The Mathematica® Journal

Computing Mixed-Design (Split-Plot) ANOVA

Sylvain Chartier
Denis Cousineau

The mixed, within-between subjects ANOVA (also called a split-plot ANOVA) is a statistical test of means commonly used in the behavioral sciences. One approach to computing this analysis is to use a corrected between-subjects ANOVA. A second approach uses the general linear model by partitioning the sum of squares and cross-product matrices. Both approaches are detailed in this article. Finally, a package called *MixedDesignANOVA* is introduced that runs mixed-design ANOVAs using the second approach and displays summary statistics as well as a mean plot.

■ Introduction

The mixed, within-between subjects design (also called split-plot or randomized blocks factorial) ANOVA is a technique that compares the means obtained by manipulating two factors, one being a repeated-measure factor. Let g be the number of independent groups, each representing one level of the between-subjects factor, let c be the number of measures corresponding to the within-subjects factor, and let n_i be the number of subjects in the i^{th} group.

The data is contained in a matrix X of the form:

$$X = \begin{pmatrix} 1 & y_{111} & y_{112} & \cdots & y_{11c} \\ 1 & y_{121} & y_{122} & \cdots & y_{12c} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & y_{1n_1 1} & y_{1n_1 2} & \cdots & y_{1n_1 c} \\ 2 & y_{211} & y_{212} & \cdots & y_{21c} \\ 2 & y_{221} & y_{222} & \cdots & y_{22c} \\ \vdots & \vdots & \vdots & & \vdots \\ 2 & y_{2n_2 1} & y_{2n_2 2} & \cdots & y_{2n_2 c} \\ \vdots & \vdots & \vdots & & \vdots \\ g & y_{g11} & y_{g12} & \cdots & y_{g1c} \\ g & y_{g21} & y_{g22} & \cdots & y_{g2c} \\ \vdots & \vdots & \vdots & & \vdots \\ g & y_{gn_g 1} & y_{gn_g 2} & \cdots & y_{gn_g c} \end{pmatrix}$$

First load the package. It is available from
www.mathematica-journal.com/data/uploads/2011/10/Chartier.zip.

```
Needs["MixedDesignANOVA`"]
```

An example taken from Howell [1] (p. 481) concerns data collected in a study by King [2]. King investigated the effect of midazolam on the motor activity of rats. The rats were measured at six different times ($c = 6$) and there were $g = 3$ equal groups of $n_i = 8$ individuals, $i = 1 \ldots, g$. Hence, the total number of rats measured was $N = 24$. The data is listed in Table 1.

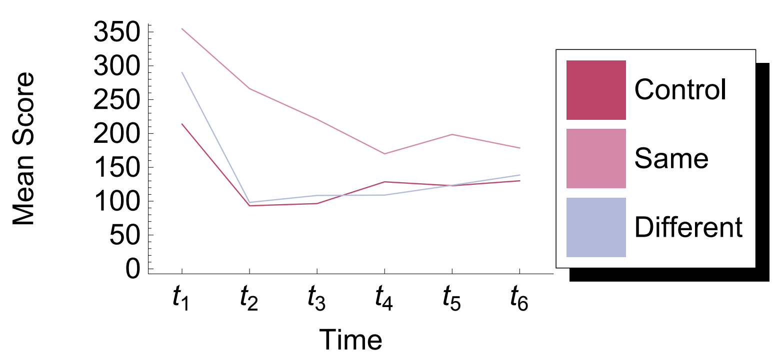
```
WSLabels = {"Time", {"t1", "t2", "t3", "t4", "t5", "t6"}};
BSLabels = {"Group", {"Control", "Same", "Different"}};
X = {
     {1, 150., 44., 71., 59., 132., 74.},
     {1, 335., 270., 156., 160., 118., 230.},
     {1, 149., 52., 91., 115., 43., 154.},
     {1, 159., 31., 127., 212., 71., 224.},
     {1, 159., 0., 35., 75., 71., 34.},
     {1, 292., 125., 184., 246., 225., 170.},
     {1, 297., 187., 66., 96., 209., 74.},
     {1, 170., 37., 42., 66., 114., 81.},
     {2, 346., 175., 177., 192., 239., 140.},
     {2, 426., 329., 236., 76., 102., 232.},
     {2, 359., 238., 183., 123., 183., 30.},
     {2, 272., 60., 82., 85., 101., 98.},
     {2, 200., 271., 263., 216., 241., 227.},
     {2, 366., 291., 263., 144., 220., 180.},
     {2, 371., 364., 270., 308., 219., 267.},
     {2, 497., 402., 294., 216., 284., 255.},
     {3, 362., 104., 144., 114., 115., 127.},
     {3, 338., 132., 91., 77., 108., 169.},
     {3, 282., 186., 225., 134., 189., 169.},
     {3, 317., 31., 85., 120., 131., 205.},
     {3, 263., 94., 141., 142., 120., 195.},
     {3, 138., 38., 16., 95., 39., 55.},
     {3, 329., 62., 62., 6., 93., 67.},
     {3, 292., 139., 104., 184., 193., 122.}
    } /. {1 -> BSLabels[[2, 1]], 2 -> BSLabels[[2, 2]],
     3 -> BSLabels[[2, 3]]};
TableForm[X,
 TableHeadings -> {None, Join[{BSLabels[[1]]}, WSLabels[[2]]]}
]
```

