PRIFYSGOL ABERYSTWYTH UNIVERSITY

ADRAN MATHEMATEG / DEPARTMENT OF MATHEMATICS

ARHOLIADAU SEMESTER 2 / SEMESTER 2 EXAMINATIONS

MAI – MEHEFIN / MAY – JUNE 2026

MA36820 – Statistical Inference and Linear Models

The questions on this paper are written in English.

Amser a ganiateir - 2 awr

- Gellir rhoi cynnig ar bob cwestiwn.
- Rhoddir mwy o ystyriaeth i berfformiad yn Rhan B wrth bennu marc dosbarth cyntaf.
- Cyfrifianellau Casio FX-83 neu FX-85 YN UNIG a ganiateir.
- Darperir tablau ystadegol.
- Mae taflen fformiwlâu wedi cael ei hatodi.
- Mae modd i fyfyrwyr gyflwyno atebion i'r papur hwn naill ai yn y Gymraeg neu'r Saesneg.

Time allowed - 2 hours

- All questions may be attempted.
- Performance in Section B will be given greater consideration in assigning a first class mark.
- Casio FX-83 or FX-85 calculators ONLY may be used.
- Statistical tables will be provided.
- A formula sheet is attached.
- Students may submit answers to this paper in either Welsh or English.

© 2026 Aberystwyth University

Section A

1. A random variable K follows the Binomial distribution $\mathrm{Bin}(20, p)$, where $0 < p < 1$ is an unknown parameter. A single observation of K was taken, with outcome $k = 15$.

(a) Write down the likelihood and the score function, and hence find the Maximum Likelihood Estimate (MLE) of p. [6 marks]

(b) A Bayesian statistician's prior beliefs for p are that it is equally likely to take any value in $(0, 1)$. Show that their posterior distribution for p is a member of the Beta family, and state the relevant parameters. [4 marks]

(c) Find the posterior mean, and confirm that it lies between the prior mean and the MLE. [3 marks]

2. The proportion p of a large electorate support candidate A in an election. Randomly selected members of the electorate were interviewed one by one until 5 supporters of A were found; this took 6 interviews.

(a) Quoting the appropriate distribution, show that the likelihood of p (based on the above information) is proportional to $p^5(1-p)$. [2 marks]

(b) Previous polling suggests a conjugate prior with mean $4/5$ and variance $4/275$. Find the parameters of this prior. [3 marks]

(c) Identify the posterior distribution as a member of the conjugate family. [3 marks]

(d) Find the posterior pdf, including the normalisation constant. [2 marks]

(e) Two additional members of the electorate are to be interviewed. Find the predictive probability that both of them support candidate A. [3 marks]

3. Random variables X and Y have their joint probability density function

$$f(x, y) = c\, e^{-2x^2 - \frac{1}{6}y^2 - xy - x}, \quad (x, y) \in \mathbb{R}^2$$

where c is the appropriate constant.

(a) Show that $\begin{pmatrix} X & Y \end{pmatrix}^T$ has dispersion matrix $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -3 \\ -3 & 12 \end{pmatrix}$, find the mean vector $\boldsymbol{\mu}$, and hence identify the distribution of $\begin{pmatrix} X & Y \end{pmatrix}^T$. [6 marks]

(b) Show that $5X + Y$ is distributed independently of $3X + 2Y$. [2 marks]

4. Let X, Y and Z be independent standard Normal variables, and

$$Q = X^2 + 9Y^2 + 25Z^2 - 6XY + 10XZ - 30YZ.$$

Show that Q is a multiple of a chi-squared random variable. [6 marks]

5. A linear model with design matrix $\mathbf{X}$ and parameter vector $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix}^T$ was fitted to thirty independent Normal homoscedastic observations with common variance $\sigma^2 = 5$, leading to

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 3 & -4 \\ -4 & 7 \end{pmatrix}, \qquad \mathbf{X}^T\mathbf{y} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Consider the parameter vector $\boldsymbol{\phi} = \begin{pmatrix} \phi_1 & \phi_2 \end{pmatrix}^T$, where $\phi_1 = \beta_1 + \beta_2$ and $\phi_2 = \beta_1 - \beta_2$.

