Module 2: Completely Randomized Designs

Statistical Model for a CRD

From ANOVA to a Statistical Model

ANOVA answers whether treatments differ.

A statistical model explains:

- Where variability comes from
- How treatment effects are represented
- What parameters/coefficients we estimate

Modeling the Response

For a CRD, each observation can be written as:

$$y_{ij} = \text{systematic part} + \text{random error}$$

The model separates signal from noise.

Example 2.1: Running Shoes

Response: Lap time (seconds)

Treatment structure:

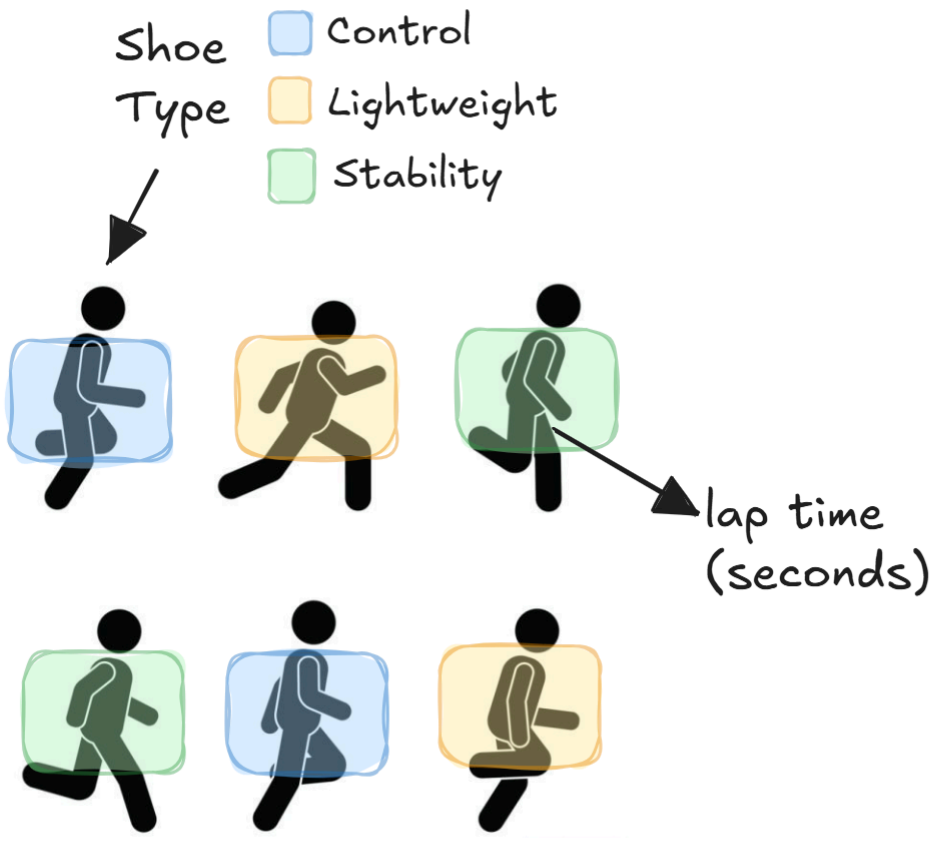
- One-way
- Factor: Shoe type
- 3 Levels: control, lightweight, and stability
- $t = 3$

Experimental structure:

- CRD
- Experimental Unit: Individual ($r = 2$)
- Measurement Unit: Individual ($N = 6$)

Goal: Determine whether shoe type affects mean lap time.

Example 2.1: Running Shoes (Blueprint)



Cell Means Model

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ with } \varepsilon_{ij} \text{ iid } \sim N(0, \sigma^2)$$

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

Where:

- y_{ij} - the *observed* lap time for the j^{th} runner wearing the i^{th} shoe type.
- μ_i - the mean lap time for runners wearing the i^{th} shoe type.
- ϵ_{ij} - the experimental error associated with the j^{th} runner wearing the i^{th} shoe type.

Effects Model

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

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

Where:

- y_{ij} - the *observed* lap time for the j^{th} runner wearing the i^{th} shoe type.
- μ - the overall mean lap time.
- τ_i - the effect of the i^{th} shoe type.
- ϵ_{ij} - the experimental error associated with the j^{th} runner wearing the i^{th} shoe type.

