Course Content

- Factorial ANOVA (2-way, 3-way, ...)
- Likelihood for models with normal errors
- Logistic and Poisson regression
- Mixed models (quite a few weeks)
  - Nested random effects, Variance components
  - Estimation and Inference
  - Prediction, BLUP's
  - Split plot designs
  - Random coefficient regression
  - Marginal and conditional models
  - Model assessment
- Nonparametric regression / smoothing
- Numeric maximization
- Extended data analysis example

Review of 1 way ANOVA

Study includes K groups (or treatments). Questions concern group means, equality of group means, differences in group means, linear combinations of group means.

The usual model(s):

$$\begin{aligned} Y_{ij} &\sim N(\mu_i, \sigma^2) \\ Y_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \mu + \alpha_i + \varepsilon_{ij} \\ &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \varepsilon_{ij} &\sim N(0, \sigma^2) \end{aligned}$$

i identifies the group, $i = 1, 2, \ldots K$.
j identifies observation within group, $j = 1, 2, \ldots n_i$.

1-way ANOVA: example

Does changing your diet help you live longer? Mice were randomly assigned to one of 5 diets and followed until they died.

- NP: No calorie restriction (ad lib).
- N/N85: 85 cal/day throughout life (usual food recommendation)
- N/R50: 85 cal/day early in life, 50 cal/day later.
- N/R40: 85 cal/day early in life, 40 cal/day later.
- R/R50: 50 cal/day early in life, 50 cal/day later.

Raised and fed individually. Response is time of death.
Original data set had between 49 and 71 mice per diet and a 6th trt. I have subsampled individuals to get 49 per diet and removed one diet. Will see why later.

1-way ANOVA: linear contrasts

The treatment structure suggests specific comparisons of treatments:

Question	Contrast
Does red. cal. early alter long.?	N/R50 - R/R 50
Does late from 50 to 40 a. l.?	N/R50 - N/R40
or	(N/R50 + R/R50)/2 - N/R40
Does late from 85 to 50 a. l.?	N/N85 - N/R50
Ave. eff. of red. late cal.?	NP - (N/N85 + N/R50 + N/R40 + R/R50)/4
linear eff. of late cal.?	80*N/R85 - 25*N/R50 - 55*N/R40
or	(80*N/R85 - 25*N/R50 - 55*N/R40)/33975

Where do the last two sets of coefficients come from?
(See next slide)

1-way ANOVA: linear contrasts

Remember equation for slope in a simple linear regression, using two subscripts: i: treatment, j observation within a treatment, $X_{ij} = X_i \; \forall j$, all treatments have same n.

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\Sigma_{i,j}(X_{ij} - \overline{X})\; Y_{ij}}{\Sigma_{i,j}(X_{ij} - \overline{X})^2} \\
&= \frac{\Sigma_i\,(X_i - \overline{X})\,\Sigma_j\, Y_{ij}}{\Sigma_i n\,(X_i - \overline{X})^2} \\
&= \frac{\Sigma_i\,(X_i - \overline{X})\; n\,\overline{Y}_i}{\Sigma_i n\,(X_i - \overline{X})^2} \\
&= \frac{\Sigma_i\,(X_i - \overline{X})}{\Sigma_i\,(X_i - \overline{X})^2}\,\overline{Y}_i = c_i\;\overline{Y}_i
\end{aligned}$$

1-way ANOVA: linear contrasts

This is a linear combination of treatment means with coefficients that depend on the X's. The last one is $(X_i - \overline{X})/\Sigma(X_i - \overline{X})^2$, so the estimate is the regression slope. The simpler set are "nice" coefficients proportional to $(X_i - \overline{X})$, so they only give a test of slope = 0.

Each of these is a linear combination of treatment means.
Each is also a linear contrast: the sum of coefficients = 0.
Each is specified before the data are collected, either explicitly in a data analysis plan or implicitly by the choice of treatments.

Review of 1 way ANOVA

Sufficient Statistics: $\overline{Y}_i$, $\Sigma(Y_{ij} - \overline{Y}_i)^2$

- They are the only part of the data relevant for model-based inference
- Depend on the model
- Can do analysis from raw data or from sufficient statistics
- Need raw data to do non-model based inference, e.g.
  Evaluation of the model (e.g. by inspection of residuals)
  Randomization based inference

Estimate σ^2 using $\Sigma(Y_{ij} - \overline{Y}_i)^2$
Estimate μ_i using $\overline{Y}_i$

Models for 1 way ANOVA

$$\text{cell means:} \quad Y_{ij} = \mu_i + \varepsilon_{ij} \tag{1}$$

$$\text{effects} \quad = \mu + \alpha_i + \varepsilon_{ij} \tag{2}$$

Model (1) has $K + 1$ parameters (K for means)
Model (2) has $K + 2$ parameters ($K + 1$ for means)

$K + 1$ sufficient statistics, so model 2 is overparameterized

Estimation for 1 way ANOVA models

Define β = vector of parameters in a model.

Models (1): β is length K, $\boldsymbol{X}$ has full column rank
Estimate $\hat{\beta}$ by $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})$

Model (2): β is length $K+1$, $\boldsymbol{X}$ has $K+1$ columns but column rank is K
Model (2) is overparameterized. We will use generalized inverses.
Estimate $\hat{\beta}$ by $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-}(\boldsymbol{X}'\boldsymbol{Y})$

SAS uses generalized inverses. R puts a restriction on one (or more) parameters.

Estimation in 1 way ANOVA

Var $\hat{\boldsymbol{\beta}} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^-$

Estimate σ^2 by:

$$\hat{\sigma}^2 = s_p^2 = \frac{\Sigma(Y_{ij} - \overline{Y}_i)^2}{N - \text{rank } \boldsymbol{X}},$$

where $N = \sum N_i$.

Estimate Var $\hat{\boldsymbol{\beta}}$ by $\hat{\Sigma}_{\hat{\beta}} = s_p^2(\boldsymbol{X}'\boldsymbol{X})^-$.

standard error of $\hat{\mu}_i = s_{\mu_i} = \sqrt{\hat{\Sigma}_{\hat{\beta},i,i}}$

Inference in 1 way ANOVA

Inference about μ_i based on

$$T = \frac{\hat{\mu}_i - \mu_{i,0}}{s_{\mu_i}}$$

Leads to tests of $H_0 : \mu_i = \mu_{i,0}$; often $\mu_{i,0} = 0$ and confidence intervals for μ_i.

Tests on multiple parameters simultaneously, e.g. all groups have the same population mean:

- model (1): $H_0 : \mu_i = \mu, \forall i$
- model (2): $H_0 : \alpha_i = 0, \forall i$

Inference in 1 way ANOVA

Three approaches:

- 1) Model comparison: Use SSE = $\Sigma(Y_{ij} - \overline{Y}_i)^2$ as a measure of how well a model fits a set of data.
  Compare SSE_{full} for full model, E $Y_{ij} = \mu + \alpha_i$ to SSE_{red} for reduced model, expressing the null hypothesis:
  Under H_0: E $Y_{ij} = \mu$

$$F = \frac{(\text{SSE}_{red} - \text{SSE}_{full})/(K-1)}{\text{SSE}_{full}/(N-K)}$$

  has a central F distribution with $K-1$, $N-K$ d.f.

Hypothesis tests: via orthogonal contrasts

- 2) Combining orthogonal contrasts.
  $\gamma = \Sigma l_i \mu_i$ is a linear combination of population means
  Is a linear contrast when $\Sigma l_i = 0$
  Estimated by $\hat{\gamma} = \Sigma l_i \overline{Y}_i$
  Est. variance is $s_p^2 \Sigma (l_i^2 / n_i)$
  Inference on one linear comb. usually by T distributions
  The SS associated with a contrast are defined as:

$$SS_\gamma = \hat{\gamma}^2 / (\Sigma_i l_i^2 / n_i)$$

  Leads to an F test of $H_0 : \gamma = 0$, df = 1, $N - K$

Hypothesis tests in 1 way ANOVA: via orthog. contrasts

- Combining orthogonal contrasts.
  $l_i\mu_i$ and $m_i\mu_i$, are orthogonal when $\Sigma l_i m_i/n_i = 0$
  When all $n_i = n$, then condition is $\Sigma l_i m_i = 0$.
  The SS associated with $K-1$ pairwise orthogonal contrasts "partition" the "between" group SS.
  Get the "between" group SS by writing $K-1$ orthogonal contrasts, calculating the SS for each, and summing.
  Many different sets of orthogonal contrasts, sum is always the same

Hypothesis tests in 1 way ANOVA: via $C\beta$ test

- 3) Can write arbitrary sets of linear combinations as $H_o : C\boldsymbol{\beta} = \mathbf{m}$ (C an $r \times (k+1)$ matrix of rank r)
  Examples of C matrices
  Model: $Y_i = \beta_0 + \beta_1 \boldsymbol{X}_{1i} + \beta_2 \boldsymbol{X}_{2i} + \epsilon_i$
  - Test $\beta_1 = 0$, $C = [0\ 1\ 0]$
  - Test $\beta_1 = \beta_2$, $C = [0\ 1\ -1]$
  - Test $\beta_1 = 0$, $\beta_2 = 0$, $C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Hypothesis tests in 1 way ANOVA: via $C\beta$ test

- 3) To test $H_0 : C\beta = \boldsymbol{m}$ compare

$$F = \frac{(C\boldsymbol{b} - \boldsymbol{m})' \left[C(\boldsymbol{X}^T\boldsymbol{X})^{-1}C' \right]^{-1} (C\boldsymbol{b} - \boldsymbol{m})}{r\, MS_{error}} \tag{3}$$

with an $F_{r,df_{error.full}}$ distribution
If the contrasts defined by the rows of C are orthogonal, $\left[C(\boldsymbol{X}'\boldsymbol{X})^{-1}C' \right]$ is diagonal. The quadratic form in the numerator of (3) is then a sum.
Hence, this approach generalizes orthogonal contrasts to any set of contrasts

1-way ANOVA: example, cont.

Summary of the data:

Treatment	N/N85	N/R40	N/R50	NP	R/R50
Average	34.3	47.3	46.4	27.4	44.8
sd	3.06	4.15	e.59	6.13	4.54

Error SS: 4687.7, MSE = $s_p = 19.53$

Want to test $H_0 : \mu_i = \mu \, \forall i$, i.e. all treatments have the same mean lifespan, or no effect of treatment on lifespan.

Reminder: treatments randomly assigned, so causal claims about treatment effects are justified.

Example: Model Comparison

Reduced model: all groups have same mean: E $Y_{ij} = \mu$
Full model: each group has a different mean:
E $Y_{ij} = \mu_i$ or E $Y_{ij} = \mu + \alpha_i$

		Error
Model	df	SS
Reduced	244	19824
Full	240	4687.7
Difference	4	15136.2

$$F = \frac{15136.2/4}{4687.7/240} = \frac{3784.0}{19.5} = 193.74$$

$$p < 0.0001$$

1-way ANOVA: model comparison

Or, in ANOVA table format:

Model	Source	df	SS	MS	F	p
Difference	Model	4	15136.2	3784.0	193.74	< 0.0001
Full	Error	240	4687.7	19.5		
Reduced	C.Total	244	19824			

Example: Orthogonal Contrasts

A convenient set of orthogonal contrasts (when $n_i = n$) are the Helmert contrasts, which for 5 treatments are:

	Coefficients				Estimate	SS
1	-1	0	0	0	-13.041	4166.5
1	1	-2	0	0	-11.171	1019.2
1	1	1	-3	0	45.859	8587.5
1	1	1	1	-4	-23.586	1362.9

The sum of the 4 SS is 15136.2
same as the numerator SS from model comparison.

