ST553 Review

Nick Sun

May 13, 2019

Chapter 3: Completely Randomized Design

An experiment has 4 main components:

- Experimental unit: smallest unit to which a treatment is applied
- Treatments
- Design (How EUs are allocated to treatments)
- Measurement on experimental units

Typical research questions involve:

- Comparing treatments
- Estimating the treatment effects

It is usually better to use **balanced** designs instead of **unbalanced** designs.

- The estimates of the treatments means in each group has the same precision
- We need to choose N large enough to get reasonable precision and power
- Given N experimental units, allocation to groups is sampling without replacement, so experimental units aren't quite independent

Model for the CRD looks like:

$$y_{ij} = \mu_i + \epsilon_{ij}$$

with the assumption that $\epsilon_{ij} \sim N(0, \sigma^2)$.

Inference in CRD:

Defining some notation first: we have

- A total sample size of N
- g treatment groups
- n_i = sample size for treatment group i

We have a point estimate $\hat{\mu_i}$ and an estimate for the variance:

$$\hat{\mu_i} = \bar{y}_i$$

$$\hat{\sigma}^2 = MSE = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2}{N - g}$$

where

$$\bar{y}_{i.} \sim N(\mu_i, \frac{\sigma^2}{n_i})$$

so $SE(\bar{y}_{i.}) = \frac{s}{\sqrt{n_i}}$

The test statistic for testing $H_0 : \mu_i = \mu_0$

$$\frac{\bar{y}_i - \mu_0}{s/\sqrt{n_i}} \sim t_{N-g} \text{ if } H_0 \text{ true}$$

and the $(1-\alpha)100\%$ for μ_i is

$$\bar{y}_{i.} \pm t_{\alpha/2,N-g} \frac{s}{\sqrt{n_i}}$$

Fun math-stats facts from 552:

$$\begin{aligned} s^2 = \frac{(N-g)s^2}{\sigma^2} &= \frac{\sum_{j=1}^{n_1}(y_{ij} - \bar{y}_{1.})^2}{\sigma^2} + \ldots + \frac{\sum_{j=1}^{n_g}(y_{ij} - \bar{y}_{g.})^2}{\sigma^2} \\ &\sim \chi^2_{(n_1-1)} + \ldots + \chi^2_{(n_g-1)} = \chi^2_{N-g} \end{aligned}$$

The terms that constitute the χ^2_{N-g} are independent because of our independence assumption.

Note another *fun math-stats fact:*

$$\frac{\bar{y}_i - \mu_i}{\sigma^2} \perp \frac{(N-g)s^2}{\sigma^2}$$

which gives us the beautiful and ugly pivotal quantity

$$\frac{\frac{\bar{y}_{i.} - \mu_i}{\sigma/\sqrt{n_i}}}{\sqrt{\frac{s^2(N-g)}{\sigma^2}/(N-g)}} \sim t_{N-g}$$

Parameterizing the CRD model

Recall the matrix formulation of the linear model:

$$Y = X\beta + \epsilon$$

and the least squares estimate of β is $\hat{\beta} = (X'X)^{-1}X'Y$

The cell means parameterization with g = 3, n_i = 2:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{pmatrix}$$

The **regression parameterization** $y_{ij} = \beta_0 + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \epsilon_i j$

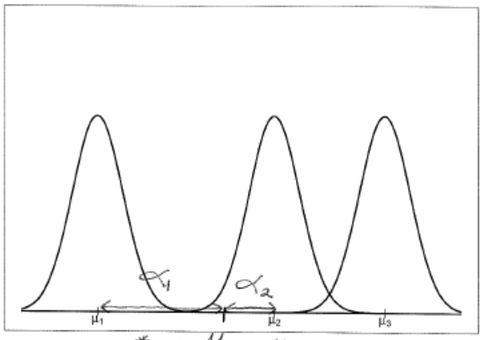
$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{pmatrix}$$

where group 3 is the reference group.

The **factor effects parameterization** is most useful with multiple factors:

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

where $\mu_i = \mu + \alpha_i$



$$\underbrace{\mu^*}_{\text{Oehlert's notation}} = \mu = \frac{\mu_1 + \mu_2 + \mu_3}{3}$$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{pmatrix}$$

However there is a problem here. The columns of $\mathbf{X}$ are not linearly independent so $(X'X)^{-1}$ doesn't exist, therefore we can't estimate the parameters uniquely. This model is *overparameterized.*

α_is are deviations from μ, so $\sum_{i=1}^{g} \alpha_i = 0$, so $\alpha_g = -\alpha_1 - \ldots - \alpha_{g-1}$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{pmatrix}$$

So to recap the three parameterizations of the same model are:

- Cell means $y_{ij} = \mu_i + \epsilon_{ij}$
- Regression $y_{ij} = \beta_0 + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \epsilon_{ij}$
- Factor effects $y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \sum_i \alpha_i = 0$

F-test as a General Linear test

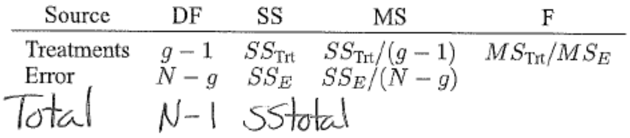
$$\text{F statistic} = (\frac{SSE_{reduced} - SSE_{full})/(df_{reduced} - df_{full})}{SSE_{full}/df_{full}}$$

The numerator is the reduction in unexplained variance due to going to the more complicated model. The statistc itself measure improvement in model fit per unit of model complexity.

We can use the F-test to test for differences among the population means. *The cell means parameterization is the most useful for one-factor CRD.*

Some general facts:

- The $SSE_{model} = \sum_{i=1}^{n}(y_i - \hat{y_i})^2$
- degrees of freedom = sample size - number of mean parameters
  - df(full) = N - g
  - df(reduced) = N - 1



Source	DF	SS	MS	F
Treatments	$g-1$	SS_{Trt}	$SS_{\text{Trt}}/(g-1)$	MS_{Trt}/MS_E
Error	$N-g$	SS_E	$SS_E/(N-g)$	
Total	N-1	SStotal		

The *error row* refers to the full model $y_{ij} = \mu_i + \epsilon_{ij}$, *total row* refers to the reduced model $y_{ij} = \mu + \epsilon_{ij}$, and the *treatments row* refers to the difference between the total and error rows.

Cochran's Theorem: Suppose $Z_1, \dots, Z_\eta \sim N(0,1)$ and $\sum_{i=1}^{\eta} Z_i^2 = Q_1 + \dots + Q_s$ where each Q_i is a sum of squares with η_i degrees of freedom. Then $Q_1, \dots, Q_s$ are independent χ^2 random variables iff $\eta_1 + \dots + \eta_s = \eta$.

$$SS_{total} = SS_{treatments} + SSE$$
$$SS_{total}/\sigma^2 = SS_{treatments}/\sigma^2 + SSE/\sigma^2$$

where the degrees of freedom on the RHS is N-1 and the degrees of freedom on the LHS is g-1+N-g.

Then by Cochran's Theorem, $SS_{treatment}/\sigma^2$ and SSE/σ^2 are independent χ^2 random variables.

So
$$\frac{SS_{treatment}/\sigma^2(1/(g-1))}{SS_{treatment}/\sigma^2(1/(N-g))} = \frac{MS_{treatment}}{MSE} \sim F_{g-1,N-g}$$

Expected value of Mean Squares:

- E(MSE) = σ^2
- $E(MS_{treatment}) = \sigma^2 + \sum_{i=1}^{g} n_i(\mu_i - \mu)^2 = \sum_{i=1}^{g} n_i(\alpha_i)^2$

If $H_0 : \mu_1 = \dots = \mu_g$ is true, then $E(MS_{treatment}) = \sigma^2$ If at least one of the pairs of μs are different, then $E(MS_{treatment}) > E(MSE)$

In more complex designs, the denominator of the F-statistic may not be the MSE! The numerator will be MS for the effect we are testing and the denominator will be the MS whose expected value will be the same as the expected value for the numerator when the null hypothesis is true.

Contrast Sum of Squares

In experimental design framework, sums of squares quantify variability due to different factors.

We have already gone over $SS_{treatment} = SSE_{reduced} - SSE_{full}$, but we can also define a contrast sum of squares.

A contrast sum of squares also has a *full* and *reduced* model where the full model is $y_{ij} = \mu_i + \epsilon_{ij}$ and the full model is the same but with $H_0 : C = 0$ e.g. $H_0 : C = \mu_1 - \mu_4 = 0$ imposed.

In the above example, the reduced model would have $\mu_1 = \mu_4$. The degrees of freedom for the full model would be **N-g**, the reduced model would be **N-g-1** and the contrast df would be 1. The contrast df will be 1 for any contrast since we should be able to write the last mean in terms of the others.

The **F-test for** $H_0 : C = 0$ is

$$\text{F-statistic} = \frac{SS_{constrast}/1}{MSE_{full}}$$

where sum of squares contrast is found using the formula:

$$\frac{(\sum_{i=1}^{g} w_i \bar{y}_i)^2}{\sum_i w_i^2 / n_i}$$

F-statistic $\sim F_{1,N-g}$ if the null is true and since the numerator degrees of freedom is 1, F-statistc $= (tstatistic)^2$

Contrast sum of squares quantifies variation in the data due to that comparison.

Orthogonal Contrasts

Contrasts $C_1 = \sum_{i=1}^{g} w_i \mu_i$ and $C_2 = \sum_{i=1}^{g} w_i^* \mu_i$ are orthogonal if $\sum_{i=1}^{g} \frac{w_i w_i^*}{n_i} = 0$

We usually only consider orthogonal constrasts with balanced data so in this case we can get rid of the n_i in the denominator.

With g treatments, we can have at most g-1 mutually orthogonal contrasts.

Why are orthogonal contrasts nice? We can partition variability due to treatments into variation due to *specific comparisons.*

7/12/2011 p.

Silage Example:

	Chem. treats		Veg. treat	
	1	2	3	4
	NaCl	Formic acid	Beet pulp	Control
	80.5	89.1	77.8	76.7
	79.3	75.7	79.5	77.2
	79.0	81.2	77.0	78.6
Means	79.6	82.0	78.1	77.5
Grand mean	79.3			

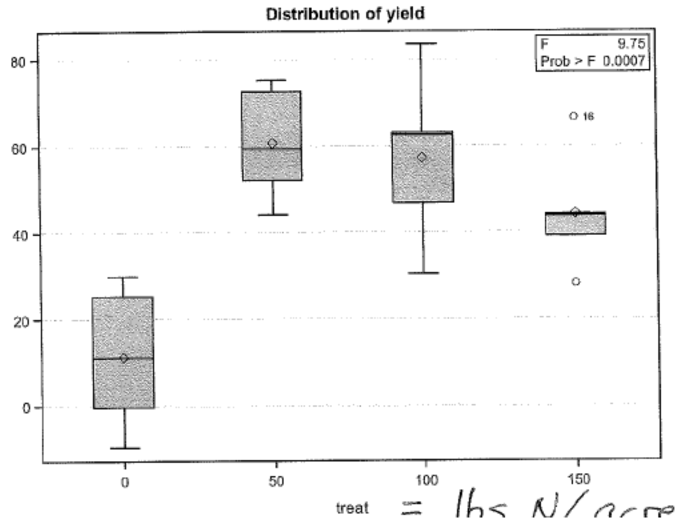
g-1 = 3 of them

A Complete Set of Orthogonal Contrasts

μ_1	μ_2	μ_3	μ_4	
1/2	1/2	-1	0	Compare chemical and vegetable treatments
1	-1	0	0	Compare chemical treatments
1/3	1/3	1/3	-1	Compare treatments to control

We can find lots of complete sets of orthogonal contrasts. Most will make no sense for the experiment design. Contrasts should be decided at the design phase to reflect research questions.

