

2x2 2-Factor Between Groups ANOVA with EMMEANS Follow-ups

The study examined the relationships of exam Review Attendance and Practice Difficulty with exam performance. Practice Difficulty was a 2-condition variable - practice problems were either easier than the exam problems (=1) or about the same difficulty as the exam problems (=2). Different sections of the course were randomly assigned to receive the two difficulty levels. The schedule showed the class meeting during which the exam review would occur & student's attendance was recorded (1= not attend, 2= attend). The dependent variable was performance on an examination.

Process:

There are a lot of steps to a complete analysis of a 2-way design. Different patterns of significant and non-significant effects will require different subsets of these. Here's a preview...

Initial Analysis

- Get descriptive means, plots & F-tests
- Determine what effects are significant
- Consider what main effects are likely to be interesting – based on the aggregations involved

2-way Interactions

- Get 2-way cell means & follow-up analyses to describe the 2-way interaction

Main Effects

- Get estimated marginal means & follow-up analyses to describe each main effect
- Why are the “Descriptive” and “Estimated” marginal means different ?

Initial Analysis

Get descriptive means, plots & F-tests

unianova TestPerf by AtndRev PractDif

```
/ method = sstype(3)
/ print descriptives
/ plot profile(PractDif * AtndRev)
/ design = PractDif AtndRev PractDif*AtndRev.
```

- ← lists DV “by” IVs
order determines left-to-right ordering of IVs in the Descriptive Statistics table
- ← corrects each effect for all other effects
- ← get descriptive cell and marginal means
- ← get plot of cell means (x-axis * “separate lines”)
- ← specify the design including the interaction that is automatically calculates from the IVs specified above)

Descriptive Statistics

Dependent Variable: TestPerf

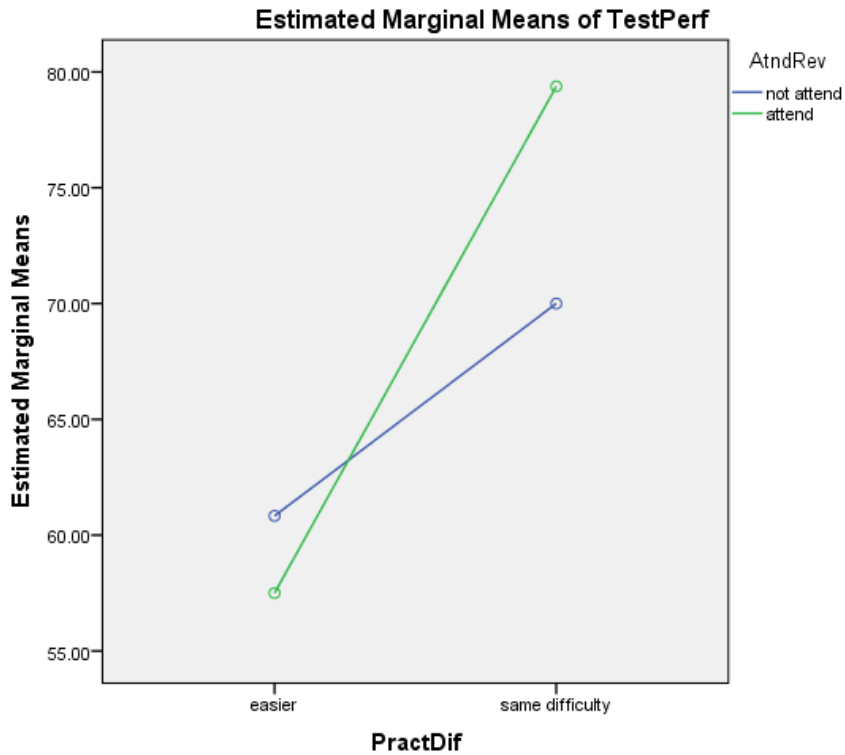
| AtndRev | PractDif | Mean | Std. Deviation | N |
|------------|-----------------|---------|----------------|----|
| not attend | easier | 60.8333 | 7.92961 | 12 |
| | same difficulty | 70.0000 | 7.55929 | 8 |
| | Total | 64.5000 | 8.87041 | 20 |
| attend | easier | 57.5000 | 7.07107 | 8 |
| | same difficulty | 79.3750 | 7.71902 | 16 |
| | Total | 72.0833 | 12.84664 | 24 |
| Total | easier | 59.5000 | 7.59155 | 20 |
| | same difficulty | 76.2500 | 8.75388 | 24 |
| | Total | 68.6364 | 11.73167 | 44 |

The “Descriptive Statistics” are the raw or “uncorrected” means.

