psyc3010 lecture 10

Mediation in MR
One and two-way within subjects anova

Before the break: moderated multiple regression
next week: mixed anova

two weeks ago → this week

- Before the break we looked at how to test interactions in multiple regression – and saw that it achieved a similar thing to factorial anova.

- this week we go back to anova to look at within subjects designs
  - **One-way**
  - **Two-way**

- **But before that, the grooviness of *mediation* in MR!**

hierarchical models are used to:

- control for nuisance variable(s)
- answer theoretical questions about the relative contribution of sets of variables
- test moderated relationships (interactions)
  $\hat{Y} = b_1X_1 + b_2X_2 + b_3X_1X_2 + a$
- test for curvilinear relationships
  $\hat{Y} = b_1X + b_2X^2 + a$
- test categorical variables with >2 levels
- test hypothesized causal order:
  - Mediation (also via path analysis, structural equation modelling, etc. – in later courses!)

what mediation means

- so far, we've considered ***direct*** relationships between predictors and a criterion

  e.g., more study time → higher exam mark

- sometimes, these relationships don't say much about underlying processes or mechanisms

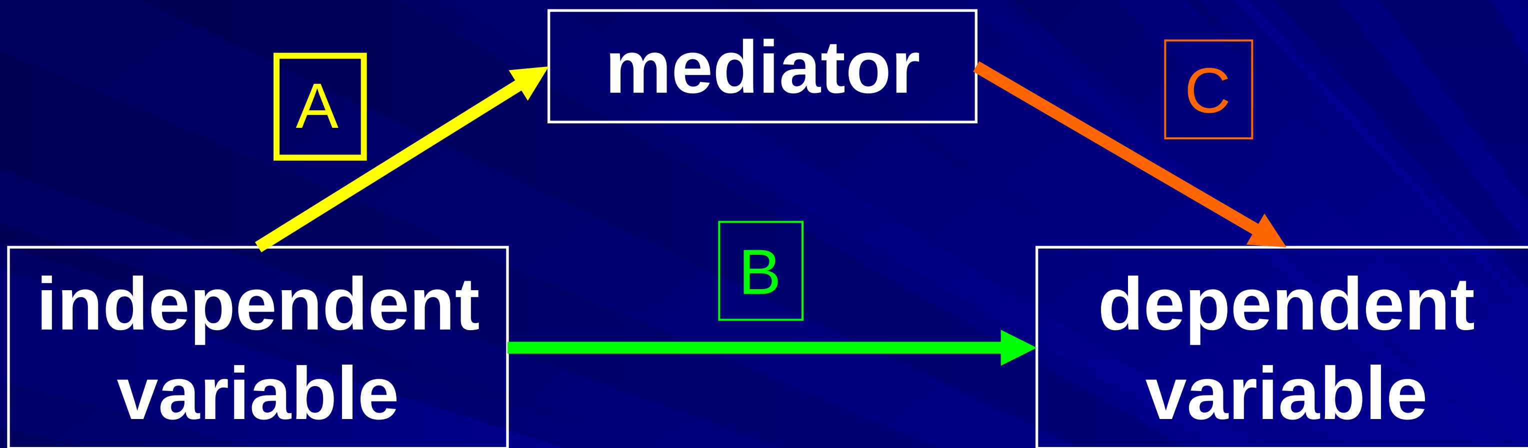
  → ***why*** does increased study time improve exam marks?

- a third variable, or ***mediator***, may ***explain*** or ***account for*** the relationship between an IV and a DV

  e.g., study time → **retention of material** → exam mark

- thus, the original predictor has an ***indirect*** relationship with the criterion

mediation:
IV causes DV indirectly through mediator



1) IV is related to ('causes') mediator (path a)
2) IV is related to DV (path b)
3) mediator is related to DV (path c)
4) IV is no longer related to DV when effect of mediator is controlled for (path b.c)

testing and reporting mediation

1. **IV is related to mediator** **(path a)**
- Conduct regression of mediator on IV. Report sig R2 and b or beta.

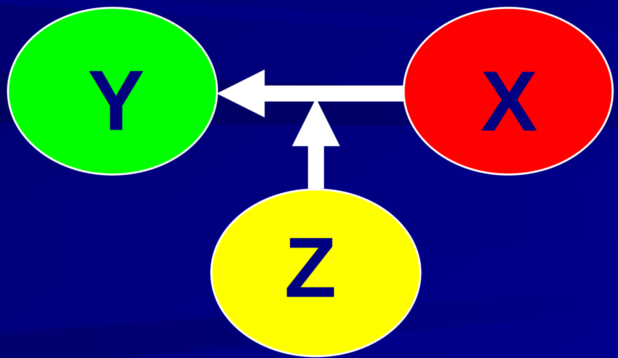
2. **IV is related to DV** **(path b)**
- Conduct HMR regressing DV on IV alone in block 1. Report sig R2 and b or beta.

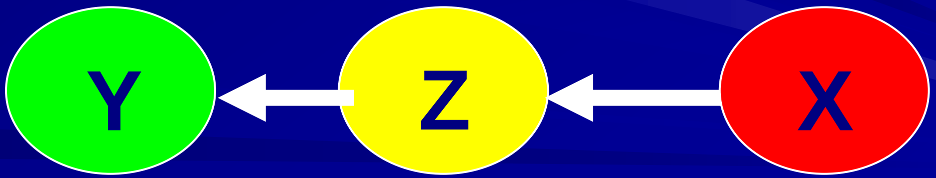
3. **mediator is related to DV** **(path c)**
4. **when paths a and c are controlled, path b is no longer significant**
- Add mediator in Block 2. R2 change need not increase significantly – if coefficient for mediator is sig, condition (ii) is met. If IV coefficient in block 2 is no longer significant, condition (iv) is met. Report sig coeff for mediator and IV ns in this block and conclude, since all 4 conditions are met, that the effect of the IV on the DV is fully mediated by the mediator.

- Often in write-up you would also present a figure, as on previous slide
- Also need to conduct "Sobel test" to see if mediation is reliable – see http://www.people.ku.edu/~preacher/sobel/sobel.htm if interested.

moderation vs mediation

- moderation and mediation are two <u>widely</u> confused terms
  - in moderation: 1. The direct X-> Y relationship is the focus. At low Z, the X->Y relationship is stronger, weaker, or reversed compared to the X->Y relationship at high Z. I.e., X & Z interact. 2. There is no "because". 3. Moderator could be (and often is) uncorrelated with IV. E.g., Exercise interacts with (weakens effect of) life hassles -> lower well-being.
  - in mediation, 1. The indirect relationship of X -> Y via Z is the focus. 2. X causes Y **because** X causes Z which in turn causes Y: X→Z→Y. 3. The mediator is associated with the IV (positively or negatively). E.g., Exercise -> lower subjective stress -> well-being.





anova – a second look

- **between-subjects designs**
  - each person serves in only one treatment/cell
  - we then assume that any difference between them is due to our experimental manipulation (or intrinsic features of the grouping variable, e.g., gender)
  - Within-cell variability is residual error
- **within-subjects (repeated-measures) designs**
  - what if each subject served in each treatment?
  - violates the assumption of independence in factorial ANOVA because scores for the participant are correlated across conditions
  - but we can calculate and ***remove*** any variance due to dependence
  - thus, we can reduce our error term and increase power

an illustration

	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

treatment means don’t differ by much – far more variability within each group than between

an illustration

	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

most of this within-group variance is caused by the fact that some subjects learn quickly, and some learn slowly – i.e., **people are different**

In between-subjects design, all within-group variance is error, whereas repeated measures design remove individual difference variation from the error term.

