Homework Team Orange 2

Minutes: September 4, 2019 - 4pm

1. Paired Check in progress, any comments?
   1. Sufyan 🡪 Grant 🡪 Evan 🡪 Cathy 🡪 Price 🡪 Sufyan
   2. Correct test stats for different types of variables
      1. Continuous: Wald chi-sq, p-value
      2. Binary & Ordinal: MH test stats, p-value
      3. Nominal: Pearson test stats, p-value
   3. Linearity Assumption (Last time)
      1. Pay attention to variables that have lots of 0 values (Box Tidwell would run only a small subset of data)
         1. Go study the data, is it really 0 or it does not have an account?
            1. “**WHERE**” statement
         2. use **GAM**. (document percentage of 0)
         3. if not converged, go back to Box Tidwell (specify it in report)
2. Data Considerations (interesting finding)
   * 1. ~ 10 important categorical variables
     2. Stratify **AGE** to categorical variables in the future
     3. **HMVAL**: significant & linearity assumption met (any other continuous variables? )
     4. **DEPAMT**: df insane ()
     5. Within Summary: multicollinearity
        1. One pair that is highly correlated (**POSAMT** & **POS**: 0.85)
        2. **CCBAL** & **MTGBAL**: 0.95
     6. **Large portion of 0 values**
3. Visualizations (Sufyan)
   1. Table for significant variables (separated by 4 types)
   2. Appendix table for all variables (separated by 4 types)
   3. Visualization for missing values: Bar chart? Pick top n?
   4. Extra: interesting findings?
      1. *Comparative horizontal bar chart to show unbalanced distribution of 0/missing data?*
4. Report draft (Due tomorrow at 5pm)
   1. Executive Summary: (**Cathy**)

*(Could write it in the end)*

* + 1. Business context
    2. briefly mention findings (how many variables)
    3. list some important variables maybe
  1. Results: (**Price**)
     1. significant variables we found
     2. **table** of Significance (by p-value)
     3. **table** of OR (By magnitude), normalize those less than 1
     4. Linearity Assumptions for continuous (var name; description; Y/N)
  2. *Concerns* & Recommendations (**Evan**)
     1. **Bar chart** for Missing values
     2. Redundant variables (correlation)
     3. *(Binary)* variables with high odds ratios
  3. Methodology & Analysis: how we get this result (eg. Box Tidwell) (**Grant**)
     1. Data set
     2. “where statement” in linearity assumption testing
     3. When GAM did not converge, we go back to Box Tidwell (with some data missing)
        1. Give 2,3 examples
     4. When MH broke, we use EXACT permutation to compute the significance
  4. Conclusion **(Sufyan)**
     1. *(Could write it in the end)*
     2. data consideration?
  5. Appendix: Table for all variables

1. Before Next Meeting (9/5/2019 at 3pm):
   1. *review visualizations*
   2. review writings
   3. discussion about the contents