# Developing Credit Risk Score using SAS programming

**Presenter: Amos Odeleye**

# DISCLAIMER

- Opinions expressed are strictly and wholly of the presenter and not of SAS, FICO, Credit Bureau, past employers, current employer or any of their affiliates
- SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries
- Other brand and product names are trademarks of their respective companies



*Credit Bureau or Credit Reporting Agency (CRA): Experian, Equifax or Transunion*

# About This Presentation

- Presented at Philadelphia-area SAS User Group (PhilaSUG) Fall 2018 meeting:
  - *Date: TUESDAY, October 30th, 2018*
  - *Venue: Janssen (Pharmaceutical Companies of Johnson & Johnson)*
- This presentation is an introductory guide on how to develop an in-house Credit Risk Score using SAS programming
- Your comment and question are welcomed, you can reach me via email: OduaTechnology@gmail.com

# Abstract


- Credit Risk Score is an analytical method of modeling the credit riskiness of individual borrowers – prospects and existing customers.
- While there are numerous generic, *one-size-fit-all* Credit Risk scores developed by vendors, there are several factors increasingly driving the development of in-house Credit Risk Score.
- This presentation will introduce the audience on how to develop an in-house Credit Risk Score using SAS programming, Reject inference methodology, and Logistic Regression.

## Why SAS for Credit Risk?

- **S**: Standardization
- **A**: Assurance of quality
- **S**: Scalability

## Some Financial Credit Risk job opportunities

- Data Analyst
- Risk Analyst
- Business Analyst
- Model Developer
- Model Validation Analyst
- Portfolio Risk Manager

# Overview of Financial Risk Management

# Risk Management

- Organization Risk Appetite:
  - Firms should select the type and level of risk that's appropriate for them to assume

- Risk Management (A-Q-M-M):
  - **A**ssess: Identify risk
  - **Q**uantify: Measuring and Estimating
  - **M**anage: Avoid, transfer, keep, mitigate, or eliminate
  - **M**onitor: Evaluate the process and make necessary adjustment

# Financial Risk Classification (C-L-O-M-O)

- **C**redit Risk:
  - *Default, Recovery, Collections, Fraud*
- **L**iquidity Risk:
  - *Funding, Trading*
- **O**perational Risk:
  - *Operational weaknesses and breakdowns*
- **M**arket Risk:
  - *Market price such as Interest rate, Foreign exchange, Equity, and Commodity price*
- **O**thers:
  - *Not already classified above.*
  - *Including but not limited to Model risk, Compliance risk, Strategic risk and Reputational risk*

# Some Financial Regulators

- **Global:**
  - *Bank for International Settlements (BIS)*
  - *BASEL Committee on Banking Supervision (BCBS)*
  - *Financial Stability Board (FSB)*
- **United States:**
  - *Federal Reserve Bank*
  - *U.S. Treasury/Office of the Comptroller of the Currency (OCC)*
  - *Federal Deposit Insurance Corporation (FDIC)*
  - *Consumer Financial Protection Bureau (CFPB)*
- **Canada:**
  - *Bank of Canada*
  - *Canada Deposit Insurance Corporation*
  - *Office of the Superintendent of Financial Institutions*
- **United Kingdom:**
  - *Bank of England*
  - *HM Treasury*
  - *Financial Services Compensation Scheme*

# Credit Risk Score

## What is It and What is Not

# CREDIT RISK SCORE:

- Is a Predictive modeling approach used to evaluate the level of credit riskiness associated with new applicants or *booked* customers
- Does *not* specifically identify "good" (positive behavior) or "bad" (negative behavior) individuals
- It is a Rank-Ordering predictive estimator that provides a *Statistical Measure* (odds or probability) that an individual with given attributes will be "GOOD" or "BAD"
- This Statistical Measure, transformed or "scaled" into a *Score* along with other business and strategy considerations are used in making financial decisions

## Credit Risk Score: The Pioneer

- Fair & Isaac company, FICO (1956) pioneered the development of credit risk score
- FICO Score is a generic score used in 80+ countries
- FICO Score is designed to *Rank-Order Customer's risk* based on the information in the consumer's credit file
- FICO Score provides an overview summary of the information on a consumer's credit file.
  - A single 3-digit number ranging from 300-850 which rank-orders consumers according to risk
  - Several versions are available including industry-tailored scores such as Bankcard or Auto score
  - The higher the score, the lower the future risk of *default* i.e., "BAD"





https://www.myfico.com/credit-education/credit-scores/fico-score-versions

## Credit Risk Score: How it is developed

- **Generic Score:**
  - Such as FICO Score, using CRA data for score development
- **Vendor Score:**
  - Developed by Vendor for the financial institution using CRA and/or in-house data for score development
- **In-house Score:**
  - Developed in-house modelers using CRA and/or in-house data

# Credit Risk Score: The Customer Lifecycle

- **Customer Marketing:**
  - Who is our target customer? Market segmentation? Pricing? Product offering/Cross-Sell? Channel?
- **Customer Originations:**
  - Also known as Underwriting or Acquisitions
  - How are customer selected and "booked"?
- **Customer Existing Account Management:**
  - Once the customer is "booked", how is the relationship maintained, controlled, and grown?
- **Customer Collections and Recovery:**
  - Some customers will default on their obligations, how and what treatments should be deployed to encourage payment and restore customers to non-delinquent status?
- **Customer Fraud Management:**
  - First Party Fraud
  - Third Party Fraud
  - Synthetic Fraud

## Risk Scores across Customer Life cycle

- Application (or Acquisition) Score
- Behavior Score
- Collection Score
- Recovery Score
- Fraud Score

# Credit Risk Score: Why develop In-house?

*Some of the factors driving increased used of in-house developed risk scores:*

- Reduced Cost
- Increased Regulation
- Improved customer experience
- Better software for building models
- Efficiency and process improvement
- Creating value and boosting profitability
- Ease of access to sizeable and reliable data
- Availability of greater educational material and training

# Risk Score: Some Modeling Approaches

- Logistic Regression
  - *Appropriate for Acquisition Risk Score used in Approve-Decline decision*
  - *Input and Outputs are clearly and well defined*
  - *Has good interpretability with business implications*
  - *Internal and External regulatory impact and compliance*
  - *Predicts Probability of Default (i.e., "BAD")*
  - *Probability is scaled into scores → Score = Offset + Factor*Ln(odds)*
- Decision Tree
- Neutral Network
- Discriminant  Analysis
- Support Vector Machine

# Logistic Regression Modeling

Population y intercept
Population Slope Coefficient
Independent Variable
Random Error term, or residual
Dependent Variable

$$y = \underbrace{\beta_0 + \beta_1 x}_{\text{Linear component}} + \underbrace{\varepsilon}_{\text{Random Error component}}$$

- Linear Regression:
  - Modeling the Outcome (*y*) vs. Predictors (*x*)
  - Statistical Assumptions: "L-I-N-E"
- From Linear Regression to Linear Models:
  - McCullagh & Nelder (1989) coined the term "**Generalized Linear Model**" and used "**LINK**" function to incorporate outcome variables that are not *Normally Distributed*
    - Logistic, Binomial, Multinomial, or Poisson
- The Logistic Regression Model "Logit":
  - Outcome is 0/1 so we model the probability "p" (bounded between 0 and 1)
    - *Probability to Odds : P/(1 – P), to remove the upper bound (1)*
    - *Odds to Logit: Log (Odds) = A + B X, to remove the lower bound (0)*

