STAT 403 Tutorial · Week 6

For this week's tutorial, I'm going to give a numerical example for the randomized block design. Then I'll take a review of what we've learned so far.

<u>Example.</u> Four different washing solutions are being compared to study their effectiveness in retarding bacteria growth. Suppose that only four trials can be run on a single day and observations (amount of bacteria after applying each washing solution) are taken for three days, as shown below.

	solution 1	solution 2	solution 3	solution 4	average
day 1	33 (Y_{11})	64 (Y_{21})	32 (Y_{31})	64 (Y_{41})	48.25 ($\bar{Y}_{\cdot 1}$)
day 2	15 (Y_{12})	26 (Y_{22})	17 (Y_{32})	25 (Y_{42})	20.75 ($\bar{Y}_{\cdot 2}$)
day 3	23 (Y_{13})	38 (Y_{23})	16 (Y_{33})	37 (Y_{43})	28.50 ($\bar{Y}_{\cdot 3}$)
average	23.67 ($\bar{Y}_{1\cdot}$)	42.67 ($\bar{Y}_{2\cdot}$)	21.67 ($\bar{Y}_{3\cdot}$)	42 ($\bar{Y}_{4\cdot}$)	32.5 ($\bar{Y}_{\cdot\cdot}$)

(i) Suppose that each day is a block. Assess whether four solutions are the same.

(ii) Answer the same question as in (i) if days are not blocks.

Solution:

(i). Denote the observation of solution i on day j by Y_{ij} $\left(\begin{smallmatrix} i=1,2,3,4 \\ j=1,2,3 \end{smallmatrix}\right)$.

Then we have model $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} N(0,\sigma^2)$.

where α_i's represent treatment effects and β_j's represent block effects.

The test of hypothesis $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ v.s. $H_a: \alpha_i$'s not same

is based on

$$F \triangleq \frac{SS_{trt}/(t-1)}{SS_{err}/(N-b-t+1)} \overset{H_0}{\sim} F_{t-1,\, N-b-t+1}$$

Note here $N = bt = 3 \times 4 = 12$, $b = 3$, $t = 4$.

$$SS_{trt} = 3\left[(\bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot})^2 + (\bar{Y}_{2\cdot} - \bar{Y}_{\cdot\cdot})^2 + (\bar{Y}_{3\cdot} - \bar{Y}_{\cdot\cdot})^2 + (\bar{Y}_{4\cdot} - \bar{Y}_{\cdot\cdot})^2\right] = 1167$$

$$SS_{blk} = 4\left[(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot\cdot})^2 + (\bar{Y}_{\cdot 2} - \bar{Y}_{\cdot\cdot})^2 + (\bar{Y}_{\cdot 3} - \bar{Y}_{\cdot\cdot})^2\right] = 1608.5$$

$$SS_{total} = \sum_{i=1}^{4}\sum_{j=1}^{3}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = 3043$$

$$SS_{err} = SS_{total} - SS_{trt} - SS_{blk} = \sum_{i=1}^{4}\sum_{j=1}^{3}(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 = 267.5$$

$$\Rightarrow \quad F_{obs} = \frac{1167/3}{267.5/6} = 8.73 .$$

$$\text{p-value} = P(F \geq F_{obs}) = P(F \geq 8.73) = 0.013 < 0.05$$

Therefore we reject the null hypothesis at 5% level that the 4 solutions have same effects.

ANOVA table:

source	SS	df
treatment	1167	$t-1=3$
block	1608.5	$b-1=2$
error	267.5	$(b-1)(t-1)=6$
total	3043	$N-1=11$

(ii). If days were not used as block factors, then the model becomes $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, which is essentially the model for complete randomized design. Compare the calculations we'll see

Recall that the testing statistic becomes $F' = \frac{SS_{trt}/(t-1)}{SS_{err}/(N-t)} \sim F_{t-1, N-t}$

$$SS'_{trt} = SS_{trt} = 1167$$

$$SS'_{tot} = SS_{tot} = 3043$$

$$SS'_{err} = SS'_{tot} - SS'_{trt} = 1876$$

$$\Rightarrow F'_{obs} = \frac{1167/4-1}{1876/12-4} = 1.66$$

$$\text{p-value} = P(F' \geq F'_{obs}) = 0.252 > 0.05$$

$\Rightarrow$ do not reject at 5% level

ANOVA table:

Source	SS	df
treatment	1167	3
error	1876	8
total	3043	11

Review of STAT 403 from WEEK 1 to WEEK 5.

- One-sample problem : $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ $\begin{cases} H_0: \mu = 0 \quad \text{v.s.} \quad H_a: \mu \neq 0 . \ (\mu \gtrless 0) \\ \text{C.I. for } \mu \end{cases}$

- Two-sample problem :

  - non-paired . $X_1, \ldots, X_n \sim N(\mu_1, \sigma^2)$
    $Y_1, \ldots, Y_m \sim N(\mu_2, \sigma^2)$

    ① $H_0: \mu_1 = \mu_2$ v.s. $H_a: \mu_1 \gtreqless \mu_2$. $\begin{cases} \text{t test} \\ \text{randomization test.} \end{cases}$

    ② C.I. for $\mu_1 - \mu_2$

  - paired version : $X_1 - Y_1, \ldots, X_n - Y_n \sim N(\mu_1 - \mu_2, \sigma^2)$

    ① $H_0: \mu_1 = \mu_2$ v.s. $H_a: \mu_1 \gtreqless \mu_2$. $\begin{cases} \text{paired t test} \\ \text{paired randomization test} \end{cases}$

    ② C.I. for $\mu_1 - \mu_2$.

- Multiple-sample (treatments) problem.

  - complete randomized designs. $Y_{ij} = \mu_i + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ $\begin{pmatrix} i = 1, \ldots, t \\ j = 1, \ldots, n_i \end{pmatrix}$

    ① Assess $\mu_1 = \mu_2 = \cdots \mu_t$: F-test & ANOVA table.

    ② Contrasts & their inferences $\theta = c_1\mu_1 + \cdots + c_t\mu_t$ $(c_1 + \cdots + c_t = 0)$.

  - randomized block designs. $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ $\begin{pmatrix} i = 1, \ldots, t \\ j = 1, \ldots, b \end{pmatrix}$.

    ① F-test & ANOVA table.