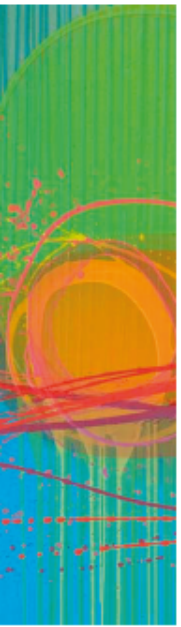
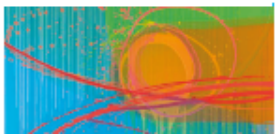
DISCOVERING STATISTICS USING SPSS THIRD EDITION



Comparing several means: ANOVA (GLM 1)

Lecture 09

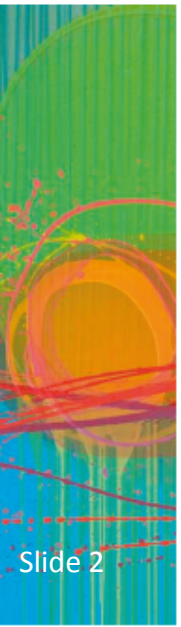
ANDY FIELD

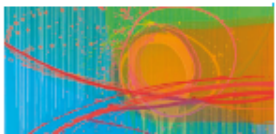


Aims

- Understand the basic principles of ANOVA
  - Why it is done?
  - What it tells us?
- Theory of one-way independent ANOVA

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



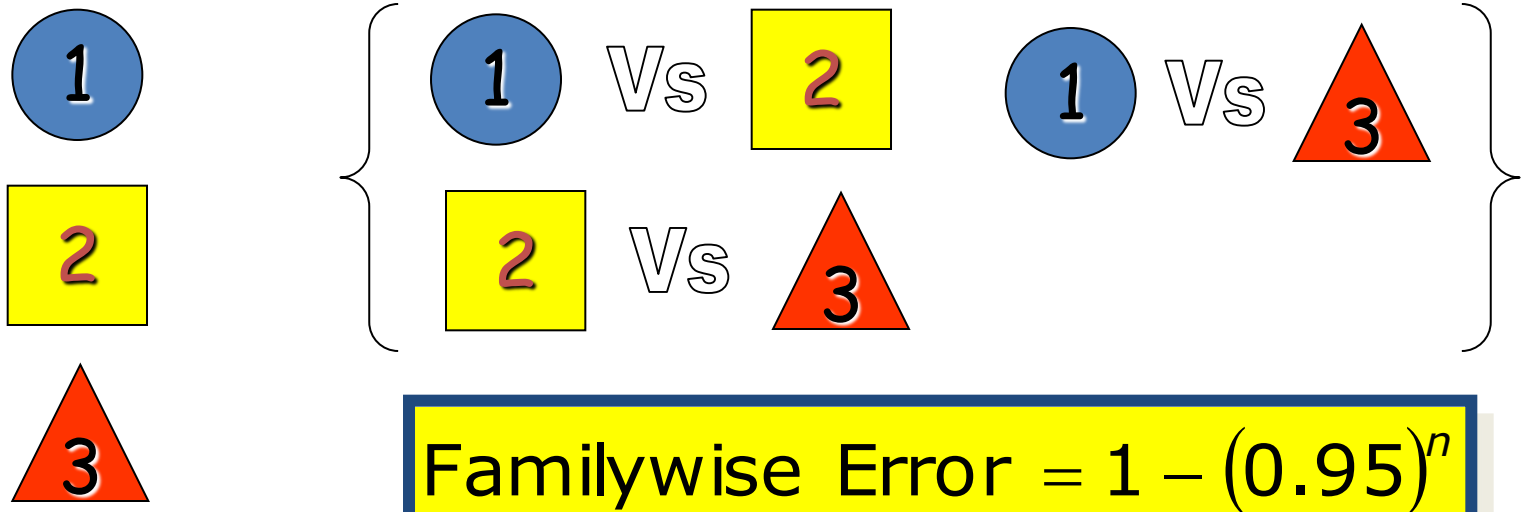


When And Why

- When we want to compare means we can use a *t*-test. This test has limitations:
  - You can compare only 2 means: often we would like to compare means from 3 or more groups.
  - It can be used only with one Predictor/Independent Variable.
- ANOVA
  - Compares several means.
  - Can be used when you have manipulated more than one Independent Variables.
  - It is an extension of regression (the General Linear Model)

Why Not Use Lots of *t*-Tests?

- If we want to compare several means why don't we compare pairs of means with *t*-tests?
  - Can't look at several independent variables.
  - Inflates the Type I error rate.



$$\text{Familywise Error} = 1 - (0.95)^n$$

What Does ANOVA Tell us?

- Null Hyothesis:
  - Like a t-test, ANOVA tests the null hypothesis that the means are the same.
- Experimental Hypothesis:
  - The means differ.
- ANOVA is an Omnibus test
  - It test for an overall difference between groups.
  - It tells us that the group means are different.
  - It doesn't tell us exactly which means differ.

ANOVA as Regression

$$\text{outcome}_i = (\text{model}) + \text{error}_i$$

$$\text{Libido}_i = b_0 + b_2\text{High}_i + b_1\text{Low}_i + \varepsilon_i$$

TABLE 10.2 Dummy coding for the three-group experimental design

Group	*Dummy Variable 1 (High)*	*Dummy Variable 2 (Low)*
Placebo	0	0
Low-Dose Viagra	0	1
High-Dose Viagra	1	0

Placebo Group

$$\text{Libido}_i = b_0 + b_2\text{High}_i + b_1\text{Low}_i + \varepsilon_i$$

$$\text{Libido}_i = b_0 + (b_2 \times 0) + (b_1 \times 0)$$

$$\text{Libido}_i = b_0$$

$$\overline{X}_{\text{Placebo}} = b_0$$

High Dose Group

$$\text{Libido}_i = b_0 + b_2\text{High}_i + b_1\text{Low}_i + \varepsilon_i$$

$$\text{Libido}_i = b_0 + (b_2 \times 1) + (b_1 \times 0)$$

$$\text{Libido}_i = b_0 + b_2$$

$$\text{Libido}_i = b_0 + b_2$$

$$\overline{X}_{\text{High}} = \overline{X}_{\text{Placebo}} + b_2$$

$$b_2 = \overline{X}_{\text{High}} - \overline{X}_{\text{Placebo}}$$

Low Dose Group

$$\text{Libido}_i = b_0 + b_2\text{High}_i + b_1\text{Low}_i + \varepsilon_i$$

$$\text{Libido}_i = b_0 + (b_2 \times 0) + (b_1 \times 1)$$

$$\text{Libido}_i = b_0 + b_1$$

$$\text{Libido}_i = b_0 + b_1$$

$$\overline{X}_{\text{Low}} = \overline{X}_{\text{Placebo}} + b_1$$

$$b_1 = \overline{X}_{\text{Low}} - \overline{X}_{\text{Placebo}}$$

Output from Regression

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.133	2	10.067	5.119	.025[a]
	Residual	23.600	12	1.967		
	Total	43.733	14			

a. Predictors: (Constant), Dummy Variable 2, Dummy Variable 1

b. Dependent Variable: Libido

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.200	.627		3.508	.004
	Dummy Variable 1	2.800	.887	.773	3.157	.008
	Dummy Variable 2	1.000	.887	.276	1.127	.282