Example: Orthogonal Contrasts - 2

A second set of 4 contrasts is:

Coefficients					Estimate	SS
-2	-1	0	1	2	0.97	4.6
2	-1	-2	-1	2	-9.43	311.6
-1	2	0	-2	1	50.35	12420.6
1	-4	-6	-4	-1	58.55	2399.4

The sum of the 4 SS is 15136.2, the same as the sum of the previous sets of contrasts and the numerator SS when doing model comparison.

Notice that the estimate of the contrast and the SS of the contrast depend on the coefficients, but if the set is orthogonal, the sum of the SS is the same.

Example: Orthogonal Contrasts - 3

Another set of 4 contrasts is:

	Coefficients				Estimate	SS
1	-1	0	0	0	-13.04	4166.5
1	0	-1	0	0	-12.10	3590.7
1	0	0	-1	0	6.90	1167.8
1	0	0	0	-1	-10.46	2679.1

This set is very easy to interpret (difference from 1st group), but are they orthogonal?

Notice that the sum of SS is 11604.2, the wrong number.

Example: $C\beta$ tests

Consider fitting model 1 (cell means model), then using the Helmert contrasts as the $\boldsymbol{C}$ matrix.

C matrix 1 (Helmert contrasts):

$\boldsymbol{C}\beta - m$	$\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]$			
13.04	0.0408	0	0	0
11.17	0	0.1224	0	0
-45.86	0	0	0.2449	0
23.58	0	0	0	0.4082

The numerator in equation (3) (slide 16) is

$$(\boldsymbol{C}\beta - m)\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C}\beta - m)' = 15136.2$$

Example: $C\beta$ tests - 2

Consider fitting model 1 (cell means model), then using the 2nd set of contrasts as the $\boldsymbol{C}$ matrix.

C matrix 2 (More orthogonal contrasts):

$\boldsymbol{C}\beta - m$	$\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]$			
0.969	0.204	0	0	0
-9.435	0	0.286	0	0
50.347	0	0	0.204	0
58.547	0	0	0	1.43

The numerator in equation (3) (slide 16) is

$$(\boldsymbol{C}\beta - m)\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C}\beta - m)' = 15136.2$$

Example: $C\beta$ tests - 3

C matrix 3 (non-orthogonal contrasts):

$\boldsymbol{C}\beta - m$	$\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]$			
-13.04	0.0408	0.0204	0.0204	0.0204
-12.11	0.0204	0.0408	0.0204	0.0204
6.90	0.0204	0.0204	0.0408	0.0204
-10.46	0.0204	0.0204	0.0204	0.0408

The numerator in equation (3) (slide 16) is

$$(\boldsymbol{C}\beta - m)\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C}\beta - m)' = 15136.2$$

Note although the components of $\boldsymbol{C}\beta$ for this $\boldsymbol{C}$ are not independent (why?), the SS arising from the $\boldsymbol{C}\beta$ approach are the same for all three $\boldsymbol{C}$ matrices.

R code for Example

```
diet <- read.csv('diet.csv', as.is=T)
diet <- diet[diet$diet !='lopro',]

# I use as.is=T to force me to explicitly declare
#   factors. default conversion does not convert
#   factors with numeric levels.

diet$diet.f <- factor(diet$diet)

# anova using lm:
diet.lm <- lm(longevity ~ diet.f, data=diet)
```

R code for Example

```
# lm() has lots of helper / reporting functions:
coef(diet.lm)
   # coefficients
vcov(diet.lm)
   # and their variance-covariance matrix
sqrt(diag(vcov(diet.lm)))
   # se's of the coefficients
anova(diet.lm)
   # ANOVA table using type I = sequential SS
summary(diet.lm)
   # lots of information
plot(predict(diet.lm), resid(diet.lm))
   # plot of residuals vs predicted values
```

R code for Example

```
# model comparison by hand:
diet.m0 <- lm(longevity ~ +1, data=diet)
   # intercept only model
anova(diet.m0, diet.lm)
   # change in SS from 1st to 2nd model

drop1(diet.lm)
   # drop one term at a time = type II SS
```

R code for Example

```
# orthogonal contrasts
diet.means <- tapply(diet$longevity, diet$diet, mean)
   # mean for each diet
diet.helm <- contr.helmert(5)
   # matrix of Helmert coeff. for 5 groups
diet.c1 <- t(diet.helm) %*% diet.means
   # estimate of each contrast
diet.ss1 <- 49*diet.c1^2/apply(diet.helm^2,2,sum)
   # SS for each contrast
sum(diet.ss1)
```

R code for Example

```
# second set of contrasts were hand entered
#  using diet.2nd <- rbind(c(-2,-1,0,1,2), ...)

# third set:
diet.trt <- -contr.treatment(5)
diet.trt[1,] <- 1

diet.c3 <- t(diet.trt) %*% diet.means
   # estimate of each contrast
diet.ss3 <- 49*diet.c3^2/apply(diet.trt^2,2,sum)
   # SS for each contrast
sum(diet.ss3)
```

R code for Example

```
# C beta tests
diet.c1 <- t(diet.helm) %*% diet.means
   # estimate of contrast

X1 <- model.matrix(~ -1 + diet.f, data=diet)
   # X matrix using cell means parameterization
diet.cm1 <- t(diet.helm) %*% solve(t(X1) %*% X1) %*%
   diet.helm

t(diet.c1) %*% solve(diet.cm1) %*% diet.c1

# how do the explicit contrasts of cell means
#   compare to R using those contrasts?
```

R code for Example

```
contrasts(diet$diet.f) <- contr.helmert
#  tell R to use helmert contrasts
#  default is contr.treatment,
#    which is drop first level of the factor
#  contr.SAS is SAS-like (drop last level)

diet.lmh <- lm(longevity ~ diet.f,data=diet)
coef(diet.lmh)
# coefficients are not the same as the
#   hand-computed contrasts!
```

R code for Example

```
# the R coefficients are related to the
#   contrasts among cell means

apply(diet.helm^2,2,sum)
   # sum of squared coefficients

coef(diet.lmh)[2:5] * c(2,6,12,20)
# these are the same as diet.c1

# although R calls them contrasts, they really are not
# they are columns of the X matrix for a regression.
model.matrix(diet.lmh)
```

R code for Example

```
# can do C beta from the lm() information

diet.mse <- 4687.7/240    # MSE = est of sigma^2
diet.clm <- coef(diet.lm)[2:5]
  # extract coefficients for diet factor
diet.vclm <- vcov(diet.lm)[2:5,2:5]
  # and their VC matrix
diet.mse * t(diet.clm) %*% solve(diet.vclm) %*%
    diet.clm
  # SS for diet
t(diet.clm) %*% solve(diet.vclm) %*% diet.clm/5
  # F statistic for diet (5 is # rows in C)
```

ANalysis Of VAriance (ANOVA) for a sequence of models

- Model comparison can be generalized to a sequence of models (not just one full and one reduced model)
- Context: usual nGM model: $\boldsymbol{y} = X\beta + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\mathbf{0},\ \sigma^2 I)$
- Let $X_1 = \mathbf{1}$ and $X_m = X$.
- But now, we have a sequence of models "in between" $\mathbf{1}$ and $\boldsymbol{X}$
- Suppose $X_2,\ \ldots, \boldsymbol{X}_{m-1}$ are design matrices satisfying
- $\mathcal{C}(X_1) < \mathcal{C}(X_2) < \ldots < \mathcal{C}(X_{m-1}) < \mathcal{C}(X_m)$.
- We'll also define $X_{m+1} = I$

Some examples

- Multiple Regression
  - $X_1 = \mathbf{1},\ X_2 = [\mathbf{1},\ \boldsymbol{x}_1],\ X_3 = [\mathbf{1},\ \boldsymbol{x}_1,\ \boldsymbol{x}_2], \ldots\ X_m = [\mathbf{1}_1\boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_{m-1}]$.
  - $SS(j+1 \mid j)$ is the decrease in SSE that results when the explanatory variable x_i is added to a model containing $1,\ x_1,\ \ldots,\ x_{j-1}$.
- Test for linear trend and test for lack of linear fit.

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \\ 1 & 4 \end{bmatrix}, X_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Context for linear lack of fit

- Let $\mu_i =$ mean surface smoothness for a piece of metal ground for i minutes ($i = 1, 2, 3, 4$).
- MS(2 | 1) / MSE can be used to test
  - $H_o : \mu_1 = \mu_2 = \mu_3 \iff \mu_i = \beta_0 \quad i = 1, 2, 3, 4$ for some $\beta_o \in I\!R$
    vs. $H_A : \mu_i = \beta_0 + \beta_1 \, i \quad i = 1, 2, 3, 4$ for some $\beta_o \in I\!R, \ \beta_1 \in I\!R \backslash \{0\}$.
  - This is the F test for a linear trend, $\beta_1 = 0$ vs. $\beta_1 \neq 0$
- MSE(3 | 2) / MSE can be used to test
  - $H_o : \mu_i = \beta_0 + \beta_1 \, i \quad i = 1, 2, 3, 4$ for some $\beta_o \beta_1 \in I\!R$
    vs. H_A : There does not exist $\beta_0, \beta_1 \in I\!R$ such that
    $\mu_i = \beta_0 + \beta_1 \, i \qquad \forall \, i = 1, 2, 3, 4.$
  - This is known as the F test for lack of linear fit.
  - Compares fit of linear regression model $\mathcal{C}(X2)$
    to fit of means model $\mathcal{C}(X3)$

- All tests can be written as full vs. reduced model tests
- Which means they could be written as tests of $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$
- But, what is $\boldsymbol{C}$?
- Especially when interpretation of $\boldsymbol{\beta}$ changes from model to model
- Example:
  - $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ slope is β_1
  - $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$ slope at X_i is $\beta_1 + 2\beta_2 X_i$
  - In grinding study, $X_i = 0$ outside range of X_i in data

- What can we say about the collection of tests in the ANOVA table?