Group	t_1	t_2	t_3	t_4	t_5	t_6
Control	150.	44.	71.	59.	132.	74.
Control	335.	270.	156.	160.	118.	230.
Control	149.	52.	91.	115.	43.	154.
Control	159.	31.	127.	212.	71.	224.
Control	159.	0.	35.	75.	71.	34.
Control	292.	125.	184.	246.	225.	170.
Control	297.	187.	66.	96.	209.	74.
Control	170.	37.	42.	66.	114.	81.
Same	346.	175.	177.	192.	239.	140.
Same	426.	329.	236.	76.	102.	232.
Same	359.	238.	183.	123.	183.	30.
Same	272.	60.	82.	85.	101.	98.
Same	200.	271.	263.	216.	241.	227.
Same	366.	291.	263.	144.	220.	180.
Same	371.	364.	270.	308.	219.	267.
Same	497.	402.	294.	216.	284.	255.
Different	362.	104.	144.	114.	115.	127.
Different	338.	132.	91.	77.	108.	169.
Different	282.	186.	225.	134.	189.	169.
Different	317.	31.	85.	120.	131.	205.
Different	263.	94.	141.	142.	120.	195.
Different	138.	38.	16.	95.	39.	55.
Different	329.	62.	62.	6.	93.	67.
Different	292.	139.	104.	184.	193.	122.

▲ **Table 1.** Data from Howell (2003) of a 3×6 design (three groups, six repeated measures).

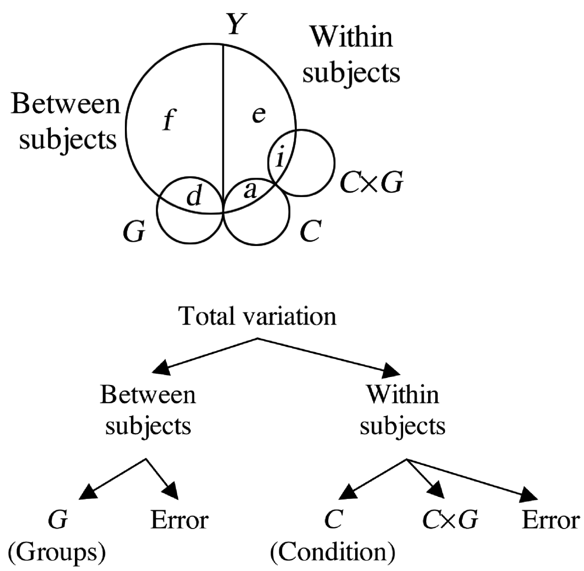
The plot of the means across conditions (Figure 1) shows evidence of a time effect, the results on the second time being generally smaller than those on the first time. In addition, there is an interaction effect caused by the "same" group that does not follow this pattern.

```
MeanPlot /. MixedDesignANOVA[X,
  Labels -> {WSLabels, BSLabels}, MeanPlot → True]
```



▲ **Figure 1.** Illustration of the example given in Table 1.

Figure 2 shows variation partitioning for the subjects by condition design. The between-subjects variation is decomposed into two parts: a source of variation due to the group effect (area d) and a source of variation due to the measurement error (area f). Within-subjects variation is decomposed into three areas: a source of variation for the repeated measures effect (area a), a source of variation for the interaction between the repeated measures and the group effect (area i), and a source of variation for error (area e). Consequently, there will be three F ratios to compute: the group effect (the ratio f / d, F_G), the repeated measure effect (the ratio a / e, F_C) and the interaction effect (the ratio i / e, $F_{C \times G}$).



▲ **Figure 2.** (top) Venn diagram for the subjects within groups by conditions design. G is the group effect, C is the repeated-measure effect and $C \times G$ is the interaction effect between the two factors. (bottom) Tree diagram showing the partitioning of the different sources of variation.

The total variation R^2 is $1 = a + d + e + f + i$. Each letter corresponds to a proportion of variation (r-squared). The between-subjects proportions of variation are:

$$d = R^2_{Y \cdot G},$$
$$f = R^2_{Y \cdot S} - R^2_{Y \cdot G},$$
$$d + f = R^2_{Y \cdot S},$$

where $R^2_{Y \cdot S}$ represents the variation explained by the between-subjects source. By dividing d and f by $d + f$, we get the mean effect, that is, the group effect:

$$\frac{d}{d + f} = R^2_{\overline{Y} \cdot G},$$
$$\frac{f}{d + f} = 1 - \frac{d}{d + f} = 1 - R^2_{\overline{Y} \cdot G}.$$

Finally, the within-subjects proportions of variation are:

$$a = R^2_{Y \cdot C},$$
$$i = R^2_{Y \cdot C \times G},$$
$$e = 1 - (d + f) - a - i = 1 - R^2_{Y \cdot S} - R^2_{Y \cdot C} - R^2_{Y \cdot C \times G}.$$

The various F statistics are ratios of the following proportions of variation weighted by their corresponding degrees of freedom:

$$F_G = \frac{d}{f} \times \frac{N-g}{g-1} = \frac{R^2_{\overline{Y} \cdot G}}{1 - R^2_{\overline{Y} \cdot G}} \times \frac{N-g}{g-1},$$
$$F_C = \frac{a}{e} \times N - g = \frac{R^2_{Y \cdot C}}{1 - R^2_{Y \cdot S} - R^2_{Y \cdot C} - R^2_{Y \cdot C \times G}} \times N - g,$$
$$F_{C \times G} = \frac{i}{e} \times \frac{N-g}{g-1} = \frac{R^2_{Y \cdot C \times G}}{1 - R^2_{Y \cdot S} - R^2_{Y \cdot C} - R^2_{Y \cdot C \times G}} \times \frac{N-g}{g-1}.$$

