Handout 3

Often we have data with two factors but there is only one observation for each treatment combination. If factor A has a levels and factor B has b levels then there are a total of ab observations. Let us denote the observations as $\{Y_{ij} : j = 1, ..., b, i = 1, ..., a\}$. If we are able to ignore the interaction, then the model is

$$Y_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + \varepsilon_{ij}, j = 1, ..., b, i = 1, ..., a,$$

where $\sum \alpha_i = 0$, $\sum \beta_j = 0$, and ε_{ij}'s are independent $N(0, \sigma^2)$. Recall that such a model without the interactions term is also called an additive model. The analysis here is basically the same as in chapter 19 for additive models with $n = 1$.

Estimation of the parameters
Let $\overline{Y}_{\cdot\cdot} = \sum\sum Y_{ij}/(ab)$, $\overline{Y}_{i\cdot} = \sum Y_{ij}/b$, $\overline{Y}_{\cdot j} = \sum Y_{ij}/a$.
Estimate of
(i) $\mu_{\cdot\cdot}$ is $\hat{\mu}_{\cdot\cdot} = \overline{Y}_{\cdot\cdot}$, (ii) α_i is $\hat{\alpha}_i = \overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}$, (iii) β_j is $\hat{\beta}_j = \overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}$.
Note that $\sum \hat{\alpha}_i = 0$ and $\sum \hat{\beta}_j = 0$.

Decomposition of total sum of squares:
Note that the fitted values and residuals are

$$\begin{aligned}
\hat{Y}_{ij} &= \hat{\mu}_{\cdot\cdot} + \hat{\alpha}_i + \hat{\beta}_j = \overline{Y}_{i\cdot} + \overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}, \\
e_{ij} &= Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu}_{\cdot\cdot} + \hat{\alpha}_i + \hat{\beta}_j) = Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot}.
\end{aligned}$$

The total sum of squares, sums of squares due to factor A, due to factor B, and the residual sum of squares are given below

$$\begin{aligned}
SSTO &= \sum\sum (Y_{ij} - \overline{Y}_{\cdot\cdot})^2, \ df(SSTO) = ab - 1, \\
SSA &= \sum\sum \hat{\alpha}_i^2 = b \sum \hat{\alpha}_i^2, \ \ df(SSA) = a - 1, \\
SSB &= \sum\sum \hat{\beta}_j^2 = a \sum \hat{\beta}_j^2, \ df(SSB) = b - 1, \\
SSE &= \sum\sum e_{ij}^2 = \sum\sum (Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})^2, df(SSE) = (a-1)(b-1).
\end{aligned}$$

The following identities hold

$$\begin{aligned}
SSTO &= SSA + SSB + SSE, \\
df(SSTO) &= df(SSA) + df(SSB) + df(SSE).
\end{aligned}$$

Mean squares:
Mean squares are defined as the sums of squares divided by their degrees of freedom.

$$\begin{aligned}
MSA &= SSA/df(SSA) = SSA/(a-1), \\
MSB &= SSB/df(SSB) = SSB/(b-1), \\
MSE &= SSE/df(SSE) = SSE/[(a-1)(b-1)].
\end{aligned}$$

The following mathematical result is useful for motivating subsequent F-tests.

Fact: (i) $E(MSE) = \sigma^2$, (ii) $E(MSA) = \sigma^2 + b\sum \alpha_i^2/(a-1)$, (iii) $E(MSB) = \sigma^2 + a\sum \beta_j^2/(b-1)$.
Note that MSE is an estimate of σ^2.
[**An important note: In your textbook,** SSE **and** MSE **are being called** $SSAB$ **and** $MSAB$]

Tests for factor effects:

Suppose that we want to test $H_0 : \alpha_i = 0$ for all i, vs. H_1 :not all a_i's are zero, at a level of significance α.
Let $F^* = MSA/MSE$. Reject H_0 if $F^* > F(1-\alpha; a-1, (a-1)(b-1))$.
Suppose that we want to test $H_0 : \beta_j = 0$ for all j, vs. H_1 :not all β_j's are zero, at a level of significance α.
Let $F^* = MSB/MSE$. Reject H_0 if $F^* > F(1-\alpha; b-1, (a-1)(b-1))$.

