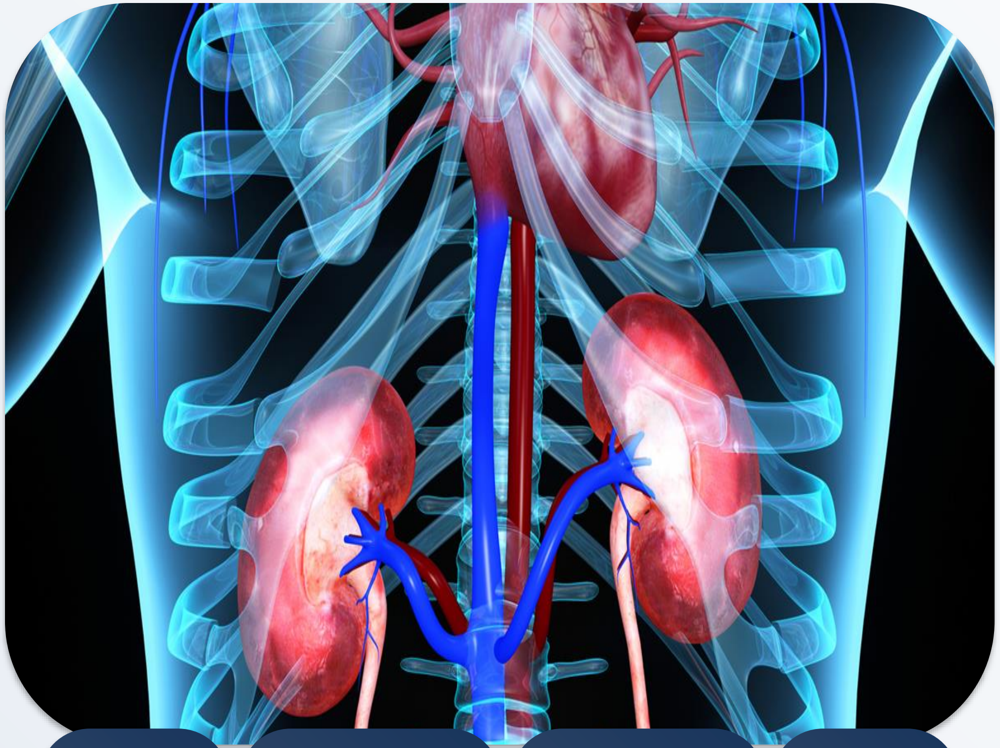
# One-Way ANOVA

## A Statistical Method for Comparing Multiple Groups

Prepared by: Dalal Qallajo

# Why do we need a one-way ANOVA?

- Which **drug type** is best to start your day?
- For **two groups**: **t-test**
- Problem: **more than two types** of coffee?
- Multiple pairwise tests = **time consuming**
- More importantly: **error accumulation**



Tacrolimus | Cyclosporine | Mycophenolate | Sirolimus

# How does error accumulate?

Four types of drug:

Results in **6 pairwise comparisons**

- Tacrolimus vs Cyclosporine
- Tacrolimus vs Mycophenolate
- Tacrolimus vs Sirolimus
- Cyclosporine vs Sirolimus
- Cyclosporine vs Mycophenolate
- Sirolimus vs Mycophenolate

**27%**

## Error Accumulation

1. Each comparison: **5%** error rate
2. No error in one test: **95%**
3. For 6 comparisons: $0.95^6 = 0.73$
4. At least one error: $1 - 0.73 = 0.27$

# What is the signal and what is the noise?

## Signal

Variance **between** groups — differences between group means

Represents **useful information** we're looking for

VS

## Noise

Variance **within** each group — differences among individuals in the **same** group

Reflects **random variation** not the main focus

ANOVA compares **signal-to-noise ratio** to determine if group differences are significant