The marginal means are weighted by the differential sizes of the cell means being aggregated.

For example, the marginal mean for the Easier PractDif is

$$((60.833 * 12) + (57.500 * 8)) / 20 = 59.500$$



From the means and the plots, it looks like performance after practice with same difficulty than with easier problems, and this effect is larger for those who attended the review.

Determine what effects are significant

Tests of Between-Subjects Effects

Dependent Variable: TestPerf

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|--------------------|-------------------------|----|-------------|----------|------|
| Corrected Model | 3582.765 ^a | 3 | 1194.255 | 20.455 | .000 |
| Intercept | 181055.373 | 1 | 181055.373 | 3101.038 | .000 |
| PractDif | 2434.320 | 1 | 2434.320 | 41.694 | .000 |
| AtndRev | 92.215 | 1 | 92.215 | 1.579 | .216 |
| PractDif * AtndRev | 408.004 | 1 | 408.004 | 6.988 | .012 |
| Error | 2335.417 | 40 | 58.385 | | |
| Total | 213200.000 | 44 | | | |
| Corrected Total | 5918.182 | 43 | | | |

a. R Squared = .605 (Adjusted R Squared = .576)

We have a significant effect for Practice Difficulty, no effect for Review Attendance and a significant interaction.

Consider what lower-order effects we will need to check for descriptive/misleading patterns

Because of the significant 2-way, the means patterns of each main effect will have to be carefully checked against the corresponding simple effects to determine if they are descriptive or misleading. Remember, this will have to be done whether the main effect is significant or not – main effect nulls can be misleading!

Consider what lower-order effects are likely to be interesting – based on the aggregations involved

PractDif

- These conditions are really pretty arbitrary.
- More importantly, it is unclear what population is represented by an average of those who attended and not attend the review session!
- So, this main effect is only likely to be interesting if that main effect is descriptive, and so, it describes the behavior of both those who did and did not attend the review.

Attend the Review

- This is a straightforward operationalization of a simple variable
- However, the marginal means are of dubious value, because the PractDif conditions are arbitrary, and so it is not clear what population would be represented by the aggregate of the easier and similar difficulty performances
- So, this main effect is only likely to be interesting if that main effect is descriptive, and so, it describes the behavior of both those who practiced with similarly difficult and easier materials.

Remember – non-significant lower-order effects that are involved in a significant higher order effect must be compared to the corresponding simple effects, to determine whether they are descriptive or misleading!!!

2-way Interaction

Pairwise Comparisons

You will usually want both sets of simple effects. One of those sets will be used to describe the pattern of the significant interaction. Each set will be used to determine if the corresponding main effect pattern is descriptive or misleading.

Select the set of simple effects that most directly addresses the research question or research hypothesis

The statement that, “We wanted to know if the relative difficulty of the practice material was related to test performance, and if this effect was different for those who did and did not attend the review session.” makes the selection of the simple effects to use to describe the interaction straightforward.

From this, we’ll want to focus on the simple effect of practice difficulty (easier, harder, similar) and then examine how this simple effect is different those who did and did not attend the review session.

Obtaining and describing the pairwise simple effects of Practice Difficulty for each level of Review Attendance

/ emmeans tables (AtnDRev by PractDif) compare (PractDif)

- ← this asks for the an analysis of the cell means for the 2-way interaction
- ← the order of the variables in parenthesis of the “table” command controls the display of the means
- ← the variable specified in the “compare” command tells which set of simple effects to test

Estimates

Dependent Variable: TestPerf

| AtnDRev | PractDif | Mean | Std. Error | 95% Confidence Interval | |
|------------|-----------------|--------|------------|-------------------------|-------------|
| | | | | Lower Bound | Upper Bound |
| not attend | easier | 60.833 | 2.206 | 56.375 | 65.291 |
| | same difficulty | 70.000 | 2.702 | 64.540 | 75.460 |
| attend | easier | 57.500 | 2.702 | 52.040 | 62.960 |
| | same difficulty | 79.375 | 1.910 | 75.514 | 83.236 |

Same cell means as in the Descriptives table above, but rearranged to match the tables command.