General framework

- Context: usual nGM model: $\boldsymbol{y} = X\beta + \epsilon,\ \epsilon \sim N(\mathbf{0},\ \sigma^2 I)$
- Let $X_1 = \mathbf{1}$ and $X_m = X$.
- Suppose $X_2,\ \ldots,\ \boldsymbol{X}_{m-1}$ are design matrices satisfying
- $\mathcal{C}(X_1) < \mathcal{C}(X_2) < \ldots < \mathcal{C}(X_{m-1}) < \mathcal{C}(X_m)$.
- We'll also define $X_{m+1} = I$
- Let $P_j = P_{X_j} = X_j(X_j'X_j)^- X_j' \quad \forall\, j = 1, \ldots, m+1$. Then
- $$\sum_{i=1}^{n}(y_i - \bar{y}_\cdot)^2 = \boldsymbol{y}'(I - P_1)\boldsymbol{y} = \boldsymbol{y}'(P_{m+1} - P_1)\boldsymbol{y}$$
  $$= \boldsymbol{y}'(P_{m+1} - P_m + P_m - P_{m-1} + \cdots + P_2 - P_1)\boldsymbol{y}$$
  $$= \boldsymbol{y}'(P_{m+1} - P_m)\boldsymbol{y} + \ldots + \boldsymbol{y}'(P_2 - P_1)\boldsymbol{y}$$
  $$= \sum_{j=1}^{m}\boldsymbol{y}'(P_{j+1} - P_j)\boldsymbol{y}.$$

Sequential SS

- The error sums of squares is a quadratic form:

$$y'(I - P_1)\boldsymbol{y} = \sum_{j=1}^{m} \boldsymbol{y}'(P_{j+1} - P_j)\boldsymbol{y}$$

- are often arranged in an ANOVA table.

Sum of Squares	
$\boldsymbol{y}'(P_2 - P_1)\boldsymbol{y}$	$SS(2 \mid 1)$
$\boldsymbol{y}'(P_3 - P_2)\boldsymbol{y}$	$SS(3 \mid 2)$
$\vdots$	$\vdots$
$\boldsymbol{y}'(P_m - P_{m-1})\boldsymbol{y}$	$SS(m \mid m-1)$
$\boldsymbol{y}'(P_{m+1} - P_m)\boldsymbol{y}$	$SSE = \boldsymbol{y}'(I - P_X)\boldsymbol{y}$
$\boldsymbol{y}'(I - P_1)\boldsymbol{y}$	$SSTot = \sum_{i=1}^{n}(y_i - \bar{y}_\cdot)^2$

- What can we say about each SS in the ANOVA table?

- 1) Each is a quadratic form, $\boldsymbol{W}'\boldsymbol{A}\boldsymbol{W}$, where $W \sim N(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I})$
- 2) Each is proportional to a Chi-square distribution, because $\forall\, j = 1, \ldots, m$, $\boldsymbol{A}\boldsymbol{\Sigma} = (P_{j+1} - P_j)/\sigma^2\ \sigma^2\boldsymbol{I}$ is idempotent

$$\begin{aligned}(P_{j+1} - P_j)(P_{j+1} - P_j) &= P_{j+1}P_{j+1} - P_{j+1}P_j - P_jP_{j+1} + P_jP_j \\ &= P_{j+1} - P_j - P_j + P_j \\ &= P_{j+1} - P_j.\end{aligned}$$

- So, (Stat 500) $\boldsymbol{y}\frac{(P_{j+1}-P_j)}{\sigma^2}\boldsymbol{y} \sim \chi^{2(ncp)}_{\nu}$ with
  - ncp $=\beta' X'(P_{j+1} - P_j)X\beta\sigma^2)$
  - d.f. $= \nu = \text{rank}(X_{j+1}) - \text{rank}(X_j)$

  for all, $j = 1, \ldots, m$

- 3) Each SS is independent, using Stat 500 fact: $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ independent of $\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}$ if $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = 0$
- i.e. $\forall\ j < \ell$

$$\begin{aligned}(P_{j+1} - P_j)(P_{\ell+1} - P_\ell) &= P_{j+1}P_{\ell+1} - P_{j+1}P_\ell - P_jP_{\ell+1} + P_jP_\ell \\ &= P_{j+1} - P_{j+1} - P_j + P_j \\ &= 0.\end{aligned}$$

- It follows that the m χ^2 random variables are all independent.
- This result sometimes called Cochran's theorem

- 4) Can add sequential SS. If it makes sense to test:
  full model X_4 vs reduced model X_2,
  SS for that test =

  $$\text{SS}(4 \mid 3) + \text{SS}(3 \mid 2) = \boldsymbol{y}'(P_4 - P_3)\boldsymbol{y} + \boldsymbol{y}'(P_3 - P_2)\boldsymbol{y}$$

- In general, 3) and 4) only true for sequential SS (type I SS)
- Applies to other SS (e.g. partial = type III SS) only when appropriate parts of $\boldsymbol{X}$ are orthogonal to each other
- For factor effects models, only when design is balanced (equal # obs. per treatment)

Connection to full vs. reduced SS

- Each comparison of models is equivalent to a full vs. reduced model comparison:
- To see this, note that:

$$\begin{aligned} SS(j+l \mid j) &= \boldsymbol{y}'(P_{j+1} - P_j)\boldsymbol{y} \\ &= \boldsymbol{y}'(P_{j+1} - P_j + I - I)\boldsymbol{y} \\ &= \boldsymbol{y}'(I - P_j - I + P_{j+1})\boldsymbol{y} \\ &= \boldsymbol{y}'(I - P_j)\boldsymbol{y} - \boldsymbol{y}'(I - P_{j+1})\boldsymbol{y} \\ &= SSE_{\text{REDUCED}} - SSE_{\text{FULL}} \end{aligned}$$

- For each test j, Ho_j is $E(\boldsymbol{y}) \in \mathcal{C}(X_j)$, Ha is $E(\boldsymbol{y}) \in \mathcal{C}(X_{j+1})$

F tests for sequential SS

- For each sequential hypothesis, $j = 1, \ldots, m-1$ we have

$$\begin{aligned} F_j &= \frac{\boldsymbol{y}'(P_{j+1} - P_j)\boldsymbol{y}/[\text{rank}(X_{j+1}) - \text{rank}(X_j)]}{\boldsymbol{y}'(I - P_X)\boldsymbol{y}\ /[n - \text{rank}(X)]} \\ &\sim F^{ncp}_{\nu_1,\nu_2} \text{ where} \\ \text{ncp} &= \beta' X'(P_{j+1} - P)j)X\beta/\sigma^2), \text{ and} \\ \nu_1 &= \text{rank}(X_{j+1}) - \text{rank}(X_j) \\ \nu_2 &= n - \text{rank}(X) \end{aligned}$$

- define MS $(j+1 \mid j) = \frac{\boldsymbol{y}'(P_{j+1} - P_j)\boldsymbol{y}}{\text{rank}(X_{j+1}) = \text{rank}(X_j)}$
- $F_j = \text{MS}\ (j+1 \mid j)$ / MSE
- Under Ho_j, noncentrality parameter for test $j = 0$

Details of non-centrality parameter

The noncentrality parameter is

$$\begin{aligned}
\beta' X'(P_{j+1} - P_j)X\beta/\sigma^2 &= \frac{\beta' X'(P_{j+1}-P_j)'(P_{j+1}-P_j)X\beta}{\sigma^2} \\
&= || \, (P_{j+1} - P_j)X\beta \, ||^2 \, /\sigma^2 \\
&= || \, P_{j+1}E(\boldsymbol{y}) - P_j E(\boldsymbol{y}) \, ||^2 \, /\sigma^2.
\end{aligned}$$

If Ho_j is true, $P_{j+1}E(\boldsymbol{y}) = P_j E(\boldsymbol{y}) = E(\boldsymbol{y})$.

Thus, the ncp. for test $j = 0$ under Ho_j.

Return to issues in examples

1. How does $\boldsymbol{X}$ matrix for $y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ relate to $\boldsymbol{X}$ matrix for $y_{ij} = \mu_i + \epsilon_{ij}$?
   What sort of $\boldsymbol{C}\beta$ test corresponds to this model comparison?
2. How to interpret tests when "meaning" of β_1 changes between $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$?

- Example: $X_i \in \{1, 2, 3, 4\}$. Consider 3 $\boldsymbol{X}$ matrices

$$X_1 \qquad X_2 \qquad X_3$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \\ 1 & 4 & 16 & 64 \end{bmatrix} \quad \begin{bmatrix} 1 & -3 & 1 & -1 \\ 1 & -3 & 1 & -1 \\ 1 & -1 & -1 & 3 \\ 1 & -1 & -1 & 3 \\ 1 & 1 & -1 & -3 \\ 1 & 1 & -1 & -3 \\ 1 & 3 & 1 & 1 \\ 1 & 3 & 1 & 1 \end{bmatrix}$$

- Columns of $\boldsymbol{X}_2$ are X's in a cubic regression:
  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_i$
- A cubic perfectly fits four points, so $\mathcal{C}(\boldsymbol{X}_2) = \mathcal{C}(\boldsymbol{X}_1)$
- So comparison of $Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}$ vs $Y_{ij} = \mu_i + \epsilon_{ij}$ is same as comparison of $Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}$ vs
  $Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij}^2 + \beta_3 X_{ij}^3 + \epsilon_{ij}$
- Now very clear that $\mathcal{C}(X_o) \in \mathcal{C}(X_1) = \mathcal{C}(X_2)$
- Model comparison test same as $\boldsymbol{C}\boldsymbol{\beta}$ test of $\beta_2 = 0$ and $\beta_3 = 0$.

- Each column of $\boldsymbol{X}_3$ can be expressed in terms of columns of $\boldsymbol{X}_2$
- Define $\boldsymbol{X}_i[j]$ as the j'th column of $\boldsymbol{X}_i$
  - $\boldsymbol{X}_3[2] = 2\boldsymbol{X}_2[2] - 5\boldsymbol{X}_3[1]$
  - $\boldsymbol{X}_3[3] = 2(\boldsymbol{X}_2[3] - \boldsymbol{X}_3[2] + 7.5)/5$
  - $\boldsymbol{X}_3[4] = 10(\boldsymbol{X}_2[4] - 7.5\boldsymbol{X}_3[3] - 10.4\boldsymbol{X}_3[2] - 25)/3$
- Why consider $\boldsymbol{X}_3$?

$$X_1'X_1 \qquad X_2'X_2 \qquad X_3'X_3$$

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 4 & 10 & 30 & 100 \\ 10 & 30 & 100 & 354 \\ 30 & 100 & 354 & 1300 \\ 100 & 354 & 1300 & 4890 \end{bmatrix} \quad \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix}$$

Orthogonal polynomials

- Columns of $\boldsymbol{X}_3$ are orthogonal, when sample sizes equal
- estimates of β's are independent ($(\boldsymbol{X}_3'\boldsymbol{X}_3)^{-1}$ is diagonal).
- Columns of $\boldsymbol{X}_3$ are one example of a set of orthogonal polynomials.
- Uses of orthogonal polynomials:
  - Historical: fitting a regression. $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ much easier to compute
  - Analysis of quantitative ("amount of") treatments:
    Decompose SS for trt into additive components due to linear, quadratic...
    Extends to interactions, e.g. linear A x linear B
  - Alternate basis for full-rank parameterization (instead of drop first)
  - Numerical stability for regressions

Orthogonal polynomials - 2

- I once tried to fit a cubic regression, X = year: 1992, 1993, ... 2006
- software complained:
  X matrix not full rank, X3 dropped from model
- Correlation matrix of estimates, $((\boldsymbol{X}'\boldsymbol{X})^{-1}$ scaled so 1's on diagonal, when X = 1, 1, 2, 2, 3, 3, 4, 4

$$\begin{array}{rrrr} 1.0000000 & -0.9871142 & 0.9652342 & -0.9421683 \\ -0.9871142 & 1.0000000 & -0.9934490 & 0.9798135 \\ 0.9652342 & -0.9934490 & 1.0000000 & -0.9960238 \\ -0.9421683 & 0.9798135 & -0.9960238 & 1.0000000 \end{array}$$

