

- As part of a quality improvement study at a semiconductor factory, a sample of wafers was drawn and cross-classified according to whether a particle was found on the die that produced the wafer and whether the wafer was good or bad (Hall 1994, Faraway 2006). The main question of interest concerning these data is whether the presence of particles on the wafer affects the quality outcome. The data are given in the following table.

	# particle	
# quality	no	yes
# good	320	14
# bad	80	36

Please answer the following questions based on the data.

- (20 points) Suppose one fits a multinomial model with the data. Under the null hypothesis that the quality and presence of a particle on the wafer are independent, what are the fitted frequencies in the four cells?
- (20 points) Find the deviance (compared to the saturated model) of this model.
- (20 points) Suppose instead one fits a Poisson model with y representing the counts in the cells:

```
mod1 <- glm(y ~ particle+quality, poisson)
```

Would the deviance be the same as that from the multinomial model? If so, why?

- (15 points) Explain the equivalence of the prospective and the retrospective study in the Bernoulli regression with the logit link.
- Consider the following random effect model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \tag{1}$$

with $i = 1, \dots, I$ and $j = 1, \dots, J$. It is assumed that $\epsilon_{ij} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ and are independent of $\alpha_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$. The parameters in this model are μ, σ_ϵ^2 , and σ_α^2 .

- (15 points) Derive the ANOVA estimates of $(\mu, \sigma_\alpha^2, \sigma_\epsilon^2)$.
- (10 points) Derive the REML estimates of $(\mu, \sigma_\alpha^2, \sigma_\epsilon^2)$.