Univariate Tests

Dependent Variable: TestPerf

| AtnDRev | | Sum of Squares | df | Mean Square | F | Sig. |
|------------|----------|----------------|----|-------------|--------|------|
| not attend | Contrast | 403.333 | 1 | 403.333 | 6.908 | .012 |
| | Error | 2335.417 | 40 | 58.385 | | |
| attend | Contrast | 2552.083 | 1 | 2552.083 | 43.711 | .000 |
| | Error | 2335.417 | 40 | 58.385 | | |

The F-tests tell us that there is a significant simple effect of Practice Difficulty for each condition of Review Attendance.

With only 2 Practice Difficulty conditions, the pairwise comparisons are redundant with the F-tests.

$$\text{Not Attend } t^2 = (9.167 / 3.488)^2 = 6.908 = F$$

$$\text{Same } t^2 = (21.875 / 3.309)^2 = 43.711 = F$$

Each F tests the simple effects of PractDif within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise Comparisons

Dependent Variable: TestPerf

| AtnDRev | (I) PractDif | (J) PractDif | Mean Difference (I-J) | Std. Error | Sig. ^b | 95% Confidence Interval for Difference ^b | |
|------------|-----------------|-----------------|-----------------------|------------|-------------------|---|-------------|
| | | | | | | Lower Bound | Upper Bound |
| not attend | easier | same difficulty | -9.167 [*] | 3.488 | .012 | -16.215 | -2.118 |
| | same difficulty | easier | 9.167 [*] | 3.488 | .012 | 2.118 | 16.215 |
| attend | easier | same difficulty | -21.875 [*] | 3.309 | .000 | -28.562 | -15.188 |
| | same difficulty | easier | 21.875 [*] | 3.309 | .000 | 15.188 | 28.562 |

Based on estimated marginal means

*. The mean difference is significant at the .050 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

The pattern of the interaction is:

Not Attend

Easier < Same

Attend

Easier << Same

This interaction pattern allows us to anticipate that the main effect of Practice Difficulty will be **descriptive**

Obtaining and describing the pairwise simple effects of Review Attendance for each level of Practice Difficulty

/ emmeans tables (PractDif by AtndRev) compare (AtndRev)

- ← this asks for the an analysis of the cell means for the 2-way interaction
- ← the order of the variables in parenthesis of the “table” command controls the display of the means
- ← the variable specified in the “compare” command tells which set of simple effects to test

Estimates

Dependent Variable: TestPerf

| PractDif | AtndRev | Mean | Std. Error | 95% Confidence Interval | |
|-----------------|------------|--------|------------|-------------------------|-------------|
| | | | | Lower Bound | Upper Bound |
| easier | not attend | 60.833 | 2.206 | 56.375 | 65.291 |
| | attend | 57.500 | 2.702 | 52.040 | 62.960 |
| same difficulty | not attend | 70.000 | 2.702 | 64.540 | 75.460 |
| | attend | 79.375 | 1.910 | 75.514 | 83.236 |

The cell means will be the same as given in the “Descriptive Statistics” above.

The F-tests tell us that the simple effect of Review Attendance is significant Same but not Easier Practice.

Univariate Tests

Dependent Variable: TestPerf

| PractDif | | Sum of Squares | df | Mean Square | F | Sig. |
|-----------------|----------|----------------|----|-------------|-------|------|
| easier | Contrast | 53.333 | 1 | 53.333 | .913 | .345 |
| | Error | 2335.417 | 40 | 58.385 | | |
| same difficulty | Contrast | 468.750 | 1 | 468.750 | 8.029 | .007 |
| | Error | 2335.417 | 40 | 58.385 | | |

With only 2 Review Attendance conditions, the pairwise comparisons are redundant with the F-tests.

$$\text{Easier } t^2 = (3.333 / 3.488)^2 = .913 = F$$

$$\text{Same } t^2 = (9.375 / 3.309)^2 = 8.029 = F$$

Each F tests the simple effects of AtndRev within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise Comparisons

Dependent Variable: TestPerf

| PractDif | (I) AtndRev | (J) AtndRev | Mean Difference (I-J) | Std. Error | Sig. ^b | 95% Confidence Interval for Difference ^b | |
|-----------------|-------------|-------------|-----------------------|------------|-------------------|---|-------------|
| | | | | | | Lower Bound | Upper Bound |
| easier | not attend | attend | 3.333 | 3.488 | .345 | -3.715 | 10.382 |
| | attend | not attend | -3.333 | 3.488 | .345 | -10.382 | 3.715 |
| same difficulty | not attend | attend | -9.375 [*] | 3.309 | .007 | -16.062 | -2.688 |
| | attend | not attend | 9.375 [*] | 3.309 | .007 | 2.688 | 16.062 |

