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

UNSW SYDNEY

SCHOOL OF MATHEMATICS AND STATISTICS

Semester 2, 2017

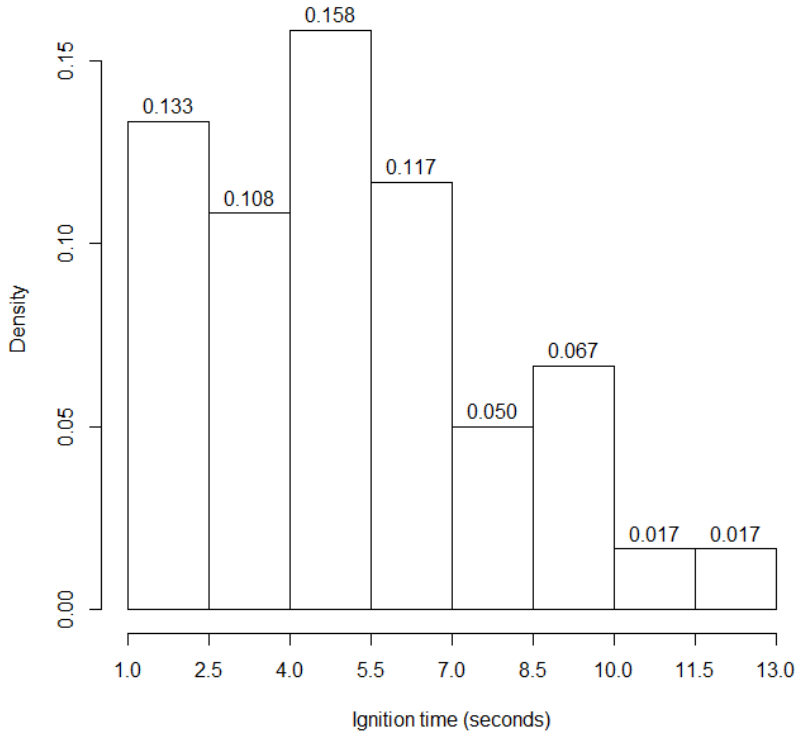
MATH2099/MATH2859

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 B – Statistics

3. Answer in a separate book marked Question 3

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

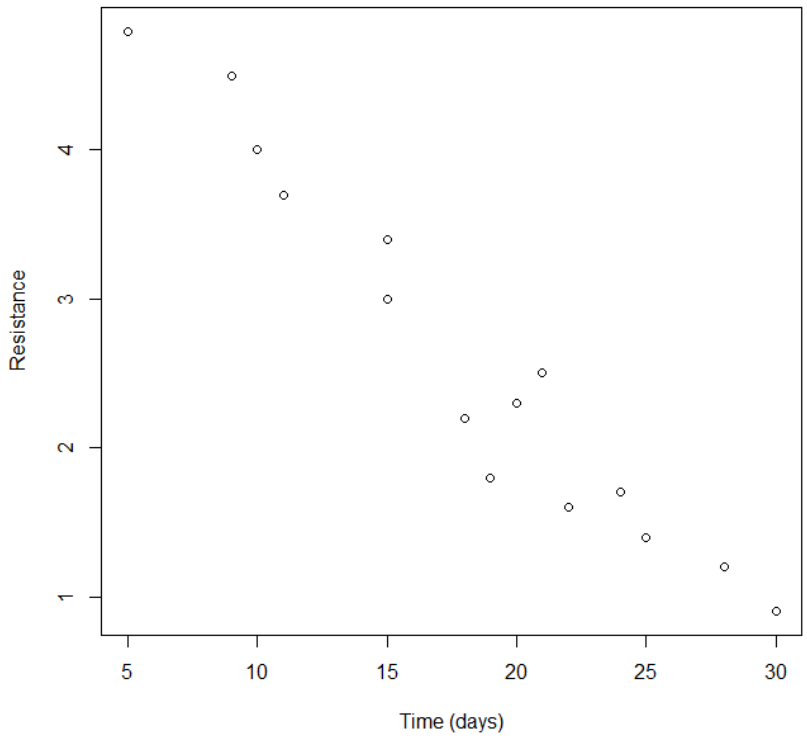
Hint: You can use the following output from Matlab:

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

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) 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 randomly selected 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, where 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.

$$\sum_{i=1}^{15} x_i = 272 \quad \text{and} \quad s_{xx} = \sum_{i=1}^{15} (x_i - \bar{x})^2 = 739.733$$

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.

