IMPERIAL

MATH50011

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2024

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Statistical Modelling 1

Date: Monday, May 13, 2024

Time: 10:00 – 12:00 (BST)

Time Allowed: 2 hours

This paper has 4 Questions.

Please Answer Each Question in a Separate Answer Booklet

Candidates should start their solutions to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

1. (a) Let γ be an unknown parameter belonging to a parameter space $\Gamma \subset \mathbb{R}$. Let $(S_n)_{n\in\mathbb{N}}$ be a sequence of estimators for γ. Provide the definition of consistency for the sequence of estimators $(S_n)_{n\in\mathbb{N}}$ with respect to the parameter $\gamma \in \Gamma$. (4 marks)

(b) Let θ be an unknown parameter belonging to a parameter space $\Theta \subset \mathbb{R}$. Let $(T_n)_{n\in\mathbb{N}}$, be a sequence of estimators for γ. Show that asymptotic normality of $(T_n)_{n\in\mathbb{N}}$ for θ implies consistency of $(T_n)_{n\in\mathbb{N}}$ for θ. (5 marks)

(c) Let $X_1, ..., X_n$, where $n \in \mathbb{N}$, be iid random variables where X_1 follows an exponential distribution with unknown parameter $\lambda > 0$, that is $X_1 \sim Exp(\lambda)$. Recall that the pdf of X_1 is $f_{X_1}(x;\lambda) = \lambda e^{-\lambda x}$ where $x > 0$. Denote by F_{X_1} the cdf of X_1 and let $p \in (0,1)$. Let $\mu_p > 0$ be the quantile of order p, that is $F_{X_1}(\mu_p;\lambda) = p$. Compute the maximum likelihood estimator $\hat{\mu}_p$. (5 marks)

(d) State and prove the Delta method. (6 marks)

(Total: 20 marks)

2. (a) Let T be an unbiased estimator of θ with parameter space $\mathbb{R}$. Let $h : \mathbb{R} \to \mathbb{R}$ be a linear function. Show that $h(T)$ is an unbiased estimator of $h(\theta)$. (4 marks)

(b) What does it mean for a statistical model to be identifiable? Further, for which values of $\alpha \in \mathbb{N}$ is the statistical model $N(\theta^\alpha, 1)$, with $\theta \in \mathbb{R}$, identifiable? Motivate your answer. (4 marks)

(c) State and prove Markov's inequality. (4 marks)

(d) Let $(Y_1, X_1), ..., (Y_n, X_n)$, where $n \in \mathbb{N}$, be iid bivariate random variables. Assume that Y_1 follows a Poisson distribution with parameter λ and that $X_1|Y_1 = y_1$ follows a Binomial distribution with parameters $y_1 \in \mathbb{N}$ and $p \in (0, 1)$. That is the pdf of (Y_1, X_1) is

$$f_{X_1,Y_1}(y_1, x_1; \lambda, p) = \frac{\lambda^{y_1}}{y_1!} e^{-\lambda} \binom{y_1}{x_1} p^{x_1} (1-p)^{y_1 - x_1}.$$

(i) Compute the maximum likelihood estimator for $\theta = (\lambda, p)$. (3 marks)

(ii) Assume that mild regularity conditions hold. Consider the region

$$S_c = \left\{ (\lambda, p) : \frac{n(\bar{Y} - \lambda)^2}{\lambda} + \frac{n\lambda(\frac{\bar{X}}{\bar{Y}} - p)^2}{p(1-p)} < c \right\}$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. For which values of c the region S_c is an approximate confidence region for θ of level 0.95? (Hint: you might want to use that $E[X_1] = E[pY_1]$.) (5 marks)

(Total: 20 marks)

3. (a) Suppose $X \sim N\left(\begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$.

