Psych 610 Handout #29, p. 1
Prof Colleen Moore

**Follow-up tests for Mixed designs** (some Grouping factors, and some within-subject factors)

An example with contrived data fictionalized from a paper by Surber & Gzesh,1984, Journal of Experimental Child Psychology. The task is a two arm balance scale with a constant weight on one arm. On each trial a weight is placed at some distance on one side. The participant says where to put the constant weight on the other side in order to make the two arms balance. The task can be done without feedback by fixing the arms. Inhelder and Piaget used this device in their studies of the development of logical reasoning, and so did Robert Siegler. The others used a choice task, whereas Surber & Gzesh had a continuous measure. The expectation is that there will be developmental differences in use of the Weight and Distance cues to make the scale balance. These developmental differences should show up as significant Grade x Wt or Grade x Dist interactions. Siegler claimed that children start by attending first to the #weights, then later learn to attend to distance, and finally combine them in the proper way. Siegler's choice task was insensitive to the difference between adding weight and distance versus multiplying. By using the continuous response measure, Surber & Gzesh had a better test of the multiplying model for subjective combination of weight and distance.

The grouping variables are Grade and Gender, and the trial factors are Weight(2) and Distance (3).

| | | | Weight 1 | | | Weight 2 | | |
|---|---|---|---|---|---|---|---|---|
| | | | D1 | D2 | D3 | D1 | D2 | D3 |
| P | grade | gend | | | | | | |
| 1 | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 5 |
| 2 | 1 | 1 | 1 | 2 | 2 | 3 | 4 | 6 |
| 3 | 1 | 1 | 3 | 4 | 4 | 2 | 5 | 5 |
| 4 | 1 | 2 | 2 | 3 | 2 | 4 | 6 | 7 |
| 5 | 1 | 2 | 1 | 3 | 3 | 4 | 5 | 4 |
| 6 | 1 | 2 | 2 | 2 | 3 | 5 | 6 | 7 |
| 7 | 2 | 1 | 3 | 3 | 5 | 4 | 7 | 8 |
| 8 | 2 | 1 | 1 | 1 | 3 | 3 | 8 | 7 |
| 9 | 2 | 1 | 2 | 2 | 4 | 3 | 5 | 7 |
| 10 | 2 | 2 | 3 | 4 | 5 | 3 | 5 | 7 |
| 11 | 2 | 2 | 1 | 2 | 4 | 2 | 5 | 6 |
| 12 | 2 | 2 | 1 | 3 | 5 | 2 | 6 | 8 |

**1. The overall anova** (from my 'legacy' DOS software, BMDP): Because Grade and Gender both begin with G, I named Grade "radeg" (moved the first letter to end).
We have sig main effects of Weight and Distance, and a Weight x Dist interaction. Then there are significant interactions of Grade x Gender x Weight, Distance x Grade.

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 1027.55556 | 1 | 1027.55556 | 435.20 | 0.0000 |
| radeg (grade) | 8.00000 | 1 | 8.00000 | 3.39 | 0.1029 |
| gend | 1.38889 | 1 | 1.38889 | 0.59 | 0.4651 |
| rg | 4.50000 | 1 | 4.50000 | 1.91 | 0.2048 |
| 1 ERROR | 18.88889 | 8 | 2.36111 | | |
| w | 102.72222 | 1 | 102.72222 | 71.12 | 0.0000 |
| wr | 0.05556 | 1 | 0.05556 | 0.04 | 0.8494 |
| wg | 0.00000 | 1 | 0.00000 | 0.00 | 1.0000 |
| wrg | 8.00000 | 1 | 8.00000 | 5.54 | 0.0464 |
| 2 ERROR | 11.55556 | 8 | 1.44444 | | |
| dist | 78.69444 | 2 | 39.34722 | 92.89 | 0.0000 |
| dr | 9.25000 | 2 | 4.62500 | 10.92 | 0.0010 |
| dg | 0.19444 | 2 | 0.09722 | 0.23 | 0.7975 |
| drg | 1.75000 | 2 | 0.87500 | 2.07 | 0.1592 |
| 3 ERROR | 6.77778 | 16 | 0.42361 | | |
| wd | 10.02778 | 2 | 5.01389 | 13.13 | 0.0004 |
| wdr | 2.69444 | 2 | 1.34722 | 3.53 | 0.0538 |
| wdg | 1.08333 | 2 | 0.54167 | 1.42 | 0.2710 |
| wdrg | 0.75000 | 2 | 0.37500 | 0.98 | 0.3961 |
| 4 ERROR | 6.11111 | 16 | 0.38194 | | |

