Section 6.2 (Cont.)

The Random-Effects Model

Fixed-Effects and Random Effects

- If treatments are specifically chosen and tests of hypotheses apply only to those factor levels. Those factors are called *fixed effects*.
- R*andom effects* is used to denote factors in an ANOVA design with levels that were not deliberately arranged by the experimenter but which were sampled from a population of possible samples instead.
- A simple criterion for deciding whether or not an effect in an experiment is random or fixed is to ask how one would select (or arrange) the levels for the factor in a replication of the study.

Examples

Ex 1. Personnel management in a large enterprise.

- Question: Does the interviewer have an effect on the rating of job candidates?
- Data: 5 interviewers selected at random, each interviews 4 candidates selected at random.
- The interviewers are random draws from a larger population. We are interested in the larger population and not these 5 specific interviewers.

Ex 2.

- Question: How does the sodium in beer differ between brands?
- Data: 6 randomly chosen brand, 8 bottles tested per brand

Random-Effects Model

- Fixed-effects model to compare treatments

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1,\ldots,k;\ j = 1,\ldots,n_i.\ \varepsilon_{ij} \sim^{iid} N\left(0,\sigma^2\right)$$

$$\text{Hypothesis}: H_0 : \tau_i = 0, \forall i \ \text{ vs } H_1 : \text{at least one } \tau_i \neq 0.\ \sum_{i=1}^{k} n_i \tau_i = 0.$$

- Random-effects model (variance component model)

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \varepsilon_{ij} \sim^{iid} N\left(0,\sigma^2\right), \tau_i \sim^{iid} N\left(0,\sigma_\tau^2\right)$$

where $\left\{\varepsilon_{ij}, i = 1,\ldots k;\ j = 1,\ldots n_i\right\}$ is independent of $\left\{\tau_i, i = 1,\ldots k\right\}$

$$\text{Hypothesis}: H_0 : \sigma_\tau^2 = 0, \text{ vs } H_1 : \sigma_\tau^2 \neq 0.$$

Important Features of Random-Effects Model

(1). $E(Y_{ij}) = \mu, \quad Var(Y_{ij}) = \sigma_\tau^2 + \sigma^2$

(2). $Y_{ij} \sim N(\mu, \sigma_\tau^2 + \sigma^2)$, but not i.i.d. for any (i, j)

$$Cov(Y_{ij}, Y_{i'j'}) = \begin{cases} 0, & \text{if } i \neq i'; \\ \sigma_\tau^2, & \text{if } i = i', \mathrm{j} \neq j'; \\ \sigma_\tau^2 + \sigma^2, & \text{if } i = i', \mathrm{j} = j'. \end{cases}$$

Coefficient of correlation between responses from same factor level

$$\rho(Y_{ij}, Y_{ij'}) = \frac{\sigma_\tau^2}{\sigma_Y^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma^2}$$

(3). $\sigma_\tau^2 = 0 \Leftrightarrow Var(\tau_i) = 0, \forall i = 1, \ldots, k$

$\Leftrightarrow \quad \tau_1 = \ldots = \tau_k = 0$

Sum of squares and mean squares

$$SSTO = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}\right)^2, SSTR = \sum_{i=1}^{k} n_i\left(\bar{Y}_i-\bar{Y}\right)^2,$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_i\right)^2 = \sum_{i=1}^{k}(n_i-1)s_i^2,$$

$$\text{Under } \mathrm{H}_0 : \sigma_\tau^2 = 0,\ \frac{SSTR}{\sigma^2} \sim \chi^2(k-1) \perp \frac{SSE}{\sigma^2} \sim \chi^2(N-k)$$

$$\text{For balanced design } n_1 = \ldots = \mathrm{n_k} = n:$$

$$E(SSTR) = n\cdot\sum_{i=1}^{k} E\left(\bar{Y}_i-\bar{Y}\right)^2 = n\cdot E\left(\sum_{i=1}^{k}(\tau_i-\bar{\tau})^2 + \sum_{i=1}^{k}(\bar{\varepsilon}_i-\bar{\varepsilon})^2\right)$$

$$= n\cdot(k-1)\sigma_\tau^2 + (k-1)\sigma^2$$

$$E(MSTR) = E(SSTR/(k-1)) = \sigma^2 + n\cdot\sigma_\tau^2$$

ANOVA for Random Effects

- Analysis of Variance Table (ANOVA)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Treatments	SSTR	k - 1	$MSTR = \dfrac{SSTR}{k-1}$	$F = \dfrac{MSTR}{MSE}$
Error	SSE	N - k	$MSE = \dfrac{SSE}{N-k}$	
Total	SST	N - 1		

- Decision rule: Reject H_0: $\sigma_\tau^2=0$ if p-value < α or $F > F_\alpha(k-1, N-k)$.