a. Dependent Variable: Libido

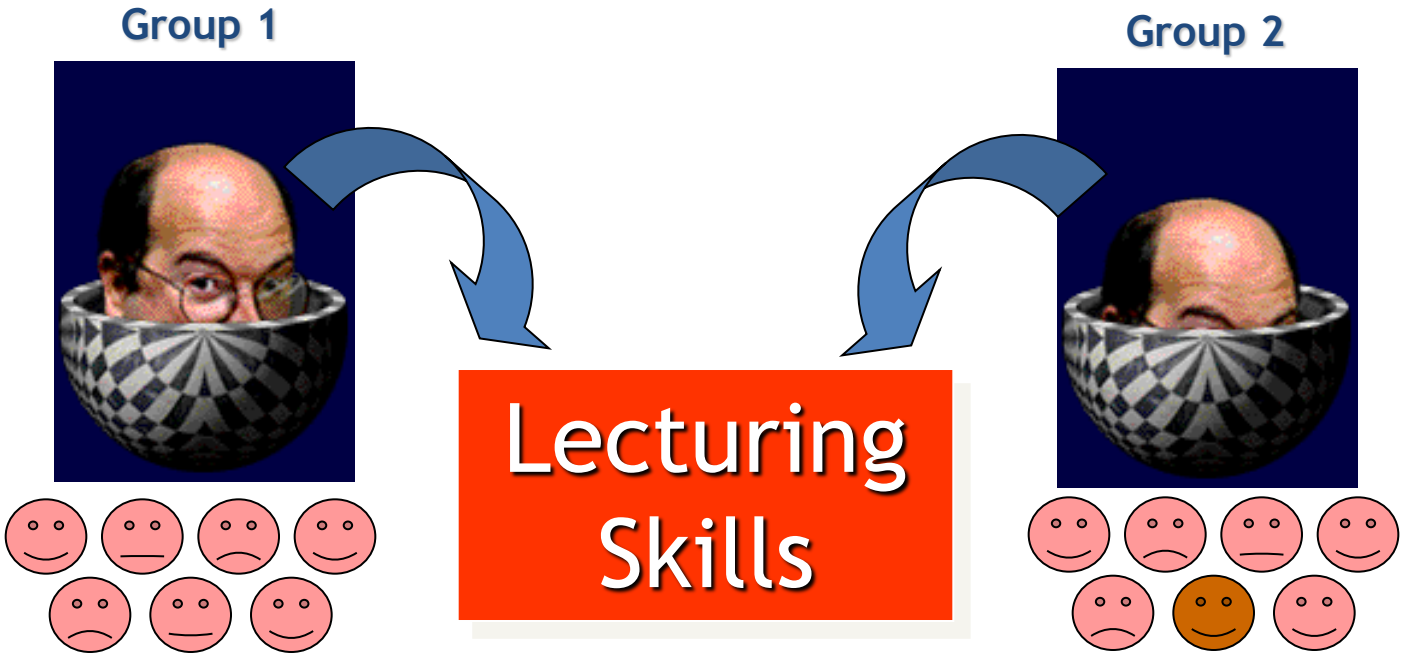
Experiments vs. Correlation

- ANOVA in Regression:
  - Used to assess whether the regression model is good at predicting an outcome.
- ANOVA in Experiments:
  - Used to see whether experimental manipulations lead to differences in performance on an outcome (DV).
    - By manipulating a predictor variable can we cause (and therefore predict) a change in behaviour?
  - Asking the same question, but in experiments we systematically manipulate the predictor, in regression we don't.

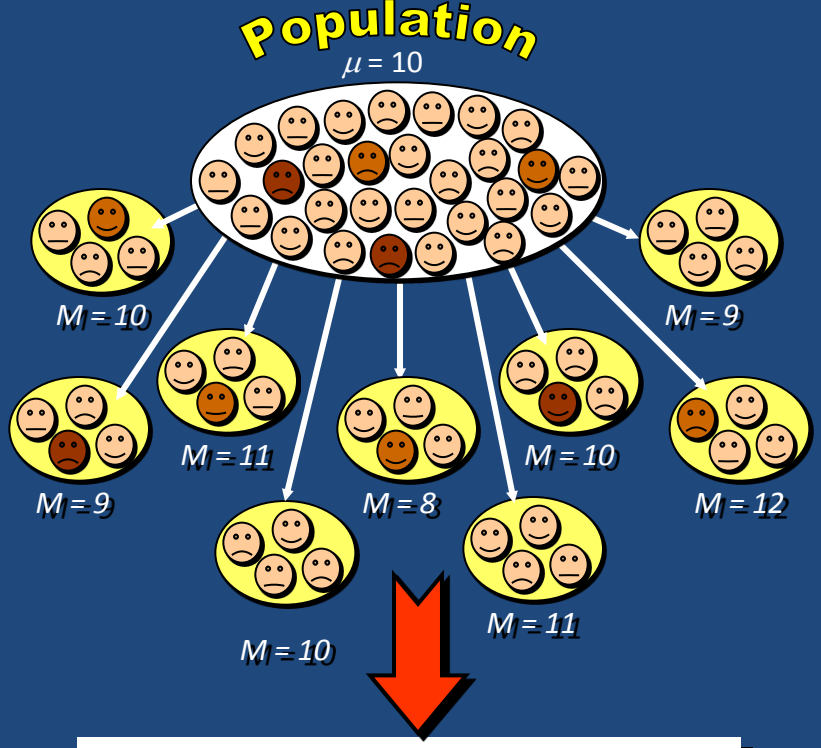
Theory of ANOVA

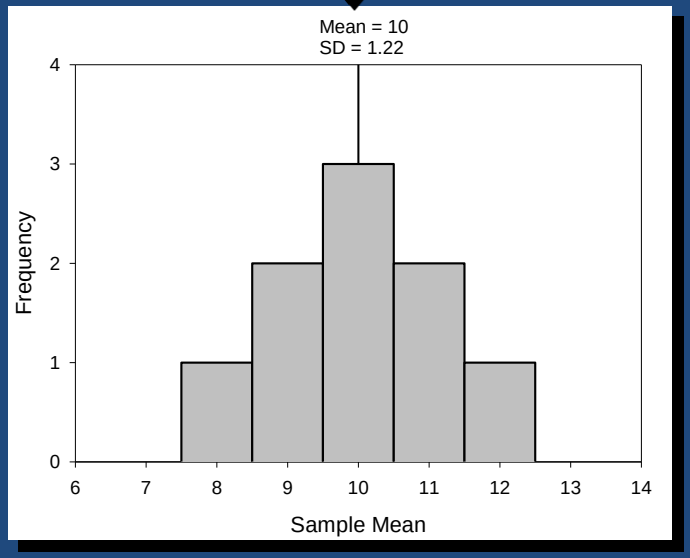
- We calculate how much variability there is between scores
  - Total Sum of squares (SS_T).
- We then calculate how much of this variability can be explained by the model we fit to the data
  - How much variability is due to the experimental manipulation, Model Sum of Squares (SS_M)...
- … and how much cannot be explained
  - How much variability is due to individual differences in performance, Residual Sum of Squares (SS_R).

Rationale to Experiments

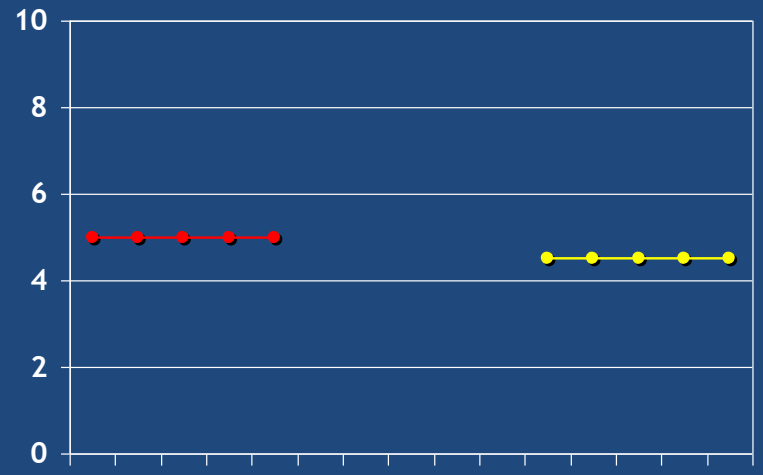


- Variance created by our manipulation
  - Removal of brain (systematic variance)
- Variance created by unknown factors
  - E.g. Differences in ability (unsystematic variance)

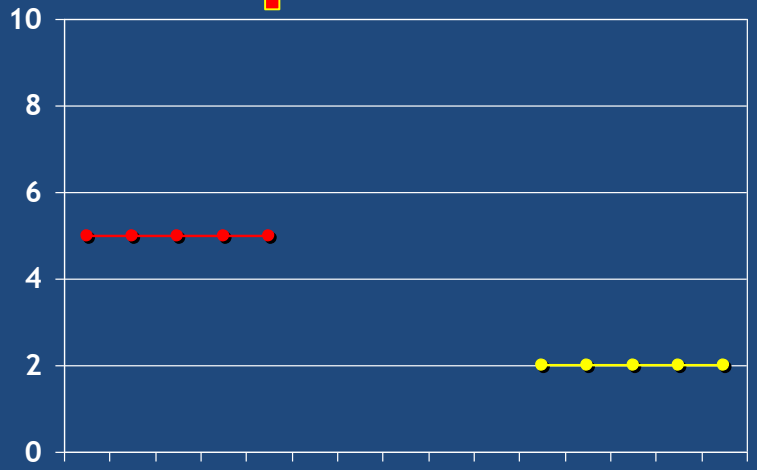




No Experiment



Experiment



Theory of ANOVA

- We compare the amount of variability explained by the Model (experiment), to the error in the model (individual differences)
  - This ratio is called the *F*-ratio.
- 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).