| SOURCE | GREENHOUSE GEISSER PROB. | HUYNH FELDT PROB. |
|---|---|---|
| dist | 0.0000 | 0.0000 |
| dr | 0.0016 | 0.0010 |
| dg | 0.7767 | 0.7975 |
| drg | 0.1649 | 0.1592 |
| wd | 0.0005 | 0.0004 |
| wdr | 0.0559 | 0.0538 |
| wdg | 0.2714 | 0.2710 |
| wdrg | 0.3942 | 0.3961 |

**2A. Interaction contrast on the within factors**. If the task is done correctly, then the Wt x Dist interaction should be Linear x Linear.

Step 1: Generate contrast coeff's

Step 2: Apply contrast coeff's to indiv data, find psi-hats for individuals.

Step 3: Analyze the psi-hats in a Grade x Gender between-groups anova. The W-linear x D-linear is the test of the grand mean.

**Step 1:** generate contrast coefficients by multiplying them together:

| | | D1 | D2 | D3 |
|---|---|---|---|---|
| | | -1 | 0 | 1 |
| W1 | 1 | -1 | 0 | 1 |
| W2 | -1 | 1 | 0 | -1 |

**Step 2**: Multiply contrast coefficients x the individual data to make individual psi-hats.

| grade | gender | psi-hat |
|---|---|---|
| 1 | 1 | -2 |
| 1 | 1 | -2 |
| 1 | 1 | -2 |
| 1 | 2 | -3 |
| 1 | 2 | 2 |
| 1 | 2 | -1 |
| 2 | 1 | -2 |
| 2 | 1 | -2 |
| 2 | 1 | -2 |
| 2 | 2 | -2 |
| 2 | 2 | -1 |
| 2 | 2 | -2 |

**Step 3:** Analyze the psi-hats in a between group 2 way between-participants anova. This gives the error term we need. It is a partitioned error. Notice that we have 8 df for error for this test, whereas in the original anova there were 16 df for error.

The test of the grand mean is significant – this says the W-linear x D-linear interaction is significant.

We need to normalize the SS to make sure it is part of the SS WxD in the original anova. $\sum$cj-squared = 4. So proper SS is 30.0833/4 = 7.521. This is a pretty good proportion of the SS W x D from the original anova: 7.521/10.028 = .75. We can also normalize the SS error = 13.3333 / 4 = 3.3333.

**DEPENDENT VARIABLE - LxL**

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| → MEAN | 30.08333 | 1 | 30.08333 | 18.05 | 0.0028 |
| radeg (grade) | 0.75000 | 1 | 0.75000 | 0.45 | 0.5212 |
| gend | 2.08333 | 1 | 2.08333 | 1.25 | 0.2960 |
| rg | 0.75000 | 1 | 0.75000 | 0.45 | 0.5212 |
| 1 ERROR | 13.33333 | 8 | 1.66667 | | |

Other things to note:

-- The Grade effect in this anova is the Grade x Wt-linear x Dist-linear interaction contrast (yes, a 3-way interaction contrast).

-- The Gender effect is the Gender x Wt-linear x Dist-linear interaction contrast (also a 3-way interaction contrast)

-- The Grade x Gender effect is the Grade x Gender x Wt-linear x Dist-linear interaction contrast (a 4-way interaction contrast).

-- If there are more than two levels of the Grouping factors, then these effects would be called ‘partial interactions’, which are no longer covered in this course.

**2B. Test of the residual** from the interaction contrast. SSresid from WxD = 10.028 – 7.521 = 2.507. Because WxD has 2 df, we can also subtract the to make a residual error SS = 6.1111- 3.3333 = 2.7778. So the residual F = 2.507 / (2.7778 / 8) = 7.220. This has df = 1, 8. The residual is also significant, p= 0.028. My interpretation would be that the results deviate from the multiplying form predicted by correct task understanding. Plot the means to see that.

