BIOS 662 Fall 2018

Analysis of Variance, Part II

David Couper, Ph.D.

david_couper@unc.edu

or

couper@bios.unc.edu

https://sakai.unc.edu/portal

Outline

- Multiple Comparisons
  - Scheffé
  - Tukey
  - Bonferroni

- Chapter 12 of the text

Multiple Comparisons

- Suppose we do n independent tests, each with probability α of making a type I error
- Suppose all n null hypotheses are true
- What is the probability of making at least one type I error?

$$1 - (1 - \alpha)^n$$

Multiple Comparisons

- Table 12.1: Probability of rejecting at least one null hypothesis when n independent tests are carried out at the α level and each null hypothesis is true

	α		
n	0.01	0.05	0.10
1	0.01	0.05	0.10
2	0.02	0.10	0.19
3	0.03	0.14	0.27
4	0.04	0.19	0.34
5	0.05	0.23	0.41
10	0.10	0.40	0.65
20	0.18	0.64	0.88
100	0.63	0.99	1.00

Multiple Comparisons

- Definition 12.2: The probability of incorrectly rejecting at least one of the true null hypotheses in an experiment involving one or more tests or comparisons is called the *per experiment error rate (PEER)*
- PEER is also known as the *family-wise error rate (FWE)*

ANOVA and Multiple Comparisons

- Rejection of $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$ does not indicate where the inequalities are

- For example,

$$H_A : \mu_1 = \mu_2 = \cdots = \mu_{K-1} \neq \mu_K$$

or

$$H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_{K-1} \neq \mu_K$$

- Usually we want to identify the inequalities

ANOVA

- Need a multiple comparisons method to test the $\binom{K}{2}$ null hypotheses

$$H_0 : \mu_i = \mu_j \quad (i \neq j)$$

- Popular methods:
  - Scheffé
  - Tukey
  - Bonferroni (Sidak, Holm, Hochberg)

ANOVA: Scheffé

- For each pair of means, compute

$$t_{ij} = \frac{\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}}{\sqrt{\mathrm{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

- Rejection region

$$C_\alpha = \left\{ t_{ij} : |t_{ij}| > \sqrt{(K-1)F_{K-1,N-K,1-\alpha}} \right\}$$

- Passive smoking example

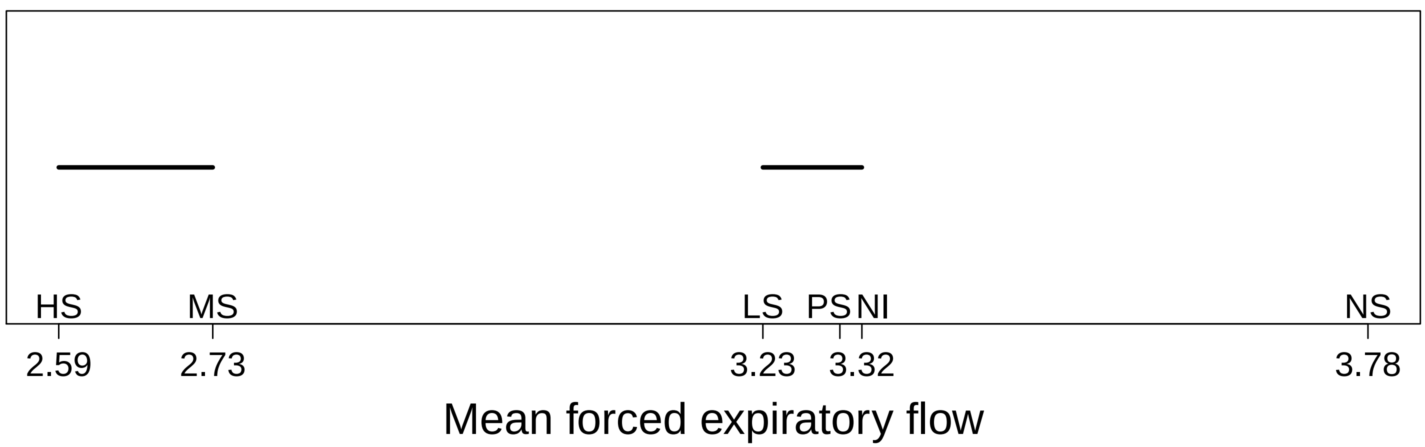
$$C_{0.05} = \left\{ t_{ij} : |t_{ij}| > \sqrt{5F_{5,1044,0.95}} = \sqrt{5(2.22)} = 3.33 \right\}$$

Scheffé: Passive Smoking Example

Comparison	t_{ij}	Significant
NS-PS	6.02	yes
NS-NI	3.65	yes
NS-LS	6.90	yes
NS-MS	13.17	yes
NS-HS	14.92	yes
PS-NI	-0.16	no
PS-LS	0.88	no
PS-MS	7.15	yes
PS-HS	8.90	yes
NI-LS	0.71	no
NI-MS	4.68	yes
NI-HS	5.79	yes
LS-MS	6.27	yes
LS-HS	8.03	yes
MS-HS	1.76	no

Scheffé: Passive Smoking Example cont.