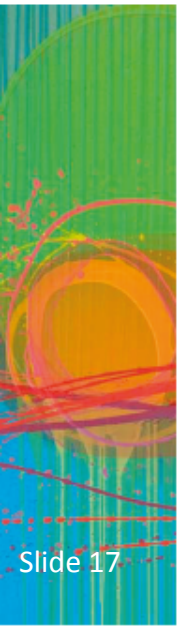
Theory of ANOVA

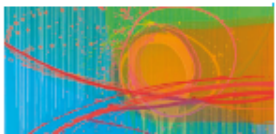
FIGURE 10.3 Partitioning variance for ANOVA

- If the experiment is successful, then the model will explain more variance than it can't
  - SS_M will be greater than SS_R

ANOVA by Hand

- Testing the effects of Viagra on Libido using three groups:
  - Placebo (Sugar Pill)
  - Low Dose Viagra
  - High Dose Viagra
- The Outcome/Dependent Variable (DV) was an objective measure of Libido.



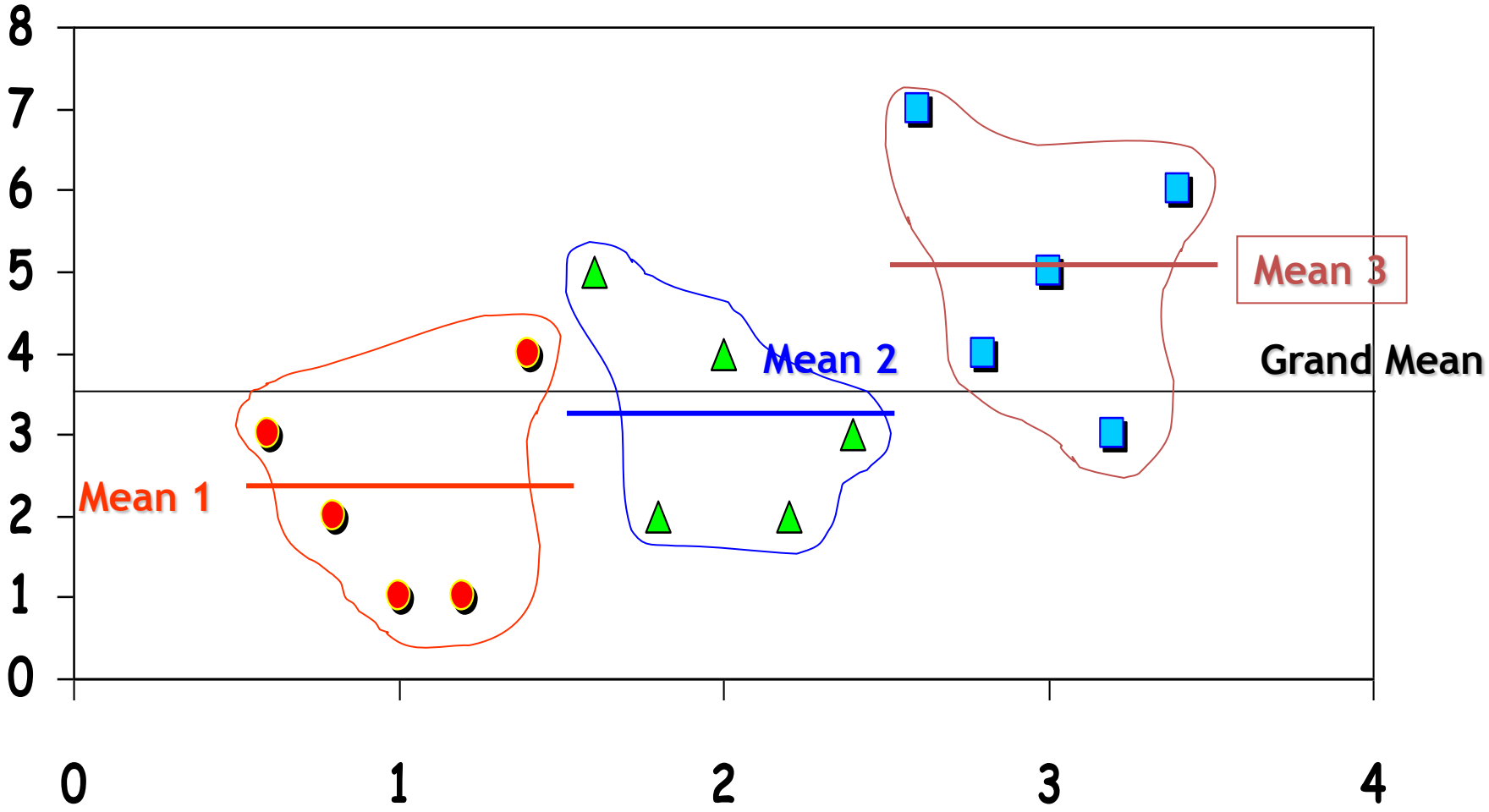


The Data

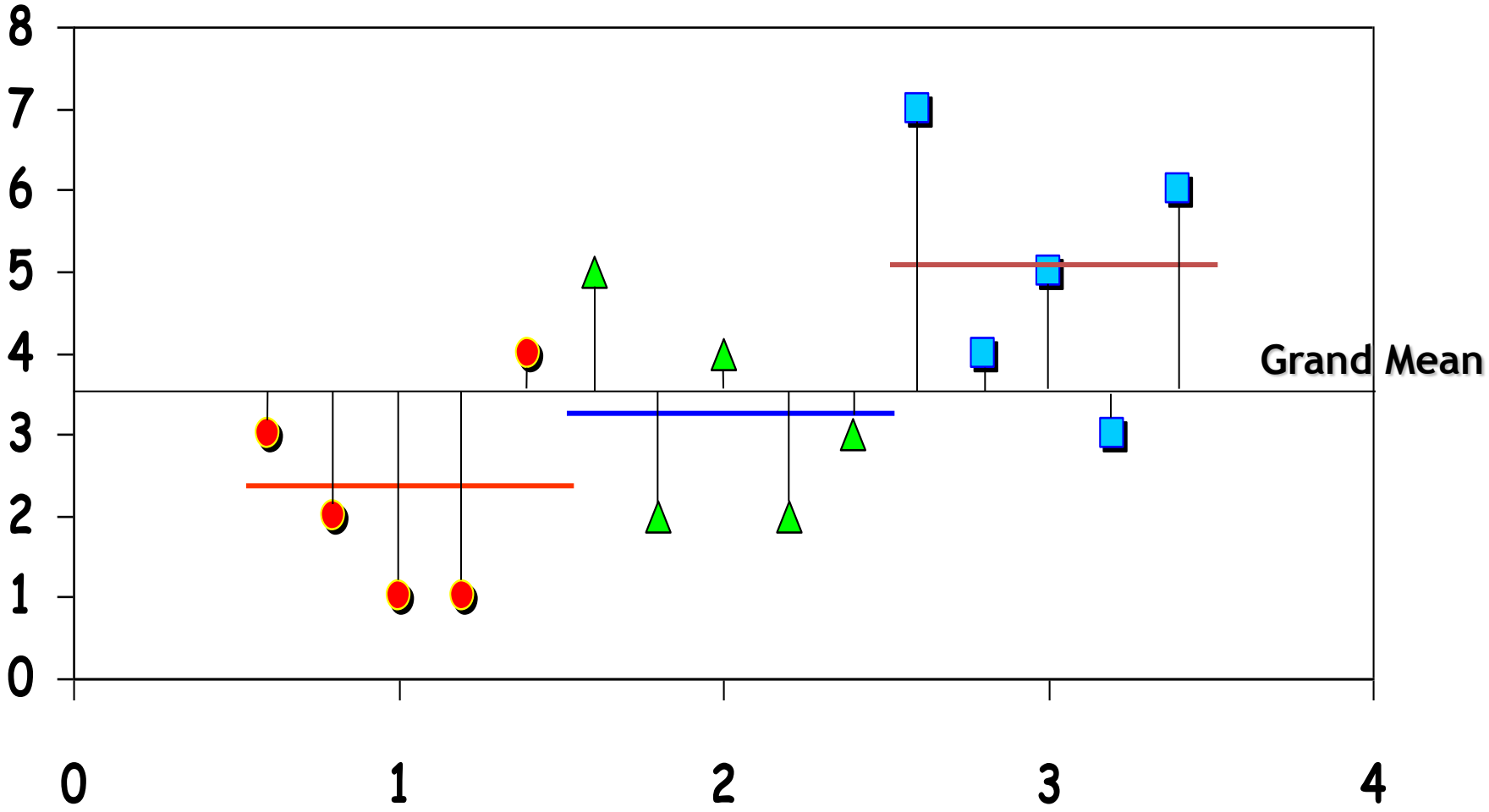
TABLE 10.1 Data in **Viagra.sav**

	Placebo	Low Dose	High Dose
	3	5	7
	2	2	4
	1	4	5
	1	2	3
	4	3	6
$\bar{X}$	2.20	3.20	5.00
s	1.30	1.30	1.58
s^2	1.70	1.70	2.50
	Grand Mean = 3.467	Grand SD = 1.767	Grand Variance = 3.124

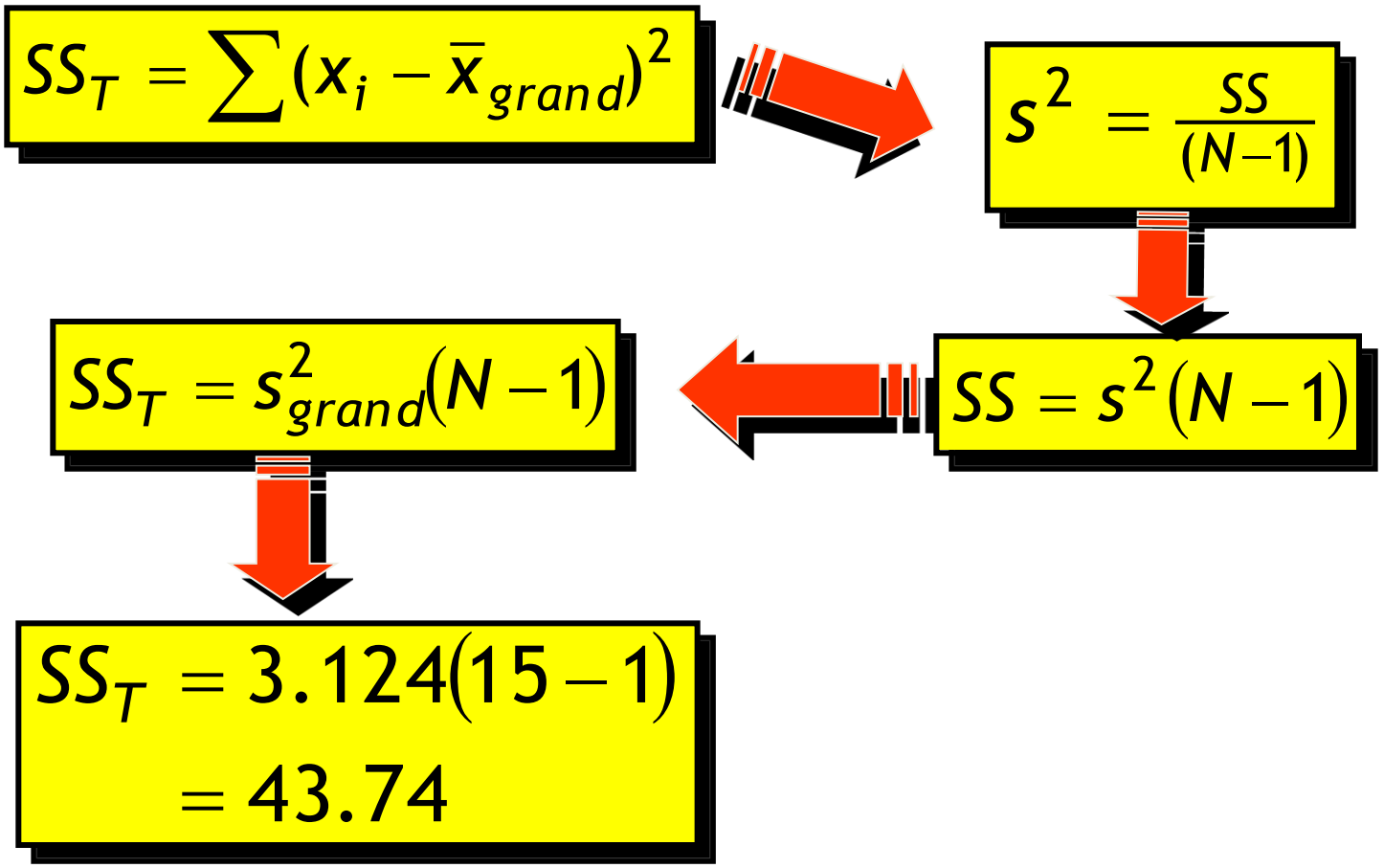
The data:



Total Sum of Squares (SS_T):



