

MATH 523 - MIDTERM EXAMINATION 2013

Calculators may be used.

- (a) The inverse Gaussian distribution is a distribution for positive continuous random variables, Y , with density

$$f_Y(y; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2 y^3} \right)^{1/2} \exp \left\{ -\frac{1}{2\mu^2 y \sigma^2} (y - \mu)^2 \right\} \quad y > 0$$

and zero otherwise, for parameters $\mu, \sigma > 0$. This density yields expectation μ .

Show that this density is a member of the Exponential-Dispersion family, and identify the canonical parameter, θ .

6 MARKS

- (b) For the density in (a), compute the variance, and hence identify the mean-variance relationship.

6 MARKS

- (c) Identify the canonical link function for a GLM based on (a), and suggest two other link functions that might be used. Of these three link functions, identify, with reasoning, which one would be most suitable for general modelling purposes.

6 MARKS

- (d) The inverse Gaussian density exhibits positive skewness. Suggest a transformation of Y that would reduce skewness; justify your suggestion.

4 MARKS

- (e) Using the canonical link function, write down the *p score equations* for $\beta_1, \dots, \beta_p$ arising from the model in (a) that facilitate the computation of maximum likelihood estimates for those parameters. Assume that the linear predictor takes the form $\eta_i = x_i \beta$, and that an independent sample $(y_i, x_i), i = 1, \dots, n$ is available.

6 MARKS

- (f) Derive the deviance for the distribution in (a) based on the sample $(y_i, x_i), i = 1, \dots, n$.

6 MARKS

- (g) The following data are 31 measurements of chlorobenzene exposure concentrations (in parts per million) on morning (AM) and afternoon (PM) shifts in a chemical plant.

AM	13.7, 10.2, 9.9, 4.3, 5.6, 45.6, 42.0, 14.1, 3.8, 9.3, 10.6, 91.3, 2.2, 3.8, 6.0, 17.8
PM	131.8, 31.0, 4.2, 2.6, 27.6, 1.7, 7.0, 2.1, 1.5, 7.5, 2.5, 2.4, 51.9, 12.9, 12.3

The results of an R analysis of these data are displayed as OUTPUT 1 on page 2. Comment on the analysis carried out and the models fitted, and interpret the results presented.

8 MARKS

- (h) From the output, determine which link function is being used as the default in R.

5 MARKS

- (i) Explain the output from lines 17-19.

3 MARKS

OUTPUT 1

```
1 > y.am<-c(13.7,10.2,9.9,4.3,5.6,45.6,42.0,14.1,3.8,9.3,10.6,91.3,2.2,3.8,6.0,17.8)
2 > y.pm<-c(131.8,31.0,4.2,2.6,27.6,1.7,7.0,2.1,1.5,7.5,2.5,2.4,51.9,12.9,12.3)
3 > y<-c(y.am,y.pm)
4 >
5 > mean(y.am)
6 [1] 18.1375
7 > mean(y.pm)
8 [1] 19.93333
9 > mean(y)
10 [1] 19.00645
11 > shift<-factor(rep(1:2,c(length(y.am),length(y.pm))),labels=c("AM","PM"))
12 >
13 > summary(glm(y~shift,family=inverse.gaussian))
14
15 Call: glm(formula = y ~ shift, family = inverse.gaussian)
16
17 Deviance Residuals:
18      Min       1Q   Median       3Q      Max
19 -0.75505  -0.40551  -0.15978  -0.03185   0.48884
20
21 Coefficients:
22             Estimate Std. Error t value Pr(>|t|)
23 (Intercept)  0.0030398  0.0022238   1.367   0.182
24 shiftPM      -0.0005231  0.0029861  -0.175   0.862
25
26 (Dispersion parameter for inverse.gaussian family taken to be 0.1180285)
27
28 Null deviance: 4.2862  on 30  degrees of freedom
29 Residual deviance: 4.2825  on 29  degrees of freedom
30 AIC: 237.58
31 Number of Fisher Scoring iterations: 8
32
33 > summary(glm(y~1,family=inverse.gaussian))
34
35 Call: glm(formula = y ~ 1, family = inverse.gaussian)
36
37 Deviance Residuals:
38      Min       1Q   Median       3Q      Max
39 -0.7521  -0.4104  -0.1675  -0.0419   0.5169
40
41 Coefficients:
42             Estimate Std. Error t value Pr(>|t|)
43 (Intercept)  0.002768   0.001494   1.853   0.0737 .
44 ---
45 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
46
47 (Dispersion parameter for inverse.gaussian family taken to be 0.1187655)
48
49 Null deviance: 4.2862  on 30  degrees of freedom
50 Residual deviance: 4.2862  on 30  degrees of freedom
51 AIC: 235.6
52 Number of Fisher Scoring iterations: 8
```