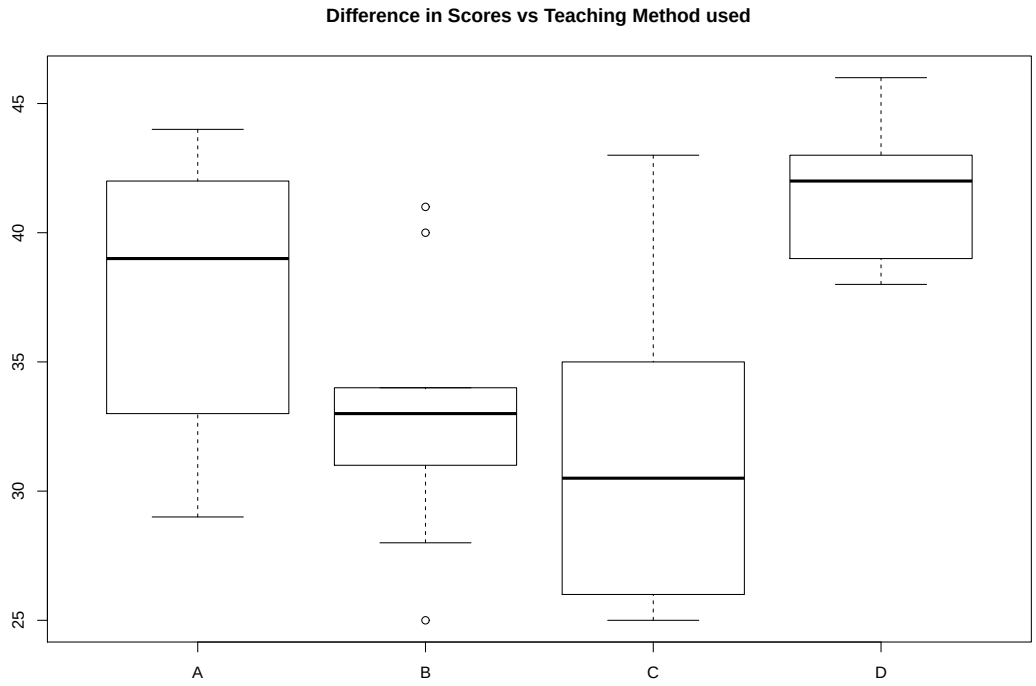
MAS programmes - Statistical Analysis (Autumn Term)

Exercises 6 - SOLUTIONS

1. Begin by entering the data into R and obtaining summary statistics and plots, as follows.

(a),(b)
```
> score <- c(33, 44, 39, 38, 29, 41, 39, 30, 42, 44,
             31, 33, 40, 34, 31, 41, 34, 28, 25, 33,
             32, 29, 34, 41, 27, 26, 43, 25, 35, 26,
             39, 42, 46, 42, 42, 46, 39, 43, 41, 38)
> method <- factor(c(rep("A", 10), rep("B", 10), rep("C", 10), rep("D", 10)))
# method<-factor(rep(c("A","B","C","D"),rep(10,4)))

> keyboard <- data.frame(score, method)
> rm(score, method)
> attach(keyboard)

> keyboard
   score method
1     33      A
2     44      A
3     39      A
4     38      A
⋮  output edited

38    43      D
39    41      D
40    38      D

> by(score, method, summary)
method: A
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  29.00   34.25   39.00   37.90   41.75   44.00
------------------------------------------------------------
method: B
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     25      31      33      33      34      41
------------------------------------------------------------
method: C
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  25.00   26.25   30.50   31.80   34.75   43.00
------------------------------------------------------------
method: D
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  38.00   39.50   42.00   41.80   42.75   46.00

> plot(method, score, main = "Difference in Scores vs Teaching Method used")
```



(c) An initial inspection of the data shows that Method D gives the most improvement. Its results are also more consistent in that the spread of the data is less than for the other methods. It remains to be seen whether this apparent improvement is statistically significant.

The established method, Method A, gives the second best results. Methods B and C give no improvement over the established method.

