Using Predictive Analytics to Detect Fraudulent Claims

May 17, 2011

Roosevelt C. Mosley, Jr., FCAS, MAAA

CAS Spring Meeting

Palm Beach, FL

Predictive Analysis for Fraud

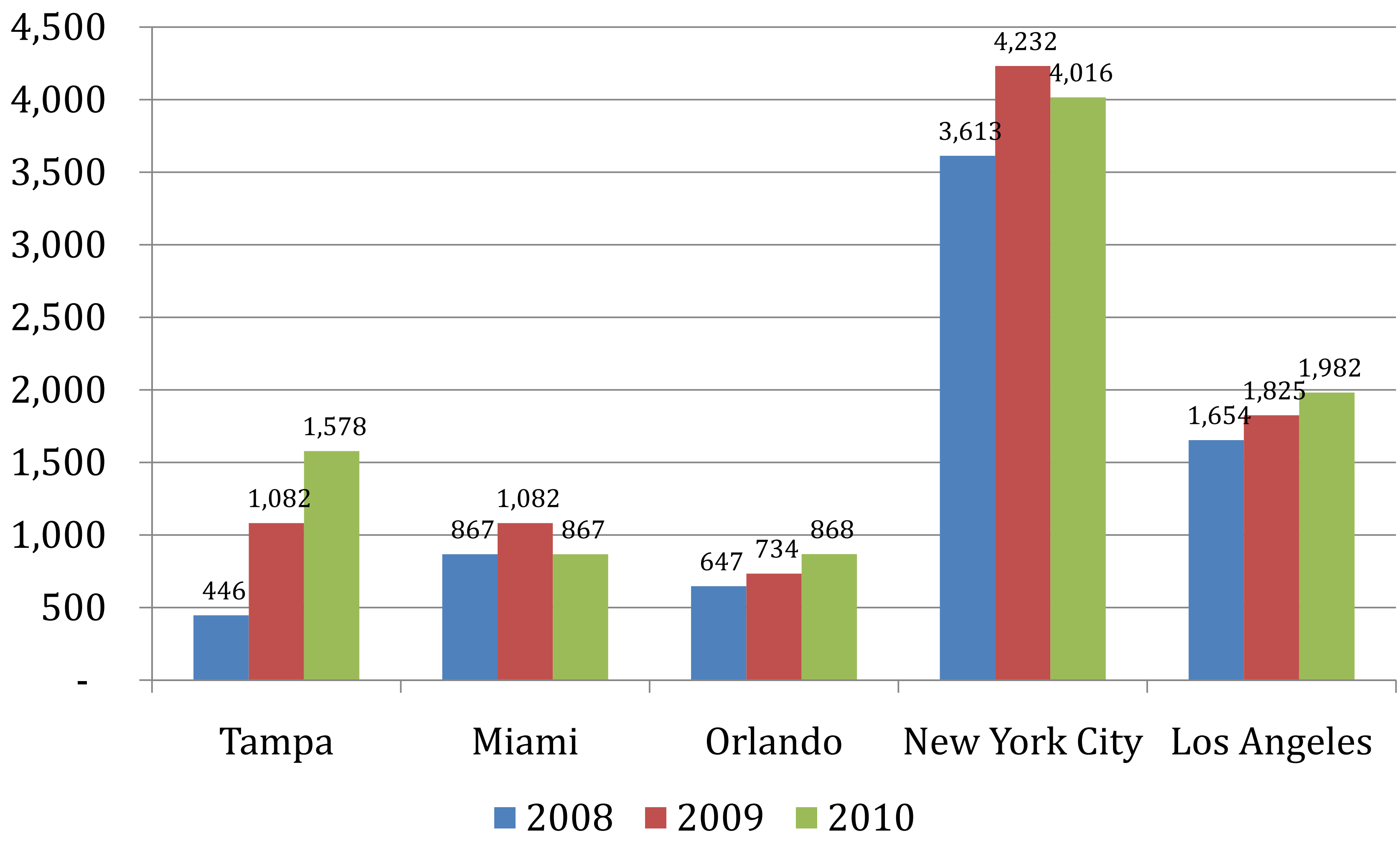
- Claim fraud is increasing, focus on fraud is magnified
- There are special investigators in the industry that are good at detecting fraud
- As good as they are, they can’t review every claim and detect all fraud
- Predictive analytics can bring the expertise to bear on all claims
- Predictive analytics can enhance the work of investigators by uncovering complexities the human eye may miss

Claim Fraud is Increasing, and the Focus on Claim Fraud is Increasing as Well

Increasing Claim Fraud – 2011 Headlines

- March 30 – Suspicious claims rise 34% in Florida
- April 17 – The Battle Against Insurance Fraud in Georgia
- April 26 – Insurance Groups Stress Need for N.Y. No-Fault Reform at Hearing
- April 26 – PIP Bills Crash in Florida
- May 2 – Four Women Booked with Insurance Fraud in Louisiana
- May 5 – Council Woman Gets Jail Time for Insurance Fraud
- May 6 - Allstate Files $4 Million Insurance Fraud Case in New York
- May 8 – Questionable Claims on the Rise in Oklahoma (+15%)
- May 12 – NY State Must Stand Against No Fault Car Insurance Fraud

Increase in Questionable Claims



Source: National Insurance Crime Bureau

Fraud Detection Process

Geneal Fraud Identification Process

- Identify triggers that alert the claim adjuster to potential fraud (fraud indicators)
- Rely on claim adjusters to identify potentially fraudulent claims (recognition, intuition)
- Potentially fraudulent claims are referred to SIU
- Smaller group of SIU investigators handle the investigation of fraudulent claims

Recognition (I’ve Seen This Before)

- Examples
  - Repeat offenders
  - Provider/patient/attorney combinations
- Approach
  - Advisory claim database
  - Experience of adjuster
- Disadvantages
  - Assumes adjuster has seen it before
  - Aliases
  - Fraud becomes smarter

Fraud Indicators

- Rules based system
- Identify known or potential fraud scenarios
- Advantages
  - Easy to implement and modify
  - Easy to understand
  - Effective to attack specific problems
- Disadvantages
  - Doesn’t detect new and unknown fraud
  - Creates smarter fraud

Fraud Indicators - Examples

- Distance between claimant's home address and medical provider
- Multiple medical opinions/providers
- Certain claim types (e.g., soft tissue)
- Changing providers for the same treatment (possibly correlated with other claim activity)
- High number of treatments for type of injury
- Abnormally long treatment time off for the type of injury
- Accident severity does not correlate with severity of injury

Intuition (Something Smells Funny)

- Something about the claim doesn’t seem right to the adjuster, and it is referred to the SIU
- Relies on ability and experience of adjuster to see suspicious cases
- Inexperienced adjusters will not have the ability to detect suspicious as well

As Good as the SIU Is…

Concerns with the Current Process

- Claim referral can be inconsistent – heavy dependence on claim adjuster
- False positives
- Claim adjuster may not be aware of all suspicious relationships
- Not all historical fraud has been identified
- Prioritization of potentially fraudulent claims

