APPENDICES BOOKLET

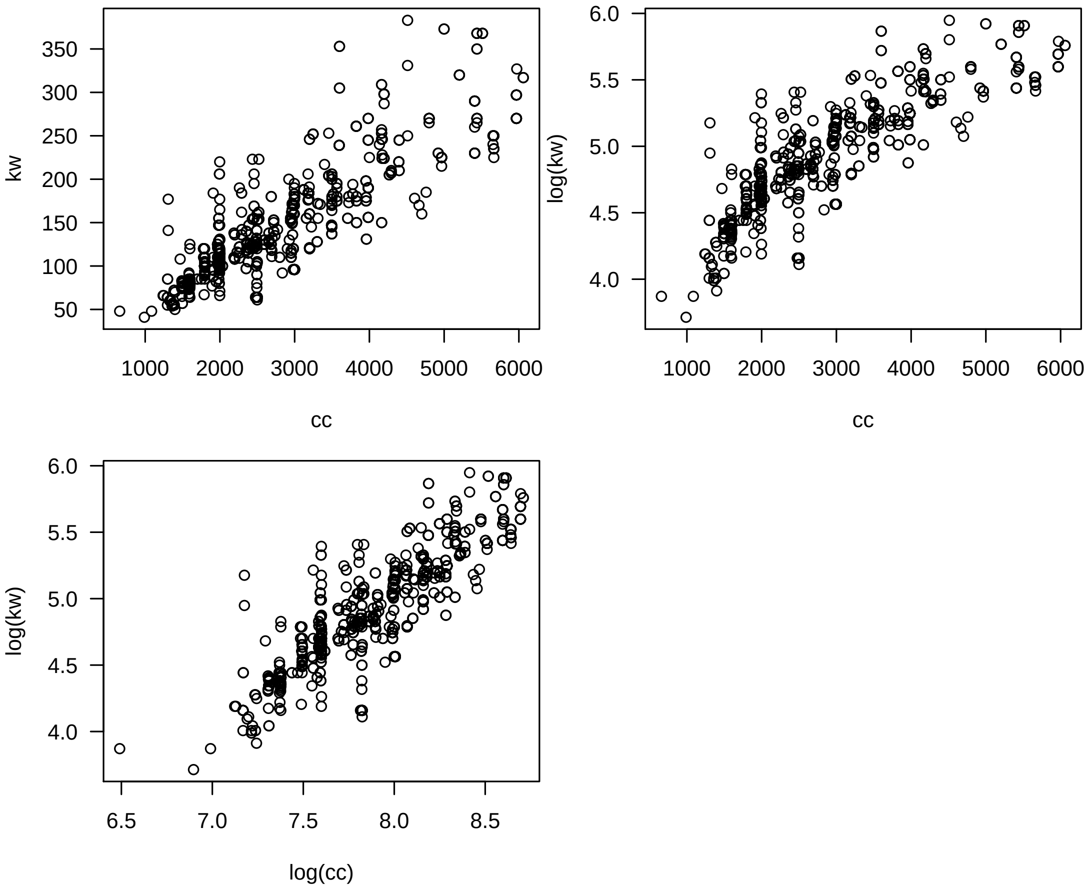
CONTENTS

NZ Car Data

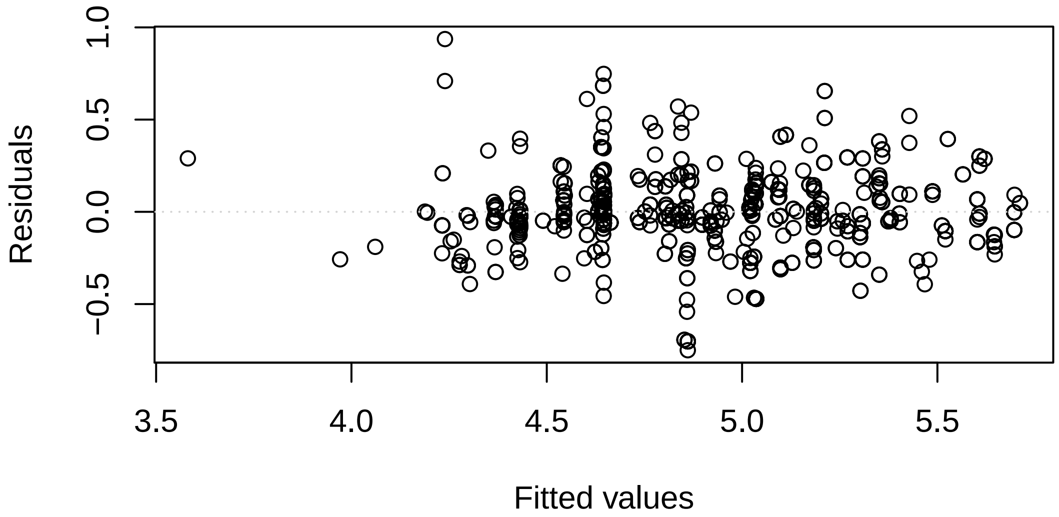
The New Zealand Automobile Association (AA) publish various data about models of car on their website. Although there are many variables available, here it is of interest to determine the relationship between the power output of the cars and the size of their engines.

