Module 4: Factorial Treatment Structure

Example: No Interaction, interpret Main Effects

Example 4.2: Bakery

Treatment Structure: 3 x 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

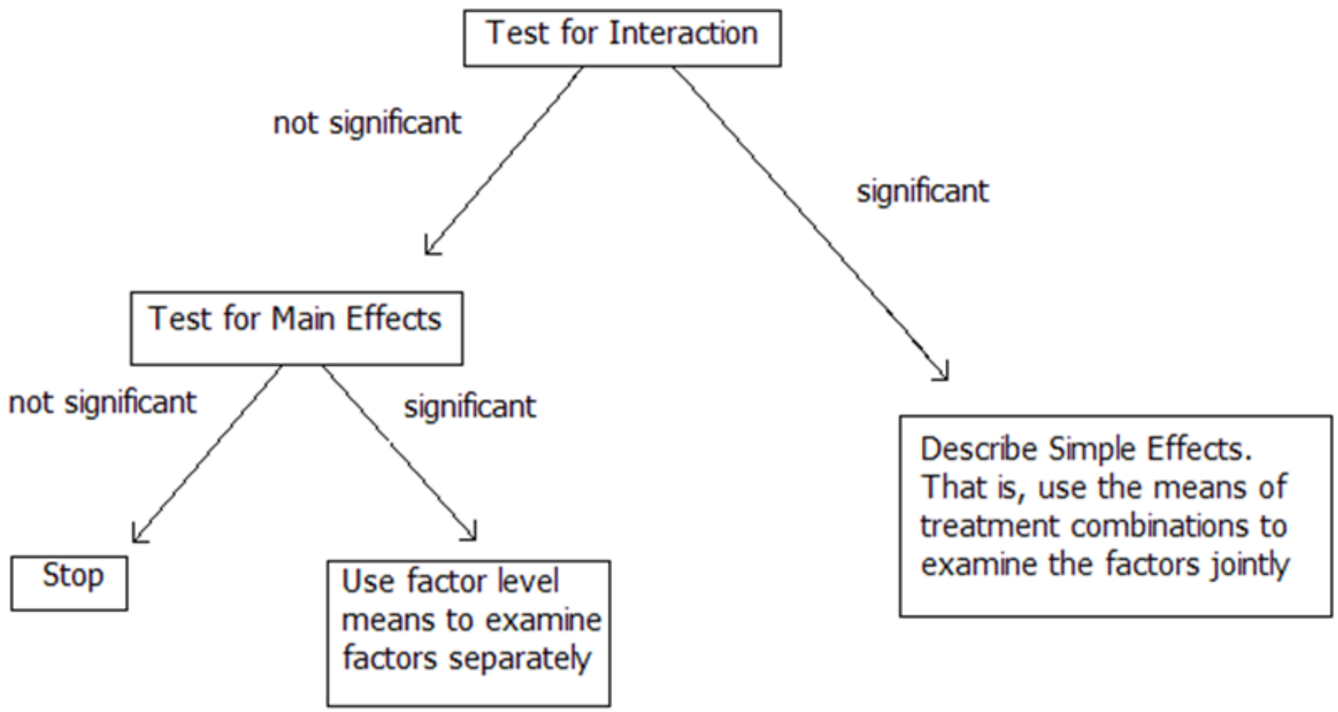
The Treatment Effects Model (Two-way Factorial)

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

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

- y_{ijk}: is the response (sales) from the k^{th} experimental unit (store) using the i^{th} level of Factor A (height) and j^{th} level of Factor B (width) combination
- μ: is the grand/overall mean sales
- α_i: is the effect of the i^{th} level of factor A (height)
- β_j: is the effect of the j^{th} level of factor B (width)
- $\alpha\beta_{ij}$: is the interaction effect between the i^{th} level of A (height) and j^{th} level of B (width)
- ϵ_{ijk}: the experimental error associated with the k^{th} experimental unit (store) using the i^{th} level of Factor A (height) and j^{th} level of Factor B (width) combination

Decision Flowchart



What are we testing?

