Module 4: Factorial Treatment Structure

Analyzing a Factorial: Model, ANOVA, and Decision Flow

Example 4.2: Bakery

Treatment Structure: 3 × 2 Full Factorial

- Shelf Height (Bottom, Middle, Top)
- Shelf Width (Regular, Wide)

Design Structure: CRD with r = 2 stores per treatment combination.

Response: Bread Sales

What changes when we add a factor?

- We add:
  - Another main effect
  - An interaction term
- We must decide what effects are present *before* interpreting the treatment means.

> The goal of the analysis is to decide which effects matter.

The Treatment Effects Model (Two-way Factorial)

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \text{ with } \epsilon_{ijk} \sim \text{ iid } N(0, \sigma^2)$$

$$\text{for } i = 1, 2, 3, \dots, a; \ j = 1, 2, \dots, b; \ k = 1, 2, \dots., r$$

ANOVA for Two-way Factorials

Just like one-way ANOVA: $SST = SSTrt + SSE$

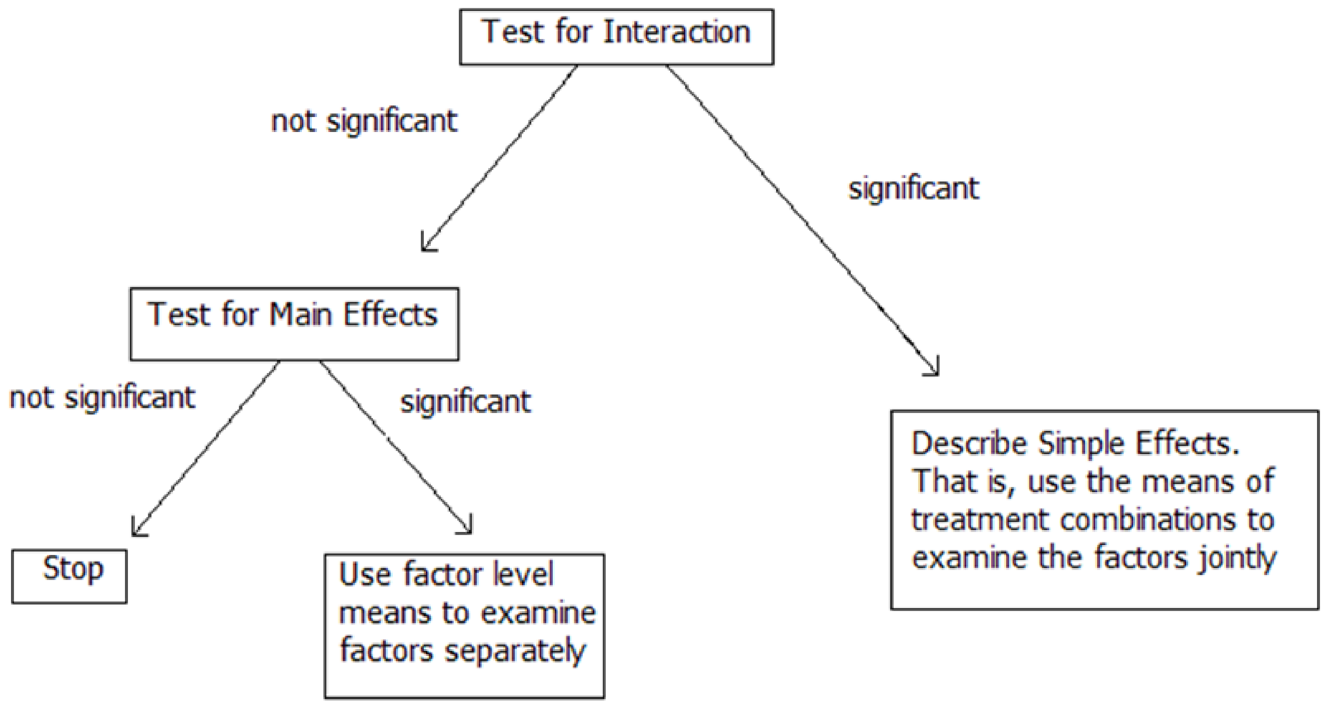
For a factorial: $SSTrt = SSA + SSB + SSAB$

- Main effect of factor A (Height): $SSA = rb\sum_i(\bar{y}_{i..} - \bar{y}_{..})^2$
- Main effect of factor B (Width): $SSB = ra\sum_j(\bar{y}_{.j.} - \bar{y}_{..})^2$
- AB Interaction Effect (Height x Width): $SSAB = SST - SSE - SSA - SSB$

Full ANOVA Table (Two-way)

Source (SV)	df	SS	MS	F	
A	$a - 1$	SSA	MSA = $\frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	Do avg sales differ across shelf heights?
B	$b - 1$	SSB	MSB = $\frac{SSB}{b-1}$	$\frac{MSB}{MSE}$	Do avg sales differ across shelf widths?
AB	$(a - 1)(b - 1)$	SSAB	MSAB = $\frac{SSAB}{(a-1)(b-1)}$	$\frac{MSAB}{MSE}$	Does the effect of width *depend* on height?
Error: e.u.(AB)	$(r - 1)ab$	SSE	MSE = $\frac{SSE}{(r-1)ab}$		
Total	$N - 1$	SST			

