STAT 516 Lec 07

Unbalanced two-way factorial design

Karl Gregory

2026-10-28

Mice example

Change in mRNA expression of esr1 in wild-type vs gen-modified mice on low- and high-fat diets.

Data from Trey Hope in Dr. Enos' lab in the USC School of Medicine.

WT LFD	ADIPO-ERα LFD	WT HFD	ADIPO-ERα HFD
493	555	595.5	1153
172	302	418.5	890.5
617	434	642	539
534	90	743.5	651.5
500.5	72	1351	585
127	489	1180	783.5
1224	453	938.5	533.5
143	552.5	670	895
852	635.5	1319	579
134	330.9	1007	551
547.5	169.5	589	800
	737.5	481.5	1026
		785	308
		1060	1415
		435	342.5
		535.5	197.5

Does the diet affect the response? The genetic modification? Is the effect of the diet the same in the wild-type and the gen-modified mice?

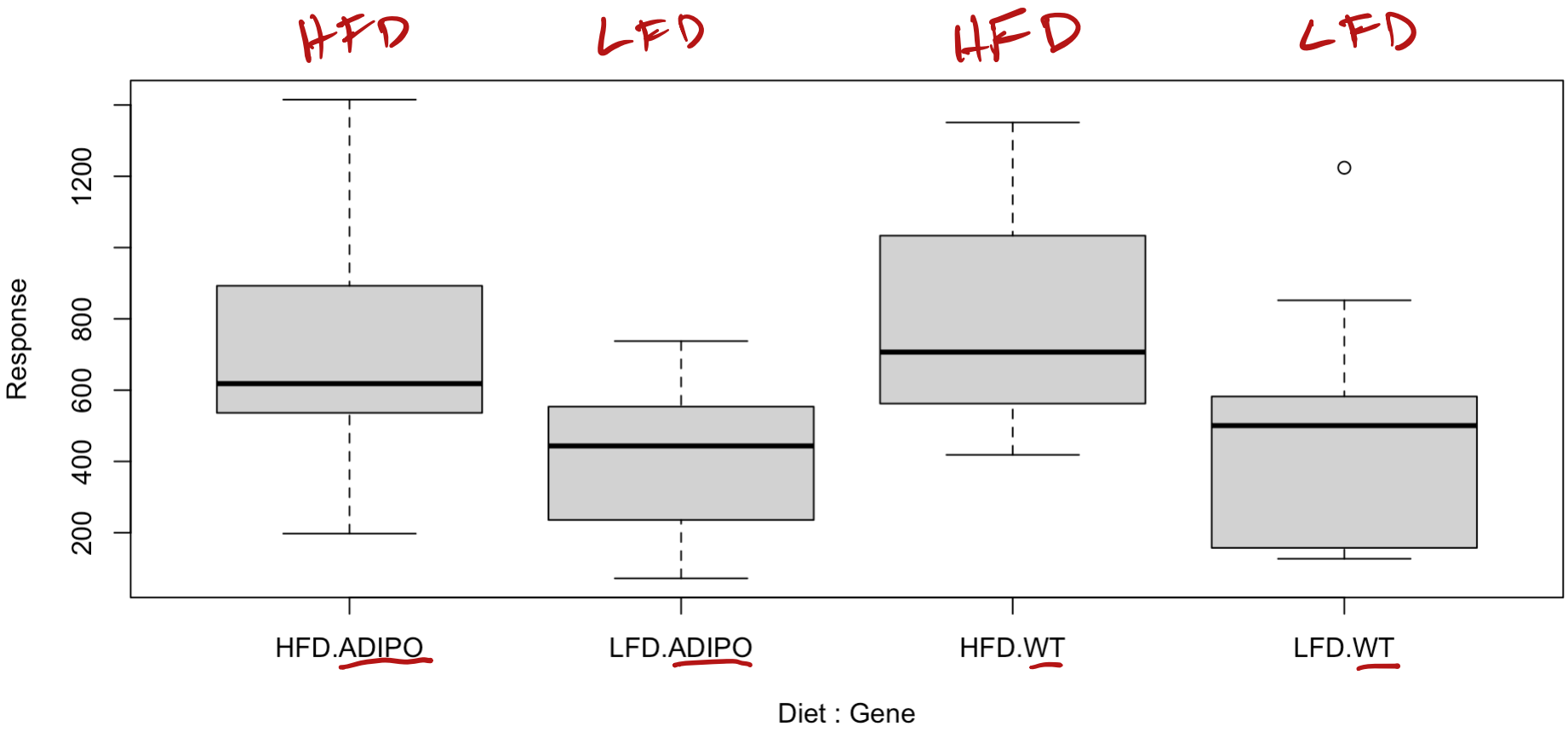
Mice

```r
link <- url("https://people.stat.sc.edu/gregorkb/data/ADIPOER.csv")
adipoer <- read.csv(link)
adipoer
```

```
   Response  Gene Diet
1       493    WT  LFD
2       172    WT  LFD
3       617    WT  LFD
4       534    WT  LFD
5       500    WT  LFD
6       127    WT  LFD
7      1224    WT  LFD
8       143    WT  LFD
9       852    WT  LFD
10      134    WT  LFD
11      548    WT  LFD
12      555 ADIPO  LFD
13      302 ADIPO  LFD
14      434 ADIPO  LFD
15       90 ADIPO  LFD
16       72 ADIPO  LFD
17      489 ADIPO  LFD
18      453 ADIPO  LFD
```