Estimation of factor effects:

The following table lists estimates of various parameters.

parameter	estimate	$\sigma^2(estimate)$	$s^2(estimate)$
$\mu_{i\cdot} = \mu_{\cdot\cdot} + \alpha_i$	$\hat{\mu}_{i\cdot} = \bar{Y}_{i\cdot}$	$\sigma^2(\hat{\mu}_{i\cdot}) = \frac{\sigma^2}{b}$	$s^2(\hat{\mu}_{i\cdot}) = \frac{MSE}{b}$
α_i	$\hat{\alpha}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$	$\sigma^2(\hat{\alpha}_i) = \frac{a-1}{ab}\sigma^2$	$s^2(\hat{\alpha}_i) = \frac{a-1}{ab}MSE$
$\mu_{\cdot j}$	$\hat{\mu}_{\cdot j} = \bar{Y}_{\cdot j}$	$\sigma^2(\hat{\mu}_{\cdot j}) = \frac{\sigma^2}{a}$	$s^2(\hat{\mu}_{\cdot j}) = \frac{MSE}{a}$
β_j	$\hat{\beta}_j = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}$	$\sigma^2(\hat{\beta}_j) = \frac{b-1}{ab}\sigma^2$	$s^2(\hat{\beta}_j) = \frac{b-1}{ab}MSE$
$D = \mu_{i\cdot} - \mu_{i'\cdot}$	$\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$	$\sigma^2(\hat{D}) = \frac{2}{b}\sigma^2$	$\sigma^2(\hat{D}) = \frac{2}{b}MSE$
$D = \mu_{\cdot j} - \mu_{\cdot j'}$	$\hat{D} = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot j'}$	$\sigma^2(\hat{D}) = \frac{2}{a}\sigma^2$	$s^2(\hat{D}) = \frac{2}{a}MSE$
$L = \sum c_i \mu_{i\cdot}$	$\hat{L} = \sum c_i \bar{Y}_{i\cdot}$	$\sigma^2(\hat{L}) = \frac{\sum c_i^2}{b}\sigma^2$	$s^2(\hat{L}) = \frac{\sum c_i^2}{b}MSE$
$L = \sum d_j \mu_{\cdot j}$	$\hat{L} = \sum d_j \bar{Y}_{\cdot j}$	$\sigma^2(\hat{L}) = \frac{\sum d_j^2}{a}\sigma^2$	$s^2(\hat{L}) = \frac{\sum d_j^2}{a}MSE$

Tukey' test for additivity.

Tukey considered a model with interaction of the form:

$$Y_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + D\alpha_i\beta_j + \varepsilon_{ij},$$

where $\sum \alpha_i = 0$ and $\sum \beta_j = 0$. Estimates of $\mu_{\cdot\cdot}$, α_i and β_j are same as before. Estimate of D is given by

$$\hat{D} = \sum\sum Y_{ij}\hat{\alpha}_i\hat{\beta}_j \Big/ \left\{ \left(\sum \hat{\alpha}_i^2\right) \left(\sum \hat{\beta}_j^2\right) \right\}.$$

We want to test $H_0 : D = 0$ vs. $H_1 : D \neq 0$.
The new interaction sum of squares is given by

$$\begin{aligned} SSAB^* &= \sum\sum (\hat{D}\hat{\alpha}_i\hat{\beta}_j)^2 = \hat{D}^2 \sum \hat{\alpha}_i^2 \sum \hat{\beta}_j^2 \\ &= (\sum\sum Y_{ij}\hat{\alpha}_i\hat{\beta}_j)^2 \Big/ \left\{ \left(\sum \hat{\alpha}_i^2\right) \left(\sum \hat{\beta}_j^2\right) \right\}, \quad df(SSAB^*) = 1. \end{aligned}$$

The new residual sum of squares is given by

$$\begin{aligned} SSE_{rem} &= SSTO - SSA - SSB - SSAB^*, \\ df(SSE_{rem}) &= df(SSTO) - df(SSA) - df(SSB) - df(SSAB^*) \\ &= (ab-1) - (a-1) - (b-1) - 1 = ab - a - b. \end{aligned}$$

where $SSTO$, SSA and SSB are same as before.