an illustration

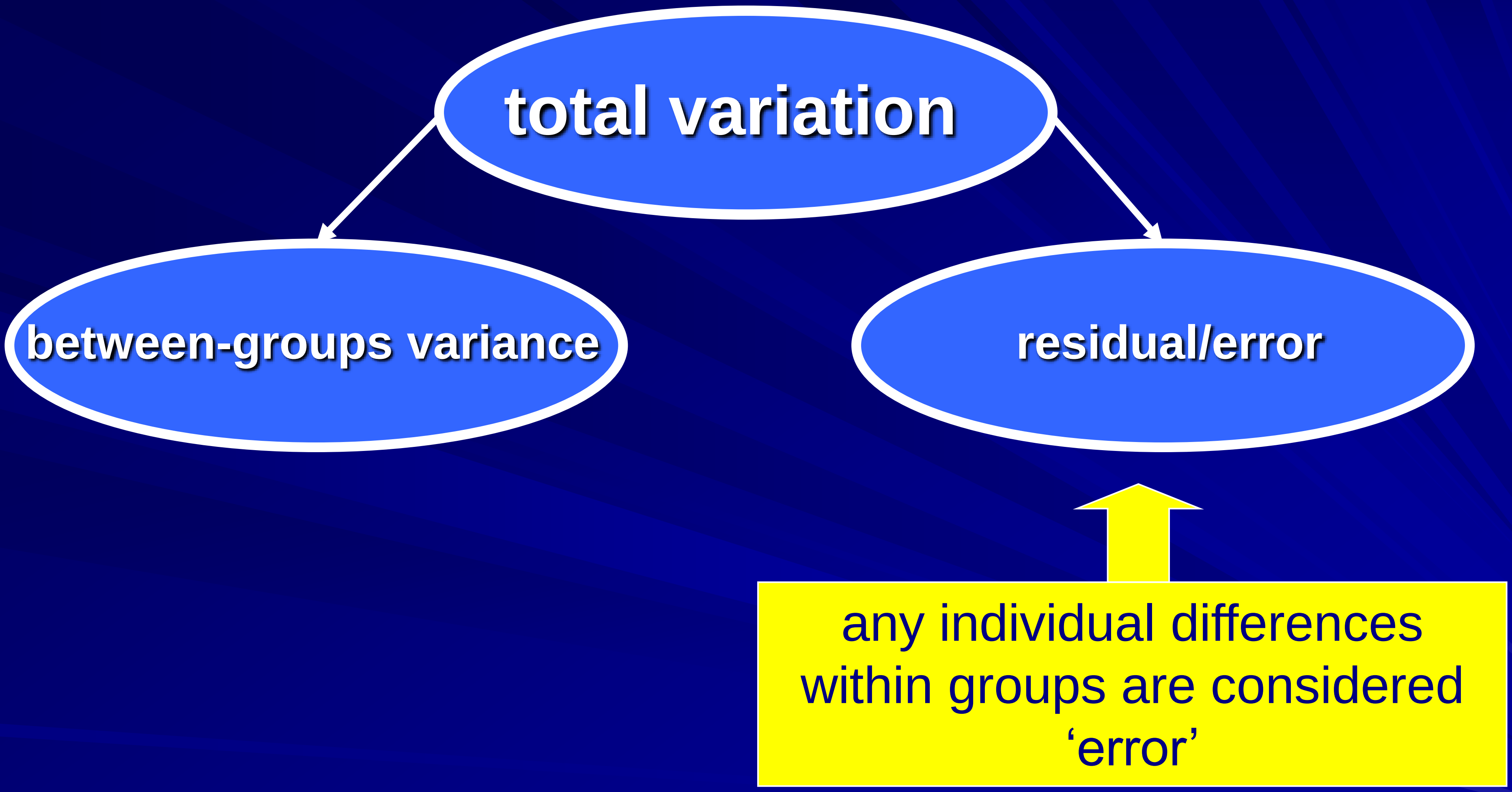
	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

solution: firstly remove the between-subjects variance (i.e., account for individual differences) and *then* compare our treatment means

Understanding RM versus BS designs

- In between subjects, assign people randomly to j conditions
  - **Total Variance = Between group + within group**
    - Treatment effect = between group variance
    - Error = within group variance
- No subject variability because each participant has only 1 data point (no variance within individual)

1-way between-subjects anova:



Understanding RM designs

- In fully within subjects design, people are tested in each of j conditions
- “subject” factor is crossed with IV (e.g., factor A)
- End up with A x S design with only 1 observation per cell
- No within-cell variance – now a cell is one observation (for person i in condition j)
- So what is error?
  - Interaction of A x S – i.e., the changes (inconsistency) in the effects of A across subjects

A x S design

	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

Overall, treatment effect for 1 = 16 – 18.67 (-2.67)
treatment effect for 2 = 19-18.67 (+0.33)
treatment effect for 3 = 21 – 18.67 (+2.33)

A x S design

	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

For S1, treatment effect for 1 = 2 – 4.33 (-2.33)
treatment effect for 2 = 4 – 4.33 (-0.33)
treatment effect for 3 = 7 – 4.33 (+2.67)

A x S design

	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

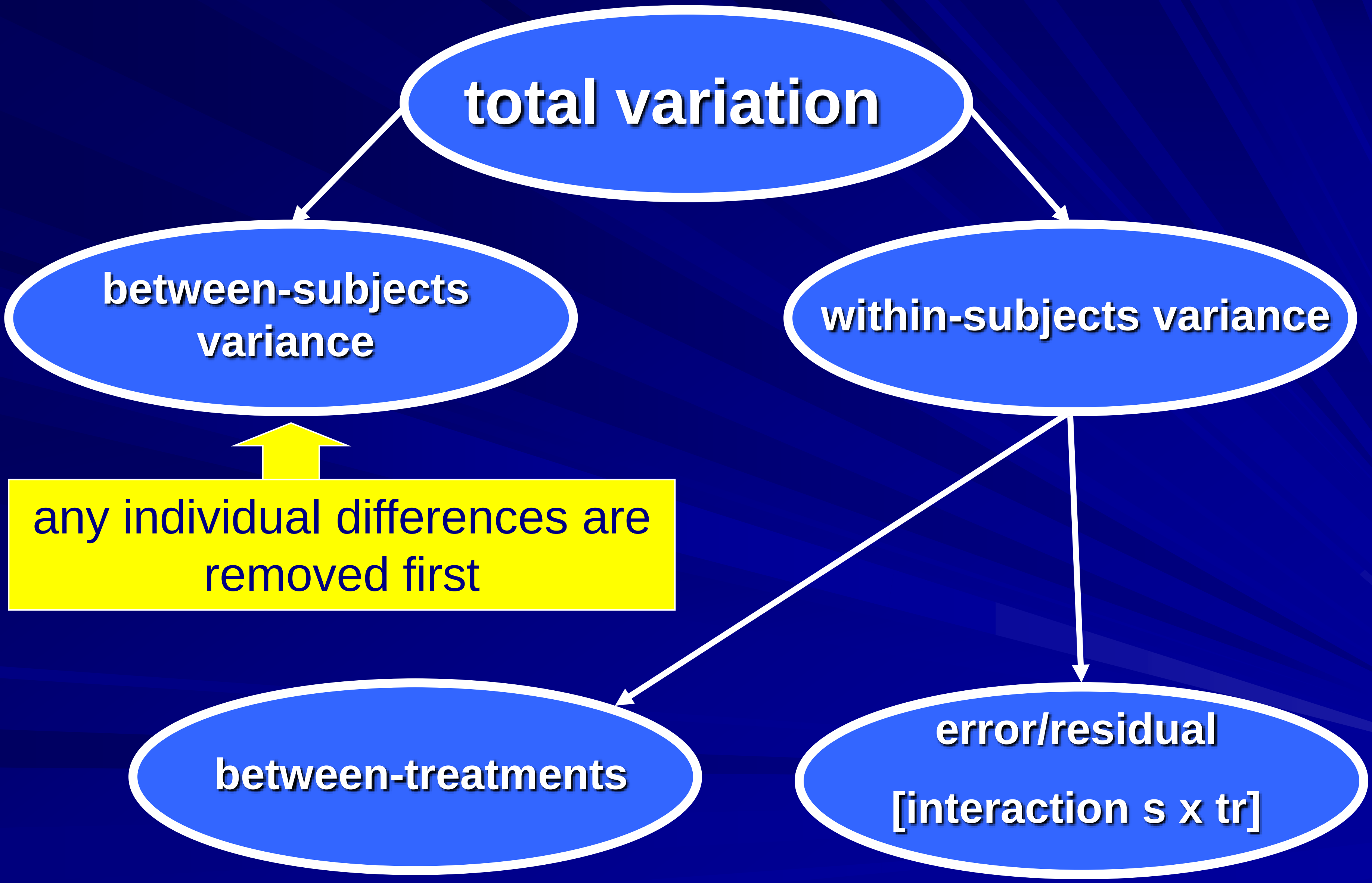
For S2, treatment effect for 1 = 10 – 11.67 (-1.67)
treatment effect for 2 = 12-11.67 (+0.33)
treatment effect for 3 = 13-11.67 (+1.33)

A x S design

	treatment			
subject	1	2	3	mean
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27.00
4	30	31	34	31.67
mean	16	19	21	18.67

For S3, treatment effect for 1 = 22 – 27 (-5)
treatment effect for 2 = 29 – 27 (+2)
treatment effect for 3 = 30 – 27 (+3)

1-way <u>within</u>-subjects anova:



Within-subjects design

- **Total Variance = Between subjects + within subjects**

Between subjects variance due to individual differences is partitioned out of error (and treatment)!