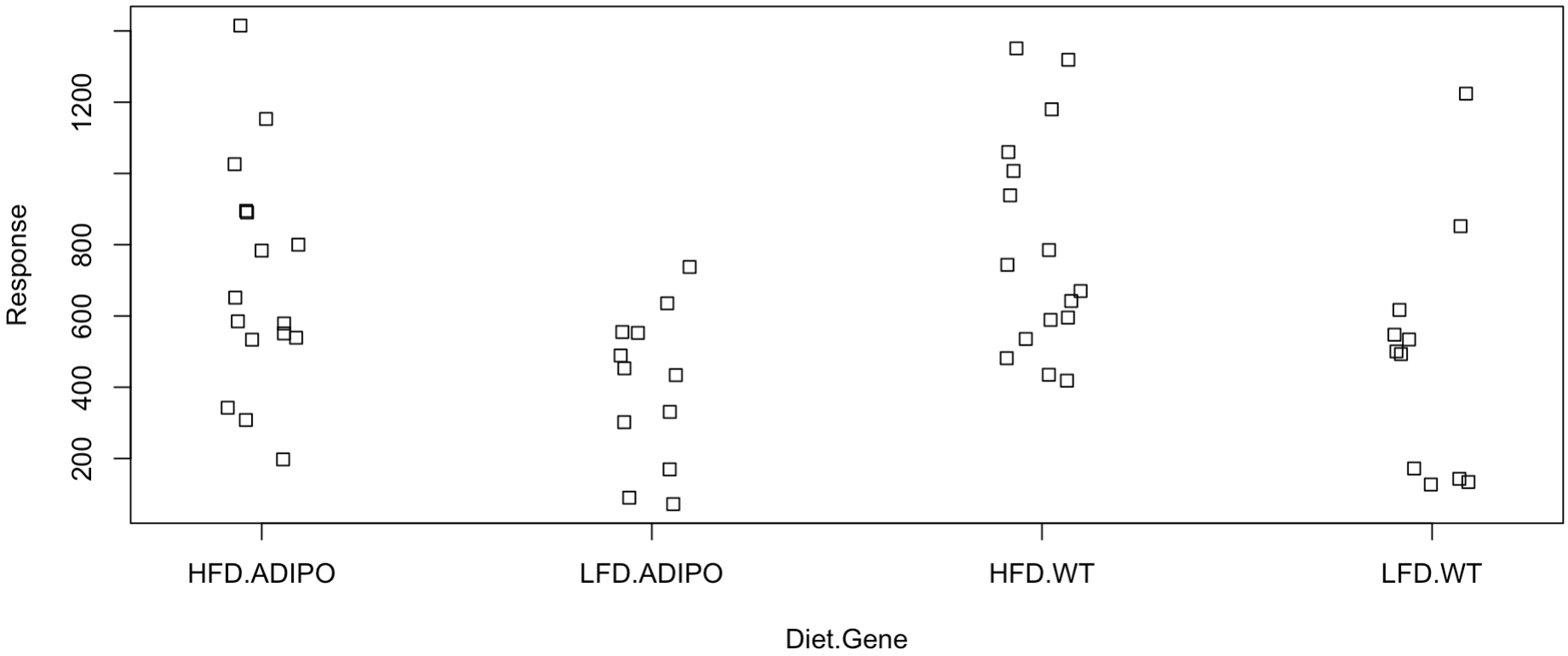
Mice

```
1 boxplot(Response ~ Diet + Gene, data = adipoer)
```



Mice

```
1 stripchart(Response ~ Diet + Gene, data = adipoer, vertical = T,
2            method = "jitter", xlab = "Diet.Gene")
```



Mice

```r
grp_means <- aggregate(Response ~ Diet + Gene, data = adipoer, mean)
grp_means
```

```
  Diet  Gene Response
1  HFD ADIPO      703
2  LFD ADIPO      402
3  HFD    WT      797
4  LFD    WT      486
```

```r
grp_counts <- aggregate(Response ~ Diet + Gene, data = adipoer, length)
grp_counts
```

```
  Diet  Gene Response
1  HFD ADIPO       16
2  LFD ADIPO       12
3  HFD    WT       16
4  LFD    WT       11
```

Two-way treatment effects model

The two-way treatment effects model assumes

$$Y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk}$$

for $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n_{ij}$, where

- Y_{ijk} is the response for EU k under level i of A and level j of B.
- μ represents a baseline or overall mean.
- The τ_i are the main effects for Factor A.
- The γ_j are the main effects for Factor B.
- The $(\tau\gamma)_{ij}$ are the interaction effects between A and B.
- The ε_{ijk} are $\mathcal{N}(0, \sigma^2)$ error terms.

Now suppose n_{ij} are *not* all the same, so the design is unbalanced.