**3. Interaction contrast across a Grouping x Within factor**: Grade x Dist-linear. This would test whether Grades 1 and 2 differ in the linear trend on Distance; i.e., it looks for developmental change in use of the Distance cue (but only in the linear trend of it). To do this, we want to average over the Wt factor.

Step 1: generate contrast coefficients
Step 2: apply contrast coefficients to individual data, and make a table of psi-hats.
Step 3: analyze the psi-hats in a

**Step 1**: The linear contrast coeff's are 1 0 –1. Because this contrast averages across the Weight factor, we can either make individual means over the weight factor, or we can make a 'stretched out' set of contrast coefficients, 1 0 –1 1 0 –1, and apply those. We apply the coeff's to the individual data because part of the contrast involves a within-participants factor.

**Step 2:** Here are the individual psi-hats, with the codes for the grouping factors:

| grade | gender | D-linear psi-hat |
|---|---|---|
| 1 | 1 | -4 |
| 1 | 1 | -4 |
| 1 | 1 | -4 |
| 1 | 2 | -3 |
| 1 | 2 | -2 |
| 1 | 2 | -3 |
| 2 | 1 | -6 |
| 2 | 1 | -6 |
| 2 | 1 | -6 |
| 2 | 2 | -6 |
| 2 | 2 | -7 |
| 2 | 2 | -10 |

**Step 3**: Do a 2-way between-participants anova on the psi-hats. The test of the Mean tests overall D-linear. The Grade ('randeg') effect is Grade x D-linear, and is what we are looking for. It is significant. Grade 2 has a stronger Distance-linear trend than Grade 1. Standardize the SS's by dividing by ∑cj-squared = 4. SS D-linear = 310.0833 / 4 = 77.521. This is a very large portion of the SS-D in original anova = 78.694. SS Grade x D-linear = 36.75 / 4 = 9.1875. This is part of Grade x Dist in the original anova, which had SS = 9.25. Most of the Grade x Dist interaction is linear on Distance.

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 310.08333 | 1 | 310.08333 | 265.79 | 0.0000 |
| → radeg (grade) | 36.75000 | 1 | 36.75000 | 31.50 | 0.0005 |
| gend | 0.08333 | 1 | 0.08333 | 0.07 | 0.7960 |
| rg | 6.75000 | 1 | 6.75000 | 5.79 | 0.0428 |
| 1 ERROR | 9.33333 | 8 | 1.16667 | | |

Note that you could also test D-linear on the grades separately and ask "Is linear trend on D significant for Grade 1?" and "Is linear trend on D significant for Grade 2"? But this test is more straight-forward.

**4. Simple main effect of Grouping factor @ each level of a within factor.**
To do this, just use the data from the one level of the within factor. Let's test Grade@Distance-1. I took only the Distance-1 data (with the two Weights). Here is the result: The test of Grade ('randeg') tests Grade@Dist-1. Notice that the error df is the same as in the overall anova, **BUT** the SS is different. This is because we are partitioning the error by only taking part of the data. The df is partitioned too (see below on 'partition').

I did all 3 simple effect tests here. We see that the Grade effect is only significant at Distance-3. This fits with the Grade x D-linear interaction. The slopes are different, so the means need not differ at all points. Plot the data and see.

**Grade @ Distance-1**

DEPENDENT VARIABLE - w1d1 w2d1

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 140.16667 | 1 | 140.16667 | 152.91 | 0.0000 |
| → radeg (grade) | 0.16667 | 1 | 0.16667 | 0.18 | 0.6811 |
| gend | 0.16667 | 1 | 0.16667 | 0.18 | 0.6811 |
| rg | 4.16667 | 1 | 4.16667 | 4.55 | 0.0656 |
| 1 ERROR | 7.33333 | 8 | 0.91667 | | |
| W | 10.66667 | 1 | 10.66667 | 25.60 | 0.0010 |
| Wr | 0.66667 | 1 | 0.66667 | 1.60 | 0.2415 |
| Wg | 0.66667 | 1 | 0.66667 | 1.60 | 0.2415 |
| Wrg | 2.66667 | 1 | 2.66667 | 6.40 | 0.0353 |
| 2 ERROR | 3.33333 | 8 | 0.41667 | | |

