Multiple Comparison Procedures
Cohen Chapter 13

For EDUC/PSY 6600

"We have to go to the deductions and the inferences," said Lestrade, winking at me. "I find it hard enough to tackle facts, Holmes, without flying away after theories and fancies."

Inspector Lestrade to Sherlock Holmes

The Boscombe Valley Mystery

ANOVA Omnibus: Significant F-ratio

- **Factor (IV) had effect on DV**
  - **Groups are not from same population**

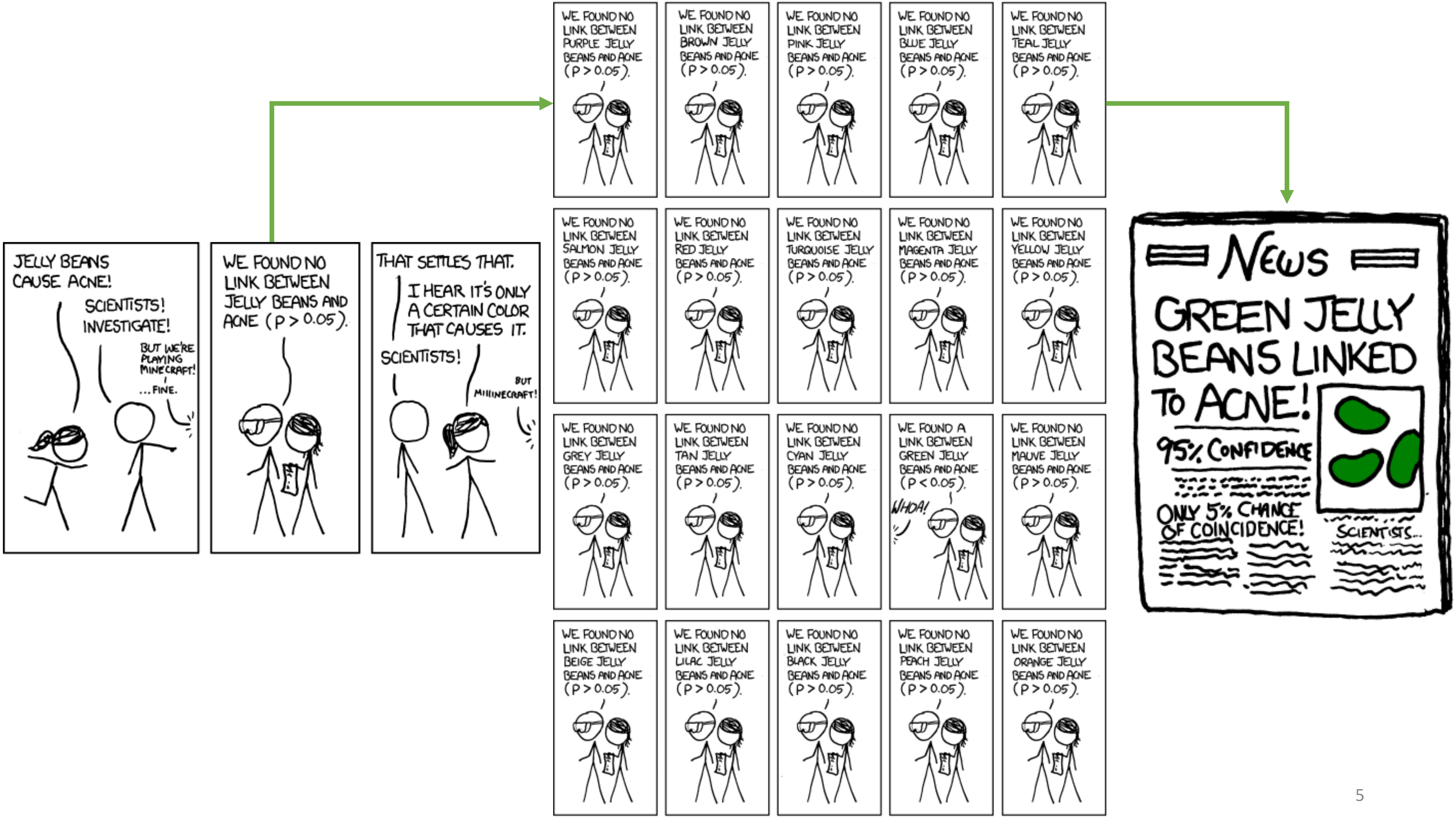
- **Which levels of factor differ?**
- **Must compare and contrast means from different levels**
- **Indicates ⩾ 1 significant difference among all <u>POSSIBLE</u> comparisons**