(a) Find (i) the OLS estimates of the parameters β_1 and β_2, and (ii) the dispersion matrix of the OLS estimator. [4 marks]

(b) Find (i) the BLUE $\widehat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$, and (ii) the dispersion matrix of $\widehat{\boldsymbol{\phi}}$. [3 marks]

(c) Obtain a 70% equal-tailed confidence interval for ϕ_1. [4 marks]

(d) Obtain a 70% confidence region for $\boldsymbol{\phi}$. [6 marks]

6. A manufacturer produces three types of lightbulbs, A, B, and C, with expected lifetimes β_1, $\beta_1 + \beta_2$, and $\beta_1 + \beta_2 + \beta_3$, respectively, where $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \end{pmatrix}^T$ is an unknown parameter vector. The lifetimes $Y_1, \ldots Y_{13}$ of 13 bulbs are observed, where the first five are type A, the next six B, and the last two C. The observations are assumed independent, homoscedastic and Normal.

We wish to test whether the expected lifetimes of the A and B bulbs are significantly different.

(a) State the null hypothesis H_0 for this test. [2 marks]

(b) Write down the design matrix for (i) the full model, and (ii) the reduced model associated to the null hypothesis. [3 marks]

(c) Suppose that the residual sums of squares (RSS) for the two models turn out 182 and 280.2, respectively. Carry out the test, displaying the analysis in an appropriate ANOVA table, and stating your conclusion using 5% significance level. [8 marks]

Section B

7. The lifetime, X, of an electronic appliance (of certain type) is found to follow the distribution whose pdf is given by

$$f(x;\theta) = 3x^2\theta e^{-\theta x^3}, \quad x > 0,$$

where $\theta > 0$ is a parameter. Consider a random sample of n such appliances with lifetimes $X_1, X_2, \ldots, X_n$.

(a) Show that the likelihood of θ is proportional to $\theta^n e^{-\theta S}$, where S is a suitable statistic (to be found). [3 marks]

(b) Find the Fisher information of θ. [2 marks]

(c) Let $n = 60$. Use Bayesian approach with Jeffreys prior, to find (in terms of S) the 80% credible posterior upper bound for θ. [7 marks]

8. Cars pass a point on a road at an average rate of λ per minute. Measured from the start of an observation period, let S be the time (in mins) of the 8th pass; then $S = \sum_{i=1}^{8} X_i$ where $X_1, \ldots, X_8$ is a random sample from $\mathrm{Expon}(\lambda)$. Consider the test of hypotheses $H_0 : \lambda = 10$ versus $H_1 : \lambda = 5$, using S as the test statistic.

(a) Find the likelihood, and write down the likelihood ratio in terms of S. [3 marks]

(b) State the distribution of S, and relate it to a chi-squared distribution. [2 marks]

(c) Show that the critical region of the most powerful test of size 0.05 is $\{S > k\}$ where k is a constant to be determined. [5 marks]

9. Consider a sample consisting of three independent variables $Y_1 \sim \mathrm{Expon}(\theta_1)$, $Y_2 \sim \mathrm{Expon}(\theta_1\theta_2)$, and $Y_3 \sim \mathrm{Expon}(\theta_2)$, where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is an unknown parameter vector. Find the Fisher Information matrix of $\boldsymbol{\theta}$. [9 marks]

(please turn page)

10. In a physics experiment, an observable quantity Y depends linearly on three control variables x_1, x_2, x_3, which can be adjusted by the experimental setup:

$$\mathbb{E}[Y] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

The quantity Y is assumed to Normally distributed, and different runs of the experiment are independent, leading to homoscedastic observations Y_i. In twelve runs, the results were as follows:

	y	x1	x2	x3		y	x1	x2	x3		y	x1	x2	x3
1	74	8	-1	1	5	15	2	0	1	9	45	7	-4	0
2	24	3	-4	1	6	72	6	-2	0	10	91	9	-2	1
3	132	13	2	0	7	23	4	3	1	11	70	7	3	0
4	150	17	-10	1	8	161	14	2	0	12	24	3	1	1

It follows that $\mathrm{TSS} = 93617$, $\mathrm{RSS} = 1239.8$, and

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 0.0016 & 0.0008 & -0.0089 \\ 0.0008 & 0.0074 & 0.0084 \\ -0.0089 & 0.0084 & 0.2171 \end{pmatrix}$$

where $\mathbf{X}$ consists of the columns of the data table corresponding to the control variables. Fitting the model to the data produces the following R output (with P-values omitted):

```
Coefficients:
   Estimate Std. Error t value Pr(>|t|)
x1  10.3231        [?]  22.113
x2   1.4381     1.0065   1.429
x3  -7.8812     5.4683     [?]
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: [?] on [?] degrees of freedom
Multiple R-squared:  [?], Adjusted R-squared:  0.9823
F-statistic: 223.5 on 3 and 9 DF,  p-value:
```

