Week5

Salary < gender (M F), race

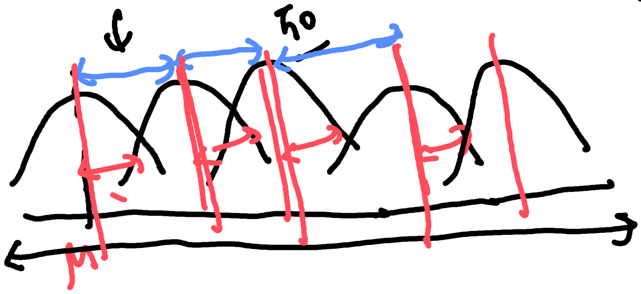
no gender effect

two-sample t-test $[H_0: \mu_M = \mu_F$ vs. $H_1: \mu_M \neq \mu_F]$ $>$, $<$

- ANOVA
  - $H_0: \mu_F = \mu_M$ $\quad H_a: \mu_F \neq \mu_M$
  - $H_0: \mu_A = \mu_B = \mu_W$ $\quad H_a:$ race
  - $H_0:$ no interaction effect

	W	B	A
F	(W, F)	(B, F)	(A, F)
M	(W, M)	(B, M)	(A, M)

μ_0



$Y_{(F,W)i}$ ← ith observ

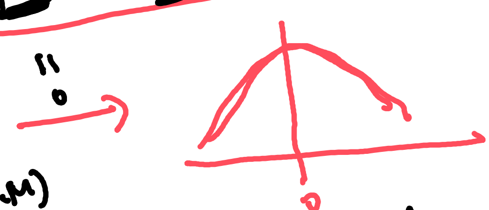
$= \mu_{(F,W)} + \text{Error}$

fixed unknown

individual-specific

$= \mu_0 + \mu_F + \mu_W + \mu_{F,W}) + \text{Error}$

$= 0$



$\mu_{(A,M)} = \mu_0 + \mu_A + \mu_M + \mu_{A \times M}$

- Model Assumptions

(1) Y is continuous numeric

(2) Data ~ Normal Each combinat ~ Normal

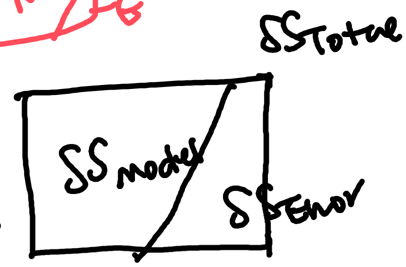
(3) Equal variance → If it is violate → Welch's ANOVA

(4) indep samples

- ANOVA
  - (1) Test main or interaction effect
  - (2) $R^2 = \dfrac{SS_{model}}{SS_{Total}}$
  - (3) post-hoc test
  - (4) Model diagnostics
    - Normality
    - Equal variance

$$SS_{Total} = SS_{Model} + SS_{Error}$$

- SS_{Total}: Variation of Y
- SS_{Model}: variation explained by Model ≈ 0 (race·gender)
- SS_{Error}: Not explained by Model



1-way ANOVA: Toothlength ~ Dose (0.5 / 1 / 2)

$$SS_{Total} = SS_{Dose} + SS_{Error}$$

$$F = \frac{SS_{Model\,(=Dose)} / df(D)}{SS_{Error} / df(Error)}$$

F ↑

$F \approx 0$

$\mu_{(F,W)}$

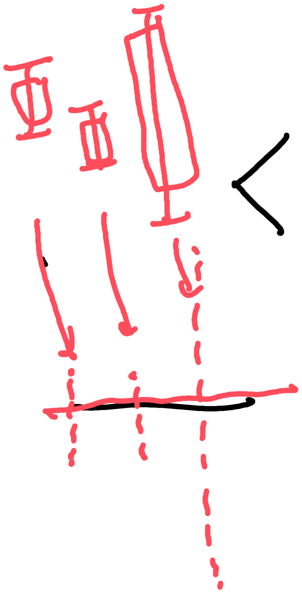
$\hat{\mu}_{(F,W)}$ sample 50

- Model diagnostics → Residuals ≈ Error

$$Y_{(F,W)i} - \hat{\mu}_{(F,W)i}$$

Estimated

(centering each group)