- Correlations even closer to ± 1 for X = 1992 ... 2006
- that $\boldsymbol{X}'\boldsymbol{X}$ matrix fails numerical test for singularity
- for fun, plot X^2 vs. X^3 or X^3 vs. X^4

Orthogonal polynomils - 3

- Consequence is numerical instability in all computations
- How can we reduce correlations among columns in $\boldsymbol{X}$ matrix?
  1. Center X's at mean X. $X_i = X_i - \bar{X}$
     Correlation matrix of estimates, $((\boldsymbol{X}'\boldsymbol{X})^{-1}$ scaled so 1's on diagonal, when X = 1, 1, 2, 2, 3, 3, 4, 4

$$\begin{array}{rrrr} 1.0000000 & 0.0000000 & -0.7808688 & 0.0000000 \\ 0.0000000 & 1.0000000 & 0.0000000 & -0.9597374 \\ -0.7808688 & 0.0000000 & 1.0000000 & 0.0000000 \\ 0.0000000 & -0.9597374 & 0.0000000 & 1.0000000 \end{array}$$

     Correlations much reduced, still have correlations between odd powers and between even powers
- Use orthogonal polynomials: all estimates uncorrelated

Coefficients for orthogonal polynomials

- Where do you find coefficients?
- Tables for statisticians, e.g. Biometrika tables, vol. 1
  Only for equally spaced X's, equal numbers of obs. per X
- General approach: n obs. X_i is vector of X^i.
- C_0 = 0'th degree orthog. poly. is a vector of 1's = X_0.
- linear orthog. poly.: want to find C_1 so that C_1 orthog. to X_0
  - X_1 is a point in n-dimensional space
  - $\mathcal{C}(C_0)$ is a subspace.
    Want to find a basis vector for the subspace $\perp \mathcal{C}(C_0)$.
  - That is $(\boldsymbol{I} - P_{C_0})X_1$, i.e. residuals from regression of X_1 on C_0
- linear coeff: proportional to residuals of regr. X_1 on C_0
- quadratic coeff. are residuals from regr. of X_2 on $[C_0, C_1]$
- C_i is prop. to residuals from regr. of X_i on $[C_0, C_1, \ldots C_{i-1}]$

Multifactor studies - Introduction

- Experiments/observational studies with more than one factor
- Examples:
  - vary price (3 levels) and advertising media (2 levels) to explore effect on sales
  - model family purchases using income (4 levels) and family stage (4 levels) as factors
  - both ex. of 2 way factorial
- Why use multifactor studies?
  - efficient (can learn about more than one factor with same set of subjects)
  - added info (can learn about factor interactions)
  - but ... too many factors can be costly, hard to analyze

Factorial designs - Introduction

- Complete factorial design - takes all possible combinations of levels of the factors as separate treatments
- Example: 3 levels of factor A (a_1, a_2, a_3) and
  2 levels of factor B (b_1, b_2) yields
  6 treatments ($a_1b_1, a_1b_2, a_2b_1, a_2b_2, a_3b_1, a_3b_2$)
- Terminology:
  - complete factorial (all combinations used) vs fractional factorial (only a subset used)
  - complete factorials widely used.
  - fractional fact. very imp. in industrial studies
    will describe concepts if time

Factorial designs - Introduction

- Outline
  - factorial design with two factors
  - factorial designs with blocking
  - factorial designs with more than two factors
  - factorials with no replication

Experimental design and Treatment design

- Experimental studies have two distinct features
  - Treatment design: what trts are used
    - complete factorial
    - 1-way layout (K unstructured treatments)
    - central composite response surface
  - Experimental design:
    how trts randomly assigned to e.u.'s
    - CRD, RCBD, Latin Square
    - split-plot (2 different sizes of e.u.'s).
  - Mix and match. e.g. 2-way factorial in a Latin Square.
  - Model will include both treatment structure and expt. design.
  - Both matter. Analysis of a 2-way factorial CRD is quite different from 2-way factorial split plot.

Crossed and nested factors

- Two factors are crossed when all combinations of one factor are matched with all combinations of the other
- price and advertising media are crossed when:

	Price		
Media	1	2	3
A	x	x	x
B	x	x	x

- Both marginal means "make sense" when crossed

Crossed and nested factors

- Notation / terminology for means
  - Cell means: means for each comb. of factor levels
  - Marginal means: means across rows or down columns
  - Dots are used to indicate averaging
	Price			
Media	1	2	3	
A	μ_{A1}	μ_{A2}	μ_{A3}	$\mu_{A.}$
B	μ_{B1}	μ_{B2}	μ_{B3}	$\mu_{B.}$
	$\mu_{.1}$	$\mu_{.2}$	$\mu_{.3}$	$\mu_{..}$

Crossed and nested factors (cont.)

- Nested factors when two media were used for price 1, 2 diff. media for price 2 and 2 diff. media for price 3.

	Media					
Price	A	B	C	D	E	F
1	x	x				
2			x	x		
3					x	x

- Has nothing to do with labels for each factor. Could label media A,C,E as 1 and B,D,F as 2. Still nested!

Crossed and nested factors (cont.)

- Marginal means for price “make sense”. Those for media do not (even if numbered 1 and 2).
- Order matters when nesting. Media nested in Price.
- Crossing often associated with treatments; nesting often associated with random effects. Not always!
- If you can change labels (e.g. swap media 1 and 2) within price, nested.
- Is there any connection between media 1 in price 1 and media 1 in price 2? yes: crossed, no: nested.

Two factor study - example 1

- These data come from a study of the palabability of a new protein supplement.
  - 75 men and 75 women were randomly assigned to taste one of three protein supplements (control, liquid, or solid).
  - The control is the product currently on the market.
    Liquid is a liquid formulation of a new product
    Solid is a solid formulation of that new product
  - 25 men and 25 women tasted each type of product
  - Participants were asked to score how well they liked the product, on a -3 to 3 scale.
- The treatment means are:

	Type of product		
Sex	Control	Liquid	Solid
Female	0.24	1.12	1.04
Male	0.20	1.24	1.08

Two factor study - example 2

- These data come from a study of the effect of the amount and type of protein on rat weight gain.
  - Rats were randomly assigned to one of 6 treatments representing all combinations of three types of protein (beef, cereal, and pork) and two amounts (high and low).
  - Rats were housed individually.
  - The response is the weight gain in grams.
  - The study started with 10 rats per treatment, but a total of five rats got sick and were excluded from the study.

Two factor study - example 2

- The sample sizes per treatment are:

	Type of protein		
Amount	beef	cereal	pork
high	7	10	10
low	10	10	8

- The treatment means are:

	Type of protein		
Amount	beef	cereal	pork
high	98.4	85.9	99.5
low	79.2	83.9	82.0

Two factor study - questions

- focus on crossed factors
  e.g. compare 3 types and 2 amounts
- 6 treatments: all comb. of 3 types and 2 amounts
  rand. assigned to one of 60 rats (n=10 per trt)
- 4 different questions that could be asked:
  - Are the 6 means (μ_{ij}) equal?
  - Is high diff from low, averaged over types?
    $\mu_{A.} - \mu_{B.} = 0$, or $\mu_{A.} = \mu_{B.}$
  - Is there an effect of type, averaged over amount?
    $\mu_{.1} = \mu_{.2} = \mu_{.3}$
  - Is the diff. between amounts the same for all types?
    $(\mu_{A1} - \mu_{B1}) = (\mu_{A2} - \mu_{B2}) = (\mu_{A3} - \mu_{B3})$?

Crossed and nested factors (cont.)

- These correspond to questions about:
  - cell means: are 6 means equal?
  - amount marginal means: (high vs. low)
  - type marginal means: (1 vs. 2 vs. 3)
  - and interactions: high-low same for all types?
- Answer using ANOVA table and F tests
- We will motivate this 3 different ways

1) 2 way ANOVA as formulae for SS

source of variation	degrees of freedom	sums of squares
factor A	$a-1$	$nb\Sigma_i(\bar{Y}_{i..}-\bar{Y}_{...})^2$
factor B	$b-1$	$na\Sigma_j(\bar{Y}_{.j.}-\bar{Y}_{...})^2$
interaction	$(a-1)(b-1)$	$n\Sigma_i\Sigma_j(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})^2$
error	$ab(n-1)$	$\Sigma_i\Sigma_j\Sigma_k(Y_{ijk}-\bar{Y}_{ij.})^2$
total	$abn-1$	$\Sigma_i\Sigma_j\Sigma_k(\bar{Y}_{ijk}-\bar{Y}_{...})^2$

- SS for A or B are variability of marginal means
- SS for AB is deviation of cell mean from "additive expectation" = $\overline{Y}_{..}+(\overline{Y}_{i.}-\overline{Y}_{..})+(\overline{Y}_{.j}-\overline{Y}_{..})=\overline{Y}_{i.}+\overline{Y}_{.j}-\overline{Y}_{..}$
- SS for Error is variability of obs around cell mean

2 way ANOVA as formulae for SS

- Expected mean squares
  E $MSA = \sigma^2 + nb\Sigma_i(\mu_{i.} - \mu_{..})^2/(a-1)$
  E $MSB = \sigma^2 + na\Sigma_j(\mu_{.j} - \mu_{..})^2/(b-1)$
  E $MSAB = \sigma^2 + n\frac{\Sigma_i\Sigma_j(\mu_{ij}-\mu_{i.}-\mu_{.j}+\mu_{..})^2}{(a-1)(b-1)}$
  E $MSE = \sigma^2$
- Intuitive justification of F test:
  - Each E *MS* has a random component: σ^2
    and a fixed component: f(μ's)
  - Appropriate denominator is the estimate of the random component
  - F = MS / $\hat{\sigma}^2$ = MS / MSE

2 way ANOVA as formulae for SS

- More formal justification of F test
- All *MS* are independent χ^2 random variables * a constant.
  - Each MS can be written as a quadratic form: $\boldsymbol{Y}' \boldsymbol{A}_k \boldsymbol{Y}$
  - with different $\boldsymbol{A}_k$ matrices for each MS
  - So $\sigma^2 \boldsymbol{Y}' \boldsymbol{A}_k \boldsymbol{Y} \sim \chi^2$ with d.f. = rank A_k
  - These $\boldsymbol{A}_k$ matrices satisfy conditions for Cochran's theoren
  - So, each pair of $\boldsymbol{Y}' \boldsymbol{A}_k \boldsymbol{Y}$ and $\boldsymbol{Y}' \boldsymbol{A}_l \boldsymbol{Y}$ are independent
- Test hypotheses about A, B, AB using F-tests, e.g.
  - test does mean of media A, averaged over prices, = mean of media B, averaged over prices
  - $H_0 : \mu_{i.} = \mu_{..}$ for all i
  - use $F = MSA/MSE$
  - compare to $F_{a,ab(n-1)}$ distn

Example 1 continued

- Cell and marginal means:

	Type of product			
Sex	Control	Liquid	Solid	Average
Female	0.24	1.12	1.04	0.80
Male	0.20	1.24	1.08	0.84
Average	0.22	1.18	1.06	0.82

