FAMILY NAME: ....
OTHER NAME(S): ....
STUDENT NUMBER: ....
SIGNATURE: ....

THE UNIVERSITY OF NEW SOUTH WALES

SCHOOL OF MATHEMATICS AND STATISTICS

Semester 2, 2015

MATH2089
Numerical Methods and Statistics

(1) TIME ALLOWED – 3 Hours

(2) TOTAL NUMBER OF QUESTIONS – 6

(3) ANSWER ALL QUESTIONS

(4) THE QUESTIONS ARE OF EQUAL VALUE

(5) THIS PAPER MAY **NOT** BE RETAINED BY THE CANDIDATE

(6) **ONLY** CALCULATORS WITH AN AFFIXED “UNSW APPROVED” STICKER MAY BE USED

(7) STATISTICAL FORMULAE ARE ATTACHED AT END OF PAPER
STATISTICAL TABLES ARE ATTACHED AT END OF PAPER

Part A – Numerical Methods consists of questions 1 – 3

Part B – Statistics consists of questions 4 – 6

Both parts must be answered

All answers must be written in ink. Except where they are expressly required pencils may only be used for drawing, sketching or graphical work.

Part A – Numerical Methods

1. Answer in a separate book marked Question 1

a) Give the results of the following MATLAB commands.

i)
```
h=1e-14;
ans1 = 1000 + h - 1000
```

ii)
```
ans2 = [1, 0]./[0, 0]
```

b) You are given that

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix},$$

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

For the matrix

$$K = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

find the following norms and condition numbers.

i) Find $||K||_1$

ii) Find $||K||_2$

iii) Find $||K||_\infty$

iv) Find $\kappa_1(K)$

v) Find $\kappa_2(K)$

vi) Find $\kappa_\infty(K)$

The computational complexity of some common operations with n by n matrices are given in Table 1.1

Operation	Flops
Matrix multiplication	$2n^3$
LU factorization	$\frac{2n^3}{3} + O(n^2)$
Cholesky factorization	$\frac{n^3}{3} + O(n^2)$
Back/forward substitution	$n^2 + O(n)$
Tridiagonal solve	$8n + O(1)$

Table 1.1: Flops for some operations with n by n matrices

In each of the remaining parts of this question, a claim is made. For each claim, state whether the claim is true or false (1 mark), and give a short reason for your answer (1 mark).

c) **Claim:** The MATLAB expression

```
ans1 = 1 / 2 * pi
```

correctly evaluates the constant $\dfrac{1}{2\pi}$.

d) You are using a computer with IEEE floating point arithmetic.
Claim: The the value of

```
ans2 = 1/10
```

is stored with relative error 0.

e) **Claim:** The following MATLAB statements accurately estimate the derivative $f'(x)$ (assume the function `f` has been previously defined):

```
x = 200;
h = eps;
ans3 = (f(x+h) - f(x-h)) / (2*h)
ans3 =
    0
```

f) Orthogonal matrices Q, which by definition satisfy $Q^TQ = I$, are widely used in numerical methods and least squares problems.
Claim: One reason for this is that orthogonal matrices have the ideal condition number $\kappa_2(Q) = 1$.

g) You are given that for two large n by n matrices, A and B, with no special structure, calculating AB takes 1000 seconds.
Claim: For an $n \times n$ symmetric positive definite matrix M, solving the linear system $M\mathbf{x}_i = \mathbf{b}_i$ for three different right-hand-side vectors $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$, will take more than 500 seconds.

h) You are given that

- the Secant method for solving $f(x) = 0$ has order of convergence $\nu \approx 1.6$,
- the error $e_k = |x^* - x_k|$ on the kth iteration is $e_k = 1.02 \times 10^{-12}$.

Claim: The error on the next iteration will be $e_{k+1} \approx 3.96 \times 10^{-21}$.

i) You are given that

- A and $\mathbf{b}$ are computed to full double precision accuracy,
- $\kappa(A) = 1.12 \times 10^4$,

Claim: The computed solution to $A\mathbf{x} = \mathbf{b}$ has at least 11 significant figures.

