

Assignment 2

Due Tuesday, 22 September, 2009

1. A department in a large U.S. university in 1973 had 585 applicants to its graduate program, with the following result:

Men accepted	Women accepted	Men rejected	Women rejected
353	17	207	8

- (a) Find the Pearson residuals for each of the cells in the table under the independence model. Present the values in a 2×2 table.
- (b) Test the hypothesis that being accepted or rejected was independent of the gender of the applicant, using a chi-square statistic.
- (c) Estimate the log odds ratio of rejection/acceptance of females to males.
- (d) Calculate the 95% confidence interval for the **odds ratio**. Use the confidence interval argument to verify your conclusion in part(b).
- (e) In 1998, a university had 80 applicants to its graduate scholarship program with the following result:

	Successful	Unsuccessful
Males	4	46
Females	1	29

Is there any evidence, based on these data, to support the claim that the incidence of successful scholarship application is lower in females? Evaluate the p -value for your test.

2. Consider a 3-way contingency table where we observe three response variables R, S and T which have r, s and t categories respectively. Let

$$p_{ijk} = P(R = i, S = j, T = k), \quad i = 1, \dots, r; j = 1, \dots, s; k = 1, \dots, t.$$

Consider the following log linear model for the $r \times s \times t$ contingency table

$$\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik},$$

for $i = 1, \dots, r; j = 1, \dots, s; k = 1, \dots, t$ and any parameter with a '1' subscript is set to 0. Prove that S is independent of T given R , that is

$$p_{ijk} = \frac{p_{ij \cdot} p_{i \cdot k}}{p_{i \cdot \cdot}} \quad \text{for all } i, j, k.$$

3. In a longitudinal study, out of 3182 people without cardiovascular disease, 2121 neither exercised regularly nor developed cardiovascular disease during the 4.5 year study. Amongst these 2121 individuals, the subjects were cross-classified by three factors: Personality type (A,B), cholesterol level (normal, high) and diastolic blood pressure (normal, high). The data are given below:

Personality(P)	Cholesterol (C)	Diastolic	B.P. (D)
		Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

- Test the hypothesis of complete independence of the three variables using the deviance.
 - Calculate the expected cell frequencies under the independence model and fit this model using Pearson's χ^2 .
 - Compare cholesterol level for personality types A and B.
 - Calculate a 95% confidence interval for the log-odds ratio comparing diastolic blood pressure for personality Types A and B.
 - Use this model to estimate the proportion of Type A personaility with high blood pressure and high cholesterol level.
4. The following table reports the number of beetles killed after a 5 hour exposure to gaseous carbon sulphide at various concentrations (Bliss (1935)).

log dose	Number of insects	Number killed
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8939	60	60

- Fit a logistic regression model for the probability of a beetle dying if exposed to a log dose x . State the fitted model and plot the fitted probabilities against log dose values (x).
- Calculate an approximate 90% confidence interval for the coefficient of x in the logistic regression model.
- What (log) dose results in a 20% success rate?
- Predict the number of beetles surviving if 70 are exposed to a log dose of 1.50.

5. The following data resulted from classifying a random sample of 80 males and a separate sample of 50 females according to their handedness.

	Left handed	Right handed
Males	4	76
Females	1	49

Use Fisher's exact test to see if there is any evidence, based on these data, to support the claim that the incidence of left-handedness is lower in females.