The variables in the data set are

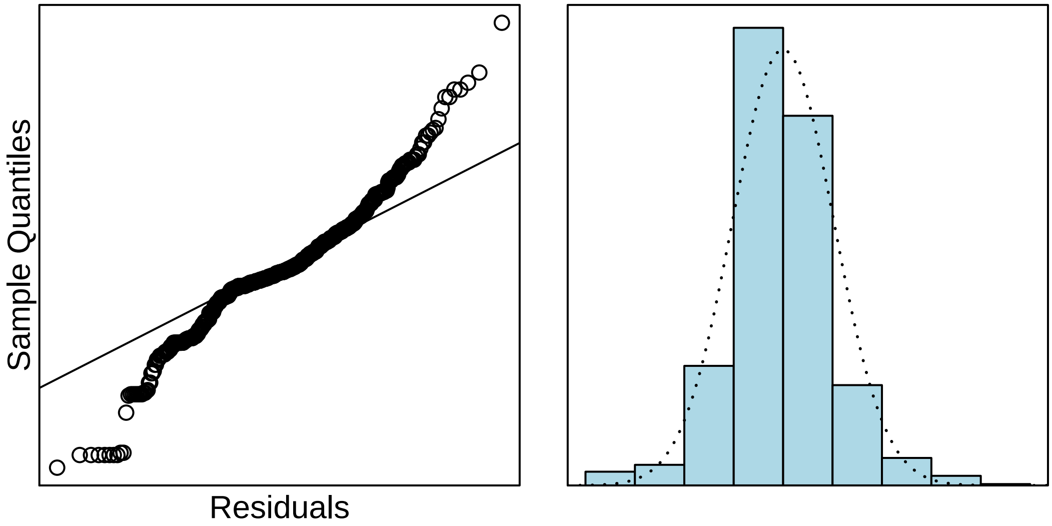
`kw`	The power output of the car measured in kilowatts (kW)
`cc`	The engine size measured in cubic centimetres (cm^3)



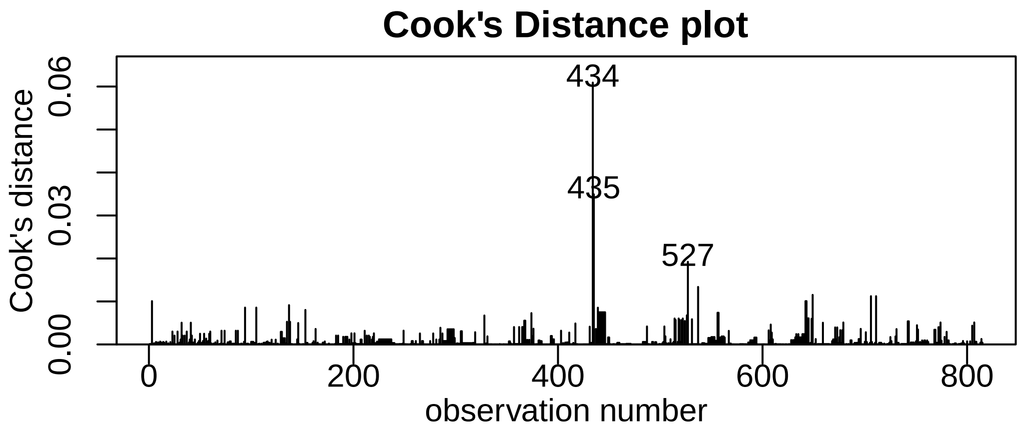
```
> nzcars.fit = lm(log(kw) ~ log(cc), data = nzcars.df)
> eovcheck(nzcars.fit)
```



