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

UNSW SYDNEY

SCHOOL OF MATHEMATICS AND STATISTICS

Semester 2, 2017

MATH2089
Numerical Methods and Statistics

(1) TIME ALLOWED – 2 Hours

(2) TOTAL NUMBER OF QUESTIONS – 4

(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

Part A – Numerical Methods consists of questions 1 – 2

Part B – Statistics consists of questions 3 – 4

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) **[18 marks]** 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 (2 marks).

i) 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.

ii) **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
```

iii) 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 $e_{k+1} \approx 3.96 \times 10^{-21}$.

iv) 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.

Please see over ...

v) You are given that

```
norm(A-A') = 9.3e-16
min(eig(A)) = 1.3e-7
max(eig(A)) = 2.6e+6
```

Claim: The condition number $\kappa_2(A)$ using the 2-norm is 2×10^{13}.

vi) You are given that

```
size(A)
ans =
    5    5
[Q,R] = qr(A);
```

Claim: For every vector $\mathbf{x} \in \mathbb{R}^5$, we have $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.

b) **[12 marks]** 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

$$E_N^{\mathrm{Trap}}(f) = O(N^{-2}), \tag{1.1}$$

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

A) Use (1.1) to estimate the ratio

$$\frac{E_N^{\mathrm{Trap}}(f)}{E_{2N}^{\mathrm{Trap}}(f)}. \tag{1.2}$$

Please see over ...

B) Use the table of errors to estimate the ratio (1.2) when $N = 512$.

C) Is the table of errors consistent with the theoretical error estimate in (1.1)?

D) Give reasons for your answer in C).

Please see over ...

2. Answer in a separate book marked Question 2

a) **[14 marks]** 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

$$m = 5, \quad c = 2, \quad \text{and} \quad k = 1.$$

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) Heun's method for solving the initial value problem $u' = g(t, u)$ with $u(t_0) = u_0$ can be summarized by the following formula: for $n = 0, 1, \ldots, N-1$, compute:

$$\begin{aligned} z_{n+1} &= u_n + hg(t_n, u_n) \\ u_{n+1} &= u_n + \frac{h}{2}[g(t_n, u_n) + g(t_{n+1}, z_{n+1})]. \end{aligned}$$