Polynomial regression



In this hypothetical example, we have an ordinal predictor variable. With g sample means, we can find a g-1 degree polynomial containing them.

For the above model, we could have two different parameterizations:

- Cell means

$$y_{ij} = \mu_i + \epsilon_{ij}$$

- and polynomial parameterization

$$y_{ij} = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_{ij}$$

Cell means and g-1 degree polynomials are two parameterizations of the same model. The ANOVA table between the two models should be identical. However, the interpretation is different. Polynomial models allow for interpolation and extrapolation.

Type I SS are sequential: full model includes all predictors up to and including that row. Reduced model includes all predictors up to the row before.

Type III SS are parial: full model includes all variables. Reduced model includes all predictors except that on that row. Sometimes the Type I and Type III sum of squares will be the same because they are examining the same models.

Lack of Fit F-tests

Lack of fit linear fit test:

- Full model: saturated model
- Reduced model: simple linear regression

Lack of quadratic fit test:

- Full model: saturated model
- Reduced model: Quadratic regression

Orthogonal Polynomial Contrasts

If X, X^2 are collinear or nearly collinear, then a good remedy will be centering X to remove multicollinearity. With higher order terms, you might to do more than just center (Gram-Schmidt orthogonalization).

Type I and III sum of squares are the same because the columns of X are orthogonal.

Orthogonal polynomial contrasts are just tidy polynomial regression. Need balanced data and incremental, equally spaced treatments.

Multiple Comparisons/Simultaneous Inference

Simultaneous inference problem: Suppose we want to do 100 independent hypothesis tests at $\alpha = .05$ and all 100 null hypotheses are true. We would expect to reject about 5 true null hypotheses.

We have some notation to consider:

- **Per comparison type I error rate**: reject H_{0i} when H_{0i} is true (our usual α)
- **Experimentwise type I error rate**: α_E = probability we reject at least one H_{0i} when all H_{0i} are true.
- **False Discovery Rate**: 0 if there are no rejections and the proportion of false rejections over the total rejections otherwise.
- **Strong familywise error rate**: Probability of at least one false rejection, or probability that FDR > 0
- **Simultaneous confidence intervals**: probability that all confidence intervals cover their respective true parameters.

Data snooping is deciding what comparisons to make after looking at the data. Type I error rate is not preserved if you pick out signficant comparisons unless you use Scheffes method.

Scheffe's Method: Reject $H_{0i} : C_i = 0$ if $\frac{SSC_i/(g-1)}{MSE} > F_{\alpha_F, g-1, N-g}$ Scheffe's method works because it tests all possible contrasts, so it doesn't matter if we snooped the data. This F-statistic is the ANOVA F-statistic to test $H_0 : \mu_1 = \ldots = \mu_g$ but with SSC_i substituted in for $SS_{treatment}$.

We can find simultaneous $(1 - \alpha_F)$100% confidence intervals using

$$\hat{C}_i \pm \sqrt{(g-1)F_{\alpha_F, g-1, N-g}}SE(C_i)$$

P-value corrections

- **Bonferroni**: reject H_{0i} if $p_i < \alpha_F/k$ where k is the number of comparisons, α_F is the strong family wise error rate
- **Holm**: reject H_{0i} if $p_i < \frac{\alpha_F}{k-i+1}$. This is a little more powerful than Bonferroni
- **FDR method**: $p_1, \ldots, p_k$ are sorted p-values. Reject H_{0i} if $p_i < \frac{iFDR}{k}$. Requires the tests to be independent

Bonferroni and Holm are both very conservative procedures. FDR method controls FDR only, not α_F

Multiple Pairwise Comparisons

- **Tukey's Honestly Significant Differences**: Method to create simultaneous $(1 - \alpha_F)$100%

$$\bar{y}_{i.} - \bar{y}_{j.} \pm q_{\alpha_F, g, N-g}\sqrt{\frac{MSE}{n}}$$

and reject $H_0 = \mu_i = \mu_j$ if

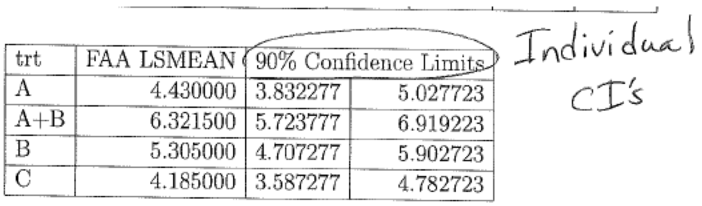
$$\frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MSE/n}} > q_{\alpha_F, g, N-g}$$

where q is the studentized range quantile which can be found using `qtukey` in R.

- **Tukey-Kramer** for slightly unbalanced data is similar to Tukey's HSD

$$\frac{MSE/n}{\rightarrow}\sqrt{MSE\frac{n_i + n_j}{2n_in_j}}$$

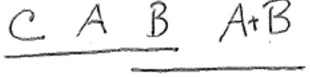
- In SAS, Tukey's HSD is often accompanied by some underline diagrams:



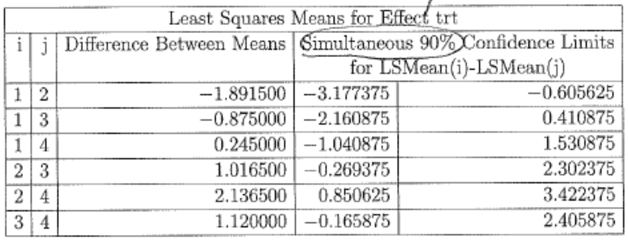
trt	FAA LSMEAN	90% Confidence Limits	
A	4.430000	3.832277	5.027723
A+B	6.321500	5.723777	6.919223
B	5.305000	4.707277	5.902723
C	4.185000	3.587277	4.782723

Individual CI's

Underline diagram: Order groups from small to large according to sample means.

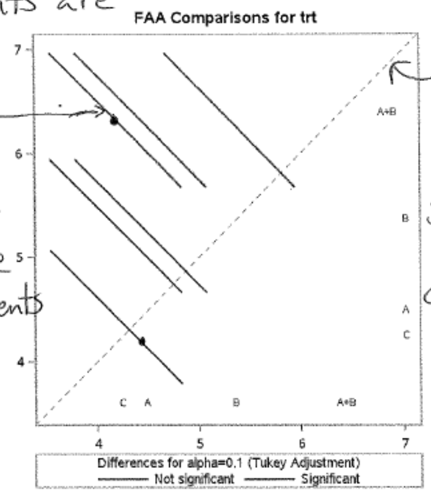


Tukey CI's 4/19/2019 p.2



Least Squares Means for Effect trt				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	−1.891500	−3.177375	−0.605625
1	3	−0.875000	−2.160875	0.410875
1	4	0.245000	−1.040875	1.530875
2	3	1.016500	−0.269375	2.302375
2	4	2.136500	0.850625	3.422375
3	4	1.120000	−0.165875	2.405875

Line segments are pairwise comparisons



Comparing μ_C & μ_{A+B}

Length of line segment represents width of CI for $\mu_C - \mu_{A+B}$

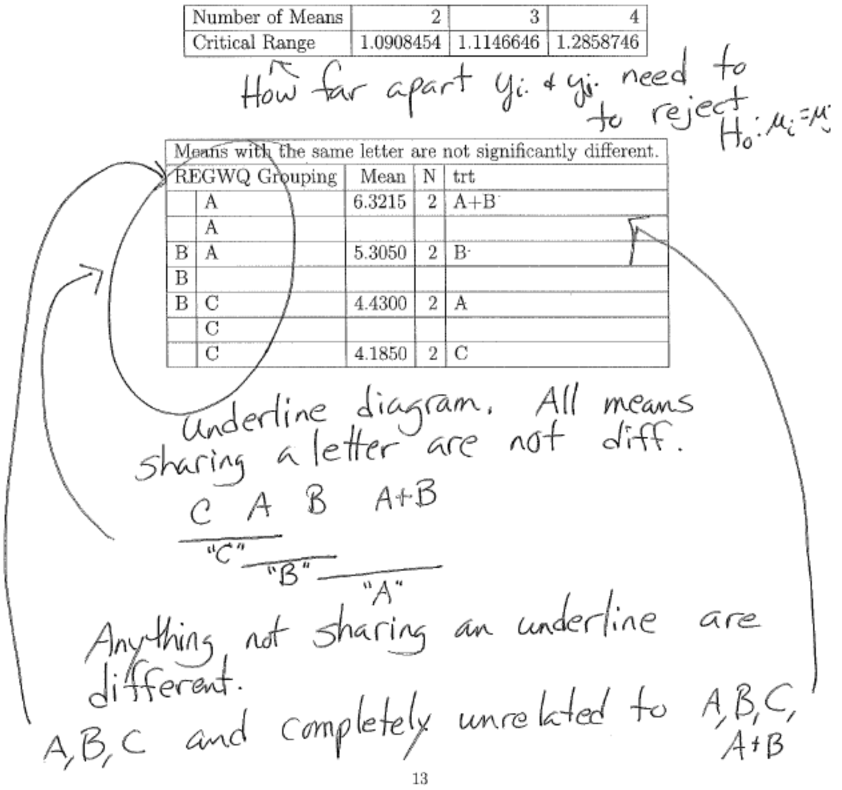
Represents when $\mu_i = \mu_j$

Segments crossing dotted line are intervals that contain 0.

↳ μ_C & μ_{A+B} are diffent since line segment doesn't touch diagonal.
μ_C & μ_A are not different.

- Additionally, you are also given a table where the t-statistics and accompanying p-values for $H_0 : \mu_A = \mu_{AB}$ are given.
  - You can use this table to figure out which pairs are signifcantly different according to the t-statistics.
- **Ryan-Einot-Gabriel-Welsch Range Test**: step-down procedure where we
  1. Order sampel means from small to large $\bar{y}_1, \ldots, \bar{y}_g$
  2. Test ranges starting with the largest $H_0 : \mu_1 = \mu_g$
  3. If we fail to reject, stop. None of the means differ.
  4. If we do reject, step down and test the next smallest range.
  5. Continue recursively. If we fail to reject, don't test the subranges.

