BIA 654 Homework 9

Recall the case study on "*Direct Mail Sales*," using Plackett-Burman (PB) design; see recent lecture slides numbered 11–18 and the published paper "*Experimental design on the front lines of marketing: Testing new ideas to increase direct mail sales*" uploaded on Canvas.

(a) Consider the Plackett-Burman design in Table shown on slides 14 and 15. Check that the design is indeed orthogonal for the main effects. That is, pick any two factors, say A and F, and check that the design includes 5 runs at each of the four factor-level combinations. (Just check this for yourself and no need to put your answer here.)

(b) Reanalyze the data shown on slide 15 by the Plackett-Burman design; Excel data file is uploaded. More specifically,

  i. Produce the normal probability plot of 19 main effects and see what factors seem to be significant.

  ii. Produce also the Pareto plot as shown on slide number 16. Using JMP: DOE-Effect Screening will do the work. It will also produce all effect estimates that you will need in part iii below.

  iii. Screen out the 19 factors by examining their respective 95% confidence intervals.

• Here, I provide some details for part iii; see also page 314 in the uploaded paper. Notice that the response variable is a *proportion* (in percent!) of customers who have responded (accepted the offer) to a particular mail sale.

- Note that there are $N = 20$ test cells (i.e., versions of mails), and each of the 20 test cells was sent to $n = 5,000$ people, resulting in the response rates listed in the last column. Therefore, the total $nN = 100,000$ number of people have received mails in this study.
- For each effect, $nN/2 = 50,000$ people will see the + level, and another $nN/2 = 50,000$ people will see the − level. Then the "effect" is the difference,

$$p_+ - p_-$$

  where p_+ is the proportion of responses resulting in orders among the $nN/2$ subjects exposed to the + level, and p_- is the proportion of responses resulting in order among the $nN/2$ subjects exposed to the − level. That is, p_+ = (total number of orders in the + level group)/50,000.
- Now, what is the standard error of effect?

$$\begin{aligned} \text{standard error (effect)} &= \sqrt{\text{estimate of Var}(p_+ - p_-)} \\ &= \sqrt{\text{estimate of Var}(p_+) + \text{estimate of Var}(p_-)}. \end{aligned}$$

  Let π_+ be the *true* (unknown) proportion of successes for customers exposed to the + level. Then,

$$\text{Var}(p_+) = \text{Var}\left(\frac{\text{Binomial}(nN/2, \pi_+)}{nN/2}\right) = \frac{(nN/2)\pi_+(1-\pi_+)}{(nN/2)^2} = \frac{\pi_+(1-\pi_+)}{nN/2}.$$

- Finally, from the above and under the null hypothesis $H_0 : \pi_+ = \pi_-$ (saying there's no effect),

$$\text{standard error (effect)} = \sqrt{\frac{\bar{p}(1-\bar{p})}{nN/2} + \frac{\bar{p}(1-\bar{p})}{nN/2}},$$

  where $\bar{p}$ is the *overall* success proportion, pooled over all runs; it is an estimate of $\pi_+ (= \pi_-)$. That is, in the case study, $\bar{p} = 1.298\%$ by noting there are 1,298 people placed an order (positive response) out of $nN = 100,000$ people.
- You can use $z_{0.05/2} = 1.96$ in this case due to large sample size (instead of t-critical values). Recall the confidence intervals formed this way can be used to *reject* the null $H_0 : \pi_+ = \pi_-$ if the interval does *not* contain zero.

• **Caution:** When you construct confidence intervals, be careful about the unit, e.g., if the response is in percent %, then standard error should also be in percent unit.