Assumptions: $\varepsilon_{ij}\ \text{iid}\ \sim N(0, \sigma^2)$

We assume:

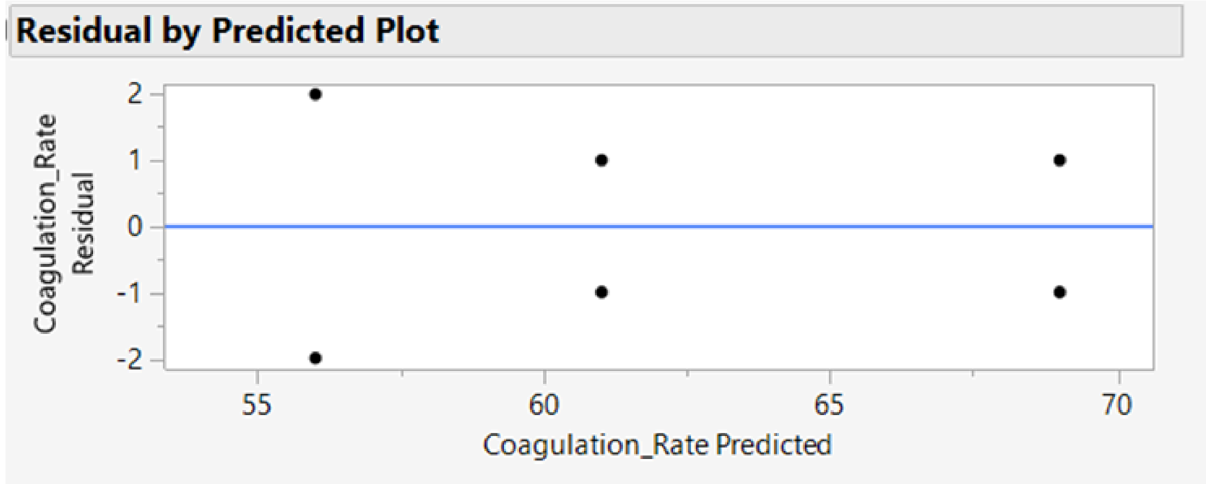
- Errors are independent and normally distributed *(iid ~ = identically and independently distributed)*
- Mean zero
- Constant variance σ^2 (e.g., pooled t-test)

All ANOVA inference depends on these assumptions.

Checking Model Assumptions: Constant Variance

Type of Plot: Look at the residuals vs fitted/predicted values

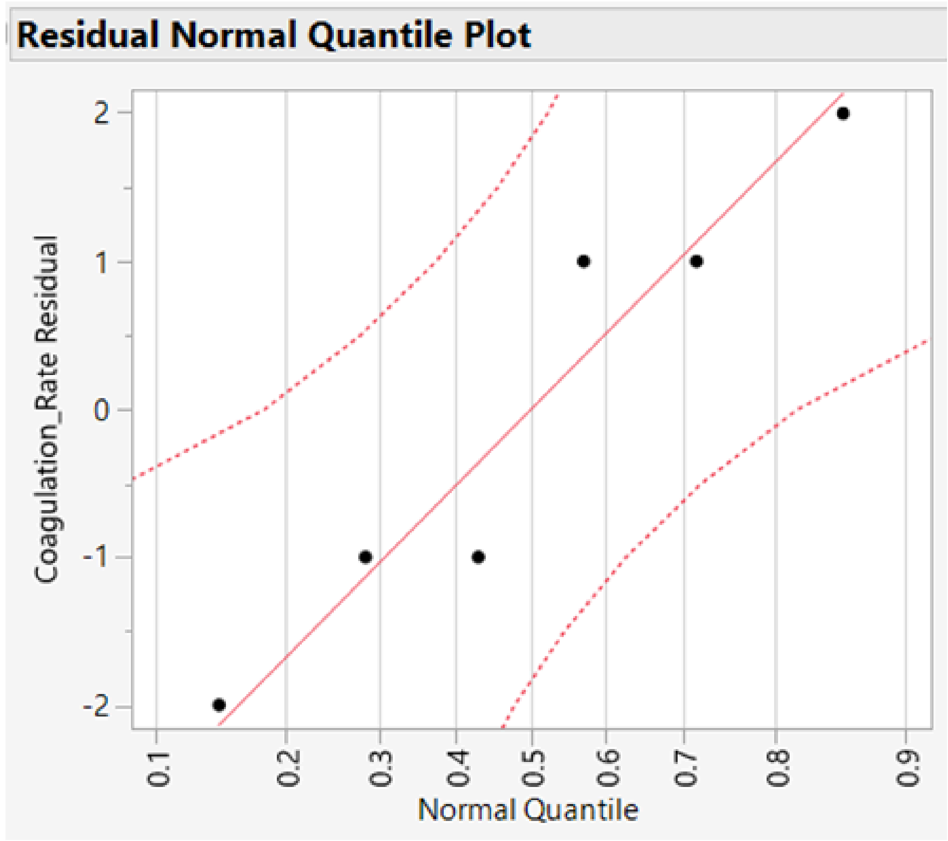
Ideal Plot: Equally spread in residuals across all treatments (e.g., horizontal band)



