Between Groups & Within-Groups ANOVA

- BG & WG ANOVA
  - Partitioning Variation
  - "making" F
  - "making" effect sizes
- Things that "influence" F
  - Confounding
  - Inflated within-condition variability
- Integrating "stats" & "methods"

ANOVA → ANalysis Of VAriance

Variance means "variation"

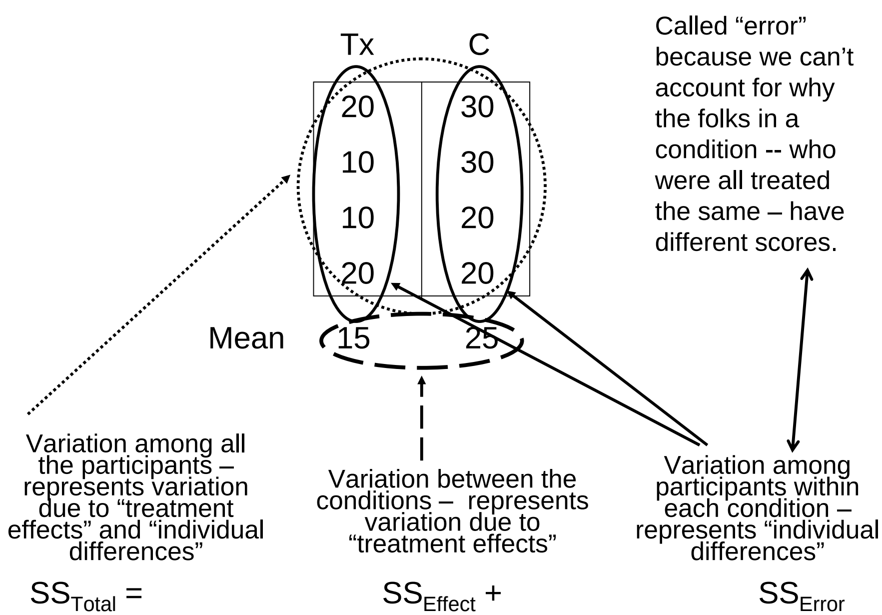
- Sum of Squares (SS) is the most common variation index
- SS stands for, "Sum of squared deviations between each of a set of values and the mean of those values"

$$SS = \sum \mathbf{(value - mean)^2}$$

So, Analysis Of Variance translates to "partitioning of SS"

In order to understand something about "how ANOVA works" we need to understand how BG and WG ANOVAs partition the SS differently and how F is constructed by each.

Variance partitioning for a BG design



Constructing BG F & r

$SS_{Total} = SS_{Effect} + SS_{Error}$

Mean Square is the SS converted to a "mean" → dividing it by "the number of things"

- $MS_{effect} = SS_{effect} / df_{effect}$ $df_{effect} = k - 1$ represents design size
- $MS_{error} = SS_{error} / df_{error}$ $df_{error} = \sum n - k$ represents sample size

F is the ratio of "effect variation" (mean difference) * "individual variation" (within condition differences)

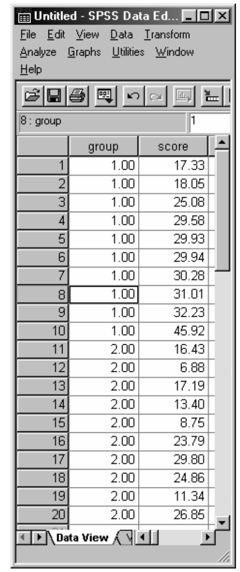
$$F = \frac{MS_{effect}}{MS_{error}}$$

r^2 = effect / (effect + error) ← conceptual formula

$= SS_{effect} / (SS_{effect} + SS_{error})$ ← definitional formula

$= F / (F + df_{error})$ ← computational forumla

An Example …



	group	score
1	1.00	17.33
2	1.00	18.05
3	1.00	25.08
4	1.00	29.58
5	1.00	29.93
6	1.00	29.94
7	1.00	30.28
8	1.00	31.01
9	1.00	32.23
10	1.00	45.92
11	2.00	16.43
12	2.00	6.88
13	2.00	17.19
14	2.00	13.40
15	2.00	8.75
16	2.00	23.79
17	2.00	29.80
18	2.00	24.86
19	2.00	11.34
20	2.00	26.85

Descriptives

SCORE

	N	Mean	Std. Deviation
1.00	10	28.9344	8.00000
2.00	10	17.9292	8.00000
Total	20	23.4318	9.61790

ANOVA

SCORE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	605.574	1	605.574	9.462	.007
Within Groups	1152.000	18	64.000		
Total	1757.574	19			