  - **Within subjects = between treatment** [treatment effect] + **treatment x subject interaction** [residual error – i.e., inconsistencies in the treatment effect]
  - **F test = TR / TR x S**

- Acknowledges reality that variability within conditions/groups and between conditions/groups are both influenced by subject factor [people doing study]

the conceptual model

$$X_{ij} = \mu + \pi_i + \tau_j + e_{ij}$$

for *i* cases and *j* treatments:

$X_{ij,}$ *any DV score* is a combination of:

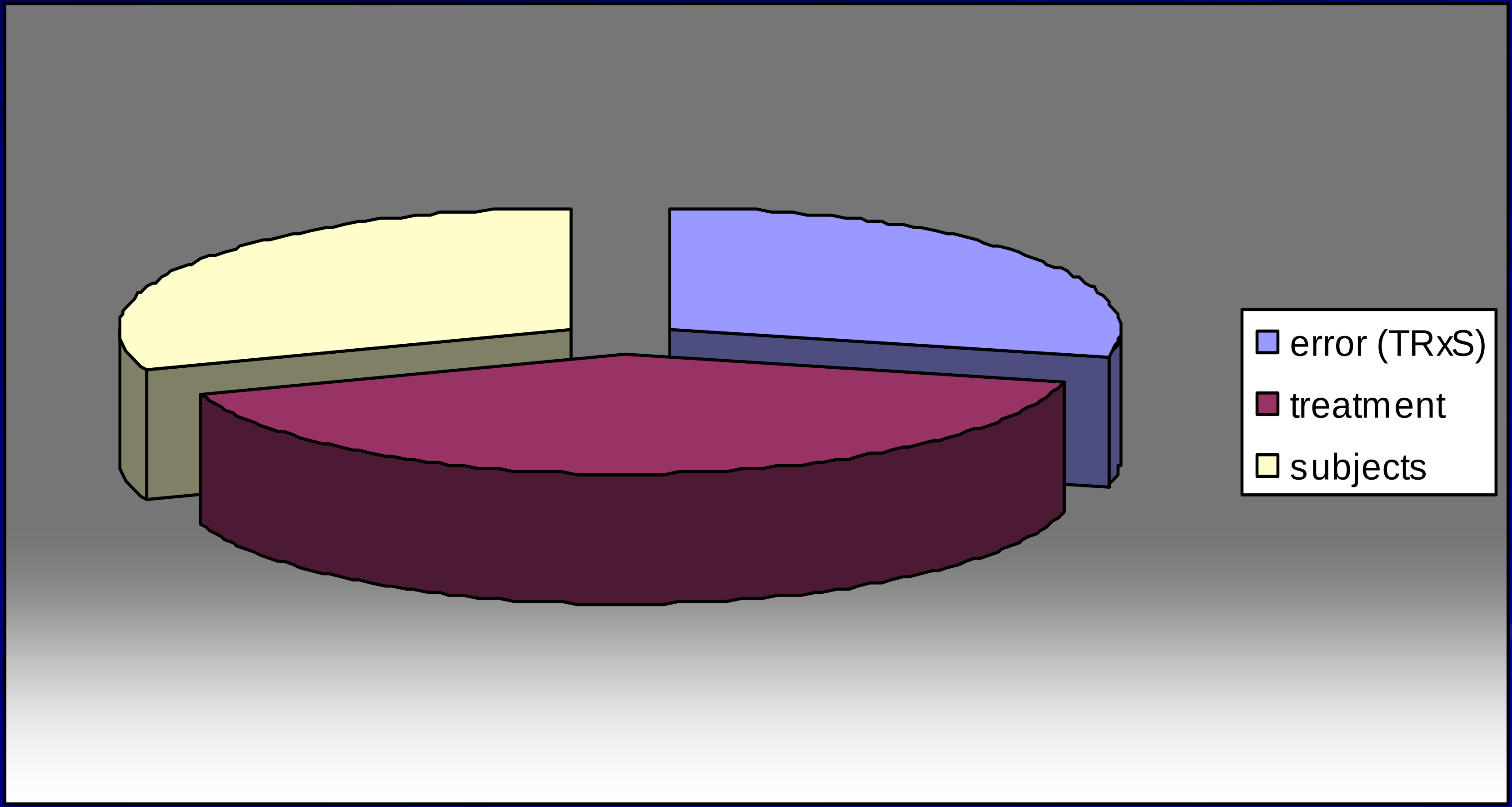
μ → *the grand mean,*

π_i → *variation due to the i-th person* $(\mu_i - \mu)$

τ_j → *variation due to the j-th treatment* $(\mu_j - \mu)$

e_{ij} → *error - variation associated with the i-th cases in the j-th treatment – error =* $\pi\tau_{ij}$ (plus chance)

partitioning the variance



worked example

basic learning study

- 1-way within-subjects design (*n*=5)
- IV: block
  - 40 trials through whole experiment
  - want to compare over 4 blocks of 10 to see if learning has occurred
- DV = number of correct responses per block

correct trials over 4 blocks of 10

	block 1	block2	block3	block 4	subj total
subject 1	4	3	6	5	**18**
subject 2	4	4	7	8	**23**
subject 3	1	2	1	3	**7**
subject 4	1	4	5	5	**15**
subject 5	5	7	6	9	**27**
block total	**15**	**20**	**25**	**30**	**90**
block mean	**3**	**4**	**5**	**6**	

$SS_A = n\sum\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2$

Definitional formulae

- Total variability – deviation of each observation from the grand mean: $SS_T = \sum\left(Y_{ij} - \bar{Y}_{..}\right)^2$
- Variability due to factor – deviation of factor group means from grand mean: $SS_A = n\sum\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2$
- Variability due to subjects – deviation of each subject's mean from the grand mean: $SS_S = a\sum\sum\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2$
- Error – changes (inconsistencies) in the effect of factor across subjects (TR x S interaction):

$$SS_{AxS} = \sum\sum\left(Y - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\right)^2$$ or $$SS_{AxS} = SS_T - SS_A - SS_S$$

computations

$$\Sigma X^2 = \mathbf{504}$$

$$\frac{(\Sigma X)^2}{N} = \mathbf{90^2 / 20 = 405}$$

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = \mathbf{504 - 405 = 99}$$

$$SS_{Ss} = \frac{\sum T_S^{\ 2}}{b} - \frac{(\Sigma X)^2}{N} = \mathbf{18^2 + 23^2 + 7^2 + 15^2 + 27^2 / 4 - 405 = 59}$$

$$SS_{TR} = \frac{\sum T_j^{\ 2}}{n} - \frac{(\Sigma X)^2}{N} = \mathbf{15^2 + 20^2 + 25^2 + 30^2 / 5 - 405 = 25}$$

$$SS_{error} = SS_{total} - SS_S - SS_{TR} = \mathbf{99 - 59 - 25 = 15}$$

degrees of freedom

number of subjects * number of conditions

$\boldsymbol{df_{total}}$ **=** nj-1 = N-1 = 19

Big N = Number of observations

$\boldsymbol{df_{s}}$ **=** n-1 = 4

$\boldsymbol{df_{tr}}$ **=** j–1 = 3

$\boldsymbol{df_{error}}$ **=** (n-1)(j-1) = 12

error *df* is different from between-subjects anova – because error is now interaction of subject factor x treatment factor

the summary table

Source	SS	df	MS	F
Between subjects (S)	59	4	14.75	
Treatment (TR)	**25**	**3**	**8.33**	**6.66***
Error	**15**	**12**	**1.25**	
Total	**99**	**19**		

*** *p* <.05** **$F_{crit}(3,12) = 3.49$**

MS_S = estimate of variance in DV attributable to **INDIVIDUAL DIFFERENCES** (averaged over treatment levels) **– but ignore this & don't report in write-up**

MS_{TR} = estimate of variance in DV attributable to **TREATMENT** (averaged over subjects)

MS_{Error} = **RESIDUAL:** estimate of variance in DV not attributable to **S or TR** (interaction - the change in the treatment effect across subjects = error)

assuming the data was obtained from a **between-subjects** design . . .

Source	SS	df	MS	F
Treatment (TR)	25	3	8.33	1.80
Error	74	16	4.63	
Total	99	19		

$F_{crit}(3,16) = 3.24$

in **between-subjects** designs, ***individual differences are inseparable from error***, hence contribute to the error term

in **within-subjects** designs it is possible to ***partial out (i.e., remove) individual differences*** from the error term

smaller error term → greater POWER ☺

a note on error terms…

- hand calculations in within-subjects anova are ***no different*** to those in between-subjects anova
  – only the error term (and df) changes

- in 1-way within-subjects the error term (and df) is the treatment x subjects interaction
  – $MS_{error} = MS_{TRxS}$

following up the main effect of treatment . .