The pattern of the interaction is:

Easier Practice

Not Attend = Attend

Same Difficulty Practice

Not Attend < Attend

This interaction pattern allows us to anticipate that the main effect of Review Attendance will be **misleading**

Based on estimated marginal means

*. The mean difference is significant at the .050 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

An Alternative Analysis of Cell Means to Describe Simple Effects and the Interaction

This is a BG model, so all the F-tests and follow-up analyses are based on a single error term ($MSE=58.385$), though the Standard Errors of the follow-ups vary with sample size. Why care? Because the follow-up analyses are based on a t-test (that isn't shown in the output) that uses the standard error in the denominator.

So, depending on whether the cells being compared have larger or smaller sample sizes, the standard error can be larger (smaller ns) or smaller (larger ns), and the same cell mean difference can be significant for one comparison and not significant for another.

An alternative is to use this "full model error term" as the basis for computing an LSD value that is then used to compare any two cell means. This is an extension of the "homogeneity of variance" assumption that is made when we compute the ANOVA error term for BG models. That assumption is that it makes sense to combine the within-group variability from the different design cells, because they each represent a sample taken from different populations that all have the same variability, so the aggregate of them all is the best estimate of the variability of each. The extension in the "full model error term" approach is that since the best estimate is derived from using the full design sample, the significance test should be based on the df from all the participants.

Why do people who like this approach like it?

1. It is based on the same estimate of variability, but larger sample size, and, so, uses a smaller standard error than the pairwise error term approach. So, it provides a more powerful significance test, and more pairwise cell mean comparisons are significantly different using this approach (though the reverse can happen on occasion).
2. This approach allows the comparison of nonadjacent cells means. Sometimes, with larger designs, there is no easy way to get SPSS to provide this significance test, but the Computators will give us an LSDmmd that we can use to compare these means.

The dialog box is titled "LSD/HSD Minimum Mean Difference Computator". It has an orange background. It contains the following fields and options:

- Number of conditions in the design: 4
- Mean Square Error (MSe): 58.385
- error degrees of freedom: 40
- Design type: 2x2 BG Factorial Design (selected), k-Between Groups Design, k-Within-Groups Design
- Buttons: "Compute LSD & HSD minimum mean differences", "LSDmmd", "HSDmmd"
- Results: LSDmmd = 6.584, HSDmmd = 8.731

The spreadsheet shows the results of the computation in a table format:

| | A | B |
|----|---|---------------|
| 1 | LSD & HSD Minimum Mean Difference | |
| 2 | | |
| 3 | Enter k (number of conditions in the effect) => | 4 |
| 4 | Enter n (average number of data points upon which each mean is based - N/k) => | 11 |
| 5 | Enter MSe (Mean Square Error) => | 58.385 |
| 6 | Select dferror (error degrees of freedom - use "next smallest" if no exact match) => | 40 |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | LSD minimum mean difference = | 6.5814 |
| 11 | HSD minimum mean difference = | 8.7316 |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |

Another Alternative Analysis of Cell Means to Describe Simple Effects and the Interaction

Another approach to testing simple effects that shows up in many examples is to use the "split file" option in SPSS and run separate analyses for each partition of the design.

sort cases by PractDif .

← sorts the cases by the selection variable

temporary.
split file layered by PractDif .

← specify that split command will only apply to the next analysis command

uninova testperf by AtndRev
/design = AtndRev,

← splits the cases by the selection variable

← specify DV "by" IV (simple effect variable)

Tests of Between-Subjects Effects

Dependent Variable: TestPerf

| PractDif | Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|-----------------|-----------------|-------------------------|----|-------------|----------|------|
| easier | Corrected Model | 53.333 ^a | 1 | 53.333 | .922 | .350 |
| | Intercept | 67213.333 | 1 | 67213.333 | 1161.446 | .000 |
| | AtndRev | 53.333 | 1 | 53.333 | .922 | .350 |
| | Error | 1041.667 | 18 | 57.870 | | |
| | Total | 71900.000 | 20 | | | |
| | Corrected Total | 1095.000 | 19 | | | |
| same difficulty | Corrected Model | 468.750 ^b | 1 | 468.750 | 7.971 | .010 |
| | Intercept | 119002.083 | 1 | 119002.083 | 2023.610 | .000 |
| | AtndRev | 468.750 | 1 | 468.750 | 7.971 | .010 |
| | Error | 1293.750 | 22 | 58.807 | | |
| | Total | 141300.000 | 24 | | | |
| | Corrected Total | 1762.500 | 23 | | | |

The SS effect (AtndRev) are the same as from the EMMEANS analyses above. Each compares the same cell means

The SS Error are different from the EMMEANS analyses above. These are based on data from two cells, while EMMEANS were based on data from all four cells.

