PS908
Research Methods & Statistics in Psychology

Analysis of Variance (3)
Within-Subjects ANOVA

Maxwell J Roberts
Department of Psychology
University of Essex
www.tubemapcentral.com

version date: 07/03/2022

Topic 3: Within-Subjects ANOVA

- Sources of variability and error
- *Individual-Differences* and *Residual* variance
- How within-subjects Analysis of Variance works
- Single-factor within-subjects Analysis of Variance
- Interpreting the ANOVA table
- Follow-up tests for three or more levels

Sources of Variability and Error

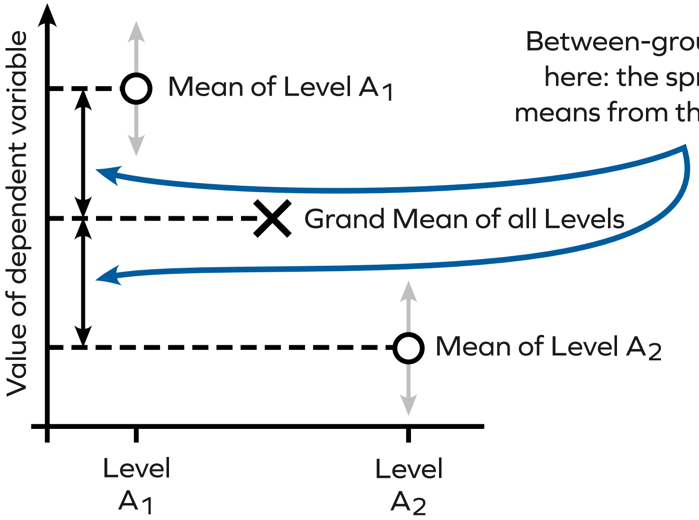
- **Between-Ss ANOVA**

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

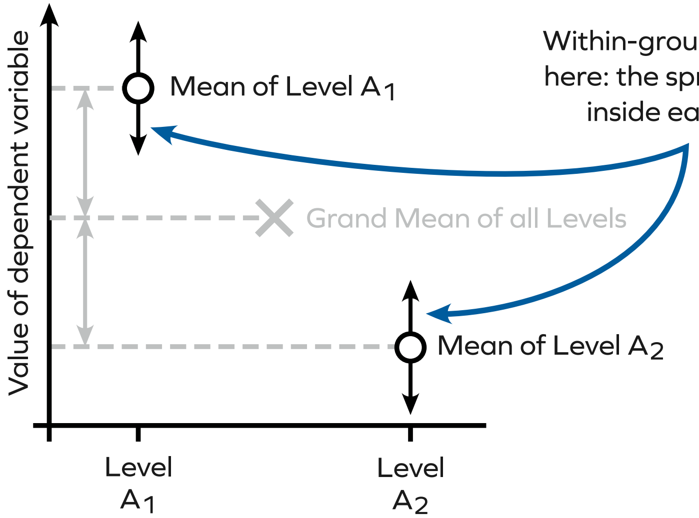
  - Between-group variance
    - Measurement of **effect**: differences between level means
  - Within-group variance
    - Measurement of **experimental error**: the extent to which people differ even though they were treated alike

Sources of Variability and Error

- **Between-Ss ANOVA**



Between-group variance is here: the spread of level means from the grand mean



Within-group variance is here: the spread of data inside each group

Sources of Variability and Error

- **Between-Ss ANOVA**
  - One measurement per person
  - Identification of causes of variability in data is limited

$$F = \frac{\begin{array}{c}\text{between-group}\\ \text{variance}\end{array} \left\{ \begin{array}{c} \text{treatment effects?} \\ \text{individual differences?} \\ \text{random errors?} \end{array} \right.}{\begin{array}{c}\text{within-group}\\ \text{variance}\end{array} \left\{ \begin{array}{c} \text{individual differences?} \\ \text{random errors?} \end{array} \right.}$$

  - Cannot distinguish between variability caused by individual differences versus random errors
  - Error term must combine both, is relatively large

Sources of Variability and Error

- **Within-Ss ANOVA**
  - Two (or more) measurements per person
  - Can now identify the variability that is uniquely caused by individual differences

	Visually Complex IQ Test Items	Visually Simple IQ Test Items
S_1	8	10
S_2	6	8
S_3	7	7
S_4	5	9
S_5	4	6
Level Means	6	8

DV: Number of items correct

Sources of Variability and Error

- **Within-Ss ANOVA**
  - An individual's level of performance is given by the ***cross-task subject mean*** across experimental levels
  - *Variability* caused by *individual differences* = variability in ***cross-task subject means*** from person to person

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8	

DV: Number of items correct

Sources of Variability and Error

- **Within-Ss ANOVA: measurement of effect**
  - Calculated as before, *between-group variance*: the extent to which the level means differ but …
  - Two (or more) measurements per person gives a purer measure of the effect

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8	

DV: Number of items correct

Sources of Variability and Error

- **Within-Ss ANOVA: measurement of effect**
  - Purer measure because individual differences variability, as defined has:
    - ***No effect*** on differences in group means
    - ***No effect*** on between-group variance

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**	
S_1	8	10	**9**	↑ *Pulling **BOTH** level means up*
S_2	6	8	**7**	} *Neutral effect on level means*
S_3	7	7	**7**	
S_4	5	9	**7**	
S_5	4	6	**5**	↓ *Pulling **BOTH** level means down*
Level Means	6	8		

DV: Number of items correct

Sources of Variability and Error

- **Within-Ss ANOVA: measurement of effect**
  - Two (or more) measurements per person
  - Gives a purer measure of effect (between-group variance) than for a between-subjects design

$$F = \frac{\begin{array}{c}\text{between-group}\\ \text{variance}\end{array} \left\{ \begin{array}{c} \text{treatment effects?} \\ \cancel{\text{individual differences?}} \\ \text{random errors?} \end{array} \right.}{\begin{array}{c}\text{within-group}\\ \text{variance}\end{array} \left\{ \begin{array}{c} \text{individual differences?} \\ \text{random errors?} \end{array} \right.}$$

