

## Question Q&A 2 – Tuesday Nov 9<sup>th</sup>

### *Questions Week 1 – Uncertainty*

- [Efe];

In the tutorial 1, under the Posterior for Bayes' Rule, when I am trying to run function below:

```
"test_data %>%  
group_by(group) %>% summarize(mean=mean(time_on_site), sd=sd(time_on_site), n=n())"
```

I am having an error, that says ;

"Error in group\_by(., group) : object 'test\_data' not found", and I cannot fix it because there is also no, data frame or data set called like that, so can you assist me with this?

- [Lucrezia];

I just checked the results of the assignment 1 and I would like to ask clarifications about the first question. Indeed, during class you said that, after doing the calculations, if the expected profit is negative, the value of not doing the test is 0.

So, 0 was my answer but it's not correct. I would like to understand it better, maybe I missed something.

- [Meike];

For assignment 1, question 4, an answer with decimals was evaluated as being false. I get that you can only have an integer amount of people, however, theoretically you should have 435.374 people. It was not stated anywhere whether you should give the practically or theoretically correct answer. I encountered the same problem last year in another course, where I did gave an integer for something that can only take integer values, but they reasoned that theoretically it should be the decimal number.

Therefore my question is whether the decimal number can also be evaluated as correct and most importantly, I would like to ask how I know what kind of answer is wanted, also across different courses?

## Questions Week 2 – RFM

- Questions from multiple people;

In the practice quiz, why do you choose to make 3 groups at the R, F and M instead of 5 as done in lecture? How do we determine how many groups to compute if it is not given in the question?

What exactly is the difference between 'quantcut' and 'ntiles'? I used 'quantcut' in order to group the customers on Recency. However, in the solutions of the Practice Quiz 'ntiles' is used, which provides a different outcome.

- [Sterre];

If you want to calculate the ROI for a segment or group, should you multiply the cost and profit with the amount of people in that group?

For example,  $(mp-c) \cdot (\text{number of people})$  for the profit for that particular group. Then sum over all groups to get the total profit.

- [Jorg];

In question 3 of Assignment 2 it says: "You can now compare the test response rate and the rollout response rates (for the 71 segments that were rolled out to). How well do test rates predict roll rates?"

Does this mean: you compare the top 71 segments from the test dataset (while only 63 segments have a response rate) and the 71 of the rollout dataset? Or does this mean: you compare all 126 segments from the test dataset with the 71 in the rollout dataset?

- [Mieke];

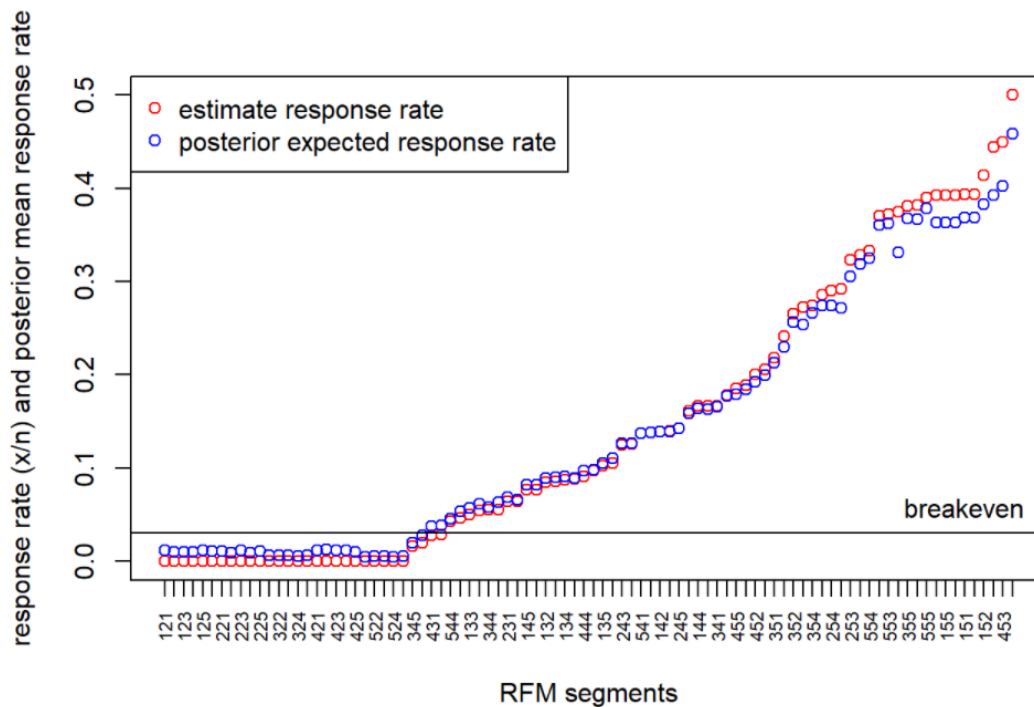
1. If we use  $E[p_z]$  (as in slide 29) in order to decide which segments we will target, do we also have to use this  $E[p_z]$  for deciding upon the profit margin (i.e.  $\text{profit margin} = m \cdot E[p_z] - c$ ). Or should we use the 'standard'  $p_z = s_z / n_z$  (as in slide 17) for deciding upon the profit margin such that  $\text{profit margin} = m \cdot p_z - c$  (even though we used  $E[p_z]$  to decide which segments we will target)?
2. I have some troubles understanding the difference between the test group, rollout group and target group. For example, if we rollout to a certain group, isn't this equal to the group we target? If not, what exactly does this rollout group mean? Could you clarify this for me?
3. For Question 7 of the assignment, they ask us to calculate the return on investment. Is it correct that we should give this answer as  $(ROI \cdot 100)$ , where  $ROI = \text{total profit} / \text{total cost}$ ? Or should we not multiply by 100 and just answer ROI?

4. Also, in question 7 it is stated: 'For the 71 segments we observe rollout outcomes for, calculate the return on investment of targeting using the posterior mean. Use the data on the actual number of customers mailed and responses, along with margin and cost per marketing, in your calculation.'

I was not sure whether we are asked to first determine which of these 71 segments we target (by use of  $E[p_z] > brk$ ) or that we should certainly rollout/target to all these 71 segments. This impacts the cost and profit needed to derive the ROI. I think this question is related to the one about my understanding of the difference between target group, rollout group and test group.

- [Noa];

1. Regarding the following plot:



In the lecture, it is argued that the red dots are shrunk to the mean (here, 0.14). However, in the Lab Session Videos, it is argued that the blue dots are shrunk to the mean. Therefore I was wondering which one is shrinking?

2. Regarding the practice quiz, question 8:

I understand how we create the coefficients  $a$  and  $b$  and assign values to them, how we create a table, how we create a plot and how we check whether the values are  $\geq brk$ . However, there is one part of the code that I don't understand:

```
tbl[agg[,1]==353,]  
  
ind = (tbl[,4]<brk)*(tbl[,5]>=brk)  
  
tbl<-cbind(tbl,ind)  
tbl
```

Could you explain what this part of the code does exactly?