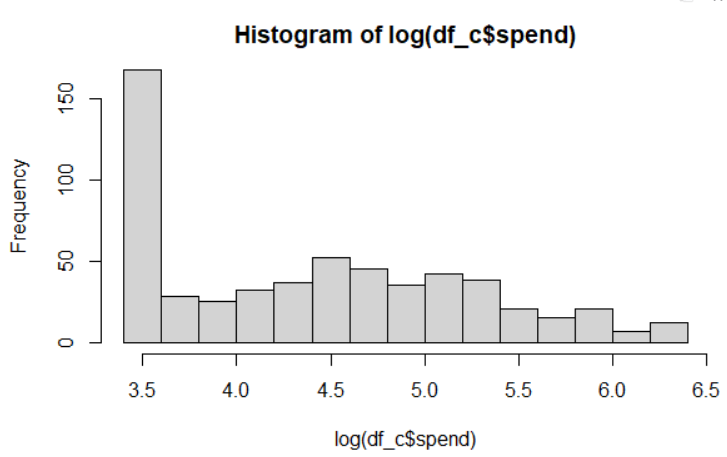
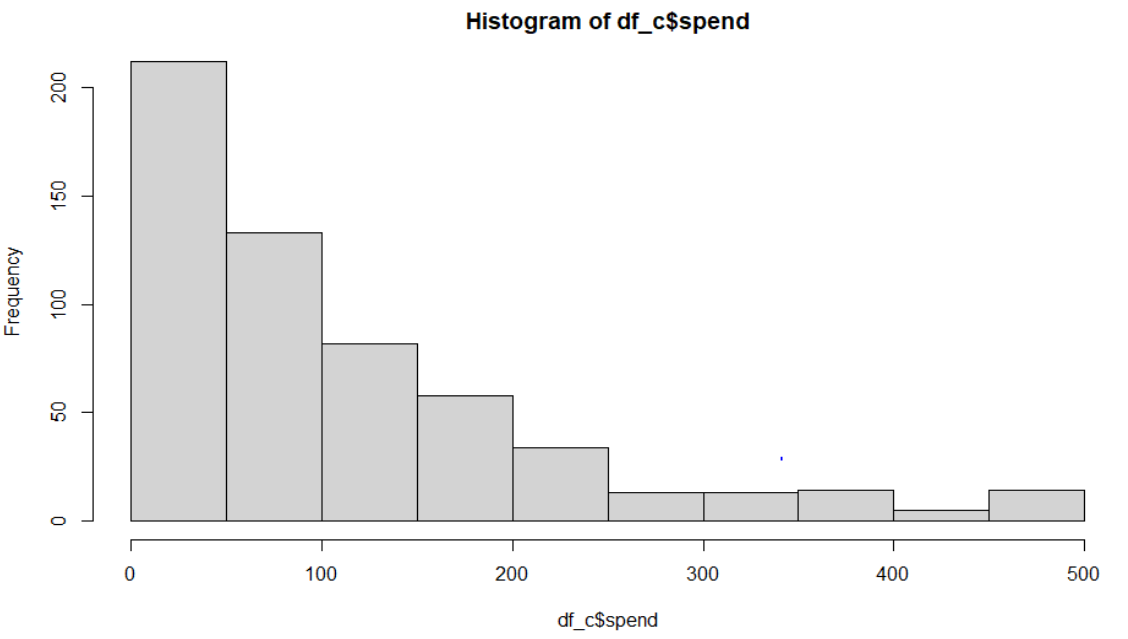
**Online Retail Promotions**

1. **Examine the "spend" variable that we want to predict and explain step-by-step what you would do to create a model to explain customer spend (bullet points are fine). What model(s) is(are) appropriate for this analysis and why. Run appropriate visualizations if necessary and document your work in your answer. Be sure to read Question 4 below to get a sense of the analysis the client is looking for.**



* Spend variable is the amount of dollars spent in two weeks by a person. The distribution looks like it follows a poisson distribution even doing a log of the variable does not make it normal. This suggests that the underlying distribution follows a poisson distribution which could be like the quantity of products bought by customers. So, I will round the amount of money spent to the nearest integer to make it discrete.
* I would run a poisson model and then check for its dispersion.
* If the dispersion is high then a negative binomial model or quasipoisson model would be appropriate.

