# Supplement to Chapter 9 of *Formulation Simplified*: <u>Categoric Components Going to Zero</u>

In Chapter 9 we laid out an experiment on an aerospace composite that not only optimized the recipe, but also settled on which materials to use. A special case of this combined mixture design arises when the range of categorical materials includes zero, that is, the experimenter removes it from the formulation. The standard crossed model produces nonsensical predictions at this point—showing differences due the type of component, even though none of it has been added to the mixture!

### *A case that illustrates the problem*

To illustrate the problem and provide a work-around, consider a food-chemical case with four components totaling to 1100 milligrams (mg). The fourth ingredient, a preservative, can be omitted from the formula (zero mg) or, if added, chosen from two alternatives. For their first try, the food chemists set up an optimal design to fit a linear mixture model crossed with a main-effect model on the categorical factor. To augment the experiment, they add 5 (unique) check-points and replicate 5 of the combinations. Table S9.1 lays out the resulting 18-run design in random order. It includes eight runs with no preservative (e.g., #s 10 and 15, highlighted). Note how it, nevertheless, spells out the preservative type (SO2-sulfur dioxide and Ca2-calcium propionate). That makes no sense—when the preservative level is zero the type becomes irrelevant.

**Table S9.1: Food chemical case—selected runs with zero preservative**

| **Run** | **A:** NaCl (mg) | **B:** KCl (mg) | **C:** BHT (mg) | **D:** Preservative (mg) | **e:** Preservative Type | **Y** Degradation (rate k) |
|---|---|---|---|---|---|---|
| 1 | 0 | 975 | 5 | 120 | SO2 | 0.1973 |
| 2 | 0 | 1000 | 0 | 100 | Ca2 | 0.2120 |
| 3 | 100 | 1000 | 0 | 0 | SO2 | 0.1863 |
| 4 | 0 | 973.33 | 120 | 6.67 | Ca2 | 0.2063 |
| 5 | 160 | 816.67 | 120 | 3.33 | SO2 | 0.2095 |
| 6 | 0 | 1000 | 0 | 100 | Ca2 | 0.2156 |
| 7 | 600 | 500 | 0 | 0 | Ca2 | 0.1937 |
| 8 | 600 | 500 | 0 | 0 | Ca2 | 0.2052 |
| 9 | 320 | 660 | 0 | 120 | SO2 | 0.1916 |
| ***10*** | ***320*** | ***660*** | ***120*** | ***0*** | ***SO2*** | ***0.2044*** |
| 11 | 560 | 500 | 0 | 40 | SO2 | 0.1853 |
| 12 | 160 | 820 | 120 | 0 | Ca2 | 0.1943 |
| 13 | 240 | 740 | 0 | 120 | Ca2 | 0.1977 |
| 14 | 470 | 500 | 65 | 65 | Ca2 | 0.1996 |
| ***15*** | ***320*** | ***660*** | ***120*** | ***0*** | ***Ca2*** | ***0.2169*** |
| 16 | 100 | 1000 | 0 | 0 | SO2 | 0.2035 |
| 17 | 470 | 500 | 65 | 65 | Ca2 | 0.1978 |
| 18 | 560 | 500 | 0 | 40 | SO2 | 0.1873 |

For these no-preservative runs, the food chemists considered ignoring factor "e" (shown in small case to differentiate this being categorical). But, before spending time and materials, they simulated degradation rate results—the data listed in the last column (reaction rate coefficient "k") —to see what repercussions might ensue.

It was good they did because, per our alert at the outset of this supplement, the predictions with no preservative (D = 0) differ, depending on which (phantom) type is selected. This can be seen in Figure S9.1a versus S9.1b.

