### Why choose a split plot design?

Rarely deliberately chosen.
- Usually done out of necessity because can't apply a treatment to the "small" size eu, e.g.:
  - too time consuming to change machinery frequently,
  - can't irrigate one row,
  - don't have enough water baths to use one per pot.
- Can be used to add treatments to ongoing study.
  - Park Grass Experiment, Rothamsted England. Effects of fertilization on hay yield.
    - Started in 1856; data collected annually since them.
    - Fertilization changes the soil pH.
    - Lime treatments (two levels: add lime, don't) started in 1903.
    - Want to continue with original treatments and plots.
  - Solution: divide each 1856 plot into two halves. One gets lime; the other does not.
  - Starting 1965, used four lime levels by dividing each 1856 plot into quarters
- Using two sizes of eu makes the study possible but complicates the analysis.

### Planning a split plot study

Two relevant aspects:
- The main plot MS is usually larger than the split plot error.
  - Because of the "mini-blocks" created by the main plots.
- The main plot error has fewer df than the split plot error (lots fewer here!).
  - So t quantiles for tests and confidence intervals will be larger
- Consequences:
  - Estimates of main plot effects have larger standard errors and wider confidence intervals than do estimates of split plot effects.
  - Tests of main plot effects have lower power than do tests of split plot effects.

### Practical advice
- If you can choose which factor to be main and which to be split,
  - put the more important factor at the split plot level.
- If possible to run a study as a 2 way factorial with one size of eu,
  - do it!
  - you're better off using a 2 way factorial instead of a split plot design.
    - The split plot has a slightly higher power for split plot effects (because of "mini-blocks").
    - But vastly smaller power for main plot effects.

**Expected Mean Squares: for concepts, not details**

One way to think about tests and appropriate error terms in an ANOVA table.

Goals of the EMS discussion:
1) Justify a choice of error term in an F test
2) Understand why denominator df is not the same for different quantities
3) Understand why you sometimes get fractional df
All at a conceptual level, equations used only to motivate concepts

Means and differences of means for the grazing study

| Factor | se | df |
|---|---|---|
| Grazing Mean | 0.074 | 9 |
| Difference | 0.104 | 9 |
| | | |
| Implant Mean | 0.045 | 11.1 |
| Difference | 0.025 | 90 |
| | | |
| Cell Mean | 0.077 | 11.1 |
| Difference within Main | 0.043 | 90 |
| Difference within Split | 0.11 | 11.1 |

Context: The data are random variables. So are any quantities computed from data.
Examples we've already seen:
Treatment mean
Difference between two treatment means
T statistic testing difference between two treatment means
F statistic testing equality of 2 or more treatment means
We've (so far) focused on the variability in that random variable, e.g.
standard error of a mean
standard error of a difference between two means

A random variable has an expected value
= theoretical mean
= value for an infinite-sized sample.
Usually written as E random variable. Examples are:
Treatment mean: E $\overline{Y}_i = \mu_i$
Difference of two treatment means: E $(\overline{Y}_i - \overline{Y}_j) = \mu_i - \mu_j$
E T statistic, when null hypothesis is true, = 0
E F statistic, when null hypothesis is true, = 1

Mean Squares in the ANOVA table are computed from data.
So each is a random variable; each has an expected value, called E $MS$.

Each Expected Mean Square, E $MS$ depends on:
population values of fixed effects
population variances of random effects,
and various sample sizes

Simple example: Two way factorial in a CRD.
The model: $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$.
The expected value of the error variance, MSE, is $\sigma^2$.
That's what justifies using $s = \sqrt{MSE}$ in T statistics.
The other terms in the ANOVA table also have EMS.
The ANOVA table:

| Source | df | EMS |
|---|---|---|
| Factor A | $(a-1)$ | $\sigma^2 + Q(\alpha_i)$ |
| Factor B | $(b-1)$ | $\sigma^2 + Q(\beta_j)$ |
| Interaction | $(a-1)(b-1)$ | $\sigma^2 + Q(\alpha\beta_{ij})$ |
| Error | $ab(n-1)$ | $\sigma^2$ |

Q() indicates a quadratic function of the indicated quantities.
Q(anything) = 0 when all the "anythings" are the same.
So $Q(\alpha_i) = 0$ when the main effects of factor A are equal.
$Q(\alpha\beta_{ij}) = 0$ when all interaction effects = 0.