In SAS, the output appears like:



Number of Means	2	3	4
Critical Range	1.0908454	1.1146646	1.2858746

How far apart $y_{i\cdot}$ & $y_{j\cdot}$ need to to reject $H_0: \mu_i = \mu_j$

Means with the same letter are not significantly different.

REGWQ Grouping		Mean	N	trt
	A	6.3215	2	A+B
	A			
B	A	5.3050	2	B
B				
B	C	4.4300	2	A
	C			
	C	4.1850	2	C

Underline diagram. All means sharing a letter are not diff.

C A B A+B

"C" "B" "A"

Anything not sharing an underline are different.

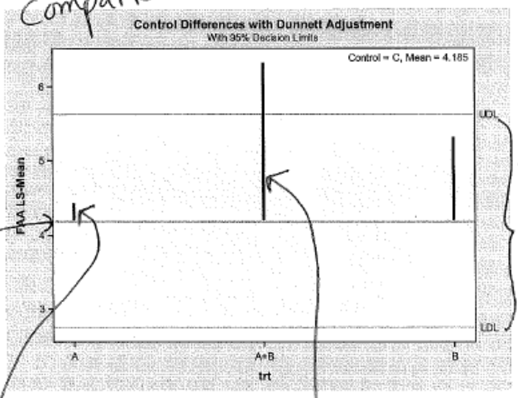
A, B, C and completely unrelated to A, B, C, A+B

13

- **Dunnett's Procedure**: Used for when we want to compare all means to a single mean, for example, a control.
  - The inner workings of this procedure aren't really touched on here, but the SAS output will contain p-values for $H_0 : \mu_i = \mu_{control}$, individual confidence intervals, and simultaneous confidence intervals.

4/19/2019 p. 7

Each vertical line is one Comparison



$\bar{y}_{control}$ = 4.185

Interval in which another mean does not differ from control.

$H_0: \mu_A = \mu_c$

$\bar{y}_A = 4.43$

$H_0: \mu_{A+B} = \mu_c$

Reject H_0 since line extends above UDL.

In the above example, treatment 2 (A+B) was seen to be significantly different from the control. The corresponding simultaneous confidence interval does not contain 0 and the chart above has the line associated with A+B extending being the upper difference limit.

- **Multiple Comparisons with the Best/Worst (MCB, MCW)**: identifies either the max or min μ_i, creates simultaneous confidence intervals for $\mu_i - \mu_{worst}$.
  - These intervals either properly contain 0 indicating that μ_i is not significantly different from the best/worst or 0 is an endpoint, indicating that μ_i is statistically different.

Diagnostics (April 22nd)

Residuals are useful for checking model assumptions. Residuals are defined in CRD as

$$r_{ij} = y_{ij} - \bar{y}_{i.}$$

and more generally:

$$r_{ij} = y_{ij} - \hat{y}_{ij}$$

Residuals are our best guess of the errors ϵ_{ij}

There are two other kinds of residuals:

- Internally studentized residuals $\frac{r_{ij}}{\sqrt{MSE(1-H_{kk})}}$ where H_{kk} reers to the diagonal element of the Hat matrix related to y_{ij}
- Externally studentized residuals $\frac{r_{ij}}{\sqrt{MSE_{(ij)}(1-H_{kk})}}$ with y_{ij} removed.

We usually use diagnostic plots instead of formal hypothesis tests to check assumptions. Hypothesis tests are generally low power.

Effects of nonnormality on σ^2:

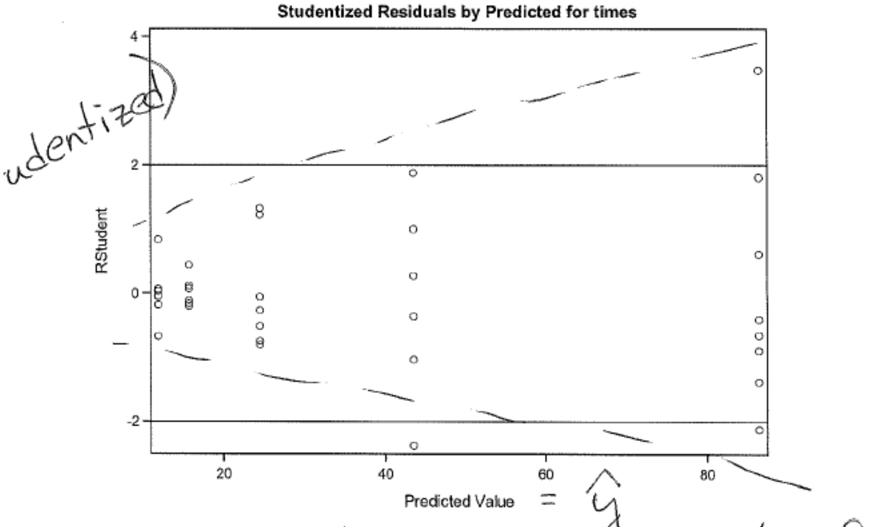
- Long tails mean too many extreme observations. Our $\hat{sigma}^2$ will be bigger than it should be, thus we don't reject as often as we should (lose power)
- Short tails mean too many type I errors (reject too often). In most settings this is the worse of the two errors.

For reasonable sample sizes and symmetric distributions, our procedures that we have discussed produce reliable inferences.

Remedies for nonnormality:

- Transformations
- Larger sample
- Different models, such as Poisson regression

Checking equal variance can either be done with Levene's test on the residuals minus the medians or using residual plots.



σ^2 represented by vertical spread of residuals. Suggests trying a log transformation (but don't log data with 0's or negative values).

The effects of nonconstant variance:

- Balanced data and not-too-different variances don't affect the t and F tests that much - they are still reliable
- Unbalanced data with large n_i and large variance mean that $\hat{\sigma}^2$ is too large so inference will be conservative

- Unbalanced data with small n_i and large variance mean that $\hat{\sigma}^2$ is too small so inference will be anti-conservative

Remedies for Non-constant variance include:

- log transformations
- Welch's ANOVA
- Weighted least squares
- More data or a more balanced design

Independence

We have several methods to check independence of different kinds

- Durbin-Watson statistic for checking serial dependence
- Residual plot vs time for serial dependence (clusters of positive and negative residuals suggest time dependence)
- Variogram plot for checking spatial dependence

Effect of non independence:

First, recall that $var(y_i + y_j) = var(y_i) + var(y_j) + 2cov(y_i, y_j)$

- Positive dependence means that $\hat{\sigma}^2$ is too small. There will be too many type I errors and inference is anti-conservative
- Negative dependence means that $\hat{\sigma}^2$ is too large. There will be too many type II errors and inference is conservative

Be careful of confounding treatments and dependence mechanisms (e.g. having two treatments administered by two different lab techs - can't separate the effect of the tech from the effect of the treatment)

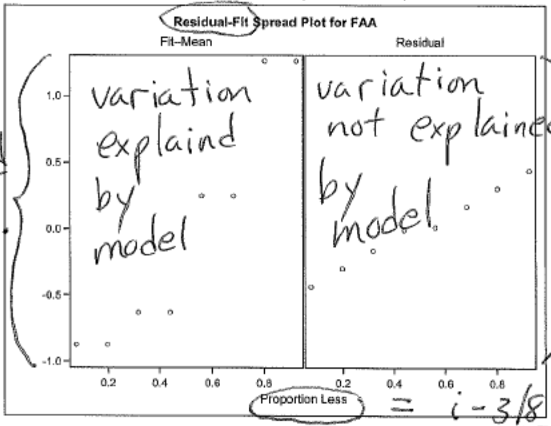
Remedies of non-independence:

- Generalized least squares
- Incorporate the dependence into the model (for example, random effects)

We can us R-F plots to diagnose model fit

Diagnostic Plot for Model Fit

Cheese data (example 5.5): R-F Plot

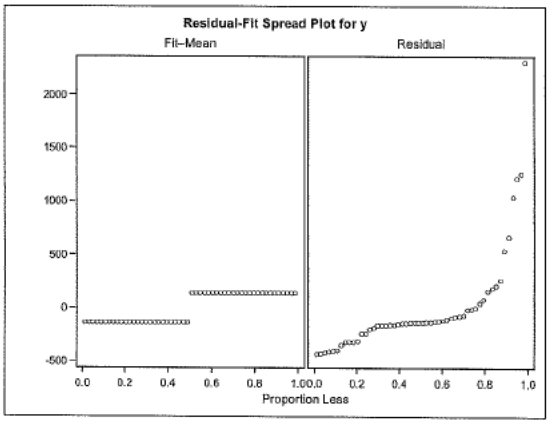


Sorted $\hat{y}_{ij} - \bar{y}_{..}$

Sorted residuals $y_{ij} - \hat{y}_{ij}$

Proportion Less $= \frac{i - 3/8}{N + 1/2}$ i = index in sorted order

Cloud-seeding experiment (example 6.1):



Not much variation explained by model.

Compare vertical spread on left & right.

Power and sample size

The sample size for a two sample t-test $H_A : \mu_1 \neq \mu_2$ can be found using this formula:

$$n \geq 2(Z_{\alpha/2} + Z_\beta)^2 \frac{\sigma^2}{\delta^2}$$

where

- n = sample size in each group
- α = Type I error rate (goes up, n goes down)
- β = Type II error rate (goes up, n goes down)
- $Z_{\alpha/2}$ = standard normal quantile related to rejection rate
- Z_β = standard normal quantile related to type II error
- σ^2 = Common variance (goes up, n goes up)
- δ = effect size $(\mu_1 - \mu_2)$ (goes down, n goes up)

Note that this depends on the distribution of $\bar{y}_{1.} - \bar{y}_{2.}$ when $H_A : \delta > 0$ is true

The sample size for a one-way ANOVA depends on the distribution of the test statistic:

- when $H_0 : \mu_1 = \ldots = \mu_g$ is true, $Fstat \sim F_{g-1,N-g}$
- when $H_A : \mu_1 = \ldots = \mu_g$ is false, $Fstat \sim F_{\zeta,g-1,N-g}$

where the non-centrality parameter ζ is found using:

$$\zeta = \frac{\sum_{i=1}^{g} n_i(\mu_i - \mu)^2}{\sigma^2} \text{ where } \mu = \frac{\sum_{i=1}^{g} \mu_i}{g}$$

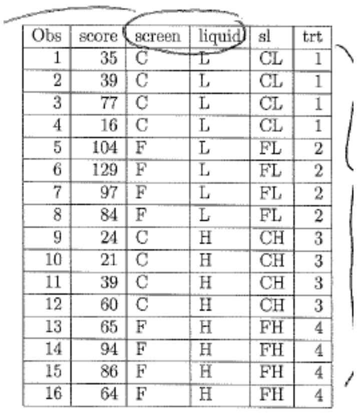
This non-central F-distribution helps quantify how false H_0 is.