Using Predictive Analytics to Address These Concerns

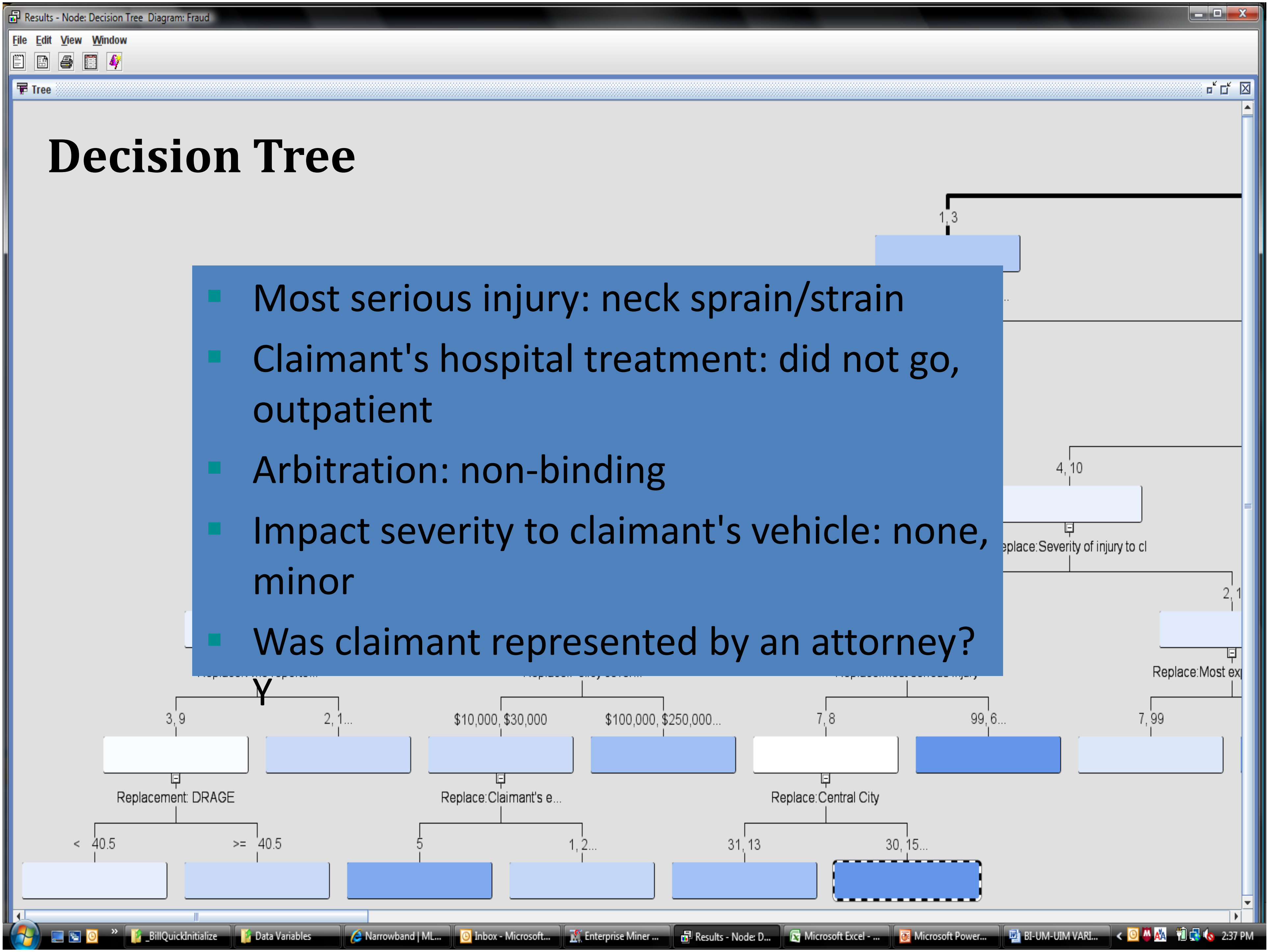
- Predictive analysis of historical referrals (consistent referrals)
- Predictive analysis of historical fraudulent claims (false positives)
- Association analysis (recognition of claim patterns)
- Clustering Methods (missed claims, prioritization)
  - K-mean clustering
  - Kohonen self-organizing maps
- PRIDIT (consistent referrals, prioritization)

Analysis of Historical Referrals

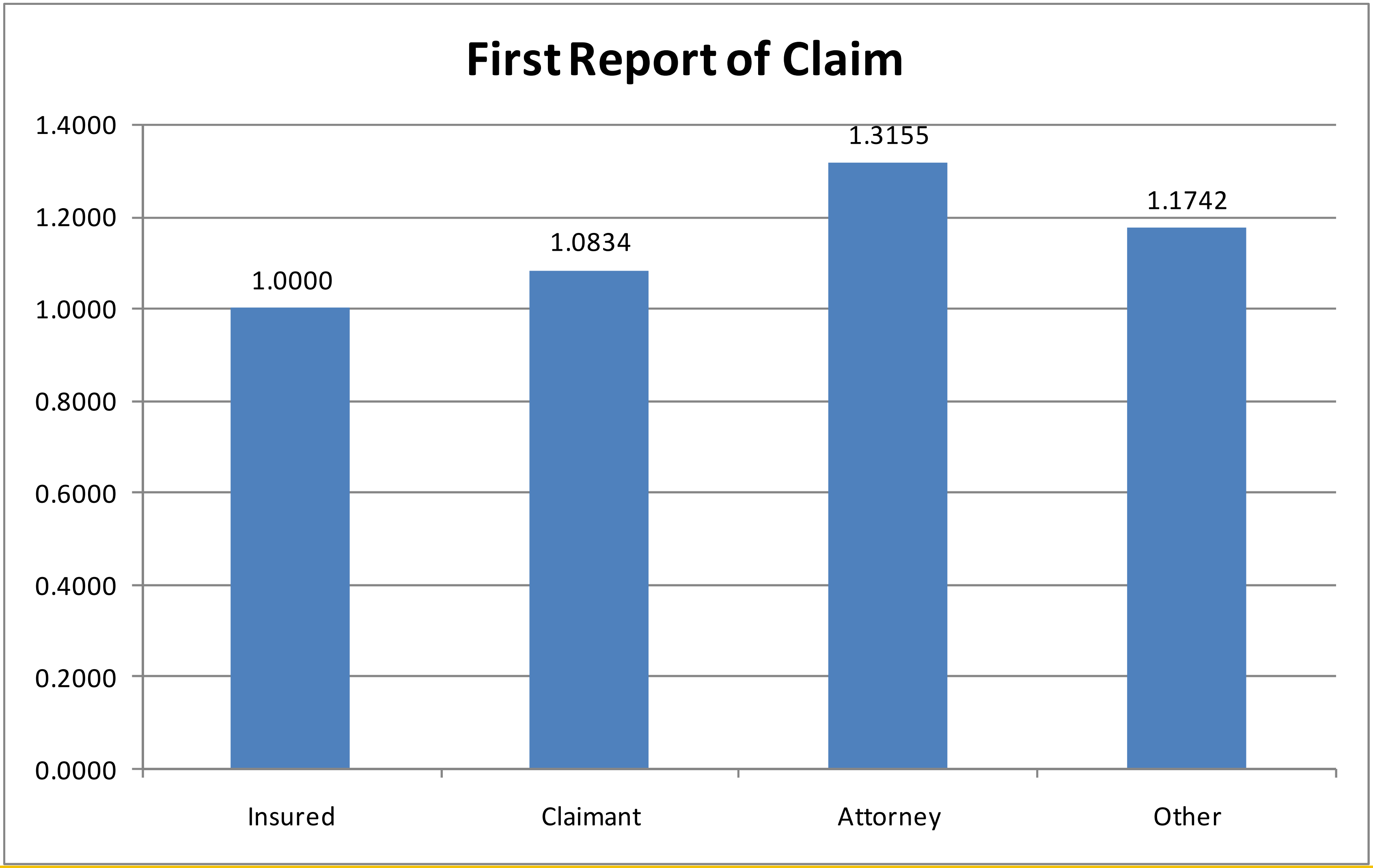
- Target: history of claim referrals to SIU
- Independent Factors: details of claim
- Models Tested
  - Decision tree
  - Neural network
  - Linear regression
  - Ensemble
- Result: given the history of claim referrals, the likelihood that a new claim should be referred to SIU based on the claim characteristics