Using EMS to decide whether you have the correct F statistic
The F statistic is a ratio of $MS_{numerator}$ divided by $MS_{denominator}$
You have the correct F statistic when

$$\frac{E\ MS_{numerator}}{E\ MS_{denominator}} = 1$$

when the null hypothesis is true.
Example: Test of Factor A in the ANOVA table above
F = $MS_A/MS_{error}$.
E $MS_A = \sigma^2 + Q(\alpha_i)$. When null true, $Q(\alpha_i) = 0$, so E $MS_A = \sigma^2$.
E $MS_{error} = \sigma^2$
E $MS_A$/E $MS_{error} = \sigma^2/\sigma^2 = 1$
F = $MS_A/MS_{error}$ is the correct F statistic.
If the ratio of E MS is not 1, you have the wrong F statistic

Grazing study: 4 pastures per grazing trt, 3 heifers per implant in each pasture

| Source | df | Expected MS |
|---|---|---|
| Grazing | 2 | $\sigma^2 + 9\sigma^2_{pasture} + Q(grazing)$ |
| Pasture(Grazing) | 9 | $\sigma^2 + 9\sigma^2_{pasture}$ |
| Implant | 2 | $\sigma^2 + Q(implant)$ |
| G*I | 4 | $\sigma^2 + Q(interaction)$ |
| Error | 90 | $\sigma^2$ |

Correct F test for grazing*implant interaction

E $MS_{g*i} = \sigma^2 + Q(interaction) = \sigma^2$ when null is true
Need to find a term with E $MS = \sigma^2$
That's the Error term
So correct F statistic is $MS_{g*i}/MS_{error}$

Correct F test for grazing
E $MS_g = \sigma^2 + 9\sigma^2_{pasture} + Q(grazing) = \sigma^2 + 9\sigma^2_{pasture}$ when null is true
Need to find a term with E $MS = \sigma^2 + 9\sigma^2_{pasture}$
That's the pasture(grazing) term
So correct F statistic is $MS_g/MS_{pasture(grazing)}$

Can go "the other direction" to find out what hypothesis tested by an F statistic.
Consider F = $MS_{grazing}/MS_{error}$.
Ratio of EMS = $\sigma^2_{error} + 9\sigma^2_{pasture} + Q(grazing)$ / $\sigma^2_{error}$.
= 1 only when $\sigma^2_{pasture} = 0$ **and** $Q(main) = 0$.
F statistic will be large either when:
the treatment means are not equal (the intended test)
or when $\sigma^2_{pasture}$ is large
or a combination of both.

**Estimates of means and differences of means.**

Goal for this block of material:
Explain why some unusual things happen, including:
Different standard errors for different comparisons of means,
even when sample sizes the same.
Fractional degrees of freedom (e.g., 24.6)

Standard errors of treatment effects (differences or contrasts among means):
Using the grazing study

| Quantity | (se of mean)$^2$ | (se of difference)$^2$ |
|---|---|---|
| Grazing | $\sigma^2_{error}/36 + \sigma^2_{pasture}/4$ | $2\left(\sigma^2_{error}/36 + \sigma^2_{pasture}/4\right)$ |
| Implant | $\sigma^2_{error}/36 + \sigma^2_{pasture}/12$ | $2\left(\sigma^2_{error}/36\right)$ |
| Cells | $\sigma^2_{error}/12 + \sigma^2_{pasture}/4$ | depends |

Why each divisor?
How many heifers and pastures contribute to a mean?
Grazing: ave. of 4 pastures, each with 9 heifers
Implant: ave. of 12 pastures, each with 3 heifers with that implant

Differences of cell means
two versions, with different se's.

Here is the conceptual explanation without formulae.
Difference between two split plot treatments in the same main plot,
e.g., C/None- C/A.
average of differences within each pasture (mini-block).
se only involves the error variance.
Difference between two main plot treatments in the same or different split plots,
e.g. C/None - R/None, or C/None - R/A.
This involves different pastures
the se involves both the error and pasture variances.

