Module 2: Completely Randomized Designs

Pairwise Comparisons

What Are Pairwise Comparisons?

> Which treatment means differ from each other?

Pairwise comparisons test:

- μ_1 vs μ_2
- μ_1 vs μ_3
- μ_2 vs μ_3

For t treatments, there are: $\binom{t}{2} = \frac{t(t-1)}{2}$ possible pairwise comparisons.

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)

Example 2.1: Shoes

With $t = 3$ shoe types:

- Control vs Lightweight $(H_0 : \mu_1 - \mu_2 = 0)$
- Control vs Stability $(H_0 : \mu_1 - \mu_3 = 0)$
- Lightweight vs Stability $(H_0 : \mu_2 - \mu_3 = 0)$

That’s ______ comparisons.

Pairwise Comparisons as Contrasts

Each pairwise comparison is a contrast:

- Control vs Lightweight: $C = (1, -1, 0)$
- Control vs Stability: $C = (1, 0, -1)$
- Lightweight vs Stability: $C = (0, 1, -1)$

The Multiple Testing Problem

Each comparison is a valid t-test.

But performing *many tests* increases the chance of a:

- Type I error (false positive)

This is called the *family-wise error rate*.

Why Error Rates Inflate

Suppose each test uses $\alpha = 0.05$. Then:

- Probability of *no* false positives (i.e., correctly fail to reject the null): $1 - \alpha = 1 - 0.05 = 0.95$
- Assuming independence, the probability of all tests leading to *no* false positives: $(1-\alpha)^{\#\text{ tests}} = (0.95)^3 = 0.857$
- Probability of *at least one* false positive (i.e., a Type I error): $1 - \left[(1-\alpha)^{\#\text{ tests}}\right] = 1 - 0.857 = 0.143$.

Note this increases with the number of tests.

What We Need

When doing many comparisons, we want to:

- Control the overall Type I error rate
- While still detecting real differences

Post Hoc Tests: adjustments to the p-values (making them larger) and/or confidence intervals (making them wider).

Fisher’s Protected LSD

1. First perform the overall ANOVA F-test
2. Only if F is significant, perform all pairwise t-tests

Note: From the ANOVA on Example 2.1 Shoe Types: ($F = 21.5$; df = 2,3; $p = 0.017$)

```
library(emmeans)
shoe_lsmeans <- emmeans(shoe_mod, specs = ~ Shoe)
pairs(shoe_lsmeans, infer = c(TRUE, TRUE), adjust = "none")
```

```
 contrast                 estimate SE df lower.CL upper.CL t.ratio p.value
 Control - Lightweight          -8  2  3   -14.36    -1.64  -4.000  0.0280
 Control - Stability             5  2  3    -1.36    11.36   2.500  0.0877
 Lightweight - Stability        13  2  3     6.64    19.36   6.500  0.0074

Confidence level used: 0.95
```

```
library(multcomp)
cld(shoe_lsmeans, adjust = "none", Letters = LETTERS, decreasing = T)
```

```
 Shoe        emmean   SE df lower.CL upper.CL .group
 Lightweight     69 1.41  3     64.5     73.5  A
 Control         61 1.41  3     56.5     65.5   B
 Stability       56 1.41  3     51.5     60.5   B

Confidence level used: 0.95
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.
```

Shoe > LSMeans Student’s t

LSMeans Differences Student’s t > Ordered Differences Report

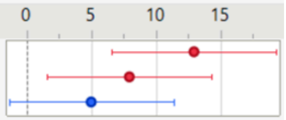
Shoe

LSMeans Differences Student's t

Level		Least Sq Mean	Std Error
Lightweight	A	69.000	1.4142
Control	B	61.000	1.4142
Stability	B	56.000	1.4142

Levels not connected by same letter are significantly different.