$SS_{total} = SS_{effect} + SS_{error}$

$1757.574 = 605.574 + 1152.000$

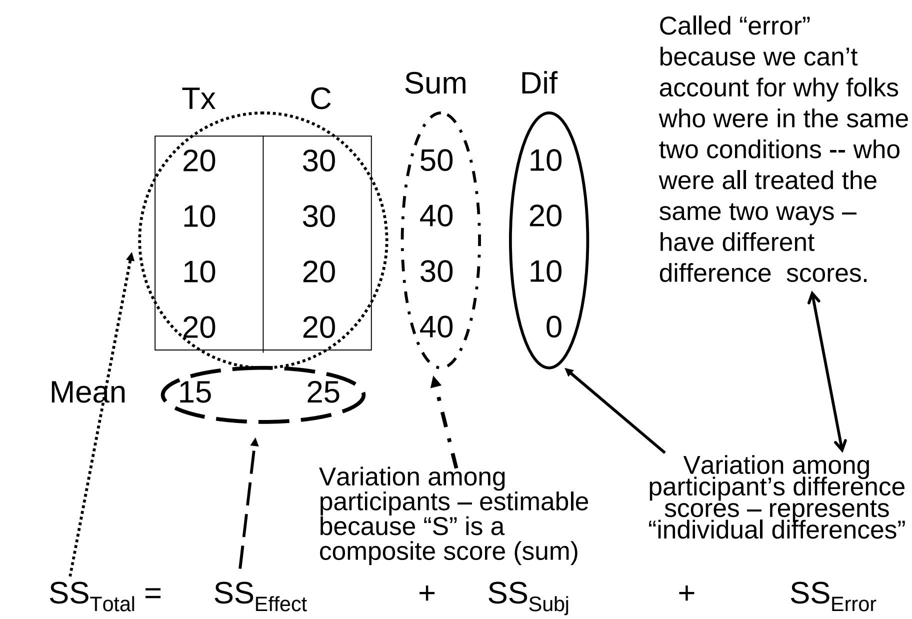
$r^2 = SS_{effect} / (SS_{effect} + SS_{error})$

$= 605.574 / (605.574 + 1152.000) = .34$

$r^2 = F / (F + df_{error})$

$= 9.462 / (9.462 + 18) = .34$

Variance partitioning for a WG design



Coinstructing WG F & r $SS_{Total} = SS_{Effect} + SS_{Subj} + SS_{Error}$

Mean Square is the SS converted to a "mean" → dividing it by "the number of things"

- $MS_{effect} = SS_{effect} / df_{effect}$ $df_{effect} = k - 1$ represents design size
- $MS_{error} = SS_{error} / df_{error}$ $df_{error} =$ **(k-1)(s-1)** represents sample size

F is the ratio of "effect variation" (mean difference) * "individual variation" (within condition differences). "Subject variation" is neither effect nor error, and is left out of the F calculation.

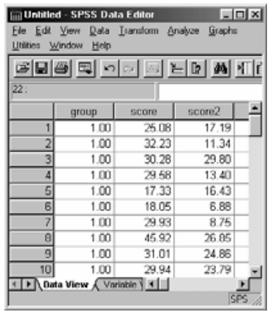
$$F = \frac{MS_{effect}}{MS_{error}}$$

r^2 = effect / (effect + error)   ← conceptual formula

$= SS_{effect} / (SS_{effect} + SS_{error})$ ← definitional formula

$= F / (F + df_{error})$ ← computational forumla

An Example …



	group	score	score2
1	1.00	25.08	17.19
2	1.00	32.23	11.34
3	1.00	30.28	29.80
4	1.00	29.58	13.40
5	1.00	17.33	16.43
6	1.00	18.05	6.88
7	1.00	29.93	8.75
8	1.00	45.92	26.05
9	1.00	31.01	24.86
10	1.00	29.94	23.79

Descriptive Statistics

	Mean	Std. Deviation	N
SCORE	28.9344	8.00000	10
SCORE2	17.9292	8.00000	10

Don't ever do this with real data !!!!!!

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	605.574	1	605.574	19.349	.002
Error(factor1)	Sphericity Assumed	281.676	9	31.297		

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	10980.981	1	10980.981	113.554	.000
Error	870.325	9	96.703		

$$SS_{total} = SS_{effect} + SS_{subj} + SS_{error}$$
$$1757.574 = 605.574 + 281.676 + 870.325$$

Professional statistician on a closed course.

$$r^2 = SS_{effect} / (SS_{effect} + SS_{error})$$
$$= 605.574 / (605.574 + 281.676) = .68$$

Do not try at home!

$$r^2 = F / (F + df_{error})$$
$$= 19.349 / (19.349 + 9) = .68$$

What happened?????   Same data. Same means & Std. Same total variance. Different F ???

BG ANOVA $SS_{Total} = SS_{Effect} + SS_{Error}$

WG ANOVA $SS_{Total} = SS_{Effect} + SS_{Subj} + SS_{Error}$

The variation that is called "error" for the BG ANOVA is divided between "subject" and "error" variation in the WG ANOVA.

Thus, the WG F is based on a smaller error term than the BG F → and so, the WG F is *generally* larger than the BG F.