(a) Find the five missing values indicated by [?] in the R output. [5 marks]

(b) Test whether β_2 is significantly different from zero at the 5% level. Make your conclusion based on the R output and statistical tables. (No further calculations/proofs required.) [4 marks]

(c) Calculate the leverage of the 4th observation. Is it unusually high? [5 marks]

(d) The ith *standardised residual* is $\hat{e}_i = \hat{\epsilon}_i / \sqrt{\hat{\sigma}^2(1 - H_{ii})}$, where $\hat{\epsilon}_i$ is the ordinary residual and H_{ii} the leverage. Calculate $\hat{e}_4$. Is it unusual? [5 marks]

Standard Distributions and their Recognition

Name	Notation	Range	Kernel	Constant	Mean	Mode	Variance	Comments
Beta	$\beta(a,b)$	$0<z<1$	$z^{a-1}(1-z)^{b-1}$	$\frac{1}{B(a,b)}=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$	$\frac{a}{(a+b)}$	$\frac{a-1}{a+b-2}$	$\frac{ab}{(a+b)^2(a+b+1)}$	$1-Z\sim\beta(b,a)$ $\frac{bZ}{a(1-Z)}\sim F(2a,2b)$
Gamma	$Ga(a,b)$	$z>0$	$z^{a-1}e^{-bz}$	$\frac{b^a}{\Gamma(a)}$	$\frac{a}{b}$	$\frac{(a-1)}{b}$	$\frac{a}{b^2}$	$cZ\sim Ga(a,b/c)$ $2bZ\sim\chi^2(2a)$
Chi-square	$\chi^2(k)$	$z>0$	$z^{(k-2)/2}e^{-z/2}$	$\frac{1}{2^{k/2}\Gamma(k/2)}$	k	$(k-2)$	$2k$	$\chi^2(k)\equiv Ga(k/2,1/2)$
Inverse chi-square	$\chi^{-2}(k)$	$z>0$	$z^{-(k-2)/2}e^{-1/(2z)}$	$\frac{1}{2^{k/2}\Gamma(k/2)}$	$\frac{1}{(k-2)}$	$\frac{1}{(k+2)}$	$\frac{2}{(k-2)^2(k-4)}$	$Z^{-1}\sim\chi^2(k)$
Inverse chi	$\chi^{-1}(k)$	$z>0$	$z^{-k-1}\exp\{-1/(2z^2)\}$	$\frac{1}{2^{\frac{1}{2}k-1}\Gamma(k/2)}$	$\frac{\Gamma\left(\frac{k-1}{2}\right)}{2^{\frac{1}{2}}\Gamma\left(\frac{k}{2}\right)}$	$\frac{1}{(k+1)^{1/2}}$	$\frac{1}{(k-2)}-(\text{mean})^2$	$Z^{-2}\sim\chi^2(k)$
Normal	$N(a,b^2)$	all z	$\exp[-\frac{1}{2}\{(z-a)/b\}^2]$	$\{2\pi b^2\}^{-1/2}$	a	a	b^2	$\frac{(Z-a)}{b}\sim N(0,1)$
Standard Normal	$N(0,1)$	all z	$\exp\{-z^2/2\}$	$\{2\pi\}^{-1/2}$	0	0	1	
(Student's) t	$t(k)$	all z	$1\Big/\left\{1+\frac{z^2}{k}\right\}^{(k+1)/2}$	$\frac{\Gamma\left(\frac{k+1}{2}\right)}{(k\pi)^{\frac{1}{2}}\Gamma\left(\frac{k}{2}\right)}$	0	0	$\frac{k}{(k-2)}$	
(Fisher's) F	$F(a,b)$	$z>0$	$\frac{z^{\frac{1}{2}a-1}}{\left(z+\frac{b}{a}\right)^{\frac{1}{2}(a+b)}}$	$\left(\frac{b}{a}\right)^{b/2}\frac{\Gamma\left(\frac{a+b}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{b}{2}\right)}$	$\frac{b}{(b-2)}$	$\frac{b(a-2)}{a(b+2)}$	$\frac{2b^2(a+b-2)}{a(b-2)^2(b-4)}$	$Z^{-1}\sim F(b,a)$ Mean>1>Mode with Median between Mode and Mean Median = 1 whenever $a=b$