STAT/MATH 415 HW#9

April 29, 2022

EXERCISES

9.1.3 In the Lottery Game, twice a day a three-digit integer is generated one digit at a time. Let p_i denote the probability of generating digit $i,\ i = 0, 1, ..., 9$. Let $\alpha = 0.05$ and use the following 50 digits to test $H_0 : p_0 = p_1 = ... = p_9 = 1/10$:

1	6	9	9	3	8	5	0	6	7	4	7	5	9	4	6	5	6	4	4	4	8	0	9	3
2	1	5	4	5	7	3	2	1	4	6	7	1	3	4	4	8	8	6	1	6	1	2	8	8

Answer: Let o_i represent the number of observations for $i,\ i = 0, 1, ..., 9$, e_i represent the expected number of observations under H_0.

i	0	1	2	3	4	5	6	7	8	9
o_i	2	6	3	4	9	5	7	4	6	4
e_i	5	5	5	5	5	5	5	5	5	5

$$X^2 = \sum_{i=0}^{9} \frac{(o_i - e_i)^2}{e_i} = \frac{9 + 1 + 4 + 1 + 16 + 0 + 4 + 1 + 1 + 1}{5} = 7.6$$

The critical region is $X^2 \geq {\chi^2}_{0.05}(10 - 1) = 16.92$, thus we do not reject H_0.

9.1.7 A rare type of heredity change causes the bacterium in *E.coli* to become resistent to streptomycin. This mutation can be detected by plating many bacteria on petri dishes containing an antibiotic medium. Any colonies that grow on this medium result from a single mutant cell. A sample of $n = 150$ petri dishes of streptomycin agar were each plated with 10^6 bacteria, and the number of colonies were counted on each dish. The observed results were that 92 dishes had 0 colonies, 46 had 1, 8 had 2, 3 had 3, and 1 dish had 4 colonies. Let X equal the number of colonies per dish. Test the hypothesis that X has a Poisson distribution. Use $\bar{x} = 0.5$ as an estimate of λ. Let $\alpha = 0.01$.

Answer: Let o_x denote the number of observations for $x = 0, 1, 2, 3, 4$ and e_x denote the expected number.

$$e_0 = 150 \times f(x = 0) = 150 \frac{0.5^0 e^{-0.5}}{0!} = 90.98$$

$$e_1 = 150 \times f(x = 1) = 150 \frac{0.5^1 e^{-0.5}}{1!} = 45.49$$

$$e_2 = 150 \times f(x = 2) = 150 \frac{0.5^2 e^{-0.5}}{2!} = 11.37$$

$$e_3 = 150 \times f(x = 3) = 150 \frac{0.5^3 e^{-0.5}}{3!} = 1.90$$

$$e_4 = 150 \times f(x \geq 4) = 150 - e_0 - e_1 - e_2 - e_3 = 0.26$$

x	0	1	2	3	4
o_x	92	46	8	3	1
e_x	90.98	45.49	11.37	1.90	0.26

$$X^2 = \sum_{x=0}^{4} \frac{(o_x - e_x)^2}{e_x} = 3.76$$

The critical region is $X^2 \geq {\chi^2}_{0.01}(5-1-1) = 11.34487$, so we do not reject H_0.

You may also group some categories, i.e. combine e_3 and e_4.

9.2.3 Each of two comparable classes of 15 students responded to two different methods of instructions, giving the following scores on a standardized test:

Class U:	91	42	39	62	55	82	67	44
	51	77	61	52	76	41	59	
Class V:	80	71	55	67	61	93	49	78
	57	88	79	81	63	51	75	

Use a chi-square test with $\alpha = 0.05$ to test the equality of the distribution of test scores by dividing the combined sample into three equal parts (low, middle, high).

Answer: First divide the combined data into three parts and then build the contigency table.

Low									
39	41	42	44	49	51	51	52	55	55
U	U	U	U	V	U	V	U	U	V
Middle									
57	59	61	61	62	63	67	67	71	75
V	U	U	V	U	V	U	V	V	V
High									
76	77	78	79	80	81	82	88	91	93
U	U	V	V	V	V	U	V	U	V

	Low	Middle	High	Total
Class U	7	4	4	15
Class V	3	6	6	15
Total	10	10	10	30

If the distributions are equal, then class (denoted by $X_i,\ i = 1, 2$) and score category (denoted by $Y_j,\ j = 1, 2, 3$) should be independent. Thus $H_0 : X \perp\!\!\!\perp Y$, that is, $P(X = i, Y = j) = P(X = i)P(Y = j)$

$$\hat{p_{1\cdot}} = \hat{p_{2\cdot}} = \frac{15}{30} = \frac{1}{2}$$

$$\hat{p_{\cdot 1}} = \hat{p_{\cdot 2}} = \hat{p_{\cdot 3}} = \frac{10}{30} = \frac{1}{3}$$

$$e_{ij} = 30 \times \hat{p_{i\cdot}} \times \hat{p_{\cdot j}} = 5 \qquad i = 1, 2 \quad j = 1, 2, 3$$

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = \frac{12}{5}$$

Since the critical region is $X^2 \geq {\chi^2}_{0.05}(5-1-2) = {\chi^2}_{0.05}(2) = 5.991$, so we do not reject H_0.

9.2.9 A survey of high school girls classified them by two attributes: whether or not they participated in sports and whether or not they had one or more older brothers. Use the following data to test the null hypothesis that these two attributes of classification are independent:

	Participated in Sports		
Older Brother(s)	Yes	No	Totals
Yes	12	8	20
No	13	27	40
Totals	25	35	60

Approximate the p-value of this test. Do we reject the null hypothesis if $\alpha = 0.05$?

Answer: Let $X_i,\ i = 1, 2$ denote the sports attribute, $Y_j,\ j = 1, 2$ denote the brother attribute (1 means yes, 2 means no). $H_0 : X \perp\!\!\!\perp Y$, that is, $P(X = i, Y = j) = P(X = i)P(Y = j)$

$$\hat{p_{1\cdot}} = P(X = 1) = \frac{5}{12}$$

$$\hat{p_{2\cdot}} = P(X = 2) = \frac{7}{12}$$

$$\hat{p_{\cdot 1}} = P(Y = 1) = \frac{1}{3}$$

$$\hat{p_{\cdot 2}} = P(Y = 2) = \frac{2}{3}$$

Then calculate the expected number of students, e_{ij} in each class under H_0

$$e_{11} = n \times P(X = 1, Y = 1) = 60 \times \frac{5}{12} \times \frac{1}{3} = \frac{25}{3}$$

$$e_{12} = n \times P(X = 1, Y = 2) = 60 \times \frac{5}{12} \times \frac{2}{3} = \frac{50}{3}$$

$$e_{21} = n \times P(X = 2, Y = 1) = 60 \times \frac{7}{12} \times \frac{1}{3} = \frac{35}{3}$$

$$e_{22} = n \times P(X = 2, Y = 2) = 60 \times \frac{7}{12} \times \frac{2}{3} = \frac{70}{3}$$

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 4.15 \sim \chi^2(3 - 1 - 1)$$

The p-value is $P(\chi^2(1) \geq 4.15) = 0.04$ and the critical region is $X^2 \geq {\chi^2}_{0.05}(1) = 3.841$, so we should reject H_0.