Statistics 347, Homework 3, due Thursday Jan 31

Discussion of homework problems among students is encouraged. However, all material handed in for credit must be your own work.

Please hand in each problem in a separate file.

1. (a) Use the data 'quine' from the MASS library. It provides data on the number of days children were absent from school in New South Wales Australia. The children were classified by four factors, Age - 4 grade levels, Eth (2 levels)- Aboriginal or non-Aboriginal, Lrn (2 levels) - slow or average learner, Sex (2-levels). Analyze the data using the Poisson model with and without dispersion. Compare the main effects model to models with two-way interactions. Which model do you prefer?

   (b) Assume that $Y \sim Poisson(Z)$ and that $Z \sim \Gamma(\mu\phi, 1/\phi)$, where $\Gamma(k,\theta)$ has the density

   $$f(z;k,\theta) = \frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-z/\theta},$$

   with mean $EZ = \mu$ and $Var(Z) = \mu/\phi$. Show that the conditional density of Z given Y is also from the Γ family. What are the values of the two parameters.

   (c) Show that the marginal of Y has negative binomial distribution: $f(y;r,p) = \frac{\Gamma(y+r)}{\Gamma(r)\Gamma(y+1)} p^y (1-p)^r$, with $0 < p < 1$ and $r > 0$. What are the values of r and p in terms of μ, ϕ and what are the mean and variance of Y. This defines a probability model with Poisson observations but a latent Γ variable yielding a variance that is not equal to the mean.

   (d) Another example: assume that $Z \sim \Gamma(\nu, \mu/\nu)$. Do the same calculations as above and find the mean and variance of the marginal of Y.

   (e) Show that if ν is known the negative binomial has exponential family form. What is the canonical link?

   (f) We want to compare the analysis of the quine data using the negative binomial family to the analysis in (a). We need to choose ν. Instead of using Maximum Likelihood do the following. Compute the mean and variance of the counts for each combination of factors. Plot the variance as a function of the mean over all factors and based on this plot choose the ν that best describes the relationship between the means and variances. Explain your choice.

   Now using the negative-binomial family in the glm function to fit the data in 'quine' and perform the same analysis on the predictors as you did for the Poisson model. Which model do you prefer, negative binomial or quasi-Poisson.

2. Table 1 gives the mean number of children born per woman, the women being classified by place, education, and years since first marriage. Any systematic variation in the number of children is of interest.

   (a) Fit an appropriate model describing how the number of children varies with marital age, mother's abode and education. Give a brief synopsis of the arguments justifying your formulation and choice of model, including checks for model adequacy.

   (b) Explain the meaning of all parameters in your model. Comment on the major factors affecting fertility.

   (c) Construct a 95% confidence interval for the mean number of children born to an urban woman with upper elementary education after ten years of marriage.

   (d) Estimate the lifetime average number of children born to rural women with secondary education. Give 90% confidence limits.

Table 1: Mean number of children born to women in Fiji of Indian race, by marital duration, type of place and education. Observed mean values and sample sizes.

Years since first marriage	Type of place: Urban, Education (1)	(2)	(3)	(4)	Type of place: Rural, Education (1)	(2)	(3)	(4)
<5	1.17	0.85	1.05	0.69	0.97	0.96	0.97	0.74
	12	27	39	51	62	102	107	47
5–9	2.54	2.65	2.68	2.29	2.44	2.71	2.47	2.24
	13	37	44	21	70	117	81	21
10–14	4.17	3.33	3.62	3.33	4.14	4.14	3.94	3.33
	18	43	29	15	88	132	50	9
15–19	4.70	5.36	4.60	3.80	5.06	5.59	4.50	2.00
	23	42	20	15	114	86	30	1
20–24	5.36	5.88	5.00	5.33	6.46	6.34	5.74	2.50
	22	25	13	3	117	68	23	2
25+	6.52	7.51	7.54	—	7.48	7.81	5.80	—
	46	45	13	0	195	59	10	0

Education categories are: (1) none, (2) lower elementary, (3) upper elementary, (4) secondary or higher. Lower figures give the number of women in the sample.