APD Statistical Validation and Documentation

../assets/APD-logo-BG.png

../assets/ISF-logo-medium.png
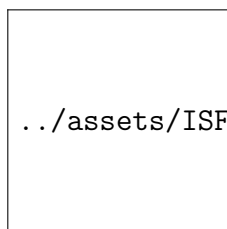
October 12, 2025

Agency for Persons with Disabilities (APD)
iBudget Algorithm Study
Procurement Office: Tamara Harrington
4030 Esplanade Way
Tallahassee, Florida 32399

ISF, Inc.
https://www.isf.com
Report prepared by Allison K.Kunerth, Ph.D.
ORCID Researcher Number 0000-0002-9035-507X
Senior Management Consultant and Data Scientist, ISF, Inc.
October 12, 2025

Full Study prepared by ISF, Inc.
Annika Baeten, Management Consultant
Abbie David, PMP, PhD, Project Manager
Jason Dillaberry, Client Partner
Matthew Fisher, JD, LLM, MBA, Principal Consultant
Juan B. Gutiérrez, PhD, Mathematician
Jessica Kemper, MPA, Senior Management Consultant
Chris Klass, MPP Finance, Senior Management Consultant
Allison Kunerth, PhD, Senior Management Consultant
Daniel Margolis, Associate Management Consultant

Contents

1   Feature Selection

Model Validation is being completed by Juan,Daniel and Allison. This document serves to track this process as much as possible given the time constraints.

1.1   Section 2.2

Analysis encompasses 6 fiscal years of data. End of FY is August 31st, 2025. Accordingly, 6- years is correct.

1.2   Section 2.3.1

Exlusions seem reasonable, including those with zero variance since they would have no impact on the outcome-and is better than excluding just low variance.

- MidYear QSI = 957, 0.5%

- Late Entry = 43,102, 20.2%

- Early Exit = 14,873, 7.0%

- No Costs = 109, 0.1%

- Negative Costs = 0, 0.0%

- Insufficient Service = 81,901, 38.5%

- No QSI = 128,791, 60.5%

- Need to double check that the min, max and mean number of records is correct

Question 1 Do we need to include the cost requirement greater than 0 since no cases meet that definition?

Answer 1 That weas filtered from data preparation for the cache. Look at sql transformation 14 0r 13

Question 2 Cardinality is number of levels of each variable being considered. Need to check whether the number 50 is correct.

Answer 2 Categorical variables with more than 50 unique levels are excluded to avoid unreliable chi-square estimates and extreme sparsity.The value 50 comes from the approximate number of 'valid' QSI categorical indicators about52 total items 2 summary variables = 50.

Question 3 Where is'OLevel' defined/derived? I know it is the ordinal mapping of the variables, but didn't immediately see it.

Answer 3 OLEVEL originates from the QSI Assessments tables $tbl_Q SIAssessments/tbl_Q SIAssessmentsLegacy$
Represents the Overall Level of need computed from the other QSI domains. It is imported into analysis as a numeric/ordinal feature after label-encoding.

2 Model 1:

TBD

3 Model 2:

TBD

4 Model 3:

TBD

5 Model 4:

TBD

6 Model 5:

TBD

7 Model 6:

TBD

8 Model 7:

TBD

9 Model 8:

TBD

10 Model 9:

TBD

11 Model 10:

TBD
'