An experiment used a factorial experimental design to compare memory in older and younger subjects in four different conditions. Part of the ANOVA table is shown below. Fill in the missing parts.

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>A</td>
<td>B</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Age</td>
<td>4</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>Condition x Age</td>
<td>3</td>
<td>3</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>Error</td>
<td>i</td>
<td>40</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluate the null hypothesis for each main effect and the interaction (alpha = 0.05).

- $F_{critical}(\alpha=.05) = 4.08$
- $df=(1,40)$
- $df=(3,40): 2.84$
An experiment used a factorial experimental design to compare memory in older and younger subjects in four different conditions. Part of the ANOVA table is shown below. Fill in the missing parts.

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>0.047*</td>
<td>2.84</td>
</tr>
<tr>
<td>Age</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0.052</td>
<td>4.08</td>
</tr>
<tr>
<td>Condition x Age</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0.403</td>
</tr>
<tr>
<td>Error</td>
<td>40</td>
<td>40</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluate the null hypothesis for each main effect and the interaction (alpha = 0.05).

The main effect of Condition is significant, but the main effect of Age and the Condition x Age interaction are not.

Consider the four sets of data. If you analyzed each set with a 1-way ANOVA, which one would yield the highest value of eta-squared (η²)?

η² = SS<sub>Group</sub>/SS<sub>Total</sub>

η² is large in cases where variations among means is high relative to variation within groups.
1. An experiment is conducted to measure the effect of playing video games on visual attention. Forty adults who had never played video games were assigned randomly to 1 of 4 groups (n=10 per group). Subjects in groups A, B, and C played Call of Duty for 1, 2, or 3 hours each day for 10 days. Subjects in group D did not play any video games. At the end of 10 days, visual attention was measured in all subjects using a standardized test, and the scores were analyzed with a 1-way between-subjects ANOVA. The effect of Group was statistically significant (F(3,36) = 3.20, p<0.05). The experimenter concludes that playing video games improves visual attention. Assuming that the statistical result is true (i.e., it is not a Type I error), is the investigator's conclusion justified? Why or why not?

No the conclusion is not justified. The omnibus F test evaluates the null hypothesis that all group means are equal. A significant omnibus F test suggests that the group means are not all equal, but it does not indicate how the means differ. It is possible, for example, that the significant F test in the above example is due to the fact that visual attention was poorer in the groups that played video games.

2. Why are measures of association strength and/or effect size useful? In other words, why not just report F and p values when describing the results of an ANOVA?

p values (and to a lesser extent F values) are affected significantly by the size of the experiment (i.e., the number of observations being analyzed) as well as the magnitude of the difference. Measures of association strength and/or effect size are less sensitive to the size of the experiment and therefore provide a more accurate measure of the magnitude of the effect being studied.
3. What null hypotheses are being evaluated by the F tests in the following ANOVA table?

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Ages)</td>
<td>1</td>
<td>240.15</td>
<td>240.15</td>
<td>25.9***</td>
<td>.000</td>
</tr>
<tr>
<td>Gender (Gender)</td>
<td>4</td>
<td>194.94</td>
<td>48.735</td>
<td>47.1***</td>
<td>.000</td>
</tr>
<tr>
<td>Age×Gender</td>
<td>4</td>
<td>190.30</td>
<td>47.575</td>
<td>5.59**</td>
<td>.003</td>
</tr>
<tr>
<td>Error</td>
<td>90</td>
<td>713.30</td>
<td>8.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>2667.79</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

The F tests for the two main effects are used to evaluate the null hypothesis that the marginal means of the levels on a factor are equal. The F test for the A×C interaction evaluates the null hypothesis that the effect of one factor does not depend on the level of the other factor.
Main Effect of Age

- **SS\text{Age}**: variation associated with marginal row (age) means
- sum of deviations of marginal means around grand mean

\[ SS_{\text{Age}} = n \times c \times \left[ \Sigma (\bar{X}_{Ai} - \bar{X}_{gm})^2 \right] \]

Sum taken across i=2 Ages
n = Ss per cell
c = levels in Condition factor
nc = observations per row

\[ SS_{\text{Age}} = 10 \times 5 \times \left[ (10.06 - 11.61)^2 + (13.16 - 11.61)^2 \right] \]

10.06

Main Effect of Condition

- **SS\text{Condition}**: variation associated with marginal column (condition) means
- sum of deviations of marginal means around grand mean

\[ SS_{\text{Condition}} = n \times a \times \left[ \Sigma (\bar{X}_{Cj} - \bar{X}_{gm})^2 \right] \]