Step 1: Calculate SS_T

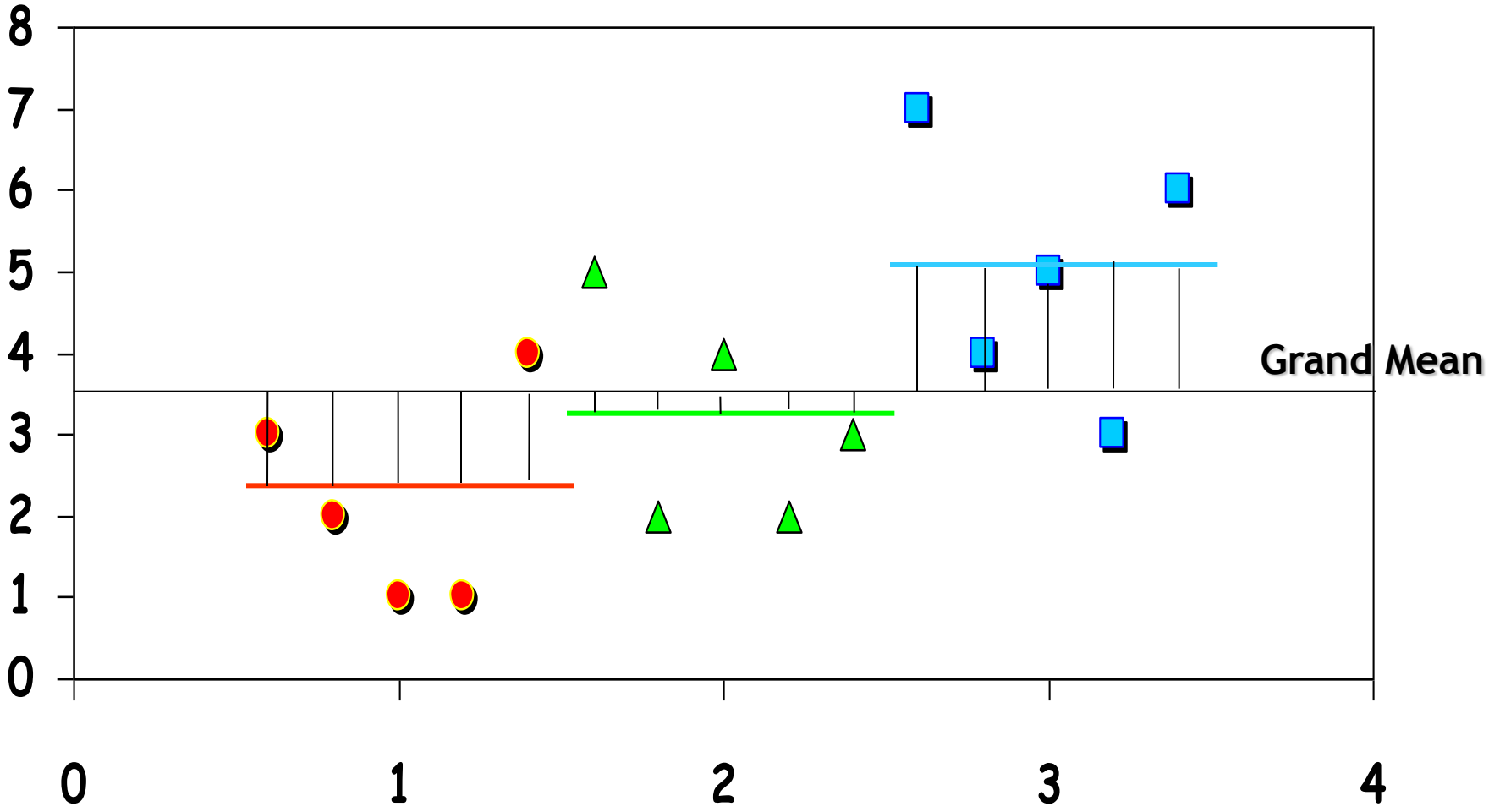


Degrees of Freedom (df)

- Degrees of Freedom (*df*) are the number of values that are free to vary.
  - Think about Rugby Teams!
- In general, the *df* are one less than the number of values used to calculate the SS.

$$df_T = (N - 1) = 15 - 1 = 14$$

Model Sum of Squares (SS_M):



Step 2: Calculate SS_M

$$SS_M = \sum n_i(\bar{x}_i - \bar{x}_{grand})^2$$

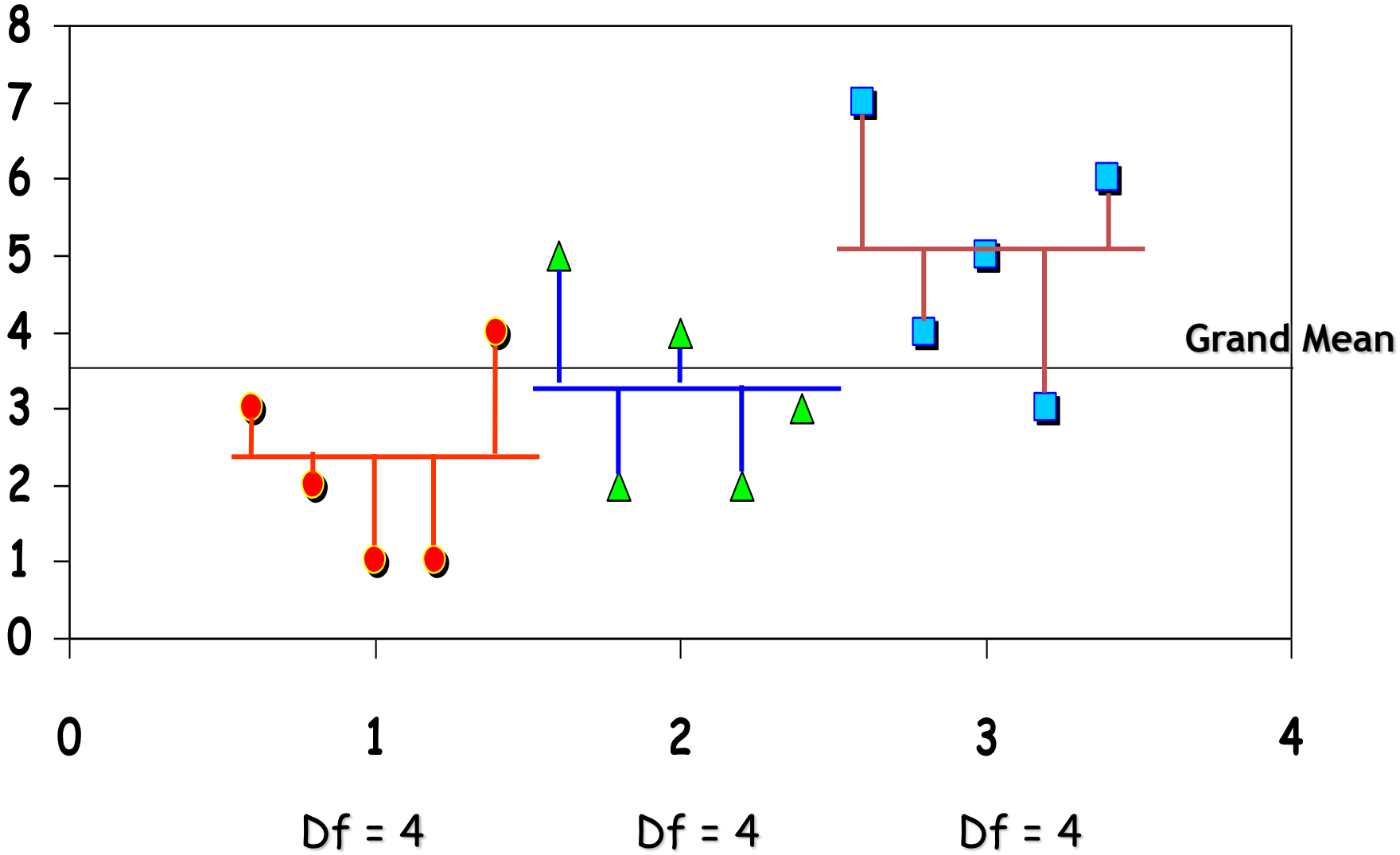
$$\begin{aligned} SS_M &= 5(2.2 - 3.467)^2 + 5(3.2 - 3.467)^2 + 5(5.0 - 3.467)^2 \\ &= 5(-1.267)^2 + 5(-0.267)^2 + 5(1.533)^2 \\ &= 8.025 + 0.355 + 11.755 \\ &= 20.135 \end{aligned}$$

Model Degrees of Freedom

- How many values did we use to calculate SS_M?
  - We used the 3 means.

$$df_M = (k - 1) = 3 - 1 = 2$$

Residual Sum of Squares (SS_R):



Step 3: Calculate SS_R

$$SS_R = \sum (x_i - \bar{x}_i)^2$$

$$s^2 = \frac{SS}{(N-1)}$$

$$SS = s^2(N-1)$$

$$SS_R = \sum s_i^2(n_i - 1)$$

$$SS_R = s^2_{group1}(n_1 - 1) + s^2_{group2}(n_2 - 1) + s^2_{group3}(n_3 - 1)$$

Step 3: Calculate SS_R

$$
\begin{aligned}
SS_R &= s^2_{group1}(n_1 - 1) + s^2_{group2}(n_2 - 1) + s^2_{group3}(n_3 - 1) \\
&= (1.70)(5 - 1) + (1.70)(5 - 1) + (2.50)(5 - 1) \\
&= (1.70 \times 4) + (1.70 \times 4) + (2.50 \times 4) \\
&= 6.8 + 6.8 + 10 \\
&= 23.60
\end{aligned}
$$

Residual Degrees of Freedom

- How many values did we use to calculate SS_R?
  - We used the 5 scores for each of the SS for each group.

$$\begin{aligned} df_R &= df_{group1} + df_{group2} + df_{group3} \\ &= (n_1 - 1) + (n_2 - 1) + (n_3 - 1) \\ &= (5 - 1) + (5 - 1) + (5 - 1) \\ &= 12 \end{aligned}$$

Double Check

$$SS_T = SS_M + SS_R$$
$$43.74 = 20.14 + 23.60$$
$$43.74 = 43.74$$

$$df_T = df_M + df_R$$
$$14 = 2 + 12$$
$$14 = 14$$

Step 4: Calculate the Mean Squared Error

$$MS_M = \frac{SS_M}{df_M} = \frac{20.135}{2} = 10.067$$

$$MS_R = \frac{SS_R}{df_R} = \frac{23.60}{12} = 1.967$$

Step 5: Calculate the *F*-Ratio

$$F = \frac{MS_M}{MS_R}$$

$$F = \frac{MS_M}{MS_R} = \frac{10.067}{1.967} = 5.12$$

Step 6: Construct a Summary Table

Source	SS	df	MS	*F*
Model	20.14	2	10.067	5.12*
Residual	23.60	12	1.967	
Total	43.74	14		

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
- Orthogonal Contrasts/Comparisons
  - Hypothesis driven
  - Planned a priori
- *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_M) 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).
  - It's like cutting up a cake (yum yum!)

