36-309/749
Experimental Design for Behavioral and Social Sciences

Nov 3, 2015
Lecture 8: Contrasts and Multiple Comparisons

Contrasts (in general)

- Context: An ANOVA rejects the "overall" null hypothesis that all *k* means of some factor are not equal, i.e., H_0: $\mu_1 = \cdots = \mu_k$. When k>2, this is not satisfying (scientifically).
- Contrasts let us ask "which are significantly different?"
- Terminology: Define a "**contrast**" or "analytic comparison" that is of scientific interest, e.g. compare μ_1 to μ_5 or compare μ_2 to the average of μ_1, μ_3 and μ_5, i.e., $\left(\frac{\mu_1+\mu_3+\mu_5}{3}\right)$
- **Contrast null hypothesis**: Express as something equal to zero. For the above examples, $\mu_1 = \mu_5 \Rightarrow \mu_1 - \mu_5 = 0$ and $\mu_2 = \left(\frac{\mu_1+\mu_3+\mu_5}{3}\right) \Rightarrow \mu_2 - \left(\frac{\mu_1+\mu_3+\mu_5}{3}\right) = 0$.

2

Contrasts, cont.

- Computer package form: Re-express in ***linear combination*** form: $C_1\mu_1 + \ldots + C_k\mu_k = 0$ which contains ***all*** of the parameters in the original "overall" null hypothesis ***in order***. (The C values are called "weights" or "coefficients". Some of the C's will be negative.) Enter just the coefficients into the computer.
  - Example 1: express $\mu_1 - \mu_5 = 0$ as $(1)\mu_1+(0)\mu_2+(0)\mu_3+(0)\mu_4+(-1)\mu_5=0$. Enter 1 0 0 0 –1 into SPSS or some other computer package.
  - Example 2: express $\mu_2 - \left(\frac{\mu_1+\mu_3+\mu_5}{3}\right) = 0$ as
    $(\ \ )\mu_1 + (\ \ )\mu_2 + (\ \ )\mu_3 + (\ \ )\mu_4 + (\ \ )\mu_5 = 0$.
    Enter into SPSS:
  - Check your work: ***Valid coefficient sets always add to zero***.

3

Planned comparisons (contrasts)

- Planned comparisons **maintain type-1 error** at α (experiment-wise) only when:
  - They are chosen in advance, i.e., truly planned.
  - They are only used if the corresponding ***overall*** p-value is $\le\alpha$.
  - They number no more than the F numerator df.
    - One-way (k level) ANOVA: k-1 planned contrasts
    - Two-way (k x m) ANOVA:
      - Interaction expected: plan (k-1)(m-1) contrasts like $\mu_{A1,B1} - \mu_{A2,B3} = 0$ or the often more informative form, such as $(\mu_{A1,B1} - \mu_{A2,B1}) - (\mu_{A1,B3} - \mu_{A2,B3}) = 0$.
      - Interaction not expected: k-1 planned contrasts for factor A; m-1 for factor B.
  - They are orthogonal (often ignored): the sum of products of corresponding coefficients equal zero, i.e., they ask ***independent*** questions.

4

Planned comparisons, cont.

- ***Optional*** technical details: See gray boxes in textbook and/or ask Howard.
- In SPSS, comparisons made using the "Contrasts" button or the LMATRIX subcommand under GeneralLinearModel/Univariate are assumed to be planned, and the p-values are wrong otherwise.

Multiple (post hoc, unplanned) comparisons

- Example 0: Darts game
- Example 1: In a study of twenty chocolate lovers vs. non-chocolate eaters (freaks of nature), researchers claimed that "higher levels of phenylacetylglutamine and citrate in the chocolate-desiring group suggest that these individuals may regulate the citric acid cycle slightly differently than those who don't fancy a daily dose of chocolate." (J Proteome Research, 6(11):4469-4477, 2007) Consider performing t-tests to see if the groups differ for each detectable compound. For 50 compounds (a low, but reasonable number), if they are all unrelated to chocolate, the chance of ***avoiding*** a false positive at a rate of 0.95 each is 0.95^{50}=0.077. [Chance of getting 1 or 2 FP is 20% and 26% respectively.] Conclusion:

Multiple Comparisons, cont.

- Example 2: In a study of the effect of magic beans on health, a carefully done, well powered, randomized clinical trial measures 12 health outcomes (BP, cholesterol, etc.).

  Assuming "magic beans" are useless, the chance of *avoiding* a false positive is:

  $0.95*0.95*\cdots*0.95 = 0.95^{12} = 0.54$

  So the chance of finding at least one (meaningless) finding is 1-0.54 = 46%.

Multiple Comparisons, cont.