- **<u>Simple *vs.* complex comparisons</u>**
  - **Simple comparisons**
    - **Comparing 2 means, pairwise**
    - **Possible for no 'pair' of group means to significantly differ**

  - **Complex comparisons**
    - **Comparing combinations of > 2 means**

Multiple Comparison Procedure

- ‘Multiple comparison procedures’ used to detect simple or complex differences
- Significant omnibus test NOT always necessary
  - Inaccurate when assumptions violated
  - Type II error
- OKAY to conduct multiple comparisons when *p*-value CLOSE to significance



JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
News
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Error Rates

- $\alpha = p$(Type I error)
  - Determined in study design
  - Generally, $\alpha = .01, .05,$ or $.10$

Experimentwise (α_{EW})

$p(\geqslant 1$ Type I error for all comparisons)

comparison error rate (α_{PC})

$\alpha = \alpha_{PC}$

α_{PC} = Error rate for any 1 comparison

Relationship between α_{PC} and α_{EW}

$\alpha_{EW} = 1 - (1 - \alpha_{PC})^c$

c = Number of comparisons

$(1 - \alpha_{PC})^c = p$(NOT making Type I error over c)

Error rates

- <u>ANOVA with 4 groups</u>
  - *F*-statistic is significant
  - Comparing each group with one another
    - $c = 6$
    - $\alpha_{PC} = .05$
    - $\alpha_{EW} =$ _____
    - α_{EW} when $c = 10$?

$\bar{X}_1 vs. \bar{X}_2$

$\bar{X}_1 vs. \bar{X}_3$

$\bar{X}_1 vs. \bar{X}_4$

$\bar{X}_2 vs. \bar{X}_3$

$\bar{X}_2 vs. \bar{X}_4$

$\bar{X}_3 vs. \bar{X}_4$

- <u>3 Options…</u>
  - Ignore α_{PC} or α_{EW}
  - Modify α_{PC}
  - Modify α_{EW}

Comparisons

Post hoc (a posteriori)	Pre Planned (a priori)
Selected after data collection and analysis	Selected before data collection
Used in exploratory research	Follow hypotheses and theory
Larger set of or all possible comparisons	Justified conducting ANY planned comparison (ANOVA doesn't need to be significant)
Inflated α_{EW}: Increased p(Type I error)	α_{EW} is much smaller than alternatives α_{EW} can slightly exceed α when planned Adjust when c is large or includes all possible comparisons?

Problems with comparisons

- Decision to statistically test certain post hoc comparisons made <u>**after**</u> examining data
  - When only 'most-promising' comparisons are selected, need to correct for inflated *p*(Type I error)
  - Biased sample data often deviates from population

- When <u>all</u> possible pairwise comparisons are conducted, *p*(Type I error) or α_{EW} is same for *a priori* and *post hoc* comparisons

For example, a significant *F*-statistic is obtained:

Assume 20 pairwise comparisons are possible

- But, in population, no significant differences exist
- Made a Type I error obtaining significant *F*-statistic
- However, a *post hoc* comparison using sample data suggests largest and smallest means differ

If we had conducted 1 planned comparison

- 1 in 20 chance (α = .05) of conducting this comparison and making a type I error