Decision Tree

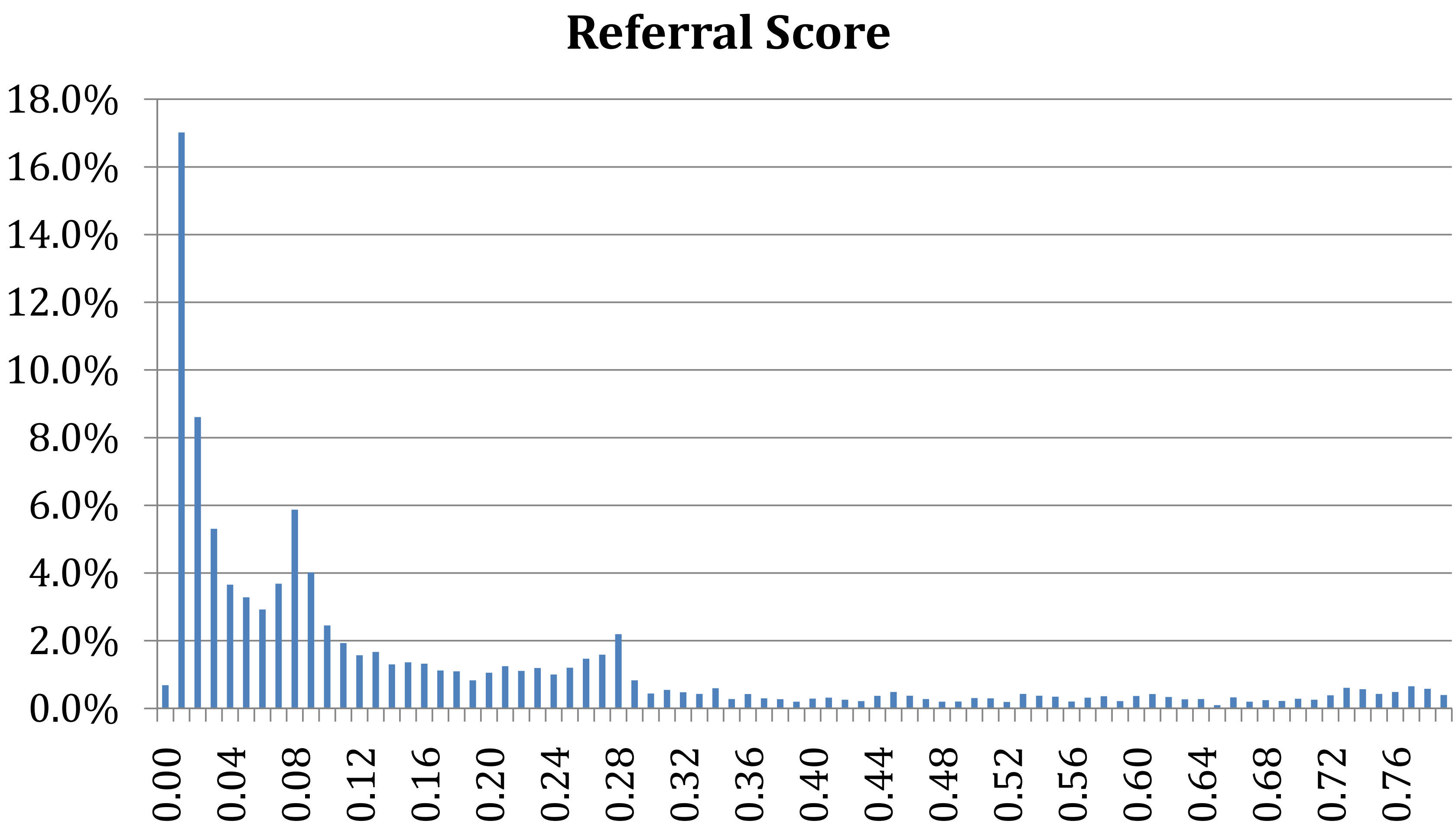
- Most serious injury: neck sprain/strain
- Claimant's hospital treatment: did not go, outpatient
- Arbitration: non-binding
- Impact severity to claimant's vehicle: none, minor
- Was claimant represented by an attorney?



Regression: First Report of Claim



Referral Score



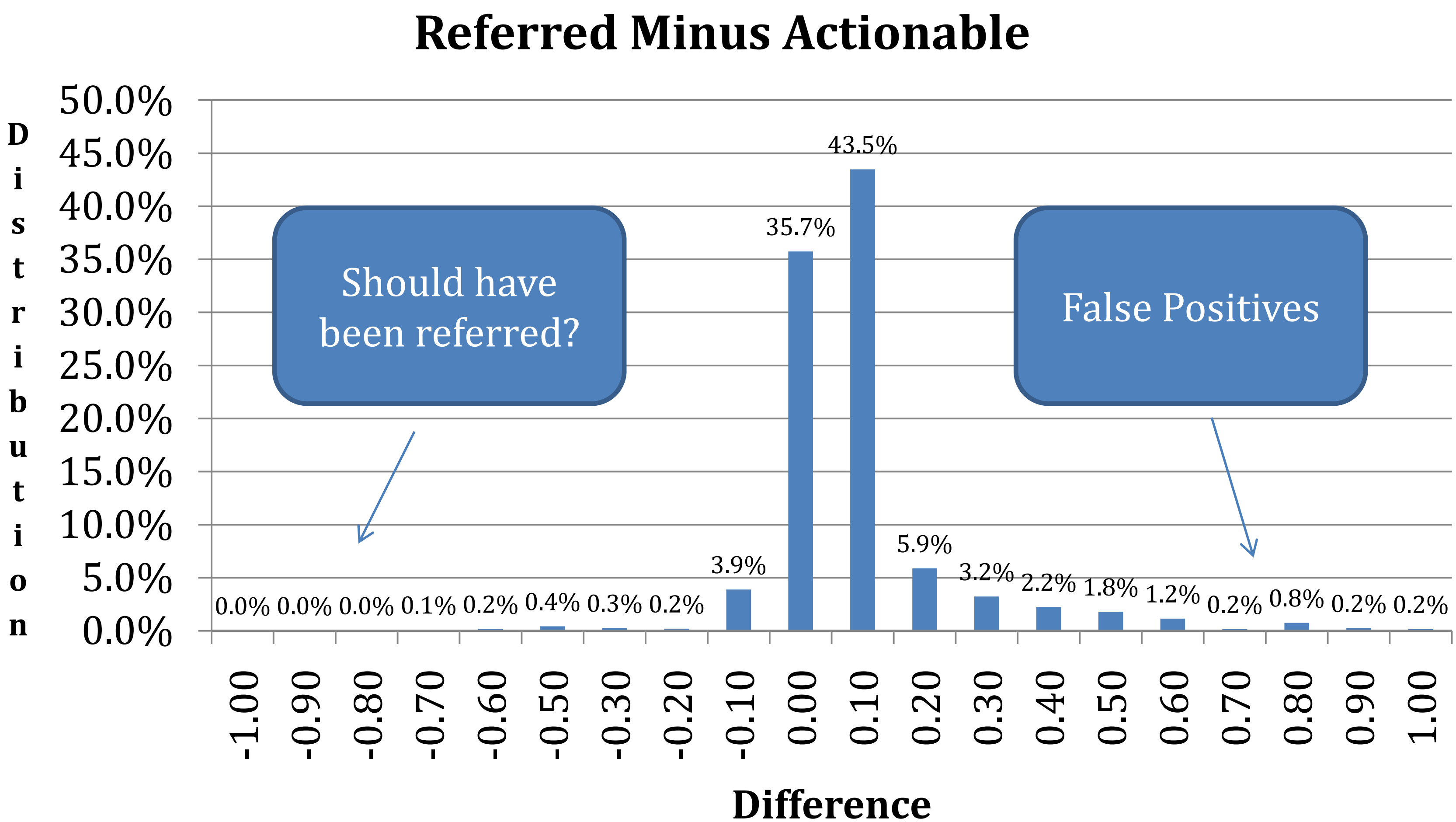
Analysis of Historical Fraudulent Claims