Normality check QQ-plot

Equal var check

(1) Leven's test

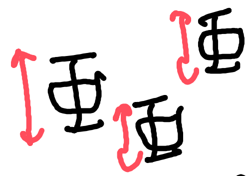
H_0: Equal var

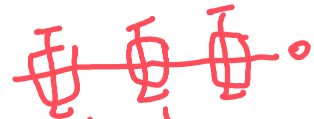
H_a: u

(2) Residual plots

1-way

2-way Balo





■ Post-hoc test

H_0: $\mu_{0.5} = \mu_1 = \mu_2$ (no Dose effect)

H_a: at least one group has different mean

5 $_5C_2$ 0.5 1 2 $_3C_2$

$$H_0: \begin{cases} \mu_{0.5} = \mu_1 \\ \mu_{0.5} = \mu_2 \\ \mu_1 = \mu_2 \end{cases}$$

$\mu_1 > \mu_{0.5}$

$\mu_2 > \mu_{0.5}$

$\mu_2 > \mu_1$

$\mu_2 > \mu_1 > \mu_{0.5}$

(eg) comparing two fruits

H_0: size ↔ H_0: size, color, sweetness, weight

p-value

$\boxed{\alpha}$

p-value

✓ 1 − 0.5 p-val 0.7 → ✓ [$\mu_1 = \mu_{0.5}$, $\mu_2 = \mu_{0.5}$] ✓

✓ 2 − 0.5 0.5

✓ 2 − 1 p 0.001 → ✓ $\mu_2 > \mu_1$

Two-way ANOVA (Balanced)

Full model:

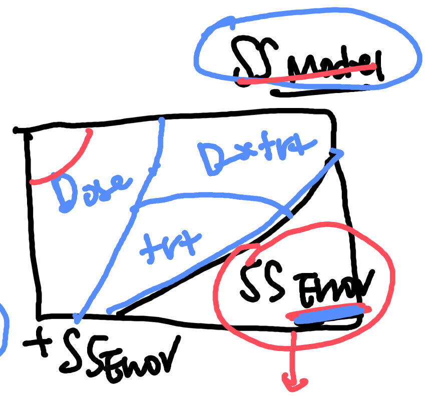
Toothlength ~ Dose + treatment +

Dose: (0.5/1/2) treatment: (OJ/VC)

$$\mu_{Toothlength} = \mu_0 + \mu_{Dose} + \mu_{trt} + \mu_{D*trt}$$

$$SS_{Total} = SS_{model} + SS_{Error}$$

$$= SS_{Dose} + SS_{supp} + SS_{D*trt} + SS_{Error}$$



rando Error

(i) Dose effect [$H_0: \mu_{0.5} = \mu_1 = \mu_2$ vs. H_a: not H_0]

$$F\text{-stat} = \frac{SS_{Dose}/df(2)}{SS_{Error}/df\ 12}$$

small p-val.

sig Dose eff.

(ii) Supplement effect

[$H_0: \mu_{OJ} = \mu_{VC}$ vs. $H_a: \mu_{OJ} \neq \mu_{VC}$]