To find the sample size, we can write power as the probability we reject a false null hypothesis: $1 - \beta = P(Fstat > F_{\alpha,g-1,N-g})$ where Fstat is the formula for the noncentral F statistic and then we solve for N by multiplying n by g

Usually for more complicated tests we have to simulate under different sample sizes and check the type I and type II error.

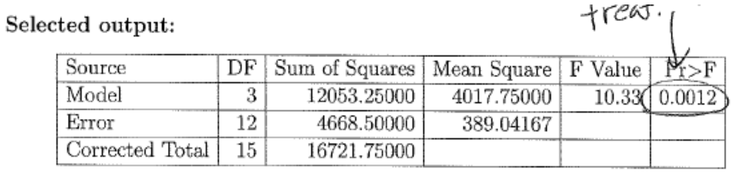
Factorial Designs

Suppose we have the following data from a 2x2 factorial design (with two factors, two levels apiece):



Obs	score	screen	liquid	sl	trt
1	35	C	L	CL	1
2	39	C	L	CL	1
3	77	C	L	CL	1
4	16	C	L	CL	1
5	104	F	L	FL	2
6	129	F	L	FL	2
7	97	F	L	FL	2
8	84	F	L	FL	2
9	24	C	H	CH	3
10	21	C	H	CH	3
11	39	C	H	CH	3
12	60	C	H	CH	3
13	65	F	H	FH	4
14	94	F	H	FH	4
15	86	F	H	FH	4
16	64	F	H	FH	4

Since we have 4 total treatments, we can start off by using a standard ANOVA F test to see if there is an overall difference



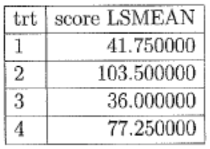
Selected output:

treas.

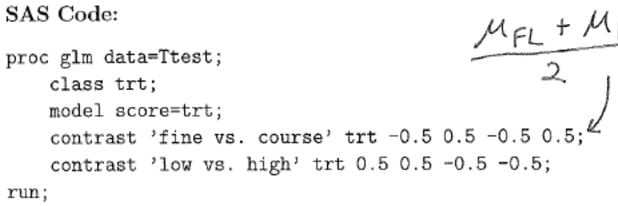
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	3	12053.25000	4017.75000	10.33	0.0012
Error	12	4668.50000	389.04167		
Corrected Total	15	16721.75000			

We see that there is in fact at least one treatment that is different

We can now use contrasts to answer research questions. Given that we have the following from the `LSMEANS` statement, we can make these contrasts:



trt	score LSMEAN	
1	41.750000	CL
2	103.500000	FL
3	36.000000	CH.
4	77.250000	FH

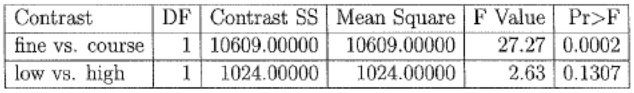


SAS Code:

$$\frac{\mu_{FL} + \mu_{FH}}{2} - \frac{\mu_{CL} + \mu_{CH}}{2}$$

```
proc glm data=Ttest;
    class trt;
    model score=trt;
    contrast 'fine vs. course' trt -0.5 0.5 -0.5 0.5;
    contrast 'low vs. high' trt 0.5 0.5 -0.5 -0.5;
run;
```

We can see that we are interested in testing fine and coarse as well as low vs high. The resulting p-values for these F tests are:



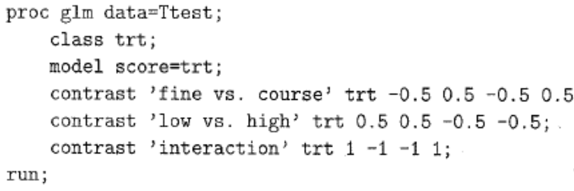
Contrast	DF	Contrast SS	Mean Square	F Value	Pr>F
fine vs. course	1	10609.00000	10609.00000	27.27	0.0002
low vs. high	1	1024.00000	1024.00000	2.63	0.1307

So we can say that we have strong evidence that mean palatability is related to screen size and no evidence liquid level is associated with mean palatability. Note that there is no Bonferroni correction here, but we could use one!

Can we make one more contrast that is orthogonal to the others?

Yes, we can compare the effect of liquid in coarse to the effect of liquid in fine. This can be notated as $C_3 = (\mu_{CL} - \mu_{CH}) - (\mu_{FL} - \mu_{FH})$. This contrast is looking at the interaction of screen and liquid (effect of liquid **depends** on screen)

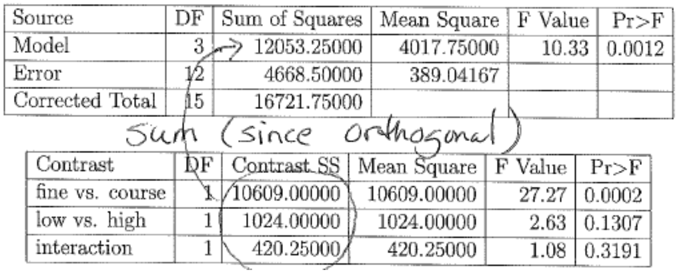
This can be coded in SAS as:



```
proc glm data=Ttest;
    class trt;
    model score=trt;
    contrast 'fine vs. course' trt -0.5 0.5 -0.5 0.5;
    contrast 'low vs. high' trt 0.5 0.5 -0.5 -0.5;
    contrast 'interaction' trt 1 -1 -1 1;
run;
```

All orthogonal (balanced data)

Output:



Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	3	12053.25000	4017.75000	10.33	0.0012
Error	12	4668.50000	389.04167		
Corrected Total	15	16721.75000			

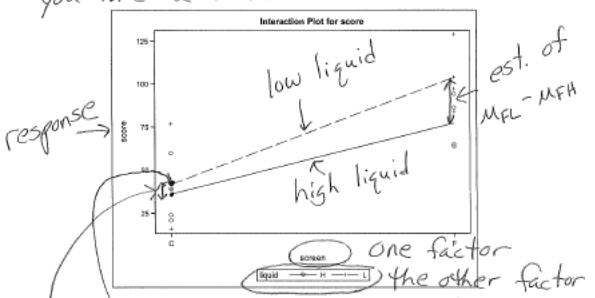
Sum (since orthogonal)

Contrast	DF	Contrast SS	Mean Square	F Value	Pr>F
fine vs. course	1	10609.00000	10609.00000	27.27	0.0002
low vs. high	1	1024.00000	1024.00000	2.63	0.1307
interaction	1	420.25000	420.25000	1.08	0.3191

We can see that we have no evidence to say that the effect of liquid on mean palatability depends on level of screen (no interaction).

We can see this visually with an interaction plot:

PROC GLM gives this plot when you have 2 factors 4/26/2019 p.7



Interaction contrast is estimated as

$\hat{C}_3 = (\hat{\mu}_{CL} - \hat{\mu}_{CH}) - (\hat{\mu}_{FL} - \hat{\mu}_{FH})$

$\bar{y}_{CL}$ = observed treat mean

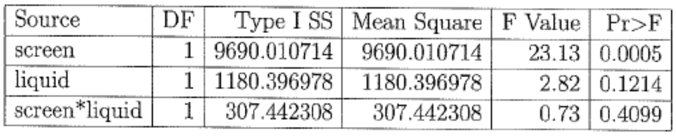
dist. here is estimate of $\mu_{CL} - \mu_{CH}$

If $C_3 = 0$, lines are parallel, after accounting for random variation in data.

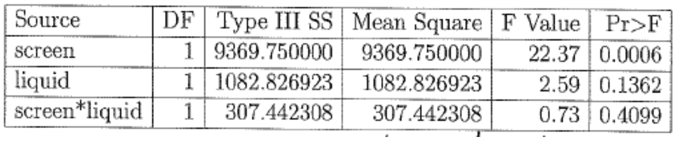
Here, we know p-value = 0.3191 for test of $H_0: C_3 = 0$. So true lines are parallel.

What if we had **unbalanced data**?

- When data are not balanced, type I and III sums of squares differ (except for the last row, which relates to the highest degree interaction term)
- Things are tidier when data is balanced
- Now contrasts are orthogonal if $\sum_{i=1}^{g} \frac{w_i w_i^*}{n_i} = 0$
  - Contrast Sum of squares agree with type III sum of squares since the same pairs of models are being compared. See below:

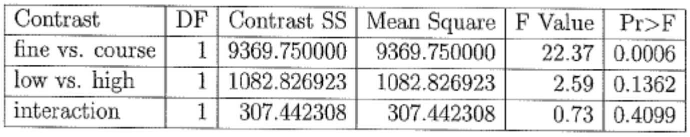


Source	DF	Type I SS	Mean Square	F Value	Pr>F
screen	1	9690.010714	9690.010714	23.13	0.0005
liquid	1	1180.396978	1180.396978	2.82	0.1214
screen*liquid	1	307.442308	307.442308	0.73	0.4099



Source	DF	Type III SS	Mean Square	F Value	Pr>F
screen	1	9369.750000	9369.750000	22.37	0.0006
liquid	1	1082.826923	1082.826923	2.59	0.1362
screen*liquid	1	307.442308	307.442308	0.73	0.4099

compared with



Contrast	DF	Contrast SS	Mean Square	F Value	Pr>F
fine vs. course	1	9369.750000	9369.750000	22.37	0.0006
low vs. high	1	1082.826923	1082.826923	2.59	0.1362
interaction	1	307.442308	307.442308	0.73	0.4099

Factor effects parameterization for factorial designs

Note that as is with every other parameterization, y_{ijk}, ϵ_{ijk} must be the same for both parameterizations and both parameterizations must have the same number of mean parameteres.