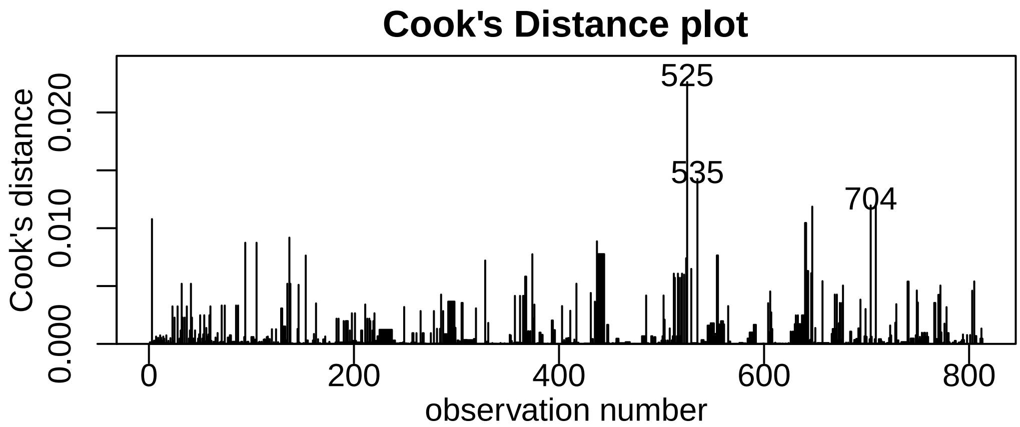

```
> normcheck(nzcars.fit, main = "", xlab = "Residuals")
```



```
> cooks20x(nzcars.fit)
```




```
> nzcars.df[434:435,]
     make model doors   cc  kw manual  auto
434 Mazda   RX8     2 1308 177  61995  <NA>
435 Mazda   RX8     2 1308 141     NA 61995
> nzcars1.fit = lm(log(kw)~log(cc),
+                  data = nzcars.df[-c(434,435), ])
> cooks20x(nzcars1.fit)
```



```
> summary(nzcars.fit)

Call:
lm(formula = log(kw) ~ log(cc), data = nzcars.df)

Residuals:
     Min       1Q   Median       3Q      Max
-0.75009 -0.07331 -0.00998  0.10614  0.93675

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.65055    0.15153  -17.49   <2e-16 ***
log(cc)      0.96010    0.01916   50.10   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2051 on 813 degrees of freedom
Multiple R-squared:  0.7553,Adjusted R-squared:  0.755
F-statistic:  2510 on 1 and 813 DF,  p-value: < 2.2e-16
> summary(nzcars1.fit)

Call:
lm(formula = log(kw) ~ log(cc), data = nzcars.df[-c(434, 435),
    ])

Residuals:
     Min       1Q   Median       3Q      Max
-0.74728 -0.06717 -0.00918  0.10694  0.75313

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.73545    0.14929  -18.32   <2e-16 ***
log(cc)      0.97060    0.01888   51.41   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2011 on 811 degrees of freedom
Multiple R-squared:  0.7652,Adjusted R-squared:  0.7649
```

```
F-statistic:  2643 on 1 and 811 DF,  p-value: < 2.2e-16


> confint(nzcars.fit)
                 2.5 %     97.5 %
(Intercept) -2.9479738 -2.3531209
log(cc)      0.9224856  0.9977227


> exp(confint(nzcars.fit))
                 2.5 %     97.5 %
(Intercept) 0.05244587 0.09507199
log(cc)     2.51553523 2.71209843


> confint(nzcars1.fit)
                 2.5 %    97.5 %
(Intercept) -3.0284991 -2.442401
log(cc)      0.9335391  1.007653


> exp(confint(nzcars1.fit))
                 2.5 %     97.5 %
(Intercept) 0.04838821 0.08695183
log(cc)     2.54349505 2.73916420
```

