ANOVA Follow-Up Comparisons:

Planned Contrasts, Multiple Comparison Procedures, and Trend Analyses

Comparing Groups in an ANOVA

- If the ANOVA is significant, then it means that there is some difference, somewhere…but it does not tell you which means are different from each other
- Two basic approaches for comparing cell means
  - **Planned contrasts** are done when you have specific hypotheses to test
    - Compare specific pairs means
  - **Multiple comparison procedures** (post hoc tests) are done when you do not have specific hypotheses
    - Compare all possible pairs of means

Why Use Follow-Up Tests?

- The *F*-ratio tells us only that the experiment was successful
  - i.e., group means were different
- It does not tell us specifically which group means differ from which
- We need additional tests to find out where the group differences lie

How?

- Multiple *t*-tests
  - We saw earlier that this is a bad idea
- Planned Contrasts
  - Hypothesis driven
  - Planned a priori
- Multiple Comparison Procedures *(Post Hoc* Tests)
  - Not Planned (no hypothesis)
  - Compare all pairs of means
- Trend Analysis

Planned Contrasts

- Basic Idea:
  - The variability explained by the Model (experimental manipulation, SS_B) is due to participants being assigned to different groups
  - This variability can be broken down further to test specific hypotheses about which groups might differ
  - We break down the variance according to hypotheses made *a priori* (before the experiment)
    - Separating the variance is similar to the idea of cutting up a cake

Rules When Choosing Contrasts

- Independent
  - contrasts must not interfere with each other (i.e., they must test unique hypotheses)
- Only 2 Chunks
  - Each contrast should compare only 2 chunks of variation
- J-1
  - You should always end up with one less possible contrast than the number of groups

Generating Hypotheses

- Example: Testing the effects of Viagra on Libido using three groups:
  - Placebo (Sugar Pill)
  - Low Dose Viagra
  - High Dose Viagra
- Dependent Variable (DV) was an objective measure of Libido
- Intuitively, what might we expect to happen?

	Placebo	Low Dose	High Dose
	3	5	7
	2	2	4
	1	4	5
	1	2	3
	4	3	6
Mean	2.20	3.20	5.00

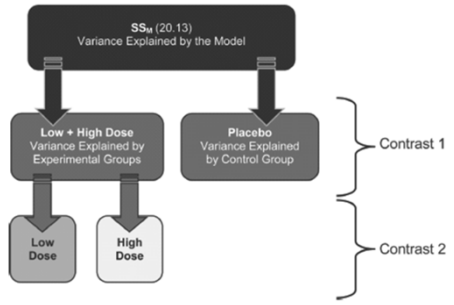
How do we choose contrasts?

- Big Hint:
  - In most experiments we usually have one or more control groups
  - The logic of control groups dictates that we expect them to be different than the groups that we have manipulated
  - The first contrast will almost always be to compare any control groups (chunk 1) with any experimental conditions (chunk 2)

Hypotheses

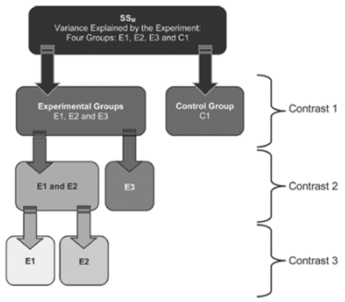
- Hypothesis 1:
  - People who take Viagra will have a higher libido than those who do not
- Hypothesis 2:
  - People taking a high dose of Viagra will have a greater libido than those taking a low dose of Viagra

Planned Comparisons

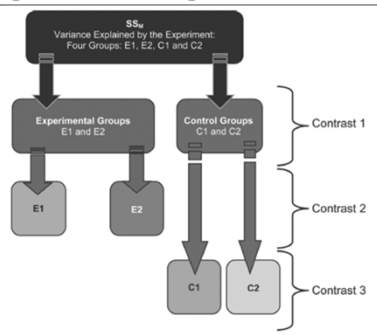


An Example of a 4 Group Design:

3 Experimental Groups and 1 Control Group

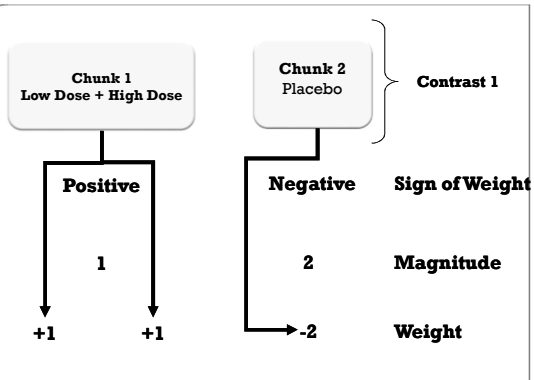


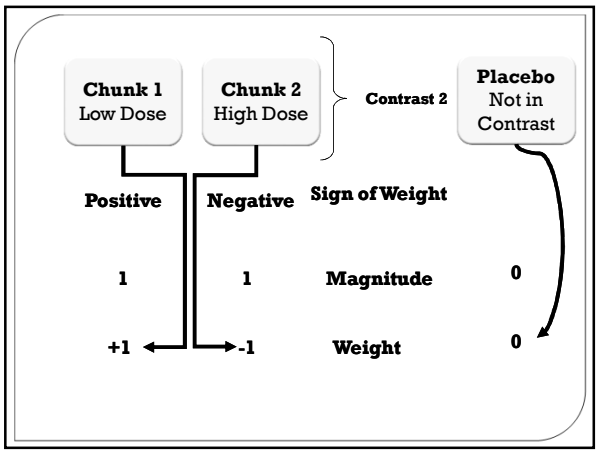
An Example of a 4 Group Design:
2 Experimental Groups and 2 Control Groups



Coding Planned Contrasts: Rules

- Rule 1
  - Groups coded with positive weights compared to groups coded with negative weights
- Rule 2
  - The sum of weights for a comparison should be zero
- Rule 3
  - If a group is not involved in a comparison, assign it a weight of zero
- Rule 4
  - For a given contrast, the weights assigned to the group(s) in one chunk of variation should be equal to the number of groups in the opposite chunk of variation
- Rule 5
  - If a group is singled out in a comparison, then that group should not be used in any subsequent contrasts





Output

Contrast Coefficients

	Dose of Viagra		
Contrast	Placebo	1 Dose	2 Doses
1	-2	1	1
2	0	-1	1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Libido	Assume equal variances	1	3.80	1.536	2.474	12	.029
		2	1.80	.887	2.029	12	.065
	Does not assume equal variances	1	3.80	1.483	2.562	8.740	.031
		2	1.80	.917	1.964	7.720	.086

Multiple Comparison Procedures: Introduction

- If the ANOVA F rejects H_o, it is favoring H_1…but H_1 merely says "any difference in the μ_j's"
  - So the F does not tell you which groups have different means, it says "some difference, somewhere"
  - As a result, F is usually not the only statistic that we need to understand a one-way design with more than two groups
  - F is an "omnibus test"
- We need tests for the multiple differences that exist between the J means
  - For example, which of the groups has the highest libido: High Dose group, Low Dose group, or Placebo group?
  - The significant F test merely says there is some difference somewhere

Multiple Comparison Procedures: Introduction

- Multiple comparisons are the many mean differences the exist when you compare J means
- Pairwise comparisons are differences in means taken two at a time
  - For J means, there are $C = \frac{J*(J-1)}{2}$ pairwise comparisons
- The hypotheses for pairwise comparisons are
  - H_o:$\mu_j=\mu_{j'}$
  - H_1:$\mu_j\neq\mu_{j'}$

Multiple Comparison Procedures: Introduction

- For the liar data, J = 3, so $C = \frac{J*(J-1)}{2} = \frac{3*2}{2} = 3$
  - There are three pairwise comparisons:
    - High Dose vs. Low Dose
    - High Dose vs. Placebo
    - Low Dose vs. Placebo
  - Error rate per comparison sets α'=.05 for each comparison, so the probability of a Type I error is about .15

Multiple Comparison Procedures: Error Rates

- Error rates:
  - **Error rate per comparison** sets α'=.05 for each comparison, so for 3 comparisons, α' would approach .15 (rather than .05)
    - It would be less than .15 because there is some overlap in the comparisons that are being made…but it would still be well above .05
  - **Error rate family-wise** controls Type I error by taking into account the number of comparisons being made in a single analysis
    - Essentially, $\frac{\alpha}{\text{number of comparisons}}$ is used for each comparison

Multiple Comparison Procedures: Error Rates

- 3 groups leads to 3 comparisons: $\frac{.05}{3} = .017$
- 4 groups leads to 6 comparisons: $\frac{.05}{6} = .008$
- 5 groups leads to 10 comparisons: $\frac{.05}{10} = .005$
- 6 groups leads to 15 comparisons: $\frac{.05}{15} = .003$
- 7 groups leads to 21 comparisons: $\frac{.05}{21} = .002$

Multiple Comparison Procedures: Tukey HSD (Honestly Significant Difference)

- We will use a t statistic for multiple comparisons. The F does not need to be significant. This t will take into account the number of comparisons when finding the degrees of freedom.