Computing a Split-Plot ANOVA from the Computations Obtained by a Between-Subjects ANOVA

Test of the Between-Subjects Effect

The group (between-subjects) effect (measured by d/f in Figure 2) can be obtained by averaging the repeated measures (so that information about the repeated measures is discarded) and submitting them to a one-way ANOVA:

$$\overline{y}_{ij} = \frac{1}{c} \sum_{k=1}^{c} y_{ijk},\ i = 1, \ldots, g,\ j = 1, \ldots, n_i.$$

Using *Mathematica*, we need to set a few constants first, and for convenience, the labels for the within- and between-subjects factors. The conditions are taken from the first column of X.

```
AllConditions = Union[X[[All, 1]]];
```

This computes the number of groups and repeated measures.

```
g = Length[AllConditions];
c = Length[X[[1]]] - 1;
```

This computes the total number of participants and the number of participants per group.

```
nTotal = Length[X];
Table[
  ni = Length[Select[X[[All, 1]], # == AllConditions[[i]] &]],
  {i, 1, g}];
```

Finally, we define the labels.

```
WSLabels = {"Time", {"t1", "t2", "t3", "t4", "t5", "t6"}};
BSLabels = {"Group", {"Control", "Same", "Different"}};
WSVariable = Time;
BSVariable = Group;
```

To get the between-subjects effect, we aggregate the data over replicated measures.

```
yG = Table[{X[[j, 1]], (1/c) Total[Drop[X[[j]], 1]]}, {j, nTotal}];
```

Using this, a one-way ANOVA is computed.

```
Needs["ANOVA`"]
```

```
resGrp = ANOVA[yG, {BSVariable}, {BSVariable},
  CellMeans → False]
```

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
Group	2	47 635.8	23 817.9	7.80059	0.00292823
Error	21	64 120.3	3053.35		
Total	23	111 756.			

In the following, we will need the between-subjects sum of squares, so we extract it from the table. To take into account the c repeated measures, the sum of squares must be multiplied by that number of measures.

```
SSBetween = c × resGrp[[2, 1, 3, 2]]
```

670 537.

□ Test of the Within-Subjects Effect

F ratios for the repeated-measure effect and the interaction effect are computed by first recoding the data matrix such that the repeated measures look like a second between-subjects factor. Thus, the bulk of the analysis simplifies into a standard factorial ANOVA. The following transforms the data.

```
yC×G = Flatten[Table[{X⟦j, 1⟧, i, X⟦j, i + 1⟧}, {i, c},
    {j, nTotal}], 1];
```

Applying a standard between-factors ANOVA, the following summary table is obtained.

```
results = ANOVA[yC×G, {WSVariable, BSVariable, All},
  {WSVariable, BSVariable}, CellMeans → False]
```

ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
Time	2	285 815.	142 908.	27.0398	1.69471×10^{-10}
Group	5	399 737.	79 947.3	15.127	1.26463×10^{-11}
Group Time	10	80 820.	8082.	1.52921	0.136377
Error	126	665 921.	5285.09		
Total	143	1.43229×10^{6}			

First, the results regarding the group effect are discarded since it has been analyzed in the previous section. Next, the results regarding the repeated measure “time” and the interaction (group × time) must be modified to obtain the corrected *F* ratios. Specifically, information regarding the error term is incorrect since it does not take into account the estimation of the between-subjects error that we obtained in the previous subsection. The corrected error sum of squares is given by

$$SS_{\text{Error}} = SS_{\text{Total}} - SS_{\text{Between}} - SS_{\text{Within}} - SS_{\text{interaction}}.$$

The corrected error degrees of freedom (df) is given by

$$df_{\text{Error}} = (N - g)\,(c - 1).$$

Using the results of the previous ANOVA, the sum of squares terms can be extracted and the corrected error term computed.

```
SSTotal = results⟦2, 1, 5, 2⟧;
SSWithin = results⟦2, 1, 2, 2⟧;
SSInteraction = results⟦2, 1, 3, 2⟧;
SSError = SSTotal - SSBetween - SSWithin - SSInteraction;

dfError = (nTotal - g) × (c - 1)

105
```

From this, the error mean square (MS) can be computed.

$$\text{MS}_{\text{Error}} = \frac{\text{SS}_{\text{Error}}}{\text{df}_{\text{Error}}}$$

```
2678.09
```

The within-subjects *F* ratios are summarized in the following table.

```
TableForm[{
  Join[results⟦2, 1, 2, {1, 2, 3}⟧,
   {results⟦2, 1, 2, 3⟧ / MSError,
    1 - CDF[FRatioDistribution[results⟦2, 1, 1, 1⟧, dfError],
      results⟦2, 1, 1, 3⟧ / MSError]}],
  Join[results⟦2, 1, 3, {1, 2, 3}⟧,
   {results⟦2, 1, 3, 3⟧ / MSError,
    1 - CDF[FRatioDistribution[results⟦2, 1, 3, 1⟧, dfError],
      results⟦2, 1, 3, 3⟧ / MSError]}],
  {dfError, SSError, MSError}
 },
 TableHeadings →
  {{WSLabels⟦1⟧, WSLabels⟦1⟧ <> " × " <> BSLabels⟦1⟧, "Error"},
   {"DF", "SS", "MS", "F", "P"}}
]
```

	DF	SS	MS	F	P
Time	5	399 737.	79 947.3	29.8524	1.11022×10^{-16}
Time × Group	10	80 820.	8082.	3.01782	0.00216428
Error	105	281 199.	2678.09		

■ Computing a Split-Plot ANOVA Using the General Linear Model Approaches

A different technique for computing a split-plot ANOVA is to use the general linear model approaches [3, 4]. By using the general linear model, all ANOVAs (factorial, repeated measures, etc.) are treated as a special case of regression analyses; the dependent variables remain the same while the predictors are generated using binary codes. Then, the various variability terms (error, within-subjects, between-subjects) can be estimated as ratios of explained to unexplained variation.