$$F\text{-stat} = \frac{SS_{supp} / df(supp)}{SS_{Error} / df(E)}$$

sig Supplement eff

(iii) Dose * Supplement effect

H_0: no interaction vs. H_a: interaction

$$F\text{-stat} = \frac{SS_{Inter}}{SS_{Error}}$$

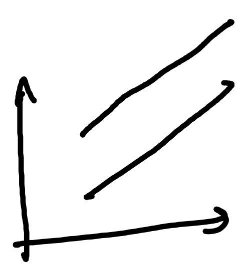
sig. Interaction eff

Dose $\mu_1 > \mu_{0.5}$ → $\mu_2 > \mu_1 > \mu_{0.5}$

$\mu_2 > \mu_{0.5}$

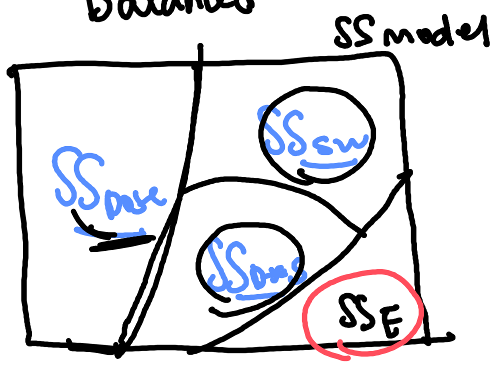
$\mu_2 > \mu_1$

Su $\mu_{VC} < \mu_{OJ}$

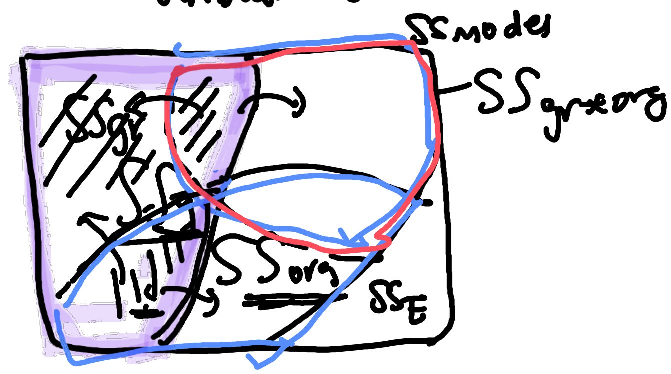


Balanced / Unbalanced ANOVA same goal

Balanced



Unbalanced



[Toothlength ~ Dose + Supp + Dose*Supp]

$SS_{Dose} + SS_{Supp} + SS_{D*S}$

"

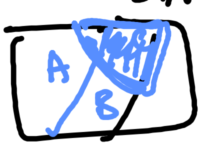
SSmodel

[days ~ gr + org + gr*org]

$SS_{gr} + SS_{org} + SS_{g*or}$

SSmodel

SSn



$Y \sim A + B + C$

$\boxed{SS(C)} = SS[C|A]$

$= SS[C|B]$

$SS(C) = SS[C|AB]$

SSError

efe

$\boxed{C}$

$SS(C) \neq SS[C|A]$

$\neq SS[C|B]$

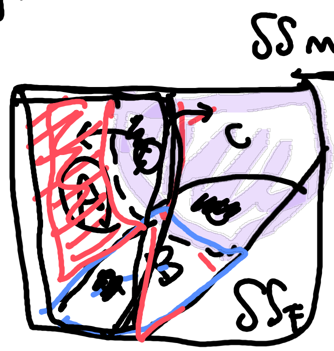
$\neq SS[C|AB]$

H_0: no C effect

H_a: sig C efe

$$F\text{-stat} = \frac{SS(C\cdot)}{SSError}$$

V.



$Y \sim A + B + C$

$SS_{model} \neq SS(A) + SS(B) + SS(C)$

$SS_{model} = SS(C) + SS(B|C) + SS(A|BC)$

$= SS(A) + SS(B|A) + SS(C|AB)$

$= SS(B) + SS(C|B) + SS(A|BC)$

$Y \sim A + B + C$

Type I "Sequential SS"

	DF	SS	MS	F	pval
A		SS(A)		F=0/	
B		+ SS(B|A)			
C		+ SS(C|AB)			

= SSmodel

Type III "Conditional SS"

	DF	SS	MS	F	pval
A		SS(A|BC)			
B		+ SS(B|AC)			
C		+ SS(C|AB)			

≠ SSmodel

$Y \sim C + A + B$

	DF	SS	MS	F	pval
C		SS(C)			
A		+ SS(A|C)			
B		+ SS(B|AC)			

= SSmodel

	DF	SS	MS	F	pval
C		SS(C|AB)			
A		SS(A|BC)			
B		SS(B|AC)			

day ~ grade + origin

SS(grade) + SS(origin | grade)

type I

grade SS(grade) 2277

origin SS(origin | grade) 2389

type II

SS(grade | origin) 2020

SS(origin | grade) 2389

grade effect

$H_0: \mu_{F0} = \mu_{F1} = \mu_{F2} = \mu_{F3}$

vs.

$H_a:$

origin effect

$H_0: \mu_A = \mu_N$

$H_a: \mu_A \neq \mu_N$