Week 4.

one-sample t-test (normality) $H_0: \mu = 50{,}000$ $H_a: \mu \neq 50{,}000$ ($>$, $<$)

two-sample ~~t-test~~ $H_0: \mu_F = \mu_M$ $H_a: \mu_F \neq \mu_M$ ($>$, $<$)

both group ~normal F M

Equal var → pool

× → Sa-

ANOVA F-test

$H_0: \mu_F = \mu_M$ (gender) → no gender $H_a: \mu_F \neq \mu_M$ → sig gender effect

→ $H_0: \mu_W = \mu_B = \mu_A$ (race) → no race effect

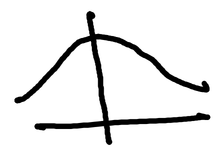
H_a: at least one race difference → sig race effect

p-value

(Y) response

• Salary ← Gender (F / M), Race (W / B / A)

parametric model



μ

(G) \	W (μ_0)	B (μ_0)	A (μ_0)
F	20 (F.W)	20 (F.B)	20 (F.A)
M	20 (M.W)	20 (M.B)	~~20~~ 19 (M.A)

obs) 1, 2, …, 20

Balanced → Balanced ANOVA

unbalanced → Unbalanced ANOVA

Equal var.

$\mu_{(F.W)}$ $\mu_{(M.A)}$

ANOVA model formulation

Salary $Y_{(F.W)i} = \mu_{(F.W)} + \varepsilon_i$ random error, $i = 1:20$

obs) … 20

$Y_{(M.A)i} = \boxed{\mu_{(M.A)}} + \varepsilon_i$

Decomposition of mean

2-way ANOVA w/ interaction

Full model

$$\mu_{(F,W)} = \underbrace{\mu_0}_{\text{baseline}} + \underbrace{\overset{\text{gender effect}}{\mu_F} + \overset{\text{race effect}}{\mu_W}}_{\text{main effect}} + \overset{\text{Interaction effect}}{\mu_{F*W}}$$

$\vdots$

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

Interaction effect

cheese 10g : cheese effect

caramel \ cheese	0	1
0	10g	20g
1	30g	60g

+20g : caramel

Response (Y) : How much eat (g)

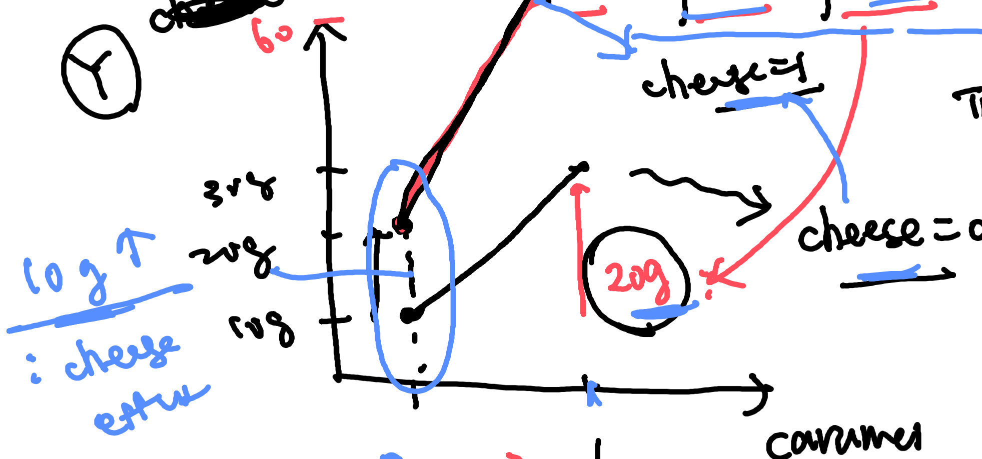
cheese < no cheese (0), cheese (1)

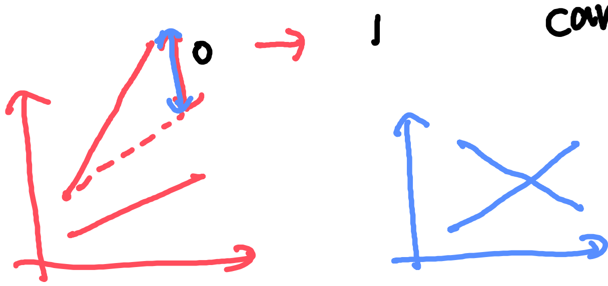
caramel: no caramel (0), caramel (1)