Checking Model Assumptions: Normality

Type of Plot: Normal probability plot of the residuals (QQ-plot)

Ideal Plot: Points fall close to the reference line



Estimating Model Parameters/Coefficients

Recall: $y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ with } \varepsilon_{ij} \text{ iid } \sim N(0, \sigma^2)$

Goal: Use data to *estimate*:

- μ (denoted $\hat{\mu}$),
- τ (denoted $\hat{\tau}$), and
- σ^2 (denoted $\hat{\sigma}^2$)

Then we can make predictions/estimates with $\hat{y}_i = \hat{\mu} + \hat{\tau}_i$.

Identifiability Problem

The model parameters are not unique unless we add a constraint.

Solutions: Impose a constraint on τ_i

- Sum to zero (JMP default) *what we will use in this class*
- Set to zero (R default) *need to change R settings when analyzing data*

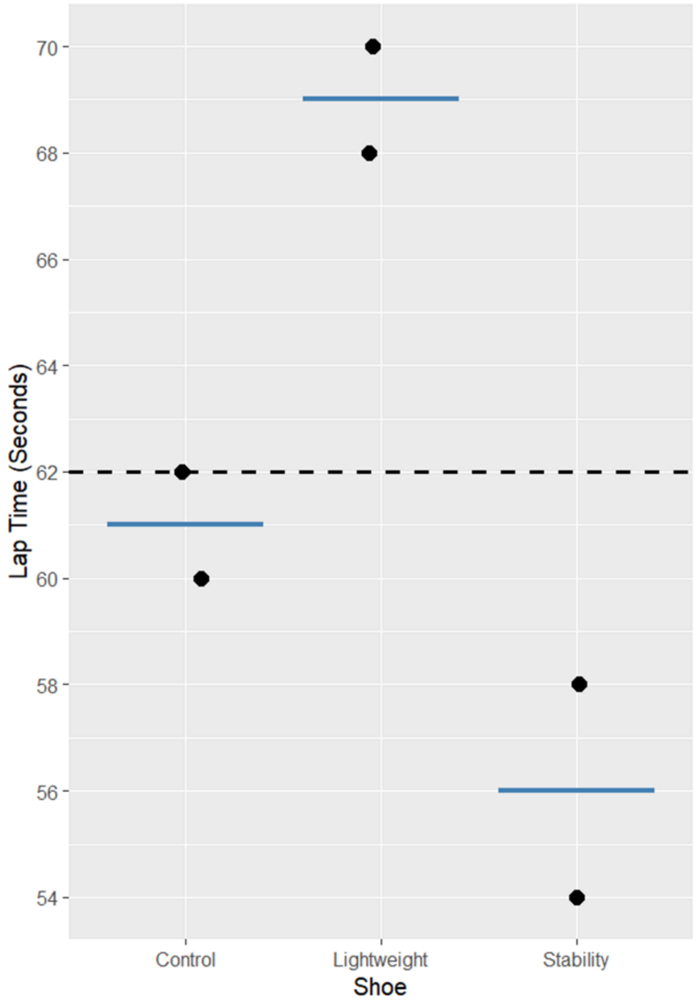
Sum to Zero

Constraint: $\sum_i^t \tau_i = 0$

- _______ (overall mean)
- _______ (deviation of trt mean from overall mean)
- _______
- _______

Then:

- _______
- _______
- _______



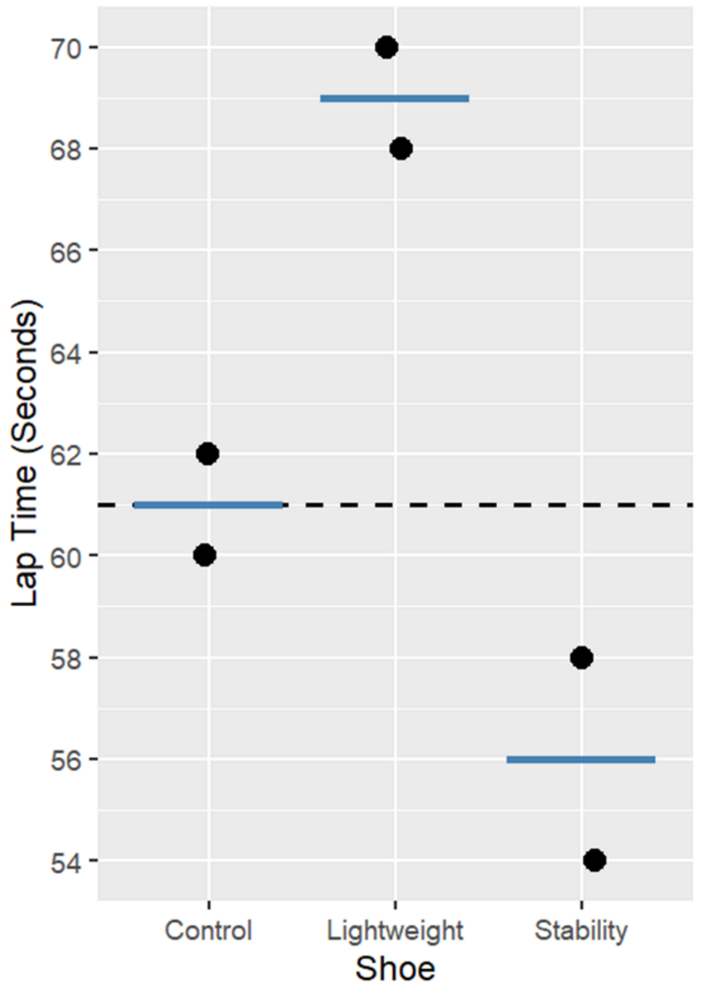
Set to Zero

Constraint: Set (or another reference level, e.g.)

- ________ (mean of reference level, e.g.)
- ________ (deviation of trt mean from mean of reference level)
- ________
- ________

Then:

- ________
- ________
- ________



R (set to zero): Estimating Parameters

```r
shoe_mod <- lm(`Lap Time (seconds)` ~ Shoe, data = shoe_data)
summary(shoe_mod)
```

```
Call:
lm(formula = `Lap Time (seconds)` ~ Shoe, data = shoe_data)

Residuals:
 1  2  3  4  5  6 
-1  1 -1  1  2 -2 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       61.000      1.414   43.13 2.74e-05 ***
ShoeLightweight    8.000      2.000    4.00   0.0280 *  
ShoeStability     -5.000      2.000   -2.50   0.0877 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2 on 3 degrees of freedom
Multiple R-squared:  0.9348,	Adjusted R-squared:  0.8913 
F-statistic:  21.5 on 2 and 3 DF,  p-value: 0.01666
```

R (sum to zero): Estimating Parameters

Change default settings to "sum to zero"

```r
options(contrasts = c("contr.sum", "contr.poly"))
shoe_mod <- lm(`Lap Time (seconds)` ~ Shoe, data = shoe_data)
summary(shoe_mod)
```

```

Call:
lm(formula = `Lap Time (seconds)` ~ Shoe, data = shoe_data)

Residuals:
 1  2  3  4  5  6 
-1  1 -1  1  2 -2 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  62.0000     0.8165  75.934 5.03e-06 ***
Shoe1        -1.0000     1.1547  -0.866  0.45018    
Shoe2         7.0000     1.1547   6.062  0.00901 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2 on 3 degrees of freedom
Multiple R-squared:  0.9348,	Adjusted R-squared:  0.8913 
F-statistic:  21.5 on 2 and 3 DF,  p-value: 0.01666
```

```r
options(contrasts = c("contr.treatment", "contr.poly"))
```

R: Estimating

```r
options(contrasts = c("contr.sum", "contr.poly"))
shoe_mod <- lm(`Lap Time (seconds)` ~ Shoe, data = shoe_data)
anova(shoe_mod)
```

```
Analysis of Variance Table

Response: Lap Time (seconds)
          Df Sum Sq Mean Sq F value  Pr(>F)
Shoe       2    172      86    21.5 0.01666 *
Residuals  3     12       4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

JMP (sum to zero): Estimating Parameters

`Analyze` > `Fit Model` > `Fill in Y (response) and Add (factor) + Effect Leverage Emphasis` > `Run`

`Expanded Estimates`

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	62	0.816497	75.93	<.0001*
Shoe[Control]	-1	1.154701	-0.87	0.4502
Shoe[Lightweight]	7	1.154701	6.06	0.0090*

Expanded Estimates

Nominal factors expanded to all levels

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	62	0.816497	75.93	<.0001*
Shoe[Control]	-1	1.154701	-0.87	0.4502
Shoe[Lightweight]	7	1.154701	6.06	0.0090*
Shoe[Stability]	-6	1.154701	-5.20	0.0138*

JMP: Estimating

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	172.00000	86.0000	21.5000
Error	3	12.00000	4.0000	Prob > F
C. Total	5	184.00000		0.0167*