Test of the Between-Subjects Effect

For this effect, a coding matrix that identifies each subject within each group could be used. However, as pointed out in [3] and detailed in [4], it is simpler to aggregate the repeated measures as in the previous section to consider only the group effect. Hence, the effect coding matrix for the groups (EC_G) has g lines. Because the last group is entirely determined by the other groups, it is not coded (otherwise the resulting matrix would be singular), resulting in $g-1$ columns.

Group	Coded As	x_1	x_2	$\cdots$	x_{g-1}
1	→	1	0	$\cdots$	0
2	→	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$g-1$	→	0	0	$\cdots$	1
g	→	−1	−1	$\cdots$	−1

Using the group coding vector for each subject and joining to it the dependent variable, we get a matrix M that contains the predictor variables in the first columns and the dependent variable in the last column.

With this matrix M, the sum of squares and cross-product (SSCP) matrix can be computed by

$$\text{SSCP} = M^{\mathrm{T}} M - (\mathbf{1}^{\mathrm{T}} M)^{\mathrm{T}} (\mathbf{1}^{\mathrm{T}} M) / N, \tag{1}$$

where $\mathbf{1}$ represents the N-dimensional vector composed of 1s. By partitioning the SSCP matrix, the coefficient of determination R^2 is obtained. The SSCP must be partitioned into four submatrices named SS_{pp}, SS_{pd}, SS_{dp}, and SS_{dd} (in which the subscript p stands for *predictors* and d stands for *dependent variable*). These matrices represent, respectively, the sum of squares of the predictors alone, the sum of cross-products between the predictors and the dependent variable, the sum of cross-products between the dependent variable and the predictors, and lastly the sum of squares of the dependent variable alone.

$$\text{SSCP} = \left(\begin{array}{c|c} \text{SS}_{pp} & \text{SS}_{pd} \\ \hline \text{SS}_{dp} & \text{SS}_{dd} \end{array}\right),$$

where the size of the SS_{pp} matrix is $(g-1)\times(g-1)$ and the size of the SS_{dd} matrix is 1×1. Finally, we verify that $\text{SS}_{pd} = \text{SS}_{dp}{}^{\mathrm{T}}$.

The coefficient of determination for the between-subjects effect ($R^2_{\overline{Y}\cdot G}$) can be obtained by the matrix multiplication

$$R^2_{\overline{Y}\cdot G} = \text{SS}_{pd}\,\text{SS}_{pp}^{-1}\,\text{SS}_{pd}\,\text{SS}_{dd}^{-1}.$$

Finally, the F value is the ratio between the explained variation and the unexplained variation, weighted by the degrees of freedom:

$$F_G = \frac{R^2_{\overline{Y}\cdot G}}{1 - R^2_{\overline{Y}\cdot G}} \times \frac{N-g}{g-1}.$$

All those operations are performed with the following commands. First we have the coding matrix.

```
ECG = Append[Table[AllConditions[[i]] → IdentityMatrix[g - 1][[i]],
     {i, g - 1}],
    AllConditions[[g]] → Table[-1, {g - 1}]];
TableForm[ECG]
```

```
Control → {1, 0}
Different → {0, 1}
Same → {-1, -1}
```

The means for each group are then computed.

```
ȳG = Table[{X[[j, 1]], 1/c Total[Drop[X[[j]], 1]]}, {j, nTotal}]
```

```
{{Control, 88.3333}, {Control, 211.5}, {Control, 100.667},
 {Control, 137.333}, {Control, 62.3333}, {Control, 207.},
 {Control, 154.833}, {Control, 85.}, {Same, 211.5},
 {Same, 233.5}, {Same, 186.}, {Same, 116.333},
 {Same, 236.333}, {Same, 244.}, {Same, 299.833},
 {Same, 324.667}, {Different, 161.}, {Different, 152.5},
 {Different, 197.5}, {Different, 148.167},
 {Different, 159.167}, {Different, 63.5},
 {Different, 103.167}, {Different, 172.333}}
```

Then the matrix M can be constructed.

```
M = Map[Flatten[{#[[1]] /. ECG, #[[2]]}] &, ȳG]
```

```
{{1, 0, 88.3333}, {1, 0, 211.5}, {1, 0, 100.667},
 {1, 0, 137.333}, {1, 0, 62.3333}, {1, 0, 207.},
 {1, 0, 154.833}, {1, 0, 85.}, {-1, -1, 211.5},
 {-1, -1, 233.5}, {-1, -1, 186.}, {-1, -1, 116.333},
 {-1, -1, 236.333}, {-1, -1, 244.}, {-1, -1, 299.833},
 {-1, -1, 324.667}, {0, 1, 161.}, {0, 1, 152.5},
 {0, 1, 197.5}, {0, 1, 148.167}, {0, 1, 159.167},
 {0, 1, 63.5}, {0, 1, 103.167}, {0, 1, 172.333}}
```

From this matrix M, the SSCP matrix is easily obtained according to equation 1.

```
ones = Array[1 &, {Length[M], 1}];
SSCP = Transpose[M].M -
   Transpose[(Transpose[ones].M)].(Transpose[ones].M) /
    Length[M];
MatrixForm[SSCP]
```

$$\begin{pmatrix} 16. & 8. & -805.167 \\ 8. & 16. & -694.833 \\ -805.167 & -694.833 & 111\,756. \end{pmatrix}$$

