Contents

Day 21 **1**

ANOVA $\rightarrow$ Analysis of VAriance **2**

Use cases . . . 2

Notation . . . 2

Hypothesis Testing . . . 3

Implicit Assumptions of Model . . . 3

Day 21

ANOVA → Analysis of VAriance

THE classic Fisher test.

Model: $y \sim N(\mu, \sigma)$

- μ and σ are unknown.
- μ may not be the same for all data points
- σ is assumed same for all data points

In one-way ANOVA:

- Data: one numerical response variable `y` and one categorical explanatory variable whose values are the "groups".
- We need ≥ 2 groups

Example situations where we use it:

- Compare control group to > 1 treatment group
- Observational study comparing 3 or more groups/populations

Use cases

- Between-group effects: variation due to changes in μ
- Within-group effects: variation due to individual differences

Notation

- $\bar{y}$ = grand mean or the mean of all data in the whole sample.
- N = total sample size
- I = number of groups
- $\bar{y}_{\mathrm{i}}$ = sample mean in group `i`
- s_{i} = sample standard deviation in group `i`
- n_{i} = sample size in group `i`
- y_{ij} = value of y for the j^{th} case in group `i`.

Hypothesis Testing

$\mathrm{H_0}$: $\mu_1 = \mu_2 = ... = \mu_{\mathrm{I}}$

- All the population means are equal $\implies$ no effect of group on response.
- Under $\mathrm{H_0}$, $y_{\mathrm{ij}} \sim N(\mu, \sigma)$
- Also: $\bar{y}_{\mathrm{i}} \sim N(\mu, \frac{\sigma}{\sqrt{n_{\mathrm{i}}}})$
- μ, σ are fixed but unknown

$\mathrm{H_a}$: <u>not</u> $\mathrm{H_0}$ [not really necessary because this is Fisher framework]

- $\implies$ effect of group on response

Implicit Assumptions of Model

- Normal population distribution
- σ is the same for all groups [not as critical]
  - robust to violations of this assumption as long as the largest $s_{\mathrm{i}} < 2 \times$ smallest s_{i}