(i) What is the distribution of $Z = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} X + \begin{pmatrix} -1 \\ -3 \\ 2 \end{pmatrix}$? (3 marks)

(ii) Are any of the components of Z independent? If yes, which are them? Motivate your answer. (3 marks)

(b) Consider a linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $E[\boldsymbol{\varepsilon}] = 0$. Define the full rank (FR) assumption, the second order assumption (SOA), and the Normal theory assumption (NTA). (5 marks)

(c) Consider the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $E[\boldsymbol{\varepsilon}] = 0$. Assume the Second Order Assumption (SOA) and the Full Rank (FR) assumption hold. State and prove the Gauss-Markov Theorem for full-rank linear models [If in the proof you use other results seen in class please state them clearly, you do not need to prove them]. (4 marks)

(d) Let X be a Bernoulli random variable with unknown parameter $\theta \in (0, 1)$, that is $P_\theta(X = 1) = \theta$ and $P_\theta(X = 0) = 1 - \theta$. Consider the random interval:

$$[L, U] = \begin{cases} [0, 1-\alpha], & \text{for } X = 0 \\ [\alpha, 1], & \text{for } X = 1. \end{cases}$$

For every $\theta_0 \in (0, 1)$, construct a test of level α (with $\alpha < 1/2$) for the hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ based on $[L, U]$ and draw the power function of such a test.

(5 marks)

(Total: 20 marks)

4. Consider the linear regression model

$$Y_i = \beta_1 + w_i\beta_2 + \varepsilon_i,$$

where ε_i is an unobserved error term and $i = 1, ..., n$ where $n = 83$. Assume the normal theory assumption (NTA) and the full rank (FR) assumption.

(a) The model can be written in the matrix form notation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Define the terms $\mathbf{Y}$, $\boldsymbol{\beta}$, $\mathbf{X}$, and $\boldsymbol{\varepsilon}$. (4 marks)

(b) Assume that we have

$$\frac{1}{n}\sum_{i=1}^{n} y_i = 2, \quad \frac{1}{n}\sum_{i=1}^{n} w_i = 2, \quad \frac{1}{n}\sum_{i=1}^{n} w_i^2 = 5, \quad \frac{1}{n}\sum_{i=1}^{n} w_i y_i = 5, \quad \frac{1}{n}\sum_{i=1}^{n} y_i^2 = 30.$$

Show that the least squares estimates are $\hat{\beta}_1 = 0$ and $\hat{\beta}_2 = 1$. (5 marks)

(c) Consider the estimator $\hat{\sigma}^2$ for σ^2 given by $\hat{\sigma}^2 := \frac{RSS}{n-2}$, where RSS is the residual sum of squares. Show that

$$\hat{\sigma}^2 = \frac{83}{81}25.$$

(4 marks)

(d) Compute the coefficient of determination R^2. (3 marks)

(e) Test the null hypothesis $H_0 : \beta_2 \leq 0$ vs $H_1 : \beta_2 > 0$ using a one-sided t-test. Can you reject H_0 at 97.5% confidence level? (Hint: $P(|Z| \geq 1.96) = 0.05$ where $Z \sim N(0, 1)$.) (4 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2024

This paper is also taken for the relevant examination for the Associateship.

MATH50011

MATH50011 (Solutions)

Setter's signature	Checker's signature	Editor's signature
....	....	....

1. (a) It means that $S_n \xrightarrow{p} \gamma$, as $n \to \infty$, for every $\gamma \in \Gamma$. seen ⇓

(b) In the notes we have defined asymptotically normality as follows. A sequence T_n of estimators for $\theta \in \mathbb{R}$ is called *asymptotically normal* if 4, A

seen ⇓

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$$

for some $\sigma^2(\theta)$. It is possible to observe that $T_n - \theta = \frac{1}{\sqrt{n}}\sqrt{n}(T_n - \theta)$. Since $\frac{1}{\sqrt{n}} \to 0$ and $\sqrt{n}(T_n - \theta) \to N(0, \sigma^2(\theta))$, by Slutsky's lemma we have that $T_n - \theta \xrightarrow{d} 0 \cdot N(0, \sigma^2(\theta)) = 0$. Since convergence in distributions to a constant is equivalent to convergence in probability to that constant, we have $T_n - \theta \xrightarrow{p} 0$. Thus, we have that $T_n \xrightarrow{d} \theta$ as $n \to \infty$, for every $\theta \in \mathbb{R}$. 5, B

(c) First, observe that $F_{X_1}(\mu_p; \lambda) = 1 - e^{-\lambda \mu_p}$. Then using $F_{X_1}(\mu_p; \lambda) = p$ we obtain that $1 - e^{-\lambda\mu_p} = p$ which implies that $\mu_p = -\frac{1}{\lambda}\log(1-p)$. It is possible to see that the function $g(x) = -\frac{1}{x}\log(1-p)$ is a bijective function from $(0, \infty)$ to $(0, \infty)$. Further, by results from lectures we know that the MLE for λ is $\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}$. Then, by the functional invariance property of the MLE we obtain that the MLE $\hat{\mu}_p$ is given by $\hat{\mu}_p = -\frac{1}{\hat{\lambda}}\log(1-p)$, that is sim. seen ⇓

$$\hat{\mu}_p = -\frac{\sum_{i=1}^n x_i}{n}\log(1-p)$$

5, C

(d) The statement is the following. Suppose that T_n is an asymptotically normal estimator of θ with unseen ⇓

