

IMC 463: Webcare Mini-Case

Professor E.C. Malthouse

Spring, 2021

Due: May 8, Noon CST.

See **webcare.csv**. I'm working on a project studying *webcare*, which is when an organization responds in to customer reviewers. This is an exploratory study where we have 7 hotels located in the city center of a mid-size European city (prior to COVID). The main question you should address is, **How should a hotel respond to customer reviews to maximize future bookings from other consumers reading the dialogue?** We know their bookings for almost 4 years. We are partnering with an on-line travel agent (OTA) that collects and displays customer reviews and the local hotel association. Each row in the data is a hotel*week combination. We know the number of bookings in the week (**bookings**) on the OTA. The **nextbook** variable is the number of bookings in the next week and will be our dependent variable. We want to know how **nextbook** is affected by different webcare strategies after controlling for some possible confounds. The variable **respond** is a dummy indicating whether or not the hotel responded to comments (webcare) during the week. We have used extensive text mining (starting with translating 40+ languages to English) to measure different how the hotel responds, including the following **webcare variables**:

- **tailoring**: webcare response mentions specifics about the review, e.g., if the review mentions the quality of the bed and the response mentions the quality of the bed that is tailoring.
- **defensive**: argue that the events described by the dissatisfied customer are untrue, e.g., "We are sorry that you considered our hotel too far from the city center. Our guests usually don't mind to walk there from the hotel, so I guess this is a matter of opinion" or "I was surprised that you found our rooms too small because they are 25m2"
- **invitecontact**: The invitation for another visit is when hotels invite their guests to return to the hotel, making their guests feel like they are welcome, e.g., Thank you for your comment. I'm glad that you enjoyed your stay. We hope to have you again in our hotel in the future!
- **explain**: explaining details regarding a specific issue mentioned by the review, e.g., construction across street make rooms noisy, or "Thank you for your review. I am sorry that you couldn't find the remote, but there was a remote available in the room. It was inside the drawer in the bedside table."
- **nonverbal**: includes non-verbal communication like emojis, ALL CAPS, emoticons, onomatopoeia, sound stretching, punctuation to pass a message.
- **apology**: apologies to reviewer, e.g., has words accept, accident*, acknowledg*, admit, agree, apolog*, approv*, assert, conced*, confess*, excus*, forgiv*, guilt, guilty, pardon, recogni*, sorry

- **compensation**: offers compensation to reviewer, e.g., has words *compensat**, *offer**, *recover**, *refund**, *reimburs**, *repay**, *restor**, *return**
- **chanchange**: channel change, e.g., reviewer invited to email or call hotel. Response has words like *call*, *chat**, *contact*, *correspondence*, *email*, *e-message*, *letter*, *mail*, *offline*, *phone*, *ring*, *visit*
- **gratitude**: offering gratitude/thanks to reviewer, e.g., *thank*, *grateful*, *appreciat**
- **info**: inquires further information about a complaint, indicated by the presence of question marks.
- **personalize**: response is personalized with reviewer's name, e.g., *Dear Mary* rather than *Dear Guest*
- **signame**: signed with a person's name
- **sigdepart**: signed by some department at hotel, e.g., "social media team"
- **sighotel**: signed by "hotel X"
- **sigmanager**: signed by "manager" (note that if response is signed by "Joe Doe, Hotel Manager" then both **signame** and **sigmanager** are incremented)
- **sigstaff**: signed by staff, e.g., front desk

I have logged all of these variables for you. When fitting models use this code:

```
library(car)
fit = glm(nextbook ~ logbook + time + respond + factor(hotelid)
+ <add your vars>, family=poisson, data=newdat2)
summary(fit); vif(fit)
```

Control variable **logbook** is the log number of bookings this week, **time** as the week number (I originally had a flexible spline on this to account for trends, seasonality, etc., but you can just use a linear term), and **hotelid** accounts for systematic difference between hotels that may not be captured by (lag) bookings. Note that the default link for a Poisson model is log, so your coefficients give the effects on the log mean of bookings next month. I originally fitted a model with all webcare variables, but the VIFs were very large. We have to do something to reduce the number of predictors and/or dimension of the data.

1. Study a correlation matrix and factor analysis of the five "sig" variables. Are there variables you are comfortable combining?
2. Study a correlation matrix of the remaining 11 "non-sig" webcare variables. What do you learn about the variables from this analysis and could combine some of the variables? We would like to retain as many original variables as possible in the analysis without having the VIFs explode (e.g., greater than 9 or 10).

3. Find correlations between all webcare variables and `nextbook`.
4. Fit a regression with all of the original variables.
5. Fit a regression model using any combined variables you've created in parts (a) and (b), possibly with some original webcare variables that are not included in factors.
6. Fit any other models you think are necessary to have robust findings, and report your favorite model. Hint: I tried stepwise and also a lasso model picking λ with cross validation, but do your own thing.
7. Make a table comparing the conclusions from your models (one row per webcare variable, one column per model, then show sign of slope +, - or 0).
8. Write a short paragraph summarizing your recommendations to the hotel association on how to respond. These are non-technical people, so keep it in English.