A factorial is a way of entering two or more treatments into an analysis.

The description of a factorial usually includes a measure of size, a 2 by 2, 3 by 4, 6 by 3 by 4, 2 by 2 by 2, etc.

Interactions were discussed.

Interactions test additivity of the main effects

Interactions are a measure of inconsistency in the behavior or the cells relative to the main effects.

Interactions are tested along with the main effects

Interactions should not be ignored if significant.

Factorial analyses can be done as two-way ANOVAs in SAS, or they can be done as contrasts.

The Randomized Block Design

This analysis is similar in many ways to a "two-way" ANOVA

The CRD is defined by the linear model, $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$. The simplest version of the CRD has one treatment and one error term. The factorial treatment arrangement discussed previously occurred within a CRD, and it had several different treatments, $Y_{ijk} = \mu + \tau_{1i} + \tau_{2j} + \tau_{1i}\tau_{2j} + \varepsilon_{ijk}$. This model has two treatments and one error. It could have many more treatments, and it would still be a factorial design. Designs having a single treatment or multiple treatments can all occur within a CRD and are referred to as different treatment arrangements.

There are other modifications of a CRD that could be done. Instead of multiple treatments we may find it necessary to subdivide the error term.

Why would we do this? Perhaps there is some variation that is not of interest. If we ignore it, that variation will go to the error term. For example, suppose we had a large agricultural experiment, and had to do our experiment in 8 different fields, or due to space limitations in a greenhouse experiment we had to separate our experiment into 3 different greenhouses or 5 different incubators. Now there is a source of variation that is due to different fields, or different greenhouses or incubators!

If we do it as a CRD, we put our treatments in the model, but if there is some variation due to field, greenhouse or incubator it will go to the error term. This would inflate our error term an make it more difficult to detect a difference (we would lose power).

How do we prevent this? First, make sure each treatment occurs in each field, greenhouse or incubator (preferably balanced). Then we would factor the new variation out of the error term by putting it in the model.

$$Y_{ijk} = \mu + \beta_i + \tau_j + \beta_i\tau_j + \varepsilon_{ijk}$$

This is not a new treatment. We will call it a BLOCK. This looks like a factorial, but it is not because the blocks are not a source of variation that we are interested in discussing.

Also, in a factorial the interaction term is likely to be something of interest. In a block design the interaction is an error term, representing random variation of experimental units across treatments.

Another difference, treatments can be either fixed or random. If both treatments are fixed, the interaction is fixed. However, blocks are usually random, and the block interaction is always random.

So why are we blocking?

It is usually used to add replication to an experiment. Additional replicates are added in another field, another greenhouse. On the one hand, the larger experiment should add power. On the other hand, if we do not take measures to keep the new variation out of the error term, we may lose power due to the larger error.

So, how does this affect our analysis?

We still have treatments with the test of treatments in the ANOVA (an F test).

We can still do post-hoc tests on the treatments.

There is only one new issue, the error term. To examine this we will need to look at the expected mean squares (EMS) for the Randomized Block Design.

RBD EMS

We will examine two possible types of models.

In the first model we have treatments and blocks and nothing else. Each treatment occurs in each block ONCE. The experiment is similar to a factorial in some regards, but not many.

The model is

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$$

In this model the error term (ε_{ij}) actually comes from the block by treatment interactions ($\beta\tau_{ij}$). This is the only error available, but that is OK. It is usually a good error term because it represents random variation among the experimental units.

Blocks \ Treatments	A1	A2	A3
Block 1	a_1b_1	a_2b_1	a_3b_1
Block 2	a_1b_2	a_2b_2	a_3b_2
Block 3	a_1b_3	a_2b_3	a_3b_3

This looks like a factorial.

The analysis is the same as the factorial, we get marginal sums or means and proceed to calculate the SS for blocks and treatments and "interaction" as before.

However, there is one big difference. If this was a factorial we would have Treatment A, Treatment B and the A*B interaction.

What would you use as an error term? We would not have one. A factorial ANOVA must have an error term for testing treatments and interaction.

However, since the "interaction" in a block design is assumed to be random variation among experimental units, it serves as an error term.

So the model works for Block designs.

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$$

The "interaction" term is a useful and respectable error term.

We do however, in this case, have one additional assumption.

Assume that there is "no interaction between the treatments and blocks". By interaction here we mean that the treatment patterns are the same in each block. We do not have a treatment behaving one way in one block, and behaving differently in another block.

So the term represents random variation in experimental units and not some interaction in the same sense as "interaction" in a factorial design.

So what about those EMS?