  - No need to change calculations, this is built in

Sources of Variability and Error

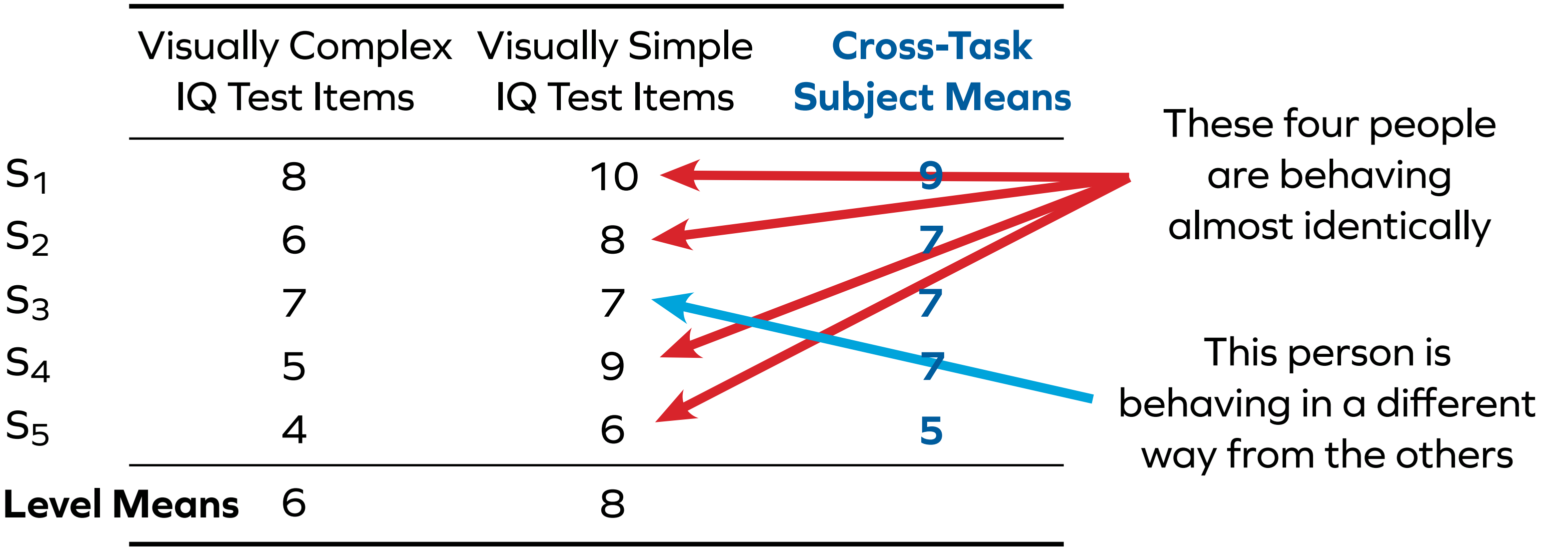
- **Within-Ss ANOVA: measurement of experimental error**
  - Within-group variance is caused by *individual differences* and *random errors of measurement*
  - But individual differences are ***irrelevant*** to hypothesis testing

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8	

DV: Number of items correct

Sources of Variability and Error

- **Within-Ss ANOVA: measurement of experimental error**
  - Experimental hypotheses entail *direction of difference*
    - Task A will have *more correct items* than Task B
    - Task A will be *faster to perform* than Task B



	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8	

These four people are behaving almost identically

This person is behaving in a different way from the others

DV: Number of items correct

Sources of Variability and Error

- **Within-Ss ANOVA: measurement of experimental error**
  - Two (or more) measurements per person
  - Gives an opportunity to remove variability caused by individual differences giving a ***smaller*** *error term*

$$F = \frac{\text{between-group variance} \left\{\begin{array}{c}\text{treatment effects?}\\ \text{\sout{individual differences?}}\\ \text{random errors?}\end{array}\right.}{\xcancel{\text{within-group variance}} \left\{\begin{array}{c}\text{\sout{individual differences?}}\\ \text{random errors?}\end{array}\right.}$$

  - Need to measure random errors of measurement to give the ***correct*** error term for a within-subjects design

Sources of Variability and Error

- **Within-Ss ANOVA**
  - An individual’s pattern of performance is given by the ***cross-task subject trend*** across experimental levels
  - *Variability* caused by *random errors* = variability in ***cross-task subject trends*** from person to person

	Visually Complex IQ Test Items	**Cross-Task Subject Trends**	Visually Simple IQ Test Items
S_1	8	↗ **2**	10
S_2	6	↗ **2**	8
S_3	7	**0**	7
S_4	5	↗ **4**	9
S_5	4	↗ **2**	6
Level Means	6		8

DV: Number of items correct

Sources of Variability and Error

- Within-group variance is the ***wrong*** measure of error

	Visually Complex IQ Test Items	Visually Simple IQ Test Items
S_1	8	10
S_2	6	8
S_3	7	7
S_4	5	9
S_5	4	6
Level Means	6	8

- Run as Between-Ss Anova
- $F(1,8) = 4$, $F_{\text{CRIT}} = 5.32$, NS
- Difference between level means is **small** in relation to variability of scores **within each level**

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	10	1	10	4	$p > .05$ NS
S/A WITHIN-GROUP	20	8	2.5		
TOTAL	30	9			

Sources of Variability and Error

- Within-group variance is the ***wrong*** measure of error

	Visually Complex IQ Test Items		Visually Simple IQ Test Items
S_1	8	↗ 2	10
S_2	6	↗ 2	8
S_3	7	0	7
S_4	5	↗ 4	9
S_5	4	↗ 2	6
Level Means	6	↔	8

