

Question1 (4 points)

In the description of the dynamics of excitons in semiconductor physics, an electromagnetic device is used to detect flaws. The reactivity Y of the device is known to be related to the thickness X of the plate used in the device. Given the following established model

$$Y_i | X_i = x_i \sim \text{Normal} \left(\beta_1 \left(1 - \exp \left(-e^{\beta_2 + \beta_3 x_i} \right) \right), \sigma^2 \right)$$

and the data in `semi.txt`.

- (a) (1 point) Find the maximum likelihood estimates of the parameters β_1 , β_2 and β_3 .
- (b) (1 point) Plot the estimated $\mathbb{E}[Y_i | X_i]$ on the same graph with the observed data.
- (c) (1 point) Suggest and compute an estimate of σ^2 .
- (d) (1 point) If β_1 and β_3 are known to be positive, then show that no matter how thin the plate used, the average reactivity can never exceed

$$\beta_1 \left(1 - \exp \left(-e^{-\beta_2} \right) \right)$$

Estimate this upper bound for the reactivity.

Question2 (3 points)

Using the data `newcar.csv` to build a logistic regression, in which `NEWCAR` is the response and two predictor variables are `INCOME` and `CAR.AGE` without any transformation. Assume there is no evidence of lack of fit, and the large-sample inferences are valid.

- (a) (1 point) Use Wald test to determine whether `CAR.AGE` can be dropped from the model.
- (b) (1 point) Use the likelihood ratio test to determine whether `CAR.AGE` is needed.
- (c) (1 point) What is the estimated probability `NEWCAR` = 1 when `INCOME` = 50 and `CAR.AGE` = 3?

Question3 (3 points)

The following table contains the counts of leukemia in five counties of California from 1996 to 2000 as well as the county population in 2000.

County	Count (n_i)	Population (x_i)
Marin	22	247289
Contra Costa	146	948816
Alameda	226	1443741
San Francisco	47	776733
San Mateo	52	70716

- (a) (1 point) Consider the following model for the data

$$N_i | X_i = x_i \sim \text{Poisson}(\theta x_i)$$

Find the maximum likelihood estimate of θ .

- (b) (2 points) Assume large-sample inferences are valid. Using a deviance based hypothesis testing to determine whether using the same θ for all counties is appropriate, that is, whether we need to introduce the categorical variable `County` into the model.