If we had conducted all possible comparisons

- 100% chance (α = 1.00) of conducting this comparison and making a type I error
- If researcher decides to make only 1 comparison after looking at data, between largest and smallest means, chance of type I error is still 100%
  - All other comparisons have been made 'in head' and this is only one of all possible comparisons
  - Testing largest *vs.* smallest means is probabilistically similar to testing all possible comparisons

Common techniques

***a priori* tests**

- Multiple t-tests
- Bonferroni (Dunn)
- *Dunn-Ŝidák**
- *Holm**
- Linear contrasts

*adjusts α_{PC}
Italicized: not covered

***post hoc* tests**

- Fisher LSD
- Tukey HSD
- Student-Newman-Keuls (SNK)
- Tukey-b
- Tukey-Kramer
- Games-Howell
- Duncan’s
- Dunnett’s
- REGWQ
- Scheffé

Common techniques

***a priori* test**[...]

- Multiple *t*-t[...]
- Bonferroni
- *Dunn-Ŝidák*
- *Holm**
- Linear cont[...]

*adjusts α_{PC}
Italicized: not covered

– REGWQ
– Scheffé

Many more comparison techniques available

Most statistical packages make no *a priori / post hoc* distinction

All called *post hoc* (SPSS) or multiple comparisons (R)

In practice, most *a priori* comparison techniques can be used as *post hoc* procedures

Called post hoc, not because they were planned after doing the study per se, but because they are conducted after an omnibus test

A Priori procedures: multiple t-tests

- **Homogeneity of variance**
  - $\boldsymbol{MS_W}$ (estimated pooled variance) and $\boldsymbol{df_W}$ (both from ANOVA) for critical value (smaller F_{crit})

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{MS_W}{n_1} + \frac{MS_W}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2MS_W}{n_j}}}$$

- Heterogeneity of variance and equal n
  - Above equation: Replace MS_W with s_j^2 and df_W with $df = 2(n_j - 1)$ for t_{crit}
- Heterogeneity of variance and unequal n
  - Above equation: Replace MS_W with s_j^2 and df_W with Welch-Satterwaite df for t_{crit}

A Priori procedures: Bonferroni (Dunn) t-test

- Bonferroni inequality
  - p(occurrence for set of events (additive) $\leqslant \sum$ of probabilities for each event)
- Adjusting α_{PC}
  - Each comparison has p(Type I error) = $\alpha_{PC} = .05$
  - $\alpha_{EW} = .05$
  - $\alpha_{EW} \leqslant c^*\alpha_{PC}$
    - $p(\geqslant 1$ Type I error) can never exceed $c^*\alpha_{PC}$
- Conduct standard independent-samples t-tests per pair

Example for 6 comparisons:
α_{PC} = .05/6 = .0083

A Priori procedures: Bonferroni (Dunn) t-test

t-tables lack Bonferroni-corrected critical values

- Software: Exact p-values
- Is exact p-value ⩽ Bonferroni-corrected α-level?

Example for 6 comparisons:
α_{PC} **= .05/6 = .0083**

More conservative: Reduced p(Type I error)

Less powerful: Increased p(Type II error)

A Priori procedures: linear contrasts - idea

- Linear combination of means:

$$L = c_1\bar{X}_1 + c_2\bar{X}_2 + \cdots + c_k\bar{X}_k = \sum_{i=1}^{k} c_j\bar{X}_j$$

  - Each group mean weighted by constant (c)
  - Products summed together

- Weights selected so means of interest are compared

- Sum of weights = 0

Example 1: 4 means

Compare M_1 to M_2, ignore others

$c_1 = 1$, $c_2 = -1$, $c_3 = 0$, $c_4 = 0$

$$L = (1)\bar{X}_1 + (-1)\bar{X}_2 + (0)\bar{X}_3 + (0)\bar{X}_4 = \bar{X}_1 - \bar{X}_2$$

