Two-way ANOVA under Model II (where both factors are random):

For example, we might be studying the effects of particular target stimuli and levels of complexity – and we randomly sample both factors to assess their (joint) effects

The Model:

$$Y_{ijk} = \mu_{..} + a_i + b_j + (ab)_{ij} + \epsilon_{ijk}$$

where

$a_i \sim N(0, \sigma_a^2)$, $b_i \sim N(0, \sigma_b^2)$,$(ab)_{ij} \sim N(0, \sigma_{ab}^2)$,

$\epsilon_{ijk} \sim N(0, \sigma^2)$, and everything is independent of everything

Because we have only sampled levels, it can't be assumed that interactions sum to zero;

thus, the expectations of the various mean-squares are different compared to Model I

We restrict our discussion to equal n's

	Model I	Model II
$E(MSA)$	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1}$	$\sigma^2 + nb\sigma_a^2 + n\sigma_{ab}^2$
$E(MSB)$	$\sigma^2 + na\frac{\sum \beta_j^2}{b-1}$	$\sigma^2 + na\sigma_b^2 + n\sigma_{ab}^2$
$E(MSAB)$	$\sigma^2 + n\frac{\sum\sum (\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\sigma^2 + n\sigma_{ab}^2$
$E(MSE)$	σ^2	σ^2

Thus, to test:

$H_o : \sigma_a^2 = 0;\ \frac{MSA}{MSAB} \sim F_{a-1,(a-1)(b-1)}$

$H_o : \sigma_b^2 = 0;\ \frac{MSB}{MSAB} \sim F_{b-1,(a-1)(b-1)}$

$H_o : \sigma_{ab}^2 = 0;\ \frac{MSAB}{MSE} \sim F_{(a-1)(b-1),(n-1)ab}$

Question: how did we do the comparable tests under Model I?

Variance component estimation:

$$\widehat{\sigma}_a^2 = \frac{(MSB - MSAB)}{an}$$

$$\widehat{\sigma}_b^2 = \frac{(MSA - MSAB)}{bn}$$

$$\widehat{\sigma}_{ab}^2 = \frac{(MSAB - MSE)}{n}$$

$$\widehat{\sigma}^2 = MSE$$

Intraclass correlations (also, think generalizability coefficients in test theory):

$$\widehat{\rho}_{I_A} \equiv \frac{\widehat{\sigma}_a^2}{\widehat{\sigma}_a^2 + \widehat{\sigma}_b^2 + \widehat{\sigma}_{ab}^2 + \widehat{\sigma}^2}$$

Model III: the Mixed Model

One factor is random and one factor is fixed

Suppose A is fixed and B is random

$Y_{ijk} = \mu_{..} + \alpha_i + b_j + (\alpha b)_{ij} + \epsilon_{ijk}$

$b_j \sim N(0, \sigma_b^2)$; $(\alpha b)_{ij} \sim N(0, \frac{a-1}{a}\sigma_{\alpha b}^2)$;

$\epsilon_{ijk} \sim N(0, \sigma^2)$; and independence (almost) everywhere

$\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{i=1}^{a} (\alpha b)_{ij} = 0$

$(\alpha b)_{ij}$ and $(\alpha b)_{i'j'}$ are dependent if $j = j'$

$$E(MSA) = \sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} + n\sigma_{\alpha b}$$

$$E(MSB) = \sigma^2 + na\sigma_b^2$$

$$E(MSAB) = \sigma^2 + n\sigma_{\alpha b}^2$$

$$E(MSE) = \sigma^2$$

Thus, to test:

$$H_o : \alpha_i = 0;\ \frac{MSA}{MSAB} \sim F_{a-1,(a-1)(b-1)}$$

$$H_o : \sigma_b^2 = 0;\ \frac{MSB}{MSE} \sim F_{b-1,(a-1)(b-1)}$$

These last two are counterintuitive given the denominators (i.e., MSE is used to test the random factor; MSAB is used to test the fixed factor)

$$H_o : \sigma_{ab}^2 = 0;\ \frac{MSAB}{MSE} \sim F_{(a-1)(b-1),(n-1)ab}$$

Some estimation is possible but no intraclass correlations

$\widehat{\alpha}_i = \widehat{\mu}_{i\cdot} - \widehat{\mu}_{\cdot\cdot}$

$\widehat{\sigma}^2_b = \frac{(MSB - MSE)}{bn}$

$\widehat{\sigma}^2_{\alpha b} = \frac{(MSAB - MSE)}{n}$

———————-

Now, let's see what happens when $n = 1$ (i.e., we don't have replication);