It is important to note that the df_{error} also changes…

BG $df_{Total} = df_{Effect} + df_{Error}$

WG $df_{Total} = df_{Effect} + df_{Subj} + df_{Error}$

So, whether F-BG or F-WG is larger depends upon how much variation is due to subject variability

What happened????? Same data. Same means & Std.
Same total variance. Different r ???

r^2 = effect / (effect + error) ← conceptual formula

= SS_{effect} / (SS_{effect} + SS_{error}) ← definitional formula

= F / (F + df_{error}) ← computational forumla

The variation that is called "error" for the BG ANOVA is divided between "subject" and "error" variation in the WG ANOVA.

Thus, the WG r is based on a smaller error term than the BG r → and so, the WG r is generally larger than the BG r.

ANOVA was designed to analyze data from studies with…

- Samples that represent the target populations
- True Experimental designs
  - proper RA
  - well-controlled IV manipulation
- Good procedural standardization
- No confounds

ANOVA is a very simple statistical model that assumes there are few sources of variability in the data

BG $SS_{Total} = SS_{Effect} + SS_{Error}$

WG $SS_{Total} = SS_{Effect} + SS_{Subj} + SS_{Error}$

$$F = \frac{SS_{effect} / df_{effect}}{SS_{error} / df_{error}}$$

However, as we've discussed, most data we're asked to analyze are not from experimental designs.

2 other sources of variation we need to consider whenever we are working with quasi- or non-experiments are…

Between-condition procedural variation -- confounds

- any source of between-condition differences other than the IV
  - subject variable confounds (initial equiv)
  - procedural variable confounds (ongoing equiv.)
- influence the numerator of F

Within-condition procedural variation

- any source of within-condition differences other than "naturally occurring population individual differences"
  - misrepresentation of target population
  - within-condition procedural variation
- influence the denominator of F

With these variation sources in mind, here are somewhat "more realistic" models of variance partitioning for BG & WG ANOVA

IndDif → individual differences

BG $SS_{Total} = SS_{IV} + SS_{confound} + SS_{IndDif} + SS_{wcvar}$

WG $SS_{Total} = SS_{IV} + SS_{confound} + SS_{Subj} + SS_{IndDif} + SS_{wcvar}$

Later we will look at "statistical control" models that attempt to include these sources of variation.

For now it is enough to see how research designs & procedures produce these sources of variation and how they in turn will influence the computed F.

Imagine an educational study that compares the effects of two types of math instruction (IV) upon performance (% - DV)

Participants were randomly assigned to conditions, treated, then allowed to practice (Prac) as many problems as they wanted to before taking the DV-producing test

Control Grp			Exper. Grp		
	Prac	DV		Prac	DV
S1	5	75	S2	10	82
S3	5	74	S4	10	84
S5	10	78	S6	15	88
S7	10	79	S8	15	89

IV
- compare Ss 5&2 - 7&4

Confounding due to Prac
- mean prac dif between cond

WG variability inflated by Prac
- wg corrrelation of Prac & DV

Individual differences
- compare Ss 1&3, 5&7, 2&4, or 6&8

The problem is that when we apply the F & r formulas, they will …

- Ignore the confounding caused by differential practice between the groups and attribute all BG variation to the type of instruction (IV) → misestimating the effect
- Ignore the inflated within-condition variation caused by differential practice within the groups and attribute all WG variation to individual differences → misestimating the error
- As a result, the F & r values won't properly reflect the relationship between type of math instruction and performance ➔ more likely to make a statistical decision error !

$$\frac{SS_{IV} + SS_{confound}}{SS_{IndDif} + SS_{wcvar}} \rightarrow \frac{SS_{effect} / df_{effect}}{SS_{error} / df_{error}} = F$$

$$r = F / (F + df_{error})$$

How research design impacts F → integrating stats & methods!

The challenge is to recognize that our best model of the sources of variation in a BG design …

$$SS_{Total} = SS_{IV} + SS_{confound} + SS_{IndDif} + SS_{wcvar}$$

doesn't match the model we use to asses the effects of the IV …

$$F = \frac{SS_{effect} / df_{effect}}{SS_{error} / df_{error}}$$

Which ignores confounds and within-condition procedural variation.

We have to be able to recognize these additional sources of variance and understand how they influence our F-test results.

How research design impacts F → integrating stats & methods!

Sources of Variation that influence the numerator of F

SS_{Effect} → "bigger" manipulations produce larger mean difference between the conditions → larger F

$SS_{confound}$ → between group differences – other than the IV -- change mean difference → changing F

- if the confound "augments" the IV → F will be inflated
- if the confound "counters" the IV → F will be underestimated

Sample size → F will generally increase with sample size

How research design impacts F → integrating stats & methods!

Sources of Variation that influence the denominator of F

SS_{IndDif} → more heterogeneous populations have larger within-condition differences → smaller F

SS_{wcvar} → within-group differences – other than natural individual differences → smaller F

- could be "procedural" → procedural difference within conditions
- could be "bad sampling" → obtain a sample that is "more/less heterogeneous than the target population"

Hint…

Confounds often also increase within-condition variability – as practice did in the earlier example.