Source	SS	df	MS	F
Treatment (TR)	25	3	8.33	6.66*
Error	15	12	1.25	
Total	40	19		

* p <.05 $F_{crit}(3,12) = 3.49$

in between-subjects anova, MS$_{error}$ is the term we would use to test <u>any</u> effect, including simple comparisons
[error = differences between subjects – expect within-cell variance is the same across conditions]

but ***within-subjects*** ANOVA we partition out and ignore the main effect of subjects and compute an error term estimating <u>inconsistency as subjects change over WS levels</u>

separate error terms:
following-up main effects

– We expect inconsistency in TR effect x subjects so in simple comparisons use only data for conditions involved in comparison & calculate **separate error terms** each time

	block 1	block2	block3	block 4	subj total
subject 1		3	6		**9**
subject 2		4	7		**11**
subject 3		2	1		**3**
subject 4		4	5		**9**
subject 5		7	6		**13**
block total		**20**	**25**		**45**
block mean		**4**	**5**		

B2 vs B3

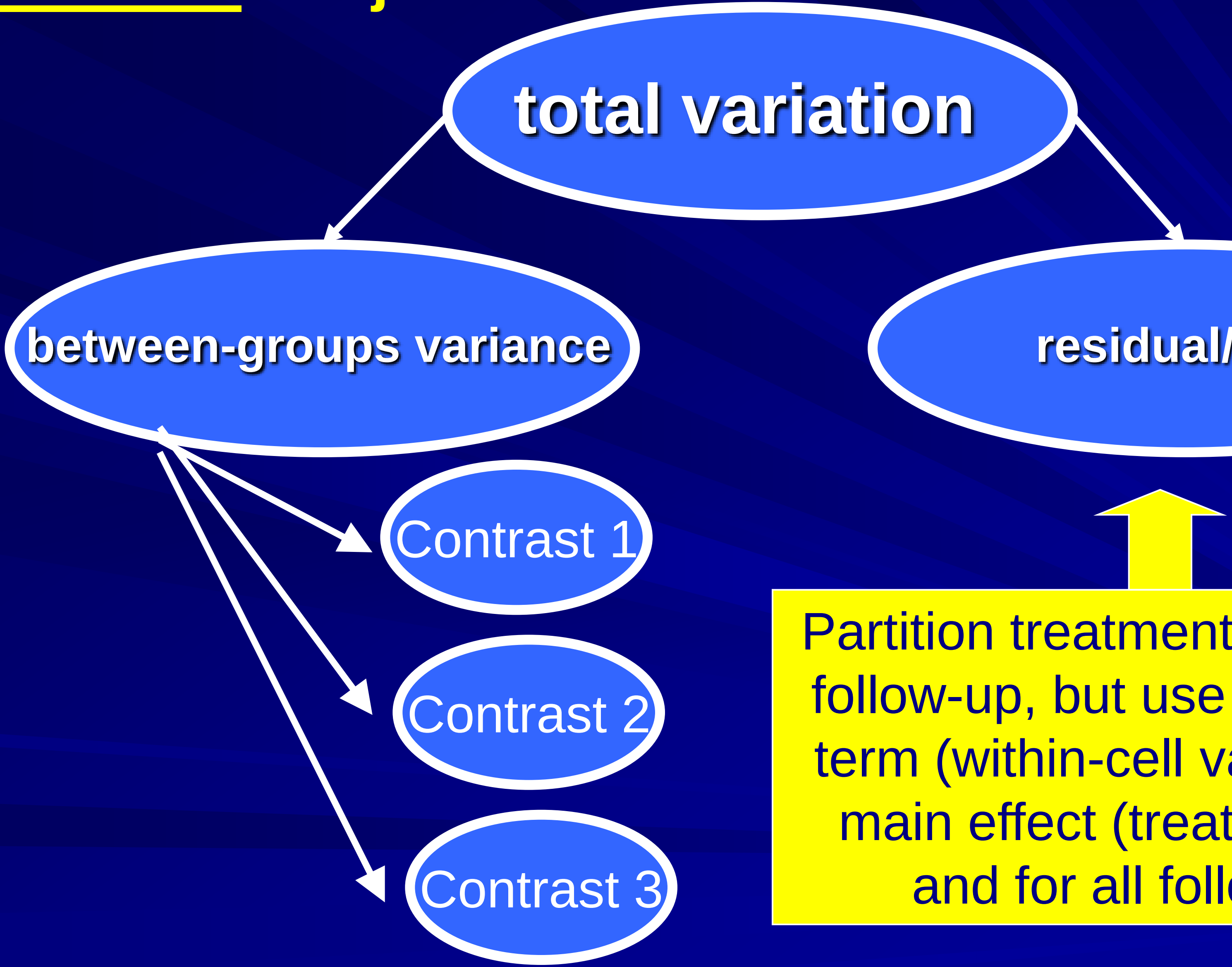
separate error terms:
following-up main effects

- We expect inconsistency in TR effect x subjects so in simple comparisons use only data for conditions involved in comparison & calculate **separate error terms** each time

	block 1	block2	block3	block 4	subj total
subject 1	4			5	**9**
subject 2	4			8	**12**
subject 3	1			3	**4**
subject 4	1			5	**6**
subject 5	5			9	**14**
block total	**15**			**30**	**45**
block mean	**3**			**6**	

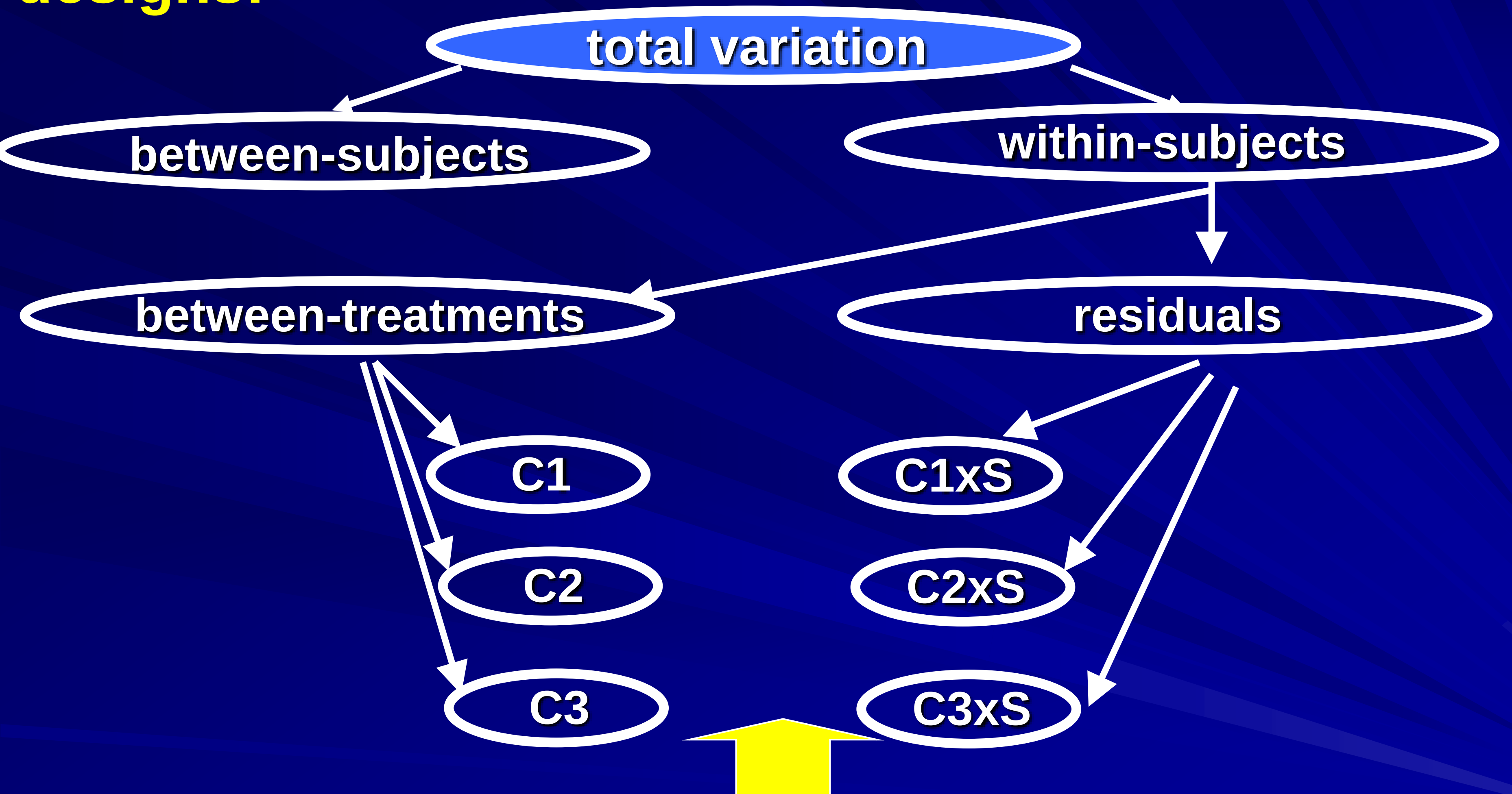
B1 vs B4

Simple comparisons in between-subjects anova:



Partition treatment variance to follow-up, but use same error term (within-cell variance) for main effect (treatment) test and for all follow-ups

Simple comparisons in RM designs:



Partition treatment variance and residual variance for follow-ups. Each contrast effect is tested against error term = C x S interaction

calculations
(contrast 1 only)

$\Sigma X^2 = \mathbf{241}$

$\dfrac{(\Sigma X)^2}{N} = \mathbf{45^2 / 10 = 202.5}$

$SS_{total_{comp}} = \Sigma X^2 - \dfrac{(\Sigma X)^2}{N} = \mathbf{241 - 202.5 = 38.5}$

$SS_{Ss_{comp}} = \dfrac{\sum T_S^{\,2}}{j} - \dfrac{(\Sigma X)^2}{N} = \mathbf{9^2 + 11^2 + 3^2 + 9^2 + 13^2 / 2 - 202.5 = 28}$

$\boldsymbol{SS_{contrast}} = \dfrac{\sum T_J^{\,2}}{n} - \dfrac{(\Sigma X)^2}{N} = \mathbf{20^2 + 25^2 / 5 - 202.5 = 2.5}$

$SS_{TR_{comp}xS} = SS_{total} - SS_S - SS_{contrast} = \mathbf{38.5 - 28 - 2.5 = 8}$

calculations
(contrast 1 only)

alternatively, use the formula from the earlier anova lectures:

$$SS_{contrast} = \frac{nL^2}{\Sigma a_j^2}$$

$$where \quad L = \sum a_j \overline{X}_j$$

$$= 4(1) + 5(-1) = -1$$

$$SS_{contrast} = \frac{5(-1^2)}{2}$$

$$= \mathbf{2.5}$$

summary table

these are the $SS_{contrasts}$ we can calculate in the same way as in between-subjects anova

df for comparison is same as usual (i.e., 1)

Source	SS	df	MS	F
B2 vs B3	2.5	1	2.5	**1.25**
Error	8	4	2	
B1 vs B4	22.5	1	22.5	**22.5***
Error	4	4	1	

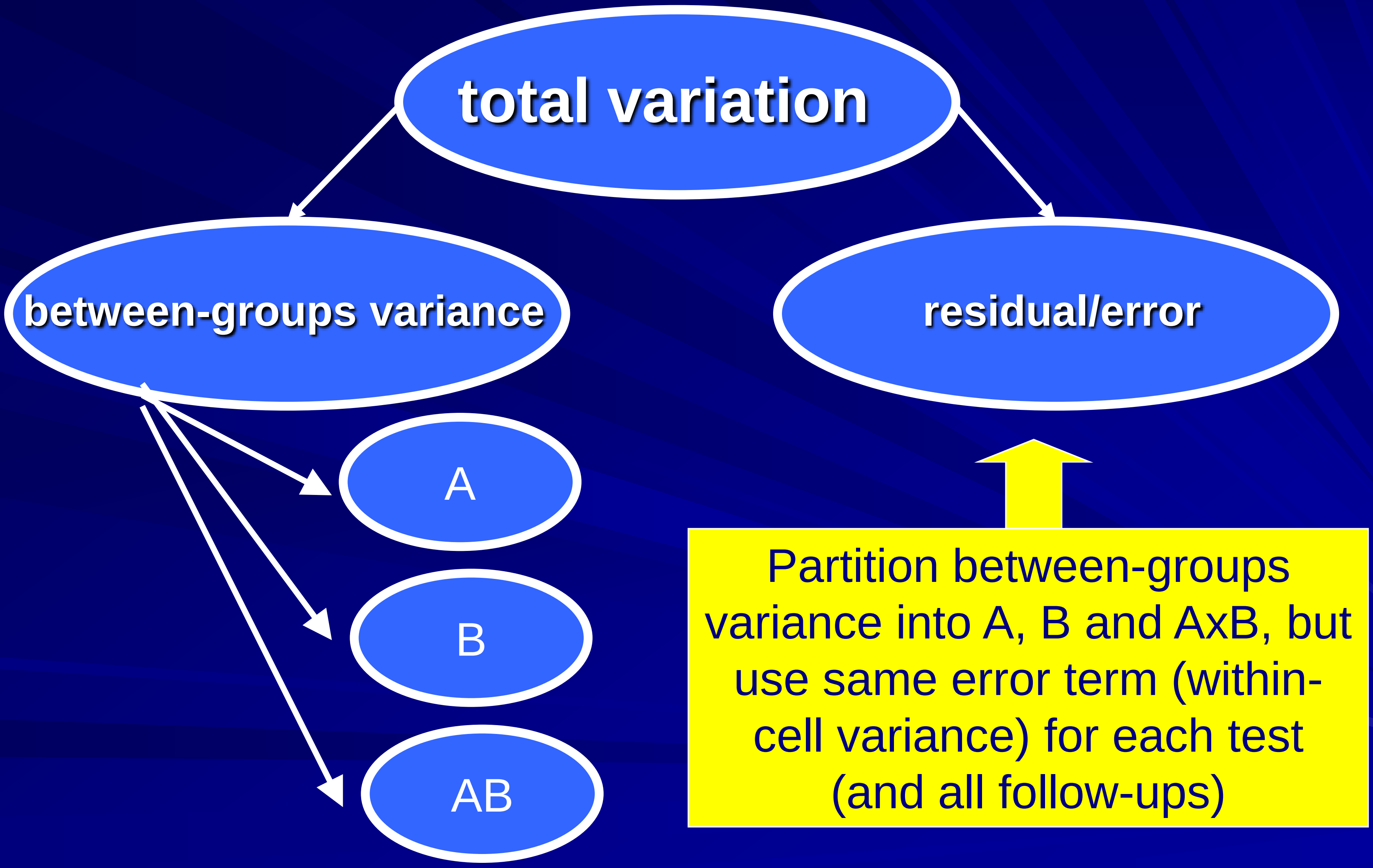
But SS_{SxL} terms we calculate separately for each within-subjects effect

$$df_{error} = (n-1)(j-1) = (5-1)(2-1) = 4$$

2-way within-subjects designs

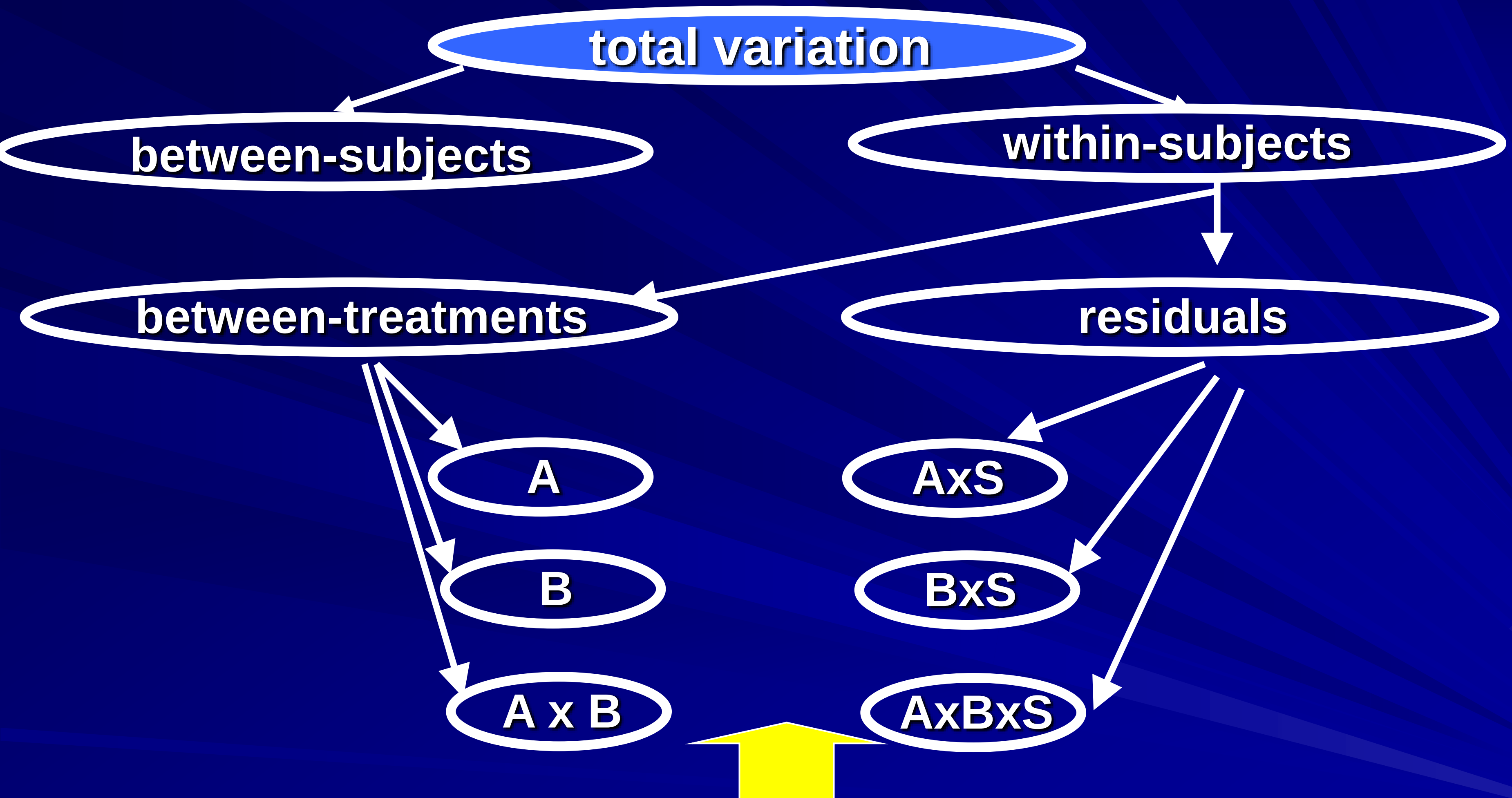
- **calculations are similar to a 2-way between-subjects ANOVA**
  - main effects for A and B are tested, as well as a AxB interaction
  - with a within-subjects design, each effect tested has a ***separate error term***
  - this error term simply corresponds to an ***interaction between the effect due to SUBJECTS, and the treatment effect***
    - **main effect of A → error term is MS_{AxS}**
    - **main effect of B → error term is MS_{BxS}**
    - **AxB interaction → error term is MS_{ABxS}**