This SSCP matrix can then be partitioned into the various sum of squares submatrices needed to compute $R^2_{\overline{Y}\cdot G}$.

```
SSpp = Take[SSCP, {1, g - 1}, {1, g - 1}];
SSpd = Take[SSCP, {1, g - 1}, {g, g}];
SSdd = Take[SSCP, {g, g}, {g, g}];
R2Y.G = (Transpose[SSpd].Inverse[SSpp].SSpd.Inverse[SSdd])[[
  1, 1]]
```

```
0.426248
```

Finally, the F ratio can be obtained.

$$F_G = \frac{R2_{\overline{Y}\cdot G}}{1 - R2_{\overline{Y}\cdot G}} \times \frac{n_{Total} - g}{g - 1}$$

```
7.80059
```

And again, the total sum of squares of the between-subjects factor $S^2_{\overline{Y}}$ is memorized for later use (area f in Figure 2).

```
SSBetween = c × SSdd;
```

□ Test of the Within-Subjects Effects

As in the previous section, the within-subjects effects are computed by dropping the repeated measures. Therefore, the subjects are considered independent and the computation is accomplished like a standard between-subjects ANOVA. However, the error term will differ from a standard ANOVA since the between-subjects variability has been evaluated and thus can be removed from the total error (Figure 2).

The first step is to compute the proportion of variance accounted for by the repeated measure effect $R^2_{Y \cdot C}$. To this end, we must create an effect coding matrix for the c repeated measures. This effect coding is created as in the previous section except that it is of size c.

```
ECC = Append[
   Table[i → IdentityMatrix[c - 1][[i]], {i, c - 1}],
   c → Table[-1, {c - 1}]];
TableForm[ECC]
```

```
1 → {1, 0, 0, 0, 0}
2 → {0, 1, 0, 0, 0}
3 → {0, 0, 1, 0, 0}
4 → {0, 0, 0, 1, 0}
5 → {0, 0, 0, 0, 1}
6 → {-1, -1, -1, -1, -1}
```

The raw data is reorganized so that for each item, the replication number is available, $\{j, y_{ijk}\}$, $i = 1, \ldots, g$, $j = 1, \ldots, c$, and $k = 1, \ldots n_i$. From it, the matrix M is computed in exactly the same way as before.

```
yC = Flatten[Table[{i, X[[j, i + 1]]}, {i, 1, c}, {j, nTotal}], 1];
M = Map[Flatten[{#[[1]] /. ECC, #[[2]]}] &, yC];
```

Next, the SSCP matrix is computed exactly as before.

```
ones = Array[1 &, {Length[M], 1}];
SSCP = Transpose[M].M -
   Transpose[(Transpose[ones].M)].(Transpose[ones].M) /
    Length[M];
MatrixForm[SSCP]
```

$$\begin{pmatrix} 48. & 24. & 24. & 24. & 24. & 3290. \\ 24. & 48. & 24. & 24. & 24. & 83. \\ 24. & 24. & 48. & 24. & 24. & -171. \\ 24. & 24. & 24. & 48. & 24. & -318. \\ 24. & 24. & 24. & 24. & 48. & -19. \\ 3290. & 83. & -171. & -318. & -19. & 1.43229 \times 10^6 \end{pmatrix}$$

Partitioning this matrix, the quantity $R^2_{Y \cdot C}$ can be obtained exactly as before.

```
SSpp = Take[SSCP, {1, c - 1}, {1, c - 1}];
SSpd = Take[SSCP, {1, c - 1}, {c, c}];
SSdd = Take[SSCP, {c, c}, {c, c}];
R2Y·C = (Transpose[SSpd].Inverse[SSpp].SSpd.Inverse[SSdd])[[
  1, 1]]
```

```
0.279089
```

The same steps are repeated one last time for the interaction effect $R^2_{Y \cdot C \times G}$. The effect coding matrix for the interaction is defined for all the combinations of the between-subjects factor and the repeated-measure factor. It is obtained with the outer product of the individual effect coding.

```
ECCG = Flatten[
   Table[{AllConditions⟦j⟧, i} →
     Flatten[Transpose[{AllConditions⟦j⟧ /. ECG}].
       ({i /. ECC})], {j, 1, g}, {i, 1, c}],
   1];
TableForm[ECCG]
```

```
{Control, 1} → {1, 0, 0, 0, 0, 0, 0, 0, 0, 0}
{Control, 2} → {0, 1, 0, 0, 0, 0, 0, 0, 0, 0}
{Control, 3} → {0, 0, 1, 0, 0, 0, 0, 0, 0, 0}
{Control, 4} → {0, 0, 0, 1, 0, 0, 0, 0, 0, 0}
{Control, 5} → {0, 0, 0, 0, 1, 0, 0, 0, 0, 0}
{Control, 6} → {-1, -1, -1, -1, -1, 0, 0, 0, 0, 0}
{Different, 1} → {0, 0, 0, 0, 0, 1, 0, 0, 0, 0}
{Different, 2} → {0, 0, 0, 0, 0, 0, 1, 0, 0, 0}
{Different, 3} → {0, 0, 0, 0, 0, 0, 0, 1, 0, 0}
{Different, 4} → {0, 0, 0, 0, 0, 0, 0, 0, 1, 0}
{Different, 5} → {0, 0, 0, 0, 0, 0, 0, 0, 0, 1}
{Different, 6} → {0, 0, 0, 0, 0, -1, -1, -1, -1, -1}
{Same, 1} → {-1, 0, 0, 0, 0, -1, 0, 0, 0, 0}
{Same, 2} → {0, -1, 0, 0, 0, 0, -1, 0, 0, 0}
{Same, 3} → {0, 0, -1, 0, 0, 0, 0, -1, 0, 0}
{Same, 4} → {0, 0, 0, -1, 0, 0, 0, 0, -1, 0}
{Same, 5} → {0, 0, 0, 0, -1, 0, 0, 0, 0, -1}
{Same, 6} → {1, 1, 1, 1, 1, 1, 1, 1, 1, 1}
```