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)). \tag{1}$$

Let $g : \Theta \to \mathbb{R}$ be a differentiable function with $g'(\theta) \neq 0$. Then,

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2\sigma^2(\theta)). \tag{2}$$

Let us now prove it. By the mean value theorem we have that

$$g(T_n) = g(\theta) + g'(\tilde{\theta})(T_n - \theta) \tag{3}$$

where $\tilde{\theta}$ lies in between T_n and θ. Since $T_n \xrightarrow{p} \theta$ (because asymptotic normality implies consistency), we have that $|\tilde{\theta} - \theta| \leq |T_n - \theta| \xrightarrow{p} 0$, which implies that $\tilde{\theta} \xrightarrow{p} \theta$, as $n \to \infty$. Since g is continuous, by the continuous mapping theorem we have

$$g(\tilde{\theta}) \xrightarrow{p} g(\theta). \tag{4}$$

By rearranging (3) and multiplying both sides by $\sqrt{n}$ we obtain

$$\sqrt{n}g(T_n) - g(\theta) = g'(\tilde{\theta})\sqrt{n}(T_n - \theta).$$

Then using (1) and (4), by Slutsky's theorem we conclude that

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2\sigma^2(\theta)). \tag{5}$$

Note: the students have seen the statement of the Delta method, but not its proof.

6, D

2. (a) The answer is true. Since the function h is linear we can write it as $h(x) = ax + b$ for some $a, b \in \mathbb{R}$. Then, using the linearity of the expectation and the unbiasedness of T we have

sim. seen ⇓

$$E[h(T)] = E[aT + b] = aE[T] + b = a\theta + b = h(\theta)$$

Since this holds for every $\theta \in \mathbb{R}$, we conclude that $h(T)$ is an unbiased estimator for $h(\theta)$.

4, A

(b) Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a statistical model with parameter space Θ. We say that $\mathcal{P}$ is identifiable if the mapping $\theta \mapsto P_\theta$ is one-to-one, that is $P_{\theta_1} = P_{\theta_2}$ implies that $\theta_1 = \theta_2$ for all $\theta_1, \theta_2 \in \Theta$.

meth seen ⇓

We have that $N(\theta^\alpha, 1)$ is identifiable for every α odd. Indeed, for α even we have that $N(\theta_1^\alpha, 1) = N(\theta_2^\alpha, 1)$ for every $\theta_2 = -\theta_1$.

4, B

(c) Statement: Let X be a random variable, then $P(|X| \geq a) \leq \frac{E[|X|]}{a}$ for $a > 0$.

seen ⇓

Proof: Since $a\mathbf{1}_{|X|\geq a} \leq |X|$, we have that $P(|X| \geq a) = E[\mathbf{1}_{|X|\geq a}] \leq \frac{1}{a}E[|X|]$

4, B

(d) (i) Since the sample is iid the likelihood function for θ is

meth seen ⇓

$$L(\theta) = \prod_{i=1}^{n} \frac{\lambda^{y_i}}{y_i!} e^{-\lambda} \binom{y_i}{x_i} p^{x_i} (1-p)^{y_i - x_i}.$$

Hence, the log likelihood is

$$l(\theta) = n\bar{y}\log\lambda - n\lambda + n\bar{x}\log p + (n\bar{y} - n\bar{x})\log(1-p).$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i$. By taking the first derivative wrt λ and setting it equal to zero we obtain

$$\frac{\partial l(\theta)}{\partial \lambda} = \frac{n\bar{y}}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \bar{y}.$$

While by taking the first derivative wrt p and setting it equal to zero we obtain

$$\frac{\partial l(\theta)}{\partial p} = \frac{n\bar{x}}{p} - \frac{(n\bar{y} - n\bar{x})}{1-p} = n\frac{\bar{x} - p\bar{y}}{p(1-p)} = 0 \Rightarrow \hat{p} = \frac{\bar{x}}{\bar{y}}.$$