Decision Flowchart



Example 4.2: Skeleton ANOVA

Source of Variation	DF = 12 stores - 1 = 11 total df

R: Fitting the Model

Let's prep the data...

```
bakery_data <- read_csv("data/04_bakery_data.csv") %>%
  mutate(height = factor(height, levels = c("bottom", "middle", "top")),
         width = factor(width, levels = c("regular", "wide")))
head(bakery_data)
```

```
# A tibble: 6 × 4
  height width   sales placement
  <fct>  <fct>   <dbl> <chr>
1 bottom regular    47 bottomregular
2 bottom regular    43 bottomregular
3 bottom wide       46 bottomwide
4 bottom wide       40 bottomwide
5 middle regular    62 middleregular
6 middle regular    68 middleregular
```

```
levels(bakery_data$height)
```

```
[1] "bottom" "middle" "top"
```

```
levels(bakery_data$width)
```

```
[1] "regular" "wide"
```

R: Fitting the Model

```r
options(contrasts = c("contr.sum", "contr.poly"))
bakery_mod <- lm(sales ~ height + width + height:width, data = bakery_data)
```

```r
anova(bakery_mod)
```

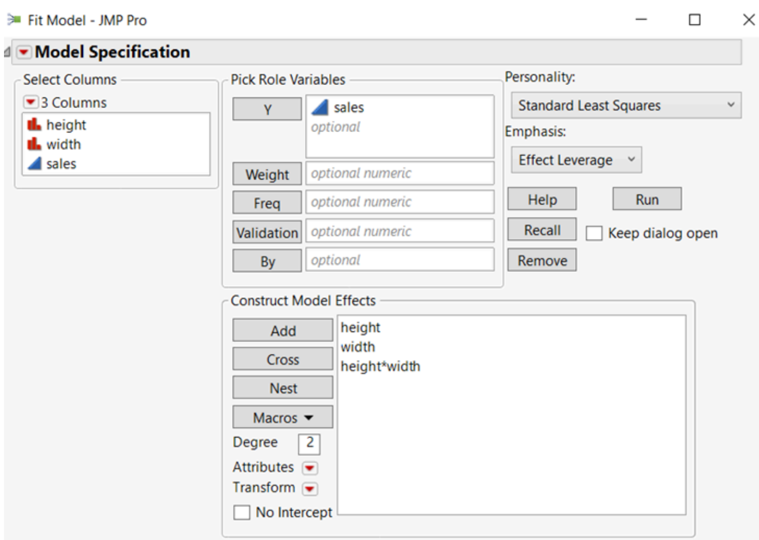
```
Analysis of Variance Table

Response: sales
             Df Sum Sq Mean Sq F value    Pr(>F)    
height        2   1544  772.00 74.7097 5.754e-05 ***
width         1     12   12.00  1.1613    0.3226    
height:width  2     24   12.00  1.1613    0.3747    
Residuals     6     62   10.33                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(bakery_mod)
```

```

Call:
lm(formula = sales ~ height + width + height:width, data = bakery_data)

Residuals:
   Min     1Q Median     3Q    Max 
    -3     -2      0      2      3 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)      51.000      0.928  54.959 2.44e-09 ***
height1          -7.000      1.312  -5.334  0.00177 ** 
height2          16.000      1.312  12.192 1.85e-05 ***
width1           -1.000      0.928  -1.078  0.32261    
height1:width1    2.000      1.312   1.524  0.17835    
height2:width1   -1.000      1.312  -0.762  0.47494    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 6 degrees of freedom
Multiple R-squared:  0.9622,	Adjusted R-squared:  0.9308 
F-statistic: 30.58 on 5 and 6 DF,  p-value: 0.0003384
```

JMP: Fitting the Model

- Analyze > Fit Model > Assign Y = Response + Highlight both treatment factors and click Macros > Full Factorial



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	1580.0000	316.000	30.5806
Error	6	62.0000	10.333	Prob > F
C. Total	11	1642.0000		0.0003*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
height	2	2	1544.0000	74.7097	<.0001*
width	1	1	12.0000	1.1613	0.3226
height*width	2	2	24.0000	1.1613	0.3747

JMP: Fitting the Model

Response > Expanded Estimates

Expanded Estimates

Nominal factors expanded to all levels

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	51	0.927961	54.96	<.0001*
height[bottom]	-7	1.312335	-5.33	0.0018*
height[middle]	16	1.312335	12.19	<.0001*
height[top]	-9	1.312335	-6.86	0.0005*
width[regular]	-1	0.927961	-1.08	0.3226
width[wide]	1	0.927961	1.08	0.3226
height[bottom]*width[regular]	2	1.312335	1.52	0.1783
height[bottom]*width[wide]	-2	1.312335	-1.52	0.1783
height[middle]*width[regular]	-1	1.312335	-0.76	0.4749
height[middle]*width[wide]	1	1.312335	0.76	0.4749
height[top]*width[regular]	-1	1.312335	-0.76	0.4749
height[top]*width[wide]	1	1.312335	0.76	0.4749