- Target: history of actionable claim referrals to SIU
- Independent Factors: details of claim
- Models Tested
  - Decision tree
  - Neural network
  - Linear regression
  - Ensemble
- Result:
  - given the history of claim referrals, the likelihood that action will be taken on a new claim based on the claim characteristics
  - Comparison to referral claims

Decision Tree Comparison – Variable Importance

Variable	Actionable Importance	SIU Referral Importance	Ratio
Central City	1.000	0.464	46.4%
Replace:Claimant's state of residence	0.967	1.000	103.5%
Impact severity to claimant's vehicle	0.962	0.828	86.2%
Was claimant represented by an attorney?	0.850	0.905	106.4%
Policy coverage limits per person	0.750	0.411	54.9%
Arbitration	0.547	0.368	67.2%
Most serious injury	0.530	0.375	70.9%
Settlement_lag	0.456	0.000	0.0%
Who reported injury to insurer	0.439	0.374	85.3%
Most expensive injury	0.423	0.239	56.5%
DRAGE	0.312	0.306	98.0%
Lawsuit status	0.295	0.000	0.0%
Driver, other violation	0.285	0.000	0.0%
Amount Spent on Medical Professionals	0.255	0.412	161.6%

Difference in Referred vs. Actionable Claims



Association Analysis (recognition of patterns)

- Technique used in market basket analysis
- Identification of items that occur together in the same record
- Produces event occurrence as well as confidence interval around the occurrence likelihood
- Can lead to sequence analysis as well, which considers timing and ordering of events

Association Analysis Measurements

- **Support** – how often items occur together

$$\frac{\text{Transactions that contain items A \& B}}{\text{All transactions}}$$

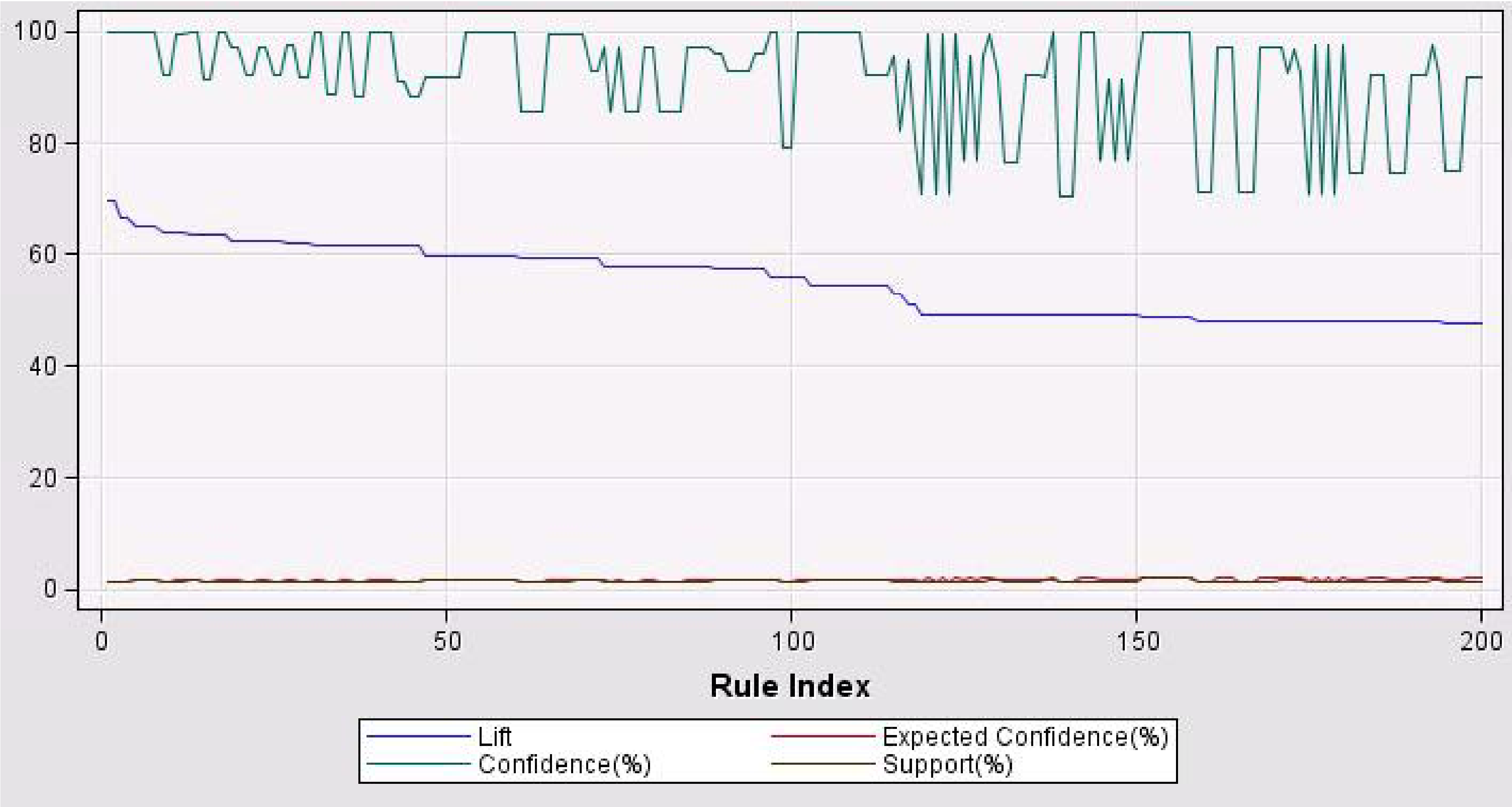
- **Confidence** – strength of association

$$\frac{\text{Transactions that contain items A \& B}}{\text{Transactions that contain item A}}$$

