

Supplementary exercise 9.20 of IPS7e

Data: One year survival of coronary heart disease (CHD) patients (with specific diagnoses) after discharge from a coronary care unit, cross-classified with respect to ownership of a pet (yes/no).

Model: The study design is not quite clear here, but it seems that the 92 patients are a SRS from some population (perhaps patients with CHD that have been treated at a particular hospital over a period of time; the paper referenced confirms this interpretation). Thus, both classification criteria in the table are response variables, and the model is a single multinomial distribution on 4 cells.

Estimation: The estimates corresponding to the single multinomial model are proportions out of the total (shown in the table below), but there may also be interest in conditional distributions (shown as well), the most obvious ones being the survival proportions within each pet ownership group.

```
XTabs 'Status' 'Ownership';
Layout 1 1;
Frequencies 'Survivors';
Counts;
RowPercents;
ColPercents;
TotPercents;
ChiSquare;
Expected;
XResiduals;
Fisher;
DMissing 'Status' 'Ownership'.
```

Tabulated Statistics: Status, Ownership			
Using frequencies in Survivors			
Rows: Status	Columns: Ownership		
	no pet	pet	All
alive	28	50	78
	35.90	64.10	100.00
	71.79	94.34	84.78
	30.43	54.35	84.78
	33.065	44.935	
	0.7759	0.5710	
dead	11	3	14
	78.57	21.43	100.00
	28.21	5.66	15.22
	11.96	3.26	15.22
	5.935	8.065	
	4.3231	3.1811	
All	39	53	92
	42.39	57.61	100.00
	100.00	100.00	100.00
	42.39	57.61	100.00
Cell Contents			
Count			
% of Row			
% of Column			
% of Total			
Expected count			
Contribution to Chi-square			
Chi-Square Test			
	Chi-Square	DF	P-Value
Pearson	8.851	1	0.003
Likelihood Ratio	9.011	1	0.003
Fisher's Exact Test			
P-Value			
0.0063747			

We will go through the analysis while answering the specific questions of the exercise.

- (a) The study is not an experiment, because no action/treatment was imposed (to either give or not give the patient a pet).

- (b) What the researchers think about causal effects, has no bearing on whether pet ownership is an explanatory variable within the collected data or not. If the study design was as described above, both variables are response variables.

To examine the probability of survival for patients with and without a pet, we compute conditional probabilities. They are in each column of the table (we condition on column values). The sample survival probability for pet-owners was $50/53 = 94\%$ and for non pet-owners $28/39 = 72\%$. It seems that pet-owners tend to survive more frequently than non pet-owners.

- (c) The hypotheses matching our single multinomial model (with two response variables) are:

H_0 : no association, or independence, between survival and pet ownership,

H_a : dependence, or association, between survival and pet ownership.

If we wanted to phrase the hypotheses in terms of the conditional probabilities within pet ownership groups, it would be more natural to start from the model with two independent binomial distributions (one for each pet ownership group). We could justify using this model even if both variables are response variables if the interest of the study was *only* in these conditional distributions. In that case we would formulate the hypotheses as

H_0 : same survival probability for pet owners and non-owners,

H_a : different survival probability for pet owners and non-owners.

In both cases we have chosen a two-sided alternative hypothesis, but because there are only two groups it would be possible to use a one-sided alternative. The wording in question (b) of the exercise suggests this may be the preferred option, with the alternative hypothesis stating that the survival probability is higher in the pet owner group.

- (d) The Minitab listing gives the Pearson test statistic and P -value as:

$$X^2 = 8.85, \quad df = 1, \quad P = 0.003.$$

Note that all expected values are > 5 , so the use of the X^2 -statistic and the chi-square distribution should be ok. (It is perhaps interesting to note that the paper in which these results were published stated the result as: $\chi^2 = 8.9$, $P < 0.002$; that is, the P -value was misstated.)

Fisher's exact test (with a two-sided alternative) gives $P = 0.006$, but even if the P -value is slightly different from that of the X^2 -statistic, the conclusion (in this case) remains the same.

- (e) We conclude the data to give a strong indication, and formal statistical evidence at $\alpha = 0.05$, of an association between survival and pet-ownership, whereby pet owners have higher survival probability. However, we should not conclude that this is a causal relation (pet-ownership affecting survival), because there may be lurking variables, for example the condition of the patients (severely ill patients may not be able to have pets). (Again, the published paper goes quite far in its discussion to indicate that the study shows pet ownership to have "health effects"; strictly speaking, this is a misleading conclusion from a study of this type.)