The data matrix is reorganized one last time so that both the group and the number of the repeated measures are available: $\{i, j, y_{ijk}\}$, $i = 1, \ldots, g$, $j = 1, \ldots, c$, and $k = 1, \ldots n_i$. The M and SSCP matrices are then computed as usual.

```
yCG = Flatten[Table[{X[[j, 1]], i, X[[j, i + 1]]}, {i, 1, c},
     {j, nTotal}], 1];
M = Map[Flatten[{{#[[1]], #[[2]]} /. ECCG, #[[3]]}] &, yCG];

ones = Array[1 &, {Length[M], 1}];
SSCP = Transpose[M].M -
   Transpose[(Transpose[ones].M)].(Transpose[ones].M) /
    Length[M];
MatrixForm[SSCP]
```

32.	16.	16.	16.	16.	16.	8.	8.	8.	8.	-738.
16.	32.	16.	16.	16.	8.	16.	8.	8.	8.	-996.
16.	16.	32.	16.	16.	8.	8.	16.	8.	8.	-608.
16.	16.	16.	32.	16.	8.	8.	8.	16.	8.	57.
16.	16.	16.	16.	32.	8.	8.	8.	8.	16.	-218.
16.	8.	8.	8.	8.	32.	16.	16.	16.	16.	-196.
8.	16.	8.	8.	8.	16.	32.	16.	16.	16.	-1024.
8.	8.	16.	8.	8.	16.	16.	32.	16.	16.	-580.
8.	8.	8.	16.	8.	16.	16.	16.	32.	16.	-168.
8.	8.	8.	8.	16.	16.	16.	16.	16.	32.	-281.
-738.	-996.	-608.	57.	-218.	-196.	-1024.	-580.	-168.	-281.	1.43229×10^6

Partitioning this last matrix, the quantity $R^2_{Y\cdot C\times G}$ can be obtained.

```
SSpp = Take[SSCP, {1, (g - 1) (c - 1)}, {1, (g - 1) (c - 1)}];
SSpd = Take[SSCP, {1, (g - 1) (c - 1)},
   {(g - 1) (c - 1) + 1, (g - 1) (c - 1) + 1}];
SSdd = Take[SSCP, {(g - 1) (c - 1) + 1, (g - 1) (c - 1) + 1},
   {(g - 1) (c - 1) + 1, (g - 1) (c - 1) + 1}];
R2Y·C×G = (Transpose[SSpd].Inverse[SSpp].SSpd.Inverse[SSdd])[[
  1, 1]]

0.056427
```

Finally, the error proportion of variance can be estimated as the within-subjects variation left unexplained:

$$R^2_{\text{Error}} = 1 - R^2_{Y \cdot S} - R^2_{Y \cdot C} - R^2_{Y \cdot C \times G},$$

in which $R^2_{Y \cdot S}$ is given by

$$R^2_{Y \cdot S} = \frac{S^2_{\overline{Y}}}{S^2_{Y}},$$

where $S^2_{\overline{Y}}$ is the between-factors sum of squares memorized in the previous subsection.

```
R2Y·S = SSBetween / SSdd [[1, 1]]
```

```
0.468156
```

```
R2Error = 1 - R2Y·S - R2Y·C - R2Y·C×G
```

```
0.196328
```

The F ratios for the repeated-measure effect and the interaction effect can be computed with the formulas:

$$F_C = \frac{R^2_{Y \cdot C}}{R^2_{\text{Error}}} \times (N - g),$$

$$F_{C \times G} = \frac{R^2_{Y \cdot C \times G}}{R^2_{\text{Error}}} \times \frac{N - g}{g - 1},$$

which we compute.

```
FC = R2Y·C / R2Error (nTotal - g)
```

```
29.8524
```

```
FCG = R2Y·C×G / R2Error (nTotal - g) / (g - 1)
```

```
3.01782
```

All the information required has been gathered; the ANOVA table can be produced just like in the section Computing a Split-Plot ANOVA from the Computations Obtained by a Between-Subjects ANOVA.