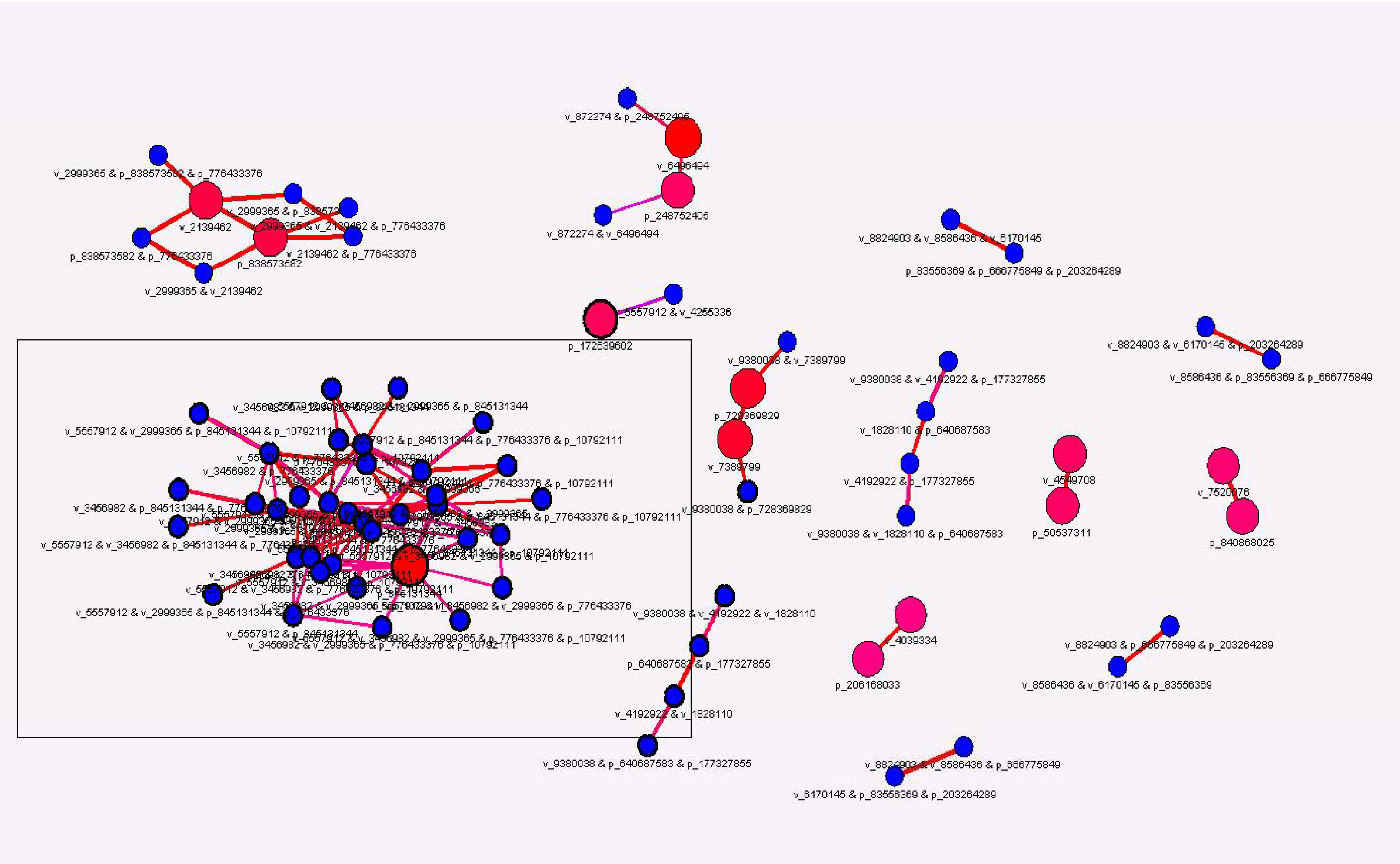
- **Expected Confidence** – proportion of items that satisfy right side of rule

$$\frac{\text{Transactions that contain item B}}{\text{All transactions}}$$

Association Analysis Output



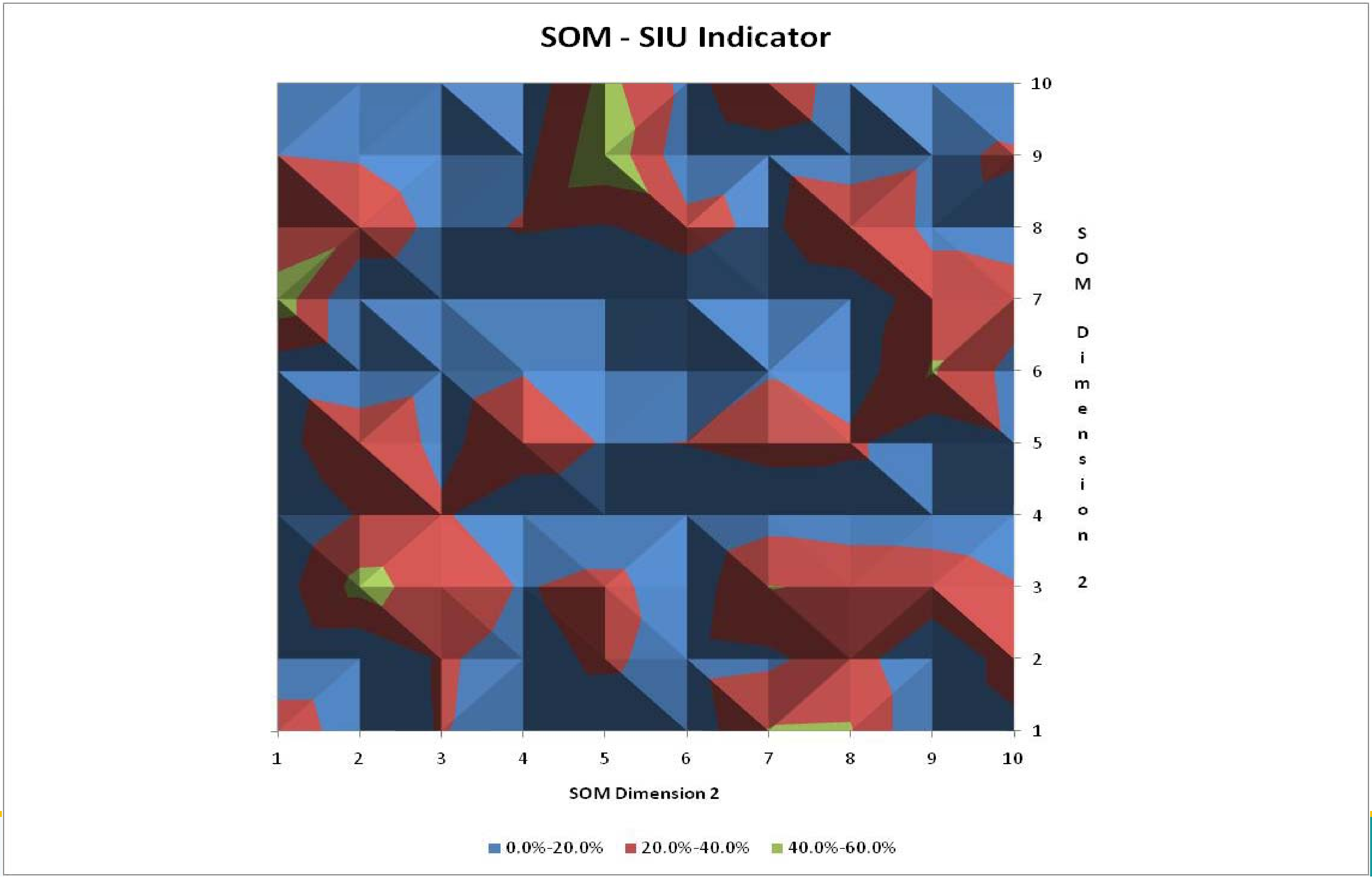
Association Output Example



Self – Organizing Maps

- Topological mapping from input space to clusters
- Observations from the input space are mapped onto an organized grids
- Neurons are determined initially, and as inputs are mapped to the grids the neurons are adjusted
- As a input is matched to the grid, all the neurons around that grid are updated