2. Answer in a separate book marked Question 2

a) Let $f : [0, 1] \to \mathbb{R}$ be defined by

$$f(x) = \frac{8}{\pi}\sqrt{x(1-x)}.$$

The function f is a probability density function so

$$I(f) = \int_0^1 f(x)\,\mathrm{d}x = 1.$$

Approximations to $I(f)$ were calculated using the Trapezoidal rule, Simpson's rule and the Gauss-Legendre rule, giving the following table of errors $E_N(f) = I(f) - Q_N(f)$:

N	E_N Trapezoidal	E_N Simpson	E_N Gauss-Legendre
32	5.8377e-03	2.2902e-03	-1.5288e-05
64	2.0659e-03	8.0865e-04	-1.9545e-06
128	7.3076e-04	2.8571e-04	-2.4714e-07
256	2.5842e-04	1.0098e-04	-3.1072e-08
512	9.1377e-05	3.5695e-05	-3.8953e-09
1024	3.2309e-05	1.2619e-05	-4.8762e-10

i) Find a linear transformation $x = \alpha + \beta z$ that maps $z \in [-1, 1]$ to $x \in [0, 1]$.

ii) **Given** the nodes $z_j, j = 1, \ldots, N$ and weights $w_j, j = 1, \ldots, N$ for the Gauss-Legendre rule for the interval $[-1, 1]$, how can you approximate $I(f)$?

iii) The error for the Trapezoidal rule satisfies

$$E_N^{\text{Trap}}(f) = O(N^{-2}), \tag{2.1}$$

provided that $f \in C^2([0, 1])$. You do **not** need to prove this.

A) Use (2.1) to estimate the ratio

$$\frac{E_N^{\text{Trap}}(f)}{E_{2N}^{\text{Trap}}(f)}. \tag{2.2}$$

B) Use the table of errors to estimate the ratio (2.2) when $N = 512$.

C) Is the table of errors consistent with the theoretical error estimate in (2.1)?

iv) You are **given** that for $x \in (0, 1)$,

$$f'(x) = \frac{4}{\pi}\frac{1-2x}{\sqrt{x(1-x)}}.$$

Explain why $I(f)$ is difficult to approximate.

v) Suppose now that the function f above is replaced by the function $g : [0, 5] \to \mathbb{R}$

$$g(x) = \frac{8}{25\pi}\sqrt{x(5-x)}$$

and the integral I is replaced by

$$J(g) = \int_0^5 g(x)\,\mathrm{d}x,$$

and a new table of errors, $E_N(g) = J(g) - Q_N(g)$, is computed. Describe how you expect the new table of errors for g to compare to the old table of errors for f.

b) The motion of a damped mass-spring system is modelled by the initial value problem

$$my'' + cy' + ky = 0, \qquad y(1) = 1, \quad y'(1) = -1,$$

where $y(t)$ is the displacement of the block at time t, m is the mass of the block, c is the damping coefficient, and k is the spring constant. Consider here the case

$$m = 2, \quad c = 4, \quad \text{and} \quad k = 6.$$

i) What is the order of the differential equation?

ii) Convert this ordinary differential equation into a system

$$\boldsymbol{x}' = \boldsymbol{f}(t, \boldsymbol{x}), \quad \text{for } t > t_0,$$

of first order differential equations.

iii) What is the initial condition $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$?

iv) Write

- **EITHER** a MATLAB anonymous function `myode`
- **OR** a MATLAB function M-file `myode.m`

to evaluate the vector valued function $\boldsymbol{f}(t, \boldsymbol{x})$.

v) Use Euler's method with a step of $h = 0.1$ to estimate $\boldsymbol{x}(1.1)$.

3. Answer in a separate book marked Question 3

The temperature $u(x,t)$ in an insulated wire of length L satisfies the partial differential equation

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2}, \quad x \in (0,L), \quad t \in (0,T]. \tag{3.1}$$