Sum taken across j=5 Conditions
n = Ss per cell
a = levels in Age factor
na = observations per column

13.16

Age x Condition Interaction

- **SS\text{Cells}**: sum of deviations of cell means around grand mean

\[ SS_{\text{Cells}} = n \times a \times \left[ \Sigma (\bar{X}_{ij} - \bar{X}_{gm})^2 \right] \]

Sum taken across all i x j cells

SS\text{Cells} reflects variation due to both main effects plus additional unique variation due to combination of age & condition
- SS\text{Cells} = SS\text{A} + SS\text{C} + SS\text{AC}
- SS\text{AC} = SS\text{Cells} - SS\text{A} - SS\text{C}

4. The F for Age is calculated by dividing MS\text{Age} by MS\text{Error}. Explain why this is a fair test of the null hypothesis.
4. The F for Age is calculated by dividing MS_{AGE} by MS_{Error}. Explain why this is a fair test of the null hypothesis.

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

**MS_{Error}** is an estimate of the population error variance. When the null hypothesis is true, MS_{AGE} also is an estimate of the population error variance. Therefore, when the null hypothesis is true the values of MS_{AGE} and MS_{Error} ought to be similar (and will be distributed as F (df_{AGE}, df_{Error})). When the null hypothesis is false, MS_{AGE} should be greater than MS_{Error}. Hence, [in the absence of additional information] rejecting the null hypothesis when F>>1 is a reasonable strategy.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Age)</td>
<td>1</td>
<td>24025</td>
<td>24025</td>
<td>29.99***</td>
<td>0.000</td>
</tr>
<tr>
<td>C (Condition)</td>
<td>4</td>
<td>1314.94</td>
<td>328.735</td>
<td>47.19***</td>
<td>0.000</td>
</tr>
<tr>
<td>AC</td>
<td>4</td>
<td>190.50</td>
<td>47.625</td>
<td>5.52**</td>
<td>0.003</td>
</tr>
<tr>
<td>Error</td>
<td>95</td>
<td>722.30</td>
<td>8.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>2667.79</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01

5. What are the 3 key assumptions about the data that must be true in order for our ANOVA to produce a valid F test?

The 3 key assumptions are:
1) normality: data within each group/cell are distributed normally
2) constant variance: the variance is constant across groups/cells
3) independence: the observations are independent

6. An experimenter conducted 2 t tests. Test 1, which compared scores in groups A1B1 and A1B2, was not significant. Test 2, which compared scores in groups A1B1 and A2B1, was significant. The investigator concluded that the effect of B (i.e., the b1-b2 difference) differed for a1 and a2. Is this conclusion justified? Explain.
6. An experimenter conducted 2 t-tests. Test 1, which compared scores in groups $A_1B_1$ and $A_1B_2$, was not significant. Test 2, which compared scores in groups $A_2B_1$ and $A_2B_2$, was significant. The investigator concluded that the effect of $B$ (i.e., the $b_1-b_2$ difference) differed for $a_1$ and $a_2$. Is this conclusion justified? Explain.

The conclusion is not justified. A proper way of testing the hypothesis that the $b_1-b_2$ difference depends on the level of $A$ is to assess the $AxB$ interaction. Separate t-tests (i.e., tests 1 & 2) do not properly assess the $AxB$ interaction.

7. The results obtained by factorial ANOVAs often lead investigators to analyze simple (main) effects. What are simple main effects and when should they be examined?
7. The results obtained by factorial ANOVA often lead investigators to analyze simple (main) effects. What are simple main effects and when should they be examined?

Simple (main) effects refer to the main effect of one factor (e.g., Age) at each level of another factor (e.g., Condition). Simple main effects are useful when the interaction between the two factors (Age x Condition) is significant, because the significant interaction implies that the effect of one factor depends on the level of the other factor.

Graphical representation of 2-way interactions

- Difference between a1 & a2 depends on level of B.
- Difference between b1 & b2 depends on level of A.

Significant 2-way interactions imply a significant deviation from parallelism.

8. Often we want to follow-up an ANOVA by using multiple comparisons to conduct more fine-grained analyses of group differences. Explain the potential problem(s) that arise when we evaluate group differences using multiple comparisons (e.g., Tukey HSD, Fisher’s LSD, multiple t-tests, etc.).

Multiple comparisons may significantly increase the probability of making a Type I error. Consider the case where we evaluate all 10 pairwise differences among 5 groups. If the null hypothesis is true in each case, and if the per-comparison alpha is 0.05, then the familywise alpha for the 10 comparisons is approximately 0.40.