BIOSTAT/STAT 570: Coursework 6

To be submitted to the course canvas site by 11:59pm Monday 29th November, 2021.

1. Table 1 reproduce data from Altham (1991) of counts of T_4 cells/mm^3 in blood samples from 20 patients in remission from Hodgkin's disease and 20 other patients in remission from disseminated malignancies. The question of interest here is: Is there a difference in the distribution of cell counts between the two diseases? A quantitative assessment of any difference is also desirable.

   (a) Carry out an exploratory data analysis and provide a summary of the two distributions (marks will be deducted for unnecessary plots/summaries).

   (b) We now examine various approaches to obtaining a 90% confidence interval for differences (on some scale) between distributions. Obtain such a confidence intervals for these data when the data are: i) on their original scale; ii) log$_e$ transformed; iii) square root transformed. What are the considerations when choosing a scale here?

   (c) We will fit Poisson, gamma and inverse Gaussian models to the cell count data assuming the canonical link with $g(\mu_i) = \boldsymbol{x}_i\boldsymbol{\beta}$, where $\boldsymbol{x}_i = [1\ 0]$ for $i = 1, \dots, n = 20$, and $\boldsymbol{x}_i = [1\ 1]$ for $i = n+1, \dots, 2n = 40$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)$. The question of interest here is whether the means of the two groups are equal. Express this question in terms of β_0 and β_1. For what function of $\boldsymbol{\beta}$ is this question answered on the scale of the original data?

   (d) Using the asymptotic distribution of the MLE, that is

   $$\boldsymbol{I}(\widehat{\boldsymbol{\beta}})^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathsf{N}_2\left(0, I_2\right),$$

   give 90% confidence intervals for each parameter. Under each of the distributional assumptions, would you conclude that the means of the two groups are equal?

Hodgkin's Disease	Non-Hodgkin's Disease
396	375
568	375
1212	752
171	208
554	151
1104	116
257	736
435	192
295	315
397	1252
288	675
1004	700
431	440
795	771
1621	688
1378	426
902	410
958	979
1283	377
2415	503

Table 1: Counts of T_4 cells/mm^3 in blood samples from 20 patients in remission from Hodgkin's disease and 20 other patients in remission from disseminated malignancies.

2. The data in Table 2, taken from Wakefield et al. (1994), were collected following the administration of a single 30mg dose of the drug Cadralazine to a cardiac failure patient. The response y_i represents the drug concentration at time x_i, $i = 1, \ldots, 8$. The most straightforward model for these data is to assume

$$\log y_i = \mu(\boldsymbol{\beta}) + \epsilon_i = \log\left[\frac{D}{V}\exp(-k_e x_i)\right] + \epsilon_i,$$

where $\epsilon_i \sim_{iid} \mathsf{N}(0, \sigma^2)$, $\boldsymbol{\beta} = [V, k_e]$ and the dose is $D = 30$. The parameters are the volume of distribution $V > 0$ and the elimination rate k_e.

i	x_i	y_i
1	2	1.63
2	4	1.01
3	6	0.73
4	8	0.55
5	10	0.41
6	24	0.01
7	28	0.06
8	32	0.02

Table 2: Concentrations y_i of the drug Cadralazine (in mg/liter) as a function of time (in hours) x_i, for $i = 1, \ldots, 8$.

(a) For this model obtain expressions for:

i. The log-likelihood function $L(\boldsymbol{\beta}, \sigma^2)$.

ii. The score function $\boldsymbol{S}(\boldsymbol{\beta}, \sigma^2)$.

iii. The expected information matrix $\boldsymbol{I}(\boldsymbol{\beta}, \sigma^2)$.

(b) Obtain the MLE, and give an asymptotic 95% confidence interval for each element of $\boldsymbol{\beta}$.

(c) Plot the data, along with the fitted curve.

(d) Using residuals, examine the appropriateness of the assumptions of the above model. Does the model seem reasonable for these data?

(e) The clearance $Cl = V \times k_e$ and elimination half-life $x_{1/2} = \log 2/k_e$ are parameters of interest in this experiment. Find the MLEs of these parameters along with asymptotic 95% confidence intervals.