One-way between-subjects ANOVA

Comparing three or more independent means

ANOVA: A Framework

- Understand the basic principles of ANOVA
  - Why it is done?
  - What it tells us?

- Theory of one-way between-subjects ANOVA

- Following up an ANOVA:
  - Planned Contrasts/Comparisons
    - Choosing Contrasts
    - Coding Contrasts
  - Post Hoc Tests

- Writing up Results

Why ANOVA?

- *t* tests are limited to situations in which there are only two levels of a single independent variable or two associated groups.

- There are many instances in which we'd like to compare more than two levels. But … performing multiple *t* tests can inflate the Type I error rate.

One-Way ANOVA

- The one-way analysis of variance is used to test the null hypothesis that three or more population means are equal

  - more precisely: test the null hypothesis that the means of the groups are not significantly different from the grand mean of all participants

One-Way ANOVA

- The *response* variable is the variable you're comparing, i.e., dependent variable
- The *factor* variable is the categorical variable being used to define the groups, i.e., independent variable
  - Usually called *k* samples (groups)
- The *one-way* is because there is one independent variable

Assumptions

- Scale dependent variable

- Normal distribution

- Equal variances

The basic principle behind ANOVA

Produces a test statistic termed the *F-ratio*

$$\frac{\text{Systematic Variance}}{\text{Unsystematic Variance}}$$

If the model explains a lot more variability than it can't explain, then the experimental manipulation has had a significant effect on the outcome (DV).

Partitioning the Variance

Libido vs Participant

Grand Mean = 3.467

$X_{Placebo} = 2.20$

$X_{LowDose} = 3.20$

$X_{HighDose} = 5.00$

b_1

b_2

Basis for SS_M

Basis for SS_R

Basis for SS_T

Placebo
Low Dose
High Dose
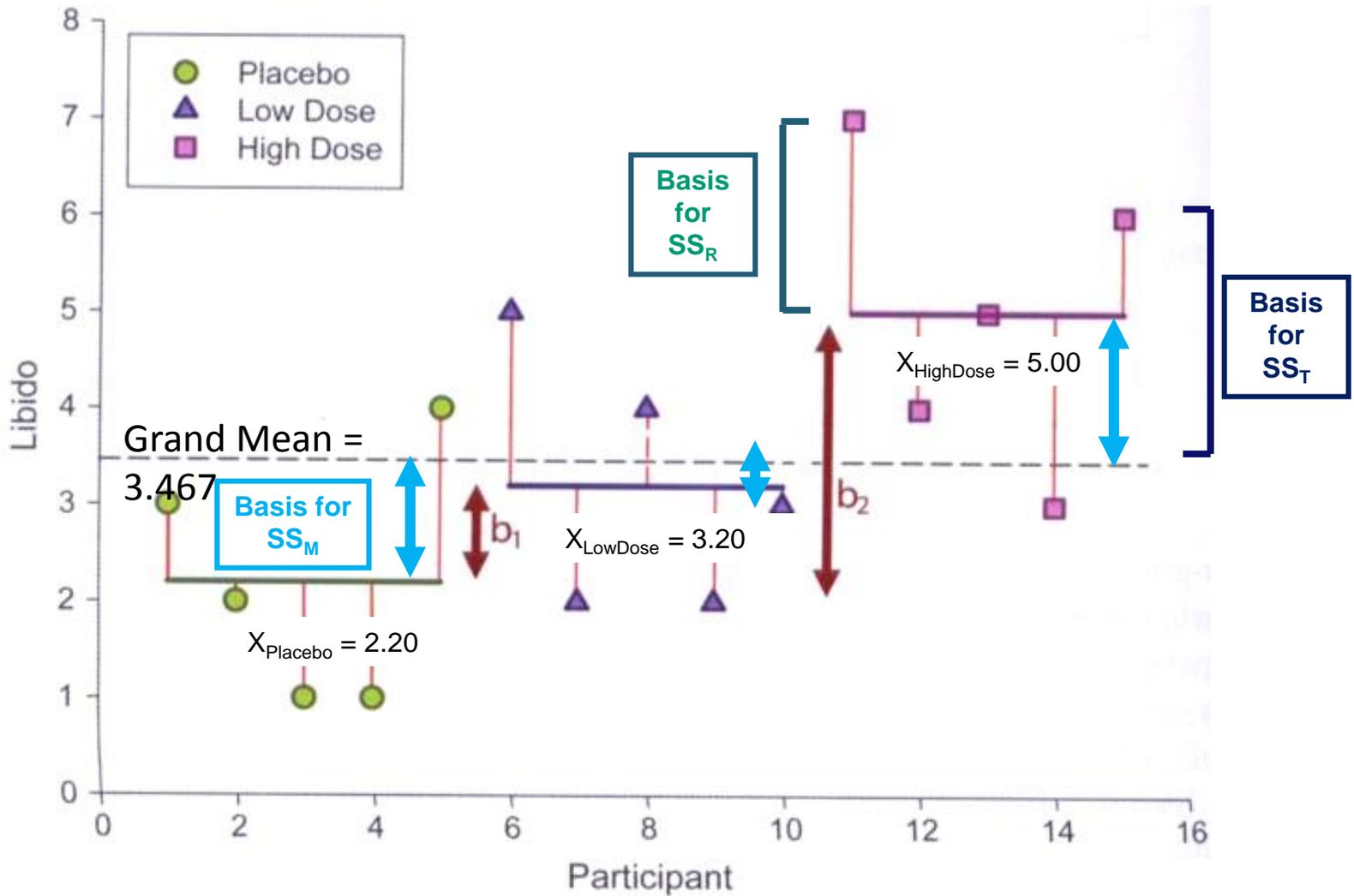
SS_M = differences between predicted values (group means) and the grand mean

