Sources of Variance & 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”
  \[ SS = \sum (value - mean)^2 \]

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

\[ SS_{Total} = SS_{Effect} + SS_{Error} \]

Variation among all the participants – represents variation due to “treatment effects” and “individual differences”
Variation between the conditions – represents variation due to “treatment effects”
Variation among participants within each condition – represents “individual differences”

Called “error” because we can’t account for why the folks in a condition -- who were all treated the same – have different scores.
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"

- $MS_{Effect} = \frac{SS_{Effect}}{df_{IV}}$
- $df_{Effect} = k - 1$ represents design size
- $MS_{Error} = \frac{SS_{Error}}{df_{Error}}$
- $df_{Error} = \sum n - k$ represents sample size

\[ F = \frac{MS_{IV}}{MS_{Error}} \]

$r^2 = \frac{Effect}{(Effect + Error)}$  \hspace{1cm} \text{conceptual formula}

$= \frac{SS_{Effect}}{(SS_{Effect} + SS_{Error})}$  \hspace{1cm} \text{definitional formula}

$= \frac{F}{(F + df_{error})}$  \hspace{1cm} \text{computational formula}

An Example …

Variance partitioning for a WG design

\[ SS_{Total} = SS_{Effect} + SS_{Error} \]

\[ 1757.574 = 605.574 + 1152.000 \]

\[ r^2 = \frac{SS_{Effect}}{(SS_{Effect} + SS_{Error})} \]

\[ = \frac{605.574}{(605.574 + 1152.000)} = .34 \]

\[ r^2 = \frac{F}{(F + df_{error})} \]

\[ = \frac{9.462}{(9.462 + 18)} = .34 \]
Constructing WG $F$ & $r$

$SS_{\text{Total}} = SS_{\text{Effect}} + SS_{\text{Subj}} + SS_{\text{Error}}$

Mean Square is the SS converted to a “mean” $\rightarrow$ dividing it by “the number of things”

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

$F$ is the ratio of “effect variation” (mean difference) $\times$ “individual variation” (within condition differences). “Subject variation” is neither effect nor error, and is left out of the $F$ calculation.

$r^2 = \frac{\text{effect}}{(\text{effect} + \text{error})}$ $\leftarrow$ conceptual formula

$r^2 = \frac{SS_{\text{effect}}}{(SS_{\text{effect}} + SS_{\text{error}})}$ $\leftarrow$ definitional formula

$r^2 = F / (F + df_{\text{error}})$ $\leftarrow$ computational formula

An Example ...

Descriptive Statistics

<table>
<thead>
<tr>
<th>SCORE</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.9344</td>
<td>8.00000</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>17.9292</td>
<td>8.00000</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects

Measure: MEASURE_1

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>605.574</td>
<td>1</td>
<td>605.574</td>
<td>19.349</td>
<td>.002</td>
</tr>
<tr>
<td>Error(factor1)</td>
<td>281.676</td>
<td>9</td>
<td>31.297</td>
<td>0</td>
<td>.000</td>
</tr>
</tbody>
</table>

Tests of Within-Subjects Effects

Measure: MEASURE_1

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor1</td>
<td>1088.581</td>
<td>1</td>
<td>1088.581</td>
<td>113.554</td>
<td>.000</td>
</tr>
</tbody>
</table>

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

Professional statistician on a closed course.

Do not try at home!

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

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 $F$ $\rightarrow$ and so, the WG $F$ is generally larger than the BG $F$.

It is important to note that the $df_{\text{error}}$ also changes...

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

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

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 ???

\[ r^2 = \frac{\text{Effect}}{\text{Effect} + \text{error}} \]  
- conceptual formula
\[ = \frac{\text{SS}_{\text{Effect}}}{\text{SS}_{\text{Effect}} + \text{SS}_{\text{error}}} \]  
- definitional formula
\[ = \frac{F}{F + df_{\text{error}}} \]  
- computational formula

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

\[ F = \frac{\text{SS}_{\text{Effect}}}{\text{df}_{\text{Effect}}} \]  
\[ \text{SS}_{\text{error}} / \text{df}_{\text{error}} \]

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 -- (sorry, no agreed-upon label)
- any source of within-condition differences other than “naturally occurring population individual differences”
  - subject variable diffs not representing the population
  - procedural variable influences on within cond variation
- influence the denominator of F
These considerations lead us to a somewhat more realistic model of the variation in the DV...

\[
SS_{Total} = SS_{IV} + SS_{confound} + SS_{indif} + SS_{wcvar}
\]

Sources of variability...