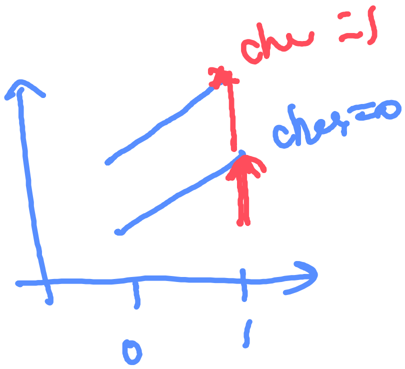
$$\overset{10\,g}{\mu_0} + \overset{10\,g}{\mu_{ch}} + \overset{20\,g}{\mu_{ca}} + \overset{20g}{\mu_{ch*ca}}$$

cheese=1

Interaction plot



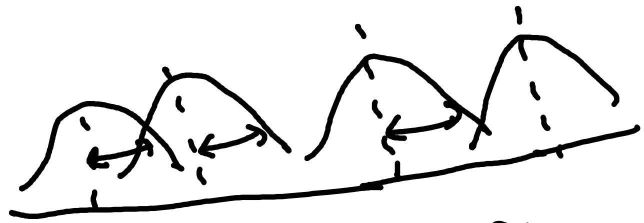




Goal: identify significant effect on Y

- ✓ main
- ✓ interaction

⇒ Ho, Ha



4 levels in race

• ANOVA Assumptions

✓ 1) Y (response) is continuous

2) Normality of each group

3) Equal variance (Homoscedasticity) Homogeneity..

✓ 4) iid sample

• How to check them?

→ manual check

→ diagnostics plot

ANOVA

ANOVA → output → diagnostics plot

1-way ANOVA : levene's test

• Balanced Data

salary ~ race

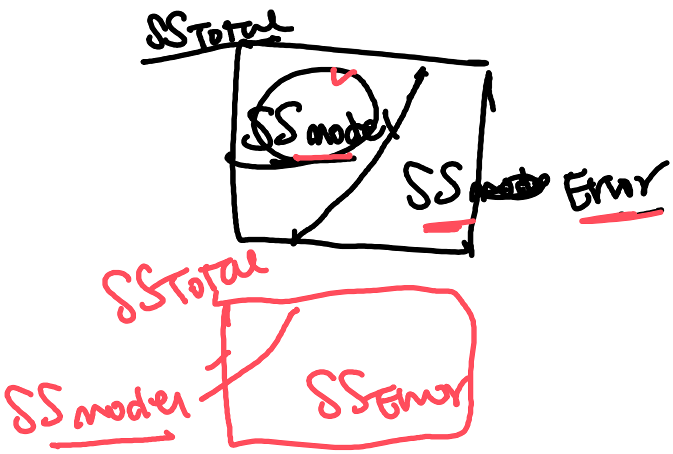
systematic variation | random variation

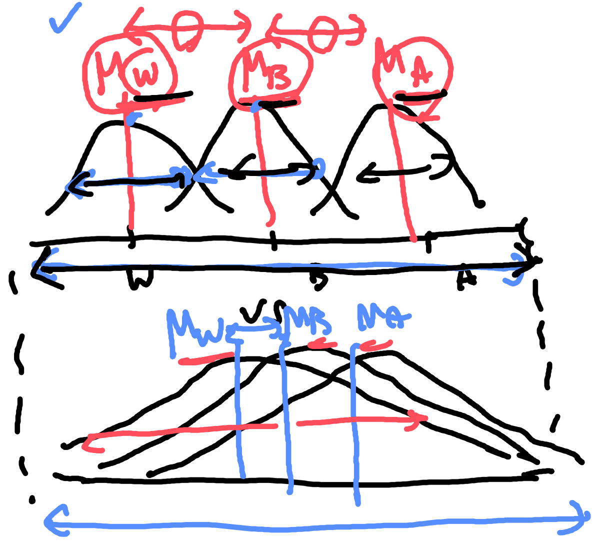
$$SS_{Total} = SS_{Model} + SS_{Error}$$

SS_{Total}: variation of Y — Total variance of Y

SS_{Model}: btw group variation — variation of Y can be explained by model (race)

SS_{Error}: Within group variation — variation of Y can NOT be explained by race model





(eg) 1way ANOVA

Salary ~ Race (W/B/A)

$H_0: \mu_W = \mu_B = \mu_A$ (no race effect)