- Differences between level means is ***large*** in relation to variability in the ***cross-task subject trends*** across levels
- There ***is*** an effect in these data

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	10	1	10	~~4~~	~~*p* > .05 NS~~
~~S/A~~ WITHIN-GROUP	~~20~~	~~8~~	~~2.5~~		
TOTAL	30	9			

Individual-Differences and Residual Variance

- **Individual-differences variance**
  - Measures variability caused by individual differences
  - Extent to which subjects differ overall *ignoring levels*
  - Calculated from the variability in the cross-task subject means [with the Grand Mean as the benchmark]

	Level A_1	Level A_2	**Cross-Task Mean**
S_1	?	?	**?**
S_2	?	?	**?**
S_3	?	?	**?**
S_4	?	?	**?**
Level Means	?	?	**GM ?**

Individual-Differences and Residual Variance

	Level A_1	Level A_2		**Cross-Task Mean**
S_1	1	9		**5**
S_2	5	5		**5**
S_3	6	4		**5**
S_4	4	6		**5**
Level Means	4	6	**GM**	**5**

- Individual-differences variance is zero

Individual-Differences and Residual Variance

	Level A_1	Level A_2	**Cross-Task Mean**
S_1	3	5	**4**
S_2	7	9	**8**
S_3	1	3	**2**
S_4	5	7	**6**
Level Means	4	6	**GM** **5**

- Individual-differences variance is greater than zero

Individual-Differences and Residual Variance

- **Residual variance**
  - Measures variability caused by random errors
  - Extent to which ***cross-task subject trends*** are ***consistent*** from subject to subject

	Level A_1	**Cross-Task Trend**	Level A_2
S_1	?	**?**	?
S_2	?	**?**	?
S_3	?	**?**	?
S_4	?	**?**	?
Level Means	?		?

Individual-Differences and Residual Variance

	Level A_1	**Cross-Task Trend**	Level A_2
S_1	3	↗ **2**	5
S_2	7	↗ **2**	9
S_3	1	↗ **2**	3
S_4	5	↗ **2**	7
Level Means	4		6

- Residual variance is zero

Individual-Differences and Residual Variance

	Level A_1	**Cross-Task Trend**	Level A_2
S_1	1	↗ **8**	9
S_2	5	**0**	5
S_3	6	↘ **2**	4
S_4	4	↗ **2**	6
Level Means	4		6

- Residual variance is greater than zero

Individual-Differences and Residual Variance

- All combinations are possible, separate and independent

Cross-task subject …

	Level A_1	Trends	Level A_2	Means
S_1	4	↗ 2	6	5
S_2	4	↗ 2	6	5
S_3	4	↗ 2	6	5
S_4	4	↗ 2	6	5
Level Means	**4**		**6**	

Individial-differences variance = zero

Residual variance = zero

Cross-task subject …

	Level A_1	Trends	Level A_2	Means
S_1	1	↗ 8	9	5
S_2	5	0	5	5
S_3	6	↘ 2	4	5
S_4	4	↗ 2	6	5
Level Means	**4**		**6**	

Individial-differences variance = zero

Residual variance = more than zero

Cross-task subject …

	Level A_1	Trends	Level A_2	Means
S_1	3	↗ 2	5	4
S_2	7	↗ 2	9	8
S_3	1	↗ 2	3	2
S_4	5	↗ 2	7	6
Level Means	**4**		**6**	

Individial-differences variance = more than zero

Residual variance = zero

Cross-task subject …

	Level A_1	Trends	Level A_2	Means
S_1	5	↘ 2	3	4
S_2	5	0	5	5
S_3	2	↗ 6	8	5
S_4	4	↗ 4	8	6
Level Means	**4**		**6**	

Individial-differences variance = more than zero

Residual variance = more than zero

How Within-Subjects ANOVA Works

$$F = \frac{\text{Variance}_{\text{EFFECT}}}{\text{Variance}_{\text{ERROR}}}$$

- Within-subjects ANOVA: error term = *residual variance*

$$F = \frac{\text{between-group variance}}{\text{residual variance}}$$

- Residual variance is caused by random errors of measurement and nothing else *by definition*

$$F = \frac{\begin{array}{c}\text{between-group}\\ \text{variance}\end{array} \left\{ \begin{array}{c}\text{treatment effects?}\\ \text{random errors?}\end{array} \right.}{\begin{array}{c}\text{residual}\\ \text{variance}\end{array} \left\{ \begin{array}{c}\text{random errors?}\end{array} \right.}$$

