

3. Multifactor Experiments

Staying with the same experiment, now think about multifactor experiments.

1. What is the full experimental design for this experiment? Tell us the dimensions, such as 2x2x3. (Hint: the full results appear in Panel 4B.)

This is a 3x3 dimension experiment with the following axes: 3 bin attributes [bin with or without sticker or no bin at all] and 3 attributes for SMS [personalized sms, generic sms or no sms at all].

The `havecell` is not included in the design dimensions because it is a pre-treatment covariate which is not included in the administration of treatment to subjects; a.k.a. not an intervention (i.e. designers cannot control who has a cell or not – similar to how they are giving out bins and text messages).

2. In the results of Table 4B, describe the baseline category. That is, in English, how would you describe the attributes of the group of people for whom all dummy variables are equal to zero?

All dummy variables equal to zero in the case of results depicted in Table 4B, represent households in the control group (i.e. households with no treatment variables for receipt of recycle bins and no SMS messages; along with households that have no cells).

Therefore this baseline category of households in control group represents the intercept.

3. In column (1) of Table 4B, interpret the magnitude of the coefficient on “bin without sticker.” What does it mean?

In column (1) of Table 4B, we find that households that received a bin but without any sticker were 3.5 percentage points more likely to turn in recyclables over the control group.

4. In column (1) of Table 4B, which seems to have a stronger treatment effect, the recycling bin with message sticker, or the recycling bin without sticker? How large is the magnitude of the estimated difference?

Recycling bin with sticker. Because of the 2% excess coefficient for recycling bin with sticker – and same level of statistical significance in both cases (as seen in the standard errors of 0.015 in both cases).

5. Is this difference you just described statistically significant? Explain which piece of information in the table allows you to answer this question.

Difference between both treatment effects is 2%. We know that the standard error for the difference between both treatment effects is the sum of the standard errors of individual treatment effects. Therefore, the standard error of the difference between both treatment effects is 0.06. We see that this is not less than 2 standard deviations from the difference in the coefficients, and hence not significant.

Another approach of checking the statistical significance of the difference between both treatment effects (with and without sticker) would be to conduct an T-test (or two-sides T-test) with a sharp null hypothesis that both treatments are the same.

Table 4b already provides us with a p-value for such an F-test $1 = 2$ (at the bottom). The p-value for F-test $1=2$ is 0.31 which indicates the differences in treatment effects are not significant at the 5% significance level.

6. Notice that Table 4C is described as results from “fully saturated” models. What does this mean? Looking at the list of variables in the table, explain in what sense the model is “saturated.”

Table 4C represents every distinct combination of bin and SMS message treatment; the omitted category being no phone and receiving no bin.