Hence, $\hat{\theta} = (\hat{\lambda}, \hat{p}) = (\bar{y}, \frac{\bar{x}}{\bar{y}})$ is a candidate for the MLE. To check whether this is indeed the MLE we need to check that the Hessian matrix is negative-definite. In particular, the cross derivatives are zero and

$$\frac{\partial^2 l(\theta)}{\partial \lambda^2} = -\frac{n\bar{y}}{\lambda^2} < 0$$

and

$$\frac{\partial^2 l(\theta)}{\partial p^2} = -\frac{n\bar{x}}{p^2} - \frac{(n\bar{y} - n\bar{x})}{(1-p)^2} < 0$$

So, $\hat{\theta}$ is indeed a MLE.

3, C

(ii) From the previous point and using the fact that $E[Y_1] = \lambda$ (because Y_1 is Poisson with parameter λ) and that $E[X_1] = \lambda p$ using the hint, we have that $\boxed{\text{unseen} \Downarrow}$

$$E\Big[\frac{\partial^2 l(\theta)}{\partial\lambda\partial p}\Big] = E\Big[\frac{\partial^2 l(\theta)}{\partial p\partial\lambda}\Big] = E[0] = 0$$

$$E\Big[\frac{\partial^2 l(\theta)}{\partial\lambda^2}\Big] = E\Big[-\frac{n\bar{y}}{\lambda^2}\Big] = -\frac{n}{\lambda^2}E[\bar{Y}] = -\frac{n}{\lambda}$$

and

$$E\Big[\frac{\partial^2 l(\theta)}{\partial p^2}\Big] = E\Big[-\frac{n\bar{x}}{p^2} - \frac{(n\bar{y}-n\bar{x})}{(1-p)^2}\Big] = -\frac{n\lambda}{p} - \frac{n\lambda - n\lambda p}{(1-p)^2}$$

$$= -\frac{n\lambda}{p} - \frac{n\lambda}{1-p} = -\frac{n\lambda}{p(1-p)}$$

Thus, the Fisher information of (Y_1, X_1) is given by

$$I(\theta) = \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & \frac{\lambda}{p(1-p)} \end{pmatrix}$$

and so $\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}\begin{pmatrix} \bar{Y} - \lambda \\ \frac{\bar{X}}{\bar{Y}} - p \end{pmatrix} \xrightarrow{d} N(0, I^{-1}(\theta))$. This implies that

$$\frac{\sqrt{n}(\bar{Y} - \lambda)}{\sqrt{\lambda}} \xrightarrow{d} Z$$

and

$$\frac{\sqrt{n}(\frac{\bar{X}}{\bar{Y}} - p)}{\sqrt{\frac{p(1-p)}{\lambda}}} \xrightarrow{d} W$$

where Z and W are independent standard Normal. Here we are using the fact that Z and W are jointly normal and uncorrelated, and so they are independent. Therefore,

$$\frac{n(\bar{Y} - \lambda)^2}{\lambda} + \frac{n\lambda(\frac{\bar{X}}{\bar{Y}} - p)^2}{p(1-p)} \xrightarrow{d} \chi^2_2.$$

Thus, c is such that $P(Q > c) = 0.95$ where $Q \sim \chi^2_2$. $\boxed{5, \text{D}}$

3. (a) (i) meth seen ⇓

$$E(Z) = \begin{pmatrix}5\\3\\2\end{pmatrix} + \begin{pmatrix}-1\\-3\\2\end{pmatrix} = \begin{pmatrix}4\\0\\4\end{pmatrix}$$

$$Cov(Z) = \begin{pmatrix}1 & 1\\0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}1 & 0 & 1\\1 & 1 & 0\end{pmatrix} = \begin{pmatrix}2 & 1 & 1\\1 & 1 & 0\\1 & 0 & 1\end{pmatrix}$$

Thus,

$$Z \sim N\left(\begin{pmatrix}4\\0\\4\end{pmatrix}, \begin{pmatrix}2 & 1 & 1\\1 & 1 & 0\\1 & 0 & 1\end{pmatrix}\right)$$

3, A

(ii) Z_2 and Z_3 are independent, because they are uncorrelated and they are jointly normal. Indeed, meth seen ⇓

$$\begin{pmatrix}Z_2\\Z_3\end{pmatrix} = \begin{pmatrix}0 & 1 & 0\\0 & 0 & 1\end{pmatrix} Z \sim N\left(\begin{pmatrix}0\\4\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$$