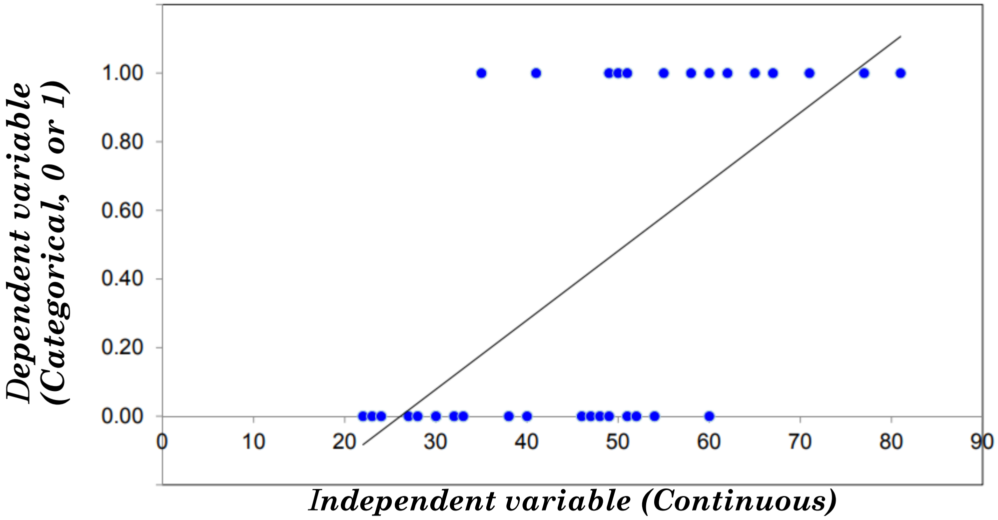




*L-I-N-E: Linearity, Independence, Normality, Equal variance*

# Linear vs. Logistic Regression Modeling

| | | Dependent Variable | |
|---|---|---|---|
| | | **Categorical** | **Continuous** |
| **Independent Variable** | **Categorical** | **Chi-squared test** | **ANOVA** |
| | **Continuous** | **Logistic Regression** | **Linear Regression** |




For Illustrative purpose only



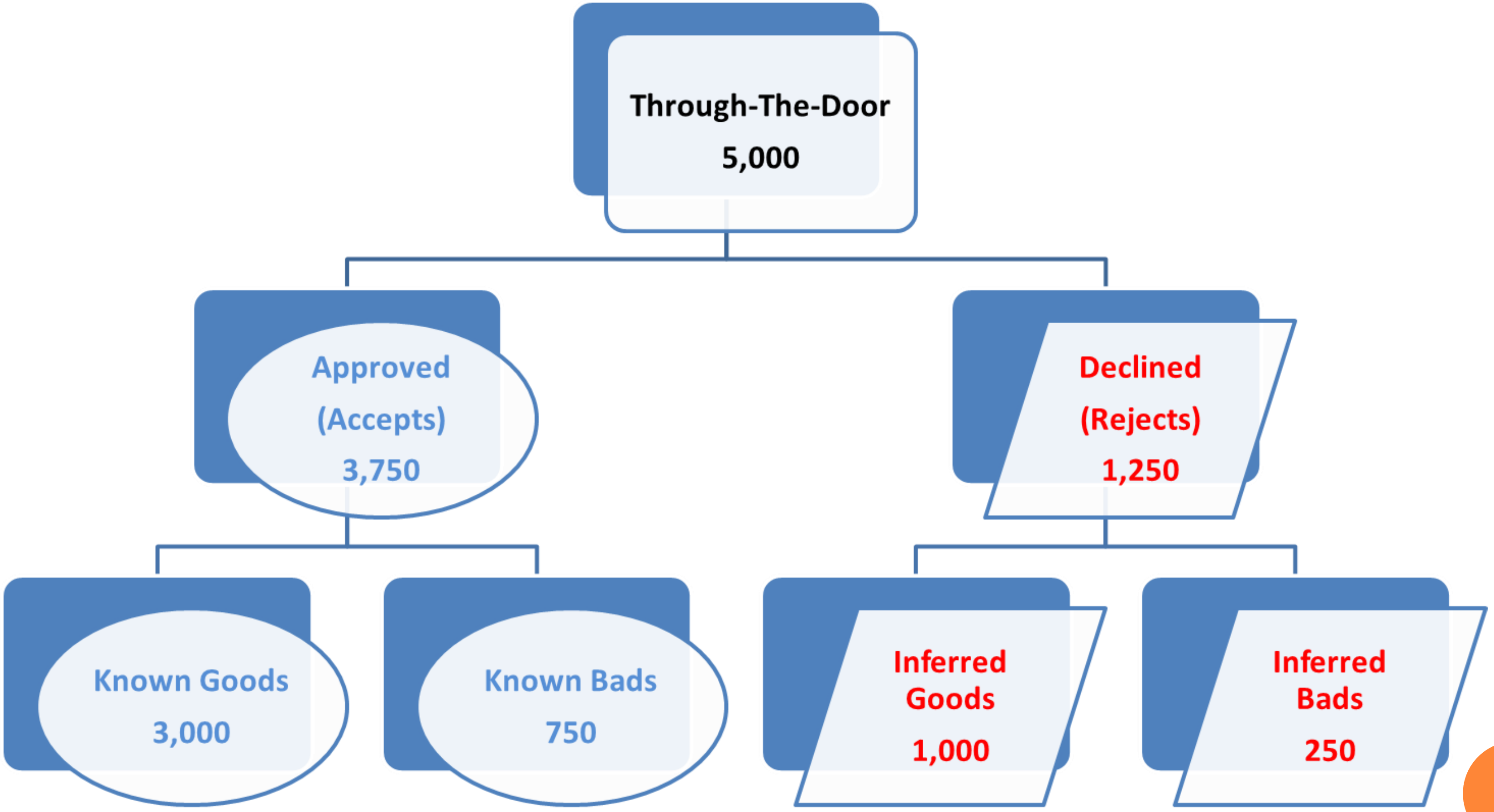


Source: http://www.feinberg.northwestern.edu/sites/bcc/docs/2011-lectures/lecture3linearlogisticregression2011.pdf

# Reject Inference Methods (*Acquisition risk score Modeling*)

- *Assign All Rejects to "Bads"*
- *Approve all applications*
- *Bureau Performance*
- *Augmentation*
- *Simple Augmentation*
- *Fuzzy Augmentation*
- *Iterative Reclassification*
- *Assign Rejects in the same % to reflect the Accepts*
- *Nearest Neighbor*
- *Memory-Based Reasoning*
- *Bureau Score Migration*
- *Manual Adjustment of Weight Of Evidence*
- *Parceling*

# Reject Inference Methods
# (*Acquisition risk score Modeling*)






For Illustrative purpose only

# The 5 Pillars of Credit Risk Score

- 35% Payment History "**Credit Performance**"
- 30% Amount Owed "**Credit Utilization**"
- 15% Length of History "**Credit History**"
- 10% Trades  "**Credit Mix**"
- 10% Inquiries "**Credit Newness**"





Source: How Scores are calculated: https://www.myfico.com/credit-education/whats-in-your-credit-score