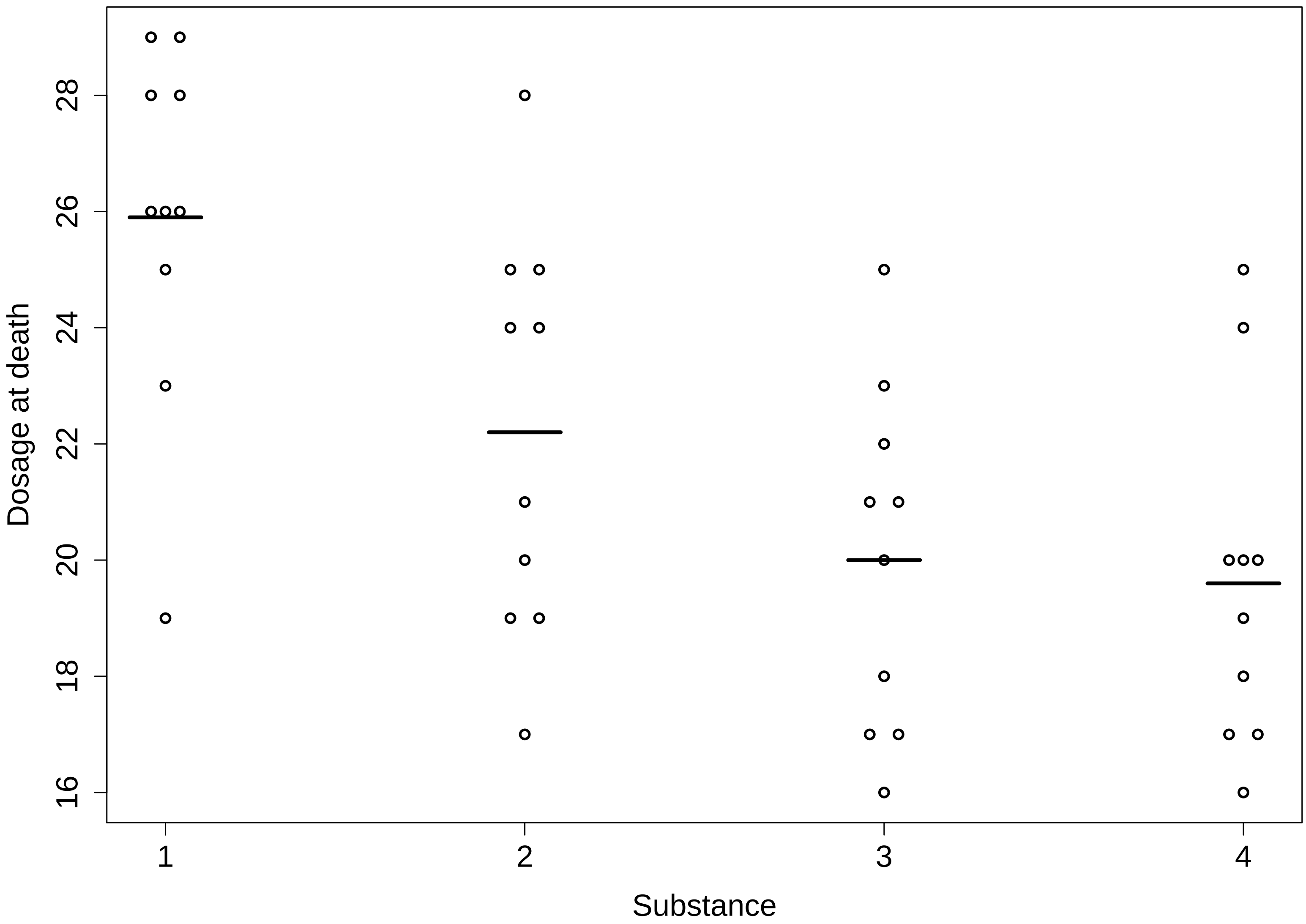
- Overall conclusions about similarities and differences across the population means indicated using schematic diagram
- Use overbars to connect means that do not differ significantly



Scheffé: Example II

- Four cardiac substances tested for relative potencies
- For each substance, ten guinea pigs anesthetized
- Outcome: dosage at death
- Data display with group means on following page

Scheffé: Example II cont.



Scheffé: Example II cont.

