107-1 Statistics
LAB14: ANOVA

助教：廖皓宇、吳家禎、賴冠宇

2018/12/28

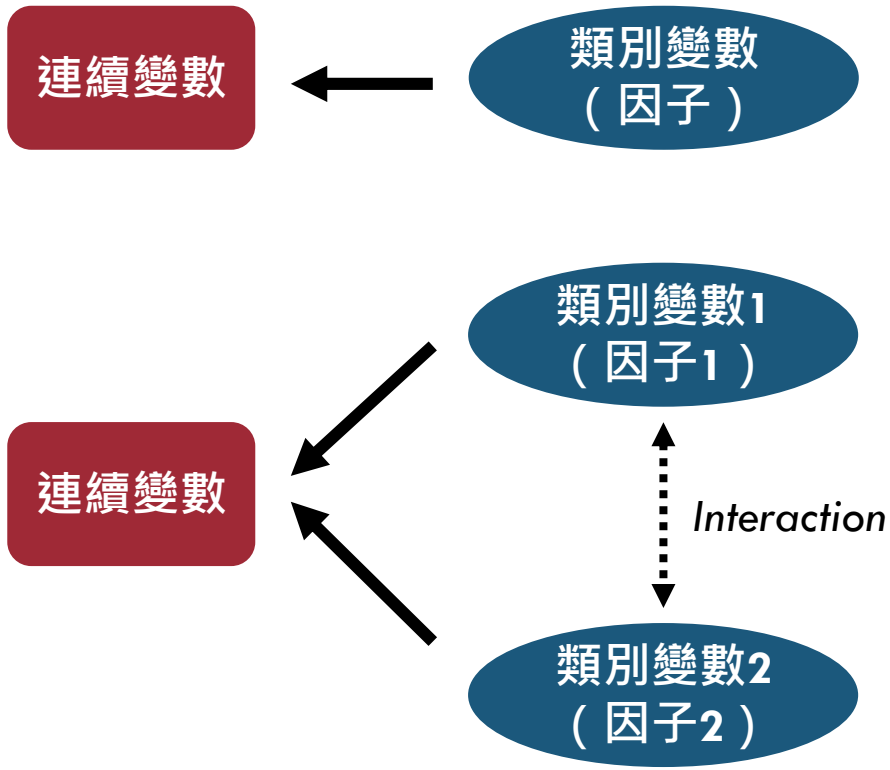
1228實習：變異數分析

- 使用提供的資料(Student.csv)
- 操作
  - Examining necessary conditions
  - One-way ANOVA
  - Two-way ANOVA
  - ANOVA via simple regression
- 期末教學意見調查

12/28 實習

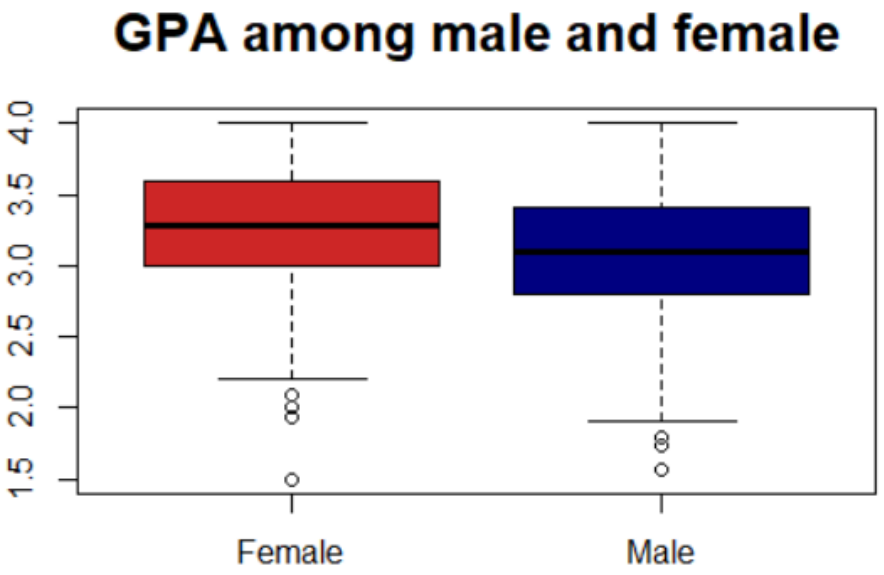
1. Examine necessary conditions
2. One-way ANOVA
   - ANOVA, simple regression
3. Two-way ANOVA
   - ANOVA, multiple regression

- *Example data: student.csv*



1. Examine necessary conditions

- Boxplot




```
# data distribution
boxplot(GPA ~ Sex, data = student,
        main = "GPA among male and female", cex.main = 1.6,
        col = c("Firebrick3", "Navy"))
```