Mice

In the mice example $a = b = 2$. We want parameter estimates $\hat{\mu}$, $\hat{\tau}_i$, $\hat{\gamma}_j$, and $\widehat{(\tau\gamma)}_{ij}$ which satisfy

$$\begin{aligned}
\hat{\mu} + \hat{\tau}_1 + \hat{\gamma}_1 + \widehat{(\tau\gamma)}_{11} &= \bar{Y}_{11\cdot} \\
\hat{\mu} + \hat{\tau}_1 + \hat{\gamma}_2 + \widehat{(\tau\gamma)}_{12} &= \bar{Y}_{12\cdot} \\
\hat{\mu} + \hat{\tau}_2 + \hat{\gamma}_1 + \widehat{(\tau\gamma)}_{21} &= \bar{Y}_{21\cdot} \\
\hat{\mu} + \hat{\tau}_2 + \hat{\gamma}_2 + \widehat{(\tau\gamma)}_{22} &= \bar{Y}_{22\cdot}
\end{aligned}$$

We have $ab = 4$ equations and $a + b + ab + 1 = 9$ unknowns.

Impose (as in balanced case) a unique solution by setting

$$\hat{\tau}_1 = 0, \quad \hat{\gamma}_1 = 0, \quad \widehat{(\tau\gamma)}_{11} = \widehat{(\tau\gamma)}_{12} = \widehat{(\tau\gamma)}_{21} = 0.$$

Mice

Then we obtain the system of equations:

$$\begin{aligned}
\hat{\mu} &= \bar{Y}_{11.} \\
\hat{\mu} + \hat{\gamma}_2 &= \bar{Y}_{12.} \\
\hat{\mu} + \hat{\tau}_2 &= \bar{Y}_{21.} \\
\hat{\mu} + \hat{\tau}_2 + \hat{\gamma}_2 + \widehat{(\tau\gamma)}_{22} &= \bar{Y}_{22.}
\end{aligned}$$

We can solve this uniquely to obtain:

$$\begin{aligned}
\hat{\mu} &= \bar{Y}_{11.} \\
\hat{\gamma}_2 &= \bar{Y}_{12.} - \bar{Y}_{11.} \\
\hat{\tau}_2 &= \bar{Y}_{21.} - \bar{Y}_{11.} \\
\widehat{(\tau\gamma)}_{22} &= \bar{Y}_{22.} - \bar{Y}_{11.} - (\bar{Y}_{12.} - \bar{Y}_{11.}) - (\bar{Y}_{21.} - \bar{Y}_{11.})
\end{aligned}$$

Mice

Construct the treatment group means from the parameter estimates:

```r
lm_out <- lm(Response ~ Diet + Gene + Diet:Gene, data = adipoer)
lm_out
```

```
Call:
lm(formula = Response ~ Diet + Gene + Diet:Gene, data = adipoer)

Coefficients:
   (Intercept)         DietLFD          GeneWT  DietLFD:GeneWT
        703.12         -301.38           93.81           -9.74
```

```r
grp_means
```

```
  Diet  Gene Response
1  HFD ADIPO      703
2  LFD ADIPO      402
3  HFD    WT      797
4  LFD    WT      486
```

Estimating the error term variance

An unbiased estimator of the error term variance σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{N-ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2,$$

where $N = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$.

Mice

The square root of the variance estimate is printed in this output:

```
1 summary(lm_out)
```

```
Call:
lm(formula = Response ~ Diet + Gene + Diet:Gene, data = adipoer)

Residuals:
   Min     1Q Median     3Q    Max
-505.6 -220.1  -11.9  170.3  738.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      703.12      75.52    9.31  1.4e-12 ***
DietLFD         -301.38     115.35   -2.61    0.012 *
GeneWT            93.81     106.80    0.88    0.384
DietLFD:GeneWT    -9.74     165.24   -0.06    0.953
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 302 on 51 degrees of freedom
Multiple R-squared:  0.229, Adjusted R-squared:  0.184
```

Sums of squares in the unbalanced design

Sum of squares	**Symbol**	**Formula**
Total	$\mathrm{SS}_{\mathrm{Tot}}$	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(Y_{ijk}-\bar{Y}_{...})^2$
Treatment	$\mathrm{SS}_{\mathrm{Trt}}$	$\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\bar{Y}_{ij.}-\bar{Y}_{...})^2$
Error	$\mathrm{SS}_{\mathrm{Error}}$	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(Y_{ijk}-\bar{Y}_{ij.})^2$

We have the decomposition $\mathrm{SS}_{\mathrm{Tot}} = \mathrm{SS}_{\mathrm{Trt}} + \mathrm{SS}_{\mathrm{Error}}$.

The Analysis of Variance (ANOVA) table

Obtain the MS values by dividing the SS values by the Df values.