If we wanted to make a factor effects parameterization for our 2x2 design, we have to make sure it has the same number of parameters as the cell means parameterization. We can do this with the following parameterization:

$$y_{ijk} = \overset{\text{overall mean}}{\mu} + \underset{\text{effect of } i^{\text{th}} \text{ level of A}}{\alpha_i} + \overset{\text{effect of } j^{\text{th}} \text{ level of B}}{\beta_j} + \underset{\text{interaction of } i^{\text{th}} \text{ level of A \& } j^{\text{th}} \text{ level of B}}{(\alpha\beta)_{ij}} + \overset{\text{random error}}{\epsilon_{ijk}}$$

Important note! We can only have four parameters, so how are we going to incorporate the different factor levels in our new parameterizaiton? We can do so with a sum to zero constraint:

$$0 = \sum_{i=1}^{2} \alpha_i = \sum_{j=1}^{2} \beta_j = \sum_{i=1}^{2} (\alpha\beta)_{ij} = \sum_{j=1}^{2} (\alpha\beta)_{ij}$$

These constraints give us

$$\mu, \quad \alpha_1, \quad \beta_1, \quad (\alpha\beta)_{11}$$

$$\alpha_2 = -\alpha_1 \quad \beta_2 = -\beta_1 \quad \begin{aligned} (\alpha\beta)_{12} &= -(\alpha\beta)_{11} \\ (\alpha\beta)_{21} &= -(\alpha\beta)_{11} \\ (\alpha\beta)_{22} &= (\alpha\beta)_{11} \end{aligned}$$

Which allows us to calculate the estimated cell means!

$$\begin{aligned} \mu_{11} &= \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} \\ \mu_{12} = \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} &= \mu + \alpha_1 - \beta_1 - (\alpha\beta)_{11} \\ \mu_{21} = \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} &= \mu - \alpha_1 + \beta_1 - (\alpha\beta)_{11} \\ \mu_{22} = \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} &= \mu - \alpha_1 - \beta_1 + (\alpha\beta)_{11} \end{aligned}$$

This can be represented as a design matrix where all of the entries are either $\pm 1, 0$ and the parameter vector is $[\mu, \alpha_1, \beta_1, (\alpha\beta)_{11}]$. The rows will repeat depending upon the number of replicates for each treatment group.

More generally, suppose factor A has a levels and factor B has b levels. Then the cell means parameterization has ab parameters, therefore our factor effects parameterization will have the following form:

$$\begin{array}{ccccccc} 0 = \sum_{i=1}^{a} \alpha_i & = & \sum_{j=1}^{b} \beta_j & = & \sum_{i=1}^{a} (\alpha\beta)_{ij} & = & \sum_{j=1}^{b} (\alpha\beta)_{ij} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ a-1 & & b-1 & & (a-1)\,(\alpha\beta)_{ij} & & (b-1)(\alpha\beta)_{ij} \\ & & & & \text{for each } j & & \text{for each } i \end{array}$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{\substack{(a-1)(b-1)\ (\alpha\beta)\text{'s} \\ \text{(not completely obvious)}}}$$

Count parameters

$$(a-1) + (b-1) + (a-1)(b-1) + \underset{\underset{\mu}{\nearrow}}{1} = ab$$

Parameter estimates in factor effects parameters

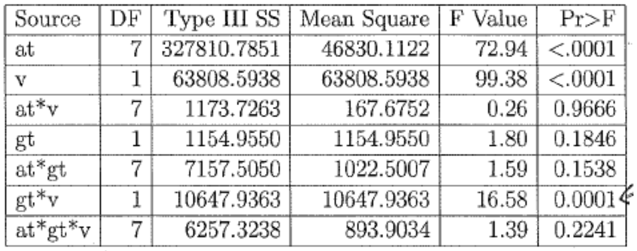
$$\begin{aligned} \hat{\mu} &= \bar{y}_{...} \\ \hat{\alpha}_i &= \hat{\mu}_{i.} - \hat{\mu} = \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\beta}_j &= \hat{\mu}_{.j} - \hat{\mu} = \bar{y}_{.j.} - \bar{y}_{...} \\ (\hat{\alpha\beta})_{ij} &= \hat{\mu}_{ij} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\mu} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \end{aligned}$$

More on Factorial Designs (May 1)

In this example, we have three factors: analysis temperature (8 levels), growth temperature (2 levels), and variety (2 levels).

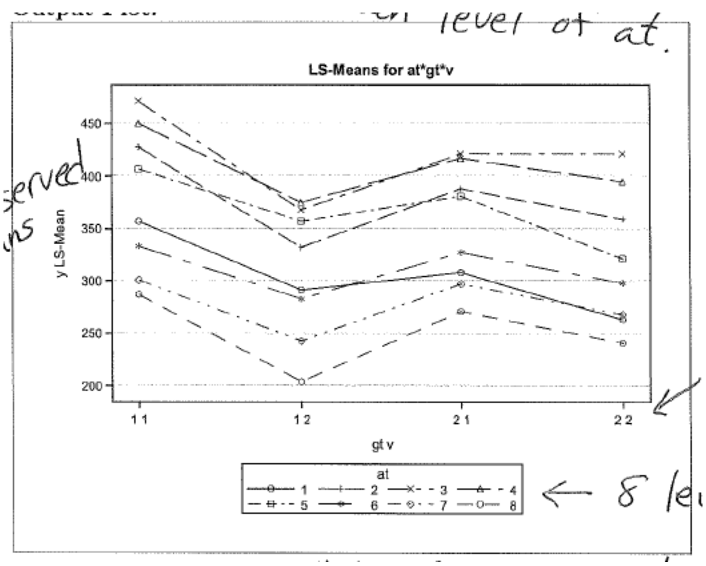
Our data is balanced so the type I and type III SS agree.

We get the following table:



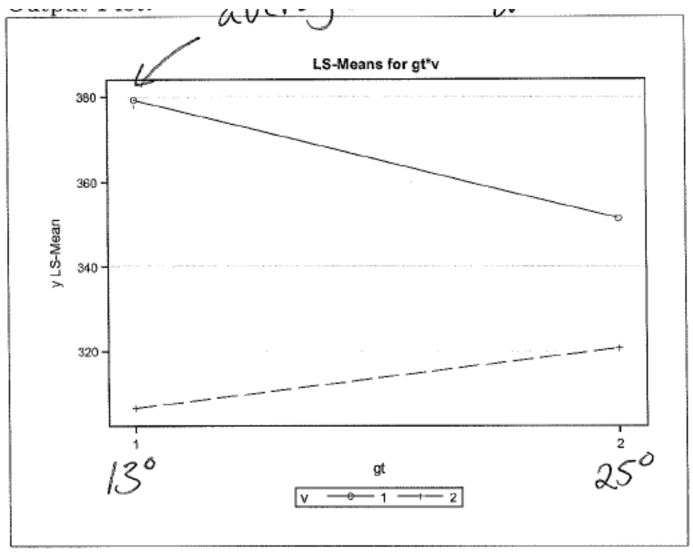
Source	DF	Type III SS	Mean Square	F Value	Pr>F
at	7	327810.7851	46830.1122	72.94	<.0001
v	1	63808.5938	63808.5938	99.38	<.0001
at*v	7	1173.7263	167.6752	0.26	0.9666
gt	1	1154.9550	1154.9550	1.80	0.1846
at*gt	7	7157.5050	1022.5007	1.59	0.1538
gt*v	1	10647.9363	10647.9363	16.58	0.0001 ←
at*gt*v	7	6257.3238	893.9034	1.39	0.2241

There is a complication here. There is a significant two way interaction term between growth temperature and variety. We can examine these interaction terms using interaction plots:



In this plot, the observed means are plotted with the combinations of v + gt listed at the x-axis and sliced by at. Notice here that the lines are approximately parallel, consistent with our large 3 way interaction term's p-value. The interpretation here is that all 2 way interactions are the same for all levels of the third factor.

If we examine the other interaction plots, they all also look pretty parallel, but there is one that stands out as being kind of interesting (and funnily enough, it corresponds to a low p-value)



Our two variables here are growth temperature (x axis) and sliced by variety. Here we can interpret this as variety 1 having a larger mean amylase activity than variety 2 for both growth temperatures, but at 13C the difference in mean amylase activity is larger than at 25C.

Note that if we were to average over variety, the means for the two growth temperatures would be in bout the same place (340ish). We should be careful of interpreting main effects if interactions exist.

We should formalize the model and the notation. Our parameterization will be:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

where:

- μ is the overall mean

- α_i is the effect of the *ith* level of analysis temperature
- β_j is the effect of the *jth* level of variety
- γ_k is the effect of the *kth* level of growth temperature
- $(\alpha\beta)_{ij}$ is the interaction effect of the *ith* level of at and *jth* level of v
- $(\alpha\gamma)_{ik}$ is the interaction effect of the *ith* level of at and *kth* level of gt
- $(\beta\gamma)_{jk}$ is the interaction effect of the *jth* variety and *kth* level of gt
- $(\alpha\beta\gamma)_{ijk}$ is the three way interaction effect of the *ith* at, *jth* v, and *kth* gt
- y_{ijkl} is the amylase activity for the *lth* replicate at the *ith* level of at, the *jth* level of v, and the *kth* level of gt
- ϵ_{ijkl} is the random error associated with the *ijkl*th replicate

The normal assumptions of iid normal errors carries over. Note that for the constraints, we have to make sure all of the interaction terms sum to 0 no matter which indices sum over!

For testing the interactions:

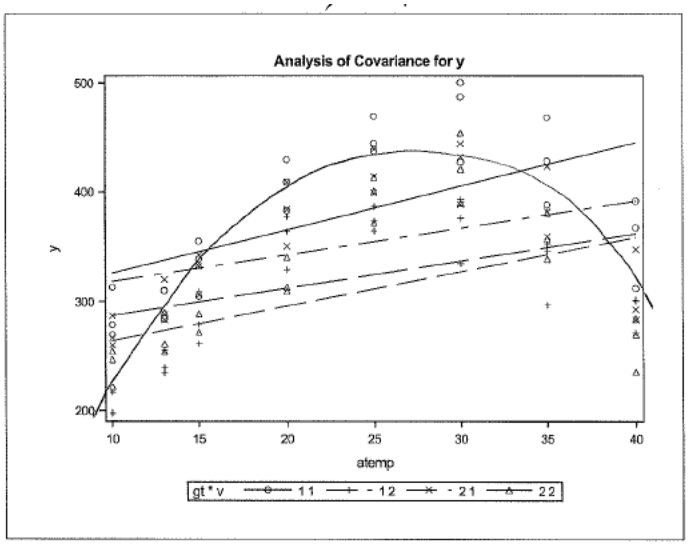
- For $H_0 : (\alpha\beta\gamma)_{ijk} = 0$, the full model will be the saturated model above and the reduced model will be the same only with the three way interaction model taken out.
  - In this example, we get F = 1.39, p = .2241 so there is no evidence of 3 way interactions
- For $H_0 : (\beta\gamma)_{jk} = 0$, we typically do Type III SS tests so we only remove this particular term in the reduced model. The full model remains the saturated model.
  - In this example, we get an F-statistic of 16.58, pvalue of .0001 so we conclude that the effect of variety is different at the two growth temperatures (alternative phrasing: effect of gt is different for different varieties)

Modeling response as a linear or quadratic function of analysis temperature

This is usually for reducing the model degrees of freedom and adding error degrees of freedom. This usually gives us a lower MSE which is a good thing. However, in our example, since the linear model doesn't fit very well, we end up having a much larger MSE than we did in the previous factor effects model.

We could in this scenario do a lack of linear fit test to compare to the previous model.

We might hypothesize that quadratic fit might be better, especially after we see this plot:



We can fit this in SAS. Using this model adds a bunch of extra parameters, but if the fit is better than we will reduce our sum of squares overall. We can also choose to use a lack of quadratic fit test in which we use the initial model to estimate the difference in means between the 2 varieties at each growth temperature.