Example 2: Same 4 means

Compare M_1, M_2, and M_3 to M_4

$c_1 = 1/3$, $c_2 = 1/3$, $c_3 = 1/3$, $c_4 = -1$

$$L = (1/3)\bar{X}_1 + (1/3)\bar{X}_2 + (1/3)\bar{X}_3 + (-1)\bar{X}_4 = \frac{(\bar{X}_1 + \bar{X}_2 + \bar{X}_3)}{3} - \bar{X}_4$$

A Priori procedures: linear contrasts - SS

- Each linear combination: $SS_{Contrast}$

<u>Equal *ns*:</u>

$$SS_{Contrast} = \frac{n_j L^2}{\sum_{j=1}^{k} c_j^2} = \frac{n_j (\sum_{j=1}^{k} c_j \bar{X}_j)^2}{\sum_{j=1}^{k} c_j^2}$$

<u>Unequal *ns*:</u>

$$SS_{Contrast} = \frac{L^2}{\sum_{j=1}^{k} \left(\frac{c_j^2}{n_j} \right)} = \frac{(\sum_{j=1}^{k} c_j \bar{X}_j)^2}{\sum_{j=1}^{k} \left(\frac{c_j^2}{n_j} \right)}$$

- $SS_{Between}$ partitioned into k $SS_{Contrasts}$
  - $SS_{Between} = SS_{Contrast\ 1} + SS_{Contrast\ 2} + \ldots + SS_{Contrast\ k}$

$$F = \frac{MS_{Contrast}}{MS_W} = \frac{nL^2 / \sum c_j^2}{MS_W} = \frac{nL^2}{\sum c_j^2 * MS_W} \text{ or } \frac{L^2}{\sum_{j=1}^{k} \left(\frac{c_j^2}{n_j} \right) * MS_W}$$

df for $SS_B = k - 1$

df for $SS_{Contrast}$ = Number of 'groups/sets' included in contrast <u>minus</u> 1

<u>***F* = $MS_{Contrast}$ / MS_W**</u>

$MS_{Contrast} = SS_{Contrast} / df_{Contrast}$

As $df = 1$, $MS_{Contrast} = SS_{Contrast}$

MS_W from omnibus ANOVA results

<u>**Max # 'legal' contrasts = df_B**</u>

Do not need to consume all available *df*
Use smaller α_{EW} if # contrasts > df_B

A Priori procedures: linear contrasts - example

Test each Contrast (ANOVA: $SS_{Between}$ = 26.53, SS_{Within} = 22.8)

Mean	N
9.2	5
6.6	5
6.2	5

Contrast 1: $M_{No\ Noise}$ versus $M_{Moderate}$ and $M_{loud,}$

$\boldsymbol{L}$ = (-2)(9.2) + (1)(6.6) + (1)(6.2) = -18.4 + 12.8 = **-5.6**
$\boldsymbol{SS_{Contrast1}}$ = 5*$(-5.6)^2$ / $(-2^2 + 1^2 + 1^2)$ = 156.8 / 6 = **26.13**

df_B = 2 – 1 = 1 → $\mathbf{MS_{Contrast1}}$ = 26.13/1 = **26.13**
df_W = 15 – 3 = 12 → $\mathbf{MS_W}$ = 22.8/12 = **1.90**

F = 26.13/1.980 = **13.75**
P< .05

α =.05 & df_W = 12 → F_{crit} = 4.75

Note: $SS_B = SS_{Contrast1} + SS_{Contrast2} = 26.13 + 0.40 = 26.53$

A Priori procedures: linear contrasts - example

Test each Contrast (ANOVA: $SS_{Between}$ = 26.53, SS_{Within} = 22.8)