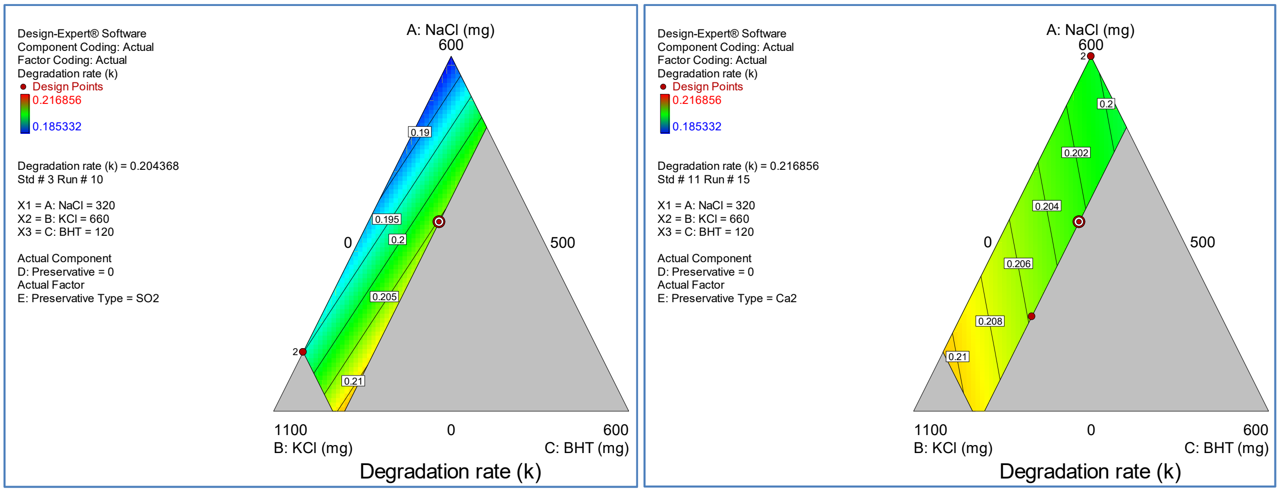


**Figure S9.1a,b: Contour graphs with one type of preservative versus the other with none of it in the mixture**

This nonsensical situation stems from the crossed model (below) including second-order terms involving "e" (shown in brackets) that do not zero out (as they should) when component D is absent.

$Y_i = 0.193\ A + 0.205\ B + 0.232\ C + 0.197\ D + [0.007\ Ae] + [0.008\ Be] – [0.038\ Ce] – 0.002\ De$

Plugging in the coded values for runs 10 and 15 (highlighted in Table 1), which lay out the same compositions (coded 0 to 1), but with differing types of preservative (coded -1 to +1), produces the following results. The terms involving D drop out due to it being at zero.

$Y_{10} = 0.193\ (0.533) + 0.205\ (0.267) + 0.232\ (0.2) + 0.007\ (0.533)\ (–1) + 0.008\ (0.267)\ (–1) – 0.038\ (0.2)\ (–1)$

$= 0.202$ for SO2

$Y_{15} = 0.193\ (0.533) + 0.205\ (0.267) + 0.232\ (0.2) + 0.007\ (0.533)\ (+1) + 0.008\ (0.267)\ (+1) – 0.038\ (0.2)\ (+1)$

$= 0.205$ for Ca2

These predictions differ (wrongly) by the minus versus plus offset between the two types of preservative.

## MAC 'N CHEESE WITH INGREDIENTS MISSING

Mark's son Hank provides a case with similar issues to what's spelled out in this supplement. While in college, he and his roommates lived largely on macaroni and cheese ("mac 'n cheese") that tasted far better with butter and milk. However, sometimes they needed to substitute margarine for butter, and, on occasion, they had no source of oil. This provides the fodder for an experiment along the same lines as that run by the food chemists on preservatives.

*"If you can't taste an ingredient, you have to ask yourself why it is there."*

- Yotam Ottolenghi, renowned chef and cook-book author, quoted by *The Telegraph*, Jan 2009, www.telegraph.co.uk/foodanddrink/recipes/4159460/More-recipes-from-Yotam-Ottolenghi.html.

## *The problem solved and experiment re-done with custom design for correct model*

The following steps provide the remedy by creating a model where "e" never appears without "D":

1. Start with the mixture model (Scheffé polynomial), in this case one that includes non-linear blending terms (e.g; AB):
$\hat{y} = \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD$
2. Add mixture terms with "D" crossed with "e" (shown in square brackets):
$\hat{y} = \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD$
$+ [\beta_{45} De]$
3. For all components other than "D", add first and second order (non-linear) terms crossed with "De":
$\hat{y} = \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD$
$+ \beta_{45} De + [\beta_{145} ADe] + [\beta_{245} BDe] + [\beta_{345} CDe] + [\beta_{1245} ABDe] + [\beta_{1345} ACDe] + [\beta_{2345} BCDe]$
4. Combine like terms (not necessary in this case).