- So SS for sex = $25 \times 2 \times \left[(0.84 - 0.82)^2 + (0.80 - 0.82)^2\right] = 0.06$
- In like fashion, SS for type = 27.36
- SS for interaction = 0.16
- and SS for error = 194.56

Example 1 continued

- Which leads to the ANOVA table

Source	df	SS	MS	F	p
sex	1	0.06	0.06	0.044	0.83
type	2	27.36	13.68	10.125	< 0.0001
sex*type	2	0.16	0.08	0.059	0.94
error	144	194.56	1.35		
c.total	149	222.14			

- BUT: these formulae only work for balanced data.
- Do not use for example 2 (unequal sample sizes)

2) 2 way ANOVA as contrasts between cell means

- The cell means model for 2 way factorial

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \qquad \epsilon_{ijk} \text{ iid } N(0, \sigma^2)$$

  where $i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n_{ij}$
  - $\mu_{ij} =$ mean of all units given level i of factor A and level j of factor B
  - $\mu_{i\cdot} = \frac{1}{b}\Sigma_j \mu_{ij}$ is mean at level i of factor A
  - $\mu_{\cdot j} = \frac{1}{a}\Sigma_i \mu_{ij}$ is mean at level j of factor B
  - $\mu_{\cdot\cdot} = \frac{1}{ab}\Sigma_i \Sigma_j \mu_{ij}$ is overall mean response
- Notice: marginal means (for A or B) defined as average of cell means

2 way ANOVA as formulae for SS

- So, comparisons of marginal means are contrasts among cell means
  - Diff between male and female: $\mu_{A.} - \mu_{B.}$
    = $(\mu_{A1} + \mu_{B1} + \mu_{C1})/3 - (\mu_{A2} + \mu_{B2} + \mu_{C2})/3$
  - Diff between liquid and solid types: $\mu_{.1} - \mu_{.2}$
    $= (\mu_{A1} + \mu_{B1})/2 - (\mu_{A2} + \mu_{B2})/2$
- We begin with assumption of equal sample size $n_{ij} = n$.
- The consequences of this will be considered later.
- Note that $N = abn$

Two factor study - contrasts/factor effects

- Start with 1-way ANOVA, 6 treatments

Source	d.f.	SS
Treatments	5	SS_{trt}
Error	$6(n-1)$	SS_{error}
Total	$6n-1$	SS_{total}

- F test for treatments answers Q 1.
- Q 2, 3 and 4 answered by contrasts
  - Sex: H0: $\mu_{A.} = \mu_{B.}$
    1 contrast: $\mu_{A.} - \mu_{B.}$
  - Type: H0: $\mu_{.1} = \mu_{.2} = \mu_{.3}$
    2 contrasts: $\mu_{.1} - \mu_{.2}$, and $(\mu_{.1} + \mu_{.2})/2 - \mu_{.3}$
  - Interactions: H0: $(\mu_{A1} - \mu_{B1}) = (\mu_{A2} - \mu_{B2}) = (\mu_{A3} - \mu_{B3})$
    2 contrasts: $(\mu_{A1} - \mu_{B1}) - (\mu_{A2} - \mu_{B2})$ and
    $[(\mu_{A1} - \mu_{B1}) + (\mu_{A2} - \mu_{B2})]/2 - (\mu_{A3} - \mu_{B3})$

Two factor study - contrasts/factor effects

- These are 5 constrasts among the 6 cell means

	H c	H l	H s	L c	L l	L s
Amount	1/3	1/3	1/3	-1/3	-1/3	-1/3
Type 1	0	1/2	-1/2	0	1/2	-1/2
Type 2	-1/2	1/4	1/4	-1/2	1/4	1/4
Interaction 1	0	1	-1	0	-1	1
Interaction 2	-1	1/2	1/2	1	-1/2	-1/2

- **H**igh, **L**ow; **c**ontrol, **l**iquid, **s**olid
- The types suggest two "natural" contrasts:
  liquid - solid = difference between new formulations
  control - (liquid+solid)/2 = ave. diff. between old and new
- Main effects are averages, so coeff. are fractions. Interactions are differences of differences.

Two factor study - contrasts/factor effects

- for tests, equivalent to

	H c	H l	H s	L c	L l	L s
Amount	1	1	1	-1	-1	-1
Type 1	0	1	-1	0	1	-1
Type 2	-2	1	1	-2	1	1
Interaction 1	0	1	-1	0	-1	1
Interaction 2	-2	1	1	2	-1	-1

- Multiplying a vector of contrast weights for A, and a vector of weights for B yields a contrast for the interaction
- When sample sizes are equal, these are orthogonal same definition as before: does $\Sigma l_i m_i / n_i = 0$?

Two factor study - contrasts/factor effects

- Q2, Q3, and Q4 are often important in a 2 factor study, common to separate those SS.
- "standard" ANOVA table for a 2 way factorial with a levels of factor A and b levels of factor B.

Source	d.f.	SS
Factor A	$a-1$	SS_A
Factor B	$b-1$	SS_B
Interaction	$(a-1)(b-1)$	SS_{AB}
Error	$ab(n-1)$	SS_{error}
Total	$abn-1$	SS_{total}

- This is "standard" only because each line corresponds to a common question.
- My way of thinking:
  treatments are real; factors are made-up constructs

Two factor study - Example 1 cont.

- For 3 products and 2 sexes:

Source	d.f.	SS
Product	2	$SS_{type} = SS_{S1} + SS_{S2}$
Sex	1	SS_{amount}
Interaction	2	$SS_{int.} = SS_{int1} + SS_{int2}$
Error	$6(n-1)$	SS_{error}
Total	$6n-1$	SS_{total}

- Error and Total lines same as in 1 way
- $SS_{trt} = SS_{sex} + SS_{product} + SS_{int}$
  $df_{trt} = df_{sex} + df_{product} + df_{int}$

Example 1 continued

- The estimates and SS for each component contrast are:

Contrast	Estimate	SS
Sex	-0.04	0.06
Type 1	0.12	0.36
Type 2	0.90	27.00
Interaction 1	-0.08	0.04
Interaction 2	-0.12	0.12

- SS for Sex: 0.06
- SS for Type: 27.00 + 0.36 = 27.36
- SS for Interaction: 0.04 + 0.12 = 0.l6
- Same as for formulae because contrasts are orthogonal

3) 2 way ANOVA as comparisons between models

- The factor effects version of the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad \epsilon_{ijk} \text{ iid } N(0, \sigma^2)$$

where $\Sigma_i \alpha_i = \Sigma_j \beta_j = \Sigma_i (\alpha\beta)_{ij} = \Sigma_j (\alpha\beta)_{ij} = 0$
and $i = 1, \ldots, a;\ j = 1, \ldots, b;\ k = 1, \ldots, n_{ij}$

  - μ = overall mean response $(= \mu_{..})$
  - α_i = effect of level i of factor A $(= \mu_{i.} - \mu_{..})$
  - β_j = effect of level j of factor B $(= \mu_{.j} - \mu_{..})$
  - $(\alpha\beta)_{ij}$ = interaction $(= \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})$

- Relationships to cell means model:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$
$$\mu_{i.} = \mu + \alpha_i + \Sigma_j \beta_j / b + \Sigma_j \alpha\beta_{ij} / b$$
$$\mu_{.j} = \mu + \beta_j + \Sigma_i \alpha_i / a + \Sigma_i \alpha\beta_{ij} / a$$
$$\mu_{..} = \mu + \Sigma_i \alpha_i / a + \Sigma_j \beta_j / b + \Sigma_{ij} \alpha\beta_{ij} / (ab)$$

Two way ANOVA as comparisons between models

- Factor effects model is much more popular
  than the cell means model
- Lots of parameters: 1 μ, 2 α's, 3β's, and 6 $(\alpha\beta)$'s
- total of 12 parameters for fixed effects + 1 for σ^2
- only 7 sufficient statistics: 6 cell means + MSE
- Find solution by using generalized inverse (SAS)
  or imposing a restriction on the parameters to create a full-rank X matrix (R)

Two factor study - interactions

- $(\alpha\beta)_{ij}$ is an interaction effect
  - Additive model: when $(\alpha\beta)_{ij} = 0$ for all i, j
    E $Y_{ijk} = \mu + \alpha_i + \beta_j$
    "effect" of factor A same at all levels of B
  - $(\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$ is the difference between the mean for factor levels i, j and what would be expected under an additive model
  - When interactions present:
    - the effect of factor A is not the same at every level of factor B
    - the effect of factor B is not the same at every level of factor A
  - can see interactions by plotting mean Y vs factor A and connect points at the same level of factor B
  - can also see interactions in tables of means
    look at differences between trts

3) Two factor study - SS as model comparison

- Q2, Q3, and Q4 answered by comparing models
- Easiest to see when use sum-to-zero constraints
- So:
  - $\Sigma_i \alpha_i = 0$
  - $\Sigma_j \beta_j = 0$
  - $\Sigma_j \alpha\beta_{ij} = 0 \; \forall i$
  - $\Sigma_i \alpha\beta_{ij} = 0 \; \forall j$
- In this case, mapping between cell means and factor effects models (slide 80) simplifies to:

$$\begin{aligned} \mu_{..} &= \mu \\ \mu_{i.} &= \mu + \alpha_i \\ \mu_{.j} &= \mu + \beta_j \\ \mu_{ij} &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} \end{aligned}$$

3) Two factor study - SS as model comparison

- But, principle applies to any choice of non-full rank or constrained full rank $\boldsymbol{X}$ matrices
- Q2: Are media / sex / amount means equal?
  - Marginal means: H0: $\mu_{A.} = \mu_{B.}$
  - Factor effects: $\mu_{A.} = \mu + \alpha_A$, so
    H0: $\alpha_A = \alpha_B = 0$ (why = 0?)
  - Full model: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  - Reduced: $Y_{ijk} = \mu + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  - Diff. in SS = SS_A, df_A = (a-1)
- Q3: Price / type / type means: Similar, dropping β_j
- Q4: Interactions: Similar, dropping $(\alpha\beta)_{ij}$

Factorial designs: SS as model comparison

- One detail:
  - Interactions: pair of models clear with interactions vs. additive model
  - Full model: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  - Reduced: $Y_{ijk} = \mu + \alpha_i + \beta_j \qquad\qquad + \epsilon_{ijk}$
- But, which pair of models to test Factor A?