Mean	N
9.2	5
6.6	5
6.2	5

Contrast 1: $M_{No\ Noise}$ versus $M_{Moderate}$ and M_{loud},

$\boldsymbol{L}$ = (-2)(9.2) + (1)(6.6) + (1)(6.2) = -18.4 + 12.8 = **-5.6**

$SS_{Contrast1}$ = 5*(-5.6)^2 / (-2^2 + 1^2 + 1^2) = 156.8 / 6 = **26.13**

$df_B = 2 - 1 = 1 \rightarrow \mathbf{MS_{Contrast1}}$ = 26.13/1 = **26.13**

$df_W = 15 - 3 = 12 \rightarrow \mathbf{MS_W}$ = 22.8/12 = **1.90**

F = 26.13/1.980 = **13.75**

P< .05

α =.05 & df_W = 12 $\rightarrow$ F_{crit} = 4.75

Note: $SS_B = SS_{Contrast1} + SS_{Contrast2}$ = 26.13 + 0.40 = 26.53

Contrast 2: $M_{Moderate}$ versus M_{loud}

$\boldsymbol{L}$ = (0)(9.2) + (-1)(6.6) + (1)(6.2) = **-0.4**

$SS_{Contrast2}$ = 5*(-0.4)^2 / (1^2 + [-1]^2) = 0.8 / 2 = **0.40**

$df_B = 2 - 1 = 1 \rightarrow \mathbf{MS_{Contrast2}}$ = 0.40/1 = **0.40**

$df_W = 15 - 3 = 12 \rightarrow \mathbf{MS_W}$ = 22.8/12 = **1.90**

F = 0.40/1.90 = **0.21**

P > .05

A Priori procedures: linear contrasts - Orthogonal

- Independent (orthogonal) contrasts
  - If M_1 is larger than average of M_2 and M_3
  - Tells us nothing about M_4 and M_5

- Dependent (non-orthogonal) contrasts
  - If M_1 is larger than average of M_2 and M_3
  - Increased probability that $M_1 > M_2$ or $M_1 > M_3$

Can conduct non-orthogonal contrasts, but…
Dependency in data
Inefficiency in analysis
Contain redundant information
Increased p(Type I error)

A Priori procedures: linear contrasts - Orthogonal

- *Orthogonality indicates $SS_{Contrasts}$ are independent partitions of SS_B*

- **Orthogonality obtained when**
  - Σ of $SS_{Contrasts} = SS_{Between}$
  - Two rules are met:
  - Rule 1: $\sum_{j=1}^{k} c_j = 0$ Rule 2: $\sum_{j=1}^{k} c_{1j} c_{2j} c_{Lj} = 0$

where c_{Lj} = Contrast weights from additional linear combinations

- From example…Orthogonal!
  - Rule 1: L_1 = (1)+(1)+(-2) = 0; L_2 = 1+(-1)+(0) = 0
  - Rule 2: -2*0 + 1*1 + 1*-1 = 1 + -1 + 0 = 0

A Priori procedures: recommendations

- **1 pairwise comparison of interest**
  - Standard *t*-test

- **Several pairwise comparisons**
  - Bonferroni, Multiple *t*-tests
  - Bonferroni is most widely used (varies by field), and can be used for multiple statistical testing situations

- **1 complex comparison**
  - Linear contrast

- **Several complex comparisons**
  - Orthogonal linear contrasts – no adjustment
  - Non-orthogonal contrasts – Bonferroni correction or more conservative α_{PC}

Post hoc procedures: Fisher's LSD Test

Aka: Fisher's Protected *t*-test = Multiple *t*-test

- Conduct as described previously: 'multiple *t*-tests'
  - 'Fisher's LSD test': Only after significant F_{stat}
  - 'Multiple *t*-test': Planned *a priori*

- One advantage is that equal *ns* are not required

Logic

If H_0 true and all means equal one another, significant overall F-statistic ensures α_{EW} is fixed at α_{PC}

Powerful: No adjustment to α_{PC}

Most liberal *post hoc* comparison

Highest p(Type I error)

Not recommended in most cases

Only use when $k = 3$

Post hoc procedures: studentized range q

- t-distribution derived under assumption of comparing only 2 sample means
  - With >2 means, sampling distribution of t is NOT appropriate as p(Type I error) > α
- Need sampling distributions based on comparing multiple means
- Studentized range q-distribution
  - k random samples (equal n) from population
  - Difference between high and low means
  - Differences divided by $\sqrt{\frac{MS_W}{n_j}}$
  - Obtain probability of multiple mean differences
  - Critical value varies to control α_{EW}