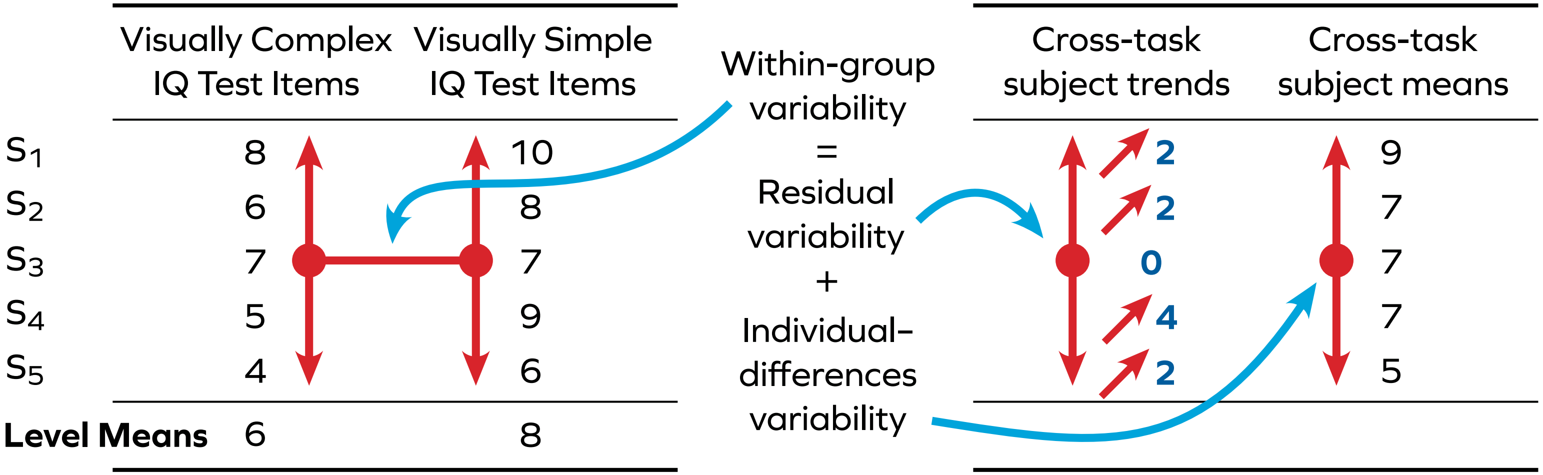
How Within-Subjects ANOVA Works

- All ANOVA error terms:
  - *Subjects treated in an identical way should behave in an identical way, the less this is the case, the more error in the experiment*
- Between-subjects ANOVA:
  Error term is within-group variance …
  - *Subjects within a level are treated the same way and therefore should get identical scores*
- Within-subjects ANOVA: Error term is residual variance …
  - *Subjects given two tasks are treated the same way and everyone should get larger scores for task x than task y*

How Within-Subjects ANOVA Works

- Within-group variability =
  - Individual-differences variability

    +
  - Residual (random error) variability



	Visually Complex IQ Test Items	Visually Simple IQ Test Items
S_1	8	10
S_2	6	8
S_3	7	7
S_4	5	9
S_5	4	6
Level Means	6	8

Within-group variability
=
Residual variability
+
Individual-differences variability

Cross-task subject trends	Cross-task subject means
2	9
2	7
0	7
4	7
2	5

How Within-Subjects ANOVA Works

- Not exactly like this but close enough …

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8	

Between-group variance = 10
Within-group variance = 2.5

How Within-Subjects ANOVA Works

- Calculate the cross-task mean for each subject
  (components of individual differences variability)
- Subtract each subject's own cross-task mean from that subject's own scores
  (removes individual differences variability)

How Within-Subjects ANOVA Works

- Individual differences variability (via cross task means) has gone but residual variability (cross-task trends) untouched
- Within-group variability much smaller, it now only contains residual variability, therefore it ***is*** the residual variability
- Between-group variability (difference in means) untouched

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	-1	+1	**0**
S_2	-1	+1	**0**
S_3	0	0	**0**
S_4	-2	+2	**0**
S_5	-1	+1	**0**
Level Means	-1	+1	

Between-group variance = 10
Within-group variance = 0.5

How Within-Subjects ANOVA Works

- Error term becomes smaller: *within-group variance has become **depolluted**, pure random errors of measurement*
- *F* ratio (almost always) becomes larger
- *F* ratio is (almost always) more likely to be significant
- Hence *increased statistical power*

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	-1	+1	**0**
S_2	-1	+1	**0**
S_3	0	0	**0**
S_4	-2	+2	**0**
S_5	-1	+1	**0**
Level Means	-1	+1	

Between-group variance = 10
Within-group variance = 0.5

Single-Factor Within-Subjects ANOVA

- New Basic Ratio: *Basic Ratio of the Subject Totals*
- Calculating Sums of Squares and degrees of freedom

Single-Factor Within-Subjects ANOVA

- One new basic ratio needed for calculation

$[S]$ = The Basic Ratio of the subject totals

$$\frac{(\Sigma S_1)^2 + (\Sigma S_2)^2 + (\Sigma S_3)^2 + (\Sigma S_4)^2 + (\Sigma S_5)^2 \text{ etc.}}{a}$$

compare

$[A]$ = The Basic Ratio of the level totals

$$\frac{(\Sigma A_1)^2 + (\Sigma A_2)^2 \text{ etc.}}{s}$$

Single-Factor Within-Subjects ANOVA

- One new basic ratio needed for calculation

$$[S] = \text{The Basic Ratio of the subject totals}$$

$$\frac{(S_1\ \text{Total})^2 + (S_2\ \text{Total})^2 + (S_3\ \text{Total})^2 + (S_4\ \text{Total})^2 + (S_5\ \text{Total})^2\ \text{etc.}}{\text{the number of scores that make up EACH } \textbf{subject} \text{ total}}$$

compare

$$[A] = \text{The Basic Ratio of the level totals}$$

$$\frac{(\text{Level Total } A_1)^2 + (\text{Level Total } A_2)^2\ \text{etc.}}{\text{the number of scores that make up EACH } \textbf{level} \text{ total}}$$

Single-Factor Within-Subjects ANOVA

- Plus the two other basic ratios, same as before

$[T]$ = The Basic Ratio of the grand total

$$\frac{(\text{Grand Total})^2}{\text{the number of scores that make up the grand total}}$$

and

$[Y]$ = The Basic Ratio of the individual scores

$$\frac{(Y_1)^2+(Y_2)^2+(Y_3)^2+(Y_4)^2+(Y_5)^2+(Y_6)^2+(Y_7)^2+(Y_8)^2+(Y_9)^2 \text{ etc.}}{1}$$

Single-Factor Within-Subjects ANOVA

- Four basic ratios to calculate

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Totals**
S_1	8	10	**18**
S_2	6	8	**14**
S_3	7	7	**14**
S_4	5	9	**14**
S_5	4	6	**10**
Level Totals	**30**	**40**	**Grand Total 70**

[Y] = The Basic Ratio of the individual scores

[A] = The Basic Ratio of the level totals

[S] = The Basic Ratio of the subject totals

[T] = The Basic Ratio of the grand total

Single-Factor Within-Subjects ANOVA

- Four basic ratios to calculate

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Totals**
S_1	8	10	**18**
S_2	6	8	**14**
S_3	7	7	**14**
S_4	5	9	**14**
S_5	4	6	**10**
Level Totals	**30**	**40**	**Grand Total 70**

