Topic 12. The Split-plot Design and its Relatives (Part II)
Repeated Measures [ST&D Ch. 16]

12.9 Repeated measures analysis

Sometimes researchers make **multiple measurements on the same experimental unit**. We have encountered such measurements before and called them **subsamples**; and in the context of the nested designs we have discussed, the primary uses of subsamples are:

1. To obtain a better estimate of the true value of an experimental unit (i.e. to reduce experimental error)
2. To estimate the components of variance in the system.

In these simple nested experiments, "subsample" was not a classification variable but merely an ID. That is, "Subsample 1" from experimental unit 1 was no more similar to "Subsample 1" from experimental unit 2 than it was to "Subsample 2" from experimental unit 2. Even though no two measurements can be made at exactly the same instant in time, we did not refer to these subsamples as repeated measures because we were not interested in characterizing the effect of time between measurements on the response variable.

Now assume that this effect of time is of interest and, again, that the measurements are made repeatedly on the **same experimental unit** (e.g. plant height **over weeks**; yield of perennial crops **over seasons**; animal growth **over months**; population dynamics **over years**, etc.). These observations are not replications because they are made on the same experimental unit (i.e. they are **not independent**). But neither are they subsamples, as defined above, because:

1. We are interested in the effect of time between measurements.
2. "Subsample" is no longer simply an ID: "Subsample 1 from e.u. 1" (e.g. Height of Plant 1 at Week 1) **is** more related to "Subsample 1 from e.u. 2" (Height of Plant 2 at Week 1) than it is to "Subsample 2 from e.u. 2" (Height of Plant 2 at Week 2).

Not replications, not subsamples, these sorts of measurements possess qualities of both and are called **repeated measurements**; and the split-plot model offers one means of analyzing such data. It is important to note in this context that an important assumption of the ANOVA is not the independence of *measurements* but the independence of *errors*. Indeed, the split-plot model implicitly assumes a correlation among measurements.

Before outlining the split-plot approach to repeated measures data, know that time series or multivariate methods can also be used to analyze such data. Time series methods are more appropriate when analyzing long series of data, with more than 20 repeated measures per individual. Consequently, such analyses are more frequently applied to stock data or weather data than to agricultural experiments. Multivariate methods are very useful for shorter time series, and researchers who consistently rely on repeated measures in their experiments should become familiar with them. That being said, the far simpler univariate analysis presented here is also quite useful.

Note that what distinguishes repeated-measures data from any other multivariate data is not so much the existence of the repeated measurements but the desire to examine *changes* in the measurements taken on each subject.

By simply designating the different points in time as "levels" of a "Time" factor, the split-plot principle can be applied to experiments where successive observations are made on the same experimental unit over a period of time. For example, a fertilizer trial or variety trial with a perennial crop like alfalfa might be harvested several times. Other examples might be repeated picking of fruit from the same trees in an orchard or repeated soil sampling of plots over time for nutrient content. In each case, the experimental units to which the treatment levels are assigned are the "main plots," and the several measurements over time are the "subplots." A subplot in this case, however, differs from the usual subplot in that it consists of data taken from *the entire main plot* rather than from some designated portion of the main plot, as is the case with the usual split-plot.

In the simple, univariate split-plot approach to repeated measures data, the covariance structure among the observations within a main plot is used to adjust the degree of freedoms in the ANOVA to give approximate tests. The approximate ANOVA for a repeated measures CRD is shown below.