- Math: The number of ways you can choose 2 items from a list of k items is called "k choose 2". We use the symbol $\binom{k}{2}$ and the answer is k(k-1)/2.
- Checking all $\binom{k}{2}$ pairs in a one-way ANOVA has exactly the same problem as for "magic beans".
- Less obviously, when we pick out the smallest and largest sample means out of k means to compare, we are implicitly performing multiple comparisons, thus increasing the chance of making a type-1 error.
- The problem is also referred to as ***post-hoc*** testing, unplanned comparisons, and data snooping.

Multiple Comparisons, cont.

- Most common goal: keep the **per-experiment type-1 error rate** at 0.05 (compare with FDR). The key to appropriate, honest post-hoc comparisons is to determine the size of the family of comparisons that you are considering, and handicap yourself (e.g., lower alpha or raise the p-value) to reduce the chances of a type 1 (FP) error, which, unfortunately, is at the expense of reduced power.
- Special example: In a two-way ANOVA with interaction, at least one of the three overall null hypotheses (two main effects plus interaction) are rejected at $p \le 0.05$ about 14% of the time for null experiments if the "corrected model" p-value is not used as a "screen".

9

Multiple Comparisons, cont.

- Appropriate methods add a "penalty" for multiple comparisons
  - Bonferroni procedure: simplest and most general, but conservative (Holm's-Bonferroni is a tiny bit better). Set $\alpha' = \alpha/m$ where m is the number of possible comparisons in the "family", then compare the p-value to α' instead of α.
  - Based on the "degree of fishing", choose one of these methods that gives adjusted p-values and/or adjusted CIs for some specific situation (generally more power than Bonferroni):
    - Tukey's procedure: test all possible pairs for one factor.
    - Dunnet's procedure: compare a control to all possible active treatments
    - Scheffé's procedure: all possible simple and complex contrasts

10

Contrasts in SPSS

- Contrasts" button gives p-values assuming that the comparisons are an appropriate set of ***planned*** comparisons.
- "Post-hoc" button gives p-values assuming post-hoc comparisons within the "family" associated with the specific post-hoc procedure. In multi-way ANOVA, a no-interaction model is assumed. Tukey (all paires) and Dunnett (baseline vs. all others) are the most useful.

11

Contrasts in SPSS, cont.

Multiple Comparisons

	(I) Color	(J) Color	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	white	red	-1.8000	.29804	.000	-2.6248	-.9752
		green	-2.4571	.29804	.000	-3.2819	-1.6324
		blue	.0952	.31021	.990	-.7632	.9537
	Red	white	1.8000	.29804	.000	.9752	2.6248
		green	-.6571	.29804	.152	-1.4819	.1676
		blue	1.8952	.31021	.000	1.0368	2.7537
	green	white	2.4571	.29804	.000	1.6324	3.2819
		red	.6571	.29804	.152	-.1676	1.4819
		blue	2.5524	.31021	.000	1.6939	3.4108
	Blue	white	-.0952	.31021	.990	-.9537	.7632
		red	-1.8952	.31021	.000	-2.7537	-1.0368
		green	-2.5524	.31021	.000	-3.4108	-1.6939

Homogeneous Subsets

	Color	N	Subset 1	Subset 2
Tukey HSD	blue	6	2.3333	
	white	7	2.4286	
	red	7		4.2286
	green	7		4.8857
	Sig.		.989	.164

Not necessarily non-overlapping!

12

Contrasts in SPSS, cont.

- SPSS LMATRIX subcommand
  - Requires syntax pasting and manual entry of contrast coefficients
  - Very flexible: ***any*** valid contrast can be specified
  - p-values are based on the comparisons being ***planned***
  - For post-hoc, calculate t=contr./SE(contr.), F=t^2, and use, e.g., Scheffé procedure

"Paste Syntax" in SPSS

- "Paste" instead of the final "OK" in SPSS causes the "syntax" to be displayed in the "Syntax Editor" instead of running the analysis.
- You can edit the syntax (following strict rules) and then "run" the syntax to run the analysis.
- Expert SPSS users often work mainly with "syntax".
- For us, syntax is used to add features to an analysis for which there are no menu items.

Contrasts in SPSS, cont.

