Assignment #9

Due on Wednesday, October 21, 2009

Read Section 5.7 on *Chi–Square Tests,* pp. 278–284, in Hogg, Craig and McKean.

Do the following problems

1. Suppose you want to test whether $1,000$ observations come from a normal$(3, 4)$ distribution. The observations are divided into the following ranges:

$$\begin{aligned} A_1 &= \{x \in \mathbb{R} \mid x \leqslant 0\} \\ A_i &= \{x \in \mathbb{R} \mid i-2 < x \leqslant i-1\} \quad \text{for } \ i = 2, 3, \ldots, 7, \text{ and} \\ A_8 &= \{x \in \mathbb{R} \mid x > 6\}. \end{aligned}$$

   It is recorded that $60, 96, 140, 210, 172, 160, 88,$ and 74 of the observations fall into $A_1, A_2, \ldots, A_8$, respectively. Do the data support the assertion that the observations come from a normal$(3, 4)$ distribution?

   (a) Compute the probability, p_i, that a given observation falls into the A_i range.

   (b) Compute the expected counts in each category under the null hypothesis, H_o, that $p_1, p_2, \ldots, p_8$ describe the true distribution of counts.

   (c) Compute the Pearson Chi–Square statistic for the data given here.

   (d) Would you reject H_o at the 5% level of significance?

2. A die was cast $n = 120$ independent times and the following data resulted If we

Spots Up	1	2	3	4	5	6
Frequency	b	20	20	20	20	$40 - b$

   use a chi–square test, for what values of b would the hypothesis that the die is unbiased be rejected at the 0.025 significance level?

3. Consider the classical problem from Mendelian genetics of crossing two types of peas. The Mendelian theory states that the probabilities of the following classifications

(a) round and yellow;
(b) wrinkled and yellow;
(c) round and green; and
(d) wrinkled and green

are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$, respectively.

Suppose that from 160 independent observations the frequencies of the respective classifications are 86, 35, 26 and 13. Are these data consistent with the Mendelian theory? Justify your answer; that is, state the hypothesis you are testing and indicate the significance level that you using to make your decision.

4. A certain genetic model suggests that the probabilities of a particular trinomial distribution are $p_1 = p^2$, $p_2 = 2p(1-p)$ and $p_3 = (1-p)^2$, respectively, where

$$0 < p < 1.$$

If X_1, X_2 and X_3 represent the respective frequencies in n independent trials, explain how we could check the adequacy of the genetic model.

5. It is proposed to fit the Poisson distribution to the following data

x	0	1	2	3	$x \geqslant 4$
Frequency	20	40	16	18	6

(a) Compute the corresponding chi–square goodness of fit statistic
Suggestion: In estimating the mean of the distribution assume that

$$\mathrm{P}(X \geqslant 4)$$

can be approximated by

$$\mathrm{P}(X = 4).$$

(b) How many degrees of freedom are associated with the chi–square distribution used in this test?

(c) Do the data support the rejection of the Poisson model at the $\alpha = 0.05$ significance level?