Dan Schumacher

9/21/23

Algorithms 1

Announcements

HW2 Due 9/30/23

Review:

Two sample t-test $H0: \mu_m = \mu_f$ vs $\neq$

ANOVA

$$\left\{\begin{array}{ll} H0: \mu_f = \mu_m & \text{v.s. } \neq \\ \text{or} & \\ H0: \mu_a = \mu_B = \mu_w & \text{@least 1} \neq \\ \text{or} & \\ H0: \text{No Interaction effect} & \text{Is Interaction} \end{array}\right\}$$

	W	B	A
F		combos	
M	combos		

changes depending on observation

$$Y_{(F,W)i} = \underbrace{\mu_{(F,W)}}_{\underset{\text{overall mean}}{\mu_0} + \mu_F + \mu_W + \mu_{F,W}} + \text{Error}$$

Assumptions

1.) Y continuous numeric 2.) Norm Dist per combo

3.) Equal Var 4.) Indep Samples

↳ if Not, Welch's ANOVA

$\boxed{\text{F-statistic}}$

Some df left out

$$F = \frac{SS_{model}}{SS_{error}}$$

if $F \approx 0$, P High No Rej H0

if $F \uparrow$ large, P Low Rej H0

Actual Test

1) Main vs Interaction
2) R^2
3) Post HOC
4) Model Diagnostics < Normality, Equal Var

ideally do first

$$\underset{\text{Var } Y}{SS_{Total}} = \underset{\text{Var explained by model}}{SS_{model}} + \underset{\text{No explained}}{SS_{error}}$$

Schedule Flu + Covid Booster

John = Black Male, 50,000$
Avg BM = 44,000$
John's Residual = 6,00$
distance of observation from their group mean

Model Diagnostics

Looks @ residuals
Use to check Assumptions
Normality check

Equal Variance check

Levene's Test → 1-Way ANOVA or 2-Way Balanced
Ho : Equal Var
Halt: Not =

otherwise

```
Par(mfrow=c(2,2))
Plot(aov.data, 1)
```

which of 2 plots want

POST HOC TEST

We know 1 group has dif mean. Which one? is it up or down?

Only when reject HO ANOVA

Pair wise test
↳ Test all @ once

$$HO: \begin{cases} \mu_B = \mu_A \\ \mu_B = \mu_W \\ \mu_W = \mu_A \end{cases} \quad Halt: \mu_x \neq \mu_y$$

Example
Ho (size, color, smell)

with more categories, more likely to rej HO, need to adjust alpha → Scheffe + Turkey's Method
↳ adjust α and p-Val acordingly

in R: SchefeTest(aov.data) (results) or TurkeyHSD(aov.data)

↳ Results look like

Dose	diff	pval
$1_\mu - S_\mu$ =	9.13	if <.05, significant
$2_\mu - S_\mu$ =	15.5	=
$2_\mu - 1_\mu$ =	6.36	=

$\mu_2 > \mu_1 > .S$

Don't mention one that isn't significance
→ Summary

Dan Schumacher
9/21/23
Alg 1

Two Way ANOVA Balanced

Toothlength ~ Dose + Treatment + $\overset{\text{interaction}}{(\text{Dose} + \text{Treatment})}$

$\mu = \mu_0 + \mu_{dose} + \mu_{treatment} + \mu_{error}$

$SSTotal = SS_{model} + SS_{error}$

Robust

$[SS_{dose} + SS_{treatment} + SS_{interaction}]$

Alt Toothlength ~ Dose + Sup + Dose*Sup

$\boxed{\text{in R}}$: aov(Toothlength ~ Dose * Suppliment, data = tooth)

Dose effect

if { F Large → Big dose effect; F ≈ 0 → No dose effect }

usually don't interpeleat interaction effect in Post Hoc

Unbalanced ANOVA

if table(data$col1) ≠ table(data$col2)

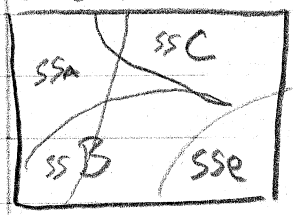
if Sample sizes aren't equal

Types sum of Squares

$SS(c) = SS(C \mid A\ B)$ if unbalanced

$SS(c) \neq SS(C \mid AB)$ if balanced

Unbalanced



$$SS_{model} = SS(A) + SS(B) + SS(C)$$
$$= SS(A) + SS(B|A) + SS(C \cdot AB)$$
or
$$SS(b) + SS(A|B)\ SS(C|AB)$$

order matters

Type 1 (sequential SS)

$Y \sim A + B + C$

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

or

$Y \sim C + A + B$

C $SS(C)$
A $SS(A|C)$
B $SS(B|AC)$

$+ = SS\,model$

order doesn't matter

Type 3 (Conditional SS)

A $SS(A|BC)$
B $SS(B|AC)$
C $SS(C|AB)$

$+ = < SS\,model$

Pros: don't have to order

Cons: doesn't use all information

in R:

Anova(aov.res, type=3)

Good to use both with diferent orders

if 2-Way ANOVA, Type 1 order only has small effect.

grade $SS(grade)$
origin $SS(origin|grade)$