	Tukey's MCP
1. Situation/hypotheses	$J \geq 2$ independent samples $H_o: \mu_j = \mu_{j'}$ All pairwise comparisons Equal n's
2. Test statistic	$t = \frac{\bar{X}_j - \bar{X}_{j'}}{\sqrt{\frac{MS_w}{n}(2)}}$
3 .Distribution	$\frac{q_{J,df,pw}}{\sqrt{2}}$ Studentized Range
4. Assumptions	1. Populations are normal 2. $\sigma^2_j = \sigma^2_{j'}$ 3. Observations are independent

Steps for the Tukey HSD

- The omnibus F-test does not have to be significant in order for the Tukey to control Type I error
- Steps
  - Obtain all possible differences between pairs of group means
  - Compute the t-statistics for all possible differences
  - Compare the absolute values of the t-statistics to the critical value
  - Reject the null hypothesis for any absolute value of t that equals or exceeds the critical value

Tukey HSD Example

- $\overline{X}_{\text{High Dose}} = 5.00$
- $\overline{X}_{\text{Low Dose}} = 3.20$
- $\overline{X}_{\text{Placebo}} = 2.20$

- $t = \frac{\overline{X}_i - \overline{X}_{i'}}{\sqrt{\frac{MS_w}{n}(2)}}$
- High Dose vs. Low Dose?
- High Dose vs. Placebo?
- Low Dose vs. Placebo?

Tukey Output

Multiple Comparisons

Dependent Variable: libido
Tukey HSD

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Placebo	Low Dose	-1.00000	.88694	.516	-3.3662	1.3662
	High Dose	-2.80000*	.88694		-5.1662	-.4338
Low Dose	Placebo	1.00000	.88694		-1.3662	3.3662
	High Dose	-1.80000	.88694	.147	-4.1662	.5662
High Dose	Placebo	2.80000*	.88694	.021	.4338	5.1662
	Low Dose	1.80000	.88694	.147	-.5662	4.1662

*. The mean difference is significant at the 0.05 level.

SPSS does not provide the t-test value for the Tukey but it can be obtained by dividing the "mean difference" by the "standard error":

$t = \frac{\text{Mean Difference}}{\text{Standard Error}} = \frac{-2.80}{0.89} = -3.15$

libido

Tukey HSD[a]

group	N	Subset for alpha = 0.05 1	2
Placebo	5	2.2000	
Low Dose	5	3.2000	3.2000
High Dose	5		5.0000
Sig.		.516	.147

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

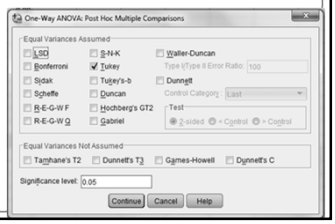
This table provides an easy way to see which group means are significantly different from each other

MCPs That Are Also Commonly Used

- **Ryan, Einor, Gabriel, and Welsh Q (REGWQ)**: Similar to Tukey in terms of Type I error control but has better power
- **Bonferroni**: Similar to Tukey
  - Has slightly more power than Tukey with small number of comparisons but less power than Tukey with large number of comparisons
  - The basic Bonferroni correction can be applied to any set of analyses (i.e., divide alpha by number of analyses)
- **Fisher-Hayter**: Similar to Tukey but is less conservative (i.e., has greater power)
  - This test is not available in SPSS

General Strategy for MCPs

- There are a lot of MCPs offered by SPSS (as well as other MCPs that it does not offer)
- If you have equal sample sizes and equal variances, then use Tukey's HSD or REGWQ
- If sample sizes are unequal, then use Gabriel or Hochberg's GT2
- If variances are unequal, then use the Games-Howell



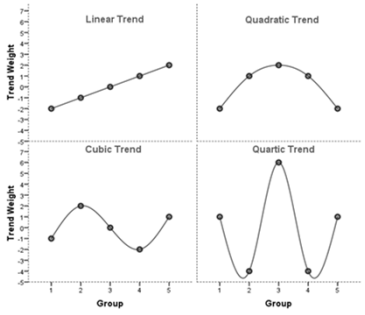
MCPs That Are Not Generally Recommended

- **Fisher LSD**: It is used quite often but it is not great because it ignores the multiple comparison issue (inflates Type I error)
- **Duncan**: Type I error rate tends to be considerably higher than it should be
- **Newman-Keuls**: This test is commonly used but it can have family-wise error rates that are greater than the researcher intended

Additional Reading About Multiple Comparison Procedures

- Toothaker, L. E. (1993). *Multiple comparison procedures*. Newbury Park, CA: Sage.

Trend Analysis



Trend Analysis: Output

ANOVA

Libido

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		20.133	2	10.067	5.119	.025
	Linear Term	Contrast	19.600	1	19.600	9.966	.008
		Deviation	.533	1	.533	.271	.612
	Quadratic Term	Contrast	.533	1	.533	.271	.612
Within Groups			23.600	12	1.967		
Total			43.733	14			