Source	Df	SS	MS	F value
Treatments	$ab-1$	$\mathrm{SS}_{\mathrm{Trt}}$	$\mathrm{MS}_{\mathrm{Trt}}$	$F_{\mathrm{stat}} = \mathrm{MS}_{\mathrm{Trt}} / \mathrm{MS}_{\mathrm{Error}}$
Error	$N-ab$	$\mathrm{SS}_{\mathrm{Error}}$	$\mathrm{MS}_{\mathrm{Error}}$	
Total	$N-1$	$\mathrm{SS}_{\mathrm{Tot}}$		

In balanced case this $ab(n-1) = abn - ab = N$

Reject H_0: μ_{ij} all the same if $F_{\mathrm{stat}} > F_{ab-1,N-ab,\alpha}$.

In the unbalanced case we have $N = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$ in place of abn.

Note that $\hat{\sigma}^2 = \mathrm{MS}_{\mathrm{Error}}$.

Main effect sums of squares under unbalancedness

- In the balanced case, we decomposed $\mathrm{SS}_{\mathrm{Trt}}$ as $\mathrm{SS}_{\mathrm{A}} + \mathrm{SS}_{\mathrm{B}} + \mathrm{SS}_{\mathrm{AB}}$.
- No such "clean" decomposition of $\mathrm{SS}_{\mathrm{Trt}}$ under unbalancedness.
- Specifically, it is not obvious how to define SS_{A} and SS_{B}.
- Three versions of SS_{A} and SS_{B} measure main effect importance:

1. Sequentially : Default in R anova( )
2. After other main effect in the *absence* of interaction
3. After other main effect in the *presence* of interaction

Recommended

These are called Type I, Type II, and Type III sums of squares.

Type I/II/III SS for main effects

Each is a full/reduced model difference in SS.

To define them, we introduce the notation: Let

- $\mathrm{SS}_{\mathrm{Error}}(A)$ denote $\mathrm{SS}_{\mathrm{Error}}$ from model with only A main effects.
- $\mathrm{SS}_{\mathrm{Error}}(B)$ denote $\mathrm{SS}_{\mathrm{Error}}$ from model with only B main effects.
- $\mathrm{SS}_{\mathrm{Error}}(A, B)$ denote $\mathrm{SS}_{\mathrm{Error}}$ from model with A,B main effects .
- $\mathrm{SS}_{\mathrm{Error}}(A, B, AB)$ denote $\mathrm{SS}_{\mathrm{Error}}$ from full model.

Note that $\mathrm{SS}_{\mathrm{Tot}}$ is $\mathrm{SS}_{\mathrm{Error}}$ from the model with no effects.

Type I SS

Type I SS values are meaningless for unbalanced designs!

The Type I sums of squares $\mathrm{SS}_{\mathrm{A}}^{\mathrm{I}}$ and $\mathrm{SS}_{\mathrm{B}}^{\mathrm{I}}$ are defined as follows:

- If A enters first and then B, then

$$\mathrm{SS}_{\mathrm{A}}^{\mathrm{I}} = \mathrm{SS}_{\mathrm{Tot}} - \mathrm{SS}_{\mathrm{Error}}(A)$$
$$\mathrm{SS}_{\mathrm{B}}^{\mathrm{I}} = \mathrm{SS}_{\mathrm{Error}}(A) - \mathrm{SS}_{\mathrm{Error}}(A, B)$$

- If B enters first and then A, then

$$\mathrm{SS}_{\mathrm{B}}^{\mathrm{I}} = \mathrm{SS}_{\mathrm{Tot}} - \mathrm{SS}_{\mathrm{Error}}(B)$$
$$\mathrm{SS}_{\mathrm{A}}^{\mathrm{I}} = \mathrm{SS}_{\mathrm{Error}}(B) - \mathrm{SS}_{\mathrm{Error}}(A, B)$$

When the design is balanced the order of entrance does not matter.

Mice

Obtain the Type I SS when Gene is added and then Diet.

Fit model with no effects:

```
lm_none <- lm(Response ~ 1, data = adipoer)
SST <- sum(lm_none$residuals^2)
```

Fit model with just Gene:

```
lm_Gene <- lm(Response ~ Gene, data = adipoer)
SSE_Gene <- sum(lm_Gene$residuals^2)
SS_Gene <- SST - SSE_Gene # take difference in SSE
```

The total SS is reduced by 1.273×10^5.

Fit model with Gene and Diet:

```
lm_Gene_Diet <- lm(Response ~ Gene + Diet, data = adipoer)
SSE_Gene_Diet <- sum(lm_Gene_Diet$residuals^2)
SS_Diet <- SSE_Gene - SSE_Gene_Diet # take difference in SSE
```

The error SS is reduced by 1.253×10^6.

Mice

Obtain the Type I SS when Diet is added and then Gene.

Fit model with no effects:

```
lm_none <- lm(Response ~ 1, data = adipoer)
SST <- sum(lm_none$residuals^2)
```

Fit model with just Diet:

```
lm_Diet <- lm(Response ~ Diet, data = adipoer)
SSE_Diet <- sum(lm_Diet$residuals^2)
SS_Diet <- SST - SSE_Diet # take difference in SSE
```

The total SS is reduced by 1.27×10^6.

Fit model with Diet and Gene:

```
lm_Gene_Diet <- lm(Response ~ Gene + Diet, data = adipoer)
SSE_Gene_Diet <- sum(lm_Gene_Diet$residuals^2)
SS_Gene <- SSE_Diet - SSE_Gene_Diet # take difference in SSE
```

The error SS is reduced by 1.107×10^5.