# Acquisition Credit Risk Score

*Probability of Default "PD" Modeling*

# Sample Portfolio Acquisition Data

| Decision | APPL_ID | Acct_ID | Status | bad_90d_24m | INT_VAR1 | INT_VAR2 | ... | INT_VAR50 | CB_VAR1 | CB_VAR2 | ... | CB_VAR500 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| APPROVE | 1700001 | 1800001 | good | 1 | 5000 | 180 | ... | 30 | 21715 | 555 | ... | 0 |
| APPROVE | 1700002 | 1800002 | good | 1 | 5000 | 450 | ... | 58 | 25755 | 8000 | ... | 3000 |
| APPROVE | 1700003 | 1800003 | bad | 0 | 3000 | 3360 | ... | 46 | 68175 | 0 | ... | 0 |
| APPROVE | 1700004 | 1800004 | good | 1 | 2500 | 0 | ... | 24 | 30300 | 148 | ... | 1 |
| APPROVE | 1700005 | 1800005 | good | 1 | 3500 | 600 | ... | 26 | 53530 | 0 | ... | 0 |
| APPROVE | 1700006 | 1800006 | good | 1 | 0 | 800 | ... | 36 | 80800 | 525 | ... | 0 |
| APPROVE | 1700007 | 1800007 | good | 1 | 10000 | 1200 | ... | 44 | 38380 | 0 | ... | 0 |
| APPROVE | 1700008 | 1800008 | good | 1 | 450 | 555 | ... | 27 | 69185 | 3000 | ... | 180 |
| APPROVE | 1700009 | 1800009 | good | 1 | 15000 | 765 | ... | 32 | 58075 | 0 | ... | 0 |
| APPROVE | 1700010 | 1800010 | bad | 0 | 16500 | 3360 | ... | 41 | 30300 | 0 | ... | 0 |
| APPROVE | 1700011 | 1800011 | good | 1 | 0 | 7896 | ... | 34 | 87365 | 0 | ... | 0 |
| APPROVE | 1700012 | 1800012 | good | 1 | 3000 | 0 | ... | 29 | 45450 | 700 | ... | 0 |
| APPROVE | 1700013 | 1800013 | good | 1 | 5000 | 2500 | ... | 30 | 184325 | 825 | ... | 0 |
| APPROVE | 1700014 | 1800014 | good | 1 | 3500 | 260 | ... | 37 | 77770 | 2500 | ... | 0 |
| APPROVE | 1700015 | 1800015 | bad | 0 | 4162 | 0 | ... | 68 | 28280 | 0 | ... | 0 |
| DECLINE | 1700016 | | | | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| DECLINE | 1700017 | | | | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| DECLINE | 1700018 | | | | 5000 | 2000 | ... | 68 | 37875 | 9000 | ... | 250 |
| DECLINE | 1700019 | | | | 750 | 500 | ... | 36 | 30300 | 47500 | ... | 3360 |
| DECLINE | 1700020 | | | | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| DECLINE | 1700021 | | | | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| DECLINE | 1700022 | | | | 10000 | 655 | ... | 22 | 27270 | 9050 | ... | 4500 |
| DECLINE | 1700023 | | | | 2000 | 500 | ... | 45 | 57065 | 8000 | ... | 8250 |
| DECLINE | 1700024 | | | | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| DECLINE | 1700025 | | | | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |

For Illustrative purpose only

# Credit Risk Score: Model Development Process

- Event definition
- Data preparation
- KGB Data Partition
- KGB Variable Analysis
- KGB Scorecard Modeling
- Reject inference, KIGB
- KIGB Data Partition
- KIGB Variable Analysis
- KIGB Scorecard Modeling
- KIGB Scorecard Scaling
- Scorecard Monitoring





For Illustrative purpose only

## EVENT "DEFAULT" DEFINITION

- When developing a credit risk score, the definition of default ("Bad") must be clearly established via:
  - Regulatory (e.g., Basel II, IFRS 9)
  - Risk analytics – Portfolio Maturity or Strategic Analysis
- Two Default components:
  - ***Default Event:*** 60 DPD, 90 DPD, Bankruptcy, or Charge off
  - ***Default Horizon:*** 12-month, 18-month, or 24-month
- Event Modeling Assumption:
  - Future performance is reflected in past performance
- Example of a Default ("Bad") definition:
  - Any account that 90 DPD (days past due) in the 24 month on book or performance window





For Illustrative purpose only

# DATA PREPARATION

- ACCEPTS:
  - Approved population data
  - KGB: Known Good Bad
- REJECTS:
  - Declined population data
  - NKGB: Not Known Good Bad

**PERFORMANCE (24-MONTH) & SAMPLE (2016'Q1) WINDOW**


For Illustrative purpose only

# KBG Data Partition

- KGB In-time sample (2016'Q1)
  - *Train dataset, 70%*
  - *Validation dataset, 30%*
- KGB Out-of-Time (OOT) sample:
  - *OOT1 – December 2015 data*
  - *OOT2 – April 2016 data*





For Illustrative purpose only

# KGB Variable Analysis

- **Variable Transformation:**
  - ***Binning (Weight of Evidence)***
    - *Widely accepted as the "Gold standard"*
    - *Has good interpretability with business implications*
  - ***Statistical Modeling:***
    - *Generalized Linear Model (GLM)*
    - *Generalized Addictive Model (GAM)*
- **Variable Reduction:**
  - *Bootstrapping*
  - *Factor Analysis*
  - *Information Value*
  - *Clustering Analysis*
  - *Kolmogorov–Smirnov*
  - *Principal Component Analysis*

# Weight of Evidence (WOE)

## SAS proc hpbin or SAS E-Miner or SAS Macro

$$\text{WOE} = \left[\ln\left(\frac{\%Bad_i}{\%Good_i}\right)\right] \times 100$$

| Score Band | # Records | %Records | # Bad | # Good | % Bad | % Good | WOE | % Bad Rate |
|---|---|---|---|---|---|---|---|---|
| 1-50 | 20,534 | 6.5% | 1,581 | 18,953 | 16.1% | 6.2% | 95.30 | 7.7% |
| 51-100 | 23,000 | 7.3% | 1,382 | 21,618 | 14.1% | 7.1% | 68.69 | 6.0% |
| 101-150 | 18,567 | 5.9% | 888 | 17,679 | 9.0% | 5.8% | 44.58 | 4.8% |
| 151-200 | 37,842 | 12.0% | 1,362 | 36,480 | 13.8% | 11.9% | 14.91 | 3.6% |
| 201-250 | 45,687 | 14.5% | 1,325 | 44,362 | 13.5% | 14.5% | -7.40 | 2.9% |
| 251-300 | 55,698 | 17.6% | 1,392 | 54,306 | 14.2% | 17.8% | -22.69 | 2.5% |
| 301-350 | 45,768 | 14.5% | 961 | 44,807 | 9.8% | 14.7% | -40.52 | 2.1% |
| 351-400 | 33,458 | 10.6% | 569 | 32,889 | 5.8% | 10.8% | -62.01 | 1.7% |
| 401-450 | 20,093 | 6.4% | 241 | 19,852 | 2.5% | 6.5% | -97.43 | 1.2% |
| 451-500 | 15,008 | 4.8% | 135 | 14,873 | 1.4% | 4.9% | 126.51 | 0.9% |
| Total | 315,655 | 100.0% | 9,836 | 305,819 | 100.0% | 100.0% | | |

*Notes:*

*~SAS E-Miner: SAS Enterprise Miner*

*~PROC HPBIN: High-Performance SAS procedure*

*~SAS Macro: There are several WOE SAS macros available freely online or in SAS papers which can be customized*

For Illustrative purpose only

# KGB Scorecard Modeling using SAS

*Logistic Regression using SAS:*

```
proc logistic data=ACCEPTS_KGB_data desc;
model bad_KGB = WOE_X1, WOE_X2, WOE_X3, etc / <options> ;
weight <sampling weights>;
run;
```



- *Alternatively, you can perform Logistic Regression modeling using any analytical software with Statistical modeling capabilities e.g., R, STATA, SPSS, MATLAB , Mathematica, Python, Julia*



For Illustrative purpose only

# Reject Inference: Fuzzy Augmentation Method

- Fuzzy Augmentation Method (FAM) is a two-step approach which incorporates not just the probability of a reject being bad, but also the possibility of being accepted in the first place.