SOM – SIU Indicator



Clustering/Segmentation

- Unsupervised classification technique
- Groups data into set of discrete clusters or contiguous groups of cases
- Performs disjoint cluster analysis on the basis of Euclidean distances computed from one or more quantitative input variables and cluster seeds
- Objects in each cluster tend to be similar, objects in different clusters tend to be dissimilar
- Can be used as a dimension reduction technique

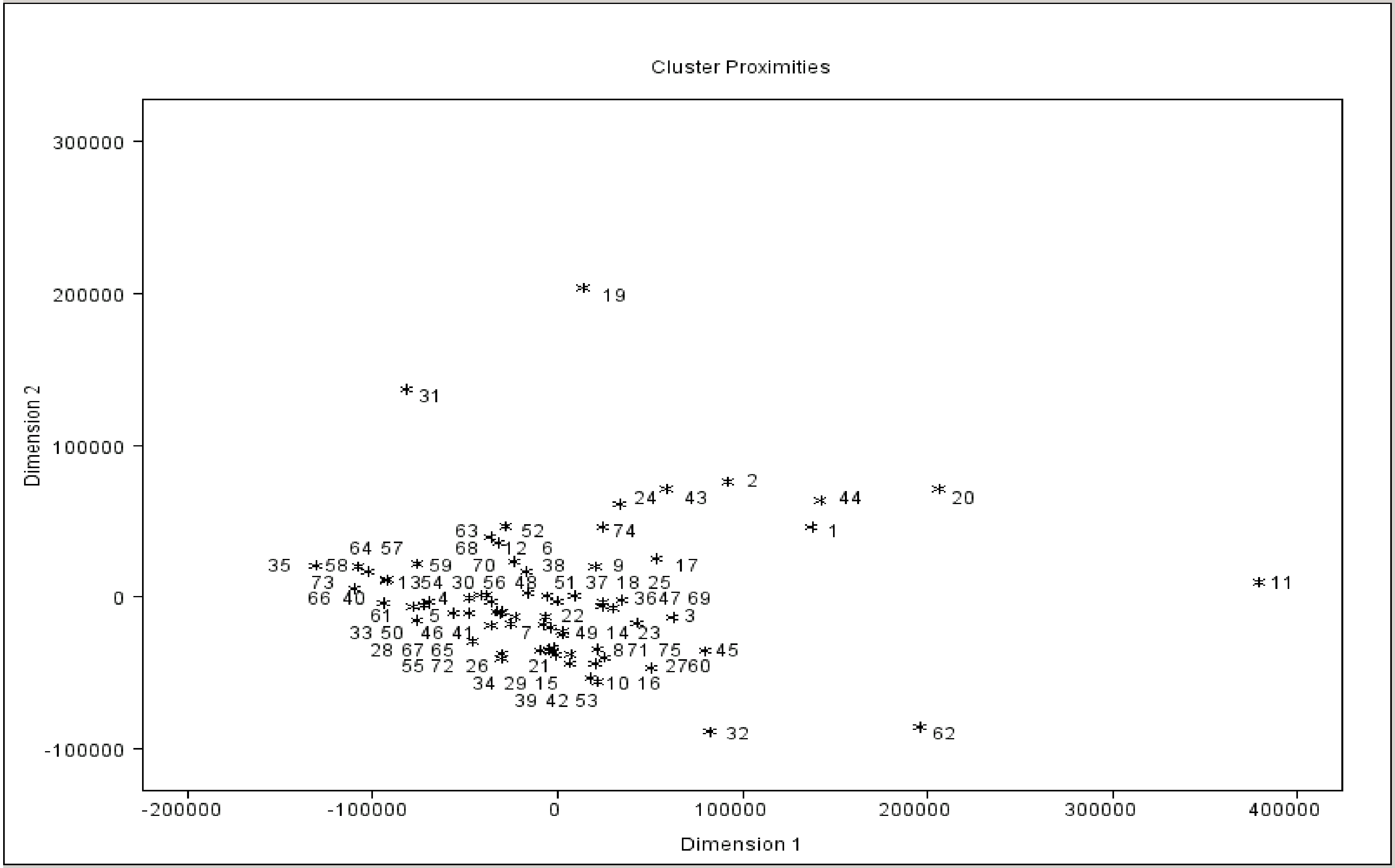
Cluster Evaluation – Suspicion Scores

- Root Mean Square Standard Deviation – variability of claims within a cluster
- Distance to Nearest Cluster – group of outlier claims
- Distance from Cluster Seed – the distance of the claim from the average
- Review of cluster summary statistics