obviously, there is no MSE but the other terms can be calculated

	Model I	Model II	Model III (A fixed)
$E(MSA)$	$\sigma^2 + b\frac{\sum \alpha_i^2}{a-1}$	$\sigma^2 + b\sigma_a^2 + \sigma_{ab}^2$	$\sigma^2 + b\frac{\sum \alpha_i^2}{a-1} + \sigma_{\alpha b}^2$
$E(MSB)$	$\sigma^2 + a\frac{\sum \beta_j^2}{b-1}$	$\sigma^2 + a\sigma_b^2 + \sigma_{ab}^2$	$\sigma^2 + a\sigma_b^2$
$E(MSAB)$	$\sigma^2 + \frac{\sum\sum (\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\sigma^2 + \sigma_{ab}^2$	$\sigma^2 + \sigma_{\alpha b}^2$

In Model I, we can test both main effects if $(\alpha\beta)_{ij} = 0$, or, if it can be assumed; they can't be tested in the usual way since there is no MSE to serve as a denominator

(We might note in passing that it is possible to test if a particular kind of interaction is present – through Tukey's one-degree-of-freedom test for nonadditivity – but we will not pursue this)

In Model II, we can test both main effects since $MSAB$ is used as a denominator in each case

In Model III, we can test the fixed factor; but only if $\sigma^2_{\alpha b}$ can be assumed zero can we test the random factor

In Model II, multiple comparisons are not done; comparisons would be (nonsensically) random variables

In Model III, we can do multiple comparisons on the fixed factor using $MSAB$ in place of MSE

When $n = 1$, we typically refer to these designs as "randomized block designs"; they extend the paired t-test, particularly under Model III where the blocks of subjects are considered the random factor

blocks consist of b subjects each that are randomly assigned to the treatments (the Fixed column factor)

The same caveat happens as before in comparing the independent and dependent t

We lose degrees-of-freedom since we are using interaction in the denominator (over doing a one-way anova); we therefore need to have a blocking or matching variable that does account for a lot of Sum of Squares

(Remember, SSE in the one-way anova is now the sum of the SS blocks and SS interaction)

In these cases it may still make sense to assume the error terms are independent but what happens if the block is a single subject observed over the treatments?

The error terms are probably not independent and have some degree of correlation

Here's the initial repeated measures design under Model III:

We relax the error term independence assumption (as we note below)

$$Y_{ij} = \mu_{..} + a_i + \beta_j + (a\beta)_{ij} + \epsilon_{ij}$$

where the subjects are the rows (Factor A, (now) assumed random); the treatments are the columns (Factor B, (now) assumed fixed)

$a_i \sim N(0, \sigma_a^2)$; β_j is fixed; $(a\beta)_{ij} \sim N(0, \frac{(b-1)}{b}\sigma_{a\beta}^2)$;

$\epsilon_{ij} \sim N(0, \sigma^2)$; the correlation between ϵ_{ij} and $\epsilon_{ij'}$ for $j \neq j'$ is a constant ρ

This is the same for *all* treatment pairs;

it is called the assumption of compound symmetry

$$E(MSA) = \sigma^2[1 + (b-1)\rho] + b\sigma_a^2$$

$$E(MSB) = \sigma^2[1-\rho] + a\frac{\sum_{j=1}^{b}\beta_j^2}{b-1} + \sigma_{a\beta}^2$$

$$E(MSAB) = \sigma^2[1-\rho] + \sigma_{a\beta}^2$$

The test we are looking for: $(H_o : \beta_j = 0)$

$$\frac{MSB}{MSAB} \sim F_{b-1,(a-1)(b-1)}$$

Source	df	SS	MS
Between Subjects (factor A)	$a-1$	SSA	MSA
Within Subjects	$a(b-1)$		
Between Treatments (factor B)	$b-1$	SSB	MSB
Subjects $\times$ Treatments (A $\times$ B)	$(a-1)(b-1)$	SSAB	MSAB

Issues in Repeated-Measures Analyses:

The analysis of repeated measures generally needs special treatment in that the usual models are not very trustworthy and can lead to erroneous conclusions.