$[Y] =$

$[A] =$

$[S] =$

$[T] =$

Single-Factor Within-Subjects ANOVA

- We will need to calculate the components of the individual-differences variance to calculate the error term
- Calculation is analogous to between-group variance

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8	**Grand Mean 7**

- Between-group variance calculated from the squared distances of the ***level*** means from the grand mean

Single-Factor Within-Subjects ANOVA

- Individual-differences variance calculated from the squared distances of the (cross task) ***subject*** means from the grand mean

	Visually Complex IQ Test Items	Visually Simple IQ Test Items	**Cross-Task Subject Means**
S_1	8	10	**9**
S_2	6	8	**7**
S_3	7	7	**7**
S_4	5	9	**7**
S_5	4	6	**5**
Level Means	6	8 **Grand Mean**	7

Single-Factor Within-Subjects ANOVA

$$\text{Between-group variance} = \frac{SS_{\text{BETWEEN}}}{df_{\text{BETWEEN}}} = \frac{[A] - [T]}{(a - 1)}$$

$$\text{Individual-differences variance} = \frac{SS_{\text{SUBJECTS}}}{df_{\text{SUBJECTS}}} = \frac{[S] - [T]}{(s - 1)}$$

- **a** = number of levels in Factor A
- **s** = number of scores in each level of Factor A

Single-Factor Within-Subjects ANOVA

- Sums of Squares are additive, can fully compartmentalise

$$SS_{TOTAL} = SS_{BETWEEN} + SS_{WITHIN}$$

and

$$SS_{WITHIN} = SS_{SUBJECTS} + SS_{RESIDUAL}$$

$$SS_{RESIDUAL} = SS_{WITHIN} - SS_{SUBJECTS}$$

$$SS_{RESIDUAL} = [Y] - [A] - ([S] - [T])$$

$$SS_{RESIDUAL} = [Y] - [A] - [S] + [T]$$

for the *error term*

Single-Factor Within-Subjects ANOVA

- Degrees of Freedom are also additive

$$df_{\text{TOTAL}} = df_{\text{BETWEEN}} + df_{\text{WITHIN}}$$

and

$$df_{\text{WITHIN}} = df_{\text{SUBJECTS}} + df_{\text{RESIDUAL}}$$

$$df_{\text{RESIDUAL}} = df_{\text{WITHIN}} - df_{\text{SUBJECTS}}$$

$$df_{\text{RESIDUAL}} = a(s - 1) - (s - 1)$$

$$df_{\text{RESIDUAL}} = (a - 1)(s - 1)$$

for the *error term*

Single-Factor Within-Subjects ANOVA

$$SS_{\mathrm{BETWEEN}} = [A] - [T]$$

$$SS_{\mathrm{SUBJECTS}} = [S] - [T]$$

$$SS_{\mathrm{RESIDUAL}} = [Y] - [A] - [S] + [T]$$

$[Y] = 520$, $[A] = 500$, $[S] = 506$, $[T] = 490$

Single-Factor Within-Subjects ANOVA

$df_{BETWEEN} = (a - 1)$

$df_{SUBJECTS} = (s - 1)$

$df_{RESIDUAL} = (a - 1)(s - 1)$

$a = 2, s = 5$

Interpreting the ANOVA Table

- Extra row added to ANOVA table

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP					
S INDIVIDUAL-DIFFS.					
AxS RESIDUAL					
TOTAL					

- Within-group variance replaced with its two components
  - **$SS_{SUBJECTS}$** abbreviated to **SS_{S}** and **$df_{SUBJECTS}$** to **df_{S}**
  - **$SS_{RESIDUAL}$** abbreviated to **SS_{AxS}** and **$df_{RESIDUAL}$** to **df_{AxS}**

Interpreting the ANOVA Table

- ***Error term*** is the measure of the experimental error

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	10	1	10	10	
S INDIVIDUAL-DIFFS.	16	4			
AxS RESIDUAL	4	4	1		
TOTAL	30	9			

$$F = \frac{\text{Variance}_{\text{EFFECT}}}{\text{Variance}_{\text{ERROR}}}$$

- Residual variance for a within-subjects ANOVA

Interpreting the ANOVA Table

- Is the *F* value large enough to be significant?

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	10	1	10	10	
S INDIVIDUAL-DIFFS.	16	4			
AxS RESIDUAL	4	4	1		
TOTAL	30	9			

- *F* value has Degrees of Freedom = (1,4)
- Check with computer or tables of critical values

Interpreting the ANOVA Table

- ACROSS 1 then DOWN 4

Critical Values of the F *Distribution*

.05 significance level in bold type

.01 significance level in plain type

degrees of freedom of numerator

edom of denominator

	1	2	3	4	5
1	**161**	**200**	**216**	**225**	**230**
	4052	4999	5403	5625	5764
2	**18.5**	**19.0**	**19.2**	**19.2**	**19.3**
	98.5	99.0	99.2	99.2	99.3
3	**10.1**	**9.55**	**9.28**	**9.12**	**9.01**
	34.1	30.8	29.5	28.7	28.2
4	**7.71**	**6.94**	**6.59**	**6.39**	**6.26**
	21.2	18.0	16.7	16.0	15.5
5	**6.61**	**5.79**	**5.41**	**5.19**	**5.05**
	16.3	13.3	12.1	11.4	11.0
6	**5.99**	**5.14**	**4.76**	**4.53**	**4.39**
	13.8	10.9	9.78	9.15	8.75
7	**5.59**	**4.74**	**4.35**	**4.12**	**3.97**
	12.2	9.55	8.45	7.85	7.46
8	**5.32**	**4.46**	**4.07**	**3.84**	**3.69**
	11.3	8.65	7.59	7.01	6.63
9	**5.12**	**4.26**	**3.86**	**3.63**	**3.48**
	10.6	8.02	6.99	6.42	6.06
10	**4.96**	**4.10**	**3.71**	**3.48**	**3.33**
	10.0	7.56	6.55	5.99	5.64
11	**4.84**	**3.98**	**3.59**	**3.36**	**3.20**
	9.65	7.21	6.22	5.67	5.32