- ***FAM Step 1 is* Classification**:
  - Build a model using Known Goods and Bads data → The KGB model
  - Score the REJECTS using the KGB model
  - Determine $p$(good) and $p$(bad) for each REJECT
  - Weigh REJECTED "good" with $p$(good) and REJECTED "bad" with $p$(bad)

- ***FAM Step 2 is* Augmentation**:
  - Combine REJECTS with ACCEPTS
    - Known Inferred Goods and Bads (KIGB) data

# Reject Inference: Fuzzy Augmentation Method

*Logistic Regression using SAS:*

```
proc logistic data=ACCEPTS_KGB_data desc;
model bad_KGB = WOE_X1, WOE_X2, WOE_X3, etc / <options> ;
weight sample_wt;
score data=REJECT_RAW_data out = rejects_scored;
run;
```



- *Alternatively, you can perform Logistic Regression modeling using any analytical software with Statistical modeling capabilities e.g., R, STATA, SPSS, MATLAB , Mathematica, Python, Julia*



For Illustrative purpose only

# REJECT INFERENCE: FUZZY AUGMENTATION METHOD

```
/* SPLIT SCORED REJECTS INTO GOOD & BADS*/
* data rejects_bad
*        rejects_good;
*     set rejects_scored;
* run;
/* CREATE WEIGHTS FOR REJECTED BAD*/
* data INFERRED_rejects_bad;
*   set rejects_bad;
*   bad_KIGB=1;
*   wgt_KIGB=sample_wt*PROB_BAD;
*   GROUP="REJECT_BAD ";
* run;
/* CREATE WEIGHTS FOR REJECTED GOOD*/
* data INFERRED_rejects_good;
*   set rejects_good;
*   bad_KIGB=0;
*   wgt_KIGB=sample_wt*(1-PROB_BAD);
*   GROUP="REJECT_GOOD";
* run;
```

```
/* KGB population */
o data KGB_data;
o    set accepts_kgb_data;
o    bad_KIGB=bad_ever90_24m;
o    wgt_KIGB=sample_wt;
o    GROUP="ACCEPT_KGB ";
o run;
/* KIGB data = KBG + Inferred
REJECTS */
o data KIGB_data;
o    set KGB_data
o        INFERRED_rejects_bad
o        INFERRED_rejects_good;
o    by id;
o run;
```



For Illustrative purpose only

# KIBG Data Partition

- KGB In-time sample (2016'Q1)
  - *Train dataset, 70%*
  - *Validation dataset, 30%*
- KGB Out-of-Time (OOT) sample:
  - *OOT1 – December 2015 data*
  - *OOT2 – April 2016 data*





For Illustrative purpose only

# KIGB Variable Analysis

- **Variable Transformation:**
  - **Binning (Weight of Evidence):**
    - *Widely accepted as the "Gold standard"*
    - *Has good interpretability with business implications*
  - **Statistical modeling:**
    - *Generalized Linear Model (GLM)*
    - *Generalized Addictive Model (GAM)*
- **Variable Reduction:**
  - Bootstrapping
  - Factor Analysis
  - Information Value
  - Clustering Analysis
  - Kolmogorov–Smirnov
  - Principal Component Analysis

# KIGB Scorecard Modeling using SAS

*Logistic Regression using SAS:*

```
proc logistic data=KIGB_data desc;
model bad_KIGB = WOE_X1, WOE_X2, WOE_X3, etc / <options> ;
weight wgt_KIGB;
output out=<output of logit, probabilities, etc.> ;
run;
```

*Alternatively, you can perform Logistic Regression modeling using any analytical software with Statistical modeling capabilities e.g., R, STATA, SPSS, MATLAB , Mathematica, Python, Julia*





For Illustrative purpose only

# KIGB Scorecard Scaling

- Probabilities transformed into *Scores* using **Scaling Parameters**:
  - Base Score
  - Base Odds
  - Point-to-Double-Odds

```
/*Sample Scaling Parameters*/
%let PDO = 20;
%let Base_Score = 200;
%let Base_Odds = 50;
/* Compute FACTOR and OFFSET */
factor= &PDO / log(2);
offset = &Base_Score - (Factor*Log(&Base_Odds));
/* Compute Risk Score */
Risk_Score = OFFSET  + FACTOR*logit_kigb;
```

For Illustrative purpose only

# Scorecard Monitoring Metrics

- Divergence (D)
- Gini's Index (GINI)
- Vintage Analysis (VA)
- Override Analysis (OA)
- Kolmogorov–Smirnov (KS)
- EvA: Expected vs. Actual (EvA)
- Population Stability Index (PSI)
- Characteristic Stability Index (CSI)
- Total Population Stability Index (TPSI)
- Log Odds & Point-to-Double Analysis (PDO)
- Area Under Receiver Operating Curve (AUROC or simply ROC)





*PSI for "Booked" or Approved population*
*TPSI for "Through-The-Door, TTD" that's both Approved and Declined population*

# APPENDIX

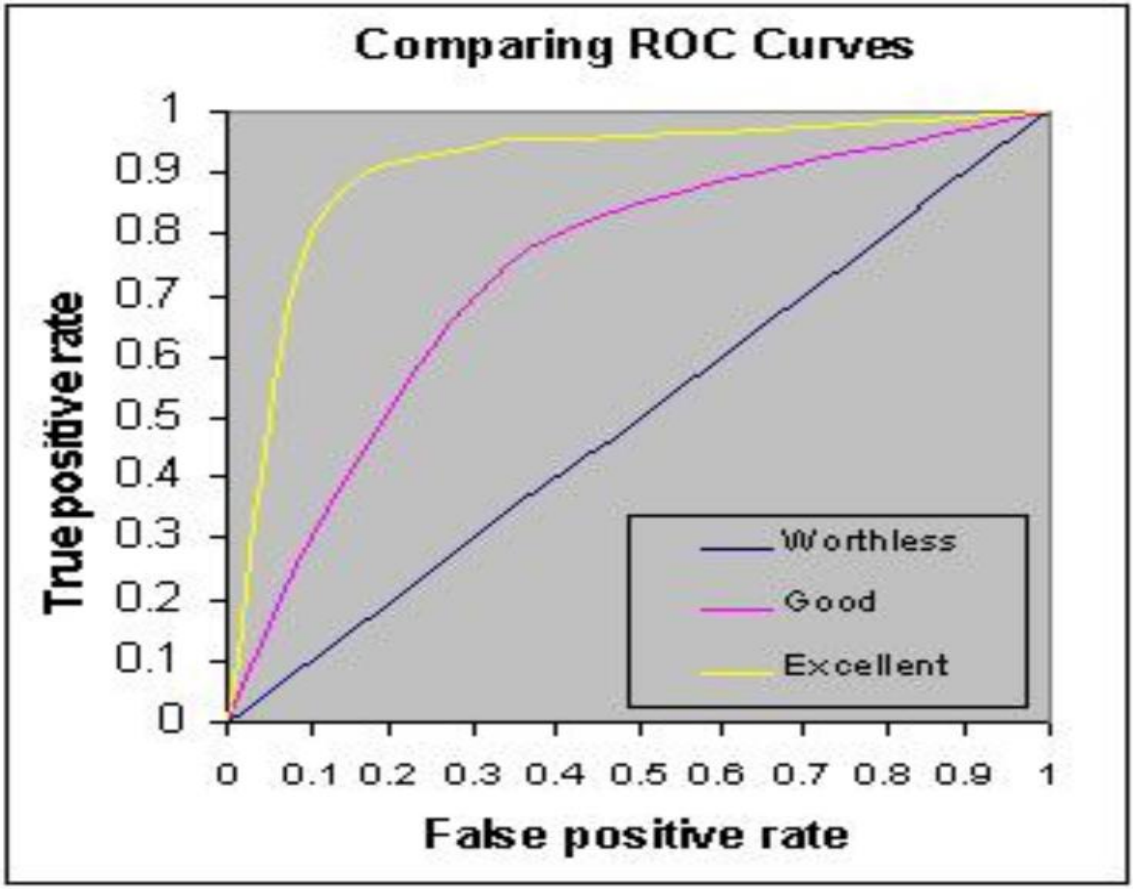
Scorecard Validation and Monitoring Charts

| Population Stability Index | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Risk Score Range | TRAIN freq | VAL freq | TRAIN % | VAL % | % Difference | % Ratio | Natural Log of Ratio | Contribution to Index |
| < 350 | 2,500 | 2,345 | 10.51% | 9.77% | 0.74% | 0.929633 | -0.072965 | 0.000540 |
| 351 -< 400 | 2,503 | 2,401 | 10.53% | 10.01% | 0.52% | 0.950694 | -0.050563 | 0.000262 |
| 401 -< 450 | 2,805 | 2,402 | 11.80% | 10.01% | 1.79% | 0.848690 | -0.164061 | 0.002929 |
| 451 -< 500 | 2,177 | 2,403 | 9.16% | 10.02% | -0.86% | 1.093965 | 0.089809 | 0.000773 |
| 501 -< 550 | 2,444 | 2,404 | 10.28% | 10.02% | 0.26% | 0.974856 | -0.025465 | 0.000066 |
| 551 -< 600 | 2,509 | 2,405 | 10.55% | 10.03% | 0.53% | 0.949995 | -0.051299 | 0.000271 |
| 601 -< 650 | 2,001 | 2,406 | 8.42% | 10.03% | -1.61% | 1.191680 | 0.175364 | 0.002829 |
| 651 -< 700 | 2,512 | 2,407 | 10.57% | 10.03% | 0.53% | 0.949651 | -0.051661 | 0.000275 |
| 701 -< 750 | 2,098 | 2,408 | 8.82% | 10.04% | -1.21% | 1.137524 | 0.128854 | 0.001564 |
| > 751 | 2,227 | 2,409 | 9.37% | 10.04% | -0.68% | 1.072074 | 0.069595 | 0.000470 |
| | | | | | | | **PSI =** | **1.00%** |