# 1 Sum of Squares

### Sum of Squares Between

Measures how much the four coffee types differ from the **overall mean**

$$SS_{\text{between}} = n\sum(\bar{X}_k - \bar{X}_G)^2$$

### Sum of Squares Within

Measures the **noise** inside each group

$$SS_{\text{within}} = \sum(X_{ik} - \bar{X}_k)^2$$

### Total sum of squares

Compares every individual score to the overall mean, giving the **total variance**

$$SS_{\text{total}} = \sum(X_{ik} - \bar{X}_G)^2$$

**The nice property of ANOVA is:** $SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$

## 2 Degrees of Freedom

**Between-groups df:** Number of groups minus one

$$k - 1 = 4 - 1 = 3$$

**Within-groups df:** Each group has 10 people → (10 - 1 = 9)
Across 4 groups:

$$9 \times 4 = 36$$

**Total df:** All participants minus one

$$40 - 1 = 39$$

**As expected:** $df_{\text{between}} + df_{\text{within}} = df_{\text{total}}$ $3 + 36 = 39$

# Mean Squares (MS)

### The Problem

We **cannot** compare these numbers directly:
• SS_between uses **3 df**
• SS_within uses **36 df**

Comparing them directly would be **unfair**.

$$SS_{\text{between}} = 27.875$$

$$SS_{\text{within}} = 101.50$$

### Mean Squares (MS)

$$MS = \frac{SS}{df}$$

We divide each Sum of Squares by its degrees of freedom to get a fair comparison between variance components

**Mean Square Between**

Between-group variance

$$\frac{27.875}{3} = 9.292$$

**Mean Square Within**

Within-group variance

$$\frac{101.50}{36} = 2.819$$

# F F-value

**The beauty of ANOVA lies here:**

$$F = \frac{\text{signal}}{\text{noise}} = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

Signal Noise

$$F = \frac{9.292}{2.819}$$

$$F = 3.296$$

# ANOVA Table

| Source | Sum of Squares | df | Mean Square | F |
|---|---|---|---|---|
| Between | 27.875 | 3 | 9.292 | **3.296** |
| Within | 101.500 | 36 | 2.819 | |
| Total | **129.375** | **39** | | |

# Is the F-value significant?

### Check the F-table at:

- df_between = 3
- df_within = 36

### Critical value

≈ 2.86

### F-value Comparison

| F_critical | F_calculated |
|---|---|
| 2.86 | 3.296 |

$$F_{\text{calculated}} > F_{\text{critical}}$$

There is a significant difference between the groups

# Expected Mean Squares (Discussion)

### You may wonder:

If there is no real difference between the groups…
why doesn't the signal become zero? And therefore F ≈ 0?

### Then, Can F be less than 1?

**In general**

F is **close to 1** when there is no true effect.

# A nice note

$$F = t^2$$

When comparing only two groups using ANOVA, the F-value equals the square of the t-value from a t-test

# What ANOVA tells us

When we run ANOVA and find a significant difference between groups, it only tells us:

**“Not all means are equal”**

**But the important question is:**

**<u>Which groups differ from each other?  And  where is the real difference?</u>**

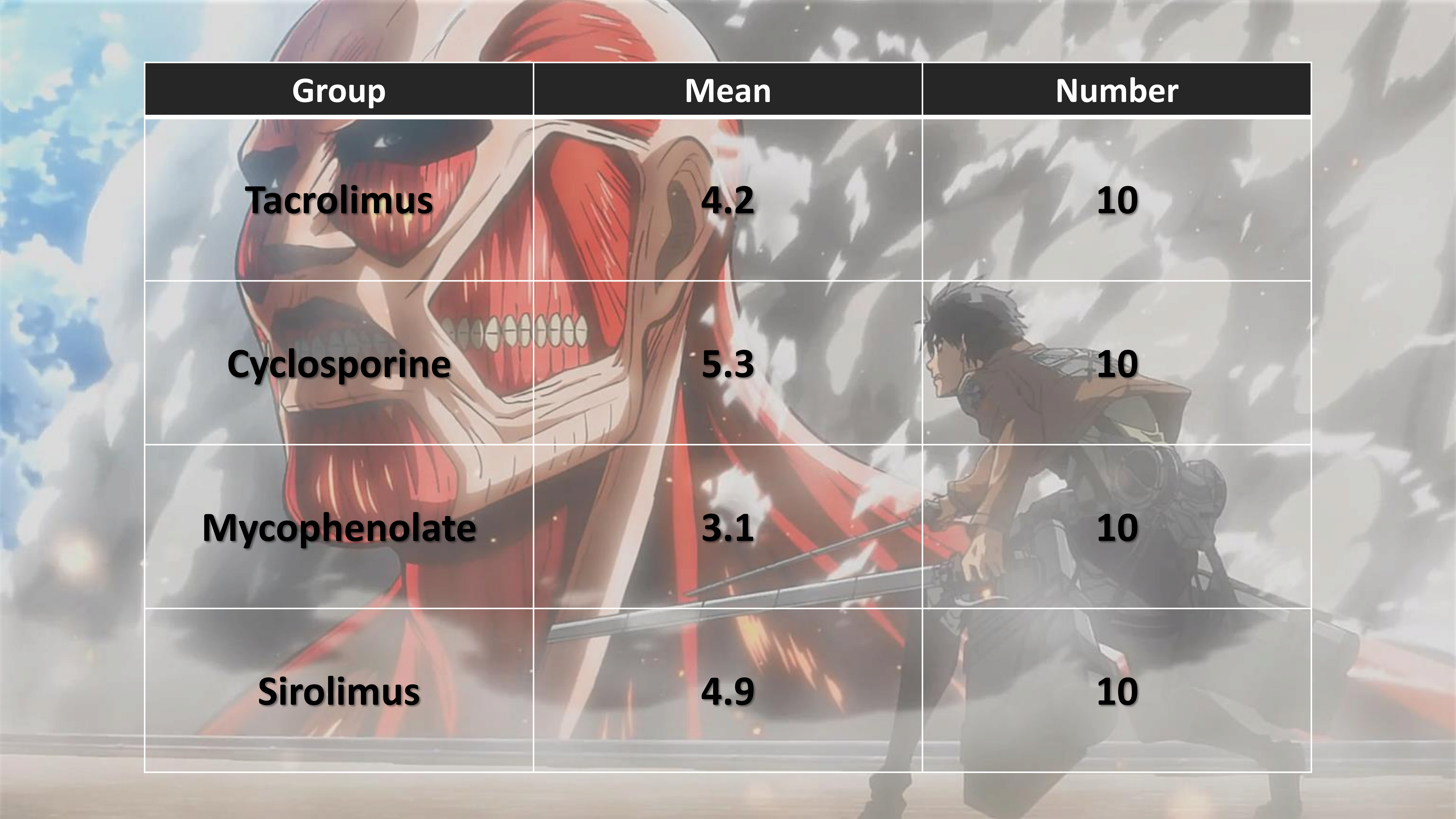
# After the ANOVA Test

### Planned Comparisons

- Determined **before** the study begins
- Researcher enters with **specific hypotheses** about certain groups only

### Post-Hoc Comparisons

- Occur **after** ANOVA tells us that a difference exists
- Used to **explore all possible** group differences



| Group | Mean | Number |
|---|---|---|
| Tacrolimus | 4.2 | 10 |
| Cyclosporine | 5.3 | 10 |
| Mycophenolate | 3.1 | 10 |
| Sirolimus | 4.9 | 10 |

# Orthogonal Planned Contrasts

We choose **3 orthogonal contrasts** (because df_between = 3):