Interpreting the ANOVA Table

- Critical *F*, df = (1,4), at $p < .05$ significance level = 7.71

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	10	1	10	10	$p < .05$ *
S INDIVIDUAL-DIFFS.	16	4			
AxS RESIDUAL	4	4	1		
TOTAL	30	9			

- Observed $F(1,4) = 10$
- Significant effect of item type, $p < .05$

Interpreting the ANOVA Table

- Mean for Visually Simple items significantly greater than mean for Visually Complex items

	Visually Complex IQ Test Items	Visually Simple IQ Test Items
S_1	8	10
S_2	6	8
S_3	7	7
S_4	5	9
S_5	4	6
Level Means	6	8

DV: Number of items correct

Follow-Up Tests for Three or More Levels

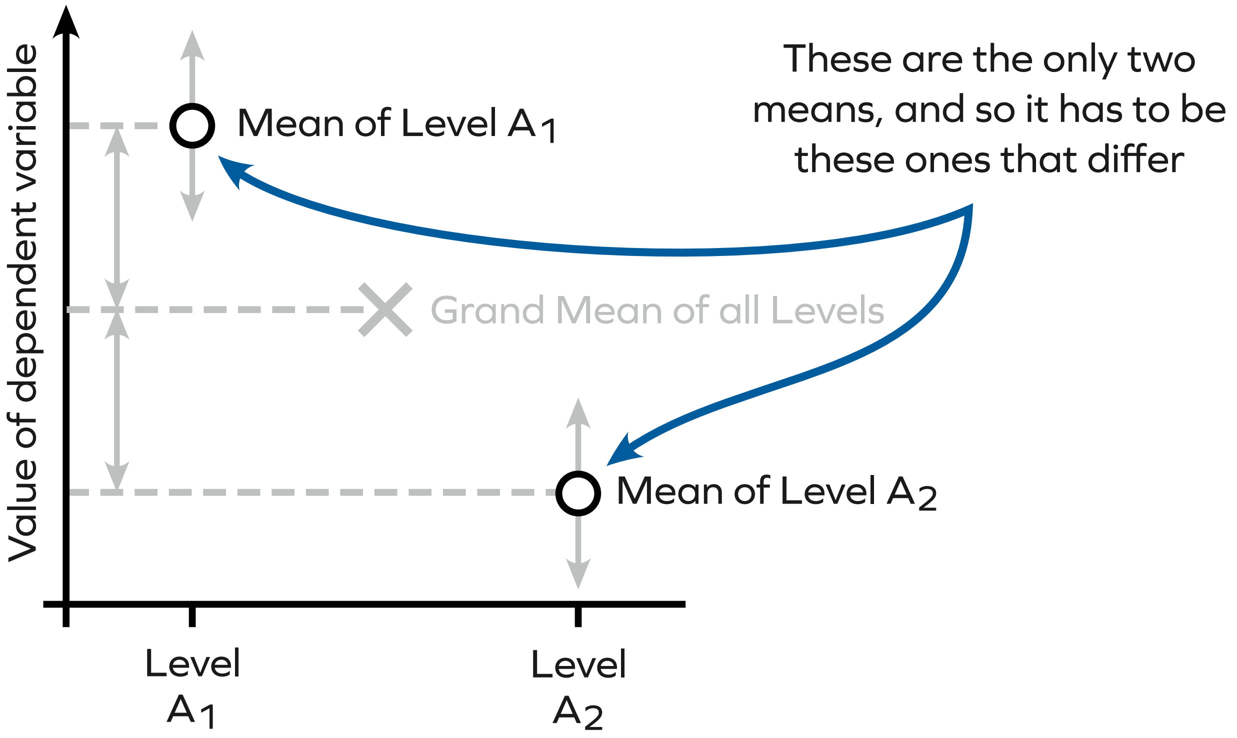
- Once ANOVA is calculated, interpretation is similar *whatever the design*
- Same *problem of multiple comparisons* and same solution strategies as for between-subjects designs
    - Correct/Protect/Select
    - *Post-hoc* tests versus Planned Comparisons
- Same formulae and methods as before
- Simply use the appropriate error term from the ANOVA table where needed

Follow-Up Tests for Three or More Levels

- Once ANOVA is calculated, interpretation is similar *whatever the design*
- Same *problem of multiple comparisons* and same solution strategies as for between-subjects designs