1. **Create a table of predictors for our dependent variable, listing all relevant predictors, the sign of their hypothesized effects, and a short 1-sentence rationale for each effect.**

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Sign** **of effect** | **Rationale** |
| recency | - | Customers who recently purchased would be spending more |
| History | + | Customers who have a history of spending more will likely spend more even after the promotions. |
| zipcode | +/- | Customers living in Urban areas are likely to spend more compared to the ones in suburban or rural areas. |
| Newcustomer | +/- | Customers who are older would tend to spend the same way and customers who are new could spend more or less. Not sure about this but I want to keep it to see its effect |
| channel | Multi - +  Phone - ?  Web - ? | Customers who use both phone and web would be spending much than customer who just use phone or the web. |
| campaign |  | This variable cannot be used be used it by itself. Interacting it with merchandise would make more sense |
| Merchandise(“Men” – customers who purchased mens merchandise, “Womens” – customers who purchased womens, “both” – customers who purchased both) | Both - +  Men - +  Women - ? | Men would purchase more online compared to women. Customers who spend on both could be couples so their spending would be much higher. |
| History\*merch |  | The effect of historical purchases on amount spent depends on the type of merchandise purchased. |
| History\*channel |  | The effect of historical purchases on amount spent could depend on the type of channel used to purchase. |
| Merch\*campaign |  | The amount spent could increase if the right customers are target so men should receive men’s promotions and vice versa. |

1. **Run alternative models to test for the effects of the hypothesized predictors. Be sure to test the assumptions of these models and modify them as necessary. Present the best 3 models and their output in a nice, compact manner. Also justify your choice of these models and include your assumptions testing results.**
   1. quassipoisson <- glm(spend ~ history + recency + merch + zipcode + newcustomer + channel + campaign + history\*merch + history\*channel + merch\*campaign, family = quasipoisson(link = log), data=df\_c)
   2. nbinom1 <- glm.nb(spend ~ history + recency + merch + zipcode + newcustomer + channel + campaign, data=df\_c)
   3. nbinom2 <- glm.nb(spend ~ history + recency + merch + zipcode + newcustomer + channel + campaign + history\*merch + history\*channel + merch\*campaign, data=df\_c) – **BEST**

stargazer::stargazer(nbinom1,nbinom2,quassipoisson, type="text", title="Model Comparison of negative binomial models for amount spent")

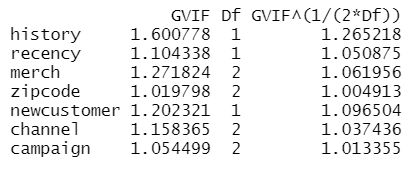
Model Justification:

I decided to use the negative binomial model or quasipoisson model because the dispersion was high on the spend variable.

Although it is recommended to do negative binomial model if the dispersion parameter is much larger than 1 in the quasipoisson model. I decided to use both to check if my model estimates are stable across my models. I used a negative binomial model with interaction terms and without interaction terms to determine how the model does with and without the interaction terms.

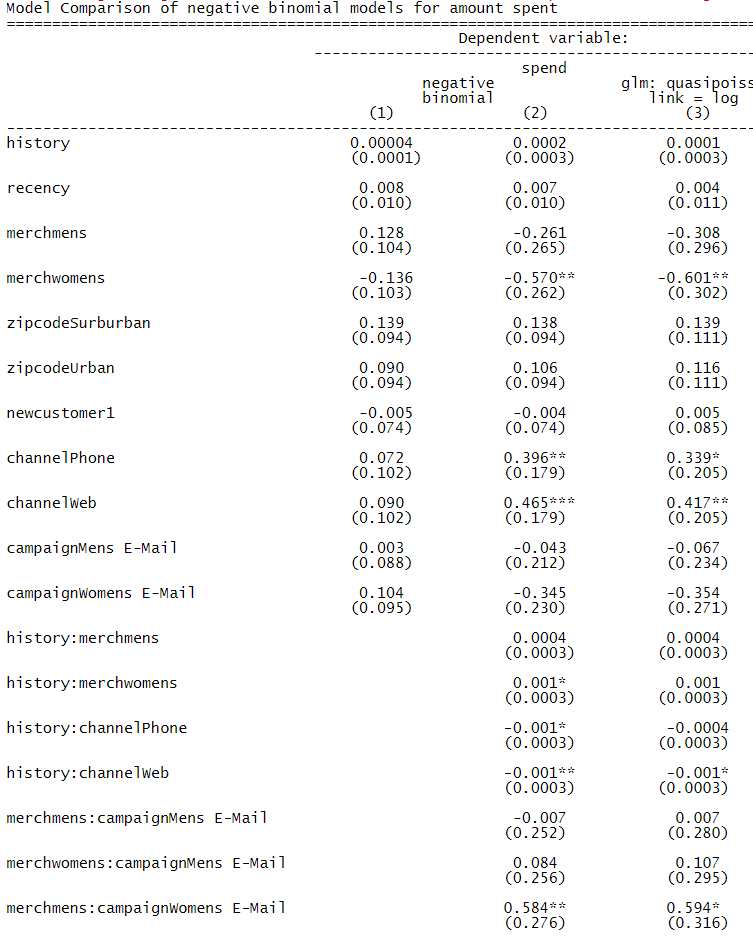
Regression Assumption tests on the best model (nbinom2):