Rules When Choosing Contrasts

- Independent
  - contrasts must not interfere with each other (they must test unique hypotheses).
- Only 2 Chunks
  - Each contrast should compare only 2 chunks of variation (why?).
- *K*-1
  - You should always end up with one less 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 I Choose Contrasts?

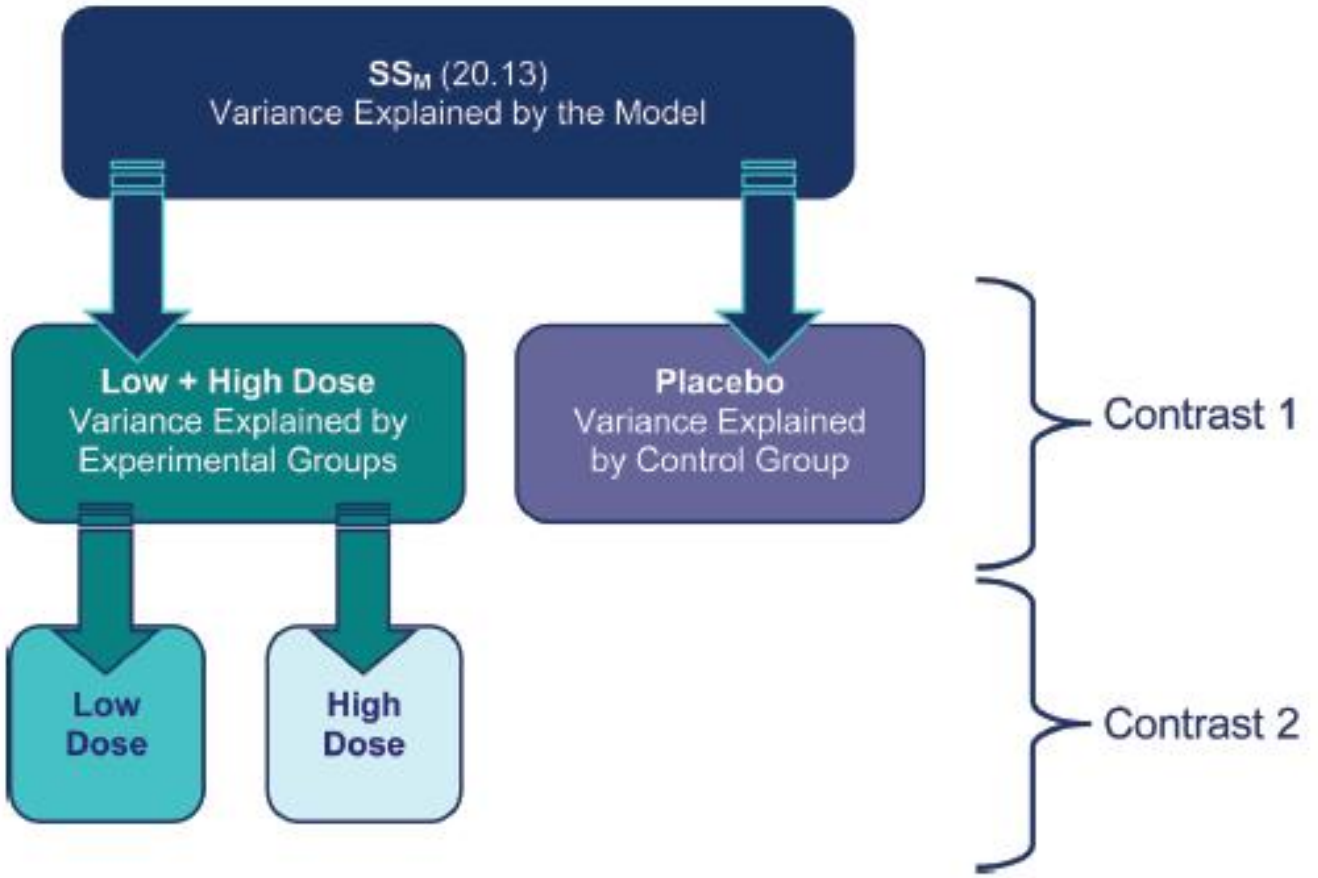
- Big Hint:
  - In most experiments we usually have one or more control groups.
  - The logic of control groups dictates that we expect them to be different to groups that we've manipulated.
  - The first contrast will always be to compare any control groups (chunk 1) with any experimental conditions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

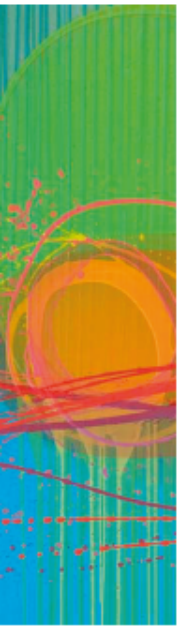
Hypotheses

- Hypothesis 1:
  - People who take Viagra will have a higher libido than those who don't.
  - Placebo ≠ (Low, High)
- Hypothesis 2:
  - People taking a high dose of Viagra will have a greater libido than those taking a low dose.
  - Low ≠ High