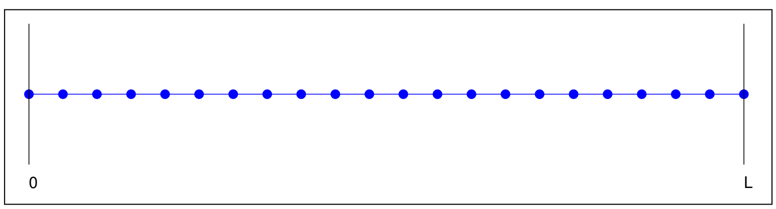
Let u_j^ℓ be your discrete finite difference approximation to the exact continuous solution $u(x_j, t_\ell)$ of (3.1) at the grid points

$$\begin{aligned} x_j &= j\Delta x \quad \text{for } j = 0,\dots,n+1, \\ t_\ell &= \ell\Delta t \quad \text{for } \ell = 0,\dots,m, \end{aligned}$$

where $L = (n+1)\Delta x$ and $T = m\Delta t$. At each time-step you solve the linear system:

$$(\boldsymbol{u}^{\ell+1} - \boldsymbol{u}^\ell)/\Delta t = -A\boldsymbol{u}^{\ell+1}/(\Delta x^2) \tag{3.2}$$

The space grid and the non-zero entries in the $n \times n$ *integer* coefficient matrix A are in Figure 3.1 for $n = 20$.



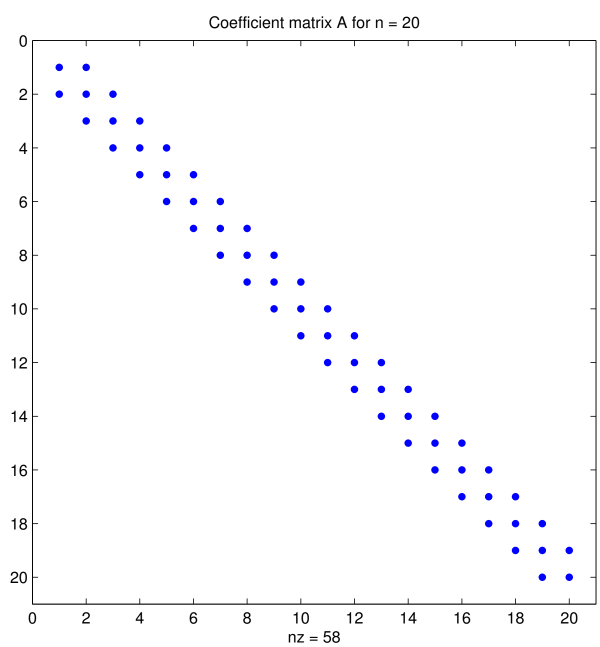


Figure 3.1: Grid and non-zero elements in coefficient matrix

You are **given** the following standard finite difference approximations for a function f of **one** variable:

$$
\begin{aligned}
f'(x) &= \frac{f(x+h)-f(x)}{h} + O(h), \\
f'(x) &= \frac{f(x+h)-f(x-h)}{2h} + O(h^2), \\
f''(x) &= \frac{f(x+h)-2f(x)+f(x-h)}{h^2} + O(h^2).
\end{aligned}
$$

The right hand side of (3.2), $-A\boldsymbol{u}^{\ell+1}/(\Delta x^2)$, is the above standard central difference approximation of $O(\Delta x^2)$ to the second derivative at the spatial grid points.

a) What additional information is needed to completely specify this problem?

b) The *jth* component of the left hand side of (3.2),

$$(u_j^{\ell+1} - u_j^{\ell})/\Delta t$$

is a standard finite difference approximation to the time derivative at the grid points for the continuous problem (3.1):

$$\frac{\partial u(x_j, t_{l+1})}{\partial t}.$$

What is the order of accuracy of this approximation ?

c) Do the finite differences in (3.2) lead to an *implicit method* or to an *explicit method*?

d) Give one advantage of an implicit method compared to an explicit method.

e) You are now given that

$$u(0,t) = u(L,t) = 0, \qquad t \in (0,T].$$

i) For $n = 20$, calculate the sparsity of A.

ii) Do you expect A^{-1} to be very sparse? (Answer yes or no.)

iii) Give the missing numbers (represented by question marks). The second row of the matrix A is $\begin{bmatrix} \textbf{?} & \textbf{?} & \textbf{?} & 0 & \dots & 0 \end{bmatrix}$

iv) For the same matrix A, give the results of the following MATLAB commands:

```
v = [1:n]';
w = A*(v.^2);
ans = w(4)
```

v) You are now also given that

$$u(x,0) = 1 - \cos(x\frac{2\pi}{L}), \qquad x \in [0, L].$$

At the first time step ($\ell = 0$), write the full equation from (3.2) that corresponds to the first row of A. Clearly indicate any values that are known, and give those values when $n = 20$.