Two things to notice:
se's are sometimes related to the E MS
If so, df are the df associated with that MS

Three examples:
1) Grazing means (Difference of means is twice quantity below)

$$\begin{aligned} se^2 &= \sigma^2_{error}/36 + \sigma^2_{pasture}/4 \\ &= \left(\sigma^2 + 9\sigma^2_{pasture}\right)/36 \\ &= MS_{pasture}/36 \end{aligned}$$

df for this se = df for $MS_{pasture}$
similar for difference of two grazing means

2) Difference between two implants:

$$\begin{aligned} se^2 &= 2\sigma^2_{error}/36 \\ &= 2MS_{error}/36 \end{aligned}$$

df = df for $MS_{error}$

3) Implant mean or cell mean (3 times quantity below)

$$\begin{aligned} se^2 &= \sigma^2_{error}/36 + \sigma^2_{pasture}/12 \\ &= \left(\sigma^2_{error} + 3\sigma^2_{pasture}\right)/36 \end{aligned}$$

Not $MS_{error}$ or $MS_{pasture}$. Don't know the df.

**Satterthwaite approximate degrees of freedom for inconvenient standard errors**

Gives approximate df for se's that are an exact E MS
e.g. Implant mean: $\text{se}^2 = \sigma^2_{error}/36 + \sigma^2_{pasture}/12 = \left(\sigma^2_{error} + 3\sigma^2_{pasture}\right)/36$
Have 2 MS in the ANOVA table: E $MS_{error} = \sigma^2_{error}$ and E $MS_{pasture} = \sigma^2_{error} + 9\sigma^2_{pasture}$

Combine these two to get $\sigma^2_{error} + 3\sigma^2_{pasture}$
$(EMS_{pasture} + 2EMS_{error})/3 = (\sigma^2_{error} + 9\sigma^2_{pasture} + 2\sigma^2_{error})/3 = \sigma^2_{error} + 3\sigma^2_{pasture}$

What is an appropriate df for this linear combination?

Satterthwaite, a quantitative psychologist, worked this out (Psychometrika, 1941).
approximate df for linear combinations of mean squares.
In case you want to know the equation, d.f. for $c_1\text{MS}_1 + c_2\text{MS}_2$:

$$\text{df} = \frac{(c_1\text{MS}_1 + c_2\text{MS}_2)^2}{\frac{(c_1\text{MS}_1)^2}{\text{df}_1} + \frac{(c_2\text{MS}_2)^2}{\text{df}_2}}$$

The result is almost always a non-integer df.
for Implant means: df = 11.1
for Cell means: df = 11.1
Computers have no problem getting F or T statistics for fractional df
e.g., for tests and confidence intervals.

**Kenward-Roger approximations**

More accurate approximation in complicated models
No simple formula, but computers have no problem
Identical to Satterthwaite in a balanced split plot.
Very close to Satterthwaite in unbalanced split plot
essentially no practical difference.
Some analyses, e.g. repeated measures analyses (covered in a few weeks),
the differences matter. KR is preferred.
A complication
there are different Kenward-Roger adjustments,
e.g., first-order, second-order, or "improved"
Often just called K-R by software
Default computing methods

| Software | Default or recommended df |
|---|---|
| JMP | Kenward Roger, 1st order |
| SAS | /ddfm=kr: Kenward-Roger, 2nd order, but can specify lots of others |
| R (emmeans) | Kenward-Roger, not clear which version |

My suggestions:
don't worry about the details.
No harm in always using K-R
Satterthwaite not wrong for split plot analyses
Definitely report what software you used.

**Estimates for the grazing study, using SAS, KR:**

| Factor | se | df |
|---|---|---|
| Grazing Mean | 0.074 | 9 |
| Difference | 0.104 | 9 |
| | | |
| Implant Mean | 0.045 | 11.1 |
| Difference | 0.025 | 90 |
| | | |
| Cell Mean | 0.077 | 11.1 |
| Difference within Main | 0.043 | 90 |
| Difference within Split | 0.11 | 11.1 |

Things to notice / confirm general patterns:

Marginal means and differences have integer df, but small because only 12 pastures

Every thing else has non-integer df because of Satterthwaite/K-R approximations

SE for difference of main plot means = $\sqrt{2}\times$ se for main plot mean

SE for difference of split plot means $< \sqrt{2}\times$ se for split plot mean

Quite a bit less here because pasture variance is twice the error variance

Two different se's and df for comparison of cell means