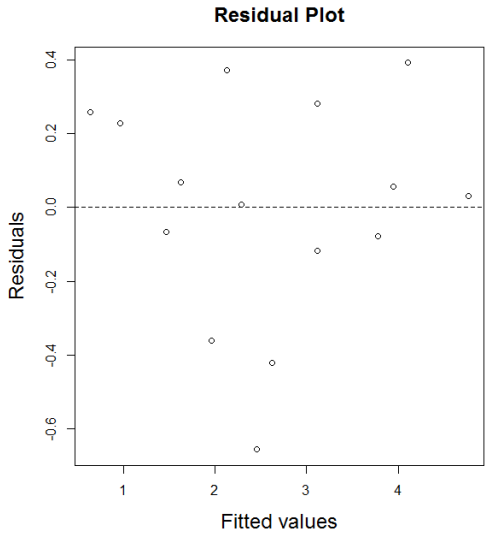
iii) **[7 marks]** Perform a hypothesis test to determine whether the variable X is significant in this model, at the 5% level of significance. (*You can use the numerical values found in the above output; however, you are required to write the details of the test: null and alternative hypotheses; rejection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

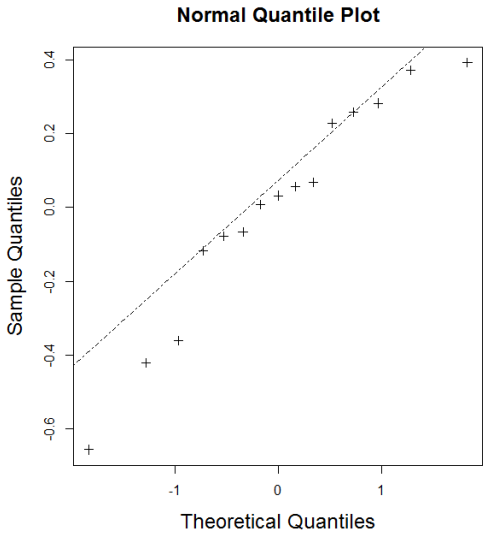
iv) **[4 marks]** Compute a two-sided 90% confidence interval for the true average `Resistance` when the `Time` an electrode spent in the graphitizing furnace is 12 days.

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 regression model must satisfy? Comment on the validity of these assumptions, given the residual versus fitted values plot and the normal quantile plot below.



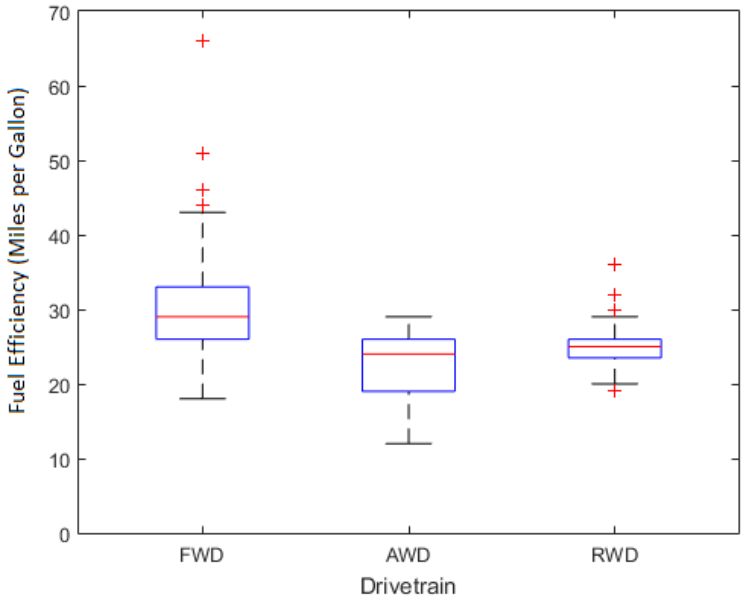


4. Answer in a separate book marked Question 4

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). These data are 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.

$\texttt{tinv}(0.9, 2) = 1.886$, $\texttt{tinv}(0.9, 307) = 1.284$, $\texttt{tinv}(0.9, 384) = 1.284$,
$\texttt{tinv}(0.95, 2) = 2.920$, $\texttt{tinv}(0.95, 307) = 1.650$, $\texttt{tinv}(0.95, 384) = 1.649$,
$\texttt{tinv}(0.975, 2) = 4.303$, $\texttt{tinv}(0.975, 307) = 1.968$, $\texttt{tinv}(0.975, 384) = 1.966$,

$\texttt{tcdf}(10.089, 2) = 0.995$, $\texttt{tcdf}(10.089, 384) = 1.000$, $\texttt{tcdf}(10.089, 387) = 1.000$,

$\texttt{finv}(0.95, 2, 381) = 3.0194$, $\texttt{finv}(0.95, 2, 384) = 3.0192$, $\texttt{finv}(0.95, 3, 387) = 3.0190$,
$\texttt{finv}(0.975, 2, 381) = 3.7248$, $\texttt{finv}(0.975, 2, 384) = 3.7245$, $\texttt{finv}(0.975, 3, 387) = 3.7243$,

$\texttt{fcdf}(59.08, 2, 381) = 1$, $\texttt{fcdf}(59.08, 2, 384) = 1$, $\texttt{fcdf}(59.08, 3, 387) = 1$.

Please see over ...

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

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 fuel efficiency among the three drivetrain types. Which of these three assumptions 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 (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).*)

Please see over ...