f) The matrix factorizes as $A = B^T B$, where B is a matrix with linearly independent columns. Prove that A is positive definite by using the quadratic form $\boldsymbol{x}^T A \boldsymbol{x}$.

g) In preparing to solve for $\boldsymbol{u}^{\ell+1}$, (3.2) can be rearranged to

$$K\boldsymbol{u}^{\ell+1} = \boldsymbol{u}^{\ell}.$$

Suppose that $\Delta t = 0.0625$ and $\Delta x = 0.25$. Eigenvalues of K are shown in Figure 3.2.

i) In terms of the matrix A, find the matrix K.

ii) In terms of K and $\boldsymbol{u}^0$, find the finite difference approximation to the exact solution at time T.

iii) How does $||\boldsymbol{u}^m||_2$ compare to $||\boldsymbol{u}^0||_2$? Give a short reason for your answer.

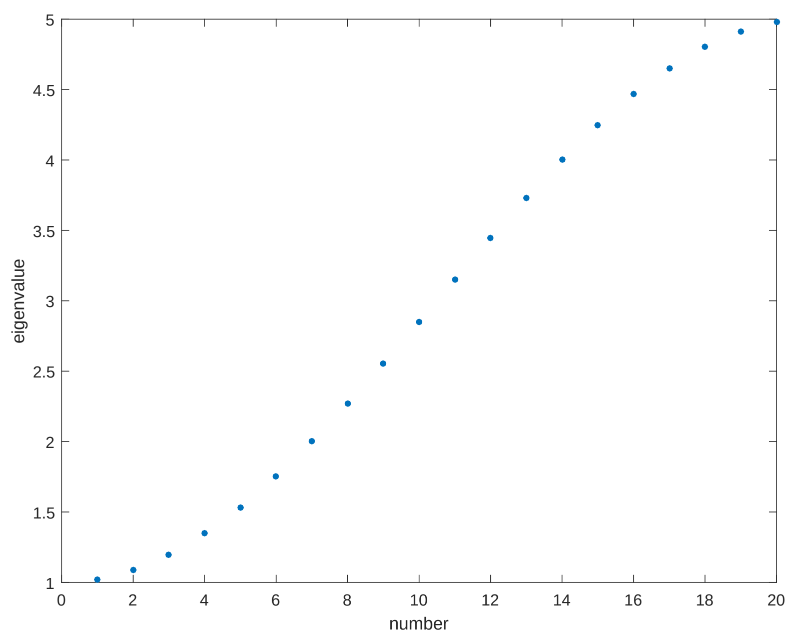
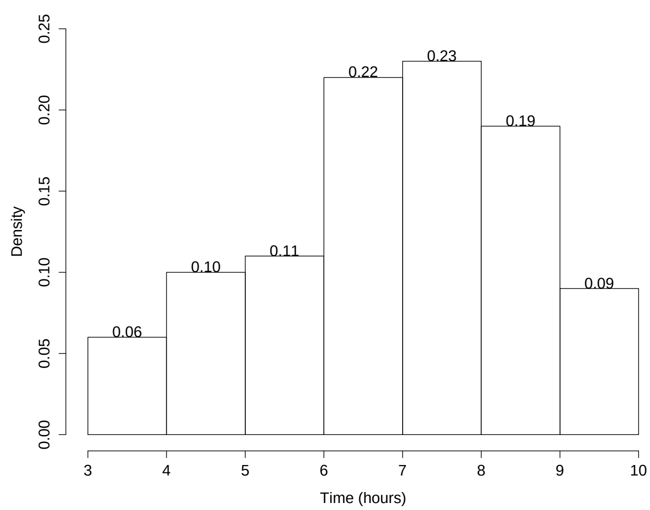


Figure 3.2: Eigenvalues of K.