- 1-way ANOVA example: factor name is "treatment", level order is Placebo, Talk, Drug, Both. Paste:

```
UNIANOVA score BY treatment
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=treatment.
```

- Edit to:

```
UNIANOVA score BY treatment
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /LMATRIX "others - control" treatment -1 1/3 1/3 1/3
  /LMATRIX "combo - (drug+talk)/2" treatment 0 1/2 1/2 -1
  /LMATRIX "drug-talk" treatment 0 -1 1 0
  /DESIGN=treatment.
```


LMATRIX for 1-way ANOVA, cont.

```
/LMATRIX "others - control" treatment -1 1/3 1/3 1/3
```

Estimate of $\frac{\mu_D+\mu_T+\mu_B}{3} - \mu_P$

Custom Hypothesis Tests #1

Contrast Results (K Matrix)[a]

Contrast			Dependent Variable score
L1	Contrast Estimate		16.333
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		16.333
	Std. Error		6.192
	Sig.		.017
	95% Confidence Interval for Difference	Lower Bound	3.324
		Upper Bound	29.343

a. Based on the user-specified contrast coefficients (L') matrix: others - placebo

LMATRIX for 1-way ANOVA, cont.

```
/LMATRIX "combo - (drug+talk)/2" treatment 0 1/2 1/2 -1
```

Estimate of $\mu_B - \frac{\mu_D+\mu_T}{2}$

Custom Hypothesis Tests #2

Contrast Results (K Matrix)[a]

Contrast			Dependent Variable score
L1	Contrast Estimate		20.400
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		20.400
	Std. Error		6.979
	Sig.		.009
	95% Confidence Interval for Difference	Lower Bound	5.738
		Upper Bound	35.062

a. Based on the user-specified contrast coefficients (L') matrix: both - (drug+talk)/2

LMATRIX for 1-way ANOVA, cont.

```
/LMATRIX "drug-talk" treatment 0 -1 1 0
```

Estimate of $\mu_D - \mu_T$

Custom Hypothesis Tests #3

Contrast Results (K Matrix)[a]

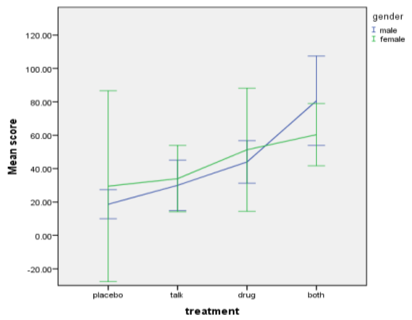
Contrast			Dependent Variable score
L1	Contrast Estimate		16.400
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		16.400
	Std. Error		7.825
	Sig.		.050
	95% Confidence Interval for Difference	Lower Bound	-.039
		Upper Bound	32.839

a. Based on the user-specified contrast coefficients (L') matrix: drug-talk

Contrasts for 1-way ANOVA without interaction

- Additive model = parallel pattern in a graph pf population means
- Valid questions: What are the effects of a specific change in level of one factor ignoring, fixing or averaging over the other factor?
- Conclusion: Analyze each factor separately as for 1-way ANOVA.

Contrasts for 2-way with Interaction



Error Bars: 95% CI

Interaction Contrasts, cont.

```
UNIANOVA score BY treatment gender
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(treatment*gender)
  /CRITERIA=ALPHA(.05)
  /LMATRIX "M-F for placebo" gender 1 -1
         treatment*gender 1 -1   0 0   0 0   0 0
  /LMATRIX "M-F for both" gender 1 -1
         treatment*gender 0 0   0 0   0 0   1 -1
  /LMATRIX "M-F for (talk+drug)/2" gender 1 -1
         treatment*gender 0 0   1/2 -1/2   1/2 -1/2  0 0
  /DESIGN=treatment gender treatment*gender.
```

21

Interaction Contrasts, cont.

```
/LMATRIX "M-F for both" gender 1 -1
 treatment*gender 0 0  0 0  0 0  1 -1
```

H_0: $\mu_{BM} = \mu_{BF}$ Estimate: $\mu_{BM} - \mu_{BF}$

Contrast Results (K Matrix)[a]

Contrast			Dependent Variable score
L1	Contrast Estimate		20.333
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		20.333
	Std. Error		7.088
	Sig.		.012
	95% Confidence Interval for Difference	Lower Bound	5.130
		Upper Bound	35.537

a. Based on the user-specified contrast coefficients (L') matrix: M-F for both

22

Interaction Contrasts, cont.

- Posthoc example: all pairs for the 2x4 ANOVA
- 8 groups = $\binom{8}{2}$ = 8*7/2 = 28 pairs
- Write up to 28 LMATRIX commands
- Compute Bonferroni α' = 0.05/28 = 0.0018
- Test (2-1)*(4-1)=3 planned contrasts using α=0.05
- Reject any others if p<0.0018

23

Summary

- Contrasts allow more useful scientific conclusions when a rejected H_0 is vague, e.g., H_0: $\mu_1=\ldots=\mu_k$ (with k>2) or H_0: additive model is good enough.
- (Remember: main effect H_0s in the presence of a significant interaction answer the wrong questions!)
- Running multiple tests increases the chance for false rejection. Beyond "df" tests, corrections must be used to "maintain the type-1 error rate at α".
- Multiple comparisons corrections reduce power. Pre-planned contrasts should be selected before running the experiment to maximize power to where it is most needed.

24