2-way between-subjects anova:



Partition between-groups variance into A, B and AxB, but use same error term (within-cell variance) for each test (and all follow-ups)

2-way within-subjects anova:



Partition treatment variance and residual variance for each effect. Each effect is tested against error term = effect x S interaction

2-way within-subjects example

another learning study:

- 2 x 4 repeated-measures factorial design (n=4)
- first factor: phase
  - phase 1: no reinforcement (100 trials)
  - phase 2: reward for correct response (100 trials)
- second factor: block
  - each phase split into four blocks of 25
  - enables us to compare performance for trials later in each phase with trials early in each phase – thereby assessing *learning*
- DV = number of correct responses per block

Phase x Block repeated measures design [phase x block x subjects]

PBS Matrix

	p1				p2			
	b1	b2	b3	b4	b1	b2	b3	b4
subject 1	3	4	3	7	5	6	7	11
subject 2	6	8	9	12	10	12	15	18
subject 3	7	13	11	11	10	15	14	15
subject 4	0	3	6	6	5	7	9	11

PS matrix

	p1	p2	
subject 1	17	29	**46**
subject 2	35	55	**90**
subject 3	42	54	**96**
subject 4	15	32	**47**

BS matrix

	b1	b2	b3	b4
subject 1	8	10	10	18
subject 2	16	20	24	30
subject 3	17	28	25	26
subject 4	5	10	15	17

PB matrix

	b1	b2	b3	b4	
p1	16	28	29	36	**109**
p2	30	40	45	55	**170**
	46	**68**	**74**	**91**	***279***

calculations . . . .

$\Sigma X^2 = \mathbf{2995} \qquad \frac{(\Sigma X)^2}{N} = \mathbf{279^2 / 32 = 2432.53}$

$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = \mathbf{2995 - 2432.53} = \underline{\mathbf{562.47}}$

BS matrix	b1	b2	b3	b4	
subject 1	8	10	10	18	**46**
subject 2	16	20	24	30	**90**
subject 3	17	28	25	26	**96**
subject 4	5	10	15	17	**47**

BETWEEN SUBJECTS EFFECT:

$SS_S = \frac{\sum T_S^2}{pb} - \frac{(\Sigma X)^2}{N} = \mathbf{46^2 + 90^2 + 96^2 + 47^2 / 8 - 2432.53} = \underline{\mathbf{272.60}}$

WITHIN SUBJECTS EFFECTS:

$SS_P = \frac{\sum T_P^2}{nb} - \frac{(\Sigma X)^2}{N} = \mathbf{109^2 + 170^2 / 16 - 2432.53} = \underline{\mathbf{116.28}}$

$SS_B = \frac{\sum T_B^2}{np} - \frac{(\Sigma X)^2}{N} = \mathbf{46^2 + 68^2 + 74^2 + 91^2 / 8 - 2432.53} = \underline{\mathbf{129.60}}$

$SS_{cellsPB} = \frac{\sum T_{PB}^2}{n} - \frac{(\Sigma X)^2}{N} = \mathbf{16^2 + 30^2 + 28^2 + 40^2 + 29^2 + 45^2 + 36^2 + 55^2 / 4 - 2432.53}$

$= \mathbf{249.22}$

$SS_{PB} = SS_{cellsPB} - SS_P - SS_B = \mathbf{249.22 - 116.28 - 129.60} = \underline{\mathbf{3.34}}$

calculations .

PS matrix	p1	p2	
subject 1	17	29	**46**
subject 2	35	55	**90**
subject 3	42	54	**96**
subject 4	15	32	**47**

$\Sigma X^2 =$ **2995** $\quad \frac{(\Sigma X)^2}{N} = \mathbf{279^2 / 32 = 2432.53}$

$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = \mathbf{2995 - 2432.53} = \underline{\mathbf{562.47}}$

BETWEEN SUBJECTS EFFECT:

$SS_S = \frac{\sum T_S^2}{pb} - \frac{(\Sigma X)^2}{N} = \mathbf{46^2 + 90^2 + 96^2 + 47^2 / 8 - 2432.53} = \underline{\mathbf{272.60}}$

WITHIN SUBJECTS EFFECTS:

$SS_P = \frac{\sum T_P^2}{nb} - \frac{(\Sigma X)^2}{N} = \mathbf{109^2 + 170^2 / 16 - 2432.53} = \underline{\mathbf{116.28}}$

$SS_B = \frac{\sum T_B^2}{np} - \frac{(\Sigma X)^2}{N} = \mathbf{46^2 + 68^2 + 74^2 + 91^2 / 8 - 2432.53} = \underline{\mathbf{129.60}}$

$SS_{cellsPB} = \frac{\sum T_{PB}^2}{n} - \frac{(\Sigma X)^2}{N} = \mathbf{16^2 + 30^2 + 28^2 + 40^2 + 29^2 + 45^2 + 36^2 + 55^2 / 4 - 2432.53}$

$= \mathbf{249.22}$

$SS_{PB} = SS_{cellsPB} - SS_P - SS_B = \mathbf{249.22 - 116.28 - 129.60} = \underline{\mathbf{3.34}}$

calculations . . . .

PB matrix

	b1	b2	b3	b4	
p1	16	28	29	36	**109**
p2	30	40	45	55	**170**
	46	**68**	**74**	**91**	***279***

$$\Sigma X^2 = \mathbf{2995} \qquad \frac{(\Sigma X)^2}{N} = \mathbf{279^2 / 32 = 2432.53}$$

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = \mathbf{2995 - 2432.53} = \underline{\mathbf{562.47}}$$

BETWEEN SUBJECTS EFFECT:

$$SS_S = \frac{\sum T_S^2}{pb} - \frac{(\Sigma X)^2}{N} = \mathbf{46^2 + 90^2 + 96^2 + 47^2 / 8 - 2432.53} = \underline{\mathbf{272.60}}$$

WITHIN SUBJECTS EFFECTS:

$$SS_P = \frac{\sum T_P^2}{nb} - \frac{(\Sigma X)^2}{N} = \mathbf{109^2 + 170^2 / 16 - 2432.53} = \underline{\mathbf{116.28}}$$

$$SS_B = \frac{\sum T_B^2}{np} - \frac{(\Sigma X)^2}{N} = \mathbf{46^2 + 68^2 + 74^2 + 91^2 / 8 - 2432.53} = \underline{\mathbf{129.60}}$$

$$SS_{cellsPB} = \frac{\sum T_{PB}^2}{n} - \frac{(\Sigma X)^2}{N} = \mathbf{16^2 + 30^2 + 28^2 + 40^2 + 29^2 + 45^2 + 36^2 + 55^2 / 4 - 2432.53}$$

$$= \mathbf{249.22}$$

$$SS_{PB} = SS_{cellsPB} - SS_P - SS_B = \mathbf{249.22 - 116.28 - 129.60} = \underline{\mathbf{3.34}}$$

calculations . . . .

