Lecture 8: Comparison Among Means

• 8. 1 What if F is significant?

- Rejection of H_0 is good to know, but not very informative

    - The null hypothesis could be false in many different ways

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$\mu_1 \neq (\mu_2 = \mu_3) \neq (\mu_4 = \mu_5)$$

$$(\mu_1 = \mu_2) \neq (\mu_3 = \mu_4) \neq \mu_5$$

- F is non-directional. We don't know which groups are different

- F is an omnibus test. The F tells us that something is going on…but it doesn't say where the effect is

    - We need more *focused comparisons*

- A solution: t tests?

    - If $\alpha = .05$, about 1 out of every 20 t-tests might result in a significant effect even if there is no effect in the population

    - Every t-test has its associated chance of a Type I error

    - With comparisons, we have a family of tests on the same data. Want to know the probability of at least 1 Type I error in the family of tests. Such a probability is called *familywise error rate*

    - Therefore, *many t*-tests can have a large *familywise error rate*

        - For independent comparisons

$$\alpha_{FW} = 1 - (1 - \alpha')^c$$

c = number of comparisons

- For $J = 3$, $\alpha_{FW} = 1 - (1 - \alpha')^c = 1 - (1 - .05)^3 = .143$

- For $J = 10$, $\alpha_{FW} = 1 - (1 - \alpha')^c = 1 - (1 - .05)^{10} = .40$

- ***Alternatives: Planned Comparisons or Post Hoc tests***

  - *Planned Comparisons or Contrasts*

    - Planned before the study; can be used instead of overall F test

    - Based on α

    - Orthogonal planned contrasts

    - Trend analysis

  - *Post Hoc or Incidental tests*

    - Based on α_{FW}

    - Use after significant overall F test to investigate specific means…no specific plan before study

    - Post hoc tests involve comparing every mean to every other mean, so they tend to have alpha inflation problems

    - The Bonferroni procedure

    - Fisher's *least significant difference* (LSD) test

    - Tukey's *honestly significant difference* (HSD) test (pairwise)

    - Scheffe (complex comparisons)

• 8. 2 Planned Comparisons

- Based on α

- A contrast is a weighted linear combination of means

$$\Psi = c_1\mu_1 + c_2\mu_2 + \ldots + c_j\mu_j = \Sigma c_j\mu_j$$

- Contrasts will be orthogonal and independent from the grand mean if

$$\Sigma c_j c'_j = c_1 c'_1 + c_2 c'_2 + \ldots + c_j c'_j = 0$$

and the sum of all weights must equal zero; $\Sigma c_j = 0$

- Estimate of a contrast

$$\hat{\psi} = \Sigma c_j \overline{X}_j = c_1 \overline{X}_1 + c_2 \overline{X}_2 + \ldots + c_j \overline{X}_j$$

- Example ($J = 3$)

- Simple contrast (comparing 2 means)

$$\hat{\psi} = \Sigma c_j \overline{X}_j = (1)\ \overline{X}_1 + (-1)\overline{X}_2 + (0)\overline{X}_3 = \overline{X}_1 - \overline{X}_2$$

- Complex contrast (comparing $\overline{X}_1$ with $\overline{X}_2$ and $\overline{X}_3$; e.g., control vs. 2 treatment groups)

$$\hat{\psi} = \Sigma c_j \overline{X}_j = (1)\overline{X}_1 + (-1/2)\overline{X}_2 + (-1/2)\ \overline{X}_3 = \overline{X}_1 - (\overline{X}_2 + \overline{X}_3) / 2$$

Example -- Data

	A_1	A_2	A_3	A_4
	22	26	28	21
	15	27	31	21
	17	24	27	26
	18	23	26	20
$\overline{X}_j$	18	25	28	22

Example – Contrasts (3 possible comparisons)

	A_1	A_2	A_3	A_4
c_1	1/2	1/2	-1/2	-1/2
c_2	1	-1	0	0
c_3	0	0	1	-1

$\Psi_1 = (.5 \text{ x } 18) + (.5 \text{ x } 25) – (.5 \text{ x } 28) – (.5 \text{ x } 22) = -3.5$

$\Psi_2 = (1 \text{ x } 18) – (1 \text{ x } 25) = -7$

$\Psi_3 = (0 \text{ x } 18) + (0 \text{ x } 25) + (1 \text{ x } 28) – (1 \text{ x } 22) = 6$

- ***Sampling Variance of Planned Comparisons***

- The sample comparison is an unbiased estimate of the population comparison

$$E(\hat{\psi}) = \Psi$$

- The variance of the sampling distribution of the comparison

$$Var(\hat{\psi}) = \sum_j c_j^2 \operatorname{var}(\overline{Y}_j) = \sigma_e^2 \sum_j \frac{c_j^2}{n_j}$$