The Problem of Multiple Comparisons

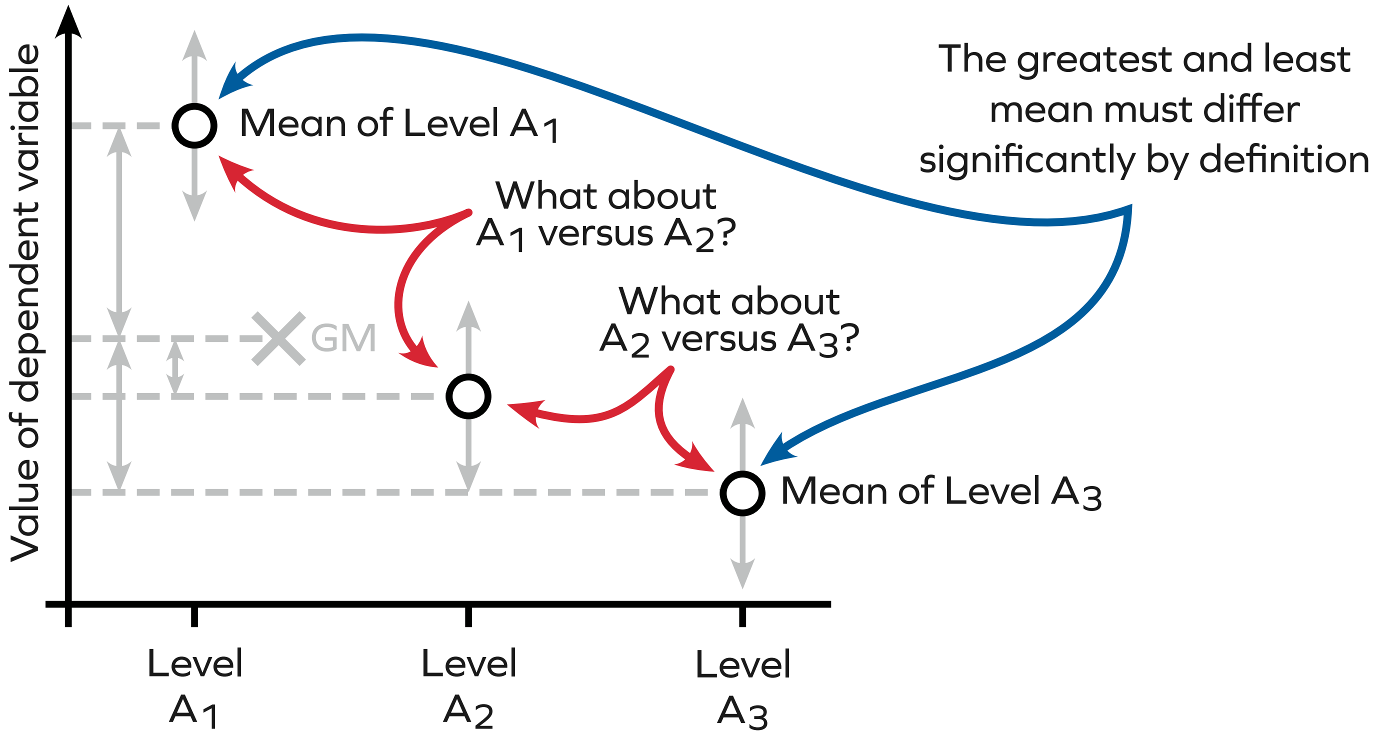
- Two levels in a factor
  - Significant ***difference*** is unambiguous
  - The two means *must* differ significantly



These are the only two means, and so it has to be these ones that differ

The Problem of Multiple Comparisons

- Three or more levels in a factor
  - Significant ***effect*** is ambiguous
  - Which exact pairs of means differ significantly?



Follow-Up Tests for Three or More Levels

- Once ANOVA is calculated, interpretation is similar *whatever the design*
- Same *problem of multiple comparisons* and same solution strategies as for between-subjects designs
  - Correct/Protect/Select
  - *Post-hoc* tests versus Planned Comparisons
- Same formulae and methods as before
- Simply use the appropriate error term from the ANOVA table where needed

Follow-Up Tests for Three or More Levels

- Example experiment

Sam is taller than James

Peter is taller than Dave

Dave is taller than Sam

Who is taller?
James or Dave?

- Suggestion: these are represented as a mental ‘ladder’

Peter

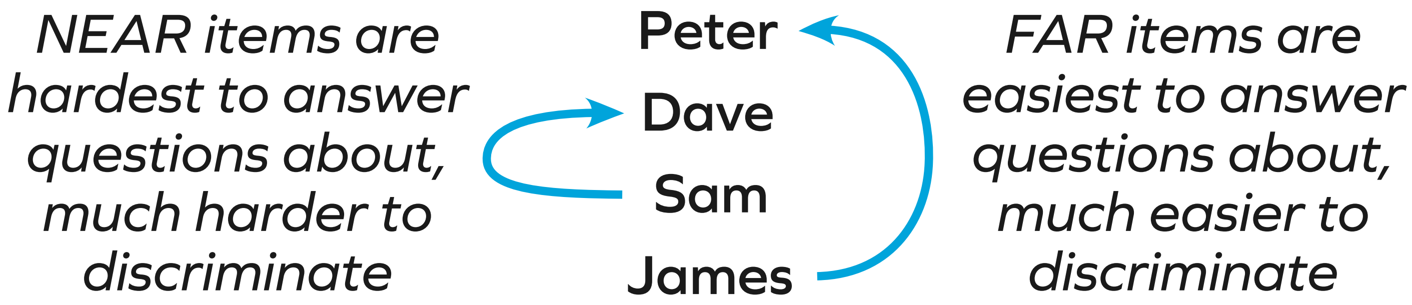
Dave

Sam

James

Follow-Up Tests for Three or More Levels

- Predicts an *inferential distance effect*



NEAR items are hardest to answer questions about, much harder to discriminate

Peter

Dave

Sam

James

FAR items are easiest to answer questions about, much easier to discriminate

  - FAR items should be easiest: P?J
  - MID items harder: P?S, D?J
  - NEAR items hardest: P?D, D?S, S?J

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

Dependent Variable: % correct trials

- 3 levels in the factor: **a = 3**
- 8 scores per level: **s = 8**

Follow-Up Tests for Three or More Levels

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	149.3	2	74.667	12.16	$p < .01$ **
S BETWEEN-SUBJECT	740.0	7			
AxS RESIDUAL	86.0	14	6.143		
TOTAL	975.3	23			

- Obtained value: $F(2,14) = 12.16$
- Critical value: $F_{\textbf{CRIT}} = 3.74$
- Good statistical evidence for treatment effects

Follow-Up Tests for Three or More Levels

- No clear predictions ***or***
  every possible comparison necessary/informative