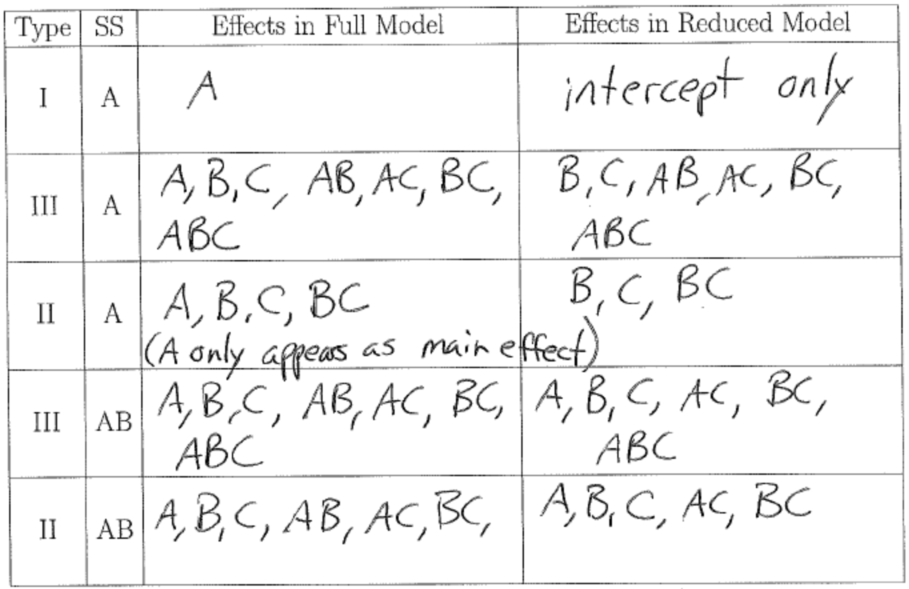
Unbalanced data (May 6)

Be careful when averaging across rows and columns with unbalanced data. Weights are different in different cells. We can still do sum of squares F tests though because SSE is calculated over all observations so the effect of the different weights doesn't appear there.

When data are unbalanced, *type I and type III sums of squares are different.* We can use type II sum of squares.

Type II sum of squares are *hierarchical* so the reduced model doesn't have higher order terms where lower order terms are missing.

Consider the following table:



Type	SS	Effects in Full Model	Effects in Reduced Model
I	A	A	intercept only
III	A	A, B, C, AB, AC, BC, ABC	B, C, AB, AC, BC, ABC
II	A	A, B, C, BC (A only appears as main effect)	B, C, BC
III	AB	A, B, C, AB, AC, BC, ABC	A, B, C, AC, BC, ABC
II	AB	A, B, C, AB, AC, BC,	A, B, C, AC, BC

Missing cells

Consider the possibilty that in a 2x3 factorial design we have lost all the data in one of our 6 cells.

We have therefore lost the factorial structure, and must use *cell means parameterization.*

Random Effects

A model with just one random effect will look like

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

where $\epsilon_{ij} \sim N(0, \sigma^2)$ and $\alpha_{ij} \sim N(0, \sigma^2_\alpha)$ and the ϵ and α are independent from each other.

Then we have

$$\begin{aligned} var(y_{ij}) &= var(\mu + \alpha_i + \epsilon_{ij}) = \sigma^2_\alpha + \sigma^2 \\ cov(y_{11}, y_{12}) &= cov(\alpha_1, \alpha_1) + cov(\alpha_1, \epsilon_{12}) + cov(\epsilon_{11}, \alpha_1) + cov(\epsilon_{11}, \epsilon_{12}) = \sigma^2_\alpha \end{aligned}$$

When should we consider *using random effects*?

- Need to model dependence among observations from the same level of a factor
  - Strength of boxes made by same machine are not independent (note that model only allows positive dependence)
- Levels of the factor are considered a sample from a larger population
  - Machines are random sample from population
- Repeating experiment would use different factor levels
  - Doing the experiment again would use different levels for the machine

Inferences for a random effects model

- Our point estimate $\hat{\mu} = \bar{y}_{..}$
- Standard error is based on the variance of the point estimate (note that $var(\bar{y}) \neq \frac{\sigma^2}{N}$ since its not a mean of independent observations):

$$
\begin{aligned}
var(\bar{y}_{..}) &= var(\frac{1}{N}\sum_i \sum_j y_{ij}) \\
&= \frac{1}{N^2} var(\sum_i \sum_j (\alpha_i + \epsilon_{ij})) \\
&= \frac{1}{N^2}\left[var(\sum_i \sum_k \alpha_i) var(\sum_i \sum_j \epsilon_{ij}) \right] \\
&= \frac{1}{N^2}\left[var(\sum_i n\alpha_i) an\sigma^2 \right] \\
&= \frac{1}{N^2}\left[n^2 a\sigma_\alpha^2 + an\sigma^2 \right] \\
&= \frac{n\sigma_\alpha^2 + \sigma^2}{N}
\end{aligned}
$$

This last point comes from the fact that $N = an$

The *intraclass correlation coefficient* is calculated as:

$$
\rho = \frac{cov(y_{ij}, y_{ij'})}{\sqrt{var(y_{ij})var(y_{ij'})}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}
$$

$\rho \approx 0$ means the observation nearly independent $\rho \approx 1$ means the observation very dependent

How can we predict α_i? Note that this is not an "estimate" since α_i is random.

Going back to our example, does box strength depend on machine? This is easy to test with a fixed effect, we just set our null hypothesis $H_0 : \alpha_i = 0$

However, in a random effects model, α_i is random so this is not an appropriate H_0 since our null hypothesis must be about parameters, not random variables. Instead, we can test the variance of the α_i so our null hypothesis will be $H_0 : \sigma_\alpha^2 = 0$ vs $H_A : \sigma_\alpha^2 > 0$. If the null is true, then there is no machine variation.

The ANOVA for the random effects is the same as the ANOVA for the one fixed-factor design.

Using **A** as notation for the model, Sum of squares of A (SSA) is $\sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2$ and dividing this by a -1 degrees of freedom gives us the mean square of A (MSA). WE then divide MSA by MSE to get our F statistic. Under $H_0 : \sigma_\alpha^2 = 0, y_{ij} \sim N(\mu, \sigma^2)$ which is the same as under fixed effects.

The expected mean squares (EMS) of these values is:

- E(MSA) $= \sigma^2 + n\sigma^2_\alpha$ with a - 1 degrees of freedom
- E(MSE) $= \sigma^2$ with N - a degrees of freedom
- SE($\bar{y}_{..}$) $= \sqrt{v\hat{a}r(\bar{y}_{..})} = \sqrt{\frac{MSA}{N}}$ since $\frac{MSA}{N}$ is an unbiased estiamte of var($\bar{y}_{..}$)

Recall that for one fixed factor and balanced data,

- E(MSE) $= \sigma^2$
- E($MS_{treatment}$) $= \sigma^2 + \frac{n}{g-1}\sum_{i=1}^{g}(\mu_i - \mu)^2 = \sigma^2 + \frac{n}{g-1}\sum_{i=1}^{g}(\alpha_i)^2$

If we wanted to test whether or not the $\alpha_i = 0$, we would put the treatment sum of squares over the MSE to make our F-statistic. Under the null hypothesis, the treatment sum of squares and the MSE have the same expectation. Under the alternative, $E(MS_{treatment}) > E(MSE)$, so we reject the null if we have a large F statistic.

The F statistic for one random factor and balanced data is very similar and follows the same logic, only with MSA instead of $MS_{treatment}$

Two random factors

We can also create a model with more than one random factor. Going back to our box example, say that we wanted to also model the machine operators as random effects. We could create a model like

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where α_is represent the machine effect, β_js represent the operator effect, and $(\alpha\beta)_{ij}$ represent the interaction between the machine and the operator.

Our assumptions for this model are that

$$\begin{aligned}\alpha_i &\sim N(0, \sigma^2_\alpha)\\ \beta &\sim N(0, \sigma^2_\beta)\\ (\alpha\beta)_{ij} &\sim N(0, \sigma^2_{\alpha\beta})\\ \epsilon_{ijk} &\sim N(0, \sigma^2)\end{aligned}$$

where the model parameters are now $\mu, \sigma^2_\alpha, \sigma^2_\beta, \sigma^2_{\alpha\beta}, \sigma^2$ and the latter four constitute the variance components.

If we further examine the sum of squares, we see that we have

- A with a - 1 degrees of freedom and $E(MSA) = \sigma^2 + n\sigma^2_{\alpha\beta} + nb\sigma^2_\alpha$
- B with a - 1 degrees of freedom and $E(MSB) = \sigma^2 + n\sigma^2_{\alpha\beta} + nb\sigma^2_\beta$
- AB with (a-1)(b-1) degrees of freedom and $E(MSAB) = \sigma^2 + n\sigma^2_{\alpha\beta}$
- error with ab(n-1) degrees of freedom and $E(MSE) = \sigma^2$

If we wanted to test whether $H_0 : \sigma^2_\alpha = 0$, then the numerator of the F-statistic would be MSA (mean square for the effect of interest) and the denominator would be MSAB (since E(MSAB) = E(MSA) under the null hypothesis).

When estimating variance components, make sure to use REML instead of MOM. REML is the method used by default in `PROC MIXED`.

```
PROC MIXED data=Exp11_2;
class m o;
model y=;
random m|o;
run;
```

If we specify method=type3, then we get MOM estimators

```
PROC MIXED data=Exp11_2 method=type3;
class m o;
model y=;
random m|o;
run;
```

The issue is that MOM estimators can fall outside of the paramter space whereas REML estimators are guaranteed to at least be in the parameter space.

DO NOT use `PROC GLM` for random effects since the F statistics there all use MSE as the denominator. The df, type I SS, and MS for all the sources will be the same, so be very careful.

Three random effects

Suppose now we have 10 random machines, 10 random operators, and 2 batches of glue. Each operator produces 4 boxes on each machine with 2 boxes per glue batch.

Our random effects here are the machine, the operator, and the batch of glue.

Our model will look like:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

where α_i denotes the machine effect, β_j denotes the operator effect, γ_k denotes the glue effect.

The assumptions are that $\alpha_i, \beta_j, \gamma_k$, the 2/3 way interactions, and ϵ_{ijkl} are all indepenent and normally distributed around 0 with variances $\sigma^2_\alpha, \sigma^2_\beta, \dots$ (8 variance components all together).

Analyzing this data in SAS can be done using

```
PROC MIXED data=Exp11_2 method=type3;
class m o g;
model y=;
random m|o|g;
run;
```

where `m|o|g` denotes the saturated model that we had above.

To test if the machine effect is nonzero, we need to test the hypothesis

$$H_0 : \sigma^2_\alpha = 0$$
$$H_A : \sigma^2_\alpha > 0$$

Our numerator is the MSA, but the denominator has to be an MS where the expected MS is the same as MSA under the H_0.