- Sampling variance will be large when within cells variance is large, the weights are large, and the number of people in each cell is small. Estimated by:

$$\text{est. } Var(\hat{\psi}) = (MS_W)\sum_j \frac{c_j^2}{n_j} \quad \text{(substitute } MS_W \text{ for } \sigma_e^2\text{)}$$

- Standard error of a contrast

$$SE_\Psi = \sqrt{MS_W\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots \frac{c_j^2}{n_j}\right)} = \sqrt{MS_W\left(\frac{\Sigma c_j^2}{n_j}\right)}$$

t-ratio ($df = N - J$)

$$t = \frac{\hat{\Psi}}{s_\Psi}$$

- Point estimate: $\hat{\psi}$

- Interval Estimate: $\hat{\psi} \pm (t_{\text{conf}})\, s_\Psi$

- ***Significance Test*** (from example)

Source	*SS*	*df*	*MS*	*F*
Between	219	3	73	12.17
Within	72	12	6	
Total	291	15		

- For the 1st comparison

$$\hat{\Psi}_1 = (.5 \times 18) + (.5 \times 25) - (.5 \times 28) - (.5 \times 22) = -3.5$$

$$est.\mathrm{var}(\hat{\Psi}) = 6\frac{.5^2 + .5^2 + (-.5)^2 + (-.5)^2}{4} = \frac{3}{2}$$

$$est.SE(\hat{\Psi}) = \sqrt{\frac{3}{2}} = 1.2247$$

$$t = \frac{\hat{\Psi}}{\sqrt{est.\mathrm{var}(\hat{\Psi})}} = \frac{-3.5}{1.2247} = -2.86$$

$df = N - J;\ 16 - 4 = 12$

$t_{(crit, \alpha=.05,\ df = 12)} = 2.18$

$t\ (12) = -2.86,\ p < .05$

95% CI = $-3.5 \pm 2.18\ (1.2247) = (-6.17, -.83)$

- For the 2nd comparison

$$\hat{\Psi}_2 = (1\times 18) - (1\times 25) + (0\times 28) + (0\times 22) = -7$$

$$est.\,\mathrm{var}(\hat{\Psi}) = 3\,;\ est.SE(\hat{\Psi}) = 1.7321$$

$$t = \frac{\hat{\Psi}}{\sqrt{est.\,\mathrm{var}(\hat{\Psi})}} = \frac{-7}{1.7321} = -4.04$$

95% CI = -7 $\pm$ 2.18 (1.7321) = (-10.77, -3.23)

- For the 3rd comparison

$$\hat{\Psi}_3 = (0\times 18) + (0\times 25) + (1\times 28) - (1\times 22) = 6$$

$$t = \frac{\hat{\Psi}}{\sqrt{est.\,\mathrm{var}(\hat{\Psi})}} = \frac{6}{1.7321} = 3.46$$

95% CI = 6 $\pm$ 2.18 (1.7321) = (2.23, 9.77)

- ***Independence of Planned Comparisons***

  - Several planned comparisons can be made from the same data

    - Some are independent; some are not

  - Two comparisons from a normal population with equal sample sizes in each cell are independent if the sum of the products of weights is zero

    $$\Sigma c_j c'_j = c_1 c'_1 + c_2 c'_2 + \ldots + c_j c'_j = 0$$

  - If the sample sizes are not equal

    $$\sum_j \frac{c_{1j} \cdot c_{2j}}{n_j} = 0$$

Example (a)

	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	Σ
c_1	1	-1	0	0
c_2	1	-1/2	-1/2	0
$c_j c'_j$	1	1/2	0	1 1/2

the contrasts are equal but not independent; they are correlated…we need to be concerned with α

Example (b)

	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	Σ
c_1	1	-1	0	0
c_2	1/2	1/2	-1	0
$c_j c'_j$	1/2	-1/2	0	0

the contrasts are orthogonal (independent from each other)

- When the tests are correlated, if we make a type I error, we are very likely to carry over more errors because the tests are correlated. So, we don't know exactly what the total alpha error is

- When the tests are independent, we know what the alpha error is so we can draw conclusions more safely

- How many orthogonal contrasts? J-1; each contrast has 1 *df*

- For orthogonal contrasts

$$SS_{\mathrm{B}} = SS_{C1} + SS_{C2} + \ldots + SS_{CJ\text{-}1}$$

- **8. 3 Trend Analysis**

- If the IV is ordered (e.g., time, age, drug dosage)

- The goal is to find whether there is a functional relation between the IV and the DV

  - linear, quadratic, cubic, quartic, mixed, etc.