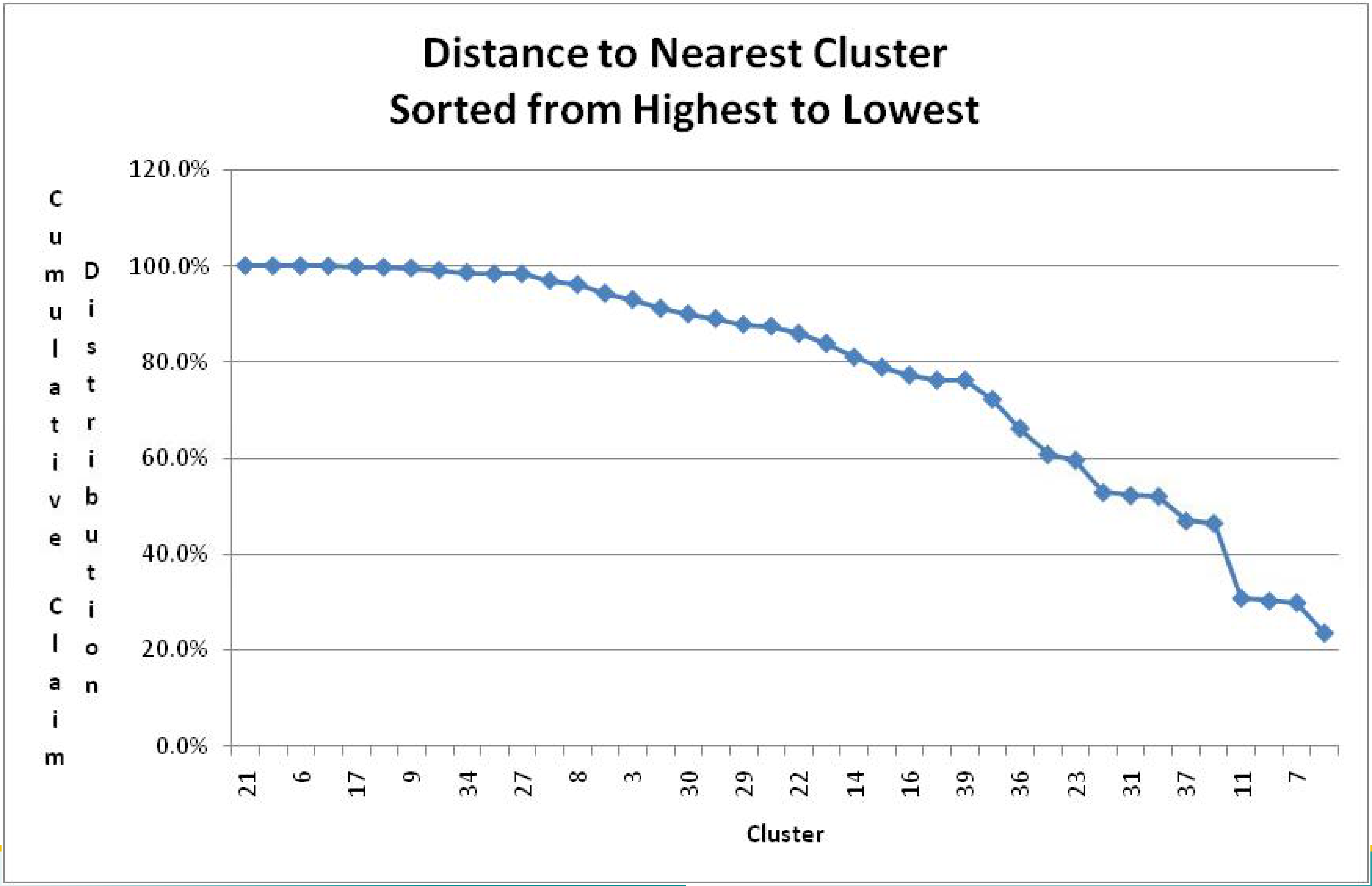
Homeowner Contents Analysis

- Claim values by detailed category
  - Replacement cost value
  - Depreciation
  - Number of items
  - Age
- Property characteristics (age, bathrooms, bedrooms)
- Coverage details (coverage C)
- Insured demographics (age, education, income)

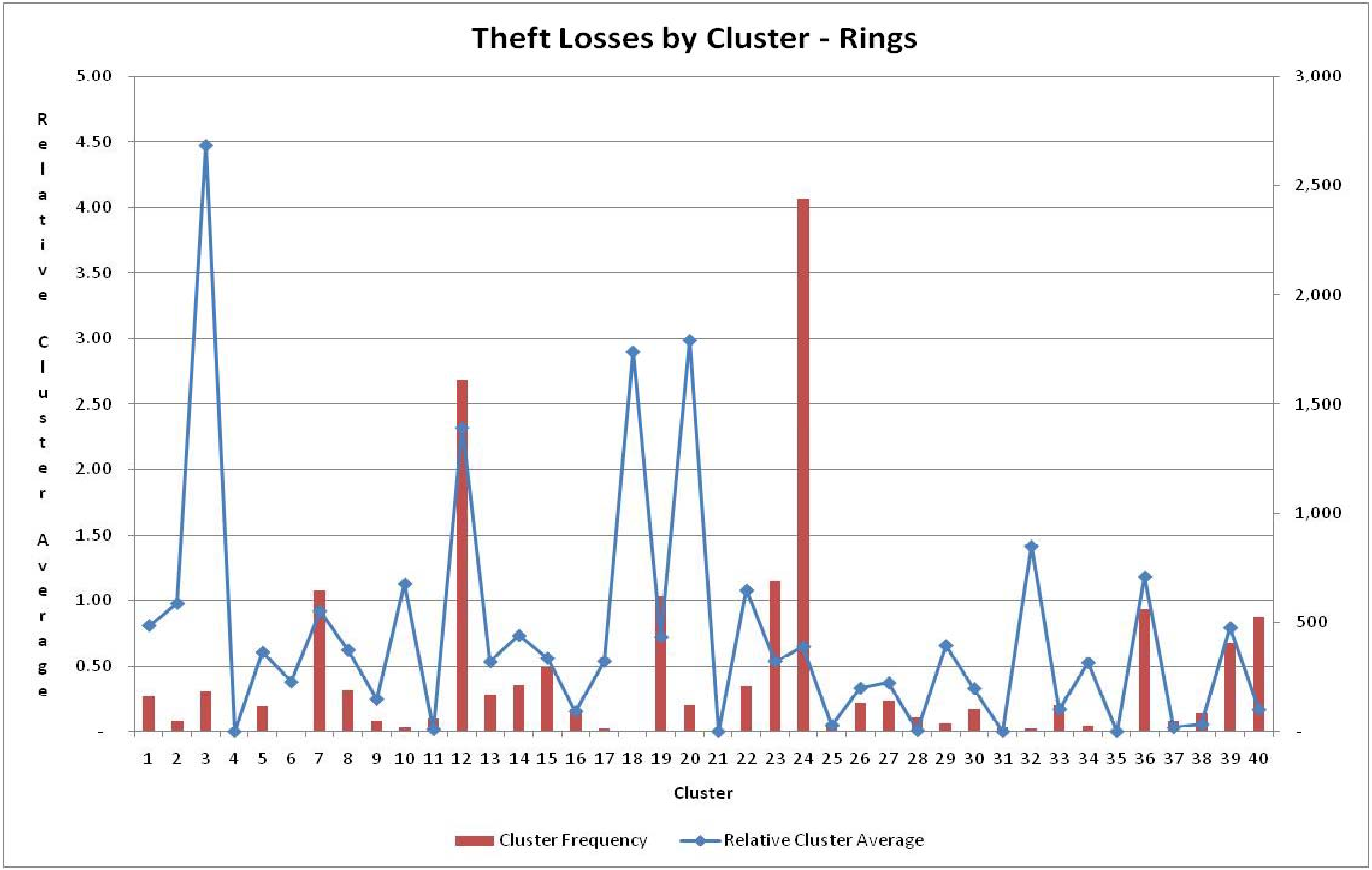
Cluster Proximities – All Causes of Loss



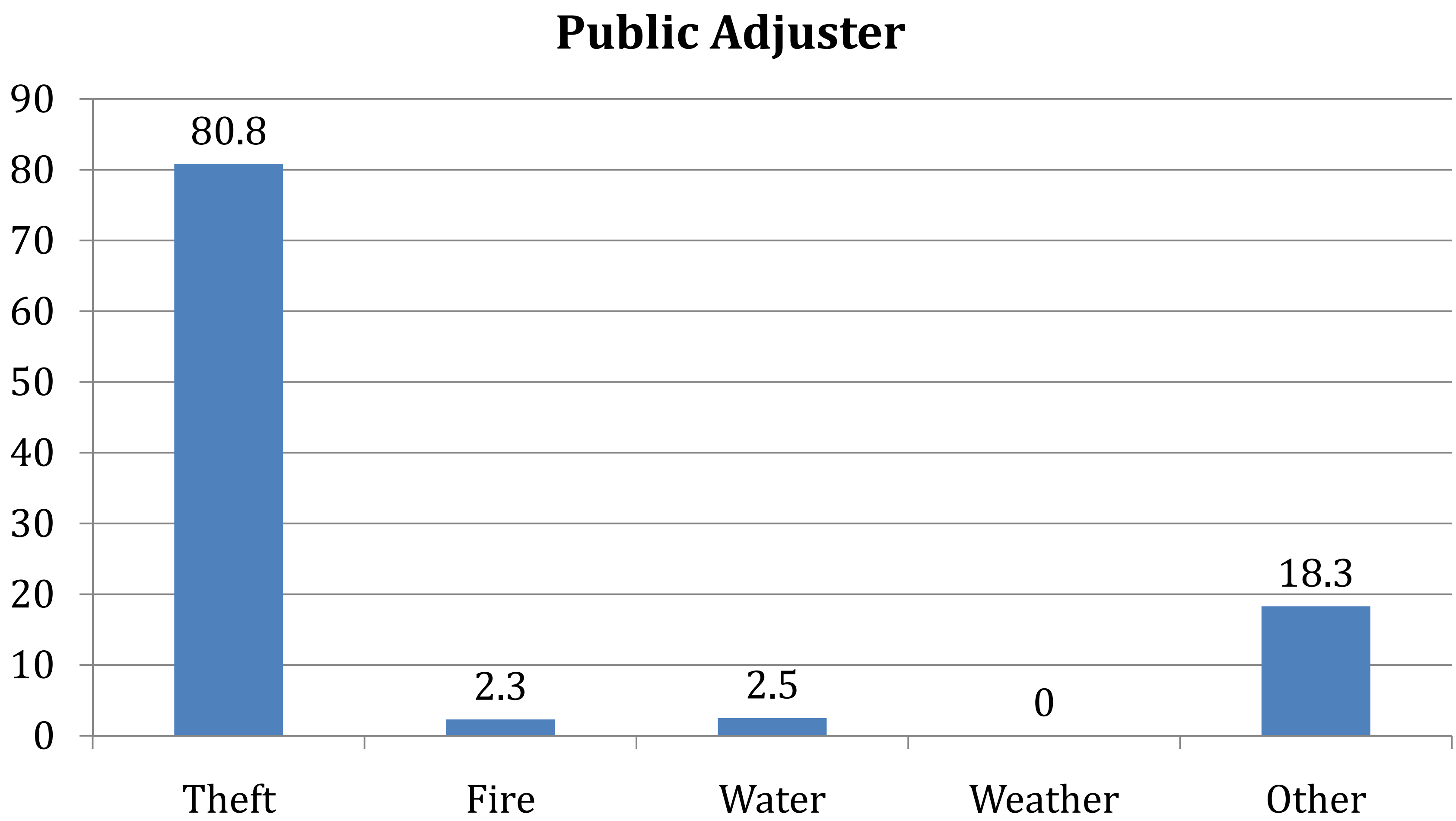
Cumulative Distribution – Distance to Nearest Cluster (Theft)