**C1: (T+ C) vs (M + S)**

+1/2 +1/2 -1/2 -1/2 = 0

**C2: T vs C**

+1 -1 0 0 = 0

**C3: M vs S**

0 0 +1 -1 = 0

(0.5×1) + (0.5×-1) + (-0.5×0) + (-0.5×0) = 0

# Computing Each Contrast Value

W = $\Sigma w_i^2 / n$

**C1: (T + C) vs (M + S)**

C1 = 1/2(4.2) + 1/2(5.3) - 1/2(3.1) - 1/2(4.9) **= 0.75**

**C1: (T + C) vs (M + S)**

W1 = $((0.5)^2 * 4)/10$ = 0.10

**C2: T vs C**

C2 = 1(4.2) - 1(5.3) **= -1.10**

**C2: T vs C**

W2 = $((1)^2 + (-1)^2)/10$ = 0.20

**C3: M vs S**

C3 = 1(3.1) - 1(4.9) **= -1.80**

**C3: M vs S**

W3 = $((1)^2 + (-1)^2)/10$ = 0.20

# Computing Sum of Squares for Each Contrast

$$SS(C) = C^2/W$$

For $C_1$: $SS(C_1) = 0.75^2/0.10 = 5.625$

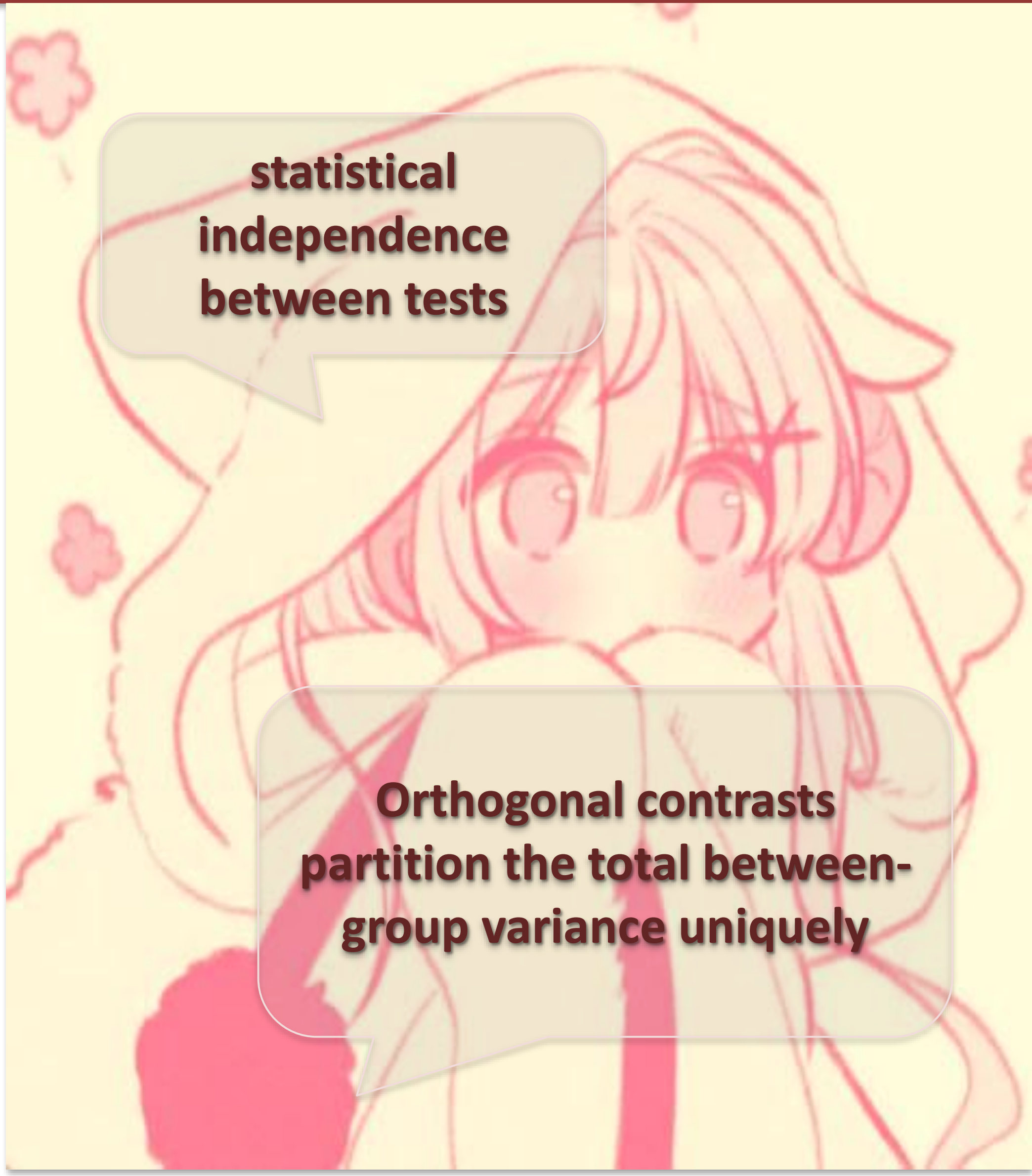
For $C_2$: $SS(C_2) = (-1.10)^2/0.20 = 6.05$