1. Examine necessary conditions (cont.)

- Descriptive statistics

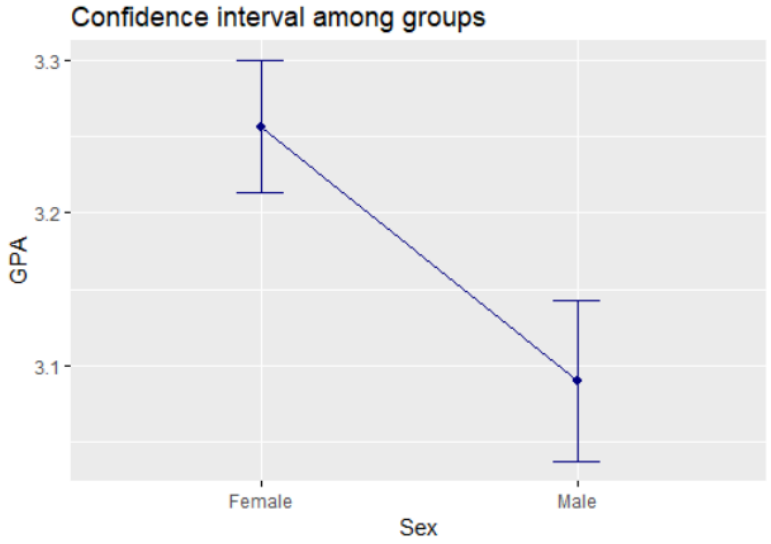
	n	mean	sd
Female	377	3.256525	0.4251325
Male	305	3.089803	0.4663943

```
# mean and sd for each group
stu.mean = tapply(student$GPA, student$Sex, mean)
stu.sd = tapply(student$GPA, student$Sex, sd)
stu.n = tapply(student$GPA, student$Sex, length)

stu = data.frame(n = stu.n, mean = stu.mean, sd = stu.sd)
stu
```

1. Examine necessary conditions (cont.)

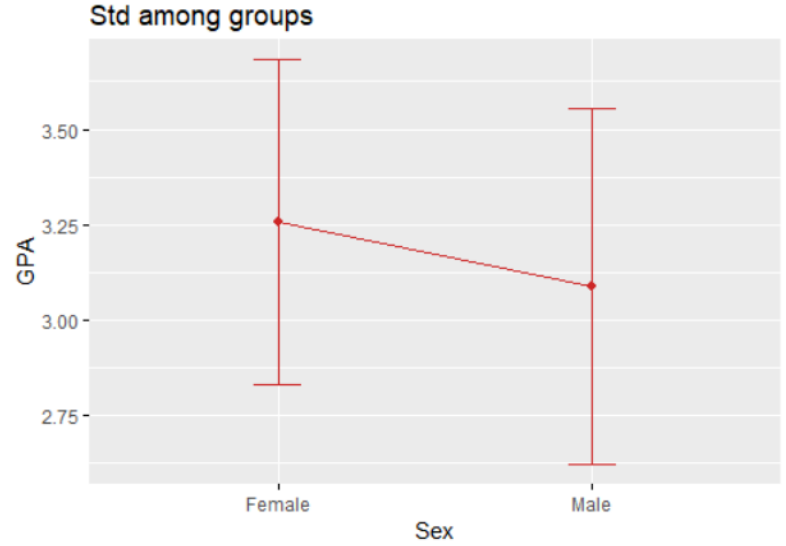
- Visualizing confidence interval




```
#install.packages("ggpubr")
library(ggpubr)

# visualize confidence interval for each group
ggline(student, x = "Sex", y = "GPA", add = c("mean_ci"),
       color = "navy", main = "Confidence interval among groups",
       ggtheme = theme_gray())
```

- Visualizing sd




```
# visualize std for each group
ggline(student, x = "Sex", y = "GPA", add = c("mean_sd"),
       color = "firebrick3", main = "Std among groups",
       ggtheme = theme_gray())
```

2. One-way ANOVA: F-test

→ Is the mean GPA different in sex? （某一類別變數X是否影響Y）

→ Is there a difference between male's GPA and female's GPA?（不同群的母體平均數是否相同）

Step1. If the factor has more than 2 classes:

H_0: $\mu_{class\ 1} = \mu_{class\ 2} = \cdots = \mu_{class\ k}$

H_1: $Not\ all\ \mu_i\ are\ equal.$

$(i = 1, 2, \ldots, k)$

Step 1.

- $Significance\ level$: $\alpha = 0.05$
- H_0: $\mu_{male} = \mu_{female}$
- H_1: $\mu_{male} \neq \mu_{female}$

Step 2.