**Grade @ Distance-2**

DEPENDENT VARIABLE - w1d2 w2d2

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 376.04167 | 1 | 376.04167 | 291.13 | 0.0000 |
| radeg (grade) | 2.04167 | 1 | 2.04167 | 1.58 | 0.2441 |
| gend | 1.04167 | 1 | 1.04167 | 0.81 | 0.3954 |
| rg | 2.04167 | 1 | 2.04167 | 1.58 | 0.2441 |
| 1 ERROR | 10.33333 | 8 | 1.29167 | | |
| W | 51.04167 | 1 | 51.04167 | 58.33 | 0.0001 |
| Wr | 2.04167 | 1 | 2.04167 | 2.33 | 0.1651 |
| Wg | 0.37500 | 1 | 0.37500 | 0.43 | 0.5311 |
| Wrg | 5.04167 | 1 | 5.04167 | 5.76 | 0.0432 |
| 2 ERROR | 7.00000 | 8 | 0.87500 | | |

**Grade @ Distance-3**

DEPENDENT VARIABLE - w1d3 w2d3

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 590.04167 | 1 | 590.04167 | 590.04 | 0.0000 |
| radeg (grade) | 15.04167 | 1 | 15.04167 | 15.04 | 0.0047 |
| gend | 0.37500 | 1 | 0.37500 | 0.37 | 0.5573 |
| rg | 0.04167 | 1 | 0.04167 | 0.04 | 0.8434 |
| 1 ERROR | 8.00000 | 8 | 1.00000 | | |
| W | 51.04167 | 1 | 51.04167 | 55.68 | 0.0001 |
| Wr | 0.04167 | 1 | 0.04167 | 0.05 | 0.8365 |
| Wg | 0.04167 | 1 | 0.04167 | 0.05 | 0.8365 |
| Wrg | 1.04167 | 1 | 1.04167 | 1.14 | 0.3175 |
| 2 ERROR | 7.33333 | 8 | 0.91667 | | |

**Verify the partition.** This helps me think about the analyses.
$\sum$SS for the grade effects in these 3 simple effect tests = .167 + 2.042 + 15.042 = 17.251, which is SSr ( or grade) + SSdxr in the original anova = 8.00 + 9.25.
The $\sum$SSerror for testing grade in the 3 analyses = 7.333 + 10.3333 + 8.000 = 25.667. This is SS s/rg + SS s/rg x d = 18.889 + 6.778, the error terms for grade and grade x dist in the original omnibus anova. The $\sum$df-error for these three tests of the grade effects = 24. This is the sum of the dferror for the grade effect, s/rg = 8, and the dferror for Distance, s/rg x D, = 16.

## 5. Simple main effect of a Within factor @ each level of a grouping factor.

Let's test Distance @ Grade 1, and Distance @ Grade 2. I used just the data from Grade 1, then just the data from Grade 2, in separate anovas. These also test Weight @ Grade, and WxD @ Grade, and Gender @ Grade. Normally we wouldn't partition error for the Gender @ Grade test because it involves only the Grouping variables. You could use the original error term from the overall anova with a hand calculation for that.

Results show that the Distance effect is significant at both Grades. The interesting part (from the point of view of the original purpose of the study) is that the WxD interaction is signif for Grade 2 but not for Grade 1. Plot the data!

## Distance @ Grade 1

DEPENDENT VARIABLE - w1d1 w1d2 w1d3 w2d1 w2d2 w2d3

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 427.11111 | 1 | 427.11111 | 187.51 | 0.0002 |
| gend | 5.44444 | 1 | 5.44444 | 2.39 | 0.1970 |
| 1 ERROR | 9.11111 | 4 | 2.27778 | | |
| W | 49.00000 | 1 | 49.00000 | 29.40 | 0.0056 |
| Wg | 4.00000 | 1 | 4.00000 | 2.40 | 0.1963 |
| 2 ERROR | 6.66667 | 4 | 1.66667 | | |
| Dist | 17.55556 | 2 | 8.77778 | 24.31 | 0.0004 |
| Dg | 0.88889 | 2 | 0.44444 | 1.23 | 0.3420 |
| 3 ERROR | 2.88889 | 8 | 0.36111 | | |
| WD | 2.66667 | 2 | 1.33333 | 2.67 | 0.1296 |
| WDg | 0.66667 | 2 | 0.33333 | 0.67 | 0.5398 |
| 4 ERROR | 4.00000 | 8 | 0.50000 | | |