Rank order group means (low to high)

- ***r*** **= <u>Range</u> or distance between groups being compared**
  - **4 means: Comparing M_1 to M_4, $r = 4$; comparing M_3 to M_4, $r = 2$**
- **Not part of calculations, used to find critical value**

q_{crit}: Use r, df_W from ANOVA, and α

- **q_{crit} always positive**

Most tests of form:

$$q = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{MS_W}{n_j}}}$$

Post hoc procedures: studentized range q

r

df_w

q_{crit}

Table A.11
Critical Values of the Studentized Range Statistic (q) for α = .05

df for Error Term	Number of Groups (or Number of Steps Between Ordered Means)																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83	59.56
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.15	10.35	10.52	10.69	10.84	10.98	11.11	11.24
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

Post hoc procedures: studentized range q

- **Note square root of 2 missing from denominator**
  - Each critical value (q_{crit}) in q-distribution has already been multiplied by square root of 2

$$q = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{MS_W}{n_j}}} \qquad \textbf{Vs.} \qquad t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{MS_W}{n_1} + \frac{MS_W}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2MS_W}{n_j}}}$$

- **Assumes all samples are of same *n***
  - Unequal *ns* can lead to inaccuracies depending on group size differences
  - If *ns* are unequal, alternatives are:
    - Compute harmonic mean (below) of *n* (if *ns* differ slightly)
    - Equal variance: Tukey-Kramer, Gabriel, Hochberg's GT2
    - Unequal variance: Games-Howell

***Post hoc* tests that rely on studentized range distribution:**

Tukey HSD

Tukey's b

S-N-K

Games-Howell

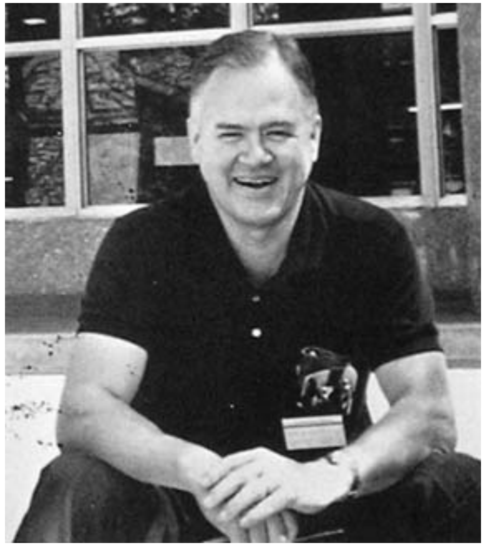
REGWQ

Duncan

Post Hoc Procedures: Tukey's HSD test

- Based on premise that Type I error can be controlled for **comparison involving largest and smallest means**, thus controlling error for all
- *Significant ANOVA <u>NOT</u> required*
- q_{crit} based on df_W, α_{EW} (table .05), and largest r
  - If we had 5 means, all comparisons would be evaluated using q_{crit} based on $r = 5$
- q_{crit} compared to q_{obt}
  - MS_W from ANOVA
- One of most conservative *post hoc* comparisons, good control of α_{EW}
- Compared to LSD…
  - HSD <u>less</u> powerful w/ 3 groups (Type II error)
  - HSD <u>more</u> conservative; less