Part B – Statistics

4. Answer in a separate book marked Question 4

The life of batteries for hand calculators produced by a particular manufacturer was tested. Each battery was put in use in a calculator that is programmed to do a continuous loop of typical calculations, and its lifetime was recorded. The lifetime (in hours) for a random sample of $n = 100$ batteries is shown by the following histogram. For convenience, the exact height of each rectangle is shown, too.



a) Which class interval contains the first quartile of the observed sample of battery lifetimes? Which class interval contains the third quartile of the observed sample of battery lifetimes?

b) Are there any outliers by the $1.5 \times$ iqr-rule? Justify your answer.

c) i) For these 100 batteries, the sample mean lifetime is $\bar{x} = 6.92$ hours with a sample standard deviation $s = 1.64$ hours. Determine a 95%-confidence interval for the true mean lifetime of the batteries produced by this manufacturer.

ii) The above histogram does not look 'bell-shaped'. Explain why this is or is not a problem for building the previous confidence interval.

d) i) What is the observed proportion of batteries which last for more than 8 hours?

ii) The manufacturer claims that more than 33% of the batteries he produces last for more than 8 hours. Does the sample provide enough evidence to contradict this claim? Carry out a suitable hypothesis test at 5% level of significance. *(Write the detail of the test: null and alternative hypotheses, rejection criterion, observed value of the test statistic, p-value, conclusion in plain language.)*

iii) Is the observed sample large enough for this test to be reliable? Justify.

e) In that factory, produced batteries are inspected for flaws by two quality inspectors. If a flaw is present, it will be detected by the first inspector with probability 0.9, and by the second inspector with probability 0.7. The inspectors function independently.

i) If a battery has a flaw, what is the probability that it will be found by at least one of the inspectors?

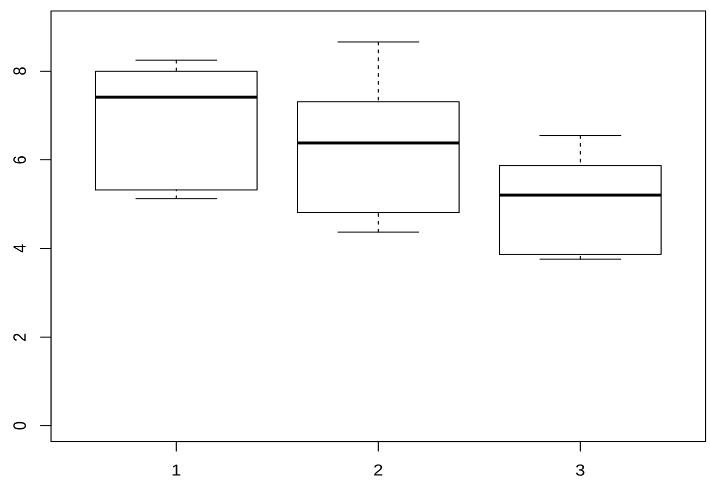
ii) Assume that both inspectors inspect every battery and that if a battery has no flaw, then neither inspector will detect a flaw. Assume also that the probability that a battery has a flaw is 0.10. If a battery is passed by both inspectors, what is the probability that it actually has a flaw?

5. Answer in a separate book marked Question 5

The table below gives the energy use of three gas ranges for six randomly selected days of a month (the units are in kilowatt-hours).

Range 1	Range 2	Range 3
8.25	8.66	6.55
5.12	4.81	3.87
5.32	4.37	3.76
8.00	6.50	5.38
6.97	6.26	5.03
7.86	7.31	5.87
$\bar{x}_1 = 6.920$	$\bar{x}_2 = 6.318$	$\bar{x}_3 = 5.077$
$s_1 = 1.387$	$s_2 = 1.586$	$s_3 = 1.103$

Comparative boxplots are given in the figure below.



a) State two features about the distribution of energy use that can be discerned from the boxplots.

b) Several assumptions need to be valid for an Analysis of Variance (ANOVA) to be an appropriate analysis. State one assumption whose validity can be assessed (at least approximately) based on the given information.