PS matrix

	p1	p2	
subject 1	17	29	**46**
subject 2	35	55	**90**
subject 3	42	54	**96**
subject 4	15	32	**47**

ERROR TERM (P):

$$SS_{cellsPxS} = \frac{\sum T_{PS}^{2}}{b} - \frac{\sum X^{2}}{N} = \mathbf{17^2 + 35^2 + \ldots + 54^2 + 32^2 / 4 - 2432.53} = \mathbf{394.72}$$

$$SS_{PxS} = SS_{cellsPxS} - SS_P - SS_S = \mathbf{394.72 - 116.28 - 272.60} = \underline{\mathbf{5.84}}$$

ERROR TERM (B):

$$SS_{cellsBxS} = \frac{\sum T_{BS}^{2}}{p} - \frac{\sum X^{2}}{N} = \mathbf{8^2 + 16^2 + \ldots + 26^2 + 17^2 / 2 - 2432.53} = \mathbf{433.97}$$

$$SS_{BxS} = SS_{cellsBxS} - SS_B - SS_S = \mathbf{433.97 - 129.6 - 272.60} = \underline{\mathbf{31.77}}$$

ERROR TERM (AB):

$$SS_{PBxS} = SS_{total} - SS_S - SS_P - SS_B - SS_{PB} - SS_{PxS} - SS_{BxS}$$

$$= 562.47 - 272.60 - 116.28 - 129.60 - 3.34 - 5.84 - 31.77 = \underline{\mathbf{3.04}}$$

NB how unlike regular between subjects ANOVA need to calculate a new error term (factor x subject) for each F test

(you'll find SS_P and SS_S on previous slide)

summary table . . .

Source	*SS*	*df*	*MS*	*F*
Between subjects	272.6	3	90.867	
P	116.28	1	116.28	**59.63***
PxS	5.84	3	1.95	
B	129.6	3	43.20	**12.24***
BxS	31.77	9	3.53	
PB	3.34	3	1.11	**3.26**
PBxS	3.04	9	0.34	

Critical $F(1,3) =$ 10.13

Critical $F(3,9) =$ 3.86

following up main effects

- as with one-way repeated measures designs, use of error term for effect (e.g., $\boldsymbol{MS_{BxS}}$) is not appropriate for follow-up comparisons
- a **separate error term** must be calculated for each comparison undertaken($\boldsymbol{MS_{B_{COMP}xS}}$)

Source	*SS*	*df*	*MS*	*F*
B_{COMP}	18.06	1	18.06	**6.62**
B_{COMPxS}	8.19	3	2.73	
Critical $F(1,3)$ =	10.13			

following up interactions . . .

- again, separate error terms must be used for each effect tested
  - ***simple effects***
    - error term is $\boldsymbol{MS}_{\boldsymbol{A\ at\ B1xS}}$
    - the interaction between the *A* treatment and *subjects*, at *B1*
  - ***simple comparisons***
    - error term is $\boldsymbol{MS}_{\boldsymbol{A_{COMP}\ at\ B1xS}}$
    - interaction between the *A* treatment (only the data contributing to the comparison, A_{COMP}), and *subjects*, at *B1*

2 approaches to within-subjects designs

- ***mixed-model* approach**
  - what we have been doing with hand calculations
  - treatment is a ***fixed*** factor, subjects is a ***random*** factor
    - Fixed factor: You chose the levels of the IV.
      - You have sampled all the levels of the IV or
      - You have selected particular levels based on a theoretical reason
    - Random factor: The levels of the IV are chosen at random
    - Random factors have different error terms: all ANOVA we have done to date has assumed the IVs are fixed. For most of you, the subject factor is the only random factor you will ever meet (be grateful). ☺ You can read up on random factor ANOVA models in advanced textbooks if you need to (e.g., as a postgrad).
  - powerful when assumptions are met
  - mathematically user-friendly
    - just like a factorial anova
  - restrictive assumptions, but adjustments available if they are violated
- ***multivariate* approach**…which we will discuss briefly later

assumptions of mixed-model approach

- **not dissimilar to between-subjects assumptions:**
  1. sample is ***randomly drawn*** from population
  2. DV scores are ***normally distributed*** in the population
  3. ***compound symmetry***
     - homogeneity of variances in levels of repeated-measures factor
     - homogeneity of covariances (equal correlations/covariances between pairs of levels)

compound symmetry

- the variance-covariance matrix:

$$\Sigma = \begin{array}{c} \\ T1 \\ T2 \\ T3 \end{array} \begin{array}{c} \begin{array}{ccc} T1 & T2 & T3 \end{array} \\ \begin{bmatrix} 158.92 & 163.33 & 163.00 \\ 163.33 & 172.67 & 170.67 \\ 163.00 & 170.67 & 170.00 \end{bmatrix} \end{array}$$

compound symmetry

- the variance-covariance matrix:

$$\Sigma = \begin{array}{c|ccc} & T1 & T2 & T3 \\ \hline T1 & 158.92 & 163.33 & 163.00 \\ T2 & 163.33 & 172.67 & 170.67 \\ T3 & 163.00 & 170.67 & 170.00 \end{array}$$

compound symmetry requires that variances are roughly equal (homogeneity of variance)

compound symmetry

- the variance-covariance matrix:

$$\Sigma = \begin{array}{c} \\ T1 \\ T2 \\ T3 \end{array} \begin{array}{ccc} T1 & T2 & T3 \\ \end{array}$$

$$\Sigma = \begin{array}{cccc} & T1 & T2 & T3 \\ T1 & 158.92 & 163.33 & 163.00 \\ T2 & 163.33 & 172.67 & 170.67 \\ T3 & 163.00 & 170.67 & 170.00 \end{array}$$

compound symmetry requires that covariances are roughly equal (homogeneity of covariance)

Mauchly’s test of sphericity

- **compound symmetry is a very restrictive assumption – often violated**
- ***sphericity* is a more broad and less restrictive assumption**
- **SPSS – Mauchley’s test of sphericity**
  - examines overall structure of covariance matrix
  - determines whether values in the main diagonal (variances) are roughly equal, and if values in the off-diagonal are roughly equal (covariances)
  - evaluated as χ^2 – if significant, sphericity assumption is violated
  - **not a robust test *<u>AT ALL</u>* – *very commonly fail to find Mauchley’s sphericity is sig even when violations of sphericity are present in the data***

violations of sphericity

when sphericity doesn't matter

- in *between-subjects designs*, because treatments are unrelated (different subjects in different treatments)
  - the assumption of homogeneity of variance still matters though
- when within-subject factors have *two levels*, because only one estimate of covariance can be computed

when it does matter

- in *all other within-subjects designs*
- when the sphericity assumption is violated, F-ratios are *positively biased*
  - **critical values of *F*** [based on df a – 1, (a – 1)(n – 1)] are ***too small***
  - therefore, probability of type-1 error increases

adjustments to degrees of freedom

- Best to assume that have a problem and make adjustment proactively – change critical *F* by adjusting degrees of freedom
- **epsilon (*ε*) adjustments**
  - epsilon is simply a value by which the degrees of freedom for the test of F-ratio is multiplied
  - equal to 1 when sphericity assumption is met (hence no adjustment), and < 1 when assumption is violated
  - the lower the epsilon value (further from 1), the more conservative the test becomes

different types of epsilon

- **Lower-bound epsilon**
  - Act as if have only 2 treatment levels with maximal heterogeneity
  - used for conditions of maximal heterogeneity, or worst-case violation of sphericity → often too conservative

- **Greenhouse-Geisser epsilon**
  - size of ε depends on degree to which sphericity is violated
  - **$1 \geq \varepsilon \geq 1/(k\text{-}1)$** : varies between 1 (sphericity intact) and lower-bound epsilon (worst-case violation)
  - ***generally recommended – not too stringent, not too lax***

different types of epsilon

- **Huynh-Feldt epsilon**
  - an adjustment applied to the GG-epsilon
  - often results in epsilon exceeding 1, in which case it is set to 1
  - used when “true value” of epsilon is believed to be $\geq .75$

spss output from our previous example

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Epsilon[a] Huynh-Feldt	Epsilon[a] Lower-bound
PHASE	1.000	.000	0	.	1.000	1.000	1.000
BLOCK	.111	3.785	5	.634	.587	1.000	.333
PHASE * BLOCK	.000	.	5	.	.348	.370	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed Tests of Within-Subjects Effects table.

b.

Design: Intercept
Within Subjects Design: PHASE+BLOCK+PHASE*BLOCK

no test for effects involving phase – only 2 levels

test for block is not significant (sphericity not violated) but ***we aren't going to trust it!***

spss output from our previous example

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
PHASE	1.000	.000	0	.	1.000	1.000	1.000
BLOCK	.111	3.785	5	.634	.587	1.000	.333
PHASE * BLOCK	.000	.	5	.	.348	.370	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed Tests of Within-Subjects Effects table.

b.

