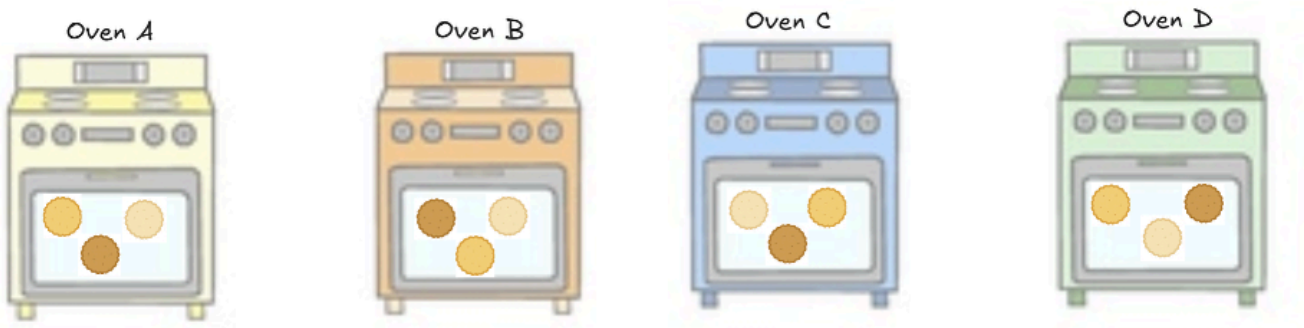
Module 5: Randomized Complete Block Designs (RCBD)

Analyzing a RCBD

Example 5.1: Cookie Recipes



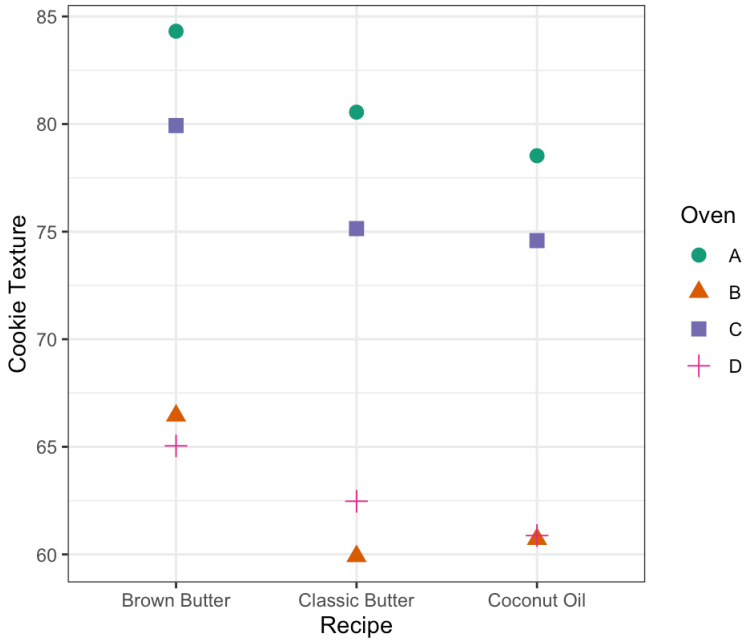
Treatment Structure: One-way with recipe (3-levels: classic butter, brown butter, coconut oil)

Design Structure: Recipe randomly assigned to tray (e.u.) in an RCBD with r = 4 ovens (blocking factor). Texture recorded for each tray (m.u.).

Cookie Recipes

```
1 cookie_data <- read_csv("data/cookie_rcbd_data.csv")
2 cookie_data
```

```
# A tibble: 12 × 3
   oven  recipe         texture
   <chr> <chr>            <dbl>
 1 A     Classic Butter    80.6
 2 A     Brown Butter      84.3
 3 A     Coconut Oil       78.5
 4 B     Classic Butter    59.9
 5 B     Brown Butter      66.4
 6 B     Coconut Oil       60.7
 7 C     Classic Butter    75.1
 8 C     Brown Butter      79.9
 9 C     Coconut Oil       74.6
10 D     Classic Butter    62.5
11 D     Brown Butter      65.0
12 D     Coconut Oil       60.9
```



What changes from a CRD?

If we analyze *ignoring* the effect of oven...