One - Way ANOVA :

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1,\ldots,k;\ j = 1,\ldots,n_i$$

(1). Fixed effects model: $Y_{ij} \sim N\left(\mu + \tau_i, \sigma^2\right)$,

$\mathrm{H_0} : \tau_i = 0, \mathrm{i} = 1,\ldots,\mathrm{k}$, vs. $\mathrm{H_1}$: at least one $\tau_i \neq 0$

$$E(MSE) = \sigma^2, E(MSTR) = \sigma^2 + \frac{1}{k-1}\sum_{i=1}^{k} n_i \tau_i^{\,2}$$

$$\hat{\mu}_i = \overline{Y}_i, \hat{\sigma}^2 = MSE$$

(2). Random effects model: $Y_{ij} \sim N\left(\mu, \sigma^2 + \sigma_\tau^{\,2}\right)$

$\mathrm{H_0} : \sigma_\tau^{\,2} = 0$ vs. $\mathrm{H_1}$: $\sigma_\tau^{\,2} > 0$

$$E(MSE) = \sigma^2, E(MSTR) = \sigma^2 + \tilde{n}\,\sigma_\tau^{\,2}, \tilde{n} = \frac{1}{k-1}\left(N - \frac{\sum_{i=1}^{k} n_i^{\,2}}{\sum_{i=1}^{k} n_i}\right)$$

$$\hat{\sigma}_\tau^2 = \frac{1}{\tilde{n}}\left(MSTR - MSE\right), \hat{\sigma}^2 = MSE$$

Estimation of Population Mean μ

$\forall$(i, j), we have $E(Y_{ij}) = \mu$, since $Y_{ij} \sim N(\mu, \sigma_\tau^2 + \sigma^2)$

Estimator for population mean :

$$\hat{\mu} = \overline{Y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}, \quad E(\hat{\mu}) = \mu$$

$$Var(\hat{\mu}) = \frac{1}{N^2} Var\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}\right) = \frac{1}{N^2}\sum_{i=1}^{k}\left\{\sum_{j=1}^{n_i} Var(Y_{ij}) + \sum_{j \neq j'} Cov(Y_{ij}, Y_{ij'})\right\}$$

$$= \frac{1}{N^2}\sum_{i=1}^{k}\left\{n_i\left(\sigma_\tau^2 + \sigma^2\right) + n_i(n_i - 1)\sigma_\tau^2\right\}$$

For balanced design : $Var(\hat{\mu}) = \dfrac{\sigma^2}{N} + \dfrac{\sigma_\tau^2}{k} = \dfrac{\sigma^2 + n\sigma_\tau^2}{N}$

Standard Error of $\hat{\mu}$: $\quad s(\hat{\mu}) = \sqrt{\dfrac{MSTR}{N}}$ $\quad$ [Note : $E(MSTR) = \sigma^2 + n\sigma_\tau^2$]

$100(1 - \alpha)\%$ Confidence Interval for mean μ : $\overline{Y} \pm t_{\frac{\alpha}{2}}(k-1) \cdot \sqrt{\dfrac{MSTR}{N}}$

Example

- The following table contains the results of the study by Apex Enterprises on the evaluation ratings (scale: 0-100) of potential employees by its personnel officers. Five officers were selected at random. Twenty job applicants were randomly assigned to the five officers, with four per personnel officer.
- Assume that the variability within each five rating distributions is approximately the same, estimate the variability and test if it differs significantly from 0.

Officer	Candiate (j)					
(i)	1	2	3	4	Mean	Si
A	76	65	85	74	75	8.2
B	59	75	81	67	70.5	9.57
C	49	63	61	46	54.75	8.5
D	74	71	85	89	79.75	8.62
E	66	84	80	79	77.25	7.8
Total					71.45	

Source	SS	DF	MS
Treatment	1579.7	4	$MSTR = 394.9$
Error	1099.3	15	$MSE = 73.3$
Total	2678.9	19	

Random-effects Model: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1,\ldots,5;\ j = 1,\ldots,4$

where factor levels $k = 5$, replications at each level $n = 4$.

Hypotheses: $H_0 : \sigma_\tau^2 = 0 \quad vs \quad H_1 : \sigma_\tau^2 > 0$

$$F = \frac{MSTR}{MSE} = \frac{394.9}{73.3} = 5.39 > F_{0.05}(4,15) = 3.06$$

or p-value $= P\{F(4,15) > 5.39\} \cong 0.007 < 0.05.$

Reject H_0, i.e. the rating variability of the officers differ from 0.

$$\hat{\sigma}_\tau^2 = \frac{MSTR - MSE}{n} = \frac{394.9 - 73.3}{4} = 80.4.$$

Model Checking for One-factor ANOVA

- Normality Check

  SAS code (proc univariate normal)

  Residual plots: (1) overall residual plot

  (2). Residual against group averages

- Equal Variances – Levene's Test

  Levene's Test is not very sensitive to the normality

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 \text{ vs } H_1 : \text{not all } \sigma_i^2 \text{ are the same}$$

- Welch's F –test (scaled F-test)