For $C_3$: $SS(C_3) = (-1.80)^2/0.20 = 16.20$

**Important Note:**

**$5.625 + 6.05 + 16.20 = 27.875 = SS\_Between$**

**→ This confirms that the contrasts are orthogonal**



# Computing Sum of Squares for Each Contrast

**(using MS_within = 2.819), F = SS(C)/ MS_within**

$F_1 = 5.625/2.819 = 1.995$

$F_2 = 6.05/2.819 = 2.146$

$F_3 = 16.20/2.819 = 5.747$

**The critical value:**

**$F_{1,36,0.05} \approx 4.11$**

Only the third contrast $C_3$ (M vs S) is statistically significant because

**5.747 > 4.11**

Contrasts 1 and 2 are not significant at $\alpha = .05$.

# Post-hoc comparisons — Pairwise tests between means

**First, we compute the absolute differences between the pairs:**

| Diff = $\lvert\Delta\rvert$ | Pair |
|---|---|
| 5.3 − 4.2 = 1.10 | **Tacrolimus vs Cyclosporine** |
| 4.2 − 3.1 = 1.10 | **Tacrolimus vs Mycophenolate** |
| 4.9 − 4.2 = 0.70 | **Tacrolimus vs Sirolimus** |
| 5.3 − 3.1 = 2.20 | **Cyclosporine vs Mycophenolate** |
| 5.3 − 4.9 = 0.40 | **Cyclosporine vs Sirolimus** |
| 4.9 − 3.1 = 1.80 | **Sirolimus vs Mycophenolate** |

# Tukey LSD-Least Significant Difference

$t$_crit ≈ 2.03 (df =36, α=.05, two-tailed) $SE = \sqrt{(MS_{within} \times (1/n + 1/n))} = \sqrt{(2.819 \times 2/10)} \approx 0.7509$

LSD = $t$_crit × SE ≈ 2.03 × 0.7509 ≈ 1.524

**Result (LSD):**

**Any difference greater than 1.524 is significant**

| S − M | C − M | T − C |
|---|---|---|
| 1.80 | 2.20 | 1.10 |
| **M − T** | **S − T** | **C − S** |
| 1.10 | 0.70 | 0.40 |

# Tukey HSD-Honestly Significant Difference

**We use the q statistic for k=4 groups and $df_{within}$=36, Approximate value: q≈3.79 (from the Q table)**

$$HSD = q \times \sqrt{(MS_{within}/n)} = 3.79 \times \sqrt{(2.819/10)} \approx 2.01$$

## Result (HSD):

**Any difference greater than 2.01 is significant**

| | | |
|---|---|---|
| S − M<br>**1.80** | C − M<br>**2.20** | S − R<br>**1.46** |
| R − T<br>**1.46** | S − U<br>**0.53** | U − R<br>**0.93** |

# t-statistic Calculations

| Pair | t-statistic = Diff / SE |
|---|---|
| T – C | 1.10/0.7509≈1.465 |
| M – T | 1.10/0.7509≈1.465 |
| T – S | 0.70/0.7509≈0.932 |
| C – M | 2.20/0.7509≈2.929 |
| S – C | 0.40/0.7509≈0.533 |
| S – M | 1.80/0.7509≈2.397 |

# Scheffé

$F_{\text{Scheffé critical}} = (k-1) \times F_{\alpha,k-1,N-k}$
$F_{\text{Scheffé critical}} = 3 \times 2.87 = 8.61$

**Decision Rule**
$F_{\text{comparison}} > F_{\text{Scheffé critical}}$

## Result (Scheffé):

Any Value greater than 8.61 is significant

| T − M | T − C | T − S |
|---|---|---|
| $t^2 = 2.13$ | $t^2 = 2.13$ | $t^2 = 0.86$ |
| **S − C** | **S − M** | **C − M** |
| $t^2 = 0.28$ | $t^2 = 5.76$ | $t^2 = 8.85$ |

# Discussion

| Test | Pair |
|---|---|
| Orthogonal Planned Contrasts | M-S |
| Tukey HSD | M-C |
| Tukey LSD | M-C & M-S |
| Scheffé | Nan |

# Thank you for your attention

Dalal Qallajo