Mice

Obtain the Type I SS on the mice data using `anova()` function.

```r
anova(lm(Response ~ Gene + Diet + Gene:Diet, data = adipoer))
```

```
Analysis of Variance Table

Response: Response
          Df  Sum Sq Mean Sq F value  Pr(>F)
Gene       1  127271  127271    1.39 0.24306
Diet       1 1253492 1253492   13.74 0.00052 ***
Gene:Diet  1     317     317    0.00 0.95325
Residuals 51 4653342   91242
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm(Response ~ Diet + Gene + Diet:Gene, data = adipoer))
```

```
Analysis of Variance Table

Response: Response
          Df  Sum Sq Mean Sq F value  Pr(>F)
Diet       1 1270105 1270105   13.92 0.00048 ***
Gene       1  110659  110659    1.21 0.27595
Diet:Gene  1     317     317    0.00 0.95325
Residuals 51 4653342   91242
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Type II SS

- Measures contribution of each main effect in *absence* of interaction.
- The Type II sums of squares $\mathrm{SS}_{\mathrm{A}}^{\mathrm{II}}$ and $\mathrm{SS}_{\mathrm{B}}^{\mathrm{II}}$ are given by

$$\mathrm{SS}_{\mathrm{A}}^{\mathrm{II}} = \mathrm{SS}_{\mathrm{Error}}(B) - \mathrm{SS}_{\mathrm{Error}}(A, B)$$
$$\mathrm{SS}_{\mathrm{B}}^{\mathrm{II}} = \mathrm{SS}_{\mathrm{Error}}(A) - \mathrm{SS}_{\mathrm{Error}}(A, B)$$

Mice

Obtain the Type II SS on the mice data.

Fit a one-way model for each factor:

```r
lm_Diet <- lm(Response ~ Diet, data = adipoer)
SSE_Diet <- sum(lm_Diet$residuals^2)
lm_Gene <- lm(Response ~ Gene, data = adipoer)
SSE_Gene <- sum(lm_Gene$residuals^2)
```

Fit a model with both main effects (but no interaction):

```r
lm_Diet_Gene <- lm(Response ~ Diet + Gene, data = adipoer)
SSE_Diet_Gene <- sum(lm_Diet_Gene$residuals^2)
SS_Diet <- SSE_Gene - SSE_Diet_Gene # take difference in SSE
SS_Gene <- SSE_Diet - SSE_Diet_Gene # take difference in SSE
```

Including Diet after Gene reduces error SS by 1.253×10^6.

Including Gene after Diet reduces error SS by 1.107×10^5.

Mice

Obtain Type II SS with `Anova()` function from R package `car`.

```r
library(car)
Anova(lm(Response ~ Diet + Gene + Diet:Gene, data = adipoer),type = "II")
```

```
Anova Table (Type II tests)

Response: Response
           Sum Sq Df F value  Pr(>F)
Diet      1253492  1   13.74 0.00052 ***
Gene       110659  1    1.21 0.27595
Diet:Gene     317  1    0.00 0.95325
Residuals 4653342 51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Type III SS

- In the unbalanced case we should use the so-called Type III SS.
- Measures contribution of each main effect in *presence* of interaction.
- The Type III sums of squares $\mathrm{SS}_{\mathrm{A}}^{\mathrm{III}}$ and $\mathrm{SS}_{\mathrm{B}}^{\mathrm{III}}$ are given by

$$\mathrm{SS}_{\mathrm{A}}^{\mathrm{III}} = \mathrm{SS}_{\mathrm{Error}}(B, AB) - \mathrm{SS}_{\mathrm{Error}}(A, B, AB)$$
$$\mathrm{SS}_{\mathrm{B}}^{\mathrm{III}} = \mathrm{SS}_{\mathrm{Error}}(A, AB) - \mathrm{SS}_{\mathrm{Error}}(A, B, AB)$$

To get Type III SS for main effects:

For SS_A (i) Fit model with B and interaction AB

(ii) Fit model with A, B, and interaction AB

(iii) Note difference between SS_{Error}.

Type III SS (cont)

- Type III SS are tedious to obtain in the unbalanced case; they are like full-reduced model differences in the error sum of squares, where the reduced model is the model without the main effect in question.
- The `anova()` or `aov()` functions in R give sequential SS, which are not meaningful in the unbalanced case; however, the Type III SS are equal to the sequential (i.e. Type I) SS when $n_{ij} = n \; \forall ij$.
- Under unbalancedness $\mathrm{SS}_{\mathrm{Trt}} \neq \mathrm{SS}_{\mathrm{A}}^{\mathrm{III}} + \mathrm{SS}_{\mathrm{B}}^{\mathrm{III}} + \mathrm{SS}_{\mathrm{AB}}$.