For the CRD we had two cases

Source	d.f.	EMS Random	EMS Fixed
Treatment	t–1	$\sigma_\varepsilon^2 + n\sigma_\tau^2$	$\sigma_\varepsilon^2 + n\dfrac{\Sigma\tau_i^2}{t-1}$
Error	t(n–1)	σ_ε^2	σ_ε^2
Total	tn–1		

For the Block design we have two cases, one with just blocks and treatments, and one with replicate observations within the cells.

Source	d.f.	EMS (no reps)	EMS (with Reps)
Treatment	t–1	$\sigma_\varepsilon^2 + b\sigma_\tau^2$	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + nb\sigma_\tau^2$
Block	b–1	$\sigma_\varepsilon^2 + t\sigma_\beta^2$	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + nt\sigma_\beta^2$
Exptl Error	(b–1)(t–1)	σ_ε^2	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2$
Rep Error	tb(n–1)		σ_ε^2
Total	tbn–1		

What is the nature of the replicates within the block by treatment cells?

Suppose the experimental unit is a plot in a field. We are evaluating plant height.
The treatment to be compared is 6 varieties of soy beans.
The experiment is done in 3 fields (blocks).
The error term is the field by variety combinations.
This experiment is unreplicated within blocks.
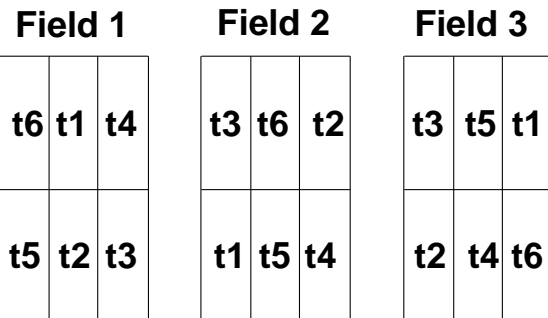
Additional replication is usually done in one of two ways.
If we have only one plot (experimental unit) for each treatment in each field, we could sample several times within each plot, sampling plant height at several places in the plot.

Our "sampling unit" is a smaller unit than the experimental unit (a plot) so we have sampling error.
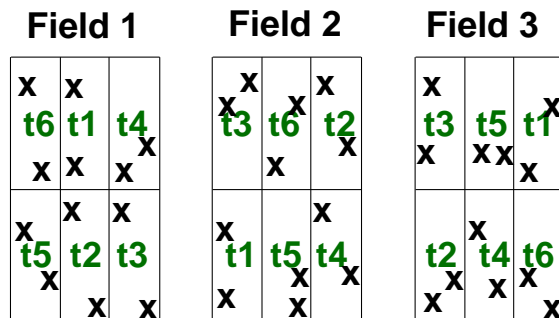
Replicated within blocks as multiple samples in an experimental unit. Error is sampling error.

Field 1   **Field 2**   **Field 3**

| t6 | t1 | t4 | | t3 | t6 | t2 | | t3 | t5 | t1 |

| t5 | t2 | t3 | | t1 | t5 | t4 | | t2 | t4 | t6 |

Field 1   **Field 2**   **Field 3**

Another type of error comes from having several plots with a given soybean variety in each plot. Here each variety of soybean has several experimental units in each field.

In this case the additional replication represents a second experimental error, one for block by treatment combinations and one for replicate plots within a block.

In this case we have replicated experimental units in each block.

Field 1			Field 2			Field 3		
t5	t3	t2	t3	t5	t1	t1	t4	t2
t1	t2	t6	t4	t3	t2	t6	t5	t4
t6	t5	t1	t6	t5	t1	t5	t1	t3
t4	t4	t3	t4	t2	t6	t3	t6	t2

Factorial EMS

I haven't mentioned factorials EMS.

Developing EMS can be pretty simple. Start with the lowest unit, and move up the source table adding additional variance components for each new term.

Source	EMS with Reps
Treatment	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + nb\sigma_\tau^2$
Block	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + nt\sigma_\beta^2$
Exptl Error	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2$
Rep Error	σ_ε^2

Interaction components occur on their own line, and on the source line for each higher effect contained in the interaction.

Each main effect gets its own source.

Now consider whether the effects are fixed or random. Modify fixed effects to show SSEffects instead of variance components.

Source	EMS with Reps
Treatment	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + nb\dfrac{\Sigma\tau_i^2}{t-1}$
Block	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + nt\sigma_\beta^2$
Exptl Error	$\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2$
Rep Error	σ_ε^2

If the model is an RBD we're done, because the interaction is always a random variable.

For factorials that are random models and mixed models were done.

Consider what the F test should be for the treatment. Surprise, SAS always uses the residual error term!