So,

$$MSAB^* = SSAB^*/df(SSAB^*) = SSAB^*,$$
$$MSE_{rem} = SSE_{rem}/df(SSE_{rem}) = SSE_{rem}/(ab - a - b).$$

Let $F^* = MSAB^*/MSE_{rem}$. Reject H_0 if $F^* > F(1-\alpha; 1, ab-a-b)$ when the level of significance is α.

Warping of copper plates

An experiment was conducted to investigate warping of copper plates. The factors studied were temperature (factor A) and copper contents (factor B) of the plates. The response variable was a measure of the amount of warping.

	Copper content				
	low	moderate	medium	high	
Temperature	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$\bar{Y}_{i.}$
low ($i = 1$)	17	16	24	28	21.25
moderate ($i = 2$)	12	18	17	27	18.50
medium ($i = 3$)	16	18	25	30	22.25
high ($i = 4$)	21	23	23	29	24.00
$\bar{Y}_{.j}$	16.50	18.75	22.25	28.50	$\bar{Y}_{..} = 21.50$

ANOVA table

Source	df	SS	MS	F	p-value
Temperature	$a - 1 = 3$	63.50	21.167	3.97	0.047
Copper	$b - 1 = 3$	328.50	109.500	20.53	0.000
Error	$(a-1)(b-1) = 9$	48.00	5.333		
Total	$ab - 1 = 15$	440.00			

The following table summarizes the parameter estimates

$\hat{\mu}_{i.}$	$s(\hat{\mu}_{i.})$	$\hat{\mu}_{.j}$	$s(\hat{\mu}_{.j})$	$\hat{\alpha}_i$	$s(\hat{\alpha}_i)$	$\hat{\beta}_j$	$s(\hat{\beta}_j)$
$\bar{Y}_{1.} = 21.25$	1.155	$\bar{Y}_{.1} = 16.50$	1.155	$\hat{\alpha}_1 = -.25$	1.000	$\hat{\beta}_1 = -5.00$	1.000
$\bar{Y}_{2.} = 18.50$	1.155	$\bar{Y}_{.2} = 18.75$	1.155	$\hat{\alpha}_2 = -3.00$	1.000	$\hat{\beta}_2 = -2.75$	1.000
$\bar{Y}_{3.} = 22.25$	1.155	$\bar{Y}_{.3} = 22.25$	1.155	$\hat{\alpha}_3 = .75$	1.000	$\hat{\beta}_3 = .75$	1.000
$\bar{Y}_{4.} = 24.00$	1.155	$\bar{Y}_{.1} = 28.50$	1.155	$\hat{\alpha}_4 = 2.50$	1.000	$\hat{\beta}_4 = 7.00$	1.000

Tukey's test for additivity: Using Tukey's model, we want to test if the interaction effect exist, i.e., we want to test $H_0 : D = 0$ against $H_1 : D \neq 0$ at level $\alpha = .05$. Here

$$\hat{D} = \frac{\sum\sum \hat{\alpha}_i \hat{\beta}_j Y_{ij}}{\sum \hat{\alpha}_i^2 \sum \hat{\beta}_j^2} = \frac{-93.5625}{(15.8750)(82.1250)} = -0.0718,$$
$$SSAB^* = \hat{D}^2 \sum \hat{\alpha}_i^2 \sum \hat{\beta}_j^2 = 6.7145,$$
$$SS\,\mathrm{Re}\,m = SSTO - SSA - SSB - SSAB^*$$
$$= 440 - 63.50 - 328.50 - 6.7145 = 41.2855,$$
$$F^* = \frac{\{SSAB^*/1\}}{\{SS\,\mathrm{Re}\,m/(ab-a-b)\}} = \frac{\{6.7145/1\}}{\{41.2855/(16-4-4)\}} = 1.3010.$$

The degrees of freedom for the F-test are $(1, ab-a-b) = (1, 8)$. The p-value here is 0.287. Clearly, the null hypothesis cannot be rejected. So we conclude that an additive model is reasonable for this data.

Figure 1: Copper Plate Data