Use Heun's method with a step of $h = 0.1$ to estimate $\boldsymbol{x}(1.1)$ for the initial value problem in question ii).

Please see over ...

b) **[16 marks]** Fick's second law predicts how diffusion causes the concentration $u(x, y)$ of a chemical to change with position $(x, y) \in \Omega$. The steady state version of Fick's second law (without interior sources of the chemical) is Laplace's equation

$$\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2} = 0. \tag{2.1}$$

Consider the rectangular domain

$$\Omega = \{(x, y) \in \mathbb{R}^2 \ : \ 0 \le x \le 2, \ 0 \le y \le 1\},$$

and discretize it using $h = 1/n$ and

$$\begin{cases} x_i = ih & \text{for } i = 0, 1, \ldots, 2n, \\ y_j = jh & \text{for } j = 0, 1, \ldots, n. \end{cases}$$

This is illustrated in Figure 2.1 for $n = 5$.

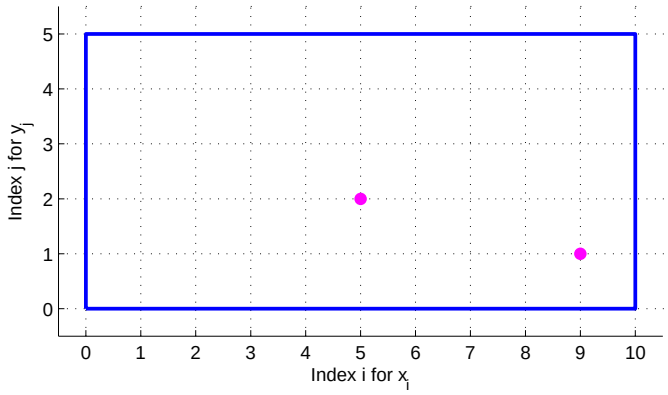


Figure 2.1: Discretization of the domain for $n = 5$ and grid points for part iv)

i) What extra information is needed to completely specify this problem?

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

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

Let $u_{i,j}$ denote the approximation to the value $u(x_i, y_j)$ of concentration at the grid point (x_i, y_j). Give central difference approximations of accuracy $O(h^2)$ to the following derivatives at the point (x_i, y_j)

A) $\dfrac{\partial^2 u(x,y)}{\partial x^2}$ *B)* $\dfrac{\partial^2 u(x,y)}{\partial y^2}$

Please see over ...

iii) Using the finite difference approximations from the previous part, show that the equation (2.1) can be approximated by

$$\beta u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} = 0, \tag{2.2}$$

and determine the value of β.

iv) Given that, in appropriate units,

$$\begin{cases} u(x,0) = u(x,1) = 10x & \text{for } 0 \le x \le 2, \\ u(0,y) = 0 & \text{for } 0 \le y \le 1, \\ u(2,y) = 20 & \text{for } 0 \le y \le 1, \end{cases}$$

write down the equation (2.2) for a discretization with $n = 5$ at the grid points (marked in Figure 2.1)

A) (x_5, y_2) *B*) (x_9, y_1)

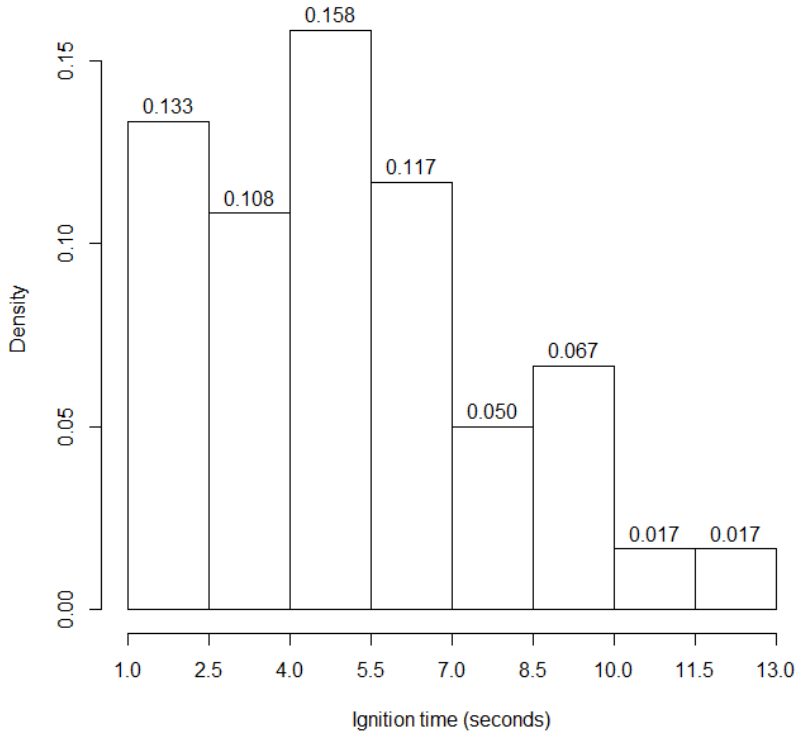
v) You are given that the coefficient matrix A is symmetric positive definite. Outline an effective way to solve the linear system $A\boldsymbol{u} = \boldsymbol{b}$.

Please see over ...

Part B – Statistics