The starting place is commonly a Mixed Model III ANOVA with a fixed treatment factor and a subject factor considered random.

To model repeated observations justifying the usual F-ratio test statistic of Mean-Square Treatments to Mean-Square Interaction, an assumption is made that the observations within a subject are correlated.

The use of the usual F-ratio, however, requires that all these correlations be the same irrespective of which pair of treatments is considered (an assumption of "compound symmetry").

The compound symmetry assumption may be reasonable when the treatment times are randomly assigned, but if the responses are obtained sequentially, then possibly not.

Treatments further apart in time are typically less correlated because of fatigue, boredom, familiarity, and so on.

Unfortunately, there is strong evidence of non-robustness in the use of the equicorrelation assumption when it is not true, with too many false rejections of the null hypothesis of no treatment differences.

A way around this nonrobustness is implemented in many software packages.

If we knew the structure of all the variances and covariances among the treatments, we could obtain a parameter, say, θ, that would give an appropriate correction for the degrees of freedom of the F-distribution against which to compare the calculated F-ratio;

that is, we would use $F_{\theta(B-1),\theta(A-1)(B-1)}$, where there are B treatments and A subjects.

Although θ is unknown, there are two possible strategies to follow:

estimate θ with Huynh–Feldt procedures (as is done, for example, in SYSTAT);

or use the greatest reduction possible with the discounting bound of $1/(B-1)$ (that is, the Geisser–Greenhouse result of $1/(B-1) \leq \theta$, also as done when an analysis is carried out using SYSTAT).

So, if a rejection occurs with the Geisser–Greenhouse method, it would also occur for the Huynh–Feldt estimation, or if you knew and used the actual value of θ.

Another approach to repeated measures, called "profile analysis," uses Hotelling's T^2 statistic and/or MANOVA on difference scores.

In fact, the only good use of a usually non-informative MANOVA may be in a repeated-measures analysis.

Three types of questions are commonly asked in a profile analysis:

Are the profiles parallel to each other? (the Interaction Test)

Are the profiles coincident? (a Main Effect Test)

And, are the profiles horizontal? (a Main Effect Test)

When done well, a profile analysis can give an informative interpretation of repeated-measures information with an associated graphical presentation.

Two possible issues with repeated measures should be noted.

First, it is assumed that the responses from our subjects are commensurable over the variables measured.

If not, an artificial transformation could be considered such as to z-scores, but by so doing, the test for horizontal profiles is not meaningful because the associated test statistic is identically zero.

Second, the number of subjects versus the number of measurement times may prevent carrying out a Hotelling T^2 comparison in a profile analysis (but not, say, a correction based on a Huynh–Feldt estimated θ).

Generally, if there are more time points than subjects, one of the degrees of freedom in the F-distribution used for the T^2 comparison is negative, and thus, the test is meaningless.

Repeated measurements obtained on the same subjects generally require a different approach than do singly occurring measurements.

As discussed above, there are various ways of not taking advantage of the usual compound symmetry assumption;

for example, carrying out the analyses of variance with Huynh–Feldt or Geisser–Greenhouse corrections, or using alternative Hotelling T^2 or MANOVA approaches.

The issues, however, go deeper than just these types of split-plot experimental designs, and the special circumstances that repeated measures offer are commonly just ignored.

Anyone analyzing repeated measures needs to remember that the variance of the difference between two means, say $\bar{X}$ and $\bar{Y}$, is not the same when $\bar{X}$ and $\bar{Y}$ are based on independent samples.

In particular, suppose $\bar{X}$ is obtained for the observations $X_1, \ldots, X_N$, and $\bar{Y}$ for $Y_1, \ldots, Y_N$.

When the samples are independent, the variance of the difference $\bar{X} - \bar{Y}$, $S^2_{\bar{X}-\bar{Y}}$, can be estimated as $S^2_{\bar{X}} + S^2_{\bar{Y}}$, where $S^2_{\bar{X}} \equiv S^2_X/N$, $S^2_{\bar{Y}} \equiv S^2_Y/N$, and S^2_X and S^2_Y are the sample variances for $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$, respectively.