Incorrect Analysis: CRD

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \text{ where } \epsilon_{ij} \text{iid} \sim N(0, \sigma^2)$$

```r
oven_crd_mod <- lm(texture ~ recipe, data = cookie_data)
anova(oven_crd_mod)
```

```
Analysis of Variance Table

Response: texture
          Df Sum Sq Mean Sq F value Pr(>F)
recipe     2  63.82  31.910  0.3459 0.7166
Residuals  9 830.24  92.248
```

Statistical Effects Model

$$y_{ij} = \mu + \tau_i + \rho_j + \epsilon_{ij} \text{ where } \epsilon_{ij} \text{iid} \sim N(0, \sigma^2)$$

for $i = 1, 2, 3$ and $j = 1, 2, 3, 4$

where:

- y_{ij} is the observed texture of the cookie tray baked in the j^{th} oven with the i^{th} recipe.
- μ is the overall mean texture
- τ_i is the effect of the i^{th} recipe
- ϵ_{ij} is the experimental error associated with the cookie tray baked in the j^{th} oven with the i^{th} recipe.

What Blocking does to the variability

$$SST = SSTrt + SSBlk + SSE$$

- $SST = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$
- $SSTrt = r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$
- $SSBlk = t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$
- $SSE = t \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$

ANOVA Table

Source of Variation	DF	SS	MS	F
Block	r - 1	SSBlk	MSBlk	MSBlk/MSE
Treatment	t - 1	SSTrt	MSTrt	MSTrt/MSE
Block x Treatment ⟶ error	(r - 1)(t - 1)	SSE	MSE ⟶ σ^2	
Total (N = rt)	**N - 1**	SST		

> Blocking is a *design tool*, not usually a research question.

Skeleton ANOVA

Source of Variation	DF

R: RCBD Analysis

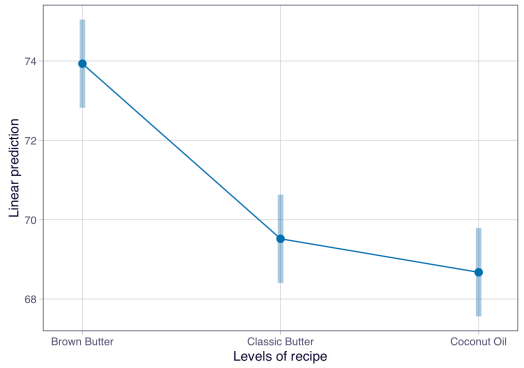
```r
oven_rcbd_mod <- lm(texture ~ recipe + oven, data = cookie_data)
anova(oven_rcbd_mod)
```

```
Analysis of Variance Table

Response: texture
          Df Sum Sq Mean Sq F value    Pr(>F)
recipe     2  63.82  31.910  38.607 0.0003749 ***
oven       3 825.28 275.092 332.829 4.651e-07 ***
Residuals  6   4.96   0.827
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

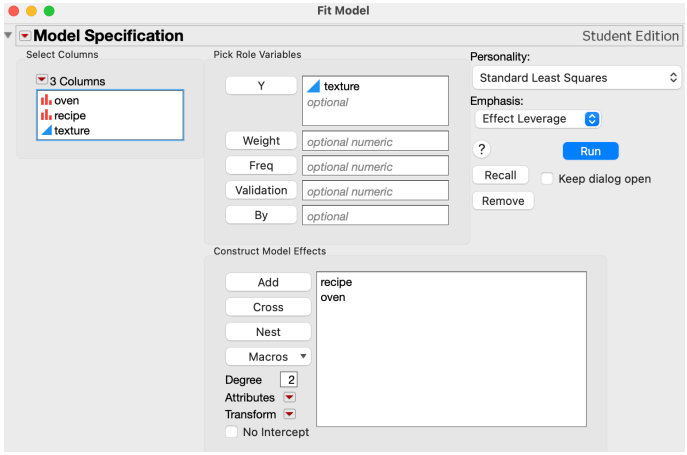
```r
emmeans(oven_rcbd_mod, ~ recipe) |>
  cld(Letters = LETTERS, decreasing = T, adjust = "tukey")
```

```
 recipe         emmean    SE df lower.CL upper.CL .group
 Brown Butter     73.9 0.455  6     72.4     75.4  A
 Classic Butter   69.5 0.455  6     68.0     71.0   B
 Coconut Oil      68.7 0.455  6     67.2     70.2   B

Results are averaged over the levels of: oven
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.
```