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Lightweight	Stability	13.00000	2.000000	6.63511	19.36489	0.0074*
Lightweight	Control	8.00000	2.000000	1.63511	14.36489	0.0280*
Control	Stability	5.00000	2.000000	-1.36489	11.36489	0.0877



Tukey's Honestly Significant Difference (HSD)

- Designed specifically for *all* pairwise comparisons
- Controls family-wise error rate

```r
library(emmeans)
shoe_lsmeans <- emmeans(shoe_mod, specs = ~ Shoe)
pairs(shoe_lsmeans, infer = c(TRUE, TRUE), adjust = "tukey")
```

```
 contrast                 estimate SE df lower.CL upper.CL t.ratio p.value
 Control - Lightweight          -8  2  3   -16.36    0.358  -4.000  0.0560
 Control - Stability             5  2  3    -3.36   13.358   2.500  0.1681
 Lightweight - Stability        13  2  3     4.64   21.358   6.500  0.0151

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
cld(shoe_lsmeans, adjust = "tukey", Letters = LETTERS, decreasing = T)
```

```
 Shoe        emmean   SE df lower.CL upper.CL .group
 Lightweight     69 1.41  3     62.2     75.8  A
 Control         61 1.41  3     54.2     67.8  AB
 Stability       56 1.41  3     49.2     62.8   B

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.
```

Shoe > LSMeans Tukey HSD

LSMeans Differences Tukey HSD > Ordered Differences Report

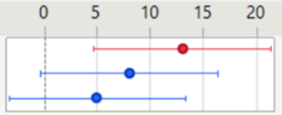
Shoe

LSMeans Differences Tukey HSD

Level			Least Sq Mean	Std Error
Lightweight	A		69.000	1.4142
Control	A	B	61.000	1.4142
Stability		B	56.000	1.4142

Levels not connected by same letter are significantly different.

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Lightweight	Stability	13.00000	2.000000	4.64258	21.35742	0.0151*
Lightweight	Control	8.00000	2.000000	-0.35742	16.35742	0.0560
Control	Stability	5.00000	2.000000	-3.35742	13.35742	0.1681



Bonferroni Adjustment

Bonferroni controls error by:

- Using a smaller significance level

$$\alpha^* = \frac{\alpha}{\text{number of tests}}$$

Simple, but often conservative.

Dunnet's Test

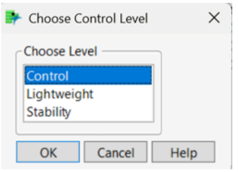
Useful when interest lies in comparing all treatments to a control.

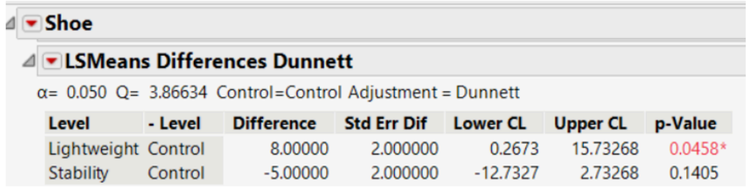
```r
library(emmeans)
shoe_lsmeans <- emmeans(shoe_mod, specs = ~ Shoe)
contrast(shoe_lsmeans,
         method = "trt.vs.ctrl", # Treatment vs. control contrast
         adjust = "dunnet", # Dunnett adjustment
         infer = c(TRUE, TRUE),
         ref = "Control"
)
```

```
 contrast             estimate SE df lower.CL upper.CL t.ratio p.value
 Lightweight - Control       8  2  3   0.0241    15.98   4.000  0.0496
 Stability - Control        -5  2  3 -12.9759     2.98  -2.500  0.1511

Confidence level used: 0.95
Conf-level adjustment: dunnettx method for 2 estimates
P value adjustment: dunnettx method for 2 tests
```

Shoe > LSMeans Dunnet > Choose Level (Control)





Shoe

LSMeans Differences Dunnett

α= 0.050 Q= 3.86634 Control=Control Adjustment = Dunnett

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Lightweight	Control	8.00000	2.000000	0.2673	15.73268	0.0458*
Stability	Control	-5.00000	2.000000	-12.7327	2.73268	0.1405