| EvA: Expected versus Actual Bad with Rank-ordering | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Risk Score Range | Expected Bads freq | Actual Bads freq | Expected % | Actual % | % Difference | % Ratio | Natural Log of Ratio | Contribution to Index |
| Low -< 350 | 8,769 | 8,339 | 36.88% | 33.09% | 3.80% | 0.897099 | -0.108589 | 0.004121 |
| 351 -< 400 | 7,261 | 6,796 | 30.54% | 26.97% | 3.58% | 0.882936 | -0.124503 | 0.004451 |
| 401 -< 450 | 3,619 | 4,292 | 15.22% | 17.03% | -1.81% | 1.118783 | 0.112241 | 0.002029 |
| 451 -< 500 | 1,982 | 2,239 | 8.34% | 8.89% | -0.55% | 1.066012 | 0.063925 | 0.000352 |
| 501 -< 550 | 818 | 1,468 | 3.44% | 5.83% | -2.39% | 1.693319 | 0.526691 | 0.012562 |
| 551 -< 600 | 523 | 1,012 | 2.20% | 4.02% | -1.82% | 1.827354 | 0.602869 | 0.010963 |
| 601 -< 650 | 323 | 326 | 1.36% | 1.29% | 0.07% | 0.951828 | -0.049371 | 0.000032 |
| 651 -< 700 | 285 | 528 | 1.20% | 2.09% | -0.89% | 1.745680 | 0.557144 | 0.004981 |
| 701 -< 750 | 140 | 141 | 0.59% | 0.56% | 0.03% | 0.951828 | -0.049371 | 0.000014 |
| 751 >- High | 56 | 62 | 0.24% | 0.25% | -0.01% | 1.035721 | 0.035098 | 0.000003 |
| | | | | | | | **PSI =** | **3.95%** |