Recall, “sum to zero”

$\hat{\mu} =$

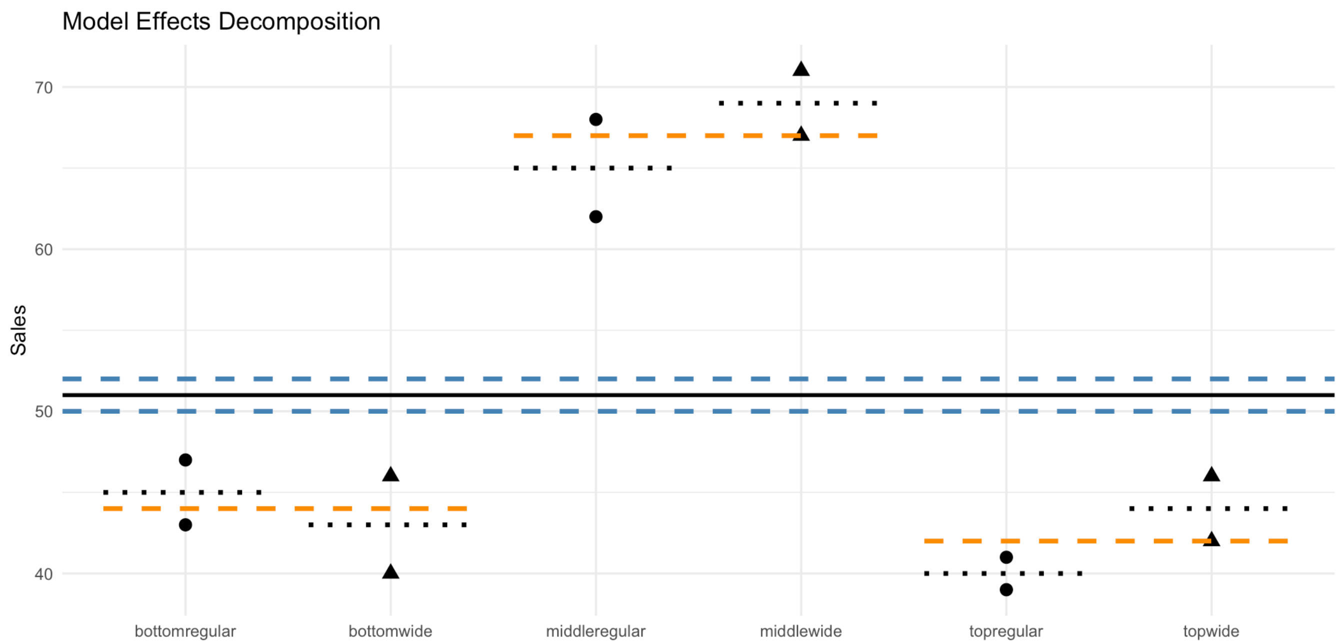
Height	
Bottom	$\hat{\alpha}_1 = 7$
Middle	$\hat{\alpha}_2 = 16$
Top	$\hat{\alpha}_3 =$

Width	
Regular	$\hat{\beta}_1 = -1$
Wide	$\hat{\beta}_2 =$

	Regular	Wide
Bottom	$\widehat{\alpha\beta}_{11} = 2$	$\widehat{\alpha\beta}_{12} =$
Middle	$\widehat{\alpha\beta}_{21} = -1$	$\widehat{\alpha\beta}_{22} =$
Top	$\widehat{\alpha\beta}_{31} =$	$\widehat{\alpha\beta}_{32} =$

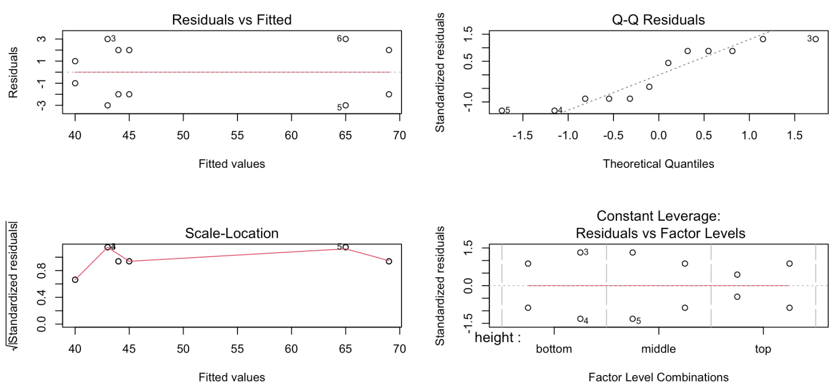
Decomposition of Model Effects

Recall our statistical effects model: $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$



Model Diagnostics

```
par(mfrow = c(2,2))
plot(bakery_mod)
```




```
par(mfrow = c(1,1))
```