Hull Damage Data

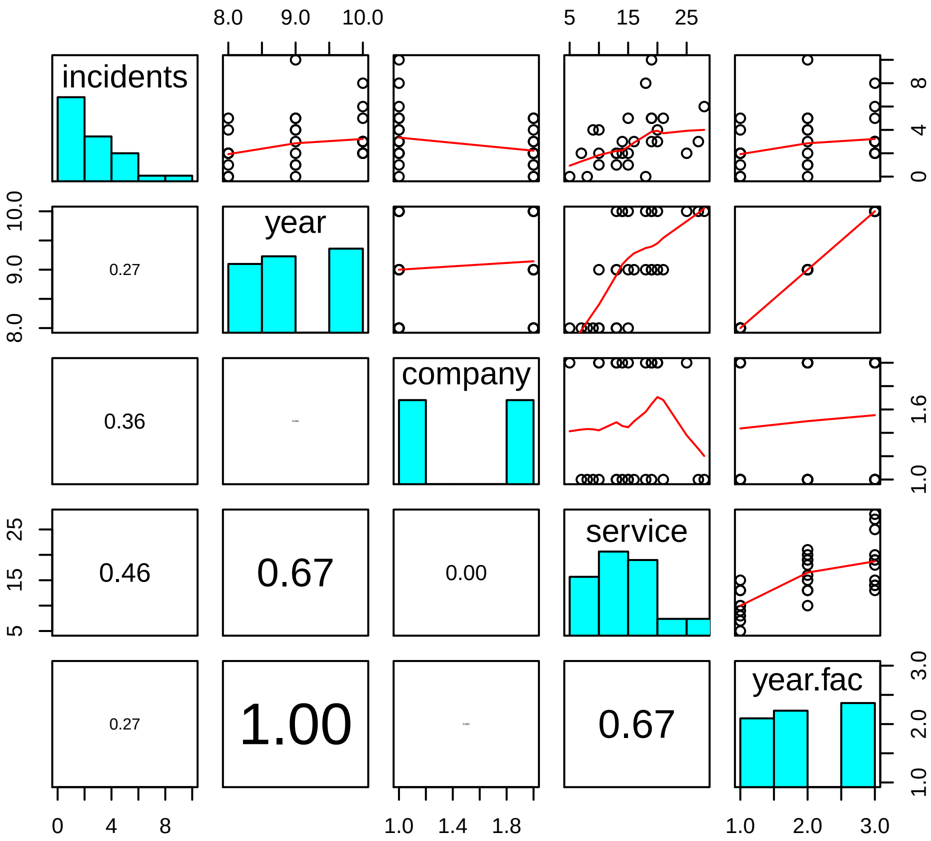
This question refers to data that come from a study investigating a particular type of minor damage caused by waves to the forward sections of ships' hulls. In total, 30 randomly selected ships were inspected for hull damage, and the number of damage incidents were recorded from each. Hull construction engineers are interested in determining if the design of the hull is related to the number of observed damage incidents. Hull designs vary across manufacturers, and potentially improve from year to year. Also, we might expect more damage incidents for ships that have been in service for longer periods of time.

The variables in the data set are

`incidents`	The number of damage incidents detected on the ship's hull
`year`	The year of construction: `8` for 2008, `9` for 2009, and `10` for 2010
`company`	The company that constructed the ship; either `A` or `B`
`service`	The number of months the ship was in service

```
> ship.df <- within(ship.df, {
+     year.fac = as.factor(year)
+ })
> summary(ship.df$incidents)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   2.500   3.067   4.000  10.000
> summary(ship.df$year.fac)
 8  9 10
 9 10 11
> summary(ship.df$company)
 A  B
15 15
> summary(ship.df$service)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.00   13.00   15.00   15.53   19.00   28.00
```

```
> pairs20x(ship.df)
```