For Illustrative purpose only

# Area Under Receiver Operating Curve (AUROC)

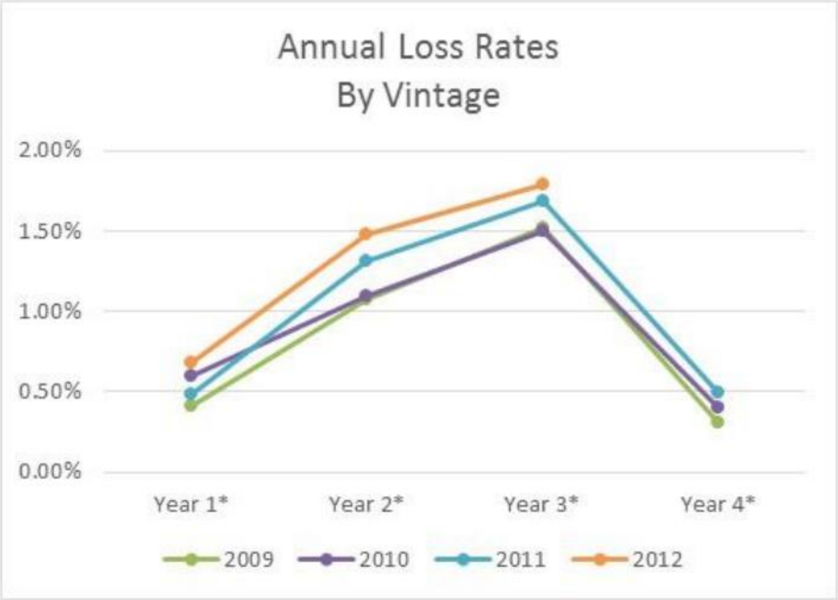


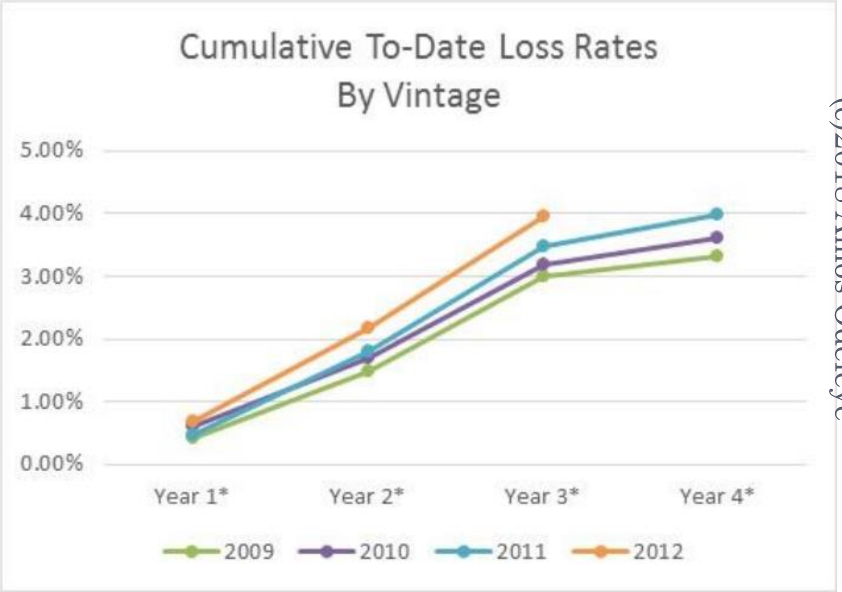

For Illustrative purpose only

(c)2018 Amos Odeleye

Source: http://gim.unmc.edu/dxtests/roc3.htm

# VINTAGE ANALYSIS



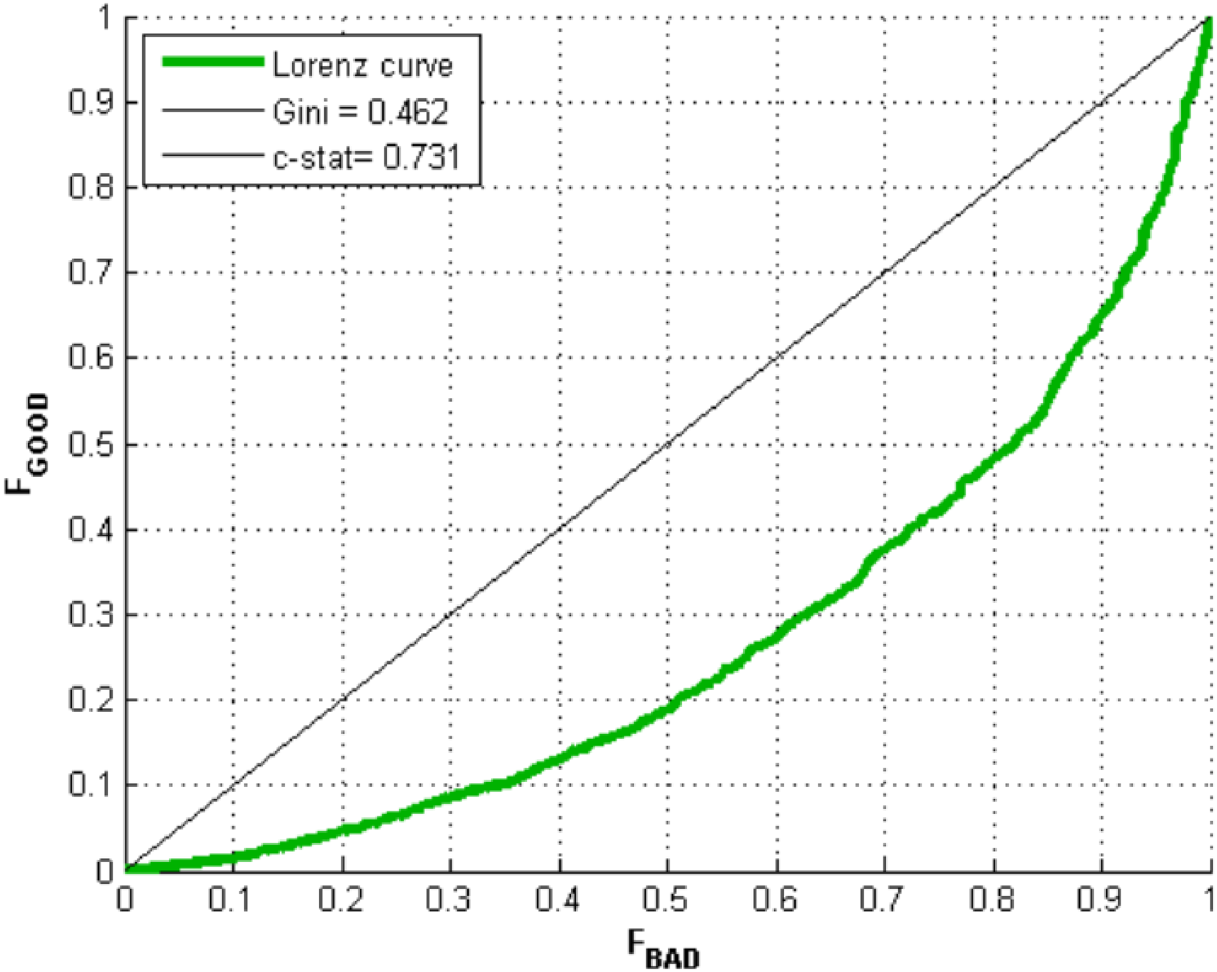




(c)2018 Amos Odeleye

Source: https://bsr.stlouisfed.org/perspectives/final_fedperspectives_cecl_october%202015.pdf

For Illustrative purpose only

# Lorenz curve

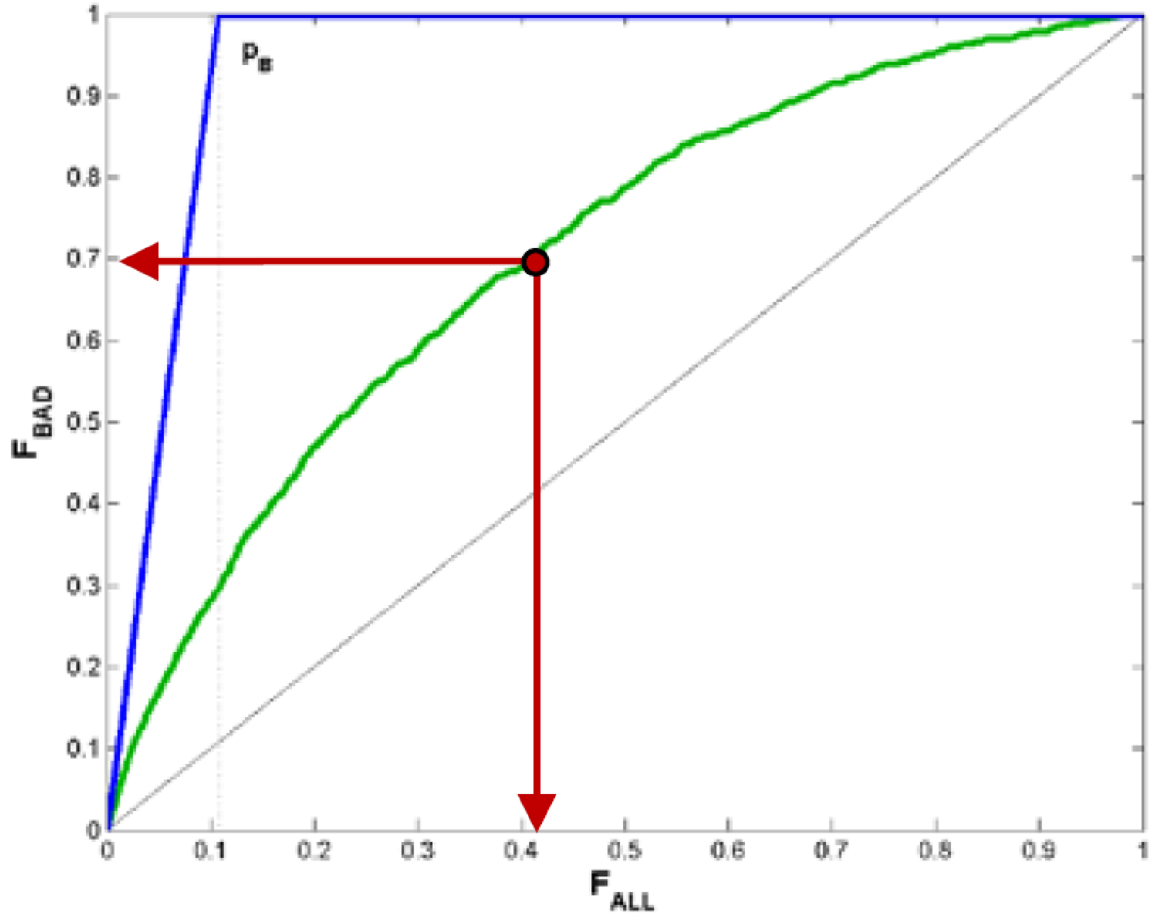



For Illustrative purpose only



Source: https://www.business-school.ed.ac.uk/crc/wp-content/uploads/sites/55/2017/03/Paper-65-Paper.pdf