| SOURCE | GREENHOUSE GEISSER PROB. | HUYNH FELDT PROB. |
|---|---|---|
| Dist | 0.0059 | 0.0014 |
| Dg | 0.3320 | 0.3395 |
| WD | 0.1533 | 0.1296 |
| WDg | 0.5038 | 0.5398 |

## Distance @ Grade 2

DEPENDENT VARIABLE - w1d1 w1d2 w1d3 w2d1 w2d2 w2d3

| SOURCE | SUM OF SQUARES | D.F. | MEAN SQUARE | F | TAIL PROB. |
|---|---|---|---|---|---|
| MEAN | 608.44444 | 1 | 608.44444 | 248.91 | 0.0001 |
| gend | 0.44444 | 1 | 0.44444 | 0.18 | 0.6918 |
| 1 ERROR | 9.77778 | 4 | 2.44444 | | |
| W | 53.77778 | 1 | 53.77778 | 44.00 | 0.0027 |
| Wg | 4.00000 | 1 | 4.00000 | 3.27 | 0.1447 |
| 2 ERROR | 4.88889 | 4 | 1.22222 | | |
| Dist | 70.38889 | 2 | 35.19444 | 72.40 | 0.0000 |
| Dg | 1.05556 | 2 | 0.52778 | 1.09 | 0.3827 |
| 3 ERROR | 3.88889 | 8 | 0.48611 | | |
| WD | 10.05556 | 2 | 5.02778 | 19.05 | 0.0009 |
| WDg | 1.16667 | 2 | 0.58333 | 2.21 | 0.1721 |
| 4 ERROR | 2.11111 | 8 | 0.26389 | | |

| SOURCE | GREENHOUSE GEISSER PROB. | HUYNH FELDT PROB. |
|---|---|---|
| Dist | 0.0000 | 0.0000 |
| Dg | 0.3797 | 0.3827 |
| WD | 0.0081 | 0.0019 |
| WDg | 0.2053 | 0.1828 |

**Verify the Partition** so we understand what we did.

a) The **between-participants** part of the design:
   i. ∑SSerror for gender@grade = 9.111 + 9.778 = 18.889 = SS s/rg in omnibus anova. ∑dferror for gender@grade = 4 + 4 = 8, the dferror for the between part of the analysis in the original anova.
   ii. ∑SSgender = 5.444 + .444 = 5.888 = SSgend + Ssgradexgend in omnibus anova. When we divide the data by Grade, we are essentially doing simple effect tests of Gender@Grade.

b) Weight, a **within**-participant variable:
   i. ∑SSerror for W@grade = 6.667 + 4.889 = 11.556 = SSerror for weight (s/rg x w) in omnibus anova. ∑dferror for W @grade = 4 + 4 = dferror for weight in omnibus.
   ii. ∑SSweight@grade = 49.0 + 53.778 = 102.778 = SSweight + SS weight x grade in omnibus anova = 102.7222 + .0556.

c) Distance
   i. ∑SSerror for D@grade = 2.889 + 3.889 = 6.778 = SSerror for distance (s/rg x d) in the omnibus anova. And the df's add up properly too.
   ii. ∑Ssdistance@grade = 17.556 + 70.389 = 87.945 = SSdist + SS dist x grade in omnibus anova = 78.694 + 9.250.

d) Within x Between partitions
   i. Weight x gender in the simple effect analyses here sum to weight x grade x gender + weight x gender in the omnibus anova (it so happens that one of these values is zero for these contrived data).
   ii. Dist x gender in the simple effects sum to dist x grade x gender + dist x gender in the omnibus anova.

**Stop and think**: Do we want partitioned error for these tests? Remember, any test that involves a within-participant (repeated measures) factor should probably use a partitioned error. Partitioned error reduces the scope over which the sphericity assumption must hold.