A12 The table below shows the three-point shooting, game by game, of Ray Allen of the Boston Celtics during the 2010 NBA playoffs (e.g., he made 0 of 4 shots in game 1). Commentators remarked that his shooting varied dramatically from game to game. In game i, suppose that $n_i y_i$ = number of three-point shots made out of n_i attempts is a $\text{Bin}(n_i, \pi_i)$ variate and the $\{y_i\}$ are independent.

Game	y_i	Game	y_i	Game	y_i	Game	y_i	Game	y_i
1	0/4	6	2/7	11	0/5	16	1/3	21	0/4
2	7/9	7	3/7	12	2/5	17	3/7	22	0/4
3	4/11	8	0/1	13	0/5	18	0/2	23	2/5
4	3/6	9	1/8	14	2/4	19	8/11	24	2/7
5	5/6	10	6/9	15	5/7	20	0/8		

(a) Fit the model $\pi_i = \beta_0$ using a binomial GLM with the logit link. Find and interpret $\hat{\beta}_0$. Does the model fit the data well?

(b) Fit the quasi-binomial model $\pi_i = \beta_0$ with $\text{var}(y_i) = \phi\pi_i(1-\pi_i)/n_i$. Is $\hat{\beta}_0$ the same as in part (a)? Why or why not?

(c) Fit the quasi-binomial model $\pi_i = \beta_0$ with $\text{var}(y_i) = \{1+\varrho(n_i-1)\}\pi_i(1-\pi_i)/n_i$. Is $\hat{\beta}_0$ the same as in part (a)? Why or why not?

(d) Fit the beta-binomial model $\pi_i = \beta_0$. Is $\hat{\beta}_0$ the same as in part (a)? Why or why not?

(e) Under models (a)–(d), compute the 95% confidence interval for β_0 and interpret.

(f) Describe a factor that could cause overdispersion. Which one of the models (b)–(e) that adjust for overdispersion do you find the most appropriate for this data?

A13 Consider the following study of automobile accident records in Florida.

Safety Equipment in Use	Whether Ejected	Injury	
		Nonfatal	Fatal
Seat belt	Yes	1,105	14
	No	411,111	483
None	Yes	4,624	497
	No	157,342	1,008

Denote safety equipment by S, whether ejected by E and fatal injury as I.

(a) Treating fatal injury as a response, fit the logistic model `S+E` to the data. Calculate the fitted counts, and decide whether this model is appropriate for the data at hand using both the likelihood-ratio and Pearson's X^2 statistic.

(b) Consider the association SI between having a fatal injury and wearing a seatbelt, and EI between having a fatal injury and being ejected. Compute the corresponding odds ratios using the model `S+E`, along with 95% confidence interval. Interpret the results and explain why the model `S+E` is called the *homogeneous association model*.

(c) Fit the model with only `E` as predictor, and recalculate the odds ratios from part (b), along with their 95% confidence intervals (if appropriate). What kind of conditional independence relationship does this model describe? Does it fit the data well?

A14 Consider the data from the US 2008 General Social Survey. For subject i, let y_i = belief in existence of heaven (1 = yes, 2 = unsure, 3 = no), x_{i1} = gender (1 = female, 0 = male) and x_{i2} = race (1 = black, 0 = white). The data is as follows:

```
cbind(race,gender,y1,y2,y3)

##      race gender  y1  y2 y3
## [1,]    1      1  88  16  2
## [2,]    1      0  54   7  5
## [3,]    0      1 397 141 24
## [4,]    0      0 235 189 39
```

(a) Fit the baseline category model

$$\log(\pi_{ij}/\pi_{i3}) = \alpha_j + \beta_j^G x_{i1} + \beta_j^R x_{i2}, \quad j = 1, 2.$$

Report the prediction equations for $\log(\pi_{i1}/\pi_{i3})$, $\log(\pi_{i2}/\pi_{i3})$ and $\log(\pi_{i1}/\pi_{i2})$.

(b) Using the "yes" and "no" response categories, interpret the conditional gender effect using a 95% confidence interval for the odds ratio.

(c) Construct a likelihood-ratio test of the hypothesis that opinion is independent of gender, given race. Interpret.

A15 For a binary explanatory variable, explain why the cumulative logit model with proportional odds structure is unlikely to fit well if, for an underlying latent response, the two groups have similar location but very different dispersion.