Review of Cluster Summary Statistics



Distance from Cluster Mean

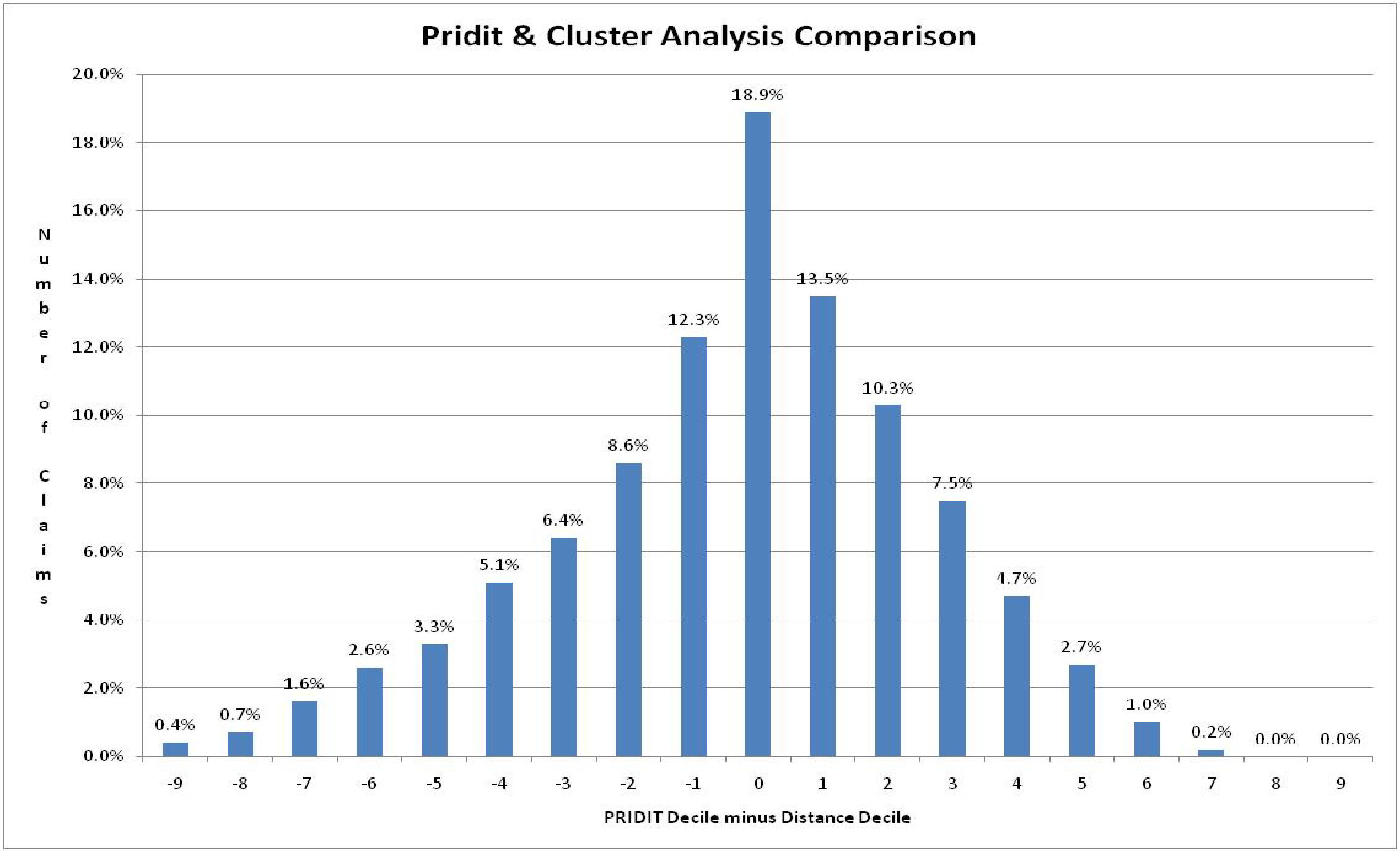


Final Fraud Calculations

Factor Name	Description	Input Value
insured_kids_2	Y, N, or U	u
peril_2	Cause of Loss	Fire
public_adjuster	0 or 1	0
IMP_REP_Coverage_C	Coverage C Amount	190,500
IMP_REP_Insured_Home_Bathrooms	Number of Bathrooms	2
IMP_REP_Insured_Home_Bedrooms	Number of Bedrooms	3
IMP_REP_Insured_Home_SqFt	Square Footage	1,412
IMP_REP_Insured_Home_YearBuilt	Year Built	1973
IMP_REP_Insured_Homeowner	Homeowner (Y or N)	Y
IMP_REP_acvloss_rcttotal	Ratio of ACV Loss to RCT Total	1.13
IMP_REP_create_lag	Delay in Creating Record	9
IMP_REP_insured_age_2	Insured Age	50
IMP_REP_insured_educationlevel_2	Years of Education	12
IMP_REP_insured_homevalue_calc_r	Home Value Calculation Rounded	149
IMP_REP_insured_yearsinhome_2	Insured Years in Home	6

Suspicion Score	
Root Mean Square Error	99.7%
Distance to Nearest Cluster	99.4%
Distance from Mean	96.5%
Combined	98.3%

PRIDIT Comparison



Wrap Up - Predictive Analysis for Fraud

- Claim fraud is increasing, focus on fraud is magnified
- There are special investigators in the industry that are good at detecting fraud
- As good as they are, they can't review every claim and detect all fraud
- Predictive analytics can bring the expertise to bear on all claims
- Predictive analytics can enhance the work of investigators by uncovering complexities the human eye may miss