3, A

(b) In a linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, the FR assumption consists in assuming that the matrix $\mathbf{X}$ has full rank, that is if $\mathbf{X}$ is a $n \times p$ matrix then $rank(\mathbf{X}) = \min(n, p)$. The SOA consists in assuming that $cov(\boldsymbol{\varepsilon}) = (cov(\varepsilon_i, \varepsilon_j))_{\substack{i=1,\dots,n\\j=1,\dots,n}} = \sigma^2 I_n$ for some $\sigma^2 > 0$, where I_n is the n-dimensional identity matrix. SOA really consists of two parts: First, that the errors of two different observations, ε_i and ε_j for $i \neq j$ are uncorrelated. Second, that the variance of all errors is identical (recall: $var(\varepsilon_i) = cov(\varepsilon_i, \varepsilon_i)$). The NTA consists in assuming that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$ for some $\sigma^2 > 0$. One could equivalently define the NTA as: $\varepsilon_1, \dots, \varepsilon_n \sim N(0, \sigma^2)$ independently. seen ⇓

5, A

(c) Statement: Assume (FR), (SOA). Let $\boldsymbol{c} \in \mathbb{R}^p$ and let $\hat{\boldsymbol{\beta}}$ be a least squares estimator of $\boldsymbol{\beta}$ in a linear model. Then the following holds: The estimator $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}$ has the smallest variance among all linear unbiased estimators for $\boldsymbol{c}^T\boldsymbol{\beta}$. seen ⇓

Proof: $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}$ is linear and unbiased

Let $\hat{\gamma} = \boldsymbol{L}^T\boldsymbol{Y}$ be any other linear unbiased estimator of $\boldsymbol{c}^T\boldsymbol{\beta}$.

$$\begin{aligned} var(\hat{\gamma}) =& var(\boldsymbol{L}^T\boldsymbol{Y}) = var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}} + \underbrace{(\boldsymbol{L}^T - \boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)}_{=:\boldsymbol{D}^T}\boldsymbol{Y}) \\ =& cov(\boldsymbol{c}^T\hat{\boldsymbol{\beta}} + \boldsymbol{D}^T\boldsymbol{Y}, \boldsymbol{c}^T\hat{\boldsymbol{\beta}} + \boldsymbol{D}^T\boldsymbol{Y}) \\ =& var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) + var(\boldsymbol{D}^T\boldsymbol{Y}) + 2cov(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}, \boldsymbol{D}^T\boldsymbol{Y}) \end{aligned}$$

$$cov(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}, \boldsymbol{D}^T\boldsymbol{Y}) = \boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\underbrace{cov(\boldsymbol{Y})}_{=\sigma^2 I_n}\boldsymbol{D} = \boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\underbrace{\boldsymbol{D}^T\boldsymbol{X}}_{\overset{(*)}{=}\mathbf{0}^T})^T\sigma^2 = 0.$$

To see $(*)$: both estimators are unbiased which implies that

$$0 = E(\hat{\gamma}) - E(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) = E(\boldsymbol{D}^T\boldsymbol{Y}) = \boldsymbol{D}^T\boldsymbol{X}\boldsymbol{\beta}$$

for all $\boldsymbol{\beta}$. Hence, $\boldsymbol{D}^T\boldsymbol{X} = \mathbf{0}^T$. Thus,

$$var(\hat{\gamma}) = var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) + var(\boldsymbol{D}^T\boldsymbol{Y}) \geq var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}).$$

4, B

(d) Let $\theta_0 \in (0,1)$. By lectures we know that $[L, U]$ is confidence interval of level $1-\alpha$. A test of level α (with $\alpha < 1/2$) for the hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ based on $[L, U]$ is given by the following: we reject H_0 if $\theta_0 \notin [L(x), U(x)]$, where $[L(x), U(x)]$ is the observed confidence interval. unseen $\Downarrow$

The power function is given by $P_\theta(\theta_0 \notin [L, U])$ or equivalently by $P_\theta(X \in R)$, where R is the rejection region. It is possible to notice that depending on the value of θ_0 we have different power functions, namely different rejection regions. In particular, if $\theta_0 < \alpha$, then the rejection region is $\{1\}$, indeed the power function is given by

$$\beta(\theta) = P_\theta(\theta_0 \notin [L, U])$$
$$= P_\theta(\theta_0 \notin [L, U] | X = 1) P_\theta(X = 1) + P_\theta(\theta_0 \notin [L, U] | X = 0) P_\theta(X = 0)$$
$$= P_\theta(X = 1) = \theta$$

Thus, for $\theta_0 < \alpha$ we have that $\beta(\theta) = \theta$. Hence, since $\theta \in [0, 1]$, the power function $\beta(\theta)$ is a straight line from $(0, 0)$ to $(1, 1)$.