Design: Intercept
Within Subjects Design: PHASE+BLOCK+PHASE*BLOCK

compare the epsilon values…

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
PHASE	Sphericity Assumed	116.281	1	116.281	59.695	.005
	Greenhouse-Geisser	116.281	1.000	116.281	59.695	.005
	Huynh-Feldt	116.281	1.000	116.281	59.695	.005
	Lower-bound	116.281	1.000	116.281	59.695	.005
Error(PHASE)	Sphericity Assumed	5.844	3	1.948		
	Greenhouse-Geisser	5.844	3.000	1.948		
	Huynh-Feldt	5.844	3.000	1.948		
	Lower-bound	5.844	3.000	1.948		
BLOCK	Sphericity Assumed	129.594	3	43.198	12.233	.002
	Greenhouse-Geisser	129.594	1.760	73.621	12.233	.011
	Huynh-Feldt	129.594	3.000	43.198	12.233	.002
	Lower-bound	129.594	1.000	129.594	12.233	.040
Error(BLOCK)	Sphericity Assumed	31.781	9	3.531		
	Greenhouse-Geisser	31.781	5.281	6.018		
	Huynh-Feldt	31.781	9.000	3.531		
	Lower-bound	31.781	3.000	10.594		
PHASE * BLOCK	Sphericity Assumed	3.344	3	1.115	3.309	.071
	Greenhouse-Geisser	3.344	1.043	3.207	3.309	.163
	Huynh-Feldt	3.344	1.109	3.016	3.309	.159
	Lower-bound	3.344	1.000	3.344	3.309	.166
Error(PHASE*BLOCK)	Sphericity Assumed	3.031	9	.337		
	Greenhouse-Geisser	3.031	3.128	.969		
	Huynh-Feldt	3.031	3.326	.911		
	Lower-bound	3.031	3.000	1.010		

spss output from our previous example

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
BLOCK	Sphericity Assumed	129.594	3	43.198	12.233	.002
	Greenhouse-Geisser	129.594	1.760	73.621	12.233	.011
	Huynh-Feldt	129.594	3.000	43.198	12.233	.002
	Lower-bound	129.594	1.000	129.594	12.233	.040
Error(BLOCK)	Sphericity Assumed	31.781	9	3.531		
	Greenhouse-Geisser	31.781	5.281	6.018		
	Huynh-Feldt	31.781	9.000	3.531		
	Lower-bound	31.781	3.000	10.594		

sphericity assumed – i.e., no adjustment

this is what we based our degrees of freedom on before, i.e., b-1 = 4-1 = 3, (n-1)(b-1) = 3 x 3 = 9 → 3,9

spss output from our previous example

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
BLOCK	Sphericity Assumed	129.594	3	43.198	12.233	.002
	Greenhouse-Geisser	129.594	1.760	73.621	12.233	.011
	Huynh-Feldt	129.594	3.000	43.198	12.233	.002
	Lower-bound	129.594	1.000	129.594	12.233	.040
Error(BLOCK)	Sphericity Assumed	31.781	9	3.531		
	Greenhouse-Geisser	31.781	5.281	6.018		
	Huynh-Feldt	31.781	9.000	3.531		
	Lower-bound	31.781	3.000	10.594		

Lower-bound – for worst case heterogeneity

i.e., *df* = 1, b-1 – here we come close to concluding non-significance (which would probably be a type-2 error)

spss output from our previous example

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
BLOCK	Sphericity Assumed	129.594	3	43.198	12.233	.002
	Greenhouse-Geisser	129.594	1.760	73.621	12.233	.011
	Huynh-Feldt	129.594	3.000	43.198	12.233	.002
	Lower-bound	129.594	1.000	129.594	12.233	.040
Error(BLOCK)	Sphericity Assumed	31.781	9	3.531		
	Greenhouse-Geisser	31.781	5.281	6.018		
	Huynh-Feldt	31.781	9.000	3.531		
	Lower-bound	31.781	3.000	10.594		

Greenhouse-Geisser
adjustment does not change significance of result

spss output from our previous example

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
BLOCK	Sphericity Assumed	129.594	3	43.198	12.233	.002
	Greenhouse-Geisser	129.594	1.760	73.621	12.233	.011
	Huynh-Feldt	129.594	3.000	43.198	12.233	.002
	Lower-bound	129.594	1.000	129.594	12.233	.040
Error(BLOCK)	Sphericity Assumed	31.781	9	3.531		
	Greenhouse-Geisser	31.781	5.281	6.018		
	Huynh-Feldt	31.781	9.000	3.531		
	Lower-bound	31.781	3.000	10.594		

Huynh-Feldt – adjusts GG

no different to ‘sphericity assumed’ – indicates that $\varepsilon > 1$

Writing up…

Changes in participants' learning with practice and with or without reinforcement were explored in a 2 [phase] x 4 [Block] repeated measures ANOVA. In these analyses, the Huynh-Feldt correction was applied to the degrees of freedom, however the full degrees of freedom are reported here. Contrary to predictions, the interaction was not significant, $F(3,9) = 3.309$, $p = .159$, eta2 = ??. However, as hypothesised, participants learned more in the phase with reinforcement ($M = 42.5$; SD = ??) than in the phase without ($M = 27.25$; SD = ??), $F(1, 3) = 59.70$, $p = .005$, eta2 = ??. A main effect of Block, $F(3,9) = 12.23$, $p = .002$, eta2 = ??, was followed up with a series of contrasts. These revealed that …

multivariate approach

- **multivariate analysis of variance (manova)**
  - creates a ***linear composite*** of multiple DVs
  - In MANOVA approach to repeated measures designs, our repeated measures variable is treated as multiple DVs and combined / weighted to maximise the difference between levels of other variables (similar to the approach regression uses to combined multiple predictors)
    - multivariate tests – *Pillai's Trace, Hotelling's Trace, Wilk's Lambda, Roy's Largest Root*
    - does not require restrictive assumptions
  - more complex and less powerful

multivariate approach

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
PHASE	Pillai's Trace	.952	59.695[a]	1.000	3.000	.005
	Wilks' Lambda	.048	59.695[a]	1.000	3.000	.005
	Hotelling's Trace	19.898	59.695[a]	1.000	3.000	.005
	Roy's Largest Root	19.898	59.695[a]	1.000	3.000	.005
BLOCK	Pillai's Trace	.992	43.017[a]	3.000	1.000	.112
	Wilks' Lambda	.008	43.017[a]	3.000	1.000	.112
	Hotelling's Trace	129.050	43.017[a]	3.000	1.000	.112
	Roy's Largest Root	129.050	43.017[a]	3.000	1.000	.112
PHASE * BLOCK	Pillai's Trace	.990	102.333[a]	2.000	2.000	.010
	Wilks' Lambda	.010	102.333[a]	2.000	2.000	.010
	Hotelling's Trace	102.333	102.333[a]	2.000	2.000	.010
	Roy's Largest Root	102.333	102.333[a]	2.000	2.000	.010

a. Exact statistic

b.
Design: Intercept
Within Subjects Design: PHASE+BLOCK+PHASE*BLOCK

Take home message

- What is MANOVA doing?
  - Weighting the DV for each level of the repeated measures IV with coefficients (like what happens to scores for each IV in multiple regression) to create a predicted DV score that maximises differences across the levels of the IV
  - **Problem:** Instead of adapting model to observed DVs, selectively weight or discount DVs based on how they fit the model.
    - Atheoretical, over-capitalises on chance
- Don’t use MANOVA approach to repeated measures
- With repeated measures designs, report the mixed model Fs not the MANOVA statistics
- Usually report GG Fs to ensure adjustment for sphericity violations which are common (regardless of Mauchley’s test, which is too conservative and may not be sig. even when there are large violations)
- Personally I always use the GG or HF adjustment (HF can be more liberal) but report full df – this is common

pros and cons

advantages of within-subjects designs:

- more efficient
  - – *n* Ss in *j* treatments generate *nj* data points
  - – simplifies procedure
- more sensitive
  - – estimate individual differences (SSsubjects) and remove from error term

pros and cons

disadvantages of within-subjects designs:

- restrictive statistical assumptions
- sequencing effects:
  - learning, practice – improved later regardless of manipulation
  - Fatigue – deteriorating later regardless of manipulation
  - Habituation – insensitivity to later manipulations
  - Sensitisation – become more responsive to later manipulations
  - Contrast – previous treatment sets standard to which react
  - Adaptation – adjustment to previous manipulations changes reaction to later
  - Direct carry-over – learn something in previous that alters later
  - Etc!
- An essential methodological practice in RM designs is to counterbalance to reduce sequencing effects
  - i.e., half participants receive order A1 then A2; half receive A2 then A1
  - But can still get treatment x order interactions

most important points

- **in within subjects anova, the error term used for ANY effect is equal to *the interaction between that effect and the effect of subjects* (a random factor)**
  - **this applies to:**
    - main effects
      - follow-up (main) comparisons
    - interactions
      - simple effects
        - follow-up (simple) comparisons
- due to problems causes by lack of compound symmetry/sphericity, adjustments (such as Greenhouse-Geisser adjustment) to our degrees of freedom are needed -- unless we used the manova approach, which we shouldn’t, because it is inferior

In class next week:

- Mixed ANOVA

In the tutes:

- This week: Within-subjects and mixed designs
- Next week: Consult for A2

Readings :

- Howell
  - chapter 14
- Field
  - Chapter 11