

MATH 523 - MIDTERM EXAMINATION 2012

Calculators may be used.

- (a) Suppose $Z \sim \text{Binomial}(m, \pi)$. Write the distribution of $Y = Z/m$ in Exponential-Dispersion family form, and identify the relationship between the canonical parameter of the distribution, θ , and $\mu = \mathbb{E}[Y]$. Also, establish the functional relationship between μ and $\text{Var}[Y]$.

5 MARKS

- (b) To construct a generalized linear model (GLM) based on p covariates, $x = (x_1, \dots, x_p)$, a linear predictor $\eta = x\beta$ based on parameter $\beta = (\beta_1, \dots, \beta_p)^T$ is used, in conjunction with a *link function*, g . Explain briefly the role of the link function, and define the *canonical* link function. For the model in (a), identify the canonical link function, and also suggest one other plausible link function.

5 MARKS

- (c) Write down the log likelihood function for the model in (a) derived from observed data $y_1, \dots, y_n$, where $m_i Y_i = Z_i \sim \text{Binomial}(m_i, \pi_i)$ are independent random variables.

3 MARKS

- (d) Using the canonical link function, write down the p *score equations* for $\beta_1, \dots, \beta_p$ for the log likelihood in (c) that facilitate the computation of maximum likelihood estimates for those parameters. Assume that the linear predictor takes the form $\eta_i = x_i\beta$.

3 MARKS

- (e) Suppose that a single factor predictor, X , taking **two** levels is used to construct the linear predictor. For data as in (c), compute the maximum likelihood estimates for the parameters in a GLM, assuming that the canonical link function is used. Explain precisely the interpretation of each parameter in the model.

6 MARKS

- (f) The following data arose from a cure-rate experiment. At each of $n = 8$ dose levels, a number (m_i) of male patients suffering from a specific illness were treated different amounts of medication, and the number of patients cured (z_i) after a course of treatment was recorded.

Dose (mg)	10	20	40	70	100	150	200	500
m_i	14	10	49	24	15	24	48	49
z_i	2	1	9	3	5	17	40	49

The results of an R analysis of these data are displayed as OUTPUT 1. Comment on the analysis carried out and the model fitted, and interpret the results presented.

6 MARKS

- (g) In a related experiment, data similar to those in (f) were analysed, but the sample contained for both male and female patients. On the basis of the OUTPUT 2, assess the evidence that the dose effect is the **different in males and females**.

Recall that for an analysis of deviance to compare two nested models, under the hypothesis that the simpler model (with p_A parameters) is an adequate simplification of the more complex model (with p_B parameters), the change in deviance is distributed as Chi-squared random variable with $p_B - p_A$ degrees of freedom.

8 MARKS