For $\theta_0 > 1 - \alpha$, by the same arguments we have that $\beta(\theta) = 1 - \theta$ and so the power function $\beta(\theta)$ is a straight line from $(0, 1)$ to $(1, 0)$.

Finally for $\theta_0 \in [\alpha, 1 - \alpha]$, by the same arguments we have that $\beta(\theta) = 0$. Thus, the power function is a straight line from $(0, 0)$ to $(1, 0)$. 5, D

4. (a) We have seen ⇓

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & w_1 \\ \vdots & \vdots \\ 1 & w_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

4, A

(b) From the information given in the question we have meth seen ⇓

$$\mathbf{X}^T\mathbf{X} = n \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}, \quad \text{and} \quad \mathbf{X}^T\mathbf{Y} = n \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Thus, the least squares estimator is

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

i.e. $\hat{\beta}_1 = 0$ and $\hat{\beta}_2 = 1$. 5, A

(c) Using that the residual e_i is given by $e_i = y_i - \hat{\beta}_1 - w_i\hat{\beta}_2$, we have meth seen ⇓

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{1}{n-2}\sum_{i=1}^n e_i^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{\beta}_1 - w_i\hat{\beta}_2)^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - w_i)^2$$

$$= \frac{1}{n-2}\sum_{i=1}^n (y_i^2 + w_i^2 - 2y_iw_i) = \frac{n}{n-2}\frac{1}{n}\sum_{i=1}^n (y_i^2 + w_i^2 - 2y_iw_i) = \frac{83}{81}(30+5-2\cdot 5) = \frac{83}{81}25.$$

4, A

(d) First, observe that meth seen ⇓

$$\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^n y_i^2 + \bar{y}^2 - 2\bar{y}\frac{1}{n}\sum_{j=1}^n y_j = \frac{1}{n}\sum_{i=1}^n y_i^2 - \bar{y}^2 = 30 - 4 = 26$$

Further, using that in point (c) we obtain that $RSS/n = 25$. Therefore, the coefficient of determination R^2 is given by

$$R^2 = 1 - \frac{RSS}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{RSS/n}{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{25}{26} = \frac{1}{26}.$$

3, B

(e) We have that part seen ⇓

$$Cov(\hat{\boldsymbol{\beta}}) = \frac{RSS}{n-r}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \frac{83}{81}25\frac{1}{83}\begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix} = \frac{25}{81}\begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$$

that is $std(\hat{\beta}_2) = \sqrt{25/81} = 5/9$. Then, the test statistic is

$$t = \frac{\hat{\beta}_2 - 0}{std(\hat{\beta}_2)} = \frac{1}{5/9} = \frac{9}{5} = 1.8.$$

Since $t \leq 1.96$ and since the student's t-distribution (with 81 degrees of freedom) have fatter tails than the Normal distribution, meaning that $P(|X| > c) \geq P(|Z| > c)$ for any c large enough where $X \sim t_{81}$ and $Z \sim N(0,1)$, then we cannot reject H_0 at level 97.5%. In particular, we cannot reject in the case of a standard Normal and so we cannot reject for the case of the student's t-distribution. 4, C

Review of mark distribution:

Total A marks: 32 of 32 marks

Total B marks: 20 of 20 marks

Total C marks: 12 of 12 marks

Total D marks: 16 of 16 marks

Total marks: 80 of 80 marks

Total Mastery marks: 0 of 20 marks

MATH50011 Statistical Modelling 1

Question	Marker's comment
4	Overall this question was answered very well, especially parts (a) - (c). A few errors made at specific parts were:- part (d): either the correct definition of R^2 was not given, or the incorrect value of RSS was used- part (e): a t-test was meant to be applied here, and given the large number of degrees of freedom the null distribution of the test-statistic could be approximated by a Normal (0,1) distribution.nbsp;