Supplement to Chapter 10 of *Formulation Simplified*:

Analysis of Combined Split-Plot Coffee Experiment

Chapter 10 describes a combined mixture-process split-plot experiment involving coffee bean blends ground at various sizes, and brewed at varying amounts. While the experimental design is relatively straightforward, the analysis of a split-plot combined design differs somewhat compared to that of a completely randomized design (CRD). In this appendix, we will point out some key differences. Let's take this a step at a time.

First, we'll fit a model to the taste preferences of one of the authors (Martin) using forward selection with AICc—the Akaike (pronounced ah-kah-ee-keh) information criterion-corrected. To put it simply, AIC is an estimator of the relative quality of statistical models. AICc is essentially AIC with a greater penalty for extra parameters, which curbs overfitting of models.

The terms selected for the model are shown in Table S10.1, which lays out the REML (restricted maximum likelihood) ANOVA (analysis of variance). This REML-derived analysis, necessitated by the split-plot error structure, differs little in appearance from that for a completely randomized design (CRD). In any case, the focus remains on the p-values and the usual rule of 0.05 being the threshold for statistical significance. Thus, for the experiment on coffee, the whole plot effects (lower-cased components in the mixture of beans) must be deemed insignificant. However, two of the subplot-grouped effects—bDE and aD^2—turn out to be very significant. Keep in mind, though, that, for illustration purposes, this data encompasses only one of the multiple tasters in the actual experiment.

Table S10.1: Analysis of variance for coffee experiment

Source	Term df	Error df	F-value	p-value
Whole-plot	4	13.89	0.6475	0.6379
Linear Mixture	2	17.30	0.9069	0.4221
ab	1	11.46	0.0130	0.9111
bc	1	11.17	0.3961	0.5418
Subplot	10	47.23	2.35	0.0242
aD	1	47.42	0.0943	0.7602
bD	1	47.99	0.0920	0.7629
bE	2	49.30	1.10	0.3408
cD	1	44.62	0.6037	0.4413
abD	1	47.83	2.38	0.1294
bcD	1	47.29	2.05	0.1586
bDE	2	47.49	5.33	0.0082
aD^2	1	51.98	4.61	0.0365

Table S10.2 presents the model-summary statistics. Most of the statistics are the same as their CRD counterparts, but there are a few key differences. First, the predicted R^2 is omitted, as it involves a bruteforce re-fit that would become computationally burdensome in larger designs. The PRESS statistic is also missing. We recommend making your decisions about model fit using the adjusted R^2.

Table S10.2: Model summary statistics

Std. Dev.	1.29	**R²**	0.3405
Mean	5.28	**Adjusted R²**	0.1336
C.V. %	24.38		

The biggest difference between the randomized and split-plot analyses stems from the sources of variance and error. In a CRD, a single estimate of the experimental error is provided, called the mean square error. This is the error that is introduced into the system from re-setting the factor levels, re-blending each mixture, measurement error, model misspecification, and so on. In a split-plot design, not all factors are changed each run, so we do not have complete independence.

Table S10.3 displays the breakdown of variance components.

Table S10.3: Variance components

Source	Variance	Standard Error	95% CI Low	95% CI High
Group	0.0596	0.2204	-0.3724	0.4916
Residual	1.60	0.3492	1.08	2.58
Total	1.66			

Notice that the experimental error at the whole-plot (group) level is 0.0596 and the error at the sub-plot level is 1.60. This indicates that only about 4% of the 1.66 total error can be attributed to something involving the coffee blends, with the great majority (96%) of the error coming from the grinding, weighing out the grounds, and measurement error. If we wanted to reduce the noise in the system, it's clear where we would see the biggest impact: the processing of the coffee, not the blending of it.

Hard-to-change terms are tested against the Group variance, while the easy-to-change (ETC) terms are tested against the residual error. Notice that, unlike for the CRD, the terms in the REML ANOVA (shown previously in Table S11.1) have differing degrees of freedom (DF), with subplot (ETC) terms having more DF.

Conclusion

A split-plot design is really like two designs in one: a whole plot design, with a sub-plot design run within each whole-plot run. Hence, it must be analyzed as such. In some cases, you may get away with incorrect analysis that's still reasonable. In other cases, however, analyzing a split-plot design as a completely randomized design will give you incorrect p-values, and will causing an increase in false alarms, generally resulting in a model that's too big and doesn't predict well.

OTHER EXPERIMENTS ON COFFEE

In 1920, Samuel Prescott, Dean of Science at MIT, got a $40,000 grant from industry to establish the Coffee Research Lab to evaluate safety and optimize taste of this newly popular beverage. The New York Times reported that a Minnesota man drank 80 cups of coffee in 8 hours and survived in "pretty good shape." Prescott fed rabbits enormous quantities of coffee and found their health unimpaired, but it made them much jumpier than usual (pun intended). The best coffee, according to the MIT experts, was made freshly ground by drip into glass or ceramic pots with water just below a boil. In his younger days as a biology professor, Prescott also searched for "growth producing" rays that would "bring forth cows the size of brontosauri." That would've satisfied demand for those liking milk in their "moo"cho grandé coffee.

(Source: "Engineering the Perfect Cup of Coffee" by Larry Owens, *Technology and Culture*, Oct. 2004, V.45 pp795-807)