12.9.1 Repeated measures ANOVA

Approximate ANOVA of repeated measurement analysis

Source	**df**	**SS**	**MS**	**Conservative df**
Among Experimental Units				
Treatment (A)	a-1	SSA	SSA/(a-1)	
*Rep * Trt (Error A)*	a (n-1)	SS(MPE)	SS(MPE)/a(n-1)	
***Within* Experimental Units**				
Time (B)	b-1	SSB	SSB /(b-1)	**1**
Treatment * Time (A*B)	(b-1)(a-1)	SSAB	SSAB/(b-1)(a-1)	**a-1**
Error B	a(b-1)(n-1)	SS(SPE)	SS(SPE)/a (b-1) (n-1)	**a(n-1)**

The analysis looks like a split-plot analysis, except that *conservative degrees of freedom* are used in all F-tests for effect of Time (i.e. the repeated measures) and any interactions with Time. No unusual problems arise when analyzing the effects of the main plot (A) because the analysis of the main plot is insensitive to the split, as seen before in the normal split-plot. However, F values generated by testing the effects of Time (subplot B) and the interaction of main plots treatments with Time (A*B) may not follow an F distribution, thereby generating erroneous results.

The normal split-plot model assumes that pairs of observations within the same main plot are equally correlated. With repeated-measures data, however, arbitrary pairs of observations on the same experimental unit are not necessarily equally correlated. Measurements close in time are often more highly correlated than measurements far apart in time. Since this unequal correlation among the repeated measurements is ignored in a simple split-plot analysis, tests derived in this manner may not be valid.

To compensate for this assumption of uniform correlation across repeated measurements, a **conservative approach** is recommended by many statisticians, "conservative" because it requires larger F values to declare significance for B and A*B effects. In this approach, it is suggested that the degrees of freedom of B (response in time) be used to scale the degrees of freedom for B, A*B, and Error B. Finally, critical *F* values should be used that are based upon these conservative degrees of freedom (previous table, right column).

The uncorrected degrees of freedom are appropriate for independent replications within main plots. The corrected ones (right column) are appropriate for totally dependent replications, a situation equivalent to having all responses represented by a single response (this explains why the corrected df is one). Total dependency is the worst theoretically-possible scenario and is therefore a severe condition to impose. The true level of dependency among repeated measurements in a real experiment will probably be somewhere between these two extremes.

12.9.2. Example of a repeated measurements experiment

An experiment was carried out to study the differences in yield of four alfalfa cultivars. Five replications of these four varieties were organized according to a CRD, and four cuttings were made of each replication over time. The data represents the repeated measurements of yield (tons/acre) of the four cultivars and is analyzed as a split-plot CRD with repeated measures:

4 levels of main plot A: **Cultivars** 1 – 4
4 levels of subplot B: **Cut times** 1 – 4 (9/10/16, 5/18/17, 6/25/17, 9/16/17)

To analyze this data, we begin by carrying out a standard split-plot analysis (the data for this example can be found at the end of this document).

```
rep_meas_mod<-aov(quality ~  A_var +
                             Error(A_var:rep) +
                             B_cut + A_var:B_cut,
                             rep_meas_dat)
summary(rep_meas_mod)
```

The output

```
Error: A_var:Rep
             Df  Sum Sq  Mean Sq  F value   Pr(>F)
A_var         3   2.841   0.9469    7.395  0.00251 **
Residuals    16   2.049   0.1280

Error: Within
              Df  Sum Sq  Mean Sq  F value  Pr(>F)
B_cut          3   37.45   12.483  130.463  <2e-16 ***
A_var:B_cut    9    0.55    0.061    0.636   0.761
Residuals     48    4.59    0.096
```

The next step is to manually adjust the relevant degrees of freedom, "penalizing" ourselves for the lack of true independence among repeated measurements:

Conservative degrees of freedom

```
Error: Within
              Df  Sum Sq  Mean Sq  F value   Pr(>F)
B_cut          1   37.45    37.45  130.463  4.2e-09 ***
A_var:B_cut    3    0.55    0.183    0.636    0.592
Residuals     16    4.59    0.287
```

The *F* test for Time (B) is significant, despite the severe penalty imposed by conservative df. We can conclude that there are significant differences among cuttings. And since no interactions were detected, subsequent analysis of the main effects is appropriate.