Factorial designs: SS as model comparison

- 1) Start without interactions or B
  - Full model: $Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk}$
  - Reduced: $Y_{ijk} = \mu + \epsilon_{ijk}$
- 2) Start without interactions
  - Full model: $Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$
  - Reduced: $Y_{ijk} = \mu + \beta_j + \epsilon_{ijk}$
- 3) Start with everything except α's
  - Full model: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  - Reduced: $Y_{ijk} = \mu + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
- similar issue with main effect of B

Factorial designs: SS as model comparison

- These model comparisons have names:
  1) is Type I SS (sequential SS).
  2) is Type II SS.
  3) is Type III SS (partial SS)
- when equal sample sizes, $n_{ij} = n$,
  choice doesn't matter
- when design is unbalanced, these are NOT the same.
- In general, prefer (US usage) partial SS = Type III
- My justification for this:
  Type III SS correspond to contrasts among cell means
  Other approaches imply that factors are real things

Example 1 continued

- Test H_0: no interaction

Model	terms	df	SS
red.	sex type	146	194.72
full	sex type sex*type	144	194.56
diff.		2	0.16

- Same result as with other methods

Example 1 continued

- Test H_0: no main effect of sex or H_0: no main effect of type
  - Concept: compare E $Y_{ij} = \alpha_i + \beta_j + \alpha\beta_{ij}$ to E $Y_{ij} = \beta_j + \alpha\beta_{ij}$
  - But, can't do by fitting models because column space of $\alpha\beta_{ij}$ includes column space of α_i.
  - Need to use a $C\beta$ test of $\alpha_i + \sum_j \alpha\beta_{ij}/b = 0$ in general, or
  - $C\beta$ test of $\alpha_i = 0$ if sum-to-zero constraints.
- Results are exactly same as earlier

Example 1 continued

- All 3 types of SS give exactly the same answer because balanced design

	Model		
Type	Full	Red.	SS
I	type	Intercept	27.36
II	type sex	sex	27.36
III	type sex type*sex	type sex	27.36

- I use this as a quick check for a balanced design.
- If I expect balanced data, but type I SS not equal to type III, then something is wrong
  - Data set isn't complete
  - Data set not read correctly
- Easy in SAS because default provides both type I and III SS
- Need to compute both separately when using R

Example 2 continued

- Example 2 has unequal sample sizes. Choice of SS type matters.
- SS for type of protein

	Model		
Type	Full	Reduced	SS
I	type	Intercept	453.01
II	type amount	amount	358.27
III	type amount type*amount	type amount	337.56

- No general pattern to change in SS
- Changing type of SS can increase or decrese SS

R code for example 1

```
food <- read.csv('food.csv', as.is=T)
food$type.f <- factor(food$type)
food$sex.f <- factor(food$sex)

# can fit using lm, but more helper functions
#   available with aov()
diet.aov <- aov(y ~ type.f + sex.f + type.f:sex.f,data=food
#   note : specifies the interaction
#   also have all the usual lm() helper functions

# a shortcut  * specifies all main effects and interaction
diet.aov <- aov(y ~ type.f*sex.f,data=food)
#   equivalent to first model
```

R code for example 1, cont.

```
anova(diet.aov)
#    gives sequential (type I) SS
#    but same as type III for balanced data

model.tables(diet.aov)
#    tables of means
```

R code for example 2

```
rat <- read.csv('ratweight.csv',as.is=T)
rat$amount.f <- factor(rat$amount)
rat$type.f <- factor(rat$type)

replications(rat)
#  gives number of replicates for each factor

table(rat$amount,rat$type)
#  2 x 3 table of counts for each treatment

rat.aov <- aov(gain ~ amount.f*type.f,data=rat)
#  BEWARE: type I (sequential SS)
```

R code for example 2, cont.

```
# to get type III SS, need to declare othogonal contrasts
#   can do that factor by factor, but the following does it
options(contrasts=c('contr.helmert','contr.poly'))
#  first string is the contrast for unordered factors
#  the second for ordered factors

rat.aov2 <- aov(gain ~ amount.f*type.f,data=rat)

drop1(rat.aov2,~.)
#  drop each term from full model => type III SS
#  second argument specifies all terms
```

R code for example 2, cont.

```
drop1(rat.aov, ~.)
#  rat.aov() was fit using default contr.treatment
#  very different and very wrong numbers if
#   forget to use an orthogonal parameterization

# getting marginal means is gruesome!
# model.tables() gives you the wrong numbers
# They are not the lsmeans and not the raw means
# I haven't taken the time to figure out what they are
```

R code for example 2, cont.

```
# easiest way I know is to fit a cell means model
#  and construct your own contrast matrices

rat.aov3 <- aov(gain ~ -1 + amount.f:type.f, data=rat)
# a cell means model (no intercept, one X column
#   for each combination of amount and type
coef(rat.aov3)

# There is at least one R package that tries to
#  calculate lsmeans automatically, but I know
#  one case where the computation is wrong.
#  (but appears correct).
```

Factorial designs: the good, the bad, and the ugly

- We have seen balanced data (almost always equal n per treatment)
- and unbalanced data (different n's)
- Balanced data is easy to analyze, in SAS or R (the good)
- Unbalanced data is as easy in SAS, requires hand work in R (the bad)
  no new concepts
- There is also ugly data, e.g.: sample sizes per treatment of

	1	2	3
A	10	10	10
B	10	10	0

- Often called missing cells. There is no data for the B3 cell.

Missing cells

- The entire analysis structure collapses. If there is one observation in the B3 cell, can estimate the cell mean and compute the marginal means and tests.
- Without any obs in B3, have no marginal mean for row B or for column 3.
- SAS is misleading:
  - prints type III SS and tests, but main effect tests are wrong.
  - the interaction test is valid, but it evaluates the only piece of the interaction that is testable, that in columns 1 and 2. Has 1 df.
  - LSMEANS for B and 3 are labelled non-est (non-estimable).
- My quick diagnosis for missing cells:
  - Fit a model including the highest possible interaction
  - Check d.f. of that interaction.
  - Should be product of main effect df.
  - If less, you have missing cells

Missing cells

- Missing cells arise quite naturally
- study of effectiveness of two types of grinding pad (A and B)
  - 4 grinding times: no grinding, 5 min, 10 min, and 20 min.
  - 7 trts, 10 replicates per trt:

Pad	—	A	A	A	B	B	B
Time	0	5	10	20	5	10	20

  - Pad is irrelevant if no grinding (0 time)
- I've seen this cast as a 3 x 4 factorial:

	Time			
Pad	0	5	10	20
None	10			
A		10	10	10
B		10	10	10

- Serious missing cell issues!

Missing cells

- I've seen this cast as a 2 x 4 factorial
  - Randomly divide the "time = 0" obs into a A/0 group and a B/0 group:

	Time			
Pad	0	5	10	20
A	5	10	10	10
B	5	10	10	10

  - Sometimes explicitly done with 0 time and two "pads", so 20 replicates of time 0
  - No missing cells, but if there is a pad effect there will be an interaction!
  - no difference between pads at time=0; some difference at other times.

Missing cells

- Best approach is to consider this as a collection of 7 treatments, do 1-way ANOVA, and write contrasts to answer interesting questions, e.g.
  - What is the difference between no grinding and the other 6 trts
  - When ground (time $>$ 0), what is the average difference between A and B?
  - When ground, what is the effect of grinding time?
  - When ground, is there an interaction between pad type and time?
- In other words, answer the usual 2-way ANOVA questions using contrasts where they make sense, and answer any other interesting questions

Missing cells

- Or, do something else relevant and interesting, e.g.
  - Which combinations of pad and time are significantly different from the control (use Dunnett's mcp)
  - Is the slope of Y vs time for pad A (or B) significantly different from 0?
  - Is the slope of Y vs time the same for pad A and pad B?
- In other words, think rather than do the usual.
- Missing cells are only a problem when model includes interactions
- No problem analyzing with an additive effects (no interaction) model
- Still need to think hard.
  - In the example, how do you code the None/0 treatment?
  - If None/0, Pad = None confounded with time=0

Factorial designs: relation to regression

- Can write either cell means or factor effects model as a regression model $\mathbf{Y} = X\beta + \varepsilon$
- Illustrate with factor effects model
- Example: $a = 2, b = 3, n = 2$

$$\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{13} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{23} \end{pmatrix} + \begin{pmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{pmatrix}$$

Factorial designs: relation to regression

- But this model is overparameterized
  (the $\boldsymbol{X}$ matrix is not of full rank)
- E.g., col 2 + col 3 = col 1,
  col 7 + col 8 + col 9 = col 2, etc.
- Can recode columns to turn into full rank $\boldsymbol{X}$

Factorial designs: relation to regression

- Rewrite factor effects model ($a = 2, b = 3, n = 2$) as full rank regression model using sum-to-zero coding

$$\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \end{pmatrix} + \begin{pmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{pmatrix}$$

Factorial designs: relation to regression

- Other parameters determined by sum to zero constraints in the model, e.g:
  $\alpha_2 = -\alpha_1,\ \beta_3 = -\beta_1 - \beta_2,\ (\alpha\beta)_{21} = -(\alpha\beta)_{11}$
- Other choices of contraints give diff. $\boldsymbol{X}$
- Cell means model can be written in regr. form
  Full rank: 6 cell means, 6 parameters
- Choice of $\boldsymbol{X}$ is arbitrary. Does it matter?
  $\hat{\alpha}_i$ depends on choice of const., not estimable
  $\hat{\mu} + \hat{\alpha}_i$ does not! estimable

Factorial designs: computing

- Different programs do things differently
- R:
  - contrast attribute of a factor specifies columns of the $\boldsymbol{X}$ matrix
  - number of and which interaction columns depend on which main effects included
  - If both main effects: $\sim$ A + B + A:B:
    interaction has $(a-1)(b-1)$ columns
    each is product of an A column and a B column
  - If only A: $\sim$ A + A:B (so B nested in A):
    interaction has $a(b-1)$ columns
  - If only B: $\sim$ B + A:B (so A nested in B):
    interaction has $(a-1)(b)$ columns
  - I don't know R determines this. Not documented (to my knowledge).

Factorial designs: computing

- R (cont.):
  - Default contrasts are “treatment” contrasts: set first level of each factor as the reference level
  - Focuses on ANOVA as regression.
  - “Contrasts” are not contrasts among cell means
  - They are columns of the $\boldsymbol{X}$ matrix in a regression
  - Estimates of coeffients are easy.
  - Marginal means are not. Require hand computing.
  - Details depend on the choice of contrasts. Need to be very aware of linear models theory.

Factorial designs: computing

- SAS:
  - Uses non-full rank (overparameterized) X matrices
  - and generalized inverses to find solutions
  - Very logical approach (to me).
    - If columns of $\boldsymbol{X}$ representing main effect of A are omitted, columns space of AB interaction automatically includes the column space of A.
    - AB interaction automatically "picks up" effects in the column space of A
    - Which is what you want if A nested in B
  - Marginal means are trivial (LSMEANS statement).
  - Contrasts really are contrasts among cell means

Factorial designs: SAS

- values in “solutions” output equivalent to “set last level to zero” constraint
- Estimates / Contrasts among marginal means are trivial (ESTIMATE and CONTRAST statements).
  - SAS automatically “pushes” interaction components onto marginal means
  - e.g. LSMEANS amount 1 -1; automatically includes the sum of appropriate interaction terms
  - LSMEANS amount 1 -1;, which looks like $\alpha_1 - \alpha_2$ is interpreted as $\alpha_1 + \Sigma_j \alpha\beta_{1j}/b - (\alpha_2 + \Sigma_j \alpha\beta_{2j}/b)$
  - New LSMESTIMATE statement in mixed and glimmix simplifies estimating simple effects.
- “model comparison” Type III SS are actually computed by $C\,\beta$ tests on appropriate sets of coefficients

Philosophy/history underlying ANOVA computing

- Some history
  - ISU had the first “statistical laboratory” in the US.
  - to help (mostly biological) researchers with statistical analysis
  - Emphasized the “treatments are real; factors are not” approach
  - Gertrude Cox hired away from ISU to found NCSU Dept. of Statistics
  - NCSU hires Jim Goodnight as a faculty member in 1972
  - In early 70’s, ANOVA computing was all specialized routines
  - Jim has inspiration for “general linear model:” (GLM)
    - NSF grant to develop SAS and PROC GLM
    - emphasized the “treatments are real; factors are approach”
    - i.e. Type III SS and non-full rank $\boldsymbol{X}$ matrices

Philosophy/history underlying ANOVA computing

- SAS became extremely successful!
- Was first general purpose ANOVA software for unbalanced data
- late 1970's: Jim forced to choose between being CEO of SAS or faculty member at NCSU
  Resigns from NSCU to be CEO of SAS.
- SAS is also an incredibly powerful data base manager
- Many businesses use SAS only for that capability
- Jim now "richest man" in NC
- Hard to write extensions to SAS procs
  - There is a matrix manipulation language (PROC IML) but I find R easier
  - And a macro facility for repetitive computations
  - But, R is much easier to cutomize

Philosophy / history underlying ANOVA computing

- British tradition dominated by John Nelder
- GENMOD program is the British equivalent to SAS
  - emphasizes sequential SS, even in unbalanced ANOVA
  - $\boldsymbol{X}$ matrices constructed by constraints on parameters
  - Nelder wrote the R code for linear models, lm()
  - So R takes a sequential approach with constraints
- In my mind, this makes it difficult to construct appropriate tests in unbalanced data, extract marginal means, or construct contrasts among marginal means.
- BUT, that's just my opinion. You can do all the above if you know what you're doing and are willing to code it. SAS just makes it easy.
- For graphics, programming, and regression, R is better.