There might not be MS with this expected value, so we need to construct an MS with this expected value. We can do that by making a linear combination of MS:

$$MS_{denominator} = \sum_s g_s MS_s$$
$$E[\sum_s g_s MS_s] = E_{H_0}[MS(\text{effect being tested})]$$

where the denominator df will be

$$\nu^* = \frac{(\sum_s g_s MS_s)^2}{\sum_s g_s^2 MS_s^2/\nu_s^2}$$

Where ν_s is the df for MS_s.

We can get the interval estimates for the variance components using SAS but keep in mind that this is not generally a good idea since the CIs are asymptotic so you need a large sample.

```
PROC MIXED data=Exp11_2 method=type3 cl covtest;
class m o g;
model y=;
random m|o|g;
run;

/* or using the REML estimates */

PROC MIXED data=Exp11_2 cl covtest;
class m o g;
model y=;
random m|o|g;
run;
```

Crossed and Nested Effects (Chapter 12, Outline 8)

Let's take the example of machines making boxes again, this time just focusing on the machine and the operator.

Suppose we have two machines and we want to compare them. A repeat of the experiment would use the same two machines, but operators are randomly selected.

A **crossed design** is where each operator sees both machines and each machine sees all operators. There are 4 operators total. The notation for each of the 8 possible treatments combinations is $A_1B_1, A_1B_2, \ldots, A_2B_1, A_2B_2, \ldots$

The model for a **crossed design** has the form

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where the strength of the *kth* box from the *ith* machine made by the *jth* operator is modeled by the overall mean, the machine effect which is fixed, the operator effect β_j which is random, $(\alpha\beta)_{ij}$ the random interaction between the machine and the operator, and the random error ϵ_{ijk}.

As per usual, the assumption is that all of the random effects are indepdent and normally distributed with their own variance components.

A **nested design** is where each machine has its *own set* of operators. There are 8 operators total. The notation for each of the 8 possible treatment combinations have the form $A_1B_{1(1)}$ which can be thought of as machine 1 getting the 1st operator nested in the 1st machine.

The model for a **nested design** has the form

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$$

where everything is almost identical to the crossed design *except* that $\beta_{j(i)}$ now refers to the random effect of the *jth* operator that is nested in the *ith* machine.

Note that there is also **no interaction** in this nested design because no operator works on both machines so we can't measure the interaction between operator and machine. Nesting has *removed an interaction term.*

In SAS code, we can represent this as

```
filename Exp11_2 'exmpl11.2.csv';
data Exp11_2;
  infile Exp11_2 firstobs=2 dlm=',';
  input m o g y;
run;

data MixBox;
  set Exp11_2;
  if m <= 2;
run;

\* For A and B being crossed *\

PROC MIXED data=MixBox method=type3;
  class m o;
  model y=m;
  random o m*o;
run;

\* For A and B being nested *\

PROC MIXED data=MixBox method=type3;
  class m o;
  model y=m;
  random o(m);
run;
```

In the Type 3 Analysis of Variance tables for either of these designs, you will see a term *Q(m)* in the expected mean square for some of the sources. This is a quadratic term which shows up when you have a fixed factor.

$$Q(m) = \frac{nb\sum\alpha_i^2}{a-1}$$

To test for the machine effect in the **crossed design**, we have to use an F statistic where the denominator is a MS whose expected value is the same as MSA when $H_0 : \alpha_i = 0$.

In our case MS_{AB} fits this criteria, so our F statistic is $\frac{91.828}{13.457} = 6.824$ which we compare to an F distribution with 1 and 9 df.

In the **nested design**, the analysis is very similar. The only differnece is that the MS in the denominator is now $MS_{B(A)}$ which gives us an F statistic of $\frac{91.828}{129.458} = .709$ which we compare to an F distribution with 1 and 18 df.

Notice that the test statistic in the nested design is smaller than the test statistic in the crossed design. This won't always be the case since the denominator degrees of freedom are larger here. Generally speaking larger denominator degrees of freedom gives us *more power.* The reason they are different is because we were able to separate the variability of the operator and the operator/machine interaction in the crossed design but we were not able to do so in the nested design.

Effectively, we have $SSB + SS_{AB} = SS_{B(A)}$ where the RHS in the crossed design and the LHS is the nested design.

Final note: we do not get to pick what analysis we do - it must match the experiment.

- Why would we use nesting?
  - Feasibility or convenience (operators are in different locations)
  - Subsampling (multiple observations on a single unit) causes the obseravations to be ensted in a single experimental unit and the experimental units are then nested in the treatments.

An example of a study that has multiple levels of nesting could be a genetics study. Suppose we have male parents (level A), female parents (level B), and our response is the offspring (C) of each male and female pair. There are multiple measurements per offspring.

The females are nested within males (ex: $B_{1(1)}$) and the offspring are ensted within each pair of parents (ex: $C_{1(1,1)}$)

The model for C nested in B nested in A is

$$y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{ijkl}$$

where in this case we choose to model all the components as random so

$$\begin{aligned}
\alpha_i &\sim N(0, \sigma^2_\alpha) \\
\beta_j &\sim N(0, \sigma^2_{\beta(\alpha)}) \\
\gamma_k &\sim N(0, \sigma^2_{\gamma(\alpha\beta)}) \\
\epsilon_{ijkl} &\sim N(0, \sigma^2)
\end{aligned}$$

and all of the random terms are independent from one another.

The covariance between two observations on the same offspring is given as: $cov(y_{ijkl}, y_{ijkl'}) = \sigma^2_\alpha + \sigma^2_{\beta(\alpha)} + \sigma^2_{\gamma(\alpha\beta)}$.

The covariance between observations of two offpsring is given as: $cov(y_{ijkl}, y_{ijk'l'}) = \sigma^2_\alpha + \sigma^2_{\beta(\alpha)}$. Notice that these are the variances of the shared random effects.

We can have **crossed and nested** factors in the same experiment. For example, assume that each machine has a unique set of four operators (for 8 operators total) and each operator make four boxes, two for each batch of glue.

- The machine and operator are *nested* because each operator sees *only one* machine
- The glue and operator are *crossed* since each operator sees both batches of glue and each batch of glue sees all 8 operators.

- The machine and glue are *crossed* since each machine sees all batchs of glue

Since operator is our nested factor, we can represent this experiment using the notation $B_{1(1)}$ which denotes the first operator nested in the first machine. $B_{1(1)}$ will appear in both treatments of glue.

Our model is

$$y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{j(i)k} + \epsilon_{ijkl}$$

where the strenght of the *lth* box from the *kth* batch of glue with the *jth* operator nested in the *ith* machine, and the *ith* machine is a function of an overall mean μ, the machine effect α_i , the nested operator effect $\beta_{j(i)}$, the glue effect γ_k, the machine glue interaction and the operator glue interaction.

Include the usual assumptions for random terms and the sum-to-0 constraints for the fixed term α_i

Estimating *means and contrasts in a mixed model* can be done by deriving the variance of the sample means in the exeperiment.

Consider an experiment with two machines (fixed), ten operators (random), machines and operators are crossed, and there are n=4 boxes for each operator per machine.

To estimate the mean box strength for machine 1, we simply need to average all observations under machine 1.

To get the standard error though, we will have to derive it.

$$\begin{aligned}
var(\bar{y_{1..}}) &= var\left[\frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}\epsilon_{ijk})\right] \\
&= \frac{1}{b^2n^2}var\left[\sum_j\sum_k \beta_j + \sum_j\sum_k(\alpha\beta)_{ij} + \sum_j\sum_k \epsilon_{ijk}\right] \\
&= \frac{1}{b^2n^2}var\left[\sum_j n\beta_j + \sum_j n(\alpha\beta)_{ij} + \sum_j\sum_k \epsilon_{ijk}\right] \\
&= \frac{1}{b^2n^2}\left[n^2var[\sum_j \beta_j] + n^2var[\sum_j(\alpha\beta)_{ij}] + var[\sum_j\sum_k \epsilon_{ijk}]\right] \\
&= \frac{1}{b^2n^2}\left[n^2b\sigma^2_\beta + n^2b\sigma^2_{\alpha\beta} + nb\sigma^2\right] \\
&= \frac{n\sigma^2_\beta + n\sigma^2_\alpha + \sigma^2}{nb}
\end{aligned}$$

Standard error is then found by $SE(\bar{y_{...}}) = \sqrt{\frac{n\sigma^2_\beta + n\sigma^2_\alpha + \sigma^2}{nb}}$

In SAS, we can calculate this using the following code:

```
PROC MIXED data=MixBox;
  class m O;
  model y=m;
  random o m*o;
  lsmeans m;
run;
```

This code outputs a covariance parameter estimates table which is used in the standard error formula we derived above.

Estimating the **difference in mean strength between machines** involves calculating $\bar{y_{2..}} - \bar{y_{1..}}$.

The variance of the point estimate is $var(\bar{y}_{2..} - \bar{y}_{1..}) = \frac{2n\sigma^2_{\alpha\beta}+2\sigma^2}{nb}$. We derive it in a similar way to the above derivation of the standard error of the sample mean for machine 1.

```
PROC MIXED data=MixBox method=type3;
  class m o;
  model y=m;
  random o m*o;
  estimate 'machine' m 1 -1;
run;
```

Complete Block Designs (Chapter 13, Outline 9)

A **block** is like a group of experimental units that are linked because of similarity.

Examples can be genetic like a litter of animals, location (like a geographical region), etc.

We should use blocking to control for anything that affects the response but is not of interest. *Anything* that affects the response but is not in the model contributes to the random error term so blocking helps reduce the error variance, giving us more power and precision.

The **randomized complete block design** or RCBD is a design where experimental units are stratified into blocks. Within each block, we randomly assign units to treatments.

Note that

- Experimental units are not randomly assigned to blocks
- *Complete* block designs are those were each block contains at least one replicate of every treatment combination (for example, Latin Squares)
- In a balanced design, each block will have the same number of replicates for each treatment combination
- There is a different randomization in each block (protects us from situations where the arrangement of the treatments matter)

The structure when drawn out resembles a factorial experiment.

The model for RCBD is

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

where we focus on fixed blocks.

The assumptions are that $\sum_{i=1}^{g} \alpha_i = \sum_{j=1}^{r} \beta_j = 0$ and $\epsilon_{ij} \sim N(0, \sigma^2)$

The ANOVA table for RCBD is

Source	df
Treatments	g-1
Blocks	r-1
Error	(g-1)(r-1)
Total	gr-1

Note that if we only have one replicate in each cell, we can't include an interaction term since we can't distinguish it from ϵ_{ij}. **(The best way to check when this is the case is to look at the subscripts)**. If $n \geq 2$, we could calculate $(\alpha\beta)_{ij}$ but we probably wouldn't.