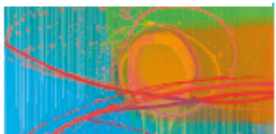
Planned Comparisons

FIGURE 10.4
Partitioning of experimental variance into component comparisons



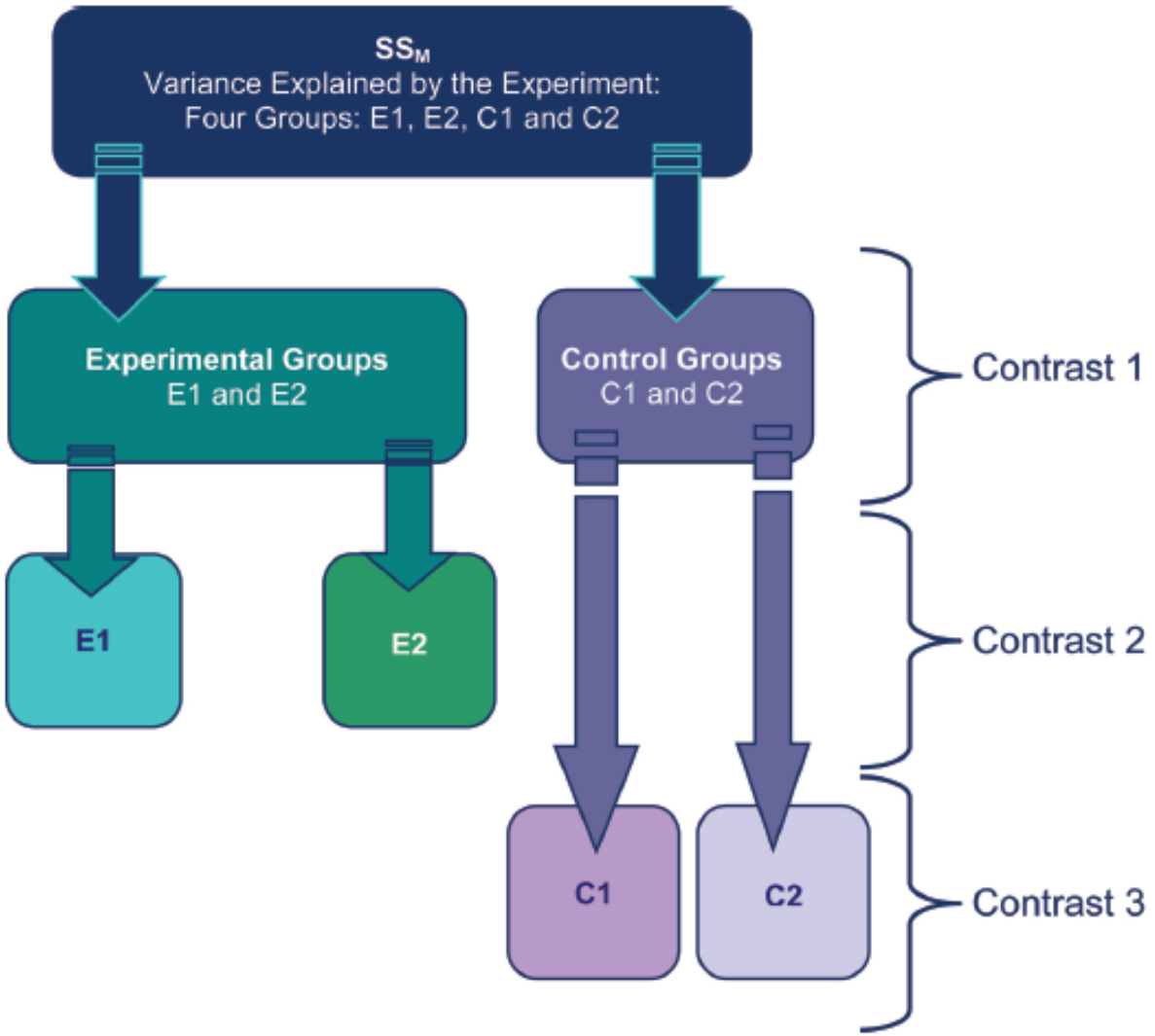


Another Example



Another Example

FIGURE 10.6
Partitioning variance for planned comparisons in a four-group experiment using two control groups

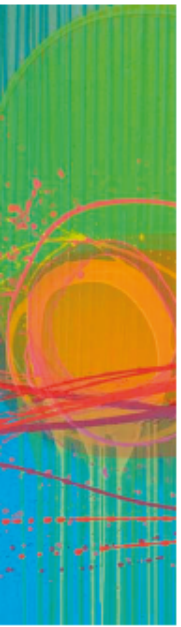


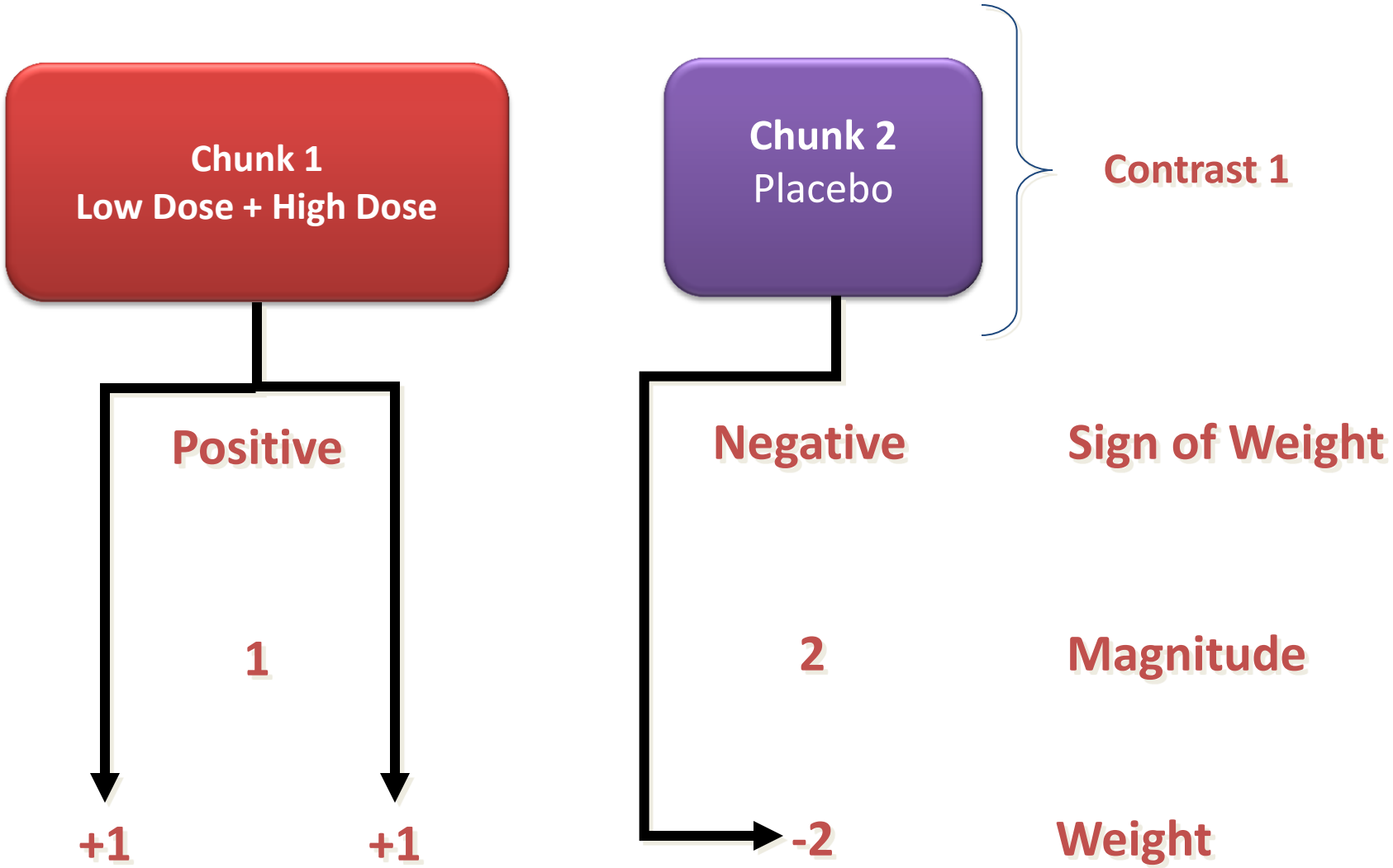
Coding Planned Contrasts: Rules

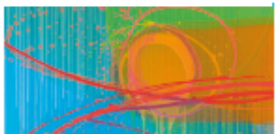
- Rule 1
  - Groups coded with positive weights compared to groups coded with negative weights.
- Rule 2
  - The sum of weights for a comparison should be zero.
- Rule 3
  - If a group is not involved in a comparison, assign it a weight of zero.