The df-error are different from the EMMEANS analyses above. These are based on n from the two cells being compared, while EMMEANS were based on n from all four cells.

The MSerror are different, because both the SSerror and df-error are different.

The F-values and p-values are different.

Here is the syntax to get the simple effects of practice difficulty for each review attendance condition.

sort cases by AtndRev.

temporary.
split file layered by AtndRev.

uninova testperf by PractDif
/design = PractDif.

Tests of Between-Subjects Effects

Dependent Variable: TestPerf

| AtndRev | Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|------------|-----------------|-------------------------|----|-------------|----------|------|
| not attend | Corrected Model | 403.333 ^a | 1 | 403.333 | 6.650 | .019 |
| | Intercept | 82163.333 | 1 | 82163.333 | 1354.754 | .000 |
| | PractDif | 403.333 | 1 | 403.333 | 6.650 | .019 |
| | Error | 1091.667 | 18 | 60.648 | | |
| | Total | 84700.000 | 20 | | | |
| | Corrected Total | 1495.000 | 19 | | | |
| attend | Corrected Model | 2552.083 ^b | 1 | 2552.083 | 45.142 | .000 |
| | Intercept | 99918.750 | 1 | 99918.750 | 1767.407 | .000 |
| | PractDif | 2552.083 | 1 | 2552.083 | 45.142 | .000 |
| | Error | 1243.750 | 22 | 56.534 | | |
| | Total | 128500.000 | 24 | | | |
| | Corrected Total | 3795.833 | 23 | | | |

a. R Squared = .270 (Adjusted R Squared = .229)

b. R Squared = .672 (Adjusted R Squared = .657)

Describing the Main Effect of Review Attendance

/ emmeans tables (AtndRev) compare (AtndRev)

Estimates

Dependent Variable: TestPerf

| AtndRev | Mean | Std. Error | 95% Confidence Interval | |
|------------|--------|------------|-------------------------|-------------|
| | | | Lower Bound | Upper Bound |
| not attend | 65.417 | 1.744 | 61.892 | 68.941 |
| attend | 68.438 | 1.654 | 65.094 | 71.781 |

Univariate Tests

Dependent Variable: TestPerf

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------|----------------|----|-------------|-------|------|
| Contrast | 92.215 | 1 | 92.215 | 1.579 | .216 |
| Error | 2335.417 | 40 | 58.385 | | |

The F tests the effect of AtndRev. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise Comparisons

Dependent Variable: TestPerf

| (I) AtndRev | (J) AtndRev | Mean Difference (I-J) | Std. Error | Sig. ^a | 95% Confidence Interval for Difference ^a | |
|-------------|-------------|-----------------------|------------|-------------------|---|-------------|
| | | | | | Lower Bound | Upper Bound |
| not attend | attend | -3.021 | 2.404 | .216 | -7.879 | 1.837 |
| attend | not attend | 3.021 | 2.404 | .216 | -1.837 | 7.879 |

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

You should notice that the means shown here are not the same as the marginal means from the "Descriptive Statistics" above (there 64.5 for Not Attend and 72.08 for Attend).

Also, the F-test for "AtndRev" in the ANOVA table above and shown below (which match) are not comparing the data means shown in the "Descriptive Statistics" above.

Because there are unequal sample sizes among the design conditions, the main effects and the interaction are all collinear (nonorthogonal, or correlated). Thus, like all other multivariate analyses using Type III SS, the model tests the unique contribution of each effect to the model, controlling for the other effects in the model.

So, in a factorial using Type III SS, the main effects being tested are different than the raw data marginal means, the same as a multiple regression including quantitative variables will test a regression weight that is not the same as the bivariate correlation between a variable and the criterion!

The overall or main effect for Review Attendance is:

$$\text{Attend} = \text{Not Attend}$$

This main effect must be communicated carefully, because it is potentially misleading.

However, we know from the pattern of the interaction that this is not descriptive for those in the Easier Practice condition.