SS_R = based on differences between person's score and their group mean.

SS_T = based on differences between each data point and the grand mean

Partitioning the variance

- SS_M = differences between predicted values (group means) and the grand mean (i.e., variation across groups) –
  if there is a significant difference between the groups, this should be a large number relative to SS_R

- SS_R = based on differences between person's score and their group mean (i.e., variation with each group)

- SS_T = based on differences between each data point and the grand mean (the total variation in the entire data set)

Mean Squares (MS$_M$ and MS$_R$)

SS_M = amount of variation explained by the model (exp. manipulation).

SS_R = amount of variation due to extraneous factors.

These are "summed" scores and will therefore be influenced by the number of scores. To eliminate this bias we calculate the average sum of squares (mean squares) by dividing by the appropriate degrees of freedom.

Calculating Degrees of Freedom
(for one-way independent groups ANOVA)

df_{total} $=$ N - 1 (number of all scores minus 1)

$df_{M \, / \, between}$ $=$ k - 1 (number of groups minus 1)

$df_{R \, / \, within}$ $=$ N - k (number of all scores minus number of groups)

The F-ratio

- We compare the amount of variability explained by the Model (MS$_M$), to the error in the model [individual differences] (MS$_R$)

  - This ratio is called the *F*-ratio $F = \dfrac{MS_M}{MS_R}$

- If the model explains a lot more variability than it can't explain, then the experimental manipulation has had a significant effect on the outcome (DV).

One-Way ANOVA

- Here is the basic one-way ANOVA table

Source	SS	df	MS	F	p
Between					
Within					
Total					

An example: Fairness in different types of societies

Fairness score: proportion of money shared in a game

	Hunter-gatherer	Farming	Natural resources	Industrial
P1	28	32	47	40
P2	36	33	43	47
P3	38	40	52	45
P4	31			
Mean	**33.25**	**35.0**	**47.33**	**44.0**
N	*4*	*3*	*3*	*3*

Grand Mean = 39.385 (The sum of all scores divided by the total N

Total SS

The sum of the squared deviation of each score from the grand total

$$SS_T = \sum (x_i - M_{grand})^2$$

Error (within-group) sum of squares (SS$_R$)

- Calculate the squared deviation of each score from its group mean
- Add the results

$$SS_R = \sum (x_i - M_i)^2$$

Model (between-group) sum of squares (SS$_M$)

1. Calculate the difference between the mean of each group and the grand mean.

    The grand mean is the mean of all scores

2. Square each of these differences

3. Multiply each result by the number of participants within that group – this is a correction (or "weighting"): a smaller sample will have less "weight" in the equation, a larger sample will have more "weight".