3. Answer in a separate book marked Question 3

a) **[9 marks]** Researchers investigated the ignition time of certain upholstery materials exposed to a flame. The ignition time (in seconds) for 80 specimens is shown by the following density histogram. The exact height of each rectangle is shown, as well.



i) **[2 marks]** What is the observed proportion of upholstery materials which have an ignition time of more than 8.5 seconds?

ii) **[4 marks]** Create a two-sided 96% confidence interval for the proportion of upholstery materials that have an ignition time of more than 8.5 seconds.

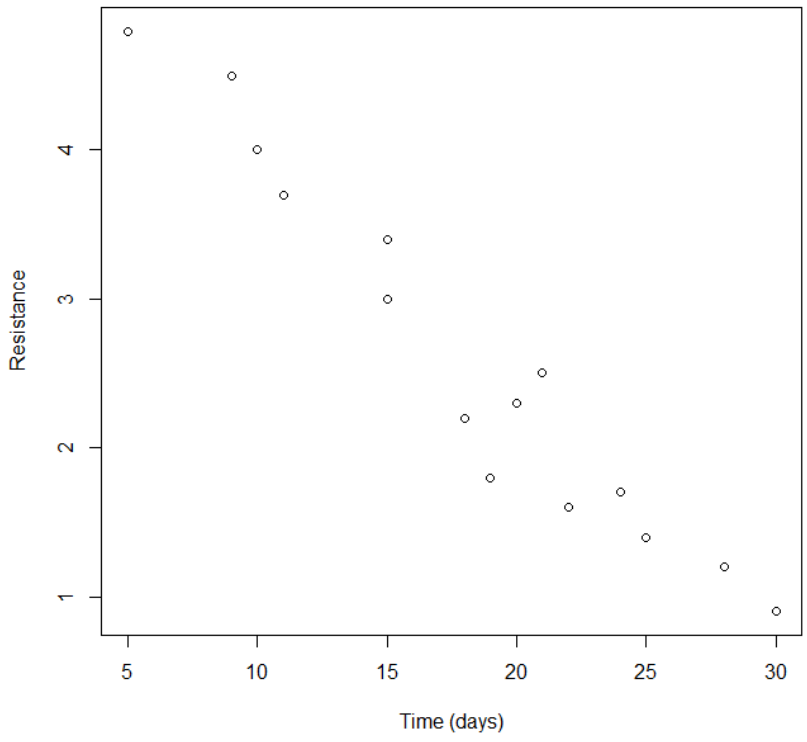
Hint: You can use the following output from Matlab:

$\texttt{norminv}(0.94) = 1.555, \texttt{norminv}(0.96) = 1.751, \texttt{norminv}(0.98) = 2.054.$

iii) **[3 marks]** Is the observed sample size large enough for the above confidence interval to be reliable? State your reasons.

Please see over ...

b) **[21 marks]** The final step in the manufacture of graphite electrodes is graphitizing in an electric furnace, which reduces the resistance of the electrode so that it will not burn up in use. The graphitizing process is a slow one and the electrodes must remain in the electric furnace for several weeks before the graphitizing is completed. In order to increase production and meet all the customers' demands, the production manager in a large producer of these electrodes carried out a study to determine how the resistance of an electrode varies with the length of time it spends in the graphitizing furnace. 15 specimens were tested, and a scatter plot of the `Time` (in days) the electrode spent in the graphitizing furnace and its `Resistance` (in the special unit used by the company for characterizing the resistance of these electrodes, smaller values represent lower resistance) is shown below.



Assume the predictor variable X is `Time` and the response variable Y is `Resistance`. The linear regression model is given by

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

Please see over ...

The following summary statistics were obtained for X.

$$\sum_{i=1}^{15} x_i = 272 \quad \text{and} \quad s_{xx} = \sum_{i=1}^{15}(x_i - \bar{x})^2 = 739.7333$$

You can use the following output from Matlab to answer the questions below.

