# Review of Weeks 5-8 (through lecture 21)



## Standard designs

- Completely randomized design
- Randomized blocks design
- Balanced factorial experiments
- Latin square designs
- Model formulas: e.g. RBD

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

- Interaction plots



## Standard designs

- Structure of ANOVA tables, degrees of freedom
- Standard error formula for pairwise comparisons;
  e.g. RBD with $n$ replicates of each treatment in each block

$$\text{var}(\bar{y}_{j\cdot} - \bar{y}_{k\cdot}) = \sigma^2 \frac{2}{bn}$$

- t-tests, confidence intervals, and LSD
- Pairwise comparisons when there is interaction versus main effects



## Linear contrasts

$$L = \sum_{i=1}^{t} c_i \mu_i \qquad \sum_{i=1}^{t} c_i = 0$$

- Variance formula:

$$\hat{L} = \sum_{i=1}^{t} c_i \bar{y}_i \qquad \text{var}\left(\hat{L}\right) = \sigma^2 \sum_{i=1}^{t} \frac{c_i^2}{n_i}$$

- Interaction and main effects hypotheses
  ($2 \times 2$ factorial experiment)
- Unbalanced data, partial versus sequential SS
- Linear and non-linear effects (of numerical factors)

## Randomized Complete Blocks Design

- Design for comparing $t$ treatments in $b$ blocks
- Statistical model (with no replication)

$$E(y_{ij}) = \mu_{ij} = \text{expected response to treatment } i \text{ in block } j$$

$$y_{ij} = \mu_{ij} + \epsilon_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

$$\begin{aligned}
\mu &= \text{overall, or reference mean} \\
\alpha_i &= \text{effect of treatment } i \\
\beta_j &= \text{effect of block } j \\
\epsilon_{ij} &= \text{random error associated with } (i,j)\text{th response}
\end{aligned}$$

- The standard assumption is that the errors are indep. $N(0, \sigma^2)$



## Randomized Complete Blocks Design

- Effect of changing the treatment within a block is constant across blocks

$$\mu_{ij} - \mu_{kj} = (\mu + \alpha_i + \beta_j) - (\mu + \alpha_k + \beta_j) = \alpha_i - \alpha_k$$

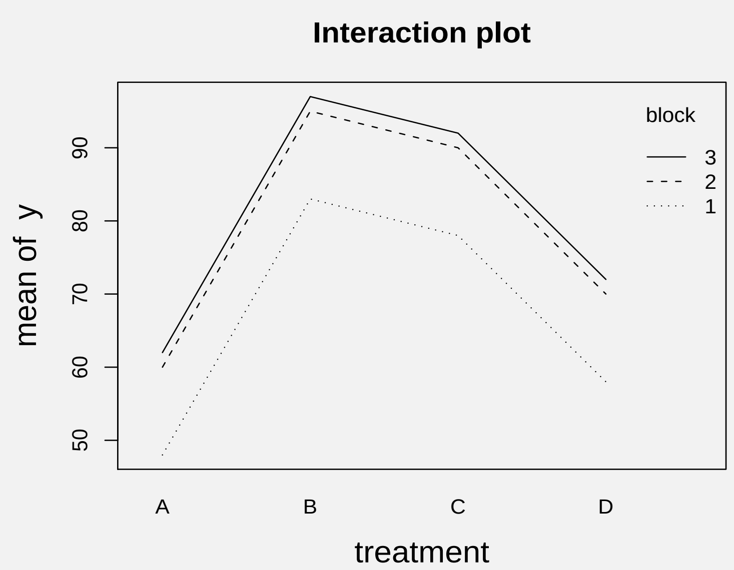
- The effect of changing blocks is the same for all treatments

$$\mu_{ij} - \mu_{ik} = (\mu + \alpha_i + \beta_j) - (\mu + \alpha_i + \beta_k) = \beta_j - \beta_k$$

- The model implies no interaction!
- Blocks have "parallel" mean profiles



## Randomized Complete Blocks Design






## Randomized Complete Blocks Design

- Primary hypothesis concerns equality of treatment means

$$H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \cdots = \mu_{t\cdot}$$

or equivalently, the treatment effects are all zero

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$$

- ANOVA decomposition

$$\begin{aligned}
\text{Total SS} &= \text{Treatment SS} + \text{Block SS} + \text{Error SS} \\
\sum_{ij} (y_{ij} - \bar{y}_{\cdot\cdot})^2 &= b \sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + t \sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 \\
&\quad + \sum_{ij} (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2
\end{aligned}$$

# Randomized Complete Blocks Design

- ANOVA Table

| Source | SS | DF | MS | F |
|---|---|---|---|---|
| Treatments | SST | $t-1$ | MST | MST/MSE |
| Blocks | SSB | $b-1$ | MSB | MSB/MSE |
| Error | SSE | $(b-1)(t-1)$ | MSE | |
| Total | TSS | $bt-1$ | | |