\(SS_{IV} \rightarrow IV \)  
\(SS_{confound} \rightarrow \) initial & ongoing equivalence problems  
\(SS_{indif} \rightarrow \) population individual differences  
\(SS_{wcvar} \rightarrow \) non-population individual differences

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

<table>
<thead>
<tr>
<th></th>
<th>Control Grp</th>
<th>Exper. Grp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prac DV</td>
<td>5 75</td>
<td>10 82</td>
</tr>
<tr>
<td>S1</td>
<td>5 75</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>5 74</td>
<td>10 84</td>
</tr>
<tr>
<td>S5</td>
<td>10 78</td>
<td>15 88</td>
</tr>
<tr>
<td>S7</td>
<td>10 79</td>
<td></td>
</tr>
</tbody>
</table>

Confounding due to Practice  
- mean prac dif between cond

WC variability due to Practice  
- wg correlation or prac & DV

IV  
- compare Ss 5&2 - 7&4

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

The problem is that the F-formula will...

\[
F = \frac{SS_{IV} + SS_{confound} / df_{IV}}{SS_{indif} + SS_{wcvar} / df_{error}}
\]

\[
≠ \frac{SS_{Effect} / df_{Effect}}{SS_{error} / df_{error}}
\]

r = \frac{F}{(F + df_{error})}

- Ignore the confounding caused by differential practice between the groups and attribute all BG variation to the type of instruction (IV) \( \rightarrow \) overestimating the SS\(_{effect}\)
- Ignore the changes in within-condition variation caused by differential practice within the groups and attribute all WG variation to individual differences \( \rightarrow \) overestimating SS\(_{error}\)
- As a result, the F & r values won't properly reflect the relationship between type of math instruction and performance \( \Rightarrow \) we will make statistical conclusion errors!
- We have 2 strategies:
  - Identify and procedurally control “inappropriate” sources
  - Include those sources in our statistical analyses
Both strategies require that we have a “richer” model for SS sources – and that we can apply that model to our research!

The “better” SS model we’ll use …

$$SS_{Total} = SS_{IV} + SS_{subcon} + SS_{proccon} + SS_{indif} + SS_{wcsubinf} + SS_{wcprocinf}$$

Sources of variability…

$$SS_{IV} \rightarrow IV$$

$$SS_{subcon} \rightarrow subject\ variable\ confounds\ (initial\ eq\ problems)$$

$$SS_{proccon} \rightarrow procedural\ variable\ confounds\ (ongoing\ eq\ pbms)$$

$$SS_{indif} \rightarrow population\ individual\ differences$$

$$SS_{wcsubinf} \rightarrow within\-condition\ subject\ variable\ influences$$

$$SS_{wcprocinf} \rightarrow within\-condition\ procedural\ variable\ influences$$

In order to apply this model, we have to understand:

1. How these variance sources relate to aspects of …
   - IV manipulation & Population selection
   - Initial equivalence & ongoing equivalence
   - Sampling & procedural standardization

2. How these variance sources befoul $$SS_{Effect}$$, $$SS_{error}$$ & F of the simple ANOVA model
   - $$SS_{IV}$$ numerator
   - $$SS$$ denominator

Sources of variability $\rightarrow$ their influence on $$SS_{Effect}$$, $$SS_{error}$$ & F…

$$SS_{IV} \rightarrow IV$$

“Bigger manipulations” produce larger mean differences

- larger $$SS_{effect} \rightarrow$$ larger numerator $\rightarrow$ larger F
- eg 0 v 50 practices instead of 0 v 25 practices
- eg those receiving therapy condition get therapy twice a week instead of once a week

“Smaller manipulations” produce smaller mean differences

- smaller $$SS_{effect} \rightarrow$$ smaller numerator $\rightarrow$ smaller F
- eg 0 v 25mg antidepressant instead of 0 v 75mg
- eg Big brother sessions monthly instead of each week
Sources of variability → their influence on $SS_{\text{Effect}}$, $SS_{\text{error}}$ & $F$...

$SS_{\text{indif}}$ → Individual differences

More homogeneous populations have smaller within-condition differences
- smaller $SS_{\text{error}}$ → smaller denominator → larger $F$
- eg studying one gender instead of both
- eg studying “4th graders” instead of “grade schoolers”

More heterogeneous populations have larger within-condition differences
- larger $SS_{\text{error}}$ → larger denominator → smaller $F$
- eg studying “young adults” instead of “college students”
- eg studying “self-reported” instead of “vetted” groups

Sources of variability → their influence on $SS_{\text{IV}}$, $SS_{\text{error}}$ & $F$...