This procedure produces a sensible model that predicts the same result for each level of the categoric component when it is absent (D=0).

Given these insights, the food chemists re-design their experiment I-optimally to the 17-term customized model laid out above. Bolstered with 5 check-points and 5 replicates, the runs come to a total of 27 as shown in Table S9.2. (Disregard the specification for the SO2 preservative (sulfur dioxide) at zero preservative—this type being arbitrarily entered as a placeholder.)

**Table S9.2: Food chemical case—a better experiment design**

| Run | A: NaCl (mg) | B: KCl (mg) | C: BHT (mg) | D: Preservative (mg) | e: Preservative Type | Degradation k |
|---|---|---|---|---|---|---|
| 1 | 455 | 525 | 0 | 120 | Ca2 | 0.2134 |
| 2 | 199 | 771 | 64 | 66 | Ca2 | 0.1859 |
| 3 | 328 | 711 | 0 | 61 | Ca2 | 0.2575 |
| 4 | 349 | 681 | 11 | 59 | SO2 | 0.1747 |
| 5 | 14 | 1000 | 47 | 38 | SO2 | 0.2505 |
| 6 | 485 | 500 | 10 | 105 | SO2 | 0.1699 |
| 7 | 0 | 1000 | 47 | 53 | Ca2 | 0.2436 |
| 8 | 478 | 525 | 46 | 51 | Ca2 | 0.1798 |
| 9 | 269 | 831 | 0 | 0 | SO2 | 0.2592 |
| 10 | 480 | 500 | 120 | 0 | SO2 | 0.2388 |
| 11 | 132 | 841 | 7 | 120 | SO2 | 0.1461 |
| 12 | 135 | 917 | 48 | 0 | SO2 | 0.2624 |
| 13 | 0 | 1000 | 0 | 100 | SO2 | 0.2266 |
| 14 | 121 | 919 | 60 | 0 | SO2 | 0.2663 |
| 15 | 349 | 681 | 11 | 59 | SO2 | 0.1552 |
| 16 | 483 | 617 | 0 | 0 | SO2 | 0.2839 |
| 17 | 269 | 831 | 0 | 0 | SO2 | 0.2648 |
| 18 | 561 | 539 | 0 | 0 | SO2 | 0.2528 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 19 | 318 | 652 | 66 | 64 | SO2 | 0.1139 |
| 20 | 223 | 747 | 120 | 10 | SO2 | 0.2110 |
| 21 | 600 | 500 | 0 | 0 | SO2 | 0.3014 |
| 22 | 131 | 849 | 0 | 120 | Ca2 | 0.2738 |
| 23 | 478 | 525 | 46 | 51 | Ca2 | 0.2203 |
| 24 | 12 | 958 | 120 | 10 | SO2 | 0.2057 |
| 25 | 0 | 1000 | 47 | 53 | Ca2 | 0.2430 |
| 26 | 318 | 717 | 65 | 0 | SO2 | 0.2452 |
| 27 | 14 | 1000 | 47 | 38 | SO2 | 0.2314 |

When modeling the results, a few considerations must be made for the categoric component going to zero:

- Do the fitting in "Real" coding, where zero represents true absence. (In the usual coding (pseudo), zero is the lowest level of a component—not necessarily absent.)
- Do not maintain model hierarchy for second-order terms involving mixture components crossed with the categorical factor. For example, if ADe is selected, do not support it with Ae. However, if ABDE is selected, terms AB, AD, BD, DE, ADE and BDE must be in the model to provide hierarchy, but not Ae or Be, because "e" cannot be in a term without D because then it does not 'zero out'.