But for factorials there is one last detail. It is perfectly possible in factorial designs that both effects are fixed, and if both effects are fixed the interaction is also fixed!

Source	EMS with Reps
Treatment A	$\sigma_\varepsilon^2 + nb \dfrac{\Sigma \tau_{Ai}^2}{(a-1)}$
Treatment B	$\sigma_\varepsilon^2 + na \dfrac{\Sigma \tau_{Bi}^2}{(b-1)}$
Interaction A*B	$\sigma_\varepsilon^2 + n \dfrac{\sum (\tau_A \tau_B)_{ij}^2}{(a-1)(b-1)}$
Error	σ_ε^2

And a FIXED effect occurs only on its own line, no other!! The fixed interaction disappears from the main effects!!!

Now what is the error term for testing treatments and interaction? Maybe SAS is right? Or maybe SAS just doesn't know what is fixed and what is random.

Testing ANOVAs in SAS

So tell SAS what is random and what is fixed.

Look for the following additions to SAS program.

How do we tell SAS which terms to test with what error term?

How do we get SAS to output EMS?

How do we get SAS to automagically test the right treatment terms with the right error terms?

Summary

Randomized Block Designs modify the model by factoring a source of variation out of the error term in order to reduce the error variance and increase power. If the basis for blocking is good, this will be effective. If the basis for blocking is not good, we lose a few degrees of freedom from the error term and may actually lose power.

The block by treatment combinations (interaction?) provide a measure of variation in the experimental units and provide an adequate error term.

We have an additional assumption that this error term represents ONLY experimental error, and not some real interaction between the treatments and blocks.

Expected mean squares for the RBD indicate that the experimental error term is the correct error term, whether there is a sampling unit or not.

Factorial designs, where effects are random or mixed are similar to RBD EMS. THE TREATMENT INTERACTION IS ACTUALLY USED AS AN ERROR TERM!

When the treatments are fixed, the main effects do not contain the interaction term, and the residual error term is the appropriate error term.

Sample size in ANOVA

Some textbooks use a slightly different expression for the equation, but it is the same as the equation discussed previously. One minor change is the expression $n \geq (t_{\alpha/2} + t_{\beta})^2 \dfrac{S^2}{\bar{d}^2}$. An alternative to the use of $\bar{d}$ is the expressing of the difference as a percentage of the mean. For example, if we wanted to test for a difference that was 10% of the mean we could use the expression $n \geq (t_{\alpha/2} + t_{\beta})^2 \dfrac{S^2}{\left(0.1\bar{Y}\right)^2}$. This expression can further be altered to express the difference in terms of the coefficient of variation $CV = S/\bar{Y}$. Calculating the sample sized needed to detect a 10% change in the mean then becomes $n \geq (t_{\alpha/2} + t_{\beta})^2 (10CV)^2$.

In analysis of variance we may also want to be able to detect a certain difference between two means (μ_1 and μ_2) out of the treatment means we are studying, so our difference will be $\mu_1 - \mu_2$. A prior analysis, or a pilot study, may provide us with an estimate of the variance (MSE in ANOVA). From here we can use a formula pretty much the same as for the t-test discussed earlier. There is one other little detail, however.

We are basically testing $H_0: \mu_1^2 - \mu_2^2 = \delta$, from the 2 sample t-test. Recall from our linear combinations we have a variance for this linear combination that is the sum of the individual variances of the mean. Therefore, the variance will be. $\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}$. Since we are usually pooling variances (ANOVA) then the formula simplifies to $MSE(\dfrac{1}{n_1} + \dfrac{1}{n_2})$

. Furthermore, since we usually attempt to have balanced experiments (equal sample size in each group) for analysis of variance the formula further simplifies to an expression similar to one seen previously, except for the addition of "2", $2MSE/n$. The additional "2" occurs when we are testing for difference in two means ($H_0: \mu_1 = \mu_2$) as opposed to testing a mean against an hypothesized value ($H_0: \mu = \mu_0$).

Note one very important thing here. In this formula "n" represents each group or population being studied, that is, each "treatment level" in an analysis of variance! So for ANOVA or two-sample t-test with equal variance and equal n, the expression for sample size is $n \geq \dfrac{2(t_{\alpha/2} + t_{\beta})^2 MSE}{\bar{d}^2}$. Note that this "n" is for each treatment. In a two sample t-test, each population would have a sample size of "n", so the total number of observations would be 2n.

In ANOVA we have "t" treatments; each would have a sample size of "n", so the total number of observations would be tn. How often are we likely to have situations with equal variance and equal n? Is this realistic? Actually, yes it is.

