Assessing Later Survey Response Based on Past Survey Completion

Ada Harris

FedCASIC Workshop

Content

- **Office of People Analytics**
- **Nonresponse Bias Analyses/Response Rates**
- **Research Objective**
- **Sample Design**
- **Research Design**
- **Weighting**
- **Methods**
- **Results**
- **Summary**
- **Next Steps**

Office of People Analytics

- **Established October 1, 2016**

- **Transitioned from Defense Manpower Data Center (DMDC):**
  - Marketing research and outreach
  - Cognitive and non cognitive testing
  - Personnel surveys and related research
  - Personnel security research

- **Provide research and analytical solutions to effectively align Total Force management programs and policies**

Nonresponse Bias Analyses

1. **Evaluated composition of sample compared to respondents**

2. **Compared estimates to known population totals**

3. **Analyzed frequency of item missing data**

4. **Analyze later survey response**

Response Rates from Active Duty Surveys

	2014	2013	2012	2011
Total	21%	24%	26%	20%
Service				
Army	19%	22%	20%	16%
Navy	21%	22%	27%	20%
Marine Corps	15%	17%	21%	15%
Air Force	29%	36%	37%	30%
Paygrade				
E1-E4	9%	11%	9%	12%
E5-E9	27%	31%	31%	32%
W1-W5	37%	42%	38%	29%
O1-O3	32%	36%	38%	28%
O4-O6	48%	51%	54%	44%

Research Objective

- **Examine current survey response using prior satisfaction and retention.**
- **Assess 'survey takers' vs. 'survey avoiders'**
- **OPA noted that as members were sampled in subsequent surveys, they became more likely that they would answer at least one.**

Number of Surveys Sampled in	Sampled Count	Respondents to = Zero	Respondents to >= One	Respondents to >= Two	Respondents to >= Three	Respondents to = Four
One	237,316	184,975 (78%)	52,341 (22%)	-	-	-
Two	16,173	10,237 (63%)	5,936 (37%)	2,032 (34%)	-	-
Three	860	423 (50%)	437 (50%)	244 (56%)	85 (35%)	-
Four	25	6 (24%)	19 (76%)	12 (63%)	8 (66%)	2 (25%)

Active Duty Sample Design

- **Annual Cross-Sectional survey (sample size about 70,000 per year)**
  - 2011, 2012, 2013, 2014

- **Target population - Active duty Army, Navy, Marine Corps, and Air Force**

- **Single-stage stratified design (Service, Paygrade, Race, Duty Location, and Family Status)**

Research Design

- **Determine the number of members that were sampled in one, two, three or all four surveys**
  - Officers are more likely to be repetitively sampled than enlisted because of necessary domain estimates and fewer military officers
  - 20.5% of members sampled three-times are O-3, 6.8% are E-3
- **4,414 sample members responded to a prior survey and were sampled in a later iteration of the survey.**

Completed Survey	Sampled in 2014	Sampled in 2013	Sampled in 2012
2013	1,042 (24%)	-	-
2012	632 (14%)	1,060 (24%)	-
2011	534 (12%)	833 (19%)	704 (16%)

Weighting

- **Created research weights for the analysis**

- **Post-stratification cells were defined by the cross-classification of paygrade and Service at final iteration of the survey the member was sampled**
  - Members service mostly remained (99%) the same from survey iterations
  - Members paygrade mostly remained (93%) the same from survey iterations

- **The final weights were adjusted to match population counts for 2014.**

Methods

- **Survey response at later survey sampling**
  - 1= responded to later survey (54%)
  - 0=did not respond to later survey (47%)
- **Dataset was split into training and validation data (70% training and 30% validation)**
- **Use Logistic Regression and Naïve Bayes to model response on later survey**
  - Satisfaction/Retention (observed from prior survey)
  - Service
  - Family Status
  - Race
  - Gender
  - Martial Status
  - Years of Service

Methods

- **Log Loss function**
  - Log-loss is an appropriate performance measure when model output is the probability of a binary outcome.
  - The log-loss measure considers confidence of the prediction when assessing how to penalize incorrect classification.
- **Satisfaction**
  - Naïve Bayes –0.677
  - Logistic Regression –0.679
- **Retention**
  - Naïve Bayes –0.71
  - Logistic Regression –0.691