Analysis of the main effect of Cultivar:

```
MP_comparison<-HSD.test(rep_meas_dat$quality, rep_meas_dat$A_var,
      DFerror = 16, MSerror = 0.1280)
MP_comparison
```

```
MSerror         HSD
  0.128 0.3236874    <-- using MS(A*rep) as the error

Tukey groups
 A_var         Mean  Grouping
     2     3.269848     a
     1     3.044449     ab
     4     2.900750      b
     3     2.760109      b
```

Analysis of the main effect of Time (subplot B):

```
SP_comparison<-LSD.test(rep_meas_dat$quality, rep_meas_dat$B_cut,
      DFerror = 16, MSerror = 0.287)
SP_comparison
```

```
MSerror       LSD
  0.287 0.3591347

Tukey groups
 B_cut      means  Grouping
     2   4.099587     a
     3   2.985440      b
     1   2.598901       c
     4   2.291229       c
```

By expliciting declaring the DFerror and MSE in the test statement, the pairwise hypothesis tests automatically follow a conservative df approach. Compare this to running the default means separation test:

```
SP_comparison_unadj<-LSD.test(lm(quality ~ A_var + A_var:rep + B_cut +
      A_var:B_cut, rep_meas_dat), "B_cut")
SP_comparison_unadj
```

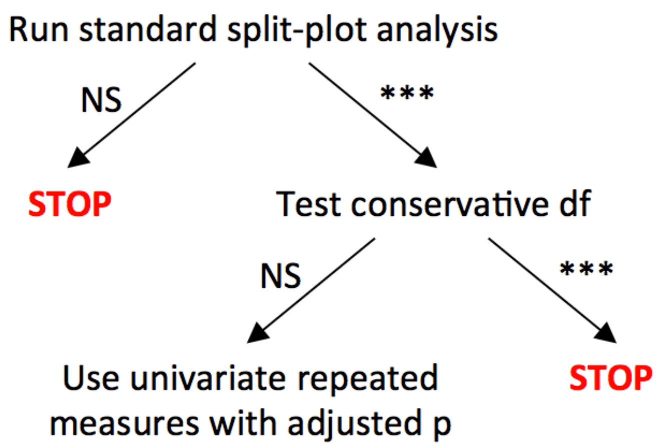
```
      Mean        CV    MSerror       LSD
  2.993789 10.33208 0.09567927 0.1966718

  Df ntr  t.value
  48   4 2.010635

  trt    means M
1   2 4.099587 a
2   3 2.985440 b
3   1 2.598901 c
4   4 2.291229 d
```

Repeated Measures Decision Tree

Below is a schematic illustrating the overall process to follow when dealing with repeated measures data.



Explanation:

1. Analyze the experiment as a simple split-plot design, treating all repeated measurements as true replications of the subplot effect (Time).

2a. If the results of this split-plot analysis are NS, STOP. Conclusion: NS.

You stop here because, if you do not detect differences with this unreasonably liberal analysis, there are no real differences due to Time.

2b. If the results of this split-plot analysis are significant, proceed to a conservative df analysis.

3a. If the results of conservative df analysis are significant, STOP. Conclusion: Significant.

You stop here because, if you detect difference with this unreasonably conservative analysis, there are real differences due to Time.

4. Finally, if the original split-plot analysis (too liberal) is significant and the conservative df analysis (too conservative) is NS, you should proceed with a more nuanced analysis, like the one available through R's ezANOVA() function (see next section).

12.9.3 Univariate repeated measures analysis using the ezANOVA() statement

In the previous example, there was no uncertainty because the adjusted and unadjusted df analyses produced the same conclusions. In some cases, however, adjusted and unadjusted df can produce different results. A more precise test, somewhere between the two extremes of complete independence and complete dependence of measurements, can be obtained using R's ezANOVA() function, designed specifically for this situation.

Repeated measures is a special case of multivariate analysis of variance (MANOVA). In a MANOVA, the four cutting times in the present example would be treated as *four response variables* (not four levels of a classification variable), with only Variety treated as a classification variable. But there is also a univariate approach to analyzing repeated measures data, which is what we will outline here.

