Improper Payments Logistic Model
Improper Payments Logistic Model

Why NSF?

- Introductory use case to begin implementing data science tools and techniques
- Open source – no licenses to procure, no hoops
- Turning data into information to drive decision making
- Professional development - upskilling for the workforce of the future
Improper Payments Logistic Model Challenges and Opportunities

- Management commitment to stewardship and oversight of grant funds
- Annual requirement for PIIA compliance and reporting
- Data sources were not fully integrated for monitoring activities
- Opportunity for continuous development for deploying data science tools

**Agency**

- Single Audit Data¹
- USASpending.gov Data
- Independent performance data
- Agency monitoring and award data

**Institution Types**

- Universities
- Collegiate systems
- Research facilities
- State governments
- Local governments
- K-12 school districts

**Financial and Operational Data**

**Federal Expenditures**

- Unallowable costs
- Inaccurate grant accounting
- Ineffective control environment
- Noncompliance with Federal Regulations

¹ Federal Audit Clearinghouse Single Audit summary data
Improper Payments Logistic Model
Building the Solution

Model Development

- Created **inventory** of available data to help **understand risks** associated with institutions receiving federal funds
- Used **R** package dplyr to combine and manipulate data sets, design, and test the model
- Conducted tests on **model validity** including reviewing for correlation and multicollinearity, and bias/anomalies

Equation for Probability of an Improper Payment Finding

\[
\frac{1}{1-p} = \beta X_1 + \beta X_2 + \cdots + \beta X_9
\]
Improper Payments Logistic Model
Key Outcomes and Next Steps

Model Results

Using the initial results as a basis for prediction, the model predicted with **87% accuracy** whether an institution would have an improper payment over a five-year period.

<table>
<thead>
<tr>
<th>Predicted Positive</th>
<th>Actual Positive</th>
<th>Actual Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>34 (8.9%)</td>
<td>25 (5.7%)</td>
<td></td>
</tr>
<tr>
<td>Predicted Negative</td>
<td>39 (7.8%)</td>
<td>339 (77.6%)</td>
</tr>
</tbody>
</table>

Compliance Summary

- No Internal Control Findings: 995
- Material Weakness or Significant Deficiency: 203
Improper Payments Logistic Model Considerations for the Community

NSF Recipient Inclusion Criteria

- Payments in last 2 FY
- Single Audit in last 5 FY
- Included in annual risk assessment
- No missing variables

Transferability Considerations

- Average Size of Award
- Grantee Portfolio Size
- Recipient Types
- IP Root Causes
- Available Data Sources

Awardees Modeled? Payments

1747 (Yes) $6.1B
328 (No) $123M

Address gaps through other analyses
Improper Payments Logistic Model
Crowdsourcing the Solution!

We have a **shared responsibility** to monitor the problem.

Why should we **stand alone** with the **solution**?

**No cost to implement!**
Talk to your IT department about R as approved software.

**Download today!**
Immediate foundation to expand with your own data.

**Review, collaborate, and share!**
We all benefit when we can share our collective experiments.


#TTS20