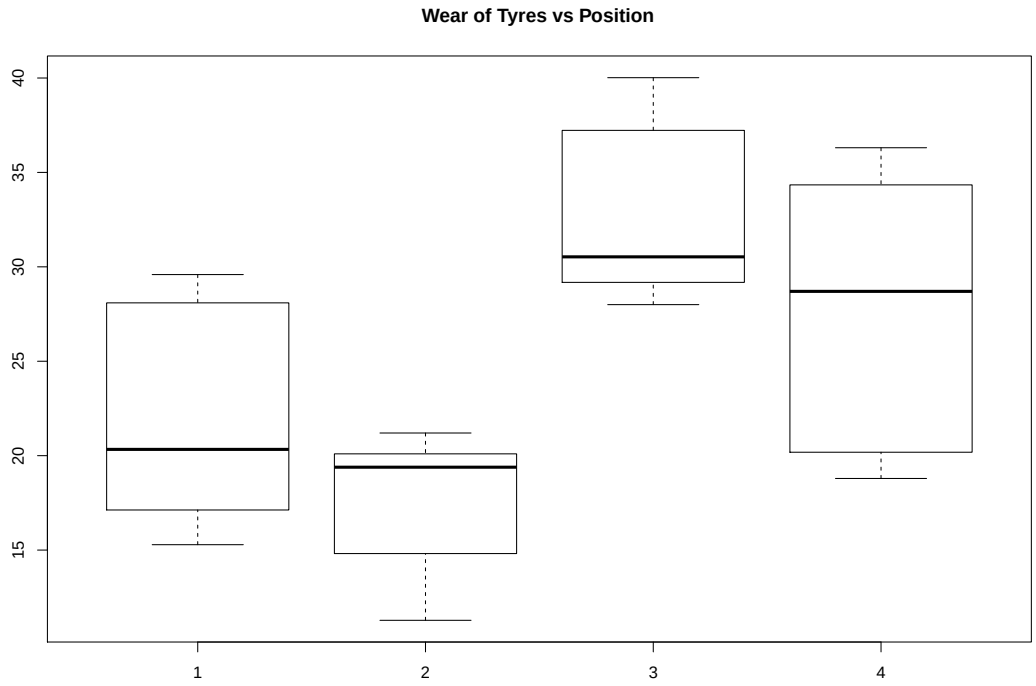
(d)
```
> keyboard.aov <- aov(score ~ method, data = keyboard)
> summary(keyboard.aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
method        3  638.3   212.8   8.343 0.000243 ***
Residuals    36  918.1    25.5
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

There is very strong evidence for overall differences between the methods ($p = 0.0002$).

2. Again, begin by entering the data in R and carrying out an exploratory analysis.

(a)
```
> wear <- c(20.935, 17.123, 29.59, 19.013, 15.919, 28.092, 20.332, 15.285, 28.304,
            18.279, 14.815, 19.973, 21.2, 11.28, 20.096, 19.389, 12.153, 20.477,
            28.535, 37.227, 30.529, 27.998, 38.853, 29.177, 30.073, 40.017, 30.795,
            20.182, 34.34, 29.023, 18.792, 34.707, 28.176, 19.203, 36.307, 28.701)
> position <- factor(c(rep(1, 9), rep(2, 9), rep(3, 9), rep(4, 9)))
# position<-factor(rep(c(1:4),rep(9,4)))

> tyres <- data.frame(wear, position)
> rm(wear, position)
> attach(tyres)

> by(wear, position, summary)
position: 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.28   17.12   20.33   21.62   28.09   29.59
------------------------------------------------------------
position: 2
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.28   14.82   19.39   17.52   20.10   21.20
------------------------------------------------------------
position: 3
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  28.00   29.18   30.53   32.58   37.23   40.02
------------------------------------------------------------
position: 4
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.79   20.18   28.70   27.71   34.34   36.31


> plot(position, wear, main = "Wear of Tyres vs Position")
```

It appears that rear tyres experience more wear than front tyres and that offside tyres experience more wear than nearside tyres.



(b)
```
> tyres.aov <- aov(wear ~ position, data = tyres)
> summary(tyres.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
position     3 1189.0   396.3   13.74 6.28e-06 ***
Residuals   32  923.1    28.8
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

There is very strong evidence of overall differences in wear among the four positions for the tyres ($p \ll 0.01$).