The univariate analysis of repeated-measures designs is similar to a split-plot analysis. The split-plot model specifies that pairs of observations on the same experimental unit are equally correlated. With repeated-measures data, pairs of observations on the same e.u. are not necessarily equally correlated. Measurements close in time are usually more highly correlated than those far apart in time. Since this unequal correlation is ignored in a split-plot analysis, the univariate tests derived in this manner may not be valid. However, in cases where many repeated measurements are missing, this might be the only way to analyze the data. In general, univariate tests are more capable of detecting differences than their multivariate counterparts; so a univariate approach to repeated-measures can be a useful strategy.

```
#library(ez)

ezANOVA(
  data = rep_meas_dat,          ← the same data set
  dv = quality,                 ← the column containing the response variable
  wid = rep,                    ← the column containing the e.u. identifier
  within = B_cut,               ← the column containing the repeated measure identifier
  between = A_var,              ← the column containing the main plot identifier
  return_aov = TRUE
)
```

The output, with commentary

1. First, R produces an ANOVA table of "Between Subjects Effects" (i.e. the effects of the Main plot factor Variety) which is equivalent to a full df analysis. Notice that the results we get here match the results from our earlier split-plot analysis *exactly*, and we did not even need to specify the correct error term for the main plot!

	Effect	DFn	DFd	F	p	p<.05	ges
2	A_var	3	16	7.3953319	2.511139e-03	*	0.29959303
3	B_cut	3	48	130.4625662	4.432360e-23	*	0.84936627
4	A_var:B_cut	9	48	0.6360617	7.605683e-01		0.07618871

2. Next, R displays the results of Mauchly's Test for Sphericity:

```
$`Mauchly's Test for Sphericity`
       Effect         W           p p<.05
3       B_cut 0.1127544 5.992606e-06     *
4 A_var:B_cut 0.1127544 5.992606e-06     *
```

As with all statistical tests, results from a repeated-measures analysis are correct *if and only if the assumptions of the analysis are met*. In addition to the standard assumptions of normality of residuals and homogeneity of variances, Mauchly's Test evaluates the additional assumption of uniform correlation among subplot (time) measurements by investigating the orthogonal components of the dataset's covariance structure.

Somewhat simplified, what is being tested is that the measurements conform to a certain pattern. Univariate tests may be too liberal when the data do not conform to the pattern tested by the sphericity test (namely, that the orthogonal components of the data are uncorrelated and have equal variance). If the sphericity test is rejected, as in this case, the message is that remedial steps should be taken. Sticking with a univariate analysis, such a remedial step involves adjusting the degrees of freedom (see next step).

3. Finally comes the results we are looking for. In the next table, R produces a table of Sphericity-Corrected F and p-values for the factors involving time (the "within e.u." effects):

```
$`Sphericity Corrections`
       Effect       GGe        p[GG] p[GG]<.05       HFe        p[HF] p[HF]<.05
3       B_cut 0.4984411 1.391816e-12         * 0.5382286 2.031184e-13         *
4 A_var:B_cut 0.4984411 6.585475e-01           0.5382286 6.694957e-01
```

The first column, labelled "GGe," presents the Greenhouse-Geisser coefficients for deflating the subplot degrees of freedom. What follows is a column of G-G adjusted p-values; that is, p-values that have been adjusted based on the actual patterns of correlation detected in this particular dataset. As we said before, repeated measures are neither perfectly correlated nor perfectly uncorrelated, but somewhere in between. Next comes a column labelled "HFe." This presents the Huynh-Feldt coefficients for deflating the subplot df. The H-F estimator provides tests that are not as conservative as the G-G estimator. We recommend using the more conservative of the two: G-G.

