University of York

Department of Health Sciences

Applied Biostatistics

Exercise: The analysis of cross-tabulations

Question 1

The table below is taken from a study investigating the cause of diarrhoea in patients with gastroenteritis and shows the relationship between foreign travel and a positive result for the organism *Providencia alcalifaciens* (Haynes and Hawkey 1989).

Recent travel abroad?	*P. alcalifaciens* positive (no.)	negative (no.)	Total
yes	25	229	254
no	5	368	373
Total	28	597	627

$\chi^2 = 23.98$, P<0.001

a) What is meant by 'chi-squared = 23.98, P<0.001?'

b) What conditions do the data have to meet for the test to be valid?

c) What conclusions can be drawn from these data?

d) What other information would be useful in deciding whether *P. alcalifaciens* was a likely cause of gastroenteritis in travellers?

Question 2

In a study of patients admitted to an otolaryngology ward, 140 patients with nose-bleeds were compared to 113 controls with other conditions. Patients were interviewed about their alcohol consumption (McGarry *et al.*, 1994). The results were:

Alcohol consumption	Nose-bleed patients (n=140)	Other patients (n=113)	Significance
Non-drinkers	47 (34%)	39 (35%)	
Occasional drinkers	30 (21%)	40 (35%)	
Regular drinkers	63 (45%)	34 (30%)	P<0.025

The authors stated that 'The proportion of non-drinkers in the patients with nose bleeds was similar to that in the controls (34% v 35%), but the proportion of regular drinkers was significantly higher (45% v 30%, P<0.025, χ^2 test of proportions.)'

a) What is wrong with this statement and what analysis should they have done?

Question 3

In each of six patients, a series of measurements of oxygen pulse and stroke volume was made to ask whether there was any relationship between these (S Jones, unpublished). The oxygen pulse/oxygen uptake graphs could be classified as rising or plateau and the stroke volume/oxygen uptake graphs as rising or falling. The oxygen pulse assessment was tabulated by the stroke volume assessment:

Oxygen Pulse	Stroke Volume	
	Rising	Falling
Rising	2	0
Plateau	0	4

a) What method could we use to test the null hypothesis that the two classifications are related, and why?

Question 4

Blood from 471 male volunteers aged 18 to 65 years was tested for antibodies to *Helicobacter pylori* (Webb *et al.*, 1994). Seroprevalence of *H pylori* increased with age as shown in the following table:

	Age group						
	<30	30-34	35-39	40-44	45-49	50-54	55-65
Seropositive	22	26	14	30	32	23	29
Seronegative	52	55	59	53	41	28	17
% seropositive	30%	32%	19%	36%	44%	45%	63%
For trend, chi-squared = 20.6, P<0.001.							

a) What is a trend test and how would you interpret the one presented here?

b) What would be the advantages and disadvantages compared to a chi-squared test for association in a contingency table?

c) Suggest an alternative way of testing the difference in age in the two seropositivity groups, assuming that the raw data were available.

References

Haynes, J. and Hawkey, P.M. (1989) *Providencia alcalifaciens* and travellers' diarrhoea *British Medical Journal* **299**, 94-5.

McGarry, G.W., Gatehouse, S., and Hinnie, J. (1994) Relation between alcohol and nose bleeds. *British Medical Journal* **309**, 640.

Webb, P.M., Knight, T., Greaves, S., Wilson, A., Newell, D.G., Elder, J., Forman, D. (1994) Relation between infection with *Helicobacter pylori* and living conditions in childhood: evidence for person to person transmission in early life. *British Medical Journal* **308**, 750-3.

Questions 1, 2, and 4 from Martin Bland and Janet Peacock: *Statistical Questions in Evidence-based Medicine*, Oxford University Press, Oxford, 2000.