarrow p < 1 - 0.99 \rightarrow p < 0.01$ therefore the results are highly significant. Thus there is a significant difference between the two proportions at $\alpha = 0.01$.

Problem (3)

A sample of 150 people from a certain industrial community showed that 80 people suffered from a lung disease. A sample of 100 people from a rural community showed that 30 suffered from the same lung disease. At $\alpha = 0.05$, is there a difference between the proportion of people who suffer from the disease in the two communities? Use normal theory test?

Answer

$\hat{p}_1 = 0.533$, $\hat{p}_2 = 0.3$, $\hat{p} = 0.44$, $\hat{q} = 1 - \hat{p} = 0.56$

Hypotheses to be tested are: H_0: $p_1 = p_2$ vs. H_1: $p_1 \neq p_2$

Test Statistic: Z = 3.64

Decision: Reject H_0

Conclusion: There is enough evidence to support the claim that there is a significant difference in the two proportions.

Problem (4)

A recent study showed that in a sample of 80 surgeons, 45 smoked. In a sample of 120 general practitioners, 63 smoked. At $\alpha = 0.05$, by using a 2 x 2 contingency-table approach is there a difference in the two proportions?

Answer

Observed Table

Practitioner	Smoking Status		Total
	Smoked	Not Smoked	
Surgeons	45	35	80
Non- Surgeons	63	57	120
Total	108	92	200

Expected Table

Practitioner	Smoking Status		Total
	Smoked	Not Smoked	
Surgeons	43.2	36.8	80
Non- Surgeons	64.8	55.2	120
Total	108	92	200

$\hat{p}_1 = 0.5625$, $\hat{p}_2 = 0.525$, $\hat{p} = 0.54$, $\hat{q} = 1 - \hat{p} = 0.46$

Hypotheses to be tested are: H_0: $p_1 = p_2$ vs. H_1: $p_1 \neq p_2$

Test Statistic: $Z = 0.521$

Decision: Do Not Reject (Accept) H_0

Conclusion: There is not enough evidence to support the claim that there is a significant difference in the two proportions.

Problem (5)

A researcher wishes to determine whether there is a relationship between the gender (sex) of an individual and the amount of headache medications consumed. A sample of 69 people is selected, and the data in the following contingency table are obtained:

Contingency Table

Gender	Headache Consumption			Total
	Low	Moderate	High	
Male	10	9	8	27
Female	13	16	12	41
Total	23	25	20	68

At $\alpha = 0.10$, can the researcher conclude headache consumption is related to gender?

Answer

H_0: The amount of headache medications consumes is independent of the individual's gender.

$$vs$$

H_1: The amount of headache medications consumes is not independent (dependent) of the individual's gender.

We have the following:

Expected Table

Gender	Headache Consumption			Total
	Low	Moderate	High	
Male	9.13	9.93	7.94	27
Female	13.87	15.07	12.06	41
Total	23	25	20	68

$X^2 = 0.283$ follows a chi-square distribution with $df = (2-1) \times (3-1) = 2$.

Decision and Conclusion

Because we get:

$$\chi^2_{(2,\ 0.95)} = 4.605 > X^2 = 0.283, \text{ we have } p < 1 - 0.95 = 0.05$$

Therefore, H_0 is not rejected (accepted) and H_1 is rejected, then the results shows that there is not enough evidence to support the claim that the amount of headache a person consumes is dependent on the individual's gender.

Exercise (6)

The frequency distribution of the weight in kg for a random sample of 200 patients ages 30−40 years selected from Jordan is given as follows:

Observed Table

Group	Observed Frequency	Expected Frequency
< 45	12	1.96
45 – 49	44	48.54
50 – 54	82	117.77
55 – 59	53	30.96
≥ 60	9	0.77
Total	200	200

Use the chi-square goodness-of-fit test to determine at $\alpha = 0.05$ if the weight data shown in the frequency distribution is normally distributed? Assume the mean and standard deviation of this hypothetical normal distribution are given by the sample mean ($\bar{x}$ = 52 kg) and the sample standard deviation (s = 3 kg).

Answer

Step (1): Hypotheses

H_0: The weights data is normally distributed with mean 52kg and standard deviation 3 kg.

vs

H_1: The weights data is not normally distributed with mean 52kg and standard deviation 3 kg.

Step (2): Chi-Square Test Statistic Value

Weight Group	Observed Frequency	Expected Frequency	X^2- Value
< 45	12	1.96	51.423
45 – 49	44	48.54	0.425
50 – 54	82	117.77	10.864
55 – 59	53	30.96	15.690
≥ 60	9	0.77	87.965
Total	200	200	166.367

Step (3):

- Two parameters have been estimated from the data (μ, σ^2), and there are 5 groups. Therefore, k = 2, g = 5.
- Under H_0, X^2 follows a chi-square distribution with df = 5 − 2 − 1 = 2.

Step (4): Decision and Conclusion

Because we get:

$$\chi^2_{(2,\ 0.95)} = 5.99 < X^2 = 166.367$$

Therefore, H_0 is rejected and H_1 is accepted, then the results are very highly significant. Thus, the normal model does not provide an adequate fit to the data.