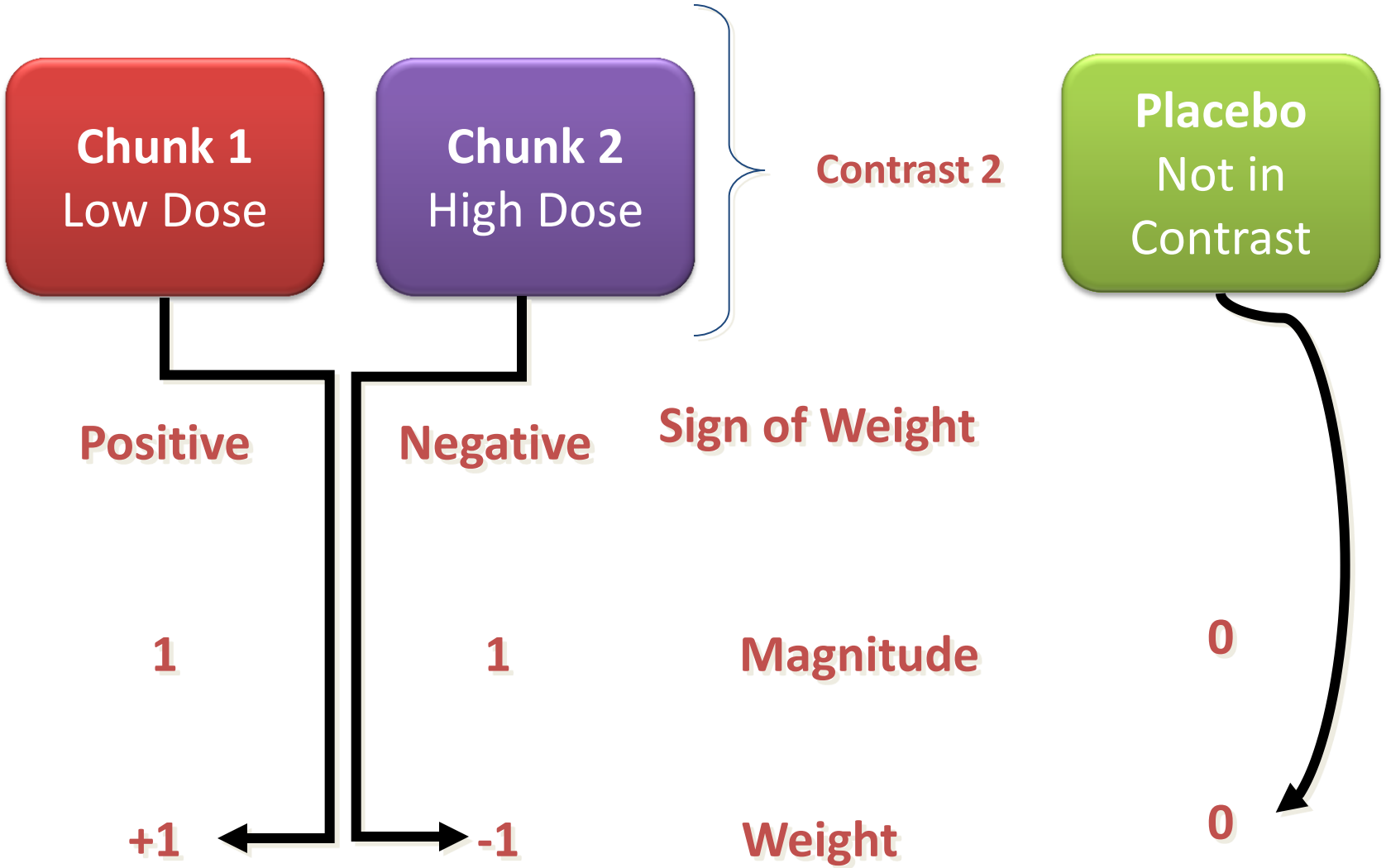
Coding Planned Contrasts: Rules

- 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.




Chunk 1
Low Dose + High Dose
Chunk 2
Placebo
Contrast 1
Positive
Negative
Sign of Weight
1
2
Magnitude
+1
+1
-2
Weight




Chunk 1
Low Dose
Chunk 2
High Dose
Contrast 2
Placebo
Not in
Contrast
Positive
Negative
Sign of Weight
1
1
Magnitude
0
+1
-1
Weight
0

Output

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

Post Hoc Tests

- Compare each mean against all others.
- In general terms they use a stricter criterion to accept an effect as significant.
  - Hence, control the familywise error rate.
  - Simplest example is the Bonferroni method:

$$\text{Bonferroni } \alpha = \frac{\alpha}{\text{Number of Tests}}$$

Post Hoc Tests Recommendations:

- SPSS has 18 types of Post hoc Test!
- Field (2009):
  - Assumptions met:
    - REGWQ or Tukey HSD.
  - Safe Option:
    - Bonferroni.
  - Unequal Sample Sizes:
    - Gabriel's (small *n*), Hochberg's GT2 (large *n*).
  - Unequal Variances:
    - Games-Howell.

Post Hoc Test Output

Multiple Comparisons

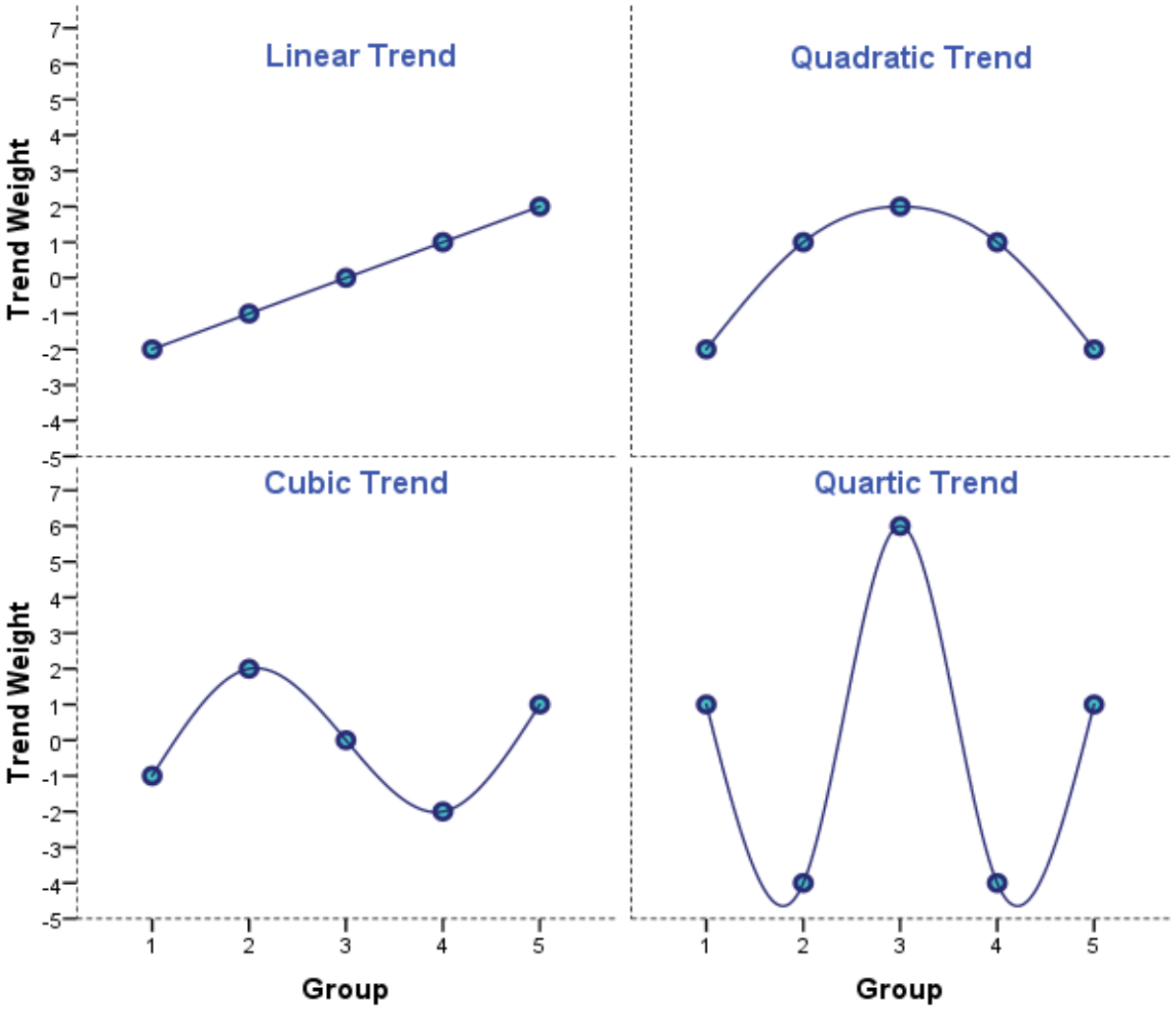
Dependent Variable:Libido

	(I) Dose of Viagra	(J) Dose of Viagra	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Placebo	Low Dose	-1.000	.887	.516	-3.37	1.37
		High Dose	-2.800*	.887	.021	-5.17	-.43
	Low Dose	Placebo	1.000	.887	.516	-1.37	3.37
		High Dose	-1.800	.887	.147	-4.17	.57
	High Dose	Placebo	2.800*	.887	.021	.43	5.17
		Low Dose	1.800	.887	.147	-.57	4.17
Games-Howell	Placebo	Low Dose	-1.000	.825	.479	-3.36	1.36
		High Dose	-2.800*	.917	.039	-5.44	-.16
	Low Dose	Placebo	1.000	.825	.479	-1.36	3.36
		High Dose	-1.800	.917	.185	-4.44	.84
	High Dose	Placebo	2.800*	.917	.039	.16	5.44
		Low Dose	1.800	.917	.185	-.84	4.44
Dunnett t (>control)[a]	Low Dose	Placebo	1.000	.887	.227	-.87	
	High Dose	Placebo	2.800*	.887	.008	.93	

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

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

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			