```r
emmip(oven_rcbd_mod, ~ recipe, CI = T)
```



JMP: RCBD Analysis



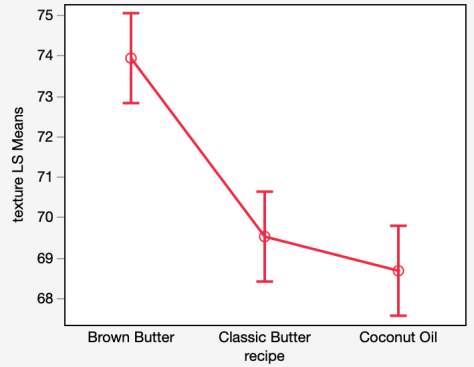
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	889.09649	177.819	215.1405
Error	6	4.95916	0.827	Prob > F
C. Total	11	894.05565		<.0001*

Parameter Estimates

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
recipe	2	2	63.81924	38.6069	0.0004*
oven	3	3	825.27725	332.8295	<.0001*



LSMeans Differences Tukey HSD

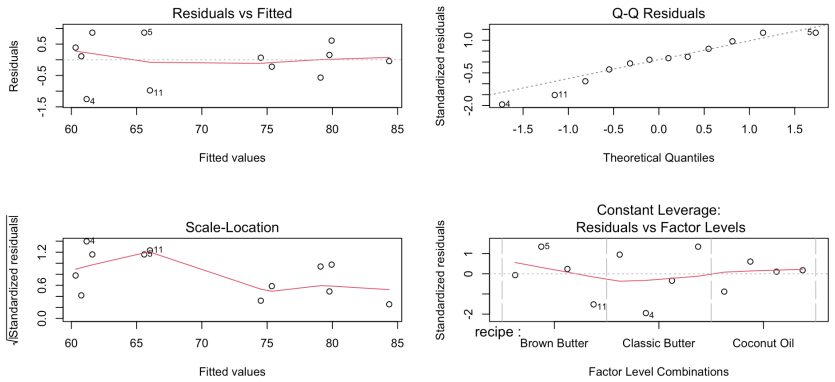
α= 0.050 Q= 3.06815

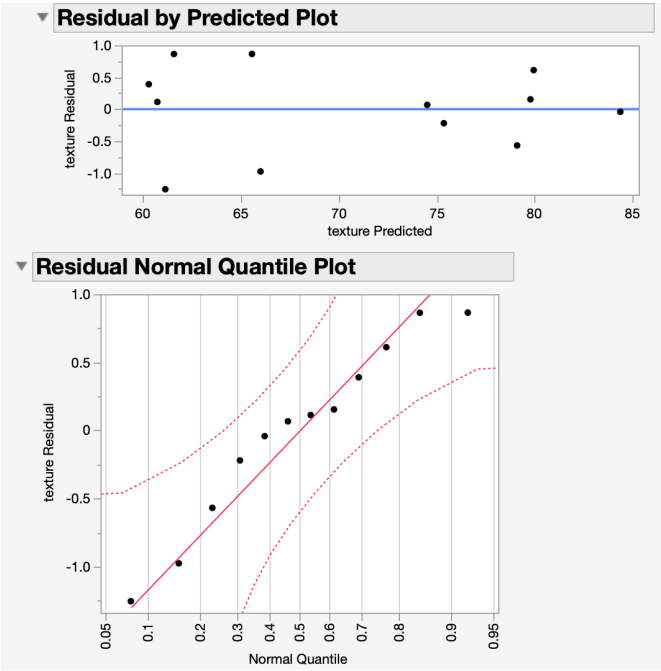
Level		Least Sq Mean	Std Error
Brown Butter	A	73.935	0.45457
Classic Butter	B	69.520	0.45457
Coconut Oil	B	68.675	0.45457

Levels not connected by same letter are significantly different.

Check Model Assumptions – $\epsilon_{ij} \text{iid} \sim N(0, \sigma^2)$

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





RCBD vs CRD

If we ignore ovens (CRD):

- Oven variability inflates MSE
- Harder to detect recipe differences

```
Analysis of Variance Table

Response: texture
          Df Sum Sq Mean Sq F value Pr(>F)
recipe     2  63.82  31.910  0.3459 0.7166
Residuals  9 830.24  92.248
```

With RCBD:

- Oven variability is removed from error
- Smaller MSE

  More power *(if blocks are different)*

```
Analysis of Variance Table

Response: texture
          Df Sum Sq Mean Sq F value    Pr(>F)
recipe     2  63.82  31.910  38.607 0.0003749 ***
oven       3 825.28 275.092 332.829 4.651e-07 ***
Residuals  6   4.96   0.827
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Why is this model additive? (no interaction)

- Each recipe appears *once* per oven
- No replication within recipe-oven combinations
- Interaction is not estimable

> Recall: The experimental error comes from the block x treatment term.