The influence a confound has on $SS_{\text{IV}}$ & $F$ depends upon the “direction of the confounding” relative to “the direction of the IV”
- if the confound “augments” the IV → $SS_{\text{IV}}$ & $F$ will be inflated
- if the confound “offsets” the IV → $SS_{\text{IV}}$ & $F$ will be deflated

$SS_{\text{subcon}}$ → subject variable confounds (initial eq problems)

Augmenting confounds
- inflates $SS_{\text{Effect}}$ → inflates numerator → inflates $F$
- eg 4th graders in Tx group & 2nd graders is Cx group
- eg run Tx early in semester and Cx at end of semester

Offsetting confounds
- deflates $SS_{\text{Effect}}$ → deflates denominator → deflates $F$
- eg 2nd graders in Tx group & 4th graders is Cx group
- eg “better neighborhoods” in Cx than in Tx

$SS_{\text{proccon}}$ → procedural variable confounds (ongoing eq problems)

Augmenting confounds
- inflates $SS_{\text{Effect}}$ → inflates numerator → inflates $F$
- eg Tx condition is more interesting/fun for assistants to run
- eg Run the Tx on the newer, nicer equipment

Offsetting confounds
- deflates $SS_{\text{Effect}}$ → deflates denominator → deflates $F$
- eg less effort put into instructions for Cx than for Tx
- eg extra practice for touch condition than for visual cond
Sources of variability → their influence on \( SS_{\text{Effect}} \), \( SS_{\text{error}} \) & \( F \)…

The influence this has on \( SS_{\text{error}} \) & \( F \) depends upon the whether the sample is more or less variable than the target pop
  if “more variable” → \( SS_{\text{error}} \) is inflated & \( F \) will be deflated
  if “less variable” → \( SS_{\text{error}} \) is deflated & \( F \) will be inflated

\( SS_{\text{wsubinf}} \rightarrow \) within-condition subject variable influences

More variation in the sample than in the population
  • inflates \( SS_{\text{error}} \) → inflates denominator → deflates \( F \)
  • eg target pop is “3rd graders” - sample is 2nd, 3rd & 4th graders
  • eg target pop is “clinically anxious” – sample is “anxious”

Less variation in the sample than in the population
  • deflates \( SS_{\text{error}} \) → deflates denominator → inflates \( F \)
  • eg target pop is “grade schoolers” – sample is 4th graders
  • eg target pop is “young adults” – sample is college students

Sources of variability → their influence on \( SS_{\text{Effect}} \), \( SS_{\text{error}} \) & \( F \)…

The influence this has on \( SS_{\text{error}} \) & \( F \) depends upon the whether the procedure leads to more or less within condition variance in the sample data than would be in the population
  • if “more variable” → \( SS_{\text{error}} \) is inflated & \( F \) will be deflated
  • if “less variable” → \( SS_{\text{error}} \) is deflated & \( F \) will be inflated

\( SS_{\text{wprocinf}} \rightarrow \) within-condition procedural variable influences

More variation in the sample data than in the population
  • inflates \( SS_{\text{error}} \) → inflates denominator → deflates \( F \)
  • eg letting participants practice as much as they want
  • eg using multiple research stations

Less variation in the sample data than in the population
  • deflates \( SS_{\text{error}} \) → deflates denominator → inflates \( F \)
  • eg DV measures that have floor or ceiling effects
  • eg time allotments that produce floor or ceiling effects

Couple of important things to note !!!

\[ SS_{\text{Total}} = SS_{\text{IV}} + SS_{\text{subcon}} + SS_{\text{proccon}} + SS_{\text{Indir}} + SS_{\text{wsubinf}} + SS_{\text{wprocinf}} \]

Most variables that are confounds also inflate within condition variation!!!
  • like in the simple example earlier
  • “Participants were randomly assigned to conditions, treated, then allowed to practice as many problems as they wanted to before taking the DV-producing test”
Couple of important things to note !!!

\[ SS_{Total} = SS_{IV} + SS_{subcon} + SS_{proccon} + SS_{Indif} + SS_{wcsubinf} + SS_{wcpocinf} \]

There is an important difference between “picking a population that has more/less variation” and “sampling poorly so that your sample has more/less variation than the population”.

• if you intend to study 3rd, 4th & 5th graders instead of just 3rd graders, your \( SS_{error} \) will increase because your \( SS_{Indif} \) will be larger – but it is a choice, not a mistake!!!

• if intend to study 3rd stage cancer patients, but can’t find enough and use 2nd, 3rd & 4th stage patients instead, your \( SS_{error} \) will be inflated because of your \( SS_{wcsubinf} \) – and that is a mistake not a choice!!!