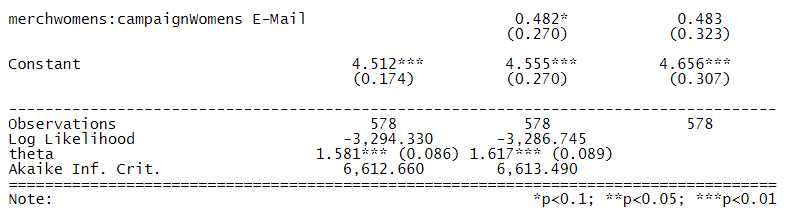
**Multicollinearity:** VIF tests indicate no value greater than 5. VIF test was done on model without interaction terms.



**Independence/Autocorrelation:** Durbin-Watson’s test (DW = 1.99, p-value =0.51) DW value is very close to 2 so it suggests there is no autocorrelation.

Please find the star gazer output below.





Note: Base level for merch is “Both” ie customers who got both men and women’s merchandise.

1. **Based on your analysis, answer the following questions (using marginal effects, not statistical significance)**

* **How did the promotion campaigns work relative to the control group?  Did the men's promotions work better than the women's promotion (or vice versa) and by how much?**

The spending decreases by 4.3% over the control group for customers who received men’s email promotion and it decreases by 34.5% over the control group for customers who received women’s email promotions given these are all customers who used to purchase both men’s and women’s merchandise. Clearly the promotions did not work for customers who purchased both men’s and women’s merchandise which could be because these would be couples who would usually spend a lot. We can see that there is about 30% decrease in spend for customers who received women’s promotions compared to men’s.

* **Should we target these promotions to new customers (who joined over the last 12 months) rather than to established customers, or vice versa?**

The amount spent decreases by 0.4% for new customers compared to older more established customers. So its better to target our old customers even though the marginal effect is low.

* **Should we target these promotions to customers who have a higher (or lower) history of spending over the last year?**

For $100 increase in spending over the last year the amount spent increases by 2% given that these are customers who used purchased using multichannel and also these are customers who purchased both men and women merchandise.

* **Did the promotions work better for phone or web channel?**

The promotions certainly worked better for web channel, because these customers spent 46.5% more than customers who used both the channels and about 6.9% more than customers who used only the phone to purchase. Purchase history had a very small effect on the spending for customers who used these different channels.

* **Will the promotions work better if the men's promotion is targeted at customers who bought men's merchandise over the last year (compared to those who purchased women's merchandise), and if the women's promotion would work better if targeted at customers who bought women's merchandise over the last year?**

Spending decreases by 0.7% if men’s promotion is targeted at customers who got men’s merchandise. Spending increases by 48.2% if women’s promotion is target at customers who got women’s merchandise. Targeting the right type of merchandise for right gender works for women but not for men.

1. **Reflect on the quality of your analysis, and comment on things you can do to further improve this analysis.**

The estimates are somewhat consistent but the estimates do not make any real world sense in some cases. For example, my best model suggests that if women’s promotions were targeted at customers who purchased men’s merchandise then the spending would go up by 58.4%. Estimates of the model cannot be relied upon. The model could further be improved by adding some more interaction terms which I might have missed possibly even three way interactions.

We could do a logistic regression analysis with the conversion or visited and see what makes a customer spend or visit a website. These insights could help the online retail company to make certain decisions on its promotion campaign.