Type III SS for unbalanced designs

ANOVA table with Type III SS and MS for A and B:

Source	Df	SS	MS	F value
A	$a-1$	$\mathrm{SS}_{\mathrm{A}}^{\mathrm{III}}$	$\mathrm{MS}_{\mathrm{A}}^{\mathrm{III}}$	$F_{\mathrm{A}} = \mathrm{MS}_{\mathrm{A}}^{\mathrm{III}} / \mathrm{MS}_{\mathrm{Error}}$
B	$b-1$	$\mathrm{SS}_{\mathrm{B}}^{\mathrm{III}}$	$\mathrm{MS}_{\mathrm{B}}^{\mathrm{III}}$	$F_{\mathrm{B}} = \mathrm{MS}_{\mathrm{B}}^{\mathrm{III}} / \mathrm{MS}_{\mathrm{Error}}$
AB	$(a-1)(b-1)$	$\mathrm{SS}_{\mathrm{AB}}$	$\mathrm{MS}_{\mathrm{AB}}$	$F_{\mathrm{AB}} = \mathrm{MS}_{\mathrm{AB}} / \mathrm{MS}_{\mathrm{Error}}$
Error	$N-ab$	$\mathrm{SS}_{\mathrm{Error}}$	$\mathrm{MS}_{\mathrm{Error}}$	
Total	$N-1$	$\mathrm{SS}_{\mathrm{Tot}}$		

1. Reject H_0: $\bar{\mu}_{1.} = \cdots = \bar{\mu}_{a.}$ if $F_{\mathrm{A}} > F_{a-1,N-ab,\alpha}$.
2. Reject H_0: $\bar{\mu}_{.1} = \cdots = \bar{\mu}_{.b}$ if $F_{\mathrm{B}} > F_{b-1,N-ab,\alpha}$.
3. Reject H_0: $\mu_{ij} = \bar{\mu}_{i.} + \bar{\mu}_{.j} - \bar{\mu}_{..} \quad \forall ij$ if $F_{\mathrm{AB}} > F_{(a-1)(b-1),N-ab,\alpha}$.

Mice

Obtain the Type III SS on the mice data.

Fit model with both main effects:

```r
lm_full <- lm(Response ~ Diet + Gene + Diet:Gene,
              contrasts = list(Gene = contr.sum,Diet = contr.sum),
              data = adipoer) # must specify constraint type!
SSE_full <- sum(lm_full$residuals^2)
```

Drop main effect of Gene from the model:

```r
drop1(lm_full,~Gene,test="F",data = adipoer)
```

```
Single term deletions

Model:
Response ~ Diet + Gene + Diet:Gene
       Df Sum of Sq     RSS AIC F value Pr(>F)
<none>              4653342 632
Gene    1    105749 4759091 631    1.16   0.29
```

Drop main effect of Diet from the model:

```r
drop1(lm_full,~Diet,test="F",data = adipoer)
```

```
Single term deletions

Model:
Response ~ Diet + Gene + Diet:Gene
       Df Sum of Sq     RSS AIC F value  Pr(>F)
<none>              4653342 632
Diet    1   1253698 5907040 643    13.7 0.00052 ***
```

Mice

Obtain Type III SS with `Anova()` from the R package `car`.

Important to specify all the options exactly!

```r
library(car)
Anova(lm(Response ~ Gene + Diet + Gene:Diet,
         contrasts = list(Gene = contr.sum, Diet = contr.sum),# specify constraint
         data = adipoer),
       type = "III")
```