- Checking conditions
  - (not skewed; no outliers; group standard deviations are not markedly different) ← *p.4-6 in this pdf.*
- F = 23.77 ← *ANOVA. aov()*

Step 3.

- $p\ value = 1.35 \times 10^{-6}$

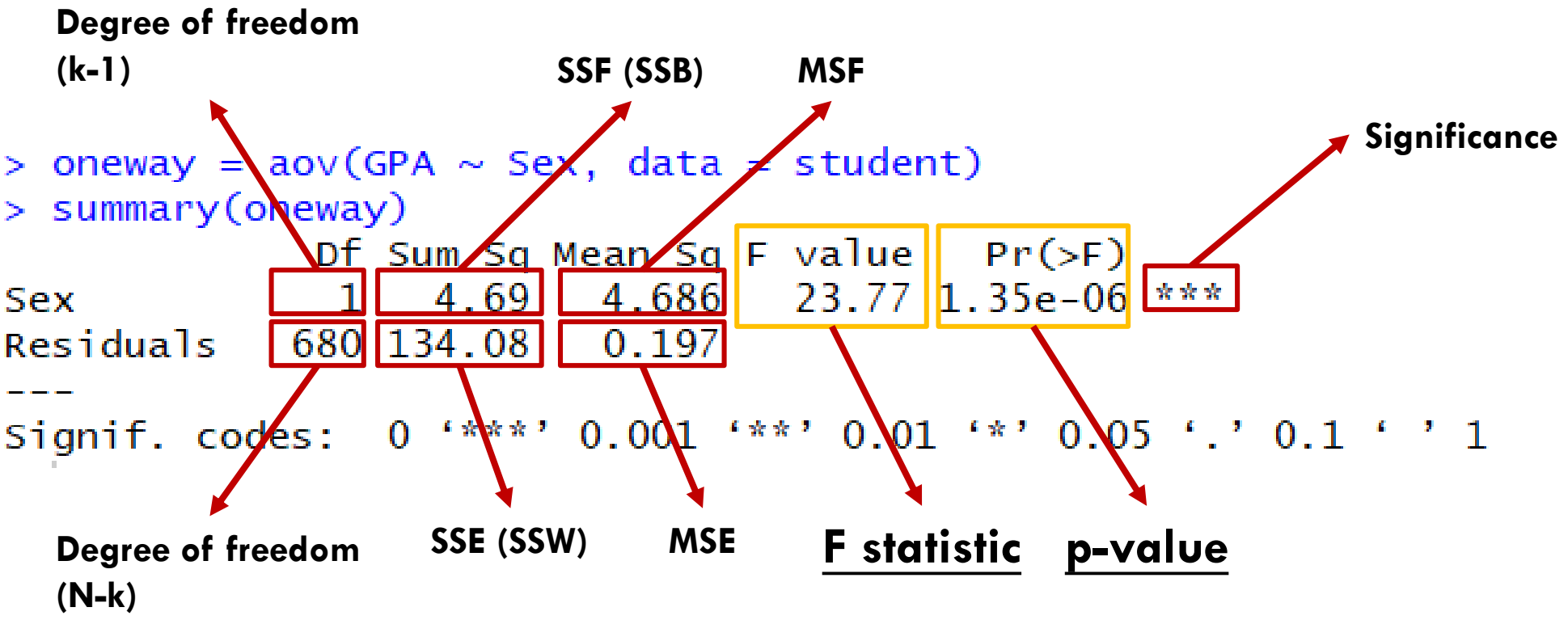
Step 4.

- $p\ value < \alpha.\ \ Reject\ H_0.$

Step 5.

- There's a statistically significant difference between male's mean GPA and female's mean GPA.

2. One-way ANOVA: Summary table



Simple regression

```
> results = lm(GPA ~ Sex, data = student)
> summary(results)

Call:
lm(formula = GPA ~ Sex, data = student)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7565 -0.2565  0.0202  0.3285  0.9102

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.25653    0.02287 142.394  < 2e-16 ***
SexMale     -0.16672    0.03420  -4.875 1.35e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4441 on 680 degrees of freedom
Multiple R-squared:  0.03377,   Adjusted R-squared:  0.03235
F-statistic: 23.77 on 1 and 680 DF,  p-value: 1.355e-06
```