After fitting the data in Table S9.1 using Real coding to the design model and then reducing it using backward selection on p-values with alpha of 0.1, the degradation rate is predicted by an equation, significant at $p<0.0001$, with the following terms: A, B, C, D, AB, AD, BD, CD, DE, ADE, BDE, ABDE. Figure S9.2 shows side by side the response surfaces for one type of preservative versus the other with none of it in the mixture (D=0). This time around they match (as they should). Now the type of preservative makes no difference when the level is zero. ☺

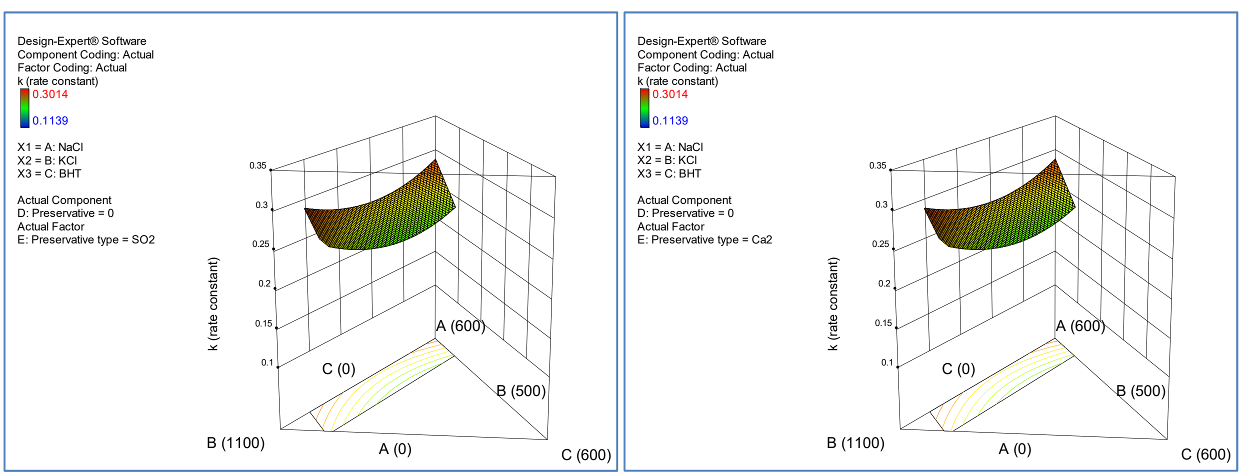


**Figure S9.2a,b: 3D graphs with one type of preservative versus the other with none of it in the mixture**

This concludes our case except for one unanswered question: What is the optimal formulation to preserve the food, that is, minimize the degradation rate? The answer is provided in Figure S9.3 in ramps view.

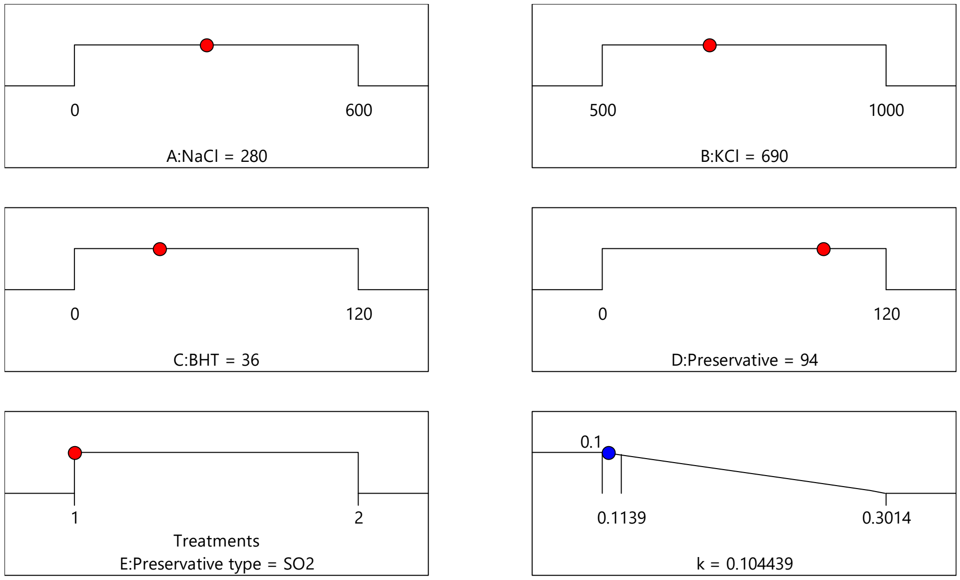


**Figure S9.3: Optimal formulation for minimizing degradation**

Note that, unsurprisingly, a relatively large proportion of preservative (D) works best. Also, of the two types, the sulfur dioxide (SO2) works best.

Mission accomplished!