- Note: The error df is

$$(bt-1)-(t-1)-(b-1)=(b-1)(t-1)$$



# Efficiency of RCB versus CR

- Blocking eliminates sources of variability in the response, making it easier to isolate treatment differences
- What is the efficiency gain due to blocking?
- In a CR experiment, with $n$ replicates per treatment

$$\text{var}(\bar{y}_{i\cdot}-\bar{y}_{k\cdot})=2\frac{\sigma^2_{CR}}{n}$$

- In a RCB experiment, with $b$ blocks

$$\text{var}(\bar{y}_{i\cdot}-\bar{y}_{k\cdot})=2\frac{\sigma^2_{RCB}}{b}$$



# Efficiency of RCB versus CR

- The two designs estimate the difference between treatments $i$ and $k$ with equal efficiency if

$$2\frac{\sigma^2_{CR}}{n}=2\frac{\sigma^2_{RCB}}{b} \qquad \text{or} \qquad \frac{\sigma^2_{CR}}{\sigma^2_{RCB}}=\frac{n}{b}$$

- The ratio $n/b$ measures the *efficiency* of the RCB design relative to a CR design.
- A rough estimate of the relative efficiency is MSE(CR)/MSE(RCB), where MSE(RCB) is the MSE from the RCB experiment, and MSE(CR) is the MSE if the blocking factor is ignored. (A better estimate is given in the notes.)



# Factorial Experiments (with replication)

- Statistical Model

$$y_{ijk}=\mu+\alpha_i+\beta_j+\alpha\beta_{ij}+\epsilon_{ijk}$$

$$\begin{aligned}
y_{ijk} &= k\text{th response at combination } (i,j)\\
\mu &= \text{overall mean or reference level}\\
\alpha_i &= \text{main effect of } i\text{th level of factor A, } i=1,\dots,a\\
\beta_j &= \text{main effect of } j\text{th level of factor B, } j=1,\dots,b\\
\alpha\beta_{ij} &= \text{interaction effect for combination } (i,j)\\
\epsilon_{ijk} &= \text{random error associated with } k\text{th response at levels } (i,j)
\end{aligned}$$

- Standard assumption is the $\epsilon_{ijk}$'s are a random sample from $N(0,\sigma^2)$

## Factorial Experiments

- The expected response at treatment combination $(i, j)$ is

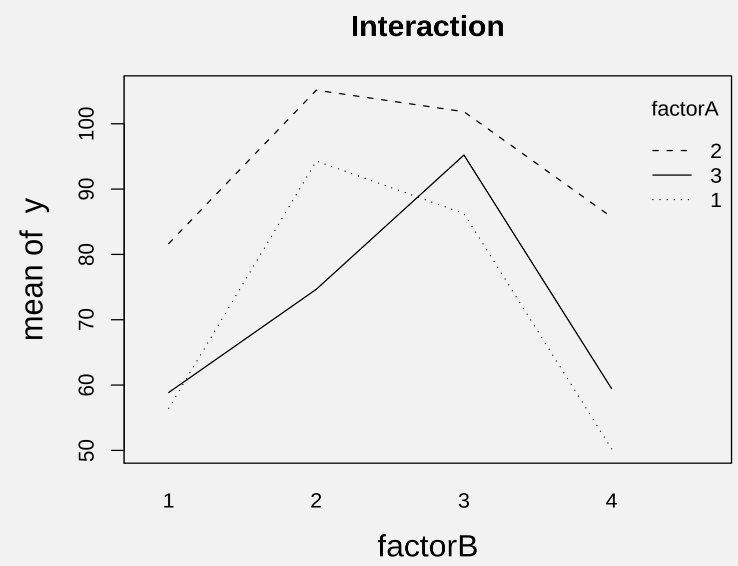
$$\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

- The difference in expected response between two levels of A (say $i$ and $i'$), controlling for B, is

$$\begin{aligned} \mu_{ij} - \mu_{i'j} &= (\alpha_i - \alpha_{i'}) + (\alpha\beta_{ij} - \alpha\beta_{i'j}) \\ &= \text{main effect} + \text{interaction} \end{aligned}$$

- The model allows the effect of changing the level of A to depend on the level of B; i.e. *interaction*
- If there is interaction, differences between levels of factor A must be examined separately at each level of factor B








## Factorial Experiments

- ANOVA table

| Source | SS | DF | MS | F |
|---|---|---|---|---|
| Factor A | SSA | $a-1$ | MSA | MSA/MSE |
| Factor B | SSB | $b-1$ | MSB | MSB/MSE |
| Interaction | SSI | $(a-1)(b-1)$ | MSI | MSI/MSE |
| Error | SSE | $ab(n-1)$ | MSE | |
| Total | TSS | $abn-1$ | | |

- The degrees of freedom add to the total

$$abn - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1)$$



## Factorial Experiments

- Test for interaction by comparing F=MSI/MSE to an F-distribution with $(a-1)(b-1)$ and $ab(n-1)$ degrees of freedom
- The MSE is a "pure" estimate of the error variance

$$\text{MSE} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} s_{ij}^2$$

where $s_{ij}^2$ is the sample variance among the replicate responses from treatment combination $(i, j)$.
- There are no degrees of freedom for error if $n = 1$!
- Ordering of factors in the model doesn't matter because the design is balanced.