    Type I error w/ > 3 groups
- Preferred with > 3 groups



Post Hoc Procedures: Tukey's HSD test

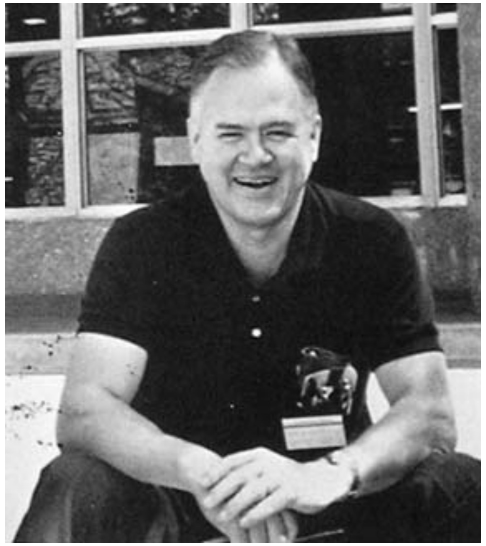
- Based on premise that Type I error can be controlled for **comparison involving largest and smallest means**, thus controlling error for all
- *Significant ANOVA*
- q_{crit} based on df_W,
  - If we had 5 me
- q_{crit} compared to q
  - MS_W from ANC
- One of most cons
- Compared to LSD
  - HSD less powerful w/ 3 groups (Type II error)
  - HSD more conservative; less

    Type I error w/ > 3 groups
- Preferred with > 3 groups

Fisher's LSD is most liberal

Tukey's HSD is nearly most conservative

Others are in-between



Post hoc: Confidence intervals: HSD

$$q = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{MS_W}{n_j}}}$$

Simultaneous Confidence Intervals for all possible pairs of populations means…at the same time!

$$\boldsymbol{\mu_i - \mu_j} = (\bar{X}_i - \bar{X}_j) \pm q\sqrt{\frac{MS_W}{n}} = (\boldsymbol{\bar{X}_i - \bar{X}_j}) \pm \boldsymbol{HSD}$$

Interval DOES INCLUDS zero → fail to reject H0: means are the same…no difference

Interval does NOT INCLUDS zero → REJECT H0 → evidence there IS a DIFFERENCE

Post hoc procedures: Scheffé Test

- Most conservative and least powerful

- Uses F- rather than t-distribution to find critical value
  - $\boldsymbol{F_{Scheffé} = (k\text{-}1)^{*}F_{crit}\ (k\text{-}1, N\text{-}k)}$
    - Scheffé recommended running his test with $\alpha_{EW} = .10$
  - $F_{Scheffé}$ is now F_{crit} used in testing

- Similar to Bonferroni; α_{PC} is computed by determining all possible linear contrasts AND pairwise contrasts

- **Not** recommended in most situations
  - Only use for **complex** ***post-hoc*** comparisons
    - Compare $F_{contrast}$ to $F_{Scheffé}$

Post hoc procedures: recommendations

- **1 pairwise comparison of interest**
  - Standard independent-samples t-test

- **Several pairwise comparisons**
  - 3 → LSD
  - > 3 → HSD or other alternatives such as Tukey-b or REGWQ
  - Control *vs*. set of Tx groups → Dunnett's

- **1 complex comparison (linear contrast)**
  - No adjustment

- **Several complex comparisons (linear contrasts)**
  - Non-orthogonal – Scheffé test
  - Orthogonal – Use more conservative α_{PC}

Analysis of trend components

- Try when the independent variable (IV) is highly ordinal or truly underlying continuous

- * LINEAR regression:
  - Run linear regression with the IV as predictor
  - Compare the F-statistic's p-value for the source=regression to the ANOVA source=between

- * CURVE-a-linear regression:
  - create a new variable that is = IV variable SQUARED
  - Run linear regression with BOTH the original IV & the squared-IV as predictors
  - Compare the F-statistic's p-value for the source=regression

Conclusion

- Not all researchers agree about best approach/methods
- <u>Method selection depends on</u>
  - Researcher preference (conservative/liberal)
  - Seriousness of making Type I *vs*. II error
  - Equal or unequal *ns*
  - Homo- or heterogeneity of variance
- Can also run mixes of pairwise and complex comparisons
- Adjusting α_{PC} to $\downarrow$ p(type I error), $\uparrow$ p(Type II error)
  - ***a priori* more powerful than *post hoc***
  - ***a priori* are better choice**
    - Fewer in number; more meaningful
    - Forces thinking about analysis in advance