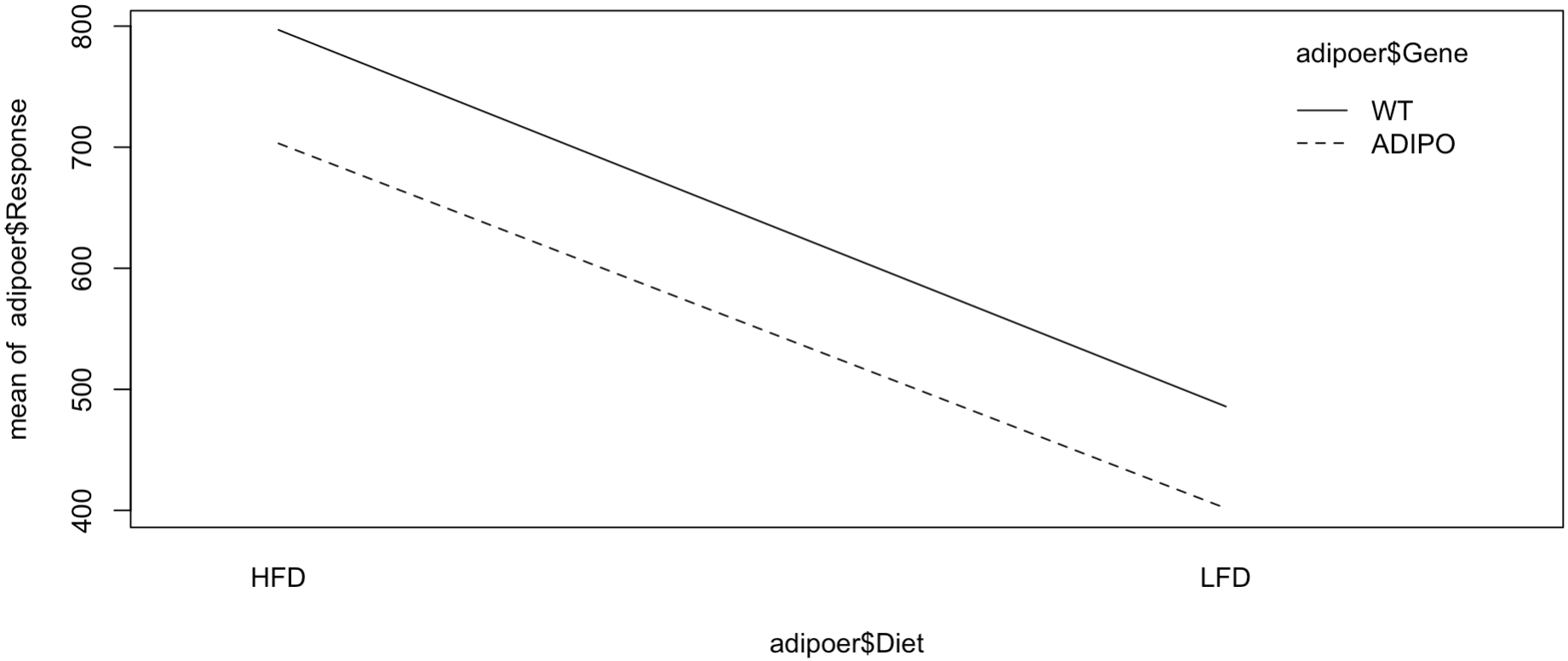
```
Anova Table (Type III tests)

Response: Response
             Sum Sq Df F value  Pr(>F)
(Intercept) 19050576  1  208.79 < 2e-16 ***
Gene          105749  1    1.16 0.28674
Diet         1253698  1   13.74 0.00052 ***
Gene:Diet        317  1    0.00 0.95325
Residuals    4653342 51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since interaction is not significant, we can focus on main effects

Mice

```
1 interaction.plot(adipoer$Diet, adipoer$Gene, adipoer$Response)
```



Estimates of cell and marginal means in unbalanced case

The estimators of the cell and marginal means are given by

- $\hat{\mu}_{ij} = \bar{Y}_{ij.}, i = 1, \ldots, a, j = 1, \ldots, b.$
- $\hat{\bar{\mu}}_{i.} = \frac{1}{b} \sum_{j=1}^{b} \hat{\mu}_{ij}, i = 1, \ldots, a.$
- $\hat{\bar{\mu}}_{.j} = \frac{1}{a} \sum_{i=1}^{a} \hat{\mu}_{ij}, j = 1, \ldots, b.$

We estimate $\hat{\bar{\mu}}_{i.}$ with $\bar{Y}_{i..}$ (and $\hat{\bar{\mu}}_{.j}$ with $\bar{Y}_{.j.}$) only when $n_{ij} = n \; \forall ij$.

Exercise: Write $\hat{\bar{\mu}}_{i.}$ in terms of Y_{ijk} and find $\mathrm{Var}\, \hat{\bar{\mu}}_{i.}$.

Some CI formulas (without familywise adjustment)

Under unbalancedness our CI formulas become more complicated.

Target	$(1-\alpha)100\%$ **confidence interval**
μ_{ij}	$\hat{\mu}_{ij} \pm t_{N-ab,\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n_{ij}}}$
$\mu_{ij} - \mu_{i'j'}$	$\hat{\mu}_{ij} - \hat{\mu}_{i'j'} \pm t_{N-ab,\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n_{ij}} + \frac{1}{n_{i'j'}}}$
$\bar{\mu}_{i.}$	$\hat{\bar{\mu}}_{i.} \pm t_{N-ab,\alpha/2}\hat{\sigma}\sqrt{\frac{1}{b^2}\sum_{j=1}^{b}\frac{1}{n_{ij}}}$
$\bar{\mu}_{.j}$	$\hat{\bar{\mu}}_{.j} \pm t_{N-ab,\alpha/2}\hat{\sigma}\sqrt{\frac{1}{a^2}\sum_{i=1}^{a}\frac{1}{n_{ij}}}$
$\bar{\mu}_{i.} - \bar{\mu}_{i'.}$	$\hat{\bar{\mu}}_{i.} - \hat{\bar{\mu}}_{i'.} \pm t_{N-ab,\alpha/2}\hat{\sigma}\sqrt{\frac{1}{b^2}\sum_{j=1}^{b}(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}})}$
$\bar{\mu}_{.j} - \bar{\mu}_{.j'}$	$\hat{\bar{\mu}}_{.j} - \hat{\bar{\mu}}_{.j'} \pm t_{N-ab,\alpha/2}\hat{\sigma}\sqrt{\frac{1}{a^2}\sum_{i=1}^{a}(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}})}$

In the above $\hat{\sigma} = \sqrt{\mathrm{MS}_{\mathrm{Error}}}$.

Mice

The interaction appears to be negligible, so we can focus on main effects.

Build 95% CIs for the differences in marginal means for

- Diet: $\hat{\bar{\mu}}_{1.} - \hat{\bar{\mu}}_{2.} \pm t_{55-4,0.05/2}\hat{\sigma}\sqrt{\frac{1}{2^2}\sum_{j=1}^{2}(\frac{1}{n_{1j}} + \frac{1}{n_{2j}})}$
- Gene: $\hat{\bar{\mu}}_{.1} - \hat{\bar{\mu}}_{.2} \pm t_{55-4,0.05/2}\hat{\sigma}\sqrt{\frac{1}{2^2}\sum_{i=1}^{2}(\frac{1}{n_{i1}} + \frac{1}{n_{i2}})},$

where $\hat{\sigma} = \sqrt{\mathrm{MS}_{\mathrm{Error}}}$.

Mice