First, ANOVA traditionally required equal variances, though more modern analytical techniques can address the lack of homogeneity. If necessary, equal variances may be achieved by a transformation or some other fix. If variance is nonhomogeneous you could use the larger estimates and get a conservative estimate of "n".

Second, the most common application for sample size calculation is in planning NEW studies, and of course in planning new studies you usually do not PLAN on unbalanced designs and non homogeneous variance.

So these situations are realistic.

Summary

Finally we saw that this formula is applicable to two-sample t-tests and ANOVA, with some modifications in the estimate of the variance. These modifications are the same ones needed for the 2-sample t-test as dictated by our study of linear combinations. However, the calculations are simplified by the common ANOVA assumption of equal variance and the prevalence of balanced experiments.

Review of Analysis of Variance procedures.

1) H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots = \mu_t = \mu$

2) H_1: some μ_i is different

3a) Assume that the observations are normally distributed about each mean, or that the residuals (i.e. deviations) are normally distributed.

b) Assume that the observations are independent

c) Assume that the variances are homogeneous

4) Set the level of type I error. Usually $\alpha = 0.05$

5) Determine the critical value. For a balanced CRD with a single factor treatment the test is an F test with t–1 and t(n–1) degrees of freedom ($F_{\alpha=0.05,\ t-1,\ t(n-1)\ d.f.}$).

6) Obtain data and evaluate.

The treatment sum of squares, as developed by Fisher, are converted to a "variance" and tested with an F test against the pooled error variance. In practice, the sum of squares are usually calculated and presented with the degrees of freedom in a table called an ANOVA table. For a balanced design (all n_i equal) the calculations are,

The uncorrected SS for treatments is $USS_{Treatments} = \dfrac{\sum\limits_{i=1}^{t}\left(\sum\limits_{j=1}^{n} Y_{ij}\right)^2}{n} = n\sum\limits_{i=1}^{t}\left(\dfrac{\sum\limits_{j=1}^{n} Y_{ij}}{n}\right)^2$.

The uncorrected SS for the total is $SS_{Total} = \sum\limits_{i}\sum\limits_{j} Y_{ij}^2$

The correction factor for both terms is $CF = \dfrac{\left(\sum\limits_{i}^{t}\sum\limits_{j}^{n} Y_{ij}\right)^2}{tn}$

Our ANOVA analyses will be done with PROC MIXED and PROC GLM. There is a PROC ANOVA, but it is a subset of PROC GLM.

LSMeans calculation

The calculations of LSMeans are different. For a balanced design, the results will be the same. However, for unbalanced designs the results will often differ.

The MEANS statement in SAS calculates a simple mean of all available observations in the treatment cells.

The LSMeans statement will calculate the mean of the treatment cell means.

Example:

The MEAN of 4 treatments, where the observations are 3,4,8 for a1, 3,5,6,7,9 for a2, 7,8,6,7 for a3 and 3,5,7 for a4 is 5.8667.

The individual cells means are 5, 6, 7 and 5 for a1, a2, a3 and a4 respectively. The mean of these 4 values is 5.75. This would be the LSMean.

Raw means

Treatments	a1	a2	a3	Means
	5	6	9	
b1	7	8		6.5
		4		
b2	7	5	5	6.6
	9		7	
Means	7	5.75	7	

LSMeans means

Treatments	a1	a2	a3	Means
b1	6	6	9	7
b2	8	5	6	6.33
Means	7	5.5	7.5	

Confidence Intervals on Treatments

Like all confidence intervals on normally distributed estimates, this will employ a t-value and will be of the form $\text{Mean} \pm t_{a/2} S_{\bar{Y}}$

The treatment mean can be obtained from a means (or LSMeans) statement, but the standard deviation provided is not the correct standard error for the interval.

The standard error in a simple CRD with fixed effects is the square root of MSE/n, where n is the number of observations used in calculating the mean.

The calculation requires other considerations when random components are involved. For example, in PROC MIXED the use of the Satterthwaite and Kenward-Roger approximations, the use of various estimation methods (the default is REML) and specifications of covariance structure are all things that can affect degrees of freedom.

The use of MSE in the numerator is the default in PROC GLM, and if a different error is desired it must be specified by the user. PROC MIXED is capable of detecting and using and error other than the MSE where appropriate.

If there are several error terms (e.g. experimental error and sampling error) use the one that is appropriate for testing the treatments. When an error term other than the residual is appropriate for testing the treatments, the degrees of freedom for the tabular t value are the d.f. from the error term used for testing. This variance term would also be used to calculate the standard error for treatment means.