- **Note** that model includes no block by treatment interaction. +The interaction term implies that treatment effect varies from block to block. +This is not generally desirable. +We should always try to select a blocking factor that we can assume doesn't interact with the treatments.
- Blocks can either be fixed or random
  - Since we're focusing on comparing treatments, so it doesn't really matter
- For RCB designs, you need at least one replicate per cell

We can compare RCBD designs with other experimental designs using **relative efficiency**. Relative efficiency is defined as

$$E_{1:2} = \frac{I_1}{I_2} = \frac{\sigma_2^2}{\sigma_1^2}$$

where by convention I_2 is the simpler design (aka the design with more error degrees of freedom).

Let's use relative efficiency to compare RCBD with CRD.

Recall that to test $H_0 : \alpha_i = 0$, the test statistic is $F = \frac{MSA}{MSE}$.

Our relative efficiency will be given by $E_{RCBD:CRD} = \frac{\sigma^2_{CRD}}{\sigma^2_{RCBD}}$.

If we perform an RCBD in SAS, the $\hat{\sigma}_{RCBD}$ is the MSE of that model.

To get $\hat{\sigma}_{CRD}$ we need to use this formula:

$$\hat{\sigma}_{CRD} = \frac{(r-1)MS_{block} + [(g-1) + (g-1)(r-1)]MSE}{(r-1) + (g-1) + (g-1)(r-1)}$$

After we compute this, we plug into this *corrected* relative efficiency:

$$\hat{E}_{RCBD:CRD} = \frac{\hat{\sigma}^2_{CRD}}{\hat{\sigma}^2_{RCBD}} \left(\frac{(\nu_{RCBD}+1)(\nu_{CRD}+3)}{(\nu_{RCBD}+3)(\nu_{CRD}+1)} \right)$$

which correct for the number of blocks. Note that if the number of blocks is not large, then the correction factor will be close to 1.

Suppose we got a relative efficiency of 2. This tells us that the RCBD is twice as efficient as the CRD, hence we need only half as many experimental units in the RCBD vs the CRD to do inference.

$E_{RCBD:CRD}$ is not a tool to decide on a model, but rather a diagnostic.

Latin Squares Design (Outline 9)

Latin square designs are employed when we have **two** blocking factors.

A popular type of design in clinical trials is the *crossover design* where we have subjects that receive all possible treatments with a washout period in between each treatment period. *Crossovers* are Latin square designs where the subjects are the first block and the time periods are the second block.

By definition, a Latin square has g treatments and two blocking factors, each with g levels. When arranging a Latin square, there is one copy of each treatment level in each row and column.

Going back to our crossover design example, the subjects can be the columns, the time periods can be the rows, and the cells can be filled in with the drug treatment for that cell.

The model for the Latin square design is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$$

where α_i is the treatment effect, β_j is the subject effect, γ_k is the time period effect.

The constraints are that $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$ and that $\epsilon_{ijk} \sim N(0, \sigma^2)$.

Latin squares by their nature are *incomplete block designs* since each block (cell in this case) doesn't contain all levels of the treatments.

We can compare Latin squares to RCBDs again using relative efficiency where

$$\hat{E}_{LS:RCBD} = \frac{\sigma^2_{RCBD}}{\sigma^2_{LS}} \left(\frac{(\nu_{LS}+1)(\nu_{RCBD}+3)}{(\nu_{LS}+3)(\nu_{RCBD}+1)} \right)$$

and $\hat{\sigma}^2_{RCBD} = \frac{(g-1)MS_{block}+[(g-1)+(g-1)(g-2)]MSE}{2(g-1)+(g-1)(g-2)}$

Lastly, in the ANOVA for Latin square designs, we have

Source	df
Treatments	g-1
Subject	g-1
Treatment Period	g-1
Error	(g-1)(g-2)
Total	$g^2 - 1$

Some notes on the Latin Square design:

- The Latin square design is restrictive because you need the same number of levels in treatments and both blocking factors.
- If a blocking factor represents a gradient such as time, you can divide that into a convenient number of levels.
- For a moderately small level g, we can make do with very few experimental units which means we can use Latin squares for pilot studies to generate hypotheses which can be later explored in larger studies

Split Plot Designs (Outline 10)

We usually employ split-plot designs when we have logical constraints on two variables. Factor A must be applied to large plots and Factor B must be applied to smaller plots (for example, sprinklers can be factor A and varieties of corn can be factor B).

The general idea of split plot designs is that we randomly assign levels of factor A to large plots and we split each of those plots into subplots which are then randomly assigned levels of B.

This is under the assumptions that:

- The whole plots are independent of one another
- Observations within a single plot may not be independent
- Whole pots are nested in treatment A (each whole plot sees one level of A) and factor B are crossed (each whole plot sees all levels of factor B)

- Treatments A and B are crossed (each level of A sees each level of B)

The model for a split plot design is given as

$$y_{ijk} = \mu + \alpha_i + \eta_{k(i)} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{k(i)j}$$

where the response is the yield for the *kth* whole plot getting sprinkler level i and variety j, μ is the overall mean, α_i is the irrigation level, $\eta_{k(i)}$ is the whole plot error, β_j is the variety effect, $(\alpha\beta)_{ij}$ is the interaction between irrigation and variety, and ϵ_{ijk} is the subplot error.

with the following assumptions:

$$\sum \alpha_i = \sum \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$
$$\eta_{k(i)} \sim N(0, \sigma_\eta^2)$$
$$\epsilon_{ijk} \sim N(0, \sigma^2)$$

The accompanying ANOVA table is:

Source	df
A	a-1
whole plot error (1)	na - a
B	b - 1
AB	(a-1)(b-1)
Subplot error (2)	a(n-1)(b-1)

To conduct the proper hypothesis tests for the following, we can use these F statistics:

Test	Null	F statistic
A	$H_0 : \alpha_i = 0$	$\frac{MSA}{MS_{error(1)}}$
B	$H_0 : \beta_j = 0$	$\frac{MSB}{MS_{error(2)}}$
AB	$H_0 : (\alpha\beta)_{ij} = 0$	$\frac{MSAB}{MS_{error(2)}}$

Some notes about the split plot design:

- Two levels of randomization are required (whole plot and subplot)
- Degrees of freddom for subplot error is larger than for the whole plot error
- Model allows positive dependence between the subplots and the same whole plot
  - $cov(y_{ijk}, y_{ij'k}) = \sigma_\eta^2$
  - $cov(y_{ijk}, y_{ij'k}) = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma^2}$
  - Model assumes that covariance pairs of subplots of the same whole plot are equal
  - Covariance doesn't decrease with distance (if this is not reasonable, use a different model)

Estimating differences: If we want to make a confidence interval for population differences in mean yield between variety 1 and variety 2 (subplot factor)

- Our point estimate will be $\bar{y}_{.1.} - \bar{y}_{.2.}$

- $SE(\bar{y}_{.1.} - \bar{y}_{.2.}) = \sqrt{\frac{2MS_{error(2)}}{an}}$
- Our 95% CI will be $\bar{y}_{.1.} - \bar{y}_{.2.} \pm t_{.975,a(n-1)(b-1)}\sqrt{\frac{2MS_{error(2)}}{an}}$

Repeated Measures (Outline 10)

In this setup, 30 babies are randomly assinged to three infant formula treatments. The response is the baby's weight which is recorded at 0 weeks, 1 week, 4 weeks, 2 months, and 6 months.

If we ignore the repeated emasures here and just analyze the gain over 6 months, we would just have a CRD. If we're interested in the treatments over time however, we have to look closer at the repeated measures.

The proposed model here is similar to a split-plot design where the baby is the whole plot, the formula is the whole plot treatment, and time is the subplot factor.

$$y_{ijk} = \mu + \alpha_i + \epsilon_{k(i)} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{k(i)j}$$

where the response is the weight of the *kth* baby in the *ith* treatment at the *jth* time, α_i is the treatment factor, β_j is the time effecet, $\epsilon_{k(i)}$ is the whole plot error (baby random effect), $(\alpha\beta)_{ij}$ is the interaction of time and treatment, and $(\epsilon\beta)_{k(i)j}$ is the residual error which is the same as the baby*time interaction (note that the subscripts help us out here).

Babies are nested within treatments.

with the following assumptions:

$$\sum \alpha_i = \sum \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$
$$\eta_{k(i)} \sim N(0, \sigma^2)$$
$$\epsilon_{ijk} \sim N(0, \sigma^2_{\alpha\beta})$$

We can explore the correlation structure by recalling from the split-plot design notes that $corr(y_{ijk}, y_{ij'k}) = \frac{\sigma^2}{\sigma^2_\eta + \sigma^2} = \rho$.

The split plot model says that the correlation matrix for the five observations on baby k nested in treatment i is

$$R_{k(i)} = \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

and the correlation matrix of Y is a matrix where the diagonals are these $R_{k(i)}$ matrices from $R_{1(1)}$ and $R_{10(3)}$.

Note that the models do not allow the covariance to decrease over time for the same baby (so the measurement on week 0 is just as related to week 2 as to 6 months). If this doesn't seem reasonable, we can use a different model that allows covariance to decrease over time.

In SAS, there is an option to use "R-side correlation" where $\epsilon \sim N(0, cov(Y))$

Analysis of Covariance (Outline 11)

In this setup, we are looking at repetitive motion pain related to keyboarding. Our main treatment of interest is different ergonomic keyboards and the response variable is hours of pain after keyboarding.

We can just model this as a regular CRD if we wanted:

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

However, here we are forgetting a covariate x_{ij} (hours spent keyboarding by the *jth* subject on the *ith* keyboard type) which potentially affects the response. This can be represented as the following model:

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + \epsilon_{ij}$$

where β is the slope parameter corresponding to hours spent keyboarding and represents the change in mean hours of pain for each hour increase in keyboarding.

The assumptions in this added covariate model is

$$\epsilon_{ij} \sim N(0, \sigma^2)$$
$$\sum(\alpha_i) = 0$$

In ST552, we generally saw this model in the context of indicator variables to distinguish the treatments. There were also **no interactions** between the indicators and the covariates.

In ST553 by converse, we are more interested in treatment means ($\mu_i = \mu + \alpha_i$) and differences between treatment means ($\alpha_i - \alpha_i'$).

The observed treatment means $\bar{y}_{i.} = \frac{1}{n}\sum y_{ij}$ has expected value $E[\bar{y}_{i.}|x_{i1}, \ldots, x_{1n}] = \mu + \alpha_i + \beta\bar{x}_{i.}$

The covariate adjusted means incorporate the average of all x_{ij}s, so they are $\mu + \alpha_i + \beta\bar{x}_{..}$.

Furthermore, we can estimate the group differences here as $\hat{\alpha}_i - \hat{\alpha}_{i'}$

Bottom line: we have already seen this shit in 552, in 553 we are just more concerned with estimating and comparing the treatment means.