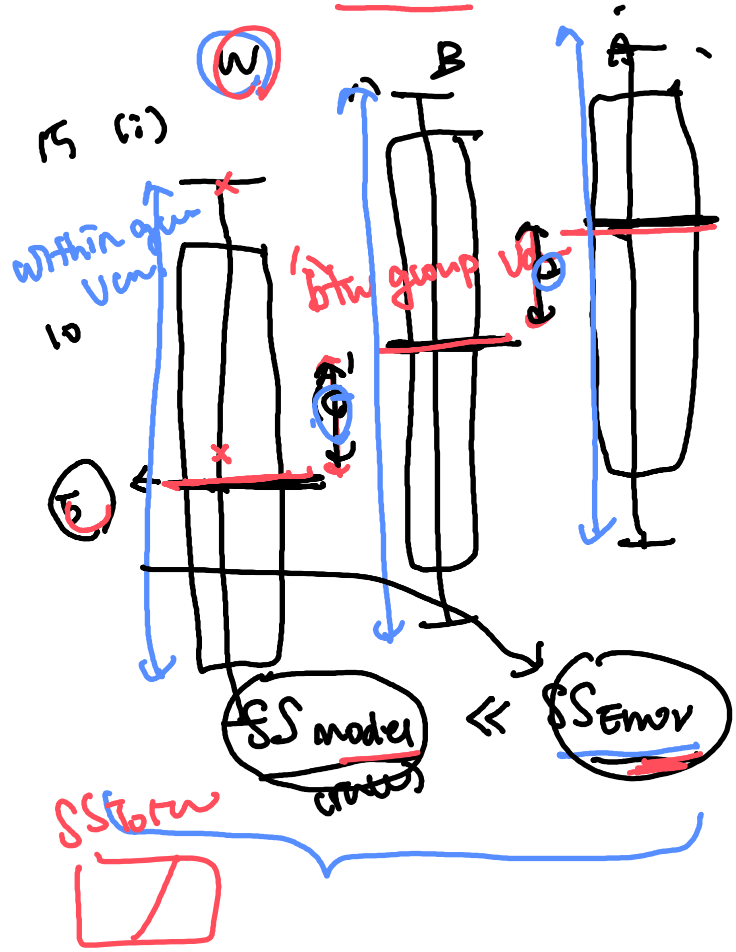
H_a: at least one ..

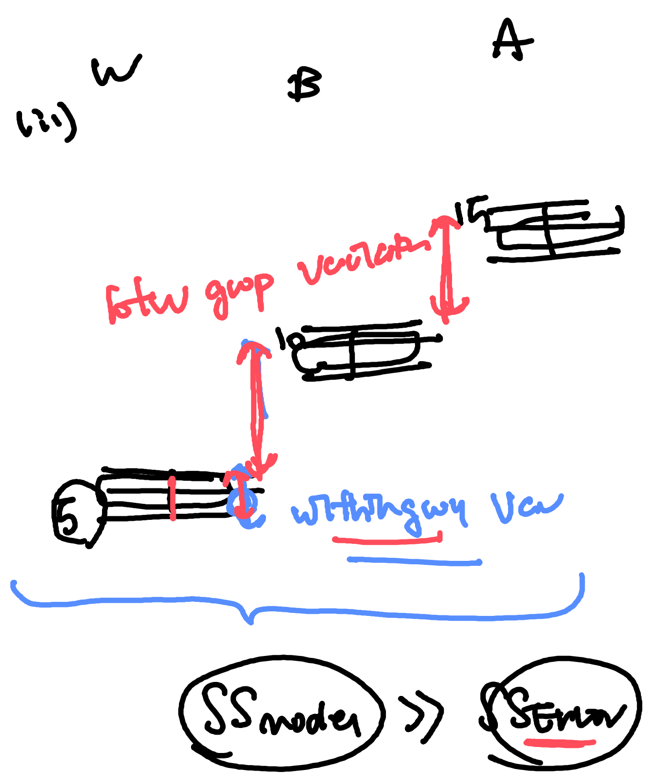
n=100

W: $\bar{Y}_W = 5$

B: $\bar{Y}_B = 10$

A: $\bar{Y}_A = 15$





$SS_{model} \ll SS_{Error}$ (i)

$SS_{model} \gg SS_{Error}$ (ii)

p-value

F-test statistic

$F \approx 0 \rightarrow$ support H_0

[F large $\rightarrow$ reject H_0

$$F = \frac{\text{btw group variation } (SS_{model}) / df_{model}}{\text{within group variation } (SS_{Error}) / df_{Error}}$$

p-value $\alpha = 0.05$

Among (i), (ii), which data lead larger F ?

smaller F ?

SS_{Total}

Y $\boxed{\cancel{SS_{model}}}$ SS_{Error}

$$R^2 = \frac{SS_{model}}{SS_{Total}} \le 1$$

$0 \le$

R^2 : R^2 5%

R^2 70%. 40%

[$R^2 = 70\%$.

$R^2 = 99\%$.

$n = 50$

Data example. 1-way

Toothlength ~ Dose (0.5) (1) / (2)

(0) Balanced ? Note: Bal Unba

(1) Test significant effect of Dose

$H_0: \mu_{0.5} = \mu_1 = \mu_2$

H_a: at least

1-way ANOVA → Levene's test

Equal var. →

Unequal var. → Welch's ANOVA

(2) R^2

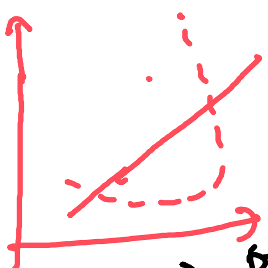
(3) Model diagnostics

- diagnostics plot
- Levene's test (for 1-way ANOVA)

→ Equal var.

H_0:

H_a:



→ p-val: 0.0001 → Unequal var.