Assume from now on that these assumptions are valid.

c) An ANOVA table was partially constructed to summarise the data:

Source	df	SS	MS	F
Treatment	**(1)**	**(4)**	**(6)**	2.811
Error	**(2)**	**(5)**	1.885	
Total	**(3)**	38.88		

Copy the ANOVA table in your answer booklet. Complete the table by determining the missing values (1)–(6), stating how you computed the missing entries.

d) Using a significance level of $\alpha = 0.05$, carry out the ANOVA F-test to determine whether there is a difference in energy use among the three different ranges. (*You can use the numerical values found in the above table; however, you are required to write the detail of the test: null and alternative hypotheses, rejection criterion or observed value of the test statistic and p-value, conclusion in plain language.*)

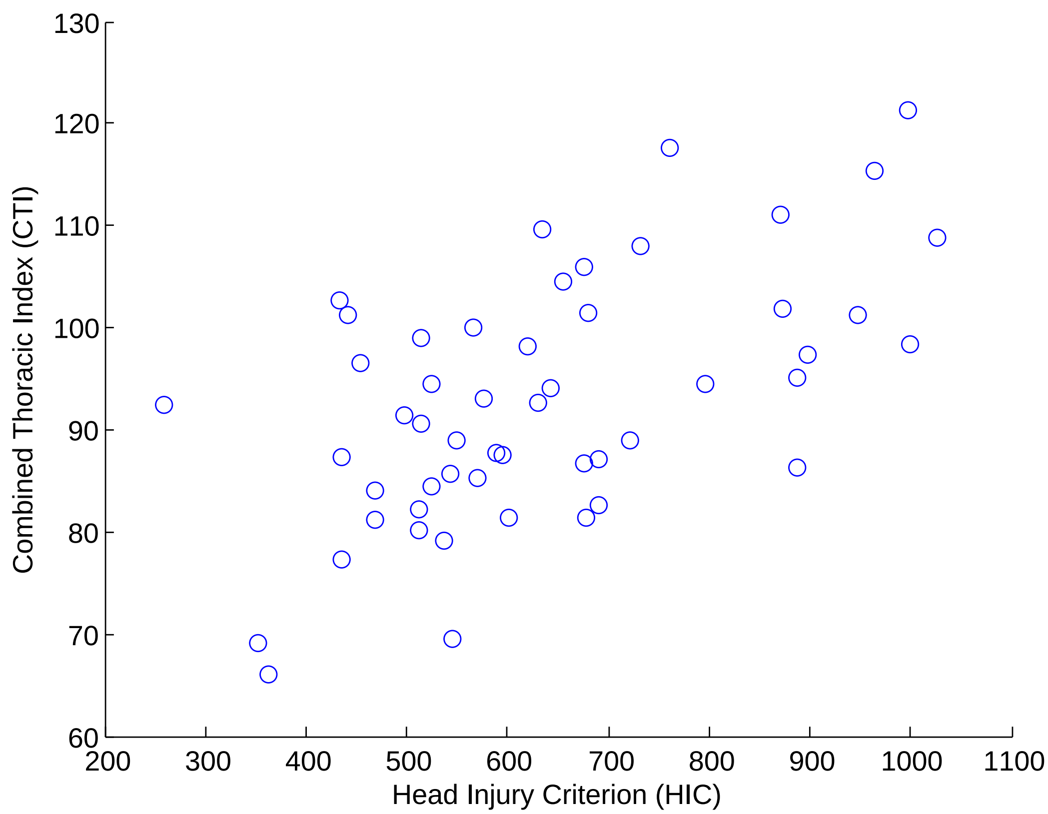
e) Using a significance level of $\alpha = 0.05$, carry out a test to see if the energy use on average for range 1 and range 2 are different. (*Write the detail of the test: null and alternative hypotheses, rejection criterion or observed value of the test statistic and p-value, conclusion in plain language.*)

f) Two two-sample t-tests are carried out for comparing range 2 and range 3, and range 1 and range 3, with p-values 0.138 and 0.035, respectively. Does simultaneously analysing the three pairwise comparisons allow you to come to the same conclusion as the ANOVA F-test in (d), at overall significance level $\alpha = 0.05$?

6. Answer in a separate book marked Question 6

In motor vehicle crash testing, several measurements are taken on hybrid III crash test dummies such as velocity, acceleration and deflection of body parts. These measures are combined into indices such as the head injury criterion (`HIC`) and combined thoracic index (`CTI`). Due to a common cause, i.e., severity of crash, it is believed `HIC` and `CTI` are related linearly.