```r
mu_hat <- grp_means$Response
nn <- grp_counts$Response
mu11_hat <- mu_hat[1] ; n11 <- nn[1] # HFD x ADIPO
mu21_hat <- mu_hat[2] ; n21 <- nn[2] # LFD x ADIPO
mu12_hat <- mu_hat[3] ; n12 <- nn[3] # HFD x WT
mu22_hat <- mu_hat[4] ; n22 <- nn[4] # LFD x WT

mu1.hat <- (mu11_hat + mu12_hat)/2 # HFD mean
mu2.hat <- (mu21_hat + mu22_hat)/2 # LFD mean
mu.1hat <- (mu11_hat + mu21_hat)/2 # ADIPO mean
mu.2hat <- (mu12_hat + mu22_hat)/2 # WT mean

N <- sum(nn)
a <- 2
b <- 2
MSE <- sum(lm_out$residuals^2)/(N - a*b)
alpha <- 0.05

me <- qt(1-alpha/2,N-a*b)*sqrt(MSE)*sqrt(1/2^2*(1/n11 + 1/n12 + 1/n11 + 1/n22))

ci_diet <- c(mu1.hat - mu2.hat - me, mu1.hat - mu2.hat + me)
ci_gene <- c(mu.1hat - mu.2hat - me, mu.1hat - mu.2hat + me)
```

We obtain the intervals:

- Diet: $[146.27, 466.24]$
- Gene: $[-248.93, 71.04]$

Tensile strength data (unbalanced) from Kuehl (2000)

Table 6.19 Tensile strength (psi) of asphaltic concrete specimens for two aggregate types with each of three kneading compaction methods

	Compaction Method			
	Kneading			
Aggregate Type	*Regular*	*Low*	*Very Low*	*Aggregate Means* ($\overline{y}_{i\cdot\cdot}$)
Basalt	106	93	56	
	108	101		
		98		
Means ($\overline{y}_{1j\cdot}$)	107.0	97.3	56	93.7
Silicious	107	63	40	
	110	60	41	
	116		44	
Means ($\overline{y}_{2j\cdot}$)	111.0	61.5	41.7	72.6
Compaction means ($\overline{y}_{\cdot j\cdot}$)	109.4	83.0	45.3	

```r
y <- c(106,108,93,101,98,56,107,110,116,63,60,40,41,44)
agg <- as.factor(c("b","b","b","b","b","b","s","s","s","s","s","s","s","s"))
comp <- as.factor(c("r","r","l","l","l","vl","r","r","r","l","l","vl","vl","vl"))
tensile <- data.frame(y = y, agg = agg, comp = comp)
```

Tensile strength (unbalanced)

Compute group means:

```
1 aggregate(y ~ agg + comp,data = tensile, mean)
```

```
  agg comp     y
1   b    l  97.3
2   s    l  61.5
3   b    r 107.0
4   s    r 111.0
5   b   vl  56.0
6   s   vl  41.7
```

Tensile strength (unbalanced)

Again sequential SS values depend on the order we put variables in:

```r
anova(lm(y ~ agg + comp + agg:comp, data = tensile))
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
agg        1   1518    1518   135.2 2.7e-06 ***
comp       2   8402    4201   374.1 1.3e-08 ***
agg:comp   2    953     477    42.5 5.5e-05 ***
Residuals  8     90      11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm(y ~ comp + agg + agg:comp, data = tensile))
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
comp       2   9159    4580   407.8 8.9e-09 ***
agg        1    761     761    67.7 3.6e-05 ***
comp:agg   2    953     477    42.5 5.5e-05 ***
Residuals  8     90      11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tensile strength (unbalanced)

Obtain Type III SS with `Anova()` from the R package `car`.

```r
library(car)
Anova(lm(y ~ agg + comp + agg:comp, data = tensile,
         contrasts = list(agg = contr.sum, comp = contr.sum)),
       type = "III")
```

```
Anova Table (Type III tests)

Response: y
            Sum Sq Df F value  Pr(>F)
(Intercept)  75050  1  6683.5 5.6e-13 ***
agg            710  1    63.3 4.6e-05 ***
comp          6806  2   303.1 2.9e-08 ***
agg:comp       953  2    42.5 5.5e-05 ***
Residuals       90  8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now we could make comparisons of means, noting the interaction.

References

Kuehl, R. O. 2000. *Design of Experiments: Statistical Principles of Research Design and Analysis*. Duxbury/Thomson Learning.