Interaction

$$H_0: \text{ All } \alpha\beta_{ij} = 0 \text{ vs } H_A: \text{At least one } \alpha\beta_{ij} \neq 0$$

Main Effect of A

$$H_0: \text{ All } \alpha_i = 0 \text{ vs } H_A: \text{At least one } \alpha_i \neq 0$$

Main Effect of B

$$H_0: \text{ All } \beta_j = 0 \text{ vs } H_A: \text{At least one } \beta_j \neq 0$$

ANOVA

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

Analysis of Variance

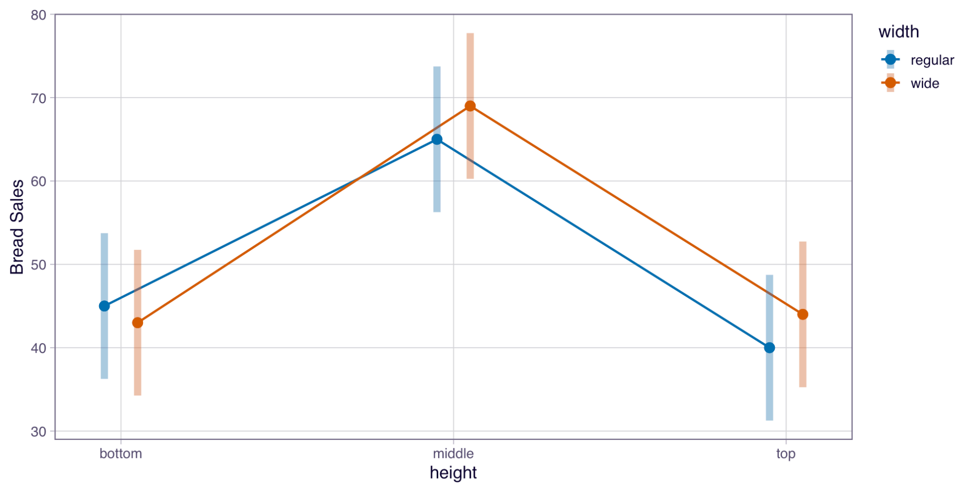
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	1580.0000	316.000	30.5806
Error	6	62.0000	10.333	**Prob > F**
C. Total	11	1642.0000		0.0003*

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

Do these results make sense?

Inspecting the interaction plot between height and width on bread sales, does there visually appear to be a significant interaction?



Where should we proceed?

1. Interaction: We do not have enough evidence of a significant interaction effect between shelf height and shelf width on bread sales ($F = 1.16$; $df = 2,6$; $p = 0.375$).
2. Width Main: We do not have enough evidence of a significant main effect of shelf width on bread sales ($F = 1.16$; $df = 1,6$; $p = 0.323$).
3. **Height Main**: We do have evidence of a significant main effect of height on bread sales ($F = 74.7$; $df = 2, 6$; $p < 0.0001$).

> What are the marginal mean sales for each height (sig main effect)? Which heights differ?

R: Marginal Height LSMEANS

```r
library(emmeans)
height_lsmeans <- emmeans(bakery_mod, specs = ~ height)
```

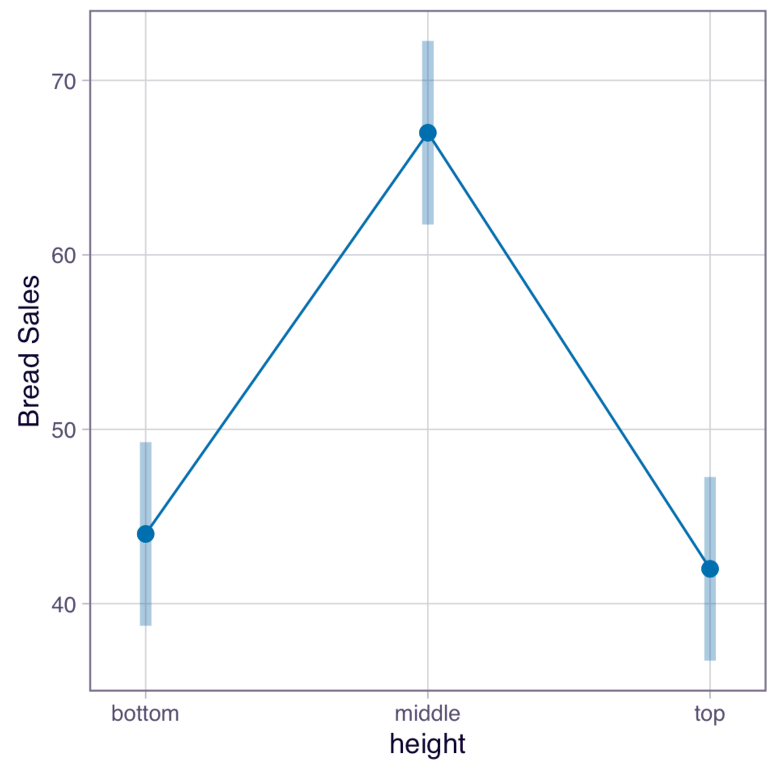
```
NOTE: Results may be misleading due to involvement in interactions
```

```r
height_lsmeans
```

```
 height emmean   SE df lower.CL upper.CL
 bottom     44 1.61  6     40.1     47.9
 middle     67 1.61  6     63.1     70.9
 top        42 1.61  6     38.1     45.9

Results are averaged over the levels of: width
Confidence level used: 0.95
```

```r
emmip(bakery_mod, ~ height, CIs = TRUE, adjust = "tukey") +
  labs(y = "Bread Sales",
       x = "height")
```