```
Estimated Coefficients:

                    Estimate      SE        tStat      pValue
                   ---------    -------    -------    ----------

   (Intercept)       5.5931     0.22314     25.066     2.164e-12
   x                -0.16506    0.011475    -14.384    2.3125e-09

Root Mean Squared Error: 0.312
R-squared: 0.941
```

$\texttt{tinv}(0.9, 13) = 1.350$,	$\texttt{tinv}(0.9, 14) = 1.345$,	$\texttt{tinv}(0.9, 15) = 1.341$,
$\texttt{tinv}(0.95, 13) = 1.771$,	$\texttt{tinv}(0.95, 14) = 1.761$,	$\texttt{tinv}(0.95, 15) = 1.753$,
$\texttt{tinv}(0.975, 13) = 2.160$,	$\texttt{tinv}(0.975, 14) = 2.145$,	$\texttt{tinv}(0.975, 15) = 2.131$,
$\texttt{tinv}(0.98, 13) = 2.282$,	$\texttt{tinv}(0.98, 14) = 2.264$,	$\texttt{tinv}(0.98, 15) = 2.249$,
$\texttt{tinv}(0.99, 13) = 2.650$,	$\texttt{tinv}(0.99, 14) = 2.625$,	$\texttt{tinv}(0.99, 15) = 2.603$.

i) [**1 mark**] What proportion of variability in the response is explained by the predictor?

ii) [**1 mark**] Assume σ is the standard deviation of the error term ϵ. Give an estimate of σ.

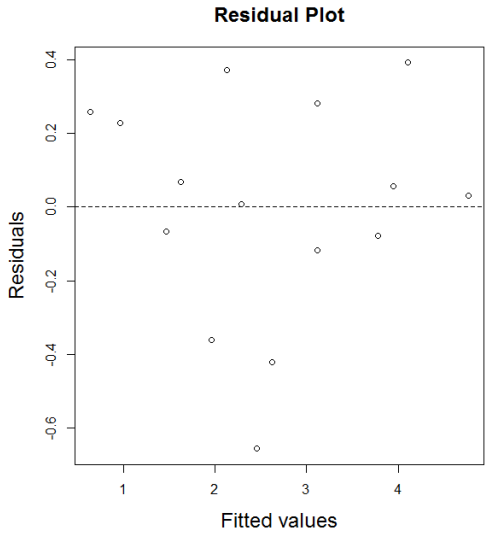
iii) [**7 marks**] Perform a hypothesis test to determine whether the variable X is significant in this model, at the 5% level of significance. (*You can use the numerical values found in the above output; however, you are required to write the details of the test: null and alternative hypotheses; rejection criterion, or observed value of the test statistics and p-value (specify the degrees of freedom if applicable); conclusion in plain language.*)

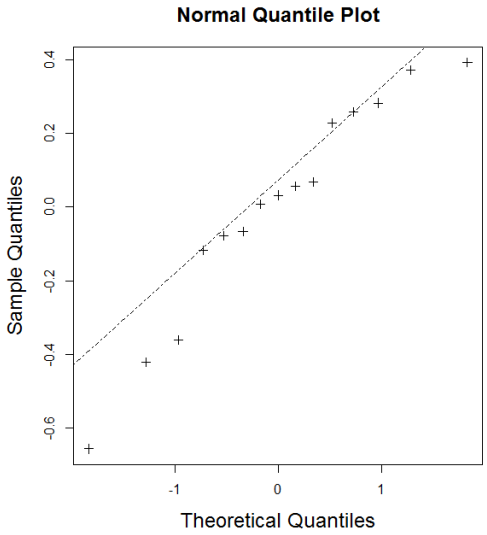
iv) [**4 marks**] Compute a two-sided 90% confidence interval for the true average `Resistance` when the `Time` an electrode spent in the graphitizing furnace is 12 days.

Please see over ...

v) **[4 marks]** Compute a two-sided 98% prediction interval for a future value of average `Resistance` that will be observed when the `Time` an electrode spent in the graphitizing furnace is 16 days.

vi) **[4 marks]** For the above regression analysis to be valid, what are the three essential assumptions that the error ϵ in the model must satisfy? Comment on the validity of these assumptions, given the residual versus fitted values plot and the normal quantile plot below.





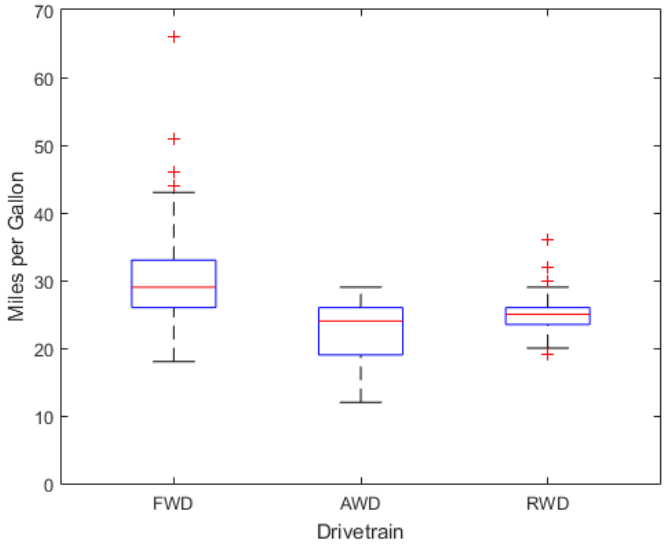
Please see over ...

4. Answer in a separate book marked Question 4

Data were collected on 387 new vehicles for year 2004. The variables include type of drivetrain (`FWD`=front wheel drive, `AWD`=all wheel drive, `RWD`=rear wheel drive) and fuel efficiency for highway travel (`HwayMPG`, larger values imply greater efficiency). This data is summarised in the table below.

FWD	AWD	RWD
$\bar{x}_1 = 29.60$	$\bar{x}_2 = 23.01$	$\bar{x}_3 = 25.46$
$s_1 = 5.91$	$s_2 = 3.97$	$s_3 = 2.71$
$n_1 = 215$	$n_2 = 78$	$n_3 = 94$

Comparative boxplots are given in the figure below.



You can use the following output from Matlab to answer the questions below.