- Global F-test strongly rejects the null of equality of the four population means ($p = 0.0002$)
- Critical region

$$C_{0.05} = \left\{ t_{ij} : |t_{ij}| > \sqrt{3F_{3,36,0.95}} = \sqrt{3 \times 2.866} = 2.93 \right\}$$

- Note that in this example the denominator of t_{ij} is always $\sqrt{\text{MSE}/5} = 1.396$ because all group sizes are 10
- So we could also write the critical region in terms of the *minimum significant difference*

$$C_{0.05} = \{ |\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| > 2.93 \times 1.396 = 4.09 \}$$

Scheffé: SAS

```
proc glm;  class group;  model dose=group;  means group/scheffe;
(proc anova with the same statements yields the same output as below)

                         Scheffe's Test for dose

       NOTE: This test controls the Type I experimentwise error rate.

             Alpha                                   0.05
             Error Degrees of Freedom                  36
             Error Mean Square                   9.747222
             Critical Value of F                  2.86627
             Minimum Significant Difference        4.0942

        Means with the same letter are not significantly different.

         Scheffe Grouping          Mean      N    group
                        A        25.900     10    1
                        A
                   B    A        22.200     10    2
                   B
                   B             20.000     10    3
                   B
                   B             19.600     10    4
```

ANOVA: Scheffé

- For each pair of means, we can also compute multiplicity adjusted confidence intervals using Scheffé's method

$$\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} \pm \sqrt{\mathrm{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \times \sqrt{(K-1)F_{K-1,N-K,1-\alpha}}$$

- The probability is at least $1 - \alpha$ that these intervals simultaneously straddle the corresponding population mean differences

- What happens when $K = 2$?

- For the cardiac substance example,

$$\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} \pm 4.09$$

Scheffé: SAS

```
proc glm;  class group;  model dose=group;  means group/scheffe cldiff;
(proc anova with the same statements yields the same output as below)
```

Scheffe's Test for dose

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	3.700	-0.394	7.794	
1 - 3	5.900	1.806	9.994	***
1 - 4	6.300	2.206	10.394	***
2 - 1	-3.700	-7.794	0.394	
2 - 3	2.200	-1.894	6.294	
2 - 4	2.600	-1.494	6.694	
3 - 1	-5.900	-9.994	-1.806	***
3 - 2	-2.200	-6.294	1.894	
3 - 4	0.400	-3.694	4.494	
4 - 1	-6.300	-10.394	-2.206	***
4 - 2	-2.600	-6.694	1.494	
4 - 3	-0.400	-4.494	3.694	

ANOVA: Tukey

- Alternative multiple comparisons approach to Scheffé
- Critical region

$$C_\alpha = \left\{ t_{ij} : |t_{ij}| > (q_{K,N-K,1-\alpha})/\sqrt{2} \right\}$$

where $q_{k,m,1-\alpha}$ is the $1-\alpha$ quantile of the *studentized range*; see qtukey in R and probmc('Range', ...) in SAS

- Multiplicity adjusted CIs

$$\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} \pm \sqrt{\text{MSE} \times 2/n} \times (q_{K,N-K,1-\alpha})/\sqrt{2}$$

Note that the multiplicity adjusted CIs here assume a balanced design, that is, $n_i = n$ for all i

ANOVA: Tukey

- What is the studentized range?
- Suppose $Y_1, \ldots, Y_k$ iid $N(\mu, \sigma^2)$
- Let s be an estimator for σ with m degrees of freedom, $s \perp Y_1, \ldots, Y_k$
- Then
$$\frac{Y_{(k)} - Y_{(1)}}{s}$$
has a studentized range distribution with parameters k and m

ANOVA: Tukey

- Cardiac substance example with $\alpha = 0.05$

$$q_{K,N-K,1-\alpha}/\sqrt{2} = q_{4,36,0.95}/\sqrt{2} = 2.69$$

- Compared with the Scheffé critical value (2.93), easier to reject; equivalently, Tukey confidence intervals will be narrower
- For this reason, Tukey is preferred to Scheffé in balanced designs where all pairwise comparisons are being considered
- Otherwise, use Scheffé or Bonferroni-type method (later in this section)

Tukey: SAS

```
proc glm;  class group;  model dose=group;  means group/tukey cldiff;
(proc anova with the same statements yields the same output as below)
```

Tukey's Studentized Range (HSD) Test for dose

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	3.700	-0.060	7.460	
1 - 3	5.900	2.140	9.660	***
1 - 4	6.300	2.540	10.060	***
2 - 1	-3.700	-7.460	0.060	
2 - 3	2.200	-1.560	5.960	
2 - 4	2.600	-1.160	6.360	
3 - 1	-5.900	-9.660	-2.140	***
3 - 2	-2.200	-5.960	1.560	
3 - 4	0.400	-3.360	4.160	
4 - 1	-6.300	-10.060	-2.540	***
4 - 2	-2.600	-6.360	1.160	
4 - 3	-0.400	-4.160	3.360	