## Linear contrasts among treatment means

- e.g. $2 \times 2$ factorial experiment

| | Factor B | | |
|---|---|---|---|
| Factor A | 1 | 2 | |
| 1 | $\mu_{11}$ | $\mu_{12}$ | $\bar{\mu}_{1\cdot}$ |
| 2 | $\mu_{21}$ | $\mu_{22}$ | $\bar{\mu}_{2\cdot}$ |
| | $\bar{\mu}_{\cdot 1}$ | $\bar{\mu}_{\cdot 2}$ | $\bar{\mu}_{\cdot\cdot}$ |

No interaction: $(\mu_{11} - \mu_{12}) - (\mu_{21} - \mu_{22}) = 0$

or $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$

No A main effect $(\mu_{11} + \mu_{12}) - (\mu_{21} + \mu_{22}) = 0$

or $\mu_{11} + \mu_{12} - \mu_{21} - \mu_{22} = 0$

No B main effect $(\mu_{11} + \mu_{21}) - (\mu_{12} + \mu_{22}) = 0$

or $\mu_{11} - \mu_{12} + \mu_{21} - \mu_{22} = 0$



## Linear contrasts among treatment means

- Estimate a linear contrast by replacing expected values by corresponding sample means

$$\hat{L} = \sum_{i=1}^{t} c_i \bar{y}_{i\cdot}$$

- The variance is given by

$$\text{var}(\hat{L}) = \sum_{i=1}^{t} c_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum_{i=1}^{t} \frac{c_i^2}{n_i}$$

- Recall: if $\text{var}(Y) = \sigma^2$, then $\text{var}(cY) = c^2\sigma^2$



## Linear contrasts among treatment means

- Test $L = 0$ using a t-statistic

$$t = \frac{\hat{L}}{\text{se}(\hat{L})} = \frac{\sum_{i=1}^{t} c_i \bar{y}_{i\cdot}}{\hat{\sigma}\sqrt{\sum_{i=1}^{t} c_i^2/n_i}}$$

- If $L = 0$, $t \sim t(\text{error df})$. Equivalently, $t^2 = F \sim F(1, \text{error df})$

$$t^2 = F = \frac{\hat{L}^2 / \left(\sum_{i=1}^{t} c_i^2/n_i\right)}{\hat{\sigma}^2}$$

- Numerator is contrast sum of squares
- Note that $\hat{\sigma}^2$ =MSE.



## Polynomial Effects

- Consider a designed experiment in which a continuous predictor, $X$, is set at $t$ *equally spaced values*, and $r$ replicate measurements are taken at each level.
- Let $x_i$ denote the $i$th treatment value.
- Let $\mu_i$ denote the expected response at the $i$th setting.
- If the mean response is linear in $X$, the slope is given by

$$\beta_1 = \frac{\sum_{i=1}^{t}(x_i - \bar{x})(\mu_i - \bar{\mu})}{\text{SSX}} = \frac{\sum_{i=1}^{t}(x_i - \bar{x})\mu_i}{\text{SSX}}$$

- Note: This is precisely the OLS estimate of the slope based on the pairs, $(x_1, \mu_1), \ldots, (x_t, \mu_t)$

## Polynomial Effects

- The slope of the linear relation between $\mu$ and $X$ is a contrast among the treatment means

$$\beta_1 \propto \sum_{i=1}^{t} c_i \mu_i \qquad \text{where} \qquad \sum_{i=1}^{t} c_i = 0$$

- We can test for a linear trend by testing that an appropriate contrast is zero. The coefficents should be equally spaced, and sum to zero.

```
 t    Contrast coefficients
---------------------------
 3    -1   0   1
 4    -3  -1   1   3
 5    -2  -1   0   1   2
---------------------------
```



## Analysis of Covariance

- Consider a completely randomized design involving a single factor, in which sampling units come with *covariate* values aready assigned.
- Treatment means need to be adjusted so they correspond to the same covariates values before comparisons can be made.
- Consider the model with no interaction:

$$y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \epsilon_{ij}$$



## Analysis of Covariance

- We can compare the treatments on the basis of their *adjusted means*, the predicted yields at the average value of the covariate(s).

$$\begin{aligned} \hat{\mu}_{i,adj} &= \hat{\beta}_0 + \hat{\alpha}_i + \hat{\beta}_1 \bar{x}_{\cdot\cdot} \\ &= \bar{y}_{i\cdot} - \hat{\beta}_1(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}) \end{aligned}$$

- These are sometimes referred to as *least squares means*.
- For pairwise comparison of means we use

$$\text{var}(\hat{\mu}_{i,adj} - \hat{\mu}_{k,adj}) = \sigma^2 \left( \frac{2}{n} + \frac{(\bar{x}_{i\cdot} - \bar{x}_{k\cdot})^2}{SXX} \right)$$