Factorial designs: after the F test

- Main effects and simple effects
- Main effect = diff. between marginal means
- Simple effect = diff. between levels of one factor at a specific level of the other factor
  e.g. diff between media in price 1 = $\mu_{1A} - \mu_{1B}$
- No interaction = equal simple effects
- If no interaction, have two est.'s of $\mu_{1A} - \mu_{1B}$
  - simple effect: $\hat{\mu}_{1A} - \hat{\mu}_{1B}$
  - main effect: $\hat{\mu}_{.A} - \hat{\mu}_{.B}$ (more precise)
- If you can justify no interaction, use main effect as estimate of each simple effect.

Factorial designs: Interpretation of marginal means

- Interpretation of marginal means very straightforward when no interaction
  - F tests of each factor:
    is there a difference (effect) of that factor either averaged over other factor **or** at each level of other factor
  - Diff. (contrasts) in marginal means:
    est. of diff or contrast on average or at each level of other factor

Factorial designs: Estimation

- The F test is just the beginning, not the end
- attention usuually focused on marginal means
  - are averages over levels of the other factor
  - Differences in marginal means = differences in averages
- Preferred estimates of cell and marginal means:
  - cell means: $\hat{\mu}_{ij} = \Sigma_k Y_{ijk}/n$
  - marginal means for A: $\hat{\mu}_{i.} = \Sigma_j \hat{\mu}_{ij}/J$
  - marginal means for B: $\hat{\mu}_{.j} = \Sigma_i \hat{\mu}_{ij}/I$
  - overall mean: $\hat{\mu}_{..} = \Sigma_{ij} \hat{\mu}_{ij}/(IJ)$

Factorial designs: Estimation

- These means have different se's (equal $n_{ij} = n$), $s = \sqrt{MSE}$
  - se $\hat{\mu}_{ij} = s/\sqrt{n}$
  - se $\hat{\mu}_{i.} = s/\sqrt{nJ}$
  - se $\hat{\mu}_{.j} = s/\sqrt{nI}$
  - se $\hat{\mu}_{..} = s/\sqrt{nIJ}$
  - se $[(\hat{\mu}_{A1} - \hat{\mu}_{A2}) - (\hat{\mu}_{B1} - \hat{\mu}_{B2})] = 2s/\sqrt{n}$
- Note: estimates of marginal means are more precise
  Especially if there are many levels of the other factor
- hidden replication:
  estimate of A marginal mean benefits from J levels of B
- estimates of interaction effects are less precise
- Tests of main effects more powerful
- Interaction tests least powerful

Estimation

- Another interpretation of the difference between two lsmeans
- Example is the difference between high and low amounts in the rat study
  - Estimate the three simple effects (high-low for beef, high-low for cereal, and high-low for pork).
  - Average these three simple effects.
  - Result is the difference in marginal means.

Other types of means for main effects

- There are two other ways to define a marginal mean
- "One-bucket" or "raw" means
  - Ignore the other factor, consider all observations with amount = high as one bucket of numbers and compute their average.
  - Why is this not the same as the lsmean in unbalanced data
  - Look at the sample sizes for the rat weight study:

	beef	cereal	pork
high	7	10	10
low	10	10	8

  - part of the amount effect "bleeds" into the type effect
  - because the beef average is 41% high, the cereal average is 50% high and the pork average is 55% high
  - Very much a concern if there is a large effect of amount

Other types of means for main effects

- A third type of marginal mean arises if you drop the interaction term from the model.
- Model now claims the population difference between high and low amounts is **the same** in all three types.
- Have three estimates of that population effect (in beef, in cereal, and in pork)
- The marginal difference is a weighted average of those estimates
- weights proportional to 1/variance of estimate
- That from cereal gets a higher weight because larger sample size
- Details in example
- Sounds appealing
  - More precise than the lsmean
  - why compute the marginal mean unless there is no interaction?

Other types of means for main effects

- But, US tradition, especially US land grant / ag tradition, is to use lsmeans
  - simple average may make sense whether or not there is an interaction
  - hypothesis being tested by lsmeans depends on the population means.
  - hypothesis tested by other (raw or weighted) means includes population means and the sample sizes (details below).
  - Including sample sizes in a hypothesis is wierd.

Example 2 continued

- What is the difference in mean weight gain between the high and low amounts?
- The cell means:

	beef	cereal	pork
high	98.43	85.9	99.5
low	79.20	83.9	82.0
diff	19.23	2.0	17.5

- LSMEANS (Type III) approach: (19.23 + 2.0 + 17.5)/3 = 12.91
  se = 0.2747 s_p

Example 2 continued

- Inv. Var. weighted mean:
  - A weighted mean is $\frac{w_1\overline{Y}_1+w_2\overline{Y}_2+w_3\overline{Y}_3}{w_1+w_2+w_3}$
  - The variances of the simple effects are: (1/7 + 1/10)σ^2, (2/10)σ^2, and (1/10 + 1/8)σ^2
  - Their inverses are (to 4 digits and ignoring σ^2 term, which cancels out): 4.1176, 5, 4.4444
  - So the Type II estimate of the difference is (4.1176*19.23 + 5 * 2.0 + 4.4444*17.5)/(4.1176 + 5 + 4.4444) = 12.31
    se = 0.2554 s_p
- The "one bucket" difference is:
  (7*98.42 + 10*85.9 + 10*99.5)/(7+10+10) - (10*79.2 + 10*83.9 + 8*82.0) = 12.50
  se = 0.2697 s_p

Why are the estimates the same when balanced?

- In example 1, all cells have $n = 25$.
- So the variances of each simple effect are 2/25, 2/25, 2/25
  The type II estimate is equally weighted
- The type I estimate is
  $(25*Y_{11}+25*Y_{12}+25*Y_{13})/75-(25*Y_{21}+25*Y_{22}+25*Y_{33})/75 =$
  $(1/3)*(Y_{11}-Y_{21})+(1/3)*(Y_{12}-Y_{22})+(1/3)*(Y_{13}-Y_{23})$
  Also an equally weighted average of the simple effects
- Balanced data are nice!
- But, unbalanced data often occur and you have to be able to handle that

Connections between type of SS and definition of marginal mean

- The F test using type III SS tests equality of lsmeans (equally weighted average of cell means).
- The F test using type I SS for the first factor in the model tests equality of raw means
  - This represents a combination of the effect of interest (e.g. type) and some bit of other effects (e.g. amount)
  - From a "treatments are real" perspective, the null hypothesis depends on the number of each treatment.
- The F test using type II SS tests equality of the inverse variance weighted average.
  - Again, the null hypothesis depends on the sample size for each treatment.

Factorial designs: sample size

- sample size can be statistically determined by se., confidence interval width, or power.
- power by far the most common
- Dr. Koehler emphasized non-central T and F distributions
- Here's another approach that provides a very good approximation and more insight.
- The non-central T distribution with n.c.p. of δ/σ is closely approximated by a central T distribution centered at δ/σ (the shifted-T distribution).
- I'll draw some pictures to motivate an approximate relationship between δ: the pop. diff. in means, *s.e.*: the pop. s.e. for the quantity of interest, α: type 1 error rate, $1-\beta$: the power.

$$\delta = (T_{1-\alpha/2,df} + T_{1-\beta,df})s.e. \tag{4}$$

Factorial designs: sample size

- Details of the s.e. depend on what difference is being considered and the trt design
  - e.g. for the difference between two marginal means averaged over three levels of the other factor, se $= \sigma\sqrt{2/(3n)}$, where n is the number of observations per cell.
  - So, if $\sigma = 14$ and $n = 10$, df = 54, and an $\alpha = 0.05$ test has 80% power to detect a difference of $(2.0049 + 0.8483) * (2 * 14/\sqrt{30}) = 10.3$.
- Solving equation (4) for n gives

$$n = \left(T_{1-\alpha/2,df} + T_{1-\beta,df}\right)^2 k^2 \left(\frac{\sigma}{\delta}\right)^2, \qquad (5)$$

where k is the constant in the s.e. formula. $k = \sqrt{2/3}$ for this problem

Factorial designs: sample size

- what n is necessary to have 80% power to detect a difference of 15?
  - df depend on n, so need to solve iteratively
  - I start with T quantiles of 2.00 and 0.85, approximately 60 df
  - $n = (2.85)^2(2/3) * (14/15)^2 = 4.7$, i.e. 5
  - n = 5 has error df = 24, Those T quantiles are 2.064 and 0.857 (approx.)
  - $n = (2.064 + 0.857)^2(2/3) * (14/15)^2 = 4.95$, i.e. n=5.
- Algorithm usually converges in 2 or 3 steps

Power of ANOVA tests

- What n is needed for 80% power for various ANOVA comparisons?
- Example: 2x2, δ=0.1, σ=0.5
- Row main effect

A		4.0	N=196
B		4.1	

- Row simple effect

A		4.0	N=392
B		4.1	

- Interaction

A	3.0	3.0	N=784
B	3.0	3.1	

Power of ANOVA tests

- The interaction line “looks” just like a main effect line in the ANOVA table.
- But, the power of that interaction test is much lower, because the s.e. of the interaction effect is much larger
- If you’re designing a study to examine interactions, you need a much larger study than if the goal is a main effect

Factorial designs: after the F test

- a-priori comparisons between marginal means
  - use contrasts between marginal means
  - usu. no adj. for multiple comp.
- post-hoc or large number of comparisons
  - adjust for multiple comp. in usual ways
    - Tukey: all pairs of marginal means
      # groups = number of marginal means
      may differ for each factor
    - Scheffe: all linear contrasts
    - Bonferroni: something else
  - Or use specialized methods for other comp.
    - Dunnett: compare many trt. to one control
    - Hsu: compare many to best in data
    - Monte-Carlo: tailor to specific family

Factorial designs: after the F test

- What if there is evidence of an interaction?
- Either because expected by science, or data suggests an interaction (F test of AB signif.).
- Usual 2 way decomposition not as useful: main effect, $\mu_{.1} - \mu_{.2}$, does not estimate simple effect, $\mu_{A1} - \mu_{A2}$.
- Are you measuring the response on the right scale?
  - Transforming Y may eliminate the interaction
  - log CFU for bacteria counts, pH for acidity
  - Works for quantitative interaction, not qualitative

Factorial designs: after the F test

- 3 approaches when there is an interaction:
  - dogma: marginal means and tests are useless
    - focus on cell means
    - split data into groups, (e.g. two 1-way ANOVA's, one for each sex or a t-test for each type of food supplement)
    - slicing: same idea, using MSE est. from all obs.
  - think (1): marginal mean = ave. simple effect
    is this interpretable in this study?
  - think (2): why is there an interaction?
    are effects additive on some other scale?
    transform responses so effects are additive

Factorial designs: model diagnostics

- Same as in earlier ANOVA/regression models
- Residuals are $e_{ijk} = Y_{ijk} - \bar{Y}_{ij\cdot}$
- Check for independence (crucial): e.u. = o.u.?
- Check for constant variance: (plot/test resids vs pred. or vs A, vs B)
- Check for normal errors. Normality is least important
- Remedies - transformation (common), weighted least squares
- Transformation changes both error properties and the model.
  - Can eliminate or introduce interactions.