■ The *MixedDesignANOVA* Package

The *MixedDesignANOVA* package performs the different analyses using the procedures outlined in the previous section. It works for equal as well as for unequal numbers of subjects per group. To use it, first load the package (adapt the path if necessary) and load some data. Optionally, you can define labels for the factors and the levels of the factors.

```
Needs["MixedDesignANOVA`"]
```

```
X = {{Control, 150., 44., 71., 59., 132., 74.},
    {Control, 335., 270., 156., 160., 118., 230.},
    {Control, 149., 52., 91., 115., 43., 154.},
    {Control, 159., 31., 127., 212., 71., 224.},
    {Control, 159., 0., 35., 75., 71., 34.},
    {Control, 292., 125., 184., 246., 225., 170.},
    {Control, 297., 187., 66., 96., 209., 74.},
    {Control, 170., 37., 42., 66., 114., 81.},
    {Same, 346., 175., 177., 192., 239., 140.},
    {Same, 426., 329., 236., 76., 102., 232.},
    {Same, 359., 238., 183., 123., 183., 30.},
    {Same, 272., 60., 82., 85., 101., 98.},
    {Same, 200., 271., 263., 216., 241., 227.},
    {Same, 366., 291., 263., 144., 220., 180.},
    {Same, 371., 364., 270., 308., 219., 267.},
    {Same, 497., 402., 294., 216., 284., 255.},
    {Different, 282., 186., 225., 134., 189., 169.},
    {Different, 317., 31., 85., 120., 131., 205.},
    {Different, 362., 104., 144., 114., 115., 127.},
    {Different, 338., 132., 91., 77., 108., 169.},
    {Different, 263., 94., 141., 142., 120., 195.},
    {Different, 138., 38., 16., 95., 39., 55.},
    {Different, 329., 62., 62., 6., 93., 67.},
    {Different, 292., 139., 104., 184., 193., 122.}};
WSLabels = {"times", {"t1", "t2", "t3", "t4", "t5", "t6"}};
BSLabels = {"Group", {"Control", "Same", "Different"}};
```

The command `MixedDesignANOVA`, with a data matrix respecting the input format `X` defined in the first section, displays the ANOVA table only, with default names (B for the between-subjects factor and A for the within-subjects factor).

```
ANOVATable /. MixedDesignANOVA[X]
```

	SC	dl	CM	F	P
Between-Subjects					
B	285 815.	2	142 908.	7.80059	0.00292823
Error	384 722.	21	18 320.1		
Total	670 537.	23			
Within-Subjects					
A	399 737.	5	79 947.3	29.8524	5.43708×10^{-11}
A × B	80 820.	10	8082.	3.01782	0.00216428
Error	281 199.	105	2678.09		
Total	761 756.	120			
Total	1.43229×10^{6}	143			

The command has seven options, as listed below. The option `Epsilons` is used to compute the Greenhouse–Geisser, the Huynh–Feldt and the lower-bound epsilons [5, 6]. The options `MeanTable` and `MeanPlot` show the mean across conditions and measures under the form of a table or visually. The option `VarCov` returns the $g\times g$ variance-covariance matrices for each group as well as the global variance-covariance matrix.

Finally, the option `SummaryStatistics` can be used to display summary statistics for each cell of the design. Default summary statistics are `None`; `Automatic` returns the mean, variance, standard deviation, length (i.e. the number of observations in the cell), unbiased skewness, and unbiased kurtosis.

```
Options[MixedDesignANOVA]

{ANOVATable → True, Epsilons → False,
 Labels → Automatic, MeanPlot → False, MeanTable → False,
 SummaryStatistics → None, VarCov → False}
```

The next command runs an analysis with all options turned on; the results are displayed one at a time afterward.

```
analysis = MixedDesignANOVA[X, Labels → {WSLabels, BSLabels},
   MeanPlot → False, MeanTable → True,
   SummaryStatistics → {Mean, Total, Variance, Skewness},
   Epsilons → True, VarCov → True];
```

```
SummaryStatistics /. analysis
```

	Mean	Total	Variance	Skewness
Control-t_1	213.875	1711.	6274.41	0.57759
Control-t_2	93.25	746.	8699.93	0.933113
Control-t_3	96.5	772.	2927.14	0.452598
Control-t_4	128.625	1029.	4943.98	0.638305
Control-t_5	122.875	983.	4256.41	0.503268
Control-t_6	130.125	1041.	5557.27	0.199142
Control	130.875	6282.	6490.66	0.614616
Same-t_1	354.625	2837.	8084.55	-0.211362
Same-t_2	266.25	2130.	12 031.4	-0.677922
Same-t_3	221.	1768.	4883.43	-0.969977
Same-t_4	170.	1360.	6100.86	0.406149
Same-t_5	198.625	1589.	4385.41	-0.554247
Same-t_6	178.625	1429.	6987.98	-0.644123
Same	231.521	11 113.	10 434.2	0.213016
Different-t_1	290.125	2321.	4805.55	-1.36696
Different-t_2	98.25	786.	2859.64	0.21114
Different-t_3	108.5	868.	3926.57	0.465401
Different-t_4	109.	872.	2759.14	-0.688462
Different-t_5	123.5	988.	2510.29	-0.0230334
Different-t_6	138.625	1109.	3137.7	-0.374307
Different	144.667	6944.	7468.31	0.83978
Grand total	169.021	24 339.	10 016.	0.63334

```
MeanTable /. analysis
```

	t_1	t_2	t_3	t_4	t_5	t_6	Mean
Control	213.875	93.25	96.5	128.625	122.875	130.125	130.875
Same	354.625	266.25	221.	170.	198.625	178.625	231.521
Different	290.125	98.25	108.5	109.	123.5	138.625	144.667
Mean	286.208	152.583	142.	135.875	148.333	149.125	169.021

```
Epsilons /. analysis
```

Greenhouse-Geisser	0.656945
Huynh-Feldt	0.867425
Lower Bound	0.2

VarCov /. analysis

{

Variance/Covariance matrices (one per group)

