

### Comparison Procedures

**LSD** Requires knowing the global F test is significant before any testing should be done. More powerful than Bonferroni, but can lead to inflated Type I error rate. Not recommended

**TUKEY** Controls experimentwise Type I error rate exactly if sample sizes between groups are equal More powerful than Bonferroni Use when only interested in all (or most) pairwise comparisons of means(-unplanned comparisons)

**BONFERRONI** Very conservative Use when only interested in a small number of planned comparisons Use when making comparisons other than between pairs.

### Individual Vs Family CI

\*\*Individual confidence level (comparisonwise,  $\alpha_{cw}$ ): Success rate of a procedure for constructing a single confidence interval (or conducting a single hypothesis test)

**Familywise confidence level (experimentwise,  $\alpha_{ew}$ ):** Success rate of a procedure for constructing a family of confidence intervals (or a family of hypothesis tests) A "successful" usage is one in which ALL intervals in the family capture their parameters

### Pairwise Example

Nematodes in tomato plants How do nematodes affect plant growth? 12 identical pots, Different# of nematodes per pot; 0, 1000, 5000, Measure increase in tomato seedling height (cm) 16 days after planting

Hypothesis Test:  $H_0: \mu_1 = \mu_2 = \mu_3$  vs  $H_a$ : Not all of the  $\mu$  are equal

Overall F test significant (p-value = 0.0027) --> not all of the group means are equal

Hypo for pairwise comparisons  $H_0: \mu_i - \mu_j = 0$   
 $H_a: \mu_i - \mu_j \neq 0 \rightarrow i \neq j$

How many pairwise comparisons are there for this example?

How many pairwise comparisons are there for this example?  $K = K - 1 = 3 - 1 = 2 \rightarrow 3(3 - 1) = 3$

### Pairwise Example (cont)

Levels not connected by the same letter are significantly different

Ordered Differences Report for lower CL and upper CI = difference of growth

Which nematode groups have significantly different ang growth? 0-5000 (p-value = 0.0045) and 5000-1000 (0.006) are significantly different Difference in growth between 0 and 5000 will be between 1.83 cm and 8.27 cm more than 5000

### Factor Effects Model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

Response = overall avg growth + main effects + interaction effect

Assumptions:  $\epsilon_{ijk} \sim N(0, \sigma^2)$  are independent Errors are normal, independent, with constant variance

Constraints:  $\sigma \alpha_i = \sigma \beta_j = \sigma (\alpha\beta)_{ij} = 0$  Required to keep model from being over-parameterized

# of test = factoralevels X factorblevels

### Assumptions Errors are

Normal Normal Probability Plot

Constant Predicted Residual vs Fitted Variance and Residual vs Factor Levels

Independence Usually Assumed and typically assumed

### Main Effects and Interaction Plot

**Main Effects Plot** Separate plot for each factor Plot of mean response for each level of the factor Gives an indication of whether a factor is important Horizontal line indicates means are the same for both levels of the factor and thus factor is not important

**Interaction Plot** Plot means for each treatment combination against levels of one factor, with different lines for the other factor Parallel Lines - No Interaction Crossing Lines - Interaction

### Simple Liner Regression Vocal

Least Squares Regression standard approach for estimating the line The line is chosen, so that the sum of squared vertical distances between points and the line is minimized.

correlation coefficient r A measure of association for two quantitative variables is the

predicted or fitted values. The points that lie on the regression line vertically above (or below) an observed value Notation  $\hat{y}_i$

residuals. (vertical) distances between observed and predicted values notatio:  $e_i = y_i - \hat{y}_i$

### Model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \text{ where } \begin{cases} i = 1, 2, \dots, n \\ \epsilon_i \sim N(0, \sigma^2) \end{cases}$$

*Handwritten annotations:  $y_i$  is response,  $\beta_0$  is intercept,  $\beta_1$  is slope,  $x_i$  is explanatory var,  $\epsilon_i$  is error.*

### F-Tests

First: Test for Interaction •  $H_0: (\alpha\beta)_{ij} = 0$  for all i, j vs.  $H_a$ : Not all  $(\alpha\beta)_{ij}$  equal zero

p-value  $< \alpha$  = Reject  $H_0$  or p-value  $> \alpha$  = fail to reject  $H_0$  This means that both factors are important (even if main effects not significant) • We may need to interpret in terms of the interaction since we may not be able to separate out the main effects for individual interpretation

Then: Test Main Effects Factor A  $H_0: \alpha_i = 0$  for all i vs.  $H_a$ : Not all  $\alpha_i$  equal zero • Main Effect of Factor B  $H_0: \beta_j = 0$  for all j vs.  $H_a$ : Not all  $\beta_j$  equal zero

$H_0$ : factor  $\alpha_i = 0$  for all i (no effect) reject  $H_0$  if p-value  $< \alpha$



## Model parameters(unknown)

$\beta_0$  = intercept  
 $\beta_1$  = slope  
 $\sigma^2$  = error variance

## Assumprions

Linear relationship between X and Y

Model assumes that the error terms are Independent, Normal, Have constant variance.

Residuals may be used to explore the legitimacy of these assumptions.



By **Hsuzc**  
[cheatography.com/hsuzc/](https://cheatography.com/hsuzc/)

Not published yet.  
Last updated 9th December, 2022.  
Page 2 of 2.

Sponsored by **ApolloPad.com**  
Everyone has a novel in them. Finish  
Yours!  
<https://apollopad.com>