**The corresponding notes are on chapter 7 (Page 216) :**

[**ftp://195.214.211.1/books/DVD-009/van\_Belle\_G.,\_Fisher\_L.D.,\_Heagerty\_P.J.\_Biostatistics.\_A\_methodology\_for\_the\_health\_sciences\_(2004)(2nd\_ed.)(en)(871s).pdf**](ftp://195.214.211.1/books/DVD-009/van_Belle_G.,_Fisher_L.D.,_Heagerty_P.J._Biostatistics._A_methodology_for_the_health_sciences_(2004)(2nd_ed.)(en)(871s).pdf)

**Following is the problem to solve:**

**PB 7.9** McKeown et al. [1952] investigate evidence that the environment is involved in infantile

pyloric stenosis. The relationship between the age at onset of the symptoms in days,

and the rank of the birth (first child, second child, etc.) was given as follows:

**Age at Onset of Symptoms (Days)**

**Birth Rank 0– 6 7–13 14–20 21–27 28–34 35–41** ≥ **42**

1 42 41 116 140 99 45 58

2 28 35 63 53 49 23 31

≥3 26 21 39 48 39 14 23

**(a)** Find the expected value (under independence) for cell (i = 2, j = 3). For this

cell compute (observed - expected)^2/ expected.

**(b)** The chi-square statistic is 13.91. What are the degrees of freedom? What can you

say about the p-value?

**(c)** In the paper, the authors present, the column percents, not the frequencies, as

above. Fill in the missing values in both arrays below. The arrangement is the

same as the first table.

44 42 53 58 53 55 52

29 36 29 ? 26 28 28

? ? 18 20 21 17 21

The adjusted residual p-values are:

0.076 0.036 0.780 0.042 0.863 0.636 0.041

0.667 0.041 0.551 0.035 0.710 0.874 0.734

0.084 0.734 ? 0.856 0.843 0.445 0.954

What can you conclude?

**(d)** The authors note that the first two weeks appear to have different patterns. They also present the data as:

**Age at Onset (Days)**

**Birth Rank 0–13** ≥**14**

1 83 458

2 63 219

≥3 47 163

For this table, X2 = 8.35.

What are the degrees of freedom? What can you say about the p-value?

**(e)** Fill in the missing values in the adjusted residual table, p-value table, and column

percent table. Interpret the data.

**Adjusted Residuals**

−2.89 2.89

? ?

1.54 −1.54

p−**Values**

0.0039 0.0039

0.065 0.065

? ?

**Column %s**

43 55

33 ?

24 ?

**(f)** Why is it crucial to know whether prior to seeing these data the investigators had

hypothesized a difference in the parity distribution between the first two weeks

and the remainder of the time period?