

# BIOS 600 · Practice Quiz for Categorical Data

Fall 2011

1. This question is inspired by last night's fantastic world series game (Games 6 between the Cardinals and the Rangers); I've actually compiled real data.

The World Series of Baseball is a best-of-7 series between the champion of the American League and the National League. The two teams compete until one of the teams wins 4 games. This means that the World Series can last 4, 5, 6, or 7 games.

If one considers the outcome of each game as a random coin-flip (that is, both teams in the world series are equally matched), then the probabilities associated with the number of games played in the world series are as follows:

$k$	Probability the World Series lasts $k$ games
4	0.1250
5	0.2500
6	0.3125
7	0.3125

The following table lists the actual number<sup>1</sup> of games played in the World Series since 1905 except for 1919, 1920, 1921, and 1994. (In 1919, 1920, and 1921, the World Series was actually a best-of-9 series. In 1994, professional baseball players were on strike.)

$k$	Number of World Series that ended in $k$ games
4	20
5	24
6	23
7	36

Test the coin-flip model with the goodness-of-fit test. As part of your solution, include

- (a) Statement of the null and alternative hypothesis
- (b) The reference distribution of the test statistic
- (c) The value of the test statistic
- (d) The p-value
- (e) The test result at an  $\alpha = 0.01$  significance level
- (f) A short interpretation of the test results

---

<sup>1</sup>Data comes from <http://www.baseball-almanac.com/ws/wsmenu.shtml>.

2. The following data reports the number and type of infections for cesarian section deliveries with and without certain risk factors (diabetes, overweight, etc.). This data was origionally published in Fahrmeir (2001).

Risk Factors	Infection		
	Type I	Type II	None
With	25	38	137
Without	4	4	43

Perform the test of no association. As part of your solution, include

- (a) Statement of the null and alternative hypothesis
- (b) The reference distribution of the test statistic
- (c) The value of the test statistic
- (d) The p-value
- (e) The test result at an  $\alpha = 0.01$  significance level
- (f) A short interpretation of the test results

3. With the data from the previous problem, combine the Type I and Type II outcomes into a single outcome. Calculate the resulting  $2 \times 2$  table. Then
- (a) Calculate the odds ratio of an infection between those pregnancies with and without risk factors
  - (b) Write the hypothesis of no association in terms of the odds ratio
  - (c) Note the reference distribution of the test statistic
  - (d) Calculate the value of the test statistic
  - (e) Calculate the p-value
  - (f) What is the test result at an  $\alpha = 0.05$  significance level
  - (g) Provide a short interpretation of the test results

4. Place the following terms in the table.

- (a) OR
- (b) Fisher's Exact Test
- (c) Hypothesis of No Association
- (d) Goodness of fit test
- (e)  $H_0 : p_1 - p_2 = \delta_0$
- (f)  $H_0 : \mu_1 - \mu_2 = \delta_0$
- (g) Linear Trend Test
- (h)  $H_0 : \mu = \mu_0$
- (i)  $H_0 : p = p_0$

2 Variables

		Y	
		Continuous	Discrete
X	Continuous		
	Discrete		

1 Variable

Continuous	Discrete

## Solutions

1. (a)  $H_0 : p_4 = 0.124 \ p_5 = 0.25 \ p_6 = 0.3125 \ p_7 = 0.3125$   
 $H_1 : p_4 \neq 0.124 \text{ or } p_5 \neq 0.25 \text{ or } p_6 \neq 0.3125 \text{ or } p_7 \neq 0.3125$   
(b) The reference distribution of the test statistic:  $\chi^2_3$ .  
(c) The value of the test statistic: 7.14  
(d) The p-value: between .1 and .05  
(e) The test result at an  $\alpha = 0.01$  significance level: Fail to reject  
(f) A short interpretation of the test results: The data are consistent with the coin-flip model.
2. (a) See notes  
(b)  $\chi^2_2$   
(c) 5.2028  
(d)  $.05 < p\text{-value} < .1$   
(e) Fail to reject  
(f)
3. (a) OR = 2.47  
(b) See notes  
(c)  $\chi^2_1$   
(d) 5  
(e)  $0.025 < p\text{-value} < 0.05$   
(f) Reject  $H_0$   
(g) The data suggests that pregnancies with risk factors that end in C-section delivery are more likely to result infection.