,

$$\left\{\begin{pmatrix} 6274.41 & 6929.32 & 2512.21 & 2386.38 & 3604.13 & 2123.16 \\ 6929.32 & 8699.93 & 2688.57 & 1933.96 & 3123.75 & 3045.68 \\ 2512.21 & 2688.57 & 2927.14 & 3462.36 & 1188.21 & 3413.64 \\ 2386.38 & 1933.96 & 3462.36 & 4943.98 & 1212.8 & 4183.63 \\ 3604.13 & 3123.75 & 1188.21 & 1212.8 & 4256.41 & -444.982 \\ 2123.16 & 3045.68 & 3413.64 & 4183.63 & -444.982 & 5557.27 \end{pmatrix},\right.$$

$$\begin{pmatrix} 8084.55 & 6094.11 & 2564. & 188.429 & 990.411 & 2144.41 \\ 6094.11 & 12031.4 & 7233.86 & 4323.14 & 3591.54 & 6710.25 \\ 2564. & 7233.86 & 4883.43 & 3169. & 2978. & 4478.57 \\ 188.429 & 4323.14 & 3169. & 6100.86 & 3876.57 & 3695.71 \\ 990.411 & 3591.54 & 2978. & 3876.57 & 4385.41 & 1982.41 \\ 2144.41 & 6710.25 & 4478.57 & 3695.71 & 1982.41 & 6987.98 \end{pmatrix},$$

$$\left.\begin{pmatrix} 4805.55 & 1065.25 & 1644.21 & -529.857 & 1458.64 & 1504.63 \\ 1065.25 & 2859.64 & 2622. & 1128. & 1959.29 & 839.679 \\ 1644.21 & 2622. & 3926.57 & 1458.71 & 2318.86 & 2001.07 \\ -529.857 & 1128. & 1458.71 & 2759.14 & 1656.43 & 1374.43 \\ 1458.64 & 1959.29 & 2318.86 & 1656.43 & 2510.29 & 1430.5 \\ 1504.63 & 839.679 & 2001.07 & 1374.43 & 1430.5 & 3137.7 \end{pmatrix}\right\},$$

Variance/Covariance common matrix

$$,\begin{pmatrix} 6388.17 & 4696.23 & 2240.14 & 681.649 & 2017.73 & 1924.07 \\ 4696.23 & 7863.64 & 4181.48 & 2461.7 & 2891.52 & 3531.87 \\ 2240.14 & 4181.48 & 3912.38 & 2696.69 & 2161.69 & 3297.76 \\ 681.649 & 2461.7 & 2696.69 & 4601.33 & 2248.6 & 3084.59 \\ 2017.73 & 2891.52 & 2161.69 & 2248.6 & 3717.37 & 989.31 \\ 1924.07 & 3531.87 & 3297.76 & 3084.59 & 989.31 & 5227.65 \end{pmatrix},$$

Variance ratio Max/Min = , 2.11538}

The two unbiased functions are available as `SkewnessU[list]` and `KurtosisU[list]` following the formulas given in [7].

```
?? SkewnessU
```

SkewnessU[list] computes the unbiased skewness over the list.

```
SkewnessU[list_] := (Sqrt[n (n-1)] Skewness[list])/(n-2) /. n → Length[list]
```

$$\text{SkewnessU[list\_]} := \frac{\sqrt{n\,(n-1)}\ \text{Skewness[list]}}{n-2} \ /.\ n \to \text{Length[list]}$$

```
?? KurtosisU
```

KurtosisU[list] computes the unbiased kurtosis excess over the list.

$$\text{KurtosisU[list\_]} := \frac{((n-1)\,(n+1))\ \text{Kurtosis[list]}}{(n-2)\,(n-3)} - \frac{3\,(n-1)^2}{(n-2)\,(n-3)} \ /.\ n \to \text{Length[list]}$$

The package *MixedDesignANOVA* works in *Mathematica* 4.0 and higher. It is available with this article or can be found at www.mapageweb.umontreal.ca/cousined/home/Others/MixedDesignAnova/Index.html.

References

[1] D. C. Howell, *Fundamental Statistics for the Behavioral Sciences*, 5th ed., Belmont, CA: Thomson–Brooks/Cole, 2004.

[2] D. A. King, "Associative Control of Tolerance to the Sedative Effects of a Short-Acting Benzodiazepine," unpublished doctoral dissertation, University of Vermont, 1986.

[3] J. Cohen and P. Cohen, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

[4] A. L. Edwards, *Multiple Regression and the Analysis of Variance and Covariance*, 2nd ed., New York: W. H. Freeman, 1985.

[5] S. W. Greenhouse and S. Geisser, "On Methods in the Analysis of Profile Data," *Psychometrika*, **24**(2), 1959 pp. 95–112. www.springerlink.com/content/220512x5l554733q.

[6] H. Huynh and L. S. Feldt, "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split Plot Designs," *Journal of Educational Statistics*, **1**(1), 1976 pp. 69–82. www.jstor.org/pss/1164736.

[7] K. Pearson, "Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder," *Biometrika*, **4**(1/2), 1905 pp. 169–212. www.jstor.org/pss/2331536.

S. Chartier and D. Cousineau "Computing Mixed-Design (Split-Plot) ANOVA," *The Mathematica Journal*, 2011. dx.doi.org/doi:10.3888/tmj.13-17.

About the Authors

Sylvain Chartier is a professor of cognitive psychology and quantitative methods at the University of Ottawa. His areas of research include artificial neural networks and nonlinear dynamic systems applied to psychology.

Denis Cousineau is a professor of cognitive psychology at the University of Ottawa. He runs research in artificial intelligence as well as on human categorization processes.

Sylvain Chartier
School of Psychology
University of Ottawa
136 Jean Jacques Lussier
Ottawa, Ontario
K1N 6N5, Canada
sylvain.chartier@uOttawa.ca

Denis Cousineau
School of Psychology
University of Ottawa
136 Jean Jacques Lussier
Ottawa, Ontario
K1N 6N5, Canada
denis.cousineau@uOttawa.ca