$\texttt{tinv}(0.9, 2) = 1.886$,	$\texttt{tinv}(0.9, 307) = 1.284$,	$\texttt{tinv}(0.9, 309) = 1.284$,
$\texttt{tinv}(0.95, 2) = 2.920$,	$\texttt{tinv}(0.95, 307) = 1.650$,	$\texttt{tinv}(0.95, 309) = 1.650$,
$\texttt{tinv}(0.975, 2) = 4.303$,	$\texttt{tinv}(0.975, 307) = 1.968$,	$\texttt{tinv}(0.975, 309) = 1.968$,
$\texttt{tcdf}(10.785, 2) = 0.996$,	$\texttt{tcdf}(10.785, 384) = 1.000$,	$\texttt{tcdf}(10.785, 387) = 1.000$,
$\texttt{finv}(0.95, 2, 381) = 3.0194$,	$\texttt{finv}(0.95, 2, 384) = 3.0192$,	$\texttt{finv}(0.95, 3, 387) = 3.0190$,
$\texttt{finv}(0.975, 2, 381) = 3.7248$,	$\texttt{finv}(0.975, 2, 384) = 3.7245$,	$\texttt{finv}(0.975, 3, 387) = 3.7243$,
$\texttt{fcdf}(59.08, 2, 381) = 1$,	$\texttt{fcdf}(59.08, 2, 384) = 1$,	$\texttt{fcdf}(59.08, 3, 387) = 1$.

a) **[3 marks]** What do the boxplots tell you about highway fuel efficiency for different drivetrains? Comment on the shape, range and location.

Please see over ...

b) **[4 marks]** List three assumptions that need to be valid for an Analysis of Variance (ANOVA) to test whether there is a difference in average downloading time among the three time periods. Which of these three can be checked by considering the accompanying summary statistics? Explain whether these verifiable assumption(s) are supported.

Assume from now on that these assumptions are valid.

c) **[5 marks]** An ANOVA table was partially constructed to summarise the data:

Source	df	SS	MS	F
Treatment	**(1)**	2885	**(4)**	**(5)**
Error	384	**(3)**	24.42	
Total	**(2)**	12261.1		

Copy the ANOVA table in your answer booklet. Complete the table by determining the missing values (1)–(5), and stating how you computed the missing entries.

d) **[7 marks]** Using a significance level of $\alpha = 0.05$, carry out the ANOVA F-test to determine whether there is a difference in highway fuel efficiency among the three types of drivetrains.

(*You can use the numerical values found in the above ANOVA table; however, you are required to write the detail of the test: null and alternative hypotheses; observed value of the test statistic; rejection criterion or the p-value (you may use bounds for the p-value, specify the degrees of freedom if applicable); conclusion in plain language.*)

e) **[4 marks]** From the previous results, construct a 90% two-sided confidence interval on the difference between the "true" average highway fuel efficiencies for FWD and RWD vehicles, that is, $\mu_1 - \mu_3$.

f) **[7 marks]** Using the Bonferroni adjustment, carry out a t-test comparing the "true" average highway fuel efficiencies for FWD and AWD vehicles. Does this allow you to come to the same conclusion as the ANOVA F-test in d), at overall level $\alpha = 0.05$? Explain.

(*You can use the numerical values found in the above ANOVA table; however, you are required to write the detail of the test: null and alternative hypotheses; observed value of the test statistic and p-value (specify the degrees of freedom if applicable).*)