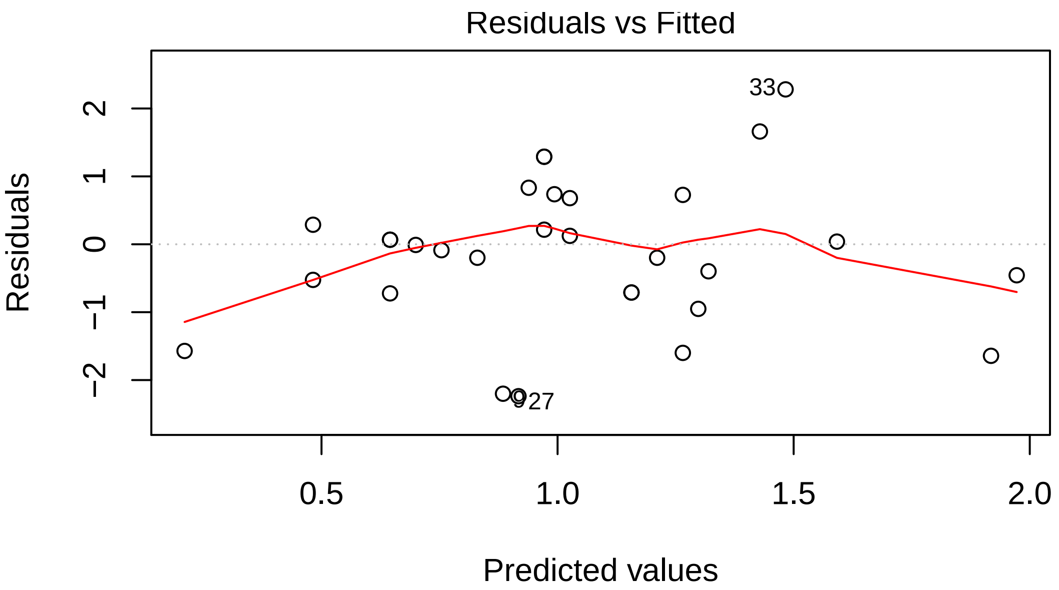



```
> ship1.fit <- glm(incidents ~ company + service + year.fac,
+                   family = "poisson", data = ship.df)
```

```
> anova(ship1.fit, test = "Chisq")
Analysis of Deviance Table

Model: poisson, link: log

Response: incidents

Terms added sequentially (first to last)


         Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                        29     50.195
company   1   6.3339        28     43.861  0.01185 *
service   1   9.4511        27     34.410  0.00211 **
year.fac  2   0.8783        25     33.532  0.64459
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> summary(ship1.fit)

Call:
glm(formula = incidents ~ company + service + year.fac, family = "poisson",
    data = ship.df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.35927  -0.70150  -0.09147   0.45431   1.99681

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.38336    0.35939   1.067   0.2861
companyB    -0.52422    0.21857  -2.398   0.0165 *
service      0.04931    0.02396   2.058   0.0396 *
year.fac9    0.27684    0.32885   0.842   0.3999
year.fac10   0.13834    0.38027   0.364   0.7160
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 50.195  on 29  degrees of freedom
Residual deviance: 33.532  on 25  degrees of freedom
AIC: 123.49

Number of Fisher Scoring iterations: 5


> ship1.fit$deviance
[1] 33.5321
> ship1.fit$df.residual
[1] 25
> 1 - pchisq(ship1.fit$deviance, ship1.fit$df.residual)
[1] 0.1182934


> ship2.fit <- glm(incidents ~ company + service,
+                  family = "poisson", data = ship.df)


> summary(ship2.fit)

Call:
glm(formula = incidents ~ company + service, family = "poisson",
    data = ship.df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.23706  -0.70999  -0.04805   0.58199   2.28167

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.44937    0.33636   1.336  0.18156
companyB    -0.51125    0.21658  -2.361  0.01825 *
service      0.05439    0.01747   3.114  0.00185 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 50.195  on 29  degrees of freedom
Residual deviance: 34.410  on 27  degrees of freedom
AIC: 120.37

Number of Fisher Scoring iterations: 5
> 1 - pchisq(ship2.fit$deviance, ship2.fit$df.residual)
[1] 0.1544394

> plot(ship2.fit, which = 1)
```




```
> confint(ship2.fit)


Waiting for profiling to be done...


                  2.5 %      97.5 %
(Intercept) -0.23128811  1.08940658
companyB    -0.94508880 -0.09286586
service      0.01990593  0.08849989
```

```
> exp(confint(ship2.fit))

Waiting for profiling to be done...

               2.5 %    97.5 %
(Intercept) 0.7935108 2.9725096
companyB    0.3886451 0.9113157
service     1.0201054 1.0925341

> ## For twelve-month change in service.
> exp(12 * confint(ship2.fit)[3, ])

Waiting for profiling to be done...

   2.5 %   97.5 %
1.269815 2.892146
```

For Part (g) Only

```
> ship3.fit <- glm(incidents ~ company + year.fac,
+                  family = "poisson", data = ship.df)


> summary(ship3.fit)

Call:
glm(formula = incidents ~ company + year.fac, family = "poisson",
    data = ship.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2483  -0.7988  -0.3150   0.5687   2.2444

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9080     0.2452   3.704 0.000213 ***
companyB     -0.5708     0.2166  -2.636 0.008392 **
year.fac9     0.5900     0.2902   2.033 0.042068 *
year.fac10    0.6285     0.2857   2.200 0.027789 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 50.195  on 29  degrees of freedom
Residual deviance: 37.809  on 26  degrees of freedom
AIC: 125.76

Number of Fisher Scoring iterations: 5
```

```
> anova(ship3.fit, test = "Chisq")
Analysis of Deviance Table

Model: poisson, link: log

Response: incidents

Terms added sequentially (first to last)


         Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                        29     50.195
company   1   6.3339        28     43.861  0.01185 *
year.fac  2   6.0522        26     37.809  0.04850 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Beer Quality Data

Researchers were interested in determining whether or not the price of beer affected its perceived quality. In total, 60 people were given identical bottles of beer to try. They were split into two groups of 30: one group was told that the beer was cheap, and the other group was told that the beer was expensive. Each person then rated the beer as either 'good' or 'poor'.

The data are shown below.

		Quality	
		Poor	Good
Price	Low	24	36
	High	12	48

```
> beer.table
           quality.poor quality.good
price.low            24           36
price.high           12           48
> chisq.test(beer.table)

Pearson's Chi-squared test with Yates' continuity correction

data:  beer.table
X-squared = 4.8016, df = 1, p-value = 0.02843
```