Lecture 33

Single Factor Experiments & ANOVA

- In this section we examine single factor experiments
- In an experiment, a "factor" is a variable that the experimentor controls; usually to gain information about a "response" variable.
- the factor takes a few number of levels, called treatments

Eg:

$\underbrace{\text{cement cure time}}_{\text{response variable}}$ — $\underbrace{\text{temperature}}_{\text{factor}}$ — $\underbrace{\text{cold, warm, hot}}_{\text{treatments}}$

$\underbrace{\text{symptoms}}_{\text{response}}$ — $\underbrace{\text{drug dosage}}_{\text{factor}}$ — $\underbrace{\text{low, med, high.}}_{\text{treatments.}}$

$\underbrace{\text{academic performance}}_{\text{response}}$ — $\underbrace{\text{weekly exercise}}_{\text{factor}}$ — $\underbrace{0, 1, 2, \dots, 7 \text{ days/week}}_{\text{treatments}}$

- Let $a = \#$ of treatments.
- For the i^{th} treatment, we get a random sample

$$\{Y_{i1}, Y_{i2}, \ldots, Y_{in_i}\}$$

where n_i is the size of the sample for the i^{th} treatment $(1 \le i \le a)$. (so Y_{ij} is the j^{th} observation at the i^{th} treatment.

- to simplify, we assume $n_i = n_j = n$ for all i, j.
- we assume a "linear statistical model" of the form

$$Y_{ij} = \underbrace{\mu + \tau_i}_{\mu_i} + \varepsilon_{ij} \quad \begin{cases} i \in \{1, \ldots, a\} \\ j \in \{1, \ldots, n\}. \end{cases}$$

↳ μ is a common mean across all treatments

↳ τ_i is a parameter associated to the i^{th} treatment called the i^{th} treament effect

↳ ε_{ij} is a random error term.

- we assume that the error terms are independent and identically distributed with normal distribution $N(0, \sigma^2)$ (this is a simplifying assumption... a completely randomized experiment).
- note: can also write $\mu_i = \mu + \tau_i$, where now

μ_i is the mean of the response associated to the i^{th} treatment.

↳ we may therefore think of the i^{th} treatment as distributed as $N(\mu_i, \sigma^2)$.

- we assume that the experimenter specifically chose the a treatments and that they want to test hypotheses about the treatment means μ_i, or estimate the treatment effects (a <u>fixed-effects model</u>)

- <u>Goal</u>: develop ANOVA for fixed-effects models.
- we assume that $\sum_{i=1}^{a} \tau_i = 0$.
- we want to test the hypothesis $\mu_1 = \mu_2 = \ldots = \mu_a$.
- since $\mu_i = \mu + \tau_i$, this is equivalent to the test given by

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0 \qquad H_1 : \tau_i \neq 0 \text{ for at least one } i.$$

- If H_0 is true, then each observation is sampled from a normal distribution $N(\mu, \sigma^2)$.
- we want to analyze the situation via the ANOVA identity $SS_T = SS_{treatments} + SS_E$

- we introduce some notation:

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij} \quad , \quad \bar{y}_{i\cdot} = y_{i\cdot}/n \qquad i = 1, \dots, a$$

$$y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/(na)$$

(i.e. "$\cdot$" means sum over that variable).

- the total variability in the data is given by

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2$$

- Then

$$\underbrace{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2}_{SS_T} = \underbrace{n \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}_{SS_{treatments}} + \underbrace{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2}_{SS_E}$$

$an - 1$ degrees of freedom — $a - 1$ degrees of freedom — $a(n-1)$ degrees of freedom.

- we define the <u>mean square for treatments</u> as

$$MS_{Treatments} = \frac{SS_{treatments}}{a-1}$$

- we define the <u>mean square for error</u> as

$$MS_E = \frac{SS_E}{a(n-1)}.$$

- we can show that

$$E(MS_{treatments}) = \sigma^2 + \frac{n \sum_{i=1}^{a} \tau_i^2}{a-1}$$

and

$$E(MS_E) = \sigma^2.$$

- MS_E and $MS_{treatments}$ are independent.

- we choose a test statistic:

$$F_0 = \frac{MS_{treatments}}{MS_E}$$

- If H_0 is true, then F_0 follows an F-distribution, $F_{a-1, a(n-1)}$.

- Note that if H_0 is true, then $E(MS_{treatments}) = \sigma^2$, so if H_1 is true, $E[MS_{treat}]/E[MS_E] \gneq 1$.

- thus, we want to reject H_0 if we find that F_0 is too large (i.e. a one sided, upper tailed test).

- For a given significance level α, we reject H_0 iff

$$f_0 > f_{\alpha, a-1, n(a-1)}$$

- Note that computations for this test are often summarized in an ANOVA table:

Source of variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{Treatments}$	$a-1$	$MS_{Treat} = SS_{Treat}/a-1$	MS_{Treat}/MS_E
Error	SS_E	$a(n-1)$	$MS_E = \frac{SS_E}{a(n-1)}$	
Total	SS_T	$an-1$		