Tukey: R

```
> group <- as.factor(group)
> fit <- aov(dose ~ group)
> TukeyHSD(fit,"group")


  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = dose ~ group)

$group
    diff        lwr         upr     p adj
2-1 -3.7  -7.460351  0.06035128 0.0551754
3-1 -5.9  -9.660351 -2.13964872 0.0008587
4-1 -6.3 -10.060351 -2.53964872 0.0003701
3-2 -2.2  -5.960351  1.56035128 0.4048758
4-2 -2.6  -6.360351  1.16035128 0.2621133
4-3 -0.4  -4.160351  3.36035128 0.9916615
```

Bonferroni Method

- Let $A_1, A_2, \ldots, A_n$ be a set of events
- Bonferroni inequality

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} \Pr(A_i)$$

- Let A_i be the event that we reject H_{0i} when H_{0i} is true for $i = 1, 2, \ldots, n$

$$\Pr(A_i) = \alpha_i$$

Bonferroni Method

- Probability of at least one Type I error

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) \;\leq\; \sum_{i=1}^{n} \alpha_i$$

- If $\alpha_i = \alpha^*$ for all i,

$$\sum_{i=1}^{n} \alpha_i = n\alpha^*$$

- If we want $\Pr(A_1 \cup \cdots \cup A_n) \leq \alpha$, choose $\alpha^* = \alpha/n$
- For ANOVA with K groups, there are $\binom{K}{2}$ tests; therefore

$$\alpha^* = \frac{\alpha}{\binom{K}{2}}$$

Bonferroni Method: Passive Smoking Example

- $K = 6;\ \binom{6}{2} = 15$
- $\alpha^* = 0.05/15 = 0.0033$
- Two-sided test,

$$\alpha^*/2 = 0.00167$$

- Rejection region

$$C_\alpha = \{|t_{ij}| > t_{N-K,1-\alpha^*/2} = t_{1044,0.9983} = 2.94\}$$

Bonferroni Method

- In SAS proc glm, **means group/bon;**
- In R, pairwise.t.test(...,p.adj="bonf")
- Sometimes called the *least-significant difference (LSD)* method (Kleinbaum et al. *Applied Regression Analysis* 3rd edition)
- Applicable well beyond ANOVA
- Choice of $\alpha_i = \alpha/\binom{K}{2}$ for all i is standard, but not necessary

Bonferroni Method

- Definition 12.1: The significance level at which each test or comparison is carried out in an experiment is call the *per comparison error rate (PCER)*
- Bonferroni uses
$$\text{PCER} = \frac{\alpha}{\binom{K}{2}}$$
to ensure
$$\text{PEER} \leq \alpha$$
- Bonferroni-type improvements (Sidak, Holm, Hochberg, Westfall and Young) available; proc glm and proc multtest; beware dependencies in test statistics

Generalizations

- Up to this point we have considered all pairwise comparisons of means
- Other parameter combinations may be of interest
- For instance ...

Factor Level Means

- Single factor level mean

$$\frac{\bar{Y}_{i\cdot} - \mu_i}{\sqrt{\text{MSE}/n_i}} \sim t_{N-K}$$

- $100(1-\alpha)\%$ CI for μ_i

$$\bar{Y}_{i\cdot} \pm t_{N-K;1-\alpha/2}\sqrt{\text{MSE}/n_i}$$

- Testing $H_0 : \mu_i = c$ vs. $H_A : \mu_i \neq c$

$$t_i = \frac{\bar{Y}_{i\cdot} - c}{\sqrt{\text{MSE}/n_i}} \sim t_{N-K}$$

$$C_\alpha = \{t_i : |t_i| > t_{N-K;1-\alpha/2}\}$$

Linear Combinations and Contrasts

- *Linear combination*

$$L = \sum_{i=1}^{K} c_i \mu_i$$

- This is a *contrast* if $\sum_i c_i = 0$
- Estimator

$$\hat{L} = \sum_{i=1}^{K} c_i \bar{Y}_{i\cdot}$$

- Compute CIs and test statistics using

$$\frac{\hat{L} - L}{\sqrt{\mathrm{MSE} \sum_i c_i^2 / n_i}} \sim t_{N-K}$$

Conclusion

- Factor level means, that is, $\mu_1, \mu_2, \ldots$: Use Bonferroni; in SAS, proc glm/anova with **means group/bon clm;**
- Pairwise comparisons: If balanced, use Tukey; otherwise, if the number of comparisons is not too large and planned *a priori*, use Bonferroni
- Contrasts: Use Scheffé or Bonferroni;

  for example, multiplicity adjusted CIs for a family of contrasts of the form

$$\hat{L} \pm \sqrt{\text{MSE} \sum_i c_i^2/n_i} \times \sqrt{(K-1)F_{K-1,N-K;1-\alpha}}$$

- Linear combinations: Use Bonferroni