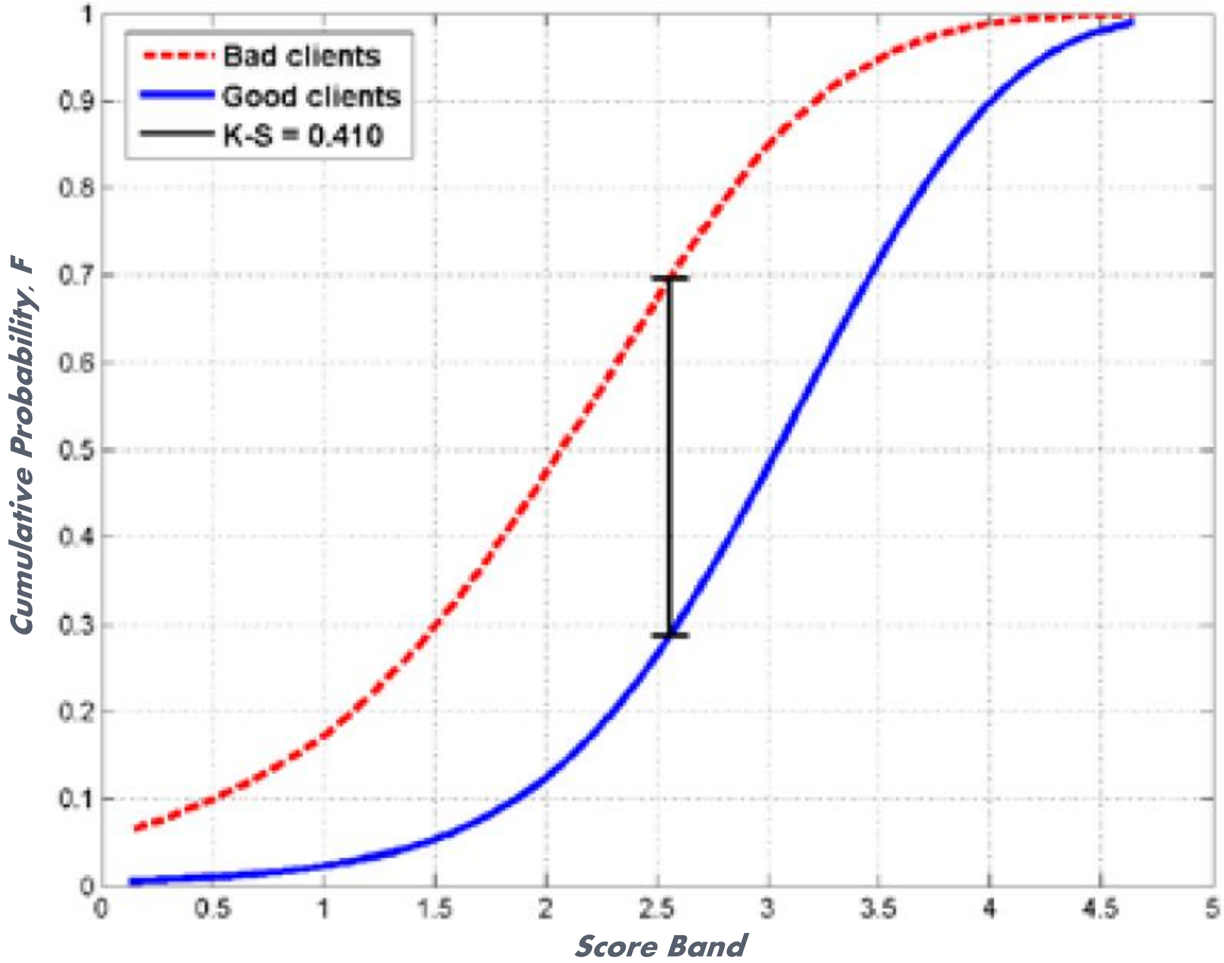
# Gains chart




*Also known as Cumulative Accuracy Plot (CAP)*

For Illustrative purpose only

# KOLMOGOROV–SMIRNOV (KS)

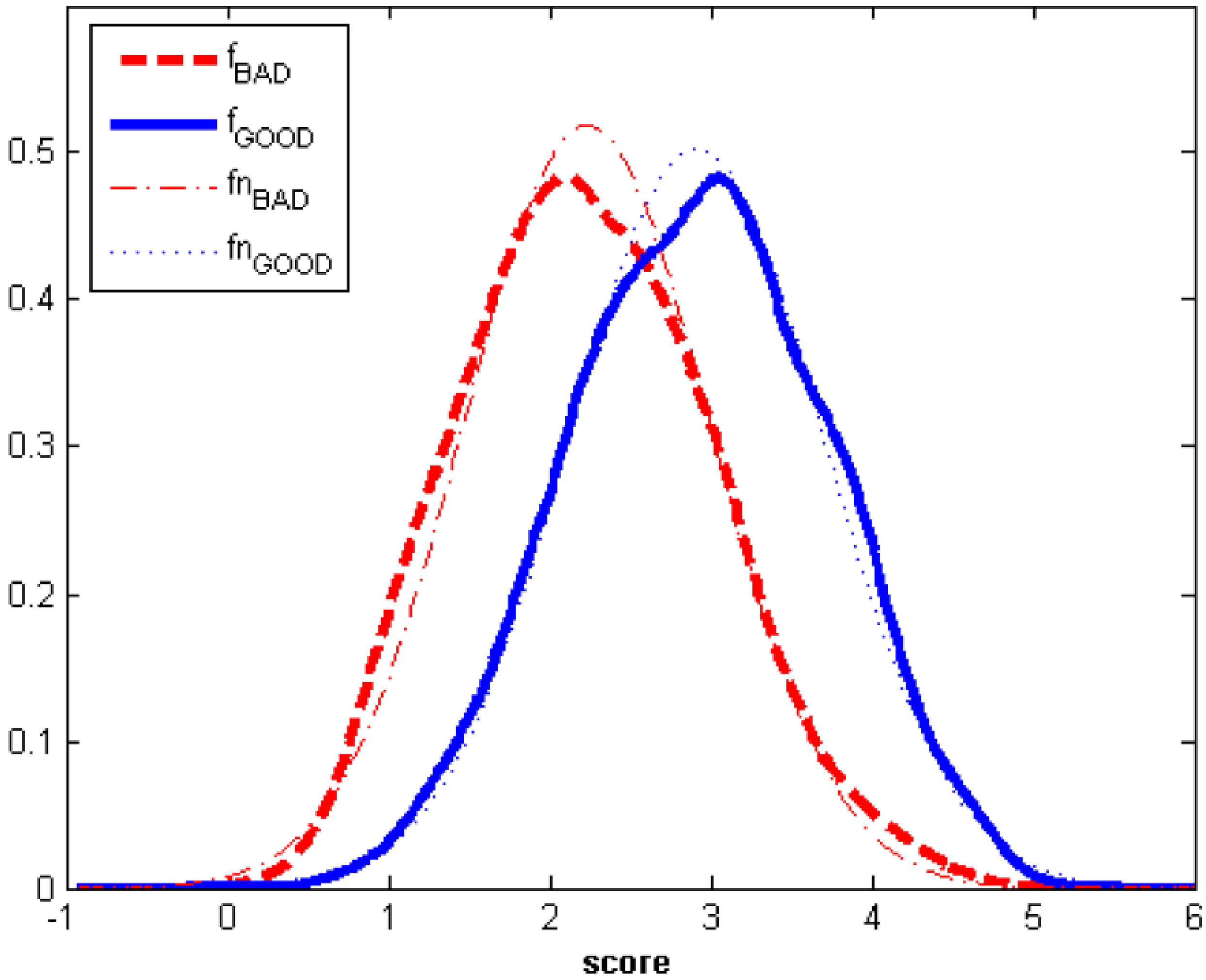



For Illustrative purpose only



Source: https://www.business-school.ed.ac.uk/crc/wp-content/uploads/sites/55/2017/03/Paper-65-Paper.pdf

# DIVERGENCE




For Illustrative purpose only

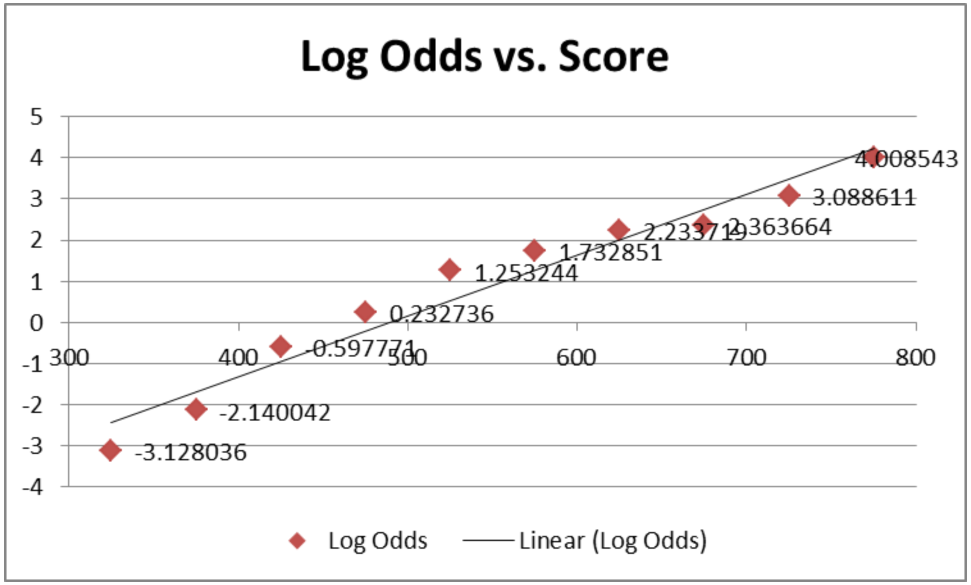
(c)2018 Amos Odeleye



Source: https://www.business-school.ed.ac.uk/crc/wp-content/uploads/sites/55/2017/03/Paper-65-Paper.pdf