Data was collected for the 1998 New Car Assessment Program (NCAP) for 52 motor vehicles. A scatter plot of `HIC` and `CTI` data is shown below



The following summary statistics were obtained for `HIC`

$$\sum_{i=1}^{52} x_i = 33023 \quad \text{and} \quad s_{xx} = \sum_{i=1}^{52}(x_i - \bar{x})^2 = 1698115$$

The regression model is given by

$$CTI = \beta_0 + \beta_1 HIC + \epsilon.$$

Use the following regression output to answer the questions below.

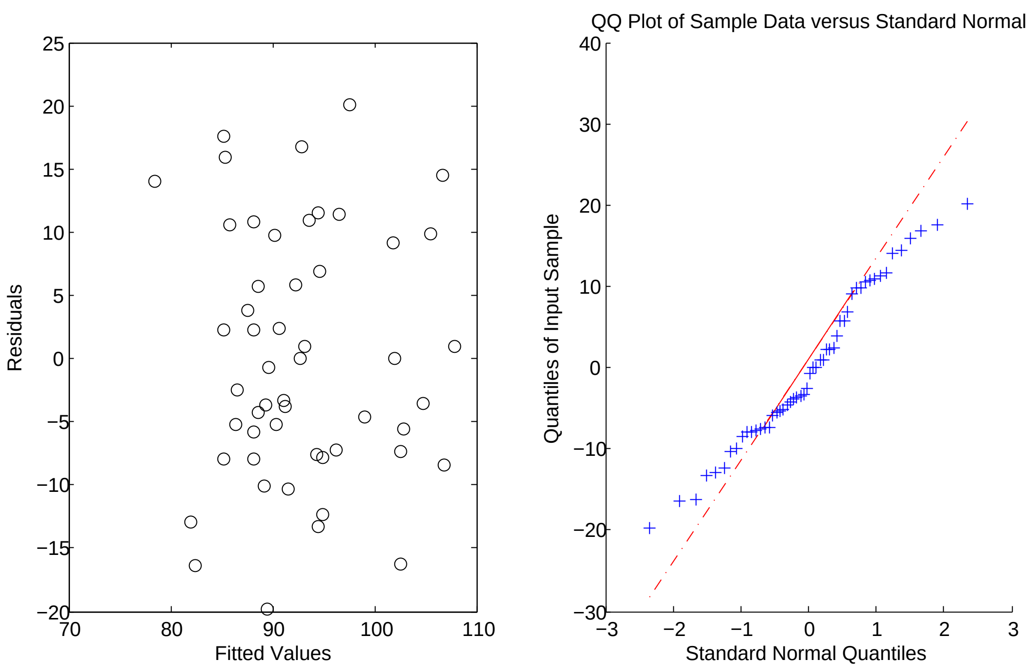
```
The regression equation is
    CTI = 68.5 + 0.0383*HIC
```

Predictor	Coef	SE Coef	T	P
Constant	68.518	5.0441	13.584	2.0897e-18
HIC	0.038321	0.0076395	5.0161	7.0309e-06

```
S = 9.96    R-Sq = 33.5%    R-Sq(adj) = 32.1%
```

a) i) For the regression analysis to be valid, what is the assumed distribution of the error ϵ?

ii) Given the plots below and other given information, explain why the statistical results are at least approximately valid for this analysis?



Assume from now on that these assumptions are valid.

b) What is the expected change in `CTI` for a 1 unit change in `HIC`?

c) What proportion of variation in the response is explained by the variation in the predictor?

d) Determine the observed sample correlation coefficient between `CTI` and `HIC`.

e) Give the estimated value of σ, the standard deviation of the error term ϵ.

f) Perform a hypothesis test to determine whether the variable X is significant in this model, at the 5% level of significance. (*You can use the numerical values found in the above output, however you are required to write the details of the test: null and alternative hypotheses, rejection criterion or observed value of the test statistics and p-value (you may present upper or lower bounds), conclusion in plain language.*)

g) Compute a 95% confidence interval for β_1.

h) Find a 99% confidence interval for the mean `CTI` when `HIC` is 1000. Explain how this interval is different to a prediction interval for `CTI` when `HIC` is 1000.