整道回歸式的顯著性檢定

H_0: $\beta_0 = \beta_1 = 0$

H_a: $Not\ all\ \beta_i = 0$

$(i = 1,2, \dots, n)$

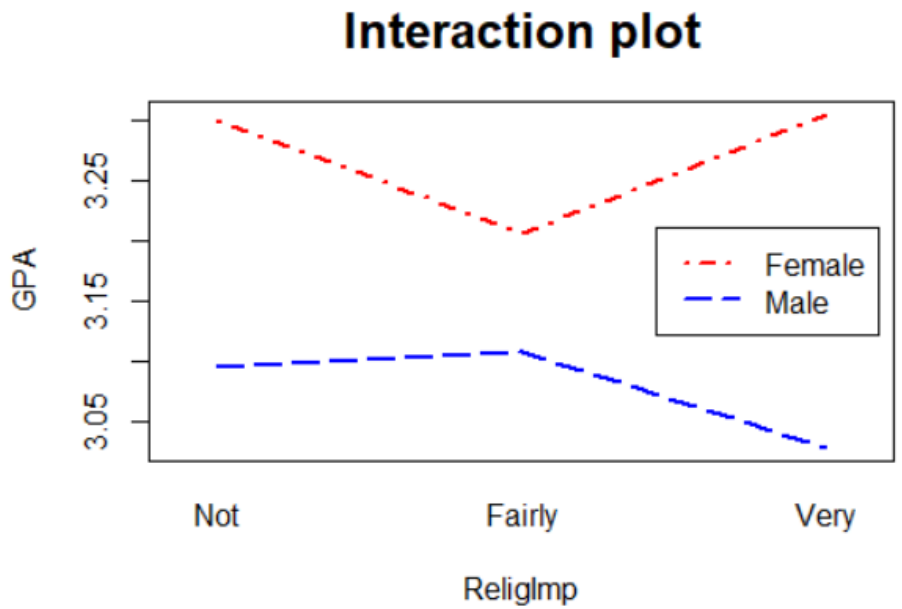
3. Two-way ANOVA

```
> twoway = aov(GPA ~ Sex + ReligImp + Sex:ReligImp, data = student)
> summary(twoway)
              Df Sum Sq Mean Sq F value   Pr(>F)
Sex            1   4.69   4.686  23.819 1.32e-06 ***
ReligImp       2   0.26   0.132   0.672    0.511
Sex:ReligImp   2   0.82   0.409   2.077    0.126
Residuals    676 133.00   0.197
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

交互作用項

```
#?interaction.plot
#interaction.plot(data$var1, data$var2, data$response)
interaction.plot(student$ReligImp, student$Sex, student$GPA,
                 #leg.bty = "o",
                 legend = F, lty = c(4,5), lwd = 2, col = c("red", "blue"),
                 xlab = "ReligImp", ylab = "GPA",
                 main = "Interaction plot", cex.main = 1.6)

legend("right", inset = 0.02,
       legend = c("Female", "Male"), col = c("red", "blue"),
       lty = c(4,5), lwd = 2)
```



Multiple regression

```r
#factor (base class for regression)
unique(student$ReligImp)
student$ReligImp = factor(student$ReligImp, levels = c("Not","Fairly", "Very"))
unique(student$ReligImp)
```

→ *Setting the base class for categorical variables*

```
> results2 = lm(GPA ~ Sex + ReligImp + Sex:ReligImp, data = student)
> summary(results2)

Call:
lm(formula = GPA ~ Sex + ReligImp + Sex:ReligImp, data = student)

Residuals:
     Min       1Q   Median       3Q      Max
-1.70687 -0.26751  0.00402  0.31525  0.90402

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           3.299038   0.043495  75.848  < 2e-16 ***
SexMale              -0.203056   0.059778  -3.397 0.000722 ***
ReligImpFairly       -0.092167   0.054690  -1.685 0.092399 .
ReligImpVery          0.005004   0.063126   0.079 0.936840
SexMale:ReligImpFairly  0.103905   0.078226   1.328 0.184539
SexMale:ReligImpVery   -0.071949   0.097212  -0.740 0.459484
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4436 on 676 degrees of freedom
Multiple R-squared:  0.04157,   Adjusted R-squared:  0.03448
F-statistic: 5.863 on 5 and 676 DF,  p-value: 2.581e-05
```

因子1 → **Base: *Female***

因子2 → **Base: *Not***

因子1及因子2的交互作用項

整道回歸式的顯著性檢定

H_0: $\beta_0 = \beta_1 = \cdots = \beta_n = 0$

H_a: $Not\ all\ \beta_i = 0$

$(i = 1,2,\ldots,n)$

期末教學意見調查

- 「myNTU/學生專區/課務資訊/期末教學意見調查」

or

- 網址： https://investea.aca.ntu.edu.tw/opinion/login.asp

- 以個人計中帳號密碼登入

作業13 變異數分析

- 練習題4題(Ch. 16)
  - 16.8; 16.24; 16.36; 16.52 (c: 5 steps)
- 準備期末考：加油！