Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

Faculty of Engineering, Mathematics and Science
School of Computer Science & Statistics

Sophisters | Trinity Term 2016

ST3452: Applied Linear Statistical Methods II

20 May 2016 | SPORTS CENTRE / GV | 14:00 – 16:00

Prof. Rozenn Dahyot

Instructions to Candidates:

You may not start this examination until you are instructed to do so by the Invigilator.

Answer all questions.

Materials permitted for this examination:

Non-programmable calculators are permitted for this examination — please indicate the make and model of your calculator on each answer book used.

1. (a) For each distribution given in table 1:

   i. Show this is a distribution.

   ii. Show this is a member of the exponential family of distributions.

   iii. Compute $\mathbb{E}[y]$. For the Weibull distribution, show that $\mathbb{E}[y] = \left(\frac{1}{\theta}\right)^{1/\lambda} \Gamma(1+\frac{1}{\lambda}))$. with $\Gamma(u) = \int_0^{+\infty} s^{u-1}\ \exp(-s)\ ds$.

   iv. Define what a link function is. Propose a link function for each distribution in table 1.

[30 marks]

(b) Explain the relation between the following distributions:

   i. Bernouilli and Binomial.

   ii. Poisson and Binomial.

   iii. Weibull and Exponential.

[10 marks]

(c) With the Weibull distribution, define and compute:

   i. the median survival time.

   ii. the hazard function.

[10 marks]

Poisson	Binomial	Weibull
$p_{y\vert\theta}(y\vert\theta) = \frac{\theta^y \exp(-\theta)}{y!}$ $y \in \mathbb{N}$ and $\theta \in \mathbb{R}^+/\{0\}$	$p_{y\vert\theta}(y\vert\theta, n) = \frac{n!}{(n-y)!y!}\ \theta^y (1-\theta)^{n-y}$ $y \in \{0, 1, \cdots, n\}$ and $\theta \in (0;1)$	$p_{y\vert\lambda\theta}(y\vert\lambda, \theta) = \lambda\ \theta\ y^{\lambda-1} \exp\left[-\theta\ y^{\lambda}\right]$ $y \in \mathbb{R}^+$ $\theta \in \mathbb{R}^+/\{0\}$ $\lambda \in \mathbb{R}^+/\{0\}$

Table 1: Distributions.

(50 marks)

2. Table 2 presents the Beetle mortality data: the variable x_i represents the dose level of toxic substance given to a group i of n_i beetles where k_i beetles died consequently. The likelihood of the proportion θ_i of dead beetles in the group i is modelled using the Binomial distribution.

   (a) What is the likelihood of the saturated model parameterised by $(\theta_1, \cdots, \theta_8)$? Explain the assumptions used.

   [3 marks]

   (b) What are the maximum likelihood estimates $\hat{\theta}_i, \forall i = 1, \cdots, 8$ for the saturated model?

   [3 marks]

   (c) A logit link function is used to model the mean of the response with the explanatory variable (dose level x). Define mathematically the logit link function and this GLM.

   [3 marks]

   (d) Write the log-likelihood function with respect to the coefficients of this GLM using the logit link function.

   [4 marks]

   (e) The R output of this logistic regression is given in table 3. Identify the estimates of the coefficients of this GLM and compute the estimates $\hat{\theta}_1$ and $\hat{\theta}_5$.

   [4 marks]

   (f) Give the definition of the deviance. Discuss the goodness of fit of this logistic model (c.f. Tab. 3 and Tab. 4).

   [5 marks]

   (g) Propose another GLM that could have been tested for this dataset.

   [3 marks]

   (h) When having several GLMs available to analyse the same dataset, what criterion can be used to select the best GLM? Explain your answer.

   [5 marks]

(30 marks)

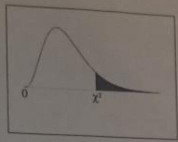
group	x_i	n_i	k_i
$i=1$	1.6907	59	6
$i=2$	1.7242	60	13
$i=3$	1.7552	62	18
$i=4$	1.7842	56	28
$i=5$	1.8113	63	52
$i=6$	1.8369	59	53
$i=7$	1.8610	62	61
$i=8$	1.8839	60	60

Table 2: Beetle mortality data.

```
Call:
glm(formula = beetle.mat ~ x, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-1.5941  -0.3944   0.8329   1.2592   1.5940

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.717      5.181  -11.72   <2e-16 ***
x             34.270      2.912   11.77   <2e-16 ***
---
Signif. codes:  0'***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.232  on 6  degrees of freedom
AIC: 41.43

Number of Fisher Scoring iterations: 4
```

Table 3: R output (logit).

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Table 4: χ^2 distribution.

3. (a) Explain the relation between the Binomial and Multinomial distributions. [5 marks]

   (b) Is the Multinomial distribution a member of the exponential family of distributions? Explain. [5 marks]

   (c) Explain how the Multinomial can be used in practice. Illustrate your comments with an example. [10 marks]

(20 marks)