4. Add the values for each group together.

$$SS_M = \sum n_i (M_i - M_{grand})^2$$

One-Way ANOVA

- After filling in the sum of squares, we have …

Source	SS	df	MS	F	p
Between	461.64				
Within	167.42				
Total	629.08				

Degrees of freedom

- The between group df is one less than the number of groups
  - We have four groups, so $df_M = 3$
- The within group df is the sum of the individual df's of each group
  - The sample sizes are 4, 3, 3, and 3
  - $df_R = 3 + 2 + 2 + 2 = 9$
- The total df is one less than the sample size
  - $df(Total) = 13 - 1 = 12$ $= (n+n+n) -1$

One-Way ANOVA

- Filling in the degrees of freedom gives this …

Source	SS	df	MS	F	p
Between	461.64	3			
Within	167.42	9			
Total	629.08	12			

Calculating the Mean Squares

Divide the SS by the corresponding df

- MS_M = 461.64 / 3 = 153.88
- MS_R = 167.42 / 9 = 18.60

One-Way ANOVA

- Completing the MS gives …

Source	SS	df	MS	F	p
Between	461.64	3	153.88		
Within	167.42	9	18.60		
Total	629.08	12			

The F ratio

- **F test statistic**
  - An F test statistic is the ratio of MS_M and MS_R
  - $F = MS_M / MS_R$
- **For our data, F = 153.88 / 18.60 = 8.27**

- **A larger F ratio means a larger difference between the group means relative to the variation within the group.**

One-Way ANOVA

- Adding F to the table …

Source	SS	df	MS	F	p
Between	461.64	3	153.88	8.27	
Within	167.42	9	18.60		
Total	629.08	12			

Hypothesis testing

- The F test statistic has an F distribution with df_M numerator df and df_R denominator df

- $P(F_{3,9} > 5.9) = 0.001$

One-Way ANOVA

- Completing the table with the p-value

Source	SS	df	MS	F	p
Between	461.64	3	153.88	8.27	0.001
Within	167.42	9	18.60		
Total	629.08	12			

Computing Effect Size

(η^2 - eta squared and ω^2 - omega squared)

Eta-squared reflects the proportion of DV variance explained by each IV in the sample data.

$$h^2 = \frac{SS_M}{SS_T}$$

What does this remind you of from regression analysis?

An alternative, omega-squared (ω^2), is more complicated but also more accurate.

$$w^2 = \frac{SS_M - (df_M)(MS_R)}{SS_T + MS_R}$$

Computing Effect Size
(η^2 eta squared vs. ω^2 omega squared)

Conventions:
Small effect: .01
Medium effect: .06
Large effect: .14

$$h^2 = \frac{SS_M}{SS_T} = \frac{153.88}{629.08} = .24$$

$$w^2 = \frac{SS_M - (df_M)(MS_R)}{SS_T + MS_R} = \frac{153.88 - 3*18.60}{629.08 + 18.60} = .15$$

Post-hoc tests

The F-ratio tells us only that the experimental manipulation has had an effect—not where the effect has occurred.

-- Planned comparisons (before you collect the data)

-- Post-hoc tests (after-the-fact "snooping").

-- Both conducted to control the overall Type I error rate at 5%.

