Generalised Linear Models – 3rd Homework Assignment

Space Shuttle (Agresti Problem 5.2)

1. For the 23 space shuttle flights before the *Challenger* mission disaster in 1986, the table below shows the temperature at the time of the flight and whether at least one primary O-ring suffered thermal distress.

Ft	°F	TD	Ft	°F	TD	Ft	°F	TD	Ft	°F	TD	Ft	°F	TD
1	66	0	2	70	1	3	69	0	4	68	0	5	67	0
6	72	0	7	73	0	8	70	0	9	57	1	10	63	1
11	70	1	12	78	0	13	67	0	14	53	1	15	67	1
16	75	0	17	70	0	18	81	0	19	76	0	20	79	0
21	75	1	22	76	0	23	58	1						

2. Use logistic regression to model the effect of temperature on the probability of thermal distress. Plot a figure of the fitted model and interpret.
3. Estimate the probability of thermal distress at 31°F, the temperature at the place and time of the *Challenger* flight.
4. Construct a confidence interval for the effect of temperature on the odds of thermal distress, and test the statistical significance of the effect. (Remark: $\text{odds}(t) = \Pr(TD = 1|t)/\Pr(TD = 0|t) = \exp(\beta_0)\exp(\beta_t)^t$, i.e. a confidence interval for $\exp(\beta_t)$).
5. Check the model fit by comparing it to a more complex model that additionally allows for a quadratic temperature effect.

Car Accidents (Agresti Problem 8.5)

1. The table below refers to automobile accidents in Florida in 1988.

Safety Equipment in use	Whether ejected	Injury Nonfatal	Injury Fatal
Seat belt	Yes	1105	14
	No	411111	483
none	Yes	4624	497
	No	157342	1008

2. Find a log-linear model that describes the data well. Interpret associations. (My proposal would be to start with a model that includes all interaction effects and to reduce as long as interactions are found not to be significant. Do not violate the model hierarchy!)
3. Compare observed counts with the respective fitted means under the following 5 models: $B + E + F$, $B : E + F$, $B : E + B : F$, $B : E + B : F + E : F$, $B : E : F$. For this comparison construct a table with the Yes/No levels of the 3 factors B (belt), E (ejected), and F (fatal) as the first 3 columns, followed by the fitted values of the 5 models. (Hint: concentrate on sums of the fitted means, you will find some exact fits with respect to these sums!)
4. Treating whether killed as the response, fit an equivalent logit model and interpret the effects. **Hint**: Let Y denote this (binary) response and let X and Z be some

explanatory factors. For a log-linear model with all two-way interactions included ($X:Y$, $X:Z$, $Y:Z$) the following holds for the mean of Y:

$$\begin{aligned}
\log \frac{\Pr(Y=1|X=i,Z=k)}{\Pr(Y=2|X=i,Z=k)} &= \log \frac{\mu_{i1k}}{\mu_{i2k}} = \log \mu_{i1k} - \log \mu_{i2k} \\
&= \left(\mu + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ}\right) \\
&\quad -\left(\mu + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ}\right) \\
&= (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ})\,.
\end{aligned}$$

The first term is constant, the second only depends on level i of X, and the third on level k of Z. Thus,

$$\operatorname{logit} \Pr(Y=1|X=i,Z=k) = \alpha + \beta_i^X + \beta_k^Z$$

is the equivalent model.

Ship Damages (McCullagh & Nelder 6.3.2)

1. The ships data from the MASS package concern a type of damage caused by waves to the forward section of cargo-carrying vessels (`library(MASS); data(ships)`). The variables are: `incidents` (number of damage incidents), `service` (aggregate months of service), `period` (period of operation: 1960–74, 75–79), `year` (year of construction: 1960–64, 65–69, 70–74, 75–79), `type` (type: "A" to "E"). Here it makes sense to model the expected number of incidents per aggregate months of service. Thus use `offset=log(service)` in the log-linear model for `incidents`.
2. First exclude ships with 0 months of service and convert the period and year variables to factors. Consider a log-linear model including all the main effects, i.e.

   ```
   log(expected number of damage incidents)
                  = beta_0 + log(aggregate months service)
                           + (effect due to ship type)
                           + (effect due to year of construction)
                           + (effect due to service period).
   ```

   Interpret the estimation results.
3. What are your findings when using `log(service)` not as an offset in the model but as an additional predictor?
4. Are there noticeable residuals?
5. Now check the necessity for any additional two-way interaction effects by the respective `anova` call.