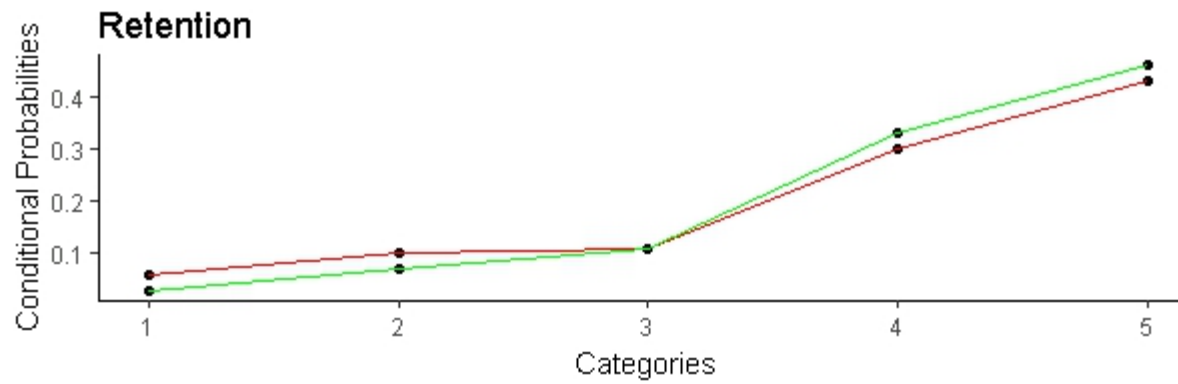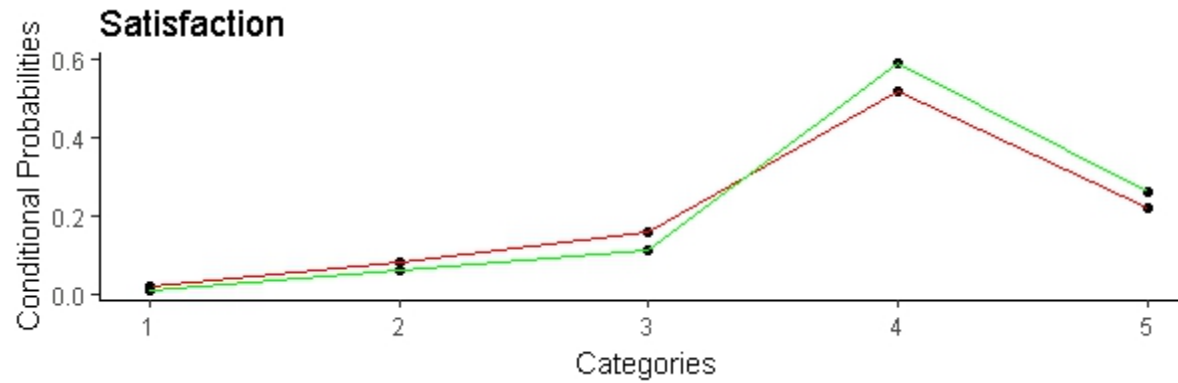- **Naïve Bayes Final Model – the whole dataset is used**

Results

- **Conditional Probabilities**

Satisfaction					
	Very dissatisfied	Dissatisfied	Neither	Satisfied	Very satisfied
Nonresponse	0.02	0.08	0.16	0.52	0.22
Response	0.01	0.06	0.11	0.59	0.26

Retention					
	Very unlikely	Unlikely	Neither	Likely	Very likely
Nonresponse	0.06	0.10	0.11	0.30	0.43
Response	0.03	0.07	0.11	0.33	0.46

Results

Summary

- **Number of times a member is sampled across iterations**
  - Response burden may be high for officers

- **Takes advantage of prior survey information that can inform types of respondents**

- **Prior satisfaction and retention show little difference between response and nonresponse on later survey**

- **Question?**

Next Steps

- **Cross-Validation**

- **Weighting strategies for combining cross-sectional surveys**

- **Improve Machine Learning techniques**

- **Investigate additional variables to model to describe likely respondents**

Thank You

Ada Harris

Statistician, Statistical Methods Branch

Office of People Analytics (OPA)

Defense Research, Surveys, and Statistics Center (RSSC)

571.372.1120

ada.b.harris.civ@mail.mil