Multiple comparisons: Post-hoc tests

- A significant F value only shows that at least one condition is significantly different from at least one other condition
  - but it does not tell us where the difference is (or differences are)
- Planned comparisons (contrasts): choose the pairs that we expect to differ significantly
  1. hunter-gatherer vs. natural resources
  2. farming vs. industrial
- Post-hoc tests compare all pairs. They are like t-tests with the difference that we make corrections for multiple comparisons
  - three common tests: LSD (does not correct), Bonferroni (very conservative), Tukey
  - When variances are not equal: Games-Howell, Dunnet's C

Computing the One-way Independent-groups ANOVA using SPSS

SpiderBG.sav

- Arachnophobia experiment. Participants are shown either
  - A real ladybird
  - A picture of a spider
  - Or a real spider
- Anxiety is measured (higher score = more anxiety)

Run arachnophobia analysis

- Note layout of data file (independent groups)
- Check labels, variable types
- Run descriptives
- Run ANOVA
  - Analyze -> Compare Means -> One-Way ANOVA
  - Enter dependent and independent variables
  - Options -> Homogeneity of Variance test, Welch (gives a corrected F value to be used if the variances are not equal)
  - Post Hoc -> choose at least one from Equal variances assumed and one from Equal variances not assumed
  - Planned comparisons:

Planned Contrasts (if we have specific hypotheses)
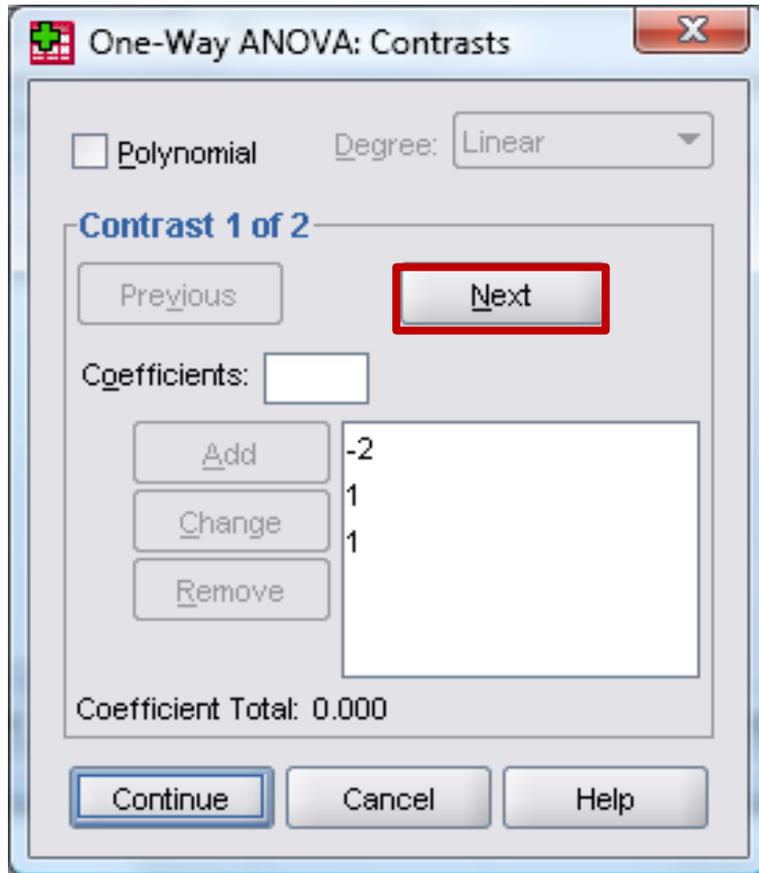
The first contrast compares the combined "real creature" groups against the control (picture) group.
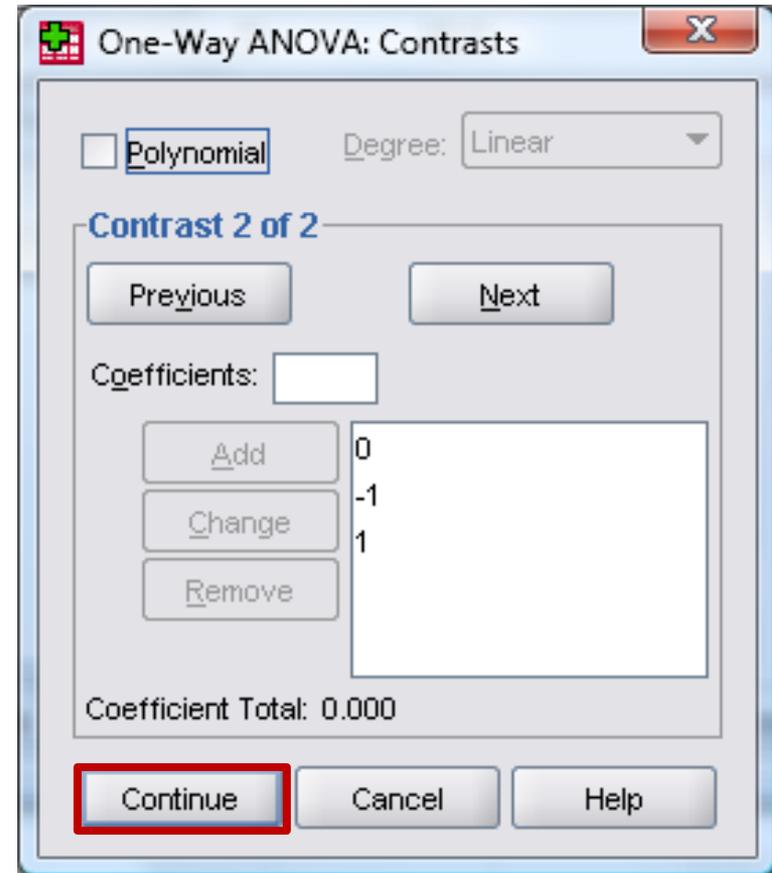
The second contrast compares the two "real creature" groups, ladybird vs. spider.