In the repeated-measures context, the variance of $\bar{X} - \bar{Y}$ can be estimated as $S^2_{\bar{X}} + S^2_{\bar{Y}} - 2(S_{XY}/N)$, where S_{XY} is the sample covariance between the observations $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$.

Thus, we have a difference in the term, $-2(S_{XY}/N)$, which in most instances will be a negative correction when the X and Y observations are positively related.

In other words, the variance of the difference, $\bar{X} - \bar{Y}$, will generally be less in the context of repeated measures compared to independent samples.

In some areas of neuroimaging, the repeated-measures nature of the data is just ignored;

we have pixels (or voxels) that are spatially arranged (and subject to various types of spatial autocorrelation) that move through time (and subject again to various types of temporal autocorrelation).

In these frameworks where the repeated measures are both spatial and temporal, it is not sufficient to just use the various multivariate general linear model extensions that assume all error terms are independent and identically distributed

(as suggested by some best-selling fMRI handbooks; for example, see Huettel, Song, & McCarthy, 2004, pp. 336–348).

Geographers have struggled with this type of spatial and temporal modeling for decades, and have documented the issues extensively

A related repeated-measures topic is in the time-series domain, where some variable is observed temporally.

Substantial modeling efforts have involved the Box–Jenkins approach of using ARIMA models

(autoregressive-integrated-moving-average).

A more subtle question in this context is to assess the effects of an intervention on the progress of such a time series.

In the case of single-subject designs, where a subject serves as his or her own control, the issue of evaluating interventions is central (see Kazdin, 1982).

A particularly elegant approach to this problem has been developed by Edgington (see Edgington & Onghena, 2007, Chapter 11: "N-of-1 Designs"), where intervention times are chosen randomly.

The same logic of analysis is possible as in a Fisherian (1971) approach to analyzing an experiment where the various units have been assigned at random to the conditions.

Issues with Matching and Blocking (Redux):

One of the main decision points in constructing an experimental design is whether to block or match subjects, and then within blocks randomly assign subjects to treatments.

Alternatively, subjects could be randomly assigned to conditions without blocking.

As discussed earlier, it is best to control for initial differences beforehand.

Intact groups can't be equated legitimately after the fact through methods such as analysis of covariance or post-hoc matching.

But the question here is whether blocking makes sense over the use of a completely randomized design.

This choice can be phrased more formally by comparing the test statistics appropriate for a two-independent or a two-dependent samples t-test.

The principle derived from this specific comparison generalizes to more complicated designs.

Suppose we have two equal-sized samples of size N, $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$.

When the two samples are independent, the two-independent samples t-statistic has the form

$$\frac{\bar{X} - \bar{Y}}{\sqrt{(S_X^2 + S_Y^2)/(N-1)}} \ ,$$

where S_X^2 and S_Y^2 are the sample variances;

this statistic is compared to a t-distribution with $2(N-1)$ degrees of freedom.

When the samples are dependent and X_i and Y_i are repeat observations on the ith subject, the paired t-statistic has the form

$$\frac{\bar{X} - \bar{Y}}{\sqrt{S_D^2/(N-1)}} \ ,$$

where S_D^2 is the sample variance of the difference scores.

Here, the paired t-statistic is compared to a t-distribution with $N - 1$ degrees of freedom.

We note the relation $S_D^2 = S_X^2 + S_Y^2 - 2S_{XY}$, where S_{XY} is the sample covariance between X and Y.

In the initial design of an experiment, there may be a choice:

match subjects and assign members within a pair to the treatments, or just assign all subjects randomly to the two treatments without matching.

Generally, if the matching variable is not very important in that the sample covariance is not that large (and positive), to compensate for the halving of the degrees of freedom in going from $2(N-1)$ to $(N-1)$, it only hurts to match.

To compensate for the loss of degrees of freedom and make the paired t-statistic sufficiently larger than the independent sample t-statistic, the variance of the differences, S_D^2, in the denominator of the paired t-statistic must be sufficiently smaller compared to $S_X^2+S_Y^2$ in the denominator of the independent samples t-statistic.

Unfortunately, unless one has some estimate of the covariance of X and Y, the choice of design must be based on a guess.

A dictum, however, may still be gleaned:

don't block or match on variables that have no possible (positive and relatively strong) relation to the type of responses being measured.