R：Statistical Programming Methods
R：程式、機率與統計

Significance Testing (2)

Type I and II Errors

	Null hypothesis H_0 is true	Null hypothesis H_0 is false
Reject Null hypothesis H_0	Type I Error (False Positive) ✗	Correct Outcome ✓
Fail to reject Null hypothesis H_0	Correct Outcome ✓	Type II Error (False Negative) ✗

Multiple Comparisons

- Every t-test will inflate type 1 error (the probability of finding a difference when in fact there is significant difference between the means)
- For Example:
  - There is only a 5% chance that the difference between the two means will be significant, and a 95% chance they will not be significant. What happens if we run the test again? The error inflation rate is modelled as:
  - $1 – (1-\alpha)^k$
- Where α is the accepted error rate, and k is the number of comparisons
- Then for three comparisons:
  - $1 – (1-0.05)^3 = 0.14$: **our error rate has been inflated**

ANOVA

- Analysis of Variance

- $F = \frac{variation\ between\ sample\ means}{variation\ within\ samples}$
  - Similar to p-value, to measures the extent to which differences among group means exceed what might be expected in a chance model

- However!
  - The ANOVA is simply an omnibus test. It only tells that there is a difference in the means, but not where.

F-ratio

- $F = \frac{MS_M}{MS_R} = \frac{(\frac{SS_M}{df_M})}{(\frac{SS_R}{df_R})}$
- $SS_M = \sum_{j=1}^{k} n_j(\overline{X}_j - \overline{X}_{gm})^2$
  - Where k is the number of groups, n_j is the number of observation in the group and gm is the grand mean (i.e., the mean of the group means)
- $SS_R = \sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2$
  - Where n is the number of observations in group j.
- $df_M = k - 1$ (number of groups - 1)
- $df_T = n - 1$ (all observations - 1)
- $df_R = df_T - df_M$
- From F-ratio to p-value, you can refer to Quick P-Value from F-Ratio Calculator (ANOVA) (socscistatistics.com).

mtcars: A data frame with 32 observations on 11 variables.

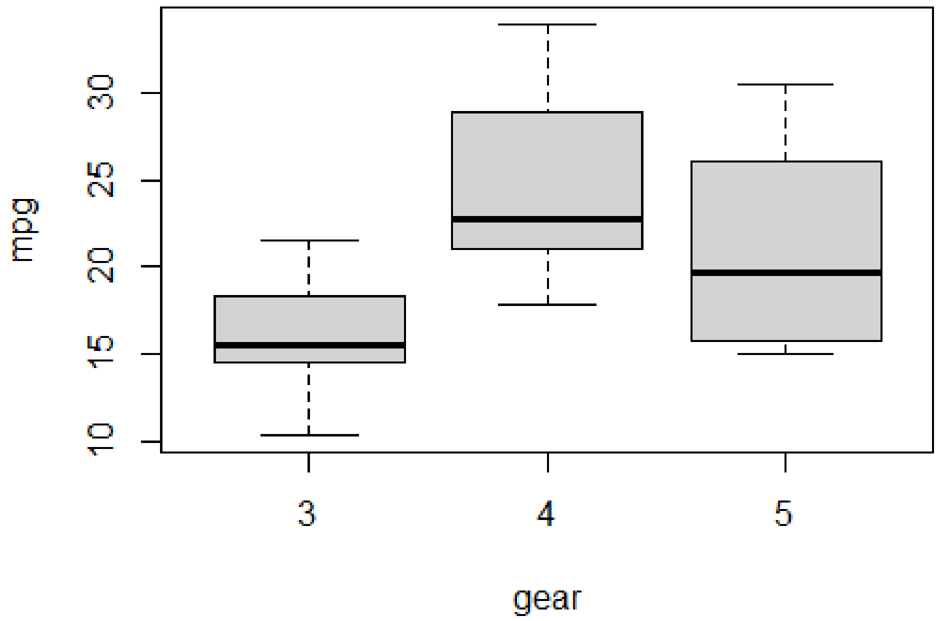
Column	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (1000 lbs)
qsec	1/4 mile time
vs	V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

Testing if mpg is difference with different gear

```
library(dplyr)
library(car)
df2 <- mtcars
#testing the homogeneity of variances
result <- leveneTest(mpg~factor(gear), data=df2)
result
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  1.4886 0.2424
##       29
```

Testing if mpg is difference with different gear

```
df2 <- mtcars
boxplot(mpg~gear, data=df2)
```



Testing if mpg is difference with different gear

```
run <- aov(mpg~factor(gear), data= df2)
summary(run)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## factor(gear)  2  483.2  241.62    10.9 0.000295 ***
## Residuals    29  642.8   22.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two-way ANOVA

- Testing if mpg is difference with different gear and transmission

```
#two-way ANOVA
run2 <- aov(mpg~factor(gear)+factor(am), data= df2)
summary(run2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(gear)  2  483.2  241.62  11.869 0.000185 ***
## factor(am)    1   72.8   72.80   3.576 0.069001 .  
## Residuals    28  570.0   20.36                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Post-hoc Tests

- ANOVA only tells that there is a difference in the means of groups, not WHERE that difference is.

- For example, the results could only that we reject null hypothesis:
  - $H0: \mu1 = \mu2 = \mu3 = \mu4 = \mu5 = \ \ldots = \mu n$

Tukey HSD

- Tukey's formula (similar to the t-test)
  - $q_s = \frac{\bar{X}_1 - \bar{X}_2}{SE}$, where X-bar 1 is the larger of the two means, X-bar 2 is the smaller of the two means, and SE is the standard error measurement derived by taking $\sqrt{\frac{MS_R}{n}}$ (mean square residuals / number of observations in the groups).

Tukey HSD – gear

```
#Tukey HSD
TukeyHSD(run)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mpg ~ factor(gear), data = df2)
##
## $`factor(gear)`
##          diff        lwr       upr     p adj
## 4-3  8.426667  3.9234704 12.929863 0.0002088
## 5-3  5.273333 -0.7309284 11.277595 0.0937176
## 5-4 -3.153333 -9.3423846  3.035718 0.4295874
```

Tukey HSD – gear and am

```
TukeyHSD(run2)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mpg ~ factor(gear) + factor(am), data = df2)
##
## $`factor(gear)`
##          diff        lwr       upr     p adj
## 4-3  8.426667  4.1028616 12.750472 0.0001301
## 5-3  5.273333 -0.4917401 11.038407 0.0779791
## 5-4 -3.153333 -9.0958350  2.789168 0.3999532
##
## $`factor(am)`
##         diff       lwr     upr     p adj
## 1-0 1.805128 -1.521483 5.13174 0.2757926
```