- It is a special case of orthogonal planned contrasts ($J-1$ orthogonal trends or polynomials)

Linear	$Y = a + bX$
Quadratic	$Y = a + bX + cX^2$
Cubic	$Y = a + bX + cX^2 + dX^3$

• 8. 4 The Bonferroni Procedure

- The Bonferroni procedure applies to any kind of comparison (i.e., orthogonal or nonorthogonal)

$$\alpha' = \frac{\alpha_{FW}}{c}$$

c = number of comparisons

- It is a "conservative" test – the Bonferroni procedure rejects too *few* hypotheses of equal means

- Good when there is a limited number of comparisons

- If there is a large number of comparisons, α_{FW} becomes extremely low

- Example, if α_{FW} is desired at .05 and c = 5

$\alpha_{FW} = .05/5 = .01$

- Use the adjusted alpha (e.g., .01) for each comparison

• 8. 5 Post Hoc Tests – Tukey's Honestly Significant Difference (HSD) Test

- Tukey's *honestly significant difference* (HSD) test does not inflate alpha very much

$$q = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{MS_W}{\tilde{n}}}}$$

where ñ is the *harmonic mean* of the two sample sizes

$$\tilde{n} = \frac{1}{\frac{1}{n_1} + \frac{1}{n_2} + \ldots \frac{1}{n_k}}$$

- It is hard to compute but not biased (not too liberal, not too conservative)

- It applies to pairwise comparisons

- Based on α_{FW}

- Studentized Range statistic (As *n* increases, so does the range)

$$q = \frac{\overline{X}_{l\arg est} - \overline{X}_{smallest}}{\sqrt{\frac{MSw}{N}}}$$

- It follows a studentized range distribution with df = (J, $N - J$)

$$HSD = \frac{critical_q}{\sqrt{\frac{MSw}{N}}}$$

If $\overline{X}_j - \overline{X}_j' \geq$ HSD, reject H_0

- HSD = honestly significant difference. For HSD, use $k = J$, the number of groups in the study. Choose α and find the *df* for error. Look up the value q_α. Then find the value:

$$HSD = q_\alpha \sqrt{\frac{MS_{error}}{n}}$$

- Compare HSD to the absolute value of the difference between all pairs of means. Any difference larger than HSD is significant

• 8. 6 Post Hoc Tests – Scheffe Test

- Applies to any kind of contrasts. Follows same calculations, but uses different critical values

$$t = \frac{\hat{\Psi}}{\sqrt{est.\,\mathrm{var}(\hat{\Psi})}}$$

- It is a very conservative test. Instead of comparing the test statistic to a critical value of *t*, use:

$$S = \sqrt{(J-1)F_\alpha}$$

where *F* comes from the overall *F* test ($J - 1$ and $N - J$ *df*)

- Test any linear combination of means

- Based on α_{FW}

- The probability of a type II error is very high but not the type I error

(previous example)

Source	*SS*	*df*	*MS*	*F*
Between	219	3	73	12.17
Within	72	12	6	
Total	291	15		

$$\hat{\Psi}_1 = (.5 \times 18) + (.5 \times 25) - (.5 \times 28) - (.5 \times 22) = -3.5$$

$$est.\,\mathrm{var}(\hat{\Psi}) = 6\frac{.5^2 + .5^2 + (-.5)^2 + (-.5)^2}{4} = \frac{3}{2}\,;\; est.SE(\hat{\Psi}) = \sqrt{\frac{3}{2}} = 1.2247$$

$$t = \frac{\hat{\Psi}}{\sqrt{est.\,\mathrm{var}(\hat{\Psi})}} = \frac{-3.5}{1.2247} = -2.86\,;\; F_{(\alpha=.05;\,3,12)} = 3.49$$

$$S = \sqrt{(J-1)F_\alpha} = \sqrt{(4-1) \times 3.49} = 3.24$$

The comparison is not significant because |-2.86| < 3.24.

- ## 8. 7 Effect Size and Practical Significance

- A significant F simply means we had enough power to reject the null hypothesis of no effect. That says nothing about how <u>important</u> the effect is – only that it is probably not due to chance

- We need a measure of *practical* significance to supplement our test of *statistical* significance

- Eta-squared $\eta^2 = \dfrac{SS_B}{SS_T}$

  - It is the proportion of the total variability of the data that is accounted for by the treatment effect (also called R^2)

  - It varies from 0 (no effect) to 1 (no error)

    in the example of drugs and anxiety $\eta^2 = \dfrac{4299.6}{7354} = .58$

  η^2 is positively biased (overestimates the true effect), with larger bias for a larger number of groups and smaller sample sizes

- Omega-squared $\omega^2 = \dfrac{SS_B - (J-1)MS_W}{SS_T + MS_W}$

  - It is the proportion of variance accounted for – with a correction factor

    in the example of $\eta^2 = \dfrac{4299.6 - (3-1)254.43}{7354 + 254.43} = .50$