Reporting the Results

Participants were exposed to a real ladybird, a picture of a spider or a real spider (N = 12 in each group). The anxiety of the three groups was measured. The mean level of anxiety was highest for the Real Spider group (M = 47.00, SD = 11.03), followed by the Picture Spider group (M = 40.00, SD = 9.29) and the Real Ladybird group showed the lowest level of anxiety (M = 28.58, SD = 3.63). A One-Way ANOVA with Welch's correction revealed that this difference was statistically significant ($F(2, 17.77) = 19.99$, $p < .001$, $\eta^2 = .46$).

If you used planned comparisons …

Planned comparisons revealed no significant difference in anxiety between the picture group and the two real animal groups, i.e., the mode of presentation did not seem to have an effect. There was, however a significant difference between the two real animal: participants in the Real Spider group showed significantly greater anxiety than participants in the Real Ladybird group, $t(13.35) = -5.49$, $p < .001$.

If you used post-hoc tests …

Games-Howell post hoc tests revealed a significant difference in anxiety between the Real Ladybird and Real Spider groups ($p < .0051$) and between the Real Ladybird and the Picture Spider groups ($p = .004$). No other comparisons were significant.

Exercise: Attractiveness

Langlois & Roggman (1990) computer-averaged photographs of faces and had their participants to rate the composite faces for attractiveness.

Independent variable: number of faces computer-averaged for the composite picture: 2, 4, 8, 16, 32

Attractiveness rating on a 5-point scale (5 = very attractive)

- Descriptives
- One-Way ANOVA
- Effect size
- Trend

Homework: Sources of self-esteem

- The self-esteem of 1500 participants was measured. The respondents were classified according to birth order: 1st born, middle born, last born. Is there a difference between the self-esteem index of the three groups? (Higher score: higher self-esteem)

- Run descriptives (remember that with a large sample size, the distribution is likely to be normal whatever the test of normality says)

- Run the One-Way ANOVA

- Interpret planned comparisons or post-hoc tests if ANOVA is significant.

- Write up the results (don't forget to include a graph with means)