R: Height pairwise comparisons

```r
pairs(height_lsmeans, adjust = "tukey", infer = c(T,T))
```

```
 contrast        estimate   SE df lower.CL upper.CL t.ratio p.value
 bottom - middle      -23 2.27  6   -29.97   -16.03 -10.119  0.0001
 bottom - top           2 2.27  6    -4.97     8.97   0.880  0.6714
 middle - top          25 2.27  6    18.03    31.97  10.999 <0.0001

Results are averaged over the levels of: width 
Confidence level used: 0.95 
Conf-level adjustment: tukey method for comparing a family of 3 estimates 
P value adjustment: tukey method for comparing a family of 3 estimates 
```

```r
library(multcomp)
cld(height_lsmeans, decreasing = F, Letters = LETTERS, adjust = "tukey")
```

```
 height emmean   SE df lower.CL upper.CL .group
 top        42 1.61  6     36.7     47.3  A    
 bottom     44 1.61  6     38.7     49.3  A    
 middle     67 1.61  6     61.7     72.3   B   

Results are averaged over the levels of: width 
Confidence level used: 0.95 
Conf-level adjustment: sidak method for 3 estimates 
P value adjustment: tukey method for comparing a family of 3 estimates 
significance level used: alpha = 0.05 
NOTE: If two or more means share the same grouping symbol,
      then we cannot show them to be different.
      But we also did not show them to be the same. 
```

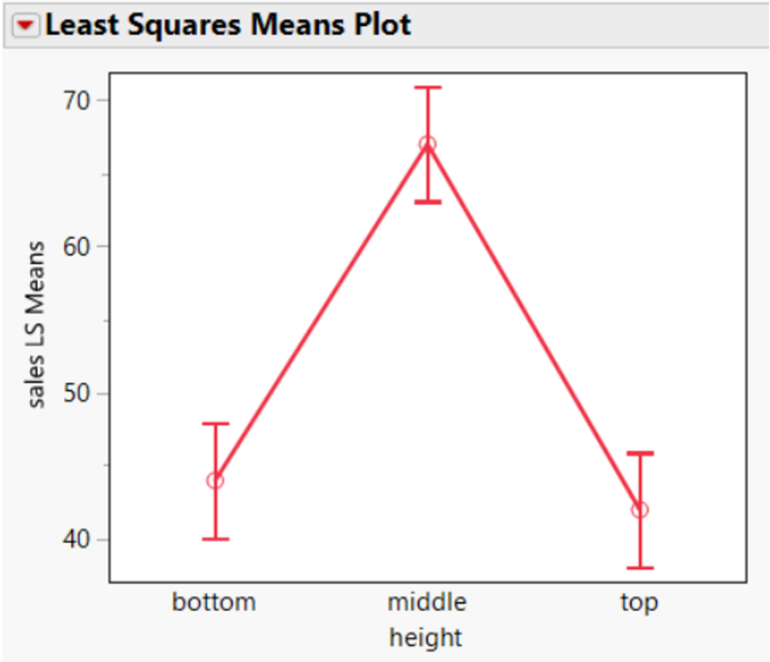
JMP: Main Effect of Height

Height > LSMeans Tukey HSD LSMeans Differences > Ordered Differences

Level			Least Sq Mean	Std Error
middle	A		67.000	1.6073
bottom		B	44.000	1.6073
top		B	42.000	1.6073

Levels not connected by same letter are significantly different.

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
middle	top	25.00000	2.273030	18.0260	31.97400	<.0001*
middle	bottom	23.00000	2.273030	16.0260	29.97400	0.0001*
bottom	top	2.00000	2.273030	-4.9740	8.97400	0.6714



What can we conclude?

- For bread placed on the middle shelf, the estimated mean bread sales is 67 (s.e. = 1.61).
- This is estimated to be 25 more sales than for bread placed on the top shelf ($t = 10.99$; $df = 6$; $p < 0.0001$) and 23 more sales than for bread placed on the bottom shelf ($t = 10.12$; $df = 6$; $p = 0.0001$).