Easier Practice Not Attend = Attend

Same Difficulty Practice Not Attend < Attend

Alternative Analyses of Marginal Means of Review Attendance

You will sometimes see folks obtain an LSDmmd value and use it to compare the marginal means, to test and describe the pattern of the main effect. That LSDmmd value will differ from the value used to compare cell means above, because the n for the marginal means is different from the n of the cell means.

The cell means spread the N = 44 participants across the 4 cells, for n = 11. The main effect of review attendance will spread those same N = 44 participants across just 2 conditions, for n = 22. Using n = 22 in the LSD Computator yields LSDmmd = 4.65. This would be used to compare the marginal means shown the "Descriptive Statistics" table: 64.5 for Not Attend vs. 72.08 for Attend, resulting is a significant difference between the marginal means!

Please note: Because this design is non-orthogonal (has unequal n), this analysis is importantly different from the approach taken using the emmeans analysis above!

- The emmeans analysis tested and described the effect of Review Attendance after correcting review attendance for the effect of practice difficulty and the interaction. That is why it compared the estimated marginal means – estimated from the model.
- This alternative approach compares raw marginal means without correction for the other effects in the model). The greater the non-orthogonality (unequal-n) of the design, the more these two analyses are likely to differ!
- Also remember to carefully consider what population, if any, is represented by the aggregated marginal means!

Which one to use? As you might expect, opinions differ, and the important things are to know what "your kind" expects and to be very clear which one you are presenting.

Describing the Main Effect of Practice Difficulty

/ emmeans tables (PractDif) compare (PractDif)

Estimates

Dependent Variable: TestPerf

| PractDif | Mean | Std. Error | 95% Confidence Interval | |
|-----------------|--------|------------|-------------------------|-------------|
| | | | Lower Bound | Upper Bound |
| easier | 59.167 | 1.744 | 55.642 | 62.691 |
| same difficulty | 74.688 | 1.654 | 71.344 | 78.031 |

Again, you should notice the means shown here are not the same as the marginal means from the "Descriptive Statistics" above (59.5 for Easier and 76.25 for Same).

Univariate Tests

Dependent Variable: TestPerf

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------|----------------|----|-------------|--------|------|
| Contrast | 2434.320 | 1 | 2434.320 | 41.694 | .000 |
| Error | 2335.417 | 40 | 58.385 | | |

The F-test matches what's in the ANOVA table above, because both are for the corrected or unique contribution of this main effect to the model. Said differently, both are testing the mean difference among the estimated marginal means of the groups, after correcting for the other effects in the model.

The F tests the effect of PractDif. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

The pairwise comparisons show the pattern of the main effect of Practice Difficulty to be:

Easier < Same

Pairwise Comparisons

Dependent Variable: TestPerf

| (I) PractDif | (J) PractDif | Mean Difference (I-J) | Std. Error | Sig. ^b | 95% Confidence Interval for Difference ^b | |
|-----------------|-----------------|-----------------------|------------|-------------------|---|-------------|
| | | | | | Lower Bound | Upper Bound |
| easier | same difficulty | -15.521 [*] | 2.404 | .000 | -20.379 | -10.663 |
| same difficulty | easier | 15.521 [*] | 2.404 | .000 | 10.663 | 20.379 |

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Also, we know from the pattern of the interaction that this is descriptive.

Not Attend Easier < Same

Attend Easier << Same

Alternative Analyses of Marginal Means of Practice Difficulty

The main effect of practice difficulty will spread the N = 44 participants across just 2 conditions, for n = 22, Using n = 22 in the LSD Computator yields LSDmmd = 4.65. This would be used to compare the marginal means shown the "Descriptive Statistics" table: 59.5 for Easier and 76.25 for Same Difficulty. Comparing these uncorrected marginal means, this way, we would reach the same conclusion than we obtained from the Type III SS analysis!

Please note: Because this design is non-orthogonal (has unequal n), this analysis is importantly different from the approach taken using the emmeans analysis above!

- The emmeans analysis tested and described the effect of practice difficulty after correcting practice difficulty for the effect of review attendance and the interaction. That is why it compared the estimated marginal means – estimated from the model.
- This approach compares the raw marginal means (without correction for the other effects in the model). The greater the non-orthogonality (unequal-n) of the design, the more these two analyses are likely to differ!
- Also remember to carefully consider what population, if any, is represented by the aggregated marginal means!

Which one to use? As you might expect, opinions differ, and the important things are to know what "your kind" expects and to be very clear which one you are presenting.