  2) *Post-hoc* comparisons
     using the ***Tukey test***

✓ **Correct:** but less stringent than Bonferroni correction

✓ **Protect:** ANOVA must be significant before proceeding

✗ ~~**Select:**~~ all possible comparisons intended to be made

[Newman-Keuls is similar, less stringent than Tukey]

Follow-Up Tests for Three or More Levels

- Need significant *F* value from main ANOVA table
- Test is designed to make all possible pairwise comparisons between means
- Special formula corrects according to number of comparisons and gives a ***critical difference***
- Less stringent than Bonferroni correction once the initial ANOVA protection has been satisfied

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

Dependent Variable: % correct trials

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value	*p*-value (sig. level)
A BETWEEN-GROUP	149.3	2	74.667	12.16	$p < .01$ **
S INDIVIDUAL-DIFFS.	740.0	7			
AxS RESIDUAL	86.0	14	6.143		
TOTAL	975.3	23			

- *F* on main ANOVA table is significant
- Sound evidence for treatment effects
- Apply Tukey test to investigate the effects

Follow-Up Tests for Three or More Levels

$$\text{Critical difference, W} = r\sqrt{\frac{\text{Variance}_{\text{ERROR}}}{s}}$$

- *r* is obtained from *studentised range statistic* tables
    - Depends on number of levels $(a = 3)$ and df_{ERROR} (14)
    - $r = 3.70$ for this experiment
    - You will never need to use the table/equation, ever!

$$W = 3.70\sqrt{\frac{6.14}{8}} = 3.24$$

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

- Critical difference = 3.24
    - FAR vs. MID
    - FAR vs. NEAR
    - MID vs. NEAR

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

- Two significant differences:
  - FAR vs. MID $p < .05$ *
  - FAR vs. NEAR $p < .05$ *
  - MID vs. NEAR $p > .05$ NS

- Not quite evidence for a ***sequential*** difficulty progression

Follow-Up Tests for Three or More Levels

- Only some of the possible comparisons will be targetted

3) Planned comparisons using ***pairwise F* tests** and (*possibly*) the ***Bonferroni correction***

   - Increase statistical power by thinking ahead
   - Use pairwise *F* tests to make only the ***crucial*** comparisons for a hypotheses
   - Advance planning/honesty/economy means no need to protect or correct the tests

Follow-Up Tests for Three or More Levels

- **Few comparisons**

  - **(a – 1)** comparisons [number of levels minus 1]

✗ ~~**Correct:**~~ no need, Type I Error is kept under control

✗ ~~**Protect:**~~ no need, Type I Error is kept under control

✓ **Select: (a – 1)** comparisons

  - Very powerful, always use this method if possible

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

- ***What is the research question?***
  - Is there a *sequential difficulty progression* from FAR to MID to NEAR?
    - Only need to test ***two of the pairwise comparisons***
    - ***Before*** collecting data, planned to compare
      - FAR vs. MID trials
      - MID vs. NEAR trials

Follow-Up Tests for Three or More Levels

- Quick formula for pairwise comparisons once main ANOVA table is computed

$$F = \frac{\frac{s}{2}\left(\overline{A}_i - \overline{A}_j\right)^2}{\text{Variance}_{\text{ERROR}}}$$

- **s** = number of scores in each level = **8**
- Error term is from main ANOVA table = **6.143**
- **df = 1** for between-group variance; **(a – 1)**
  df = 14 for the error term (same as original table)

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

Dependent Variable: % correct trials

$$F = \frac{\frac{s}{2}\left(\overline{A}_i - \overline{A}_j\right)^2}{\text{Variance}_{\text{ERROR}}} = \frac{4\left(\overline{A}_i - \overline{A}_j\right)^2}{6.143}$$

- FAR vs. MID trials

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

Dependent Variable: % correct trials

$$F = \frac{\frac{s}{2}\left(\overline{A}_i - \overline{A}_j\right)^2}{\text{Variance}_{\text{ERROR}}} = \frac{4\left(\overline{A}_i - \overline{A}_j\right)^2}{6.143}$$

- MID vs. NEAR trials

Follow-Up Tests for Three or More Levels

- Obtained values:
  - FAR vs. MID: $F(1,14) = 10.4$
  - MID vs. NEAR: $F(1,14) = 2.59$
- Critical value for df = (1,14) at **.05** sig. level: $F_{\mathbf{CRIT}} = \mathbf{4.60}$
- **(a – 1)** comparisons, so no need for Bonferroni correction

Follow-Up Tests for Three or More Levels

	FAR trials	MID trials	NEAR trials
Level Means	87.0	83.0	81.0

Dependent Variable: % correct trials

- Obtained values:
  - FAR vs. MID: $F(1,14) = 10.4$ $p < .05$, Sig
  - MID vs. NEAR: $F(1,14) = 2.59$ $p > .05$, NS

- Insufficient evidence for sequential difficulty progression
- Perhaps there is something special about the FAR trials

Single-Factor Within-Subjects ANOVA

Source	Sum of Squares	Degrees of Freedom	Variance (Mean Square)	*F*-value
A BETWEEN-GROUP	$[A] - [T]$	$(a - 1)$	$\frac{[A] - [T]}{(a - 1)}$	$\frac{\text{VARIANCE}_{\text{A}}}{\text{VARIANCE}_{\text{AxS}}}$
S INDIVIDUAL-DIFFS.	$[S] - [T]$	$(s - 1)$		
AxS RESIDUAL	$[Y] - [A] - [S] + [T]$	$(a - 1)(s - 1)$	$\frac{[Y] - [A] - [S] + [T]}{(a - 1)(s - 1)}$	
TOTAL	$[Y] - [T]$	$(as) - 1$		

- $[T]$ is the basic ratio of the grand total
  $[A]$ is the basic ratio of the level totals
  $[S]$ is the basic ratio of the subject totals
  $[Y]$ is the basic ratio of the individual scores
- a is the number of levels in Factor A
  s is the number of scores in *each level* of Factor A