This analysis produces the same F values as the previous split-plot analysis but **corrected** p-values for TIME (previous B_cut) and TIME*VAR (previous B_cut*A_var)

Each of the adjustment methods estimates a quantity known as "Epsilon" and then multiplies the numerator and denominator degrees of freedom by this estimate before determining the significance level for the test, labeled "p[GG]" and "p[HF]" in the output. This reduces the degrees of freedom less drastically than the conservative df approach, which simply divides the degrees of freedom of B, B*A, and Error by the degrees of freedom of B. At the same time, these adjusted probabilities are more conservative than the original split-plot analysis with unadjusted degrees of freedom.

As a final point, it is worth noting that even when corrections are applied to the probability (GG or HF), the univariate approach may be too liberal when the data stray far from sphericity. In a case where sphericity is rejected so dramatically ($p<0.0001$), the corrected univariate test should be interpreted cautiously. In such cases, it is worth considering a multivariate (MANOVA) approach, which makes no assumption about the pattern of correlation among subplot measurements.

Which to use, Split-Plot or REPEATED?

There is no simple answer as to which of these methods is most appropriate. If responses are significant even when the conservative df are used, those conclusions are sound.

In situations where different conclusions are obtained with the adjusted and unadjusted df, one might think that the ezANOVA() function would solve all of the problems. It can solve some problems of correlation, but the interpretation of the results requires a number of assumptions or tests that are beyond the scope of this course. You can use this approach, but *only* after making sure you understand and test all the assumptions of the model.

Missing values: In cases where many repeated measurements are missing, the split-plot approach might offer the only way of analyzing a repeated-measures design because univariate repeated-measures approach will simply remove from the analysis any level of the repeated measures factor where data are missing.

Alfalfa data:

rep	A_var	B_cut	quality
1	1	1	2.80191
1	2	1	2.76212
1	3	1	2.29151
1	4	1	2.56631
2	1	1	2.96602
2	2	1	3.09636
2	3	1	2.54027
2	4	1	2.3163
3	1	1	2.43232
3	2	1	3.09917
3	3	1	2.41199
3	4	1	2.65834
4	1	1	2.93509
4	2	1	2.65256
4	3	1	2.3042
4	4	1	2.47877
5	1	1	2.42277
5	2	1	2.63666
5	3	1	2.36941
5	4	1	2.23595
1	1	2	3.73092
1	2	2	5.4053
1	3	2	3.8114
1	4	2	4.9607
2	1	2	4.43545
2	2	2	3.90683
2	3	2	3.82716
2	4	2	3.96629
3	1	2	4.32311
3	2	2	4.08859
3	3	2	4.08317
3	4	2	3.71856
4	1	2	3.99711
4	2	2	5.42879
4	3	2	3.27852
4	4	2	3.92048
5	1	2	3.85657
5	2	2	3.77458
5	3	2	3.44835
5	4	2	4.02985

1	1	3	3.09856
1	2	3	3.82431
1	3	3	2.92575
1	4	3	2.81734
2	1	3	3.10607
2	2	3	3.26229
2	3	3	2.86727
2	4	3	2.91461
3	1	3	2.8103
3	2	3	3.13148
3	3	3	3.03906
3	4	3	2.92922
4	1	3	2.77971
4	2	3	2.70891
4	3	3	2.72711
4	4	3	3.06191
5	1	3	3.24914
5	2	3	3.09734
5	3	3	2.50562
5	4	3	2.85279
1	1	4	2.50965
1	2	4	2.72992
1	3	4	2.39863
1	4	4	2.05752
2	1	4	2.57299
2	2	4	2.58614
2	3	4	2.16287
2	4	4	2.15764
3	1	4	2.07966
3	2	4	2.60316
3	3	4	2.07076
3	4	4	2.15684
4	1	4	2.44033
4	2	4	2.30163
4	3	4	2.04933
4	4	4	2.35822
5	1	4	2.34131
5	2	4	2.30082
5	3	4	2.0898
5	4	4	1.85736