Factorial designs: randomized block design

- Reminder. An experiment has:
  - Treatment design (or treatment structure): what is done to each e.u., e.g. 2 way factorial.
  - Experimental design: (CRD, RCB, ??) how treatments are assigned to e.u.'s
- Can perform the two factor study in blocks i.e. repeat full factorial experiment ($r = IJ$ treatments) in each block
- Assume no block and treatment interactions

Factorial designs: randomized block design

- ANOVA table: combines blocking ideas and 2-way trt design

source of variation	degrees of freedom	
block	$n-1$	
treatments	$ab-1$	
factor A		$a-1$
factor B		$b-1$
interaction AB		$(a-1)(b-1)$
error	$(ab-1)(n-1)$	
total	$abn-1$	

- One variation you may encounter:
  - Block*treatment SS (and d.f.) can be partitioned:

$$SS_{Block*Trt} = SS_{Block*A} + SS_{Block*B} + SS_{Block*AB}$$

  - which leads to the following ANOVA table:

Factorial designs: randomized block design

source of variation	degrees of freedom	
block	$n-1$	
treatments	$ab-1$	
factor A		$a-1$
factor B		$b-1$
interaction AB		$(a-1)(b-1)$
block*trt	$(n-1)(ab-1)$	
error for A=block*A		$(a-1)(n-1)$
error for B=block*B		$(b-1)(n-1)$
error for AB=block*AB		$(a-1)(b-1)(n-1)$
total	$abn-1$	

- When subdivided, the appropriate F test for A is $MS_A/MS_{block*A}$
- The F test for B is $MS_B/MS_{block*B}$
- And the F test for AB is $MS_{AB}/MS_{block*AB}$

Factorial designs: subdividing errors

- I don't like this, at least for an experimental study.
- A*B treatments are randomly assigned to e.u.'s.
  block*trt is a measure of variability between e.u.'s.
  - Why should block*A error differ than block*B error?
  - What is magic about A? It may represent multiple contrasts, so why not divide further into block * 1 d.f. contrasts?
- One extreme example had ca. 30 error terms, each 1 d.f.
- Tests using small error d.f. are not powerful.
- Best to think: what is appropriate to pool?
  Is there any subject-matter reason to believe that $MS_{block*A}$ differs from $MS_{block*B}$.
- In an observational study, think hard, because you don't have randomization to help.

Factorial designs: More than two factors

- Introduce factor C with c levels, only one new concept
- ANOVA table

source of variation	degrees of freedom	sums of squares
factor A	$a-1$	$nbc\Sigma_i(\bar{Y}_{i\cdots}-\bar{Y}_{\cdots})^2$
factor B	$b-1$	$nac\Sigma_j(\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdots})^2$
factor C	$c-1$	$nab\Sigma_k(\bar{Y}_{\cdot\cdot k}-\bar{Y}_{\cdots})^2$
interaction AB	$(a-1)(b-1)$	$nc\Sigma_i\Sigma_j(\bar{Y}_{ij\cdot}-\bar{Y}_{i\cdots}-\bar{Y}_{\cdot j\cdot}+\bar{Y}_{\cdots})^2$
interaction AC	$(a-1)(c-1)$	$nb\Sigma_i\Sigma_k(\bar{Y}_{i\cdot k}-\bar{Y}_{i\cdots}-\bar{Y}_{\cdot\cdot k}+\bar{Y}_{\cdots})^2$
interaction BC	$(b-1)(c-1)$	$na\Sigma_j\Sigma_k(\bar{Y}_{\cdot jk}-\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdot\cdot k}+\bar{Y}_{\cdots})^2$
interaction ABC	$(a-1)(b-1)(c-1)$	$SS(ABC)$
error	$abc(n-1)$	$\Sigma_i\Sigma_j\Sigma_k\Sigma_l(Y_{ijkl}-\bar{Y}_{ijk\cdot})^2$
total	$abcn-1$	$\Sigma_i\Sigma_j\Sigma_k\Sigma_l(Y_{ijkl}-\bar{Y}_{\cdots})^2$

Factorial designs: More than two factors

- Effects are averages over everything omitted from that term.
  - F test for A compares averages for each level of A, averaged over all levels of B, all levels of C, and all replicates
  - F test for AB is average over levels of C and reps
- Contrasts also are straightforward extensions
- new concept: what is the ABC interaction?
- Reminder: AB interaction when effect of B depends on level of A, here averaged over all levels of C
- ABC interaction: effect of AB interaction depends on level of C

Example of a 3 way interaction

- Consider a complete 2 x 2 x 2 treatment design
- The population means are:

	C=1		C=2	
A	B=1	B=2	B=1	B=2
1	9.3	7.3	9.3	9.2
2	8.3	6.3	8.3	10.2

- For C=1, the interaction effect is (9.3 - 8.3)-(7.3-6.3) = 0
  For C=2, the interaction effect is (9.3-8.3) - (9.2-10.2) = 2
- The AB interaction effect is different in the two levels of C, so there is a ABC interaction
- The AB interaction is that between A and B, ave. levels of C
- Here the AB interaction = 1, i.e. (0+2)/2

Factorial designs: Studies with no replication

- Suppose we have two factors (A with a levels and B with b levels) but only ab experimental units
- because limited by cost or practical constraints
- Randomized block design is an example with a design factor and a treatment factor
- If try to fit the full two factor factorial model with interactions, no df left to estimate error
- Resolution: hope for no interactions, use $MS(AB)$ to estimate σ^2
- Or, replicate some but not all treatments

Factorial designs: Studies with no replication

- ANOVA table

source of variation	degrees of freedom	sums of squares
factor A	$a-1$	$b\Sigma_i(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2$
factor B	$b-1$	$a\Sigma_j(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot})^2$
error	$(a-1)(b-1)$	$\Sigma_i\Sigma_j(Y_{ij}-\bar{Y}_{i\cdot}-\bar{Y}_{\cdot j}+\bar{Y}_{\cdot\cdot})^2$
total	$ab-1$	$\Sigma_i\Sigma_j(Y_{ij}-\bar{Y}_{\cdot\cdot})^2$

Factorial designs: no replication

- ANOVA table on previous slide
- Expected mean squares
  $E(MSA) = \sigma^2 + b\Sigma_i(\mu_{i.} - \mu_{..})^2/(a-1)$
  $E(MSB) = \sigma^2 + a\Sigma_j(\mu_{.j} - \mu_{..})^2/(b-1)$
  $E(MSE) = \sigma^2$
- Usual tests
- Model checking: assumption of additivity very important.
  - Plot obs vs one factor mean: are lines approximately parallel?
  - Tukey's test for non-additivity

Factorial designs: no repl., 2^K trts

- Assume K factors, each at two levels
- Known as 2^K factorial
- One application: factors are really continuous and we want to explore response to factors. leads to response surface designs
- Or, screening lots of 'yes/no' factors
- Some special features
  - all main effects, all interactions are 1 d.f.
  - the regression approach works nicely
    - 1 column of X for each main effect (with +1/-1 coding)
    - interaction columns by multiplication
    - all columns are orthogonal
- With replication, no new issues
- With no replication same problem as discussed previously but with some new solutions

Factorial designs: 2^K studies

- Estimating σ^2 in 2^K study without replication
  - pool SS from nonsignif factors/interactions to estimate σ^2; if we pool p terms, then

$$\hat{\sigma}^2 = (N\Sigma_q b_q^2)/p$$

    b_q is regression coefficient, $N = 2^K$= # obs.
  - normal probability plot
    rank est. coeff b_q and plot against N quantiles
    all b_q have same s.e.; if $\beta_q = 0$, $b_q \sim N(0, \sigma_b^2)$
    effects far from line are "real"
    those close to line $\rightarrow \hat{\sigma}$

Factorial designs: 2^K studies

- center point replication
  - construct one new treatment at intermediate levels of each factor - called a center point
  - take n_o observations at this new center point
  - est. σ from center points = pure error
  - can test interactions; pool with some

Factorial designs: fractional factorials

- Assume K factors, each at two levels
- Sometimes 2^K is too many treatments
- Can run 2^{K-J} fractional factorial
  (2^{-J} fraction of a full factorial)
- Can't estimate all 2^K effects
- Introduce confounding by carefully selecting those treatments to use
- Note still have problem estimating σ^2
  unless there is some replication
- Example of fractional factorial on next slide

Factorial designs: fractional factorials

- 2^K study with $K = 3$ (call factors A, B, C)
- Design matrix for full factorial regression (no rep)

obs	μ	A	B	C	AB	AC	BC	ABC
1	1	1	1	1	1	1	1	1
2	1	1	1	-1	1	-1	-1	-1
3	1	1	-1	1	-1	1	-1	-1
4	1	1	-1	-1	-1	-1	1	1
5	1	-1	1	1	-1	-1	1	-1
6	1	-1	1	-1	-1	1	-1	1
7	1	-1	-1	1	1	-1	-1	1
8	1	-1	-1	-1	1	1	1	-1

Factorial designs: fractional factorials

- Consider $J = 1$, i.e. half-fraction

obs	μ	A	B	C	AB	AC	BC	ABC
1	1	1	1	1	1	1	1	1
4	1	1	-1	-1	-1	-1	1	1
6	1	-1	1	-1	-1	1	-1	1
7	1	-1	-1	1	1	-1	-1	1

  - can't distinguish between μ and *ABC*, confounded
  - so are *A* and *BC*, *B* and *AC*, *C* and *AB*
  - significant *A* effect may actually be *BC* effect
  - use only main effects in analysis
  - very useful if no interactions
  - other half-factorials will confound different effects
  - concept extends to quarter-factorials
- Most useful when many factors, so main eff. and 2-way int's are confounded with very high order (e.g. 6-way, 5-way) int's