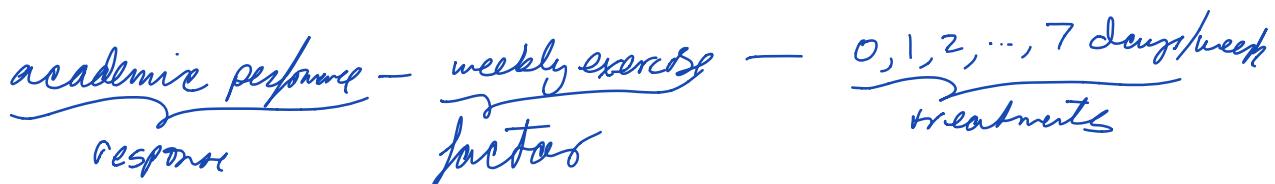


Lecture 33

Single Factor Experiments & ANOVA

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

Eg:



- Let $a = \#$ of treatments.
- For the i^{th} treatment, we get a random sample $\{Y_{i1}, Y_{i2}, \dots, Y_{in_i}\}$ where n_i is the size of the sample for the i^{th} treatment ($1 \leq i \leq a$). (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

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

- ↳ μ is a common mean across all treatments
- ↳ γ_i is a parameter associated to the i^{th} treatment called the i^{th} treatment 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 $\mu_i = \mu + \gamma_i$, 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 fixed-effects model)

- Goal: develop ANOVA for fixed-effects models.

- we assume that $\sum_{i=1}^a \gamma_i = 0$.
- we want to test the hypothesis $\mu_1 = \mu_2 = \dots = \mu_a$.
- since $\mu_i = \mu + \gamma_i$, this is equivalent to the test given by

$$H_0: \gamma_1 = \gamma_2 = \dots = \gamma_a = 0 \quad H_1: \gamma_i \neq 0 \text{ for at least one } i.$$

- 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 $SS_T = SS_{\text{treatments}} + SS_E$

- we introduce some notation:

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

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

(i.e. ".·" 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

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

\uparrow \uparrow \uparrow
 $an - 1$ $a - 1$ $a(n - 1)$
 degrees of freedom degrees of freedom degrees of freedom

- we define the mean square for treatments as

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

- we define the mean square for error as

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

- we can show that

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

and

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

- MS_E and $MS_{\text{treatments}}$ are independent.
- we choose a test statistic:

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

- If H_0 is true, then F_0 follows an F-distribution, $F_{a-1, n(a-1)}$.
- Note that if H_0 is true, then $E(MS_{\text{treatments}}) = \sigma^2$, so if H_1 is true, $E[MS_{\text{treat}}]/E[MS_E] \geq 1$.
- 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	$SS_{\text{Treatments}}$	$a - 1$	$MS_{\text{Treat}} = \frac{SS_{\text{Treat}}}{a-1}$	MS_{Treat} / MS_E
Error	SS_E	$a(n-1)$	$MS_E = \frac{SS_E}{a(n-1)}$	Y Y Y Y
Total	SS_T	$an - 1$	Y Y Y Y	Y Y Y Y