| Log Odd and Point-to-Double-Odds (PDO) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Risk Score Range** | **Score** | **Bads freq** | **Goods freq** | **Total freq** | **Bads %** | **Goods %** | **Odds** | **Log Odds** |
| Low -< 350 | 325 | 8,769 | 1,231 | 10,000 | 36.88% | 1.62% | 0.043804 | -3.128036 |
| 351 -< 400 | 375 | 7,261 | 2,739 | 10,000 | 30.54% | 3.59% | 0.117650 | -2.140042 |
| 401 -< 450 | 425 | 3,619 | 6,381 | 10,000 | 15.22% | 8.37% | 0.550036 | -0.597771 |
| 451 -< 500 | 475 | 1,982 | 8,018 | 10,000 | 8.34% | 10.52% | 1.262048 | 0.232736 |
| 501 -< 550 | 525 | 818 | 9,182 | 10,000 | 3.44% | 12.05% | 3.501686 | 1.253244 |
| 551 -< 600 | 575 | 523 | 9,477 | 10,000 | 2.20% | 12.43% | 5.656758 | 1.732851 |
| 601 -< 650 | 625 | 323 | 9,677 | 10,000 | 1.36% | 12.69% | 9.334518 | 2.233719 |
| 651 -< 700 | 675 | 285 | 9,715 | 10,000 | 1.20% | 12.75% | 10.629828 | 2.363664 |
| 701 -< 750 | 725 | 140 | 9,860 | 10,000 | 0.59% | 12.94% | 21.946574 | 3.088611 |
| 751 >- High | 775 | 56 | 9,944 | 10,000 | 0.24% | 13.05% | 55.066565 | 4.008543 |

| Score | Odds |
|---|---|
| 200 | 50 |
| 220 | 100 |
| 240 | 150 |
| 260 | 200 |
| 280 | 250 |
| 300 | 300 |
| 320 | 350 |
| 340 | 400 |
| 360 | 450 |
| 380 | 500 |
| 400 | 550 |
| 420 | 600 |
| 440 | 650 |
| 460 | 700 |




**Scaling Parameters:**

| | |
|---|---|
| **PDO** | 20 |
| **Base Odds** | 50 |
| **Base Score** | 200 |

**PDO = Ln(2) / Slope**





For Illustrative purpose only

# References

# Conference + Papers + Lecture notes


- FICO World 2018: Credit Boot Camp Course
- SAS Paper 344-2010: Improving Credit Risk Scorecards with Memory-Based Reasoning to Reject Inference with SAS® Enterprise Miner™ by Billie Anderson, Susan Haller, Naeem Siddiqi, James Cox, and David Duling, SAS Institute Inc.
- SAS Paper: Building Better Credit Scores using Reject Inference and SAS by Steve Fleming, Clarity Services Inc.
- SAS Paper ST-160: Reject Inference Methodologies in Credit Risk Modeling by Derek Montrichard, Canadian Imperial Bank of Commerce, Toronto, Canada
- Guoping Zeng, Qi Zhao, A rule of thumb for reject inference in credit scoring, Mathematical Finance Letters, Vol 2014 (2014), Article ID 2
- How to Measure Quality of Credit Scoring Models by Martin Řezáč and František Řezáč
- Population Stability and Model Performance Metrics Replication for Business Model at SunTrust Bank by Bogdan Gadidov and Benjamin McBurnett, )
- ***FED***Perspectives: An Overview of the Current Expected Credit Loss Model (CECL) and Supervisory Expectations by Steve Merriett, Joanne Wakim, Shuchi Satwah
- Expanding the Use of Weight of Evidence and Information Value to Continuous Dependent Variables for Variable Reduction and Scorecard Development by Alec Zhixiao Lin and Tung-Ying Hsieh
- Introduction to Linear Regression and Logistic Regression by Jungwha “Julia” Lee, PhD, MPH (Biostatistics Collaboration Center – Northwestern University)

# SAS Books & Resources

- Griffin, Hill, & Lim: Using SAS for Econometrics
- SAS Programming 1: Essentials, SAS Institute
- SAS Programming 2: Data Manipulation Techniques, SAS Institute
- Littell, Stroup, & Freund (2002). SAS for Linear Model
- Slaughter & Delwiche (2008). The Little SAS Book: A Primer
- Schreier (2008). PROC SQL by Example: Using SQL within SAS®
- Allison: Logistic Regression using SAS (Theory & Applications)
- Stokes, Davis, & Koch: Categorical Data analysis using the SAS System
- Carpenter (2004). Carpenter's Complete Guide to the SAS Macro Language
- Cody Ron (2018). Learning SAS by Examples
- Stokes, Davis, & Koch: Categorical Data analysis using the SAS System
- Carpenter (2004). Carpenter's Complete Guide to the SAS Macro Language
- Littell, Milliken, Stroup, Wolfinger, & Schabenberger (2006). SAS® for Mixed Models
- SAS® Certification Prep Guide: Base Programming for SAS®9, Fourth Edition
- SAS® Certification Prep Guide: Advanced Programming for SAS®9, Fourth Edition
- SAS Documentation: https://support.sas.com/documentation/
- SAS Procedure by Name search: https://support.sas.com/documentation/cdl/en/allprodsproc/63875/HTML/default/viewer.htm#a003135046.htm
- SAS Procedure by Products: https://support.sas.com/documentation/cdl/en/allprodsproc/63875/HTML/default/viewer.htm#a003178332.htm
- SAS Global Forum (Annual event)
- SAS User Groups – Local and Regional

# Textbooks

- Rencher: Methods of Multivariate Analysis
- Baesens, Rosch, & Schenle: Credit Risk Analytics
- Hosmer & Lemeshow: Applied Logistic Regression
- Crouhy, Galai, & Mark: Essential of Risk Management
- Kleinbaum & Klein: Logistic Regression, A Self Learning Text
- Lawrence & Solomon: Managing Customer Lending Business
- Johnson & Wichern (2007). Applied Multivariate Statistical Analysis
- Lyn C. Thomas: Consumer Credit Models (Pricing, Profits and Portfolios)
- Mays & Lynas: Credit Scoring for Risk Managers (Handbook for Lenders)
- Wooldridge (2010). Econometric analysis of Cross sectional & Panel data
- Lyn Thomas , Jonathan Crook & David Edelman. Credit Scoring and Its Applications
- Naeem Siddiqi: Intelligent Credit Scoring (Building and Implementing Better Credit Risk Scorecards)
- Sharda, Delan, & Turban: Business Intelligent and Business Analytics (Systems for Decision Support)
- (Electronic Version): StatSoft, Inc. (2013). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/

# Textbooks

- Hayashi: Econometrics
- Allen: Financial Risk Management
- Greene (2011). Econometric Analysis
- Kuhn & Johnson: Predictive Modeling
- Lange: Numerical Analysis for Statisticians
- Monahans: Numerical methods of Statistics
- Hull: Risk management & Financial institutions
- Lam: Enterprise Risk Management: From Incentives to Controls
- Bluhm, Overbeck, & Wagner: Introduction to Credit Risk Modeling
- Bouteille & Coogan-Pushner: Handbook of Credit Risk Management
- Friedman, Hastie, & Tibshirani (2009). The Elements of Statistical Learning
- Kutner, Nachtsheim , Neter, & Li (2004). Applied Linear Statistical Models
- Stroup (2012). Generalized Linear Mixed Models
- P. McCullagh & J.A. Nelder (1989). Generalized Linear Models
- T. J. Hastie & R. J. Tibshirani (1990). Generalized Additive Models
- James, Witten, Hastie, & Tibshirani (2009). An Introduction to Statistical Learning

## OCC Lending handbooks:

- OCC Comptroller's Handbook: *Retail Lending*
  - https://www.occ.treas.gov/publications/publications-by-type/comptrollers-handbook/retail-lending/index-ch-retail-lending.html
- OCC Comptroller's Handbook: *Credit Card Lending*
  - https://www.occ.treas.gov/publications/publications-by-type/comptrollers-handbook/credit-card-lending/index-ch-credit-card-lending.html

# Question or Comment

You can contact me at:

Amos Odeleye

OduaTechnology@gmail.com