SCORECARD CREATION

Dr. Aric LaBarr

Institute for Advanced Analytics

Process Flow

Data Collection
- Variable Selection
- Sample Size
- Sample / Performance Window

Data Cleaning
- Eliminate Duplicates
- Examine / Remove Outliers

Variable Grouping and Selection
- Weights of Evidence (WOE)
- Information Value (IV)
- Gini Criterion

Initial Scorecard Creation
- Logistic Regression
- Accuracy
- Threshold
- Assessment

Reject Inference
- Remove bias resulting from exclusion of rejects

Final Scorecard Creation
- Final Model Assessment

INITIAL SCORECARD CREATION

Initial Scorecard Model

- The scorecard is (typically) based on a logistic regression model:

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

- p is the posterior probability of default (PD) given the inputs.

Blasphemy!!!

- Wait, so I'm going through all that math just to throw things back into the logistic regression I was trying to avoid in the first place?!?!?!?!

Different Inputs

- Instead of using the original variables for the model, scorecard models have the binned variables as their foundation.

- 2 Differing Approaches (with similar results):
  1. Traditional approach – use WOE scores as new variables
  2. Another approach – use binned variables as new variables

Different Inputs

- Instead of using the original variables for the model, scorecard models have the binned variables as their foundation.

Observation	Target	Bureau Score	Bureau Score Bin (R)	Bureau Score WOE (R)
1	0	757	716 – 765	1.0914
2	1	NA	Missing	-0.6972
3	0	626	605 – 629	-0.9586
4	0	693	665 – 716	0.1776
5	0	706	665 – 716	0.1776
6	0	673	665 – 716	0.1776
7	0	730	716 – 765	1.0914

Different Inputs

- Instead of using the original variables for the model, scorecard models have the binned variables as their foundation.

Traditional Approach

Observation	Target	Bureau Score	Bureau Score Bin (R)	Bureau Score WOE (R)
1	0	757	716 – 765	1.0914
2	1	NA	Missing	-0.6972
3	0	626	605 – 629	-0.9586
4	0	693	665 – 716	0.1776
5	0	706	665 – 716	0.1776
6	0	673	665 – 716	0.1776
7	0	730	716 – 765	1.0914

Different Inputs

- Instead of using the original variables for the model, scorecard models have the binned variables as their foundation.

- Inputs are still treated as **continuous**.
- All variables now on the same scale.
- Model **coefficients** are desired output for the scorecard.
- **Coefficients** now serve as measures of variable importance.
- Fewer number of variables (not a ton of categorical variables).

Bureau Score WOE (R)
1.0914
-0.6972
-0.9586
0.1776
0.1776
0.1776
1.0914

Different Inputs

```
smbinning.gen(df = train, ivout = result_all_sig$bureau_score,
              chrname = “bureau_score_bin")
```

Different Inputs

```
smbinning.gen(df = train, ivout = result_all_sig$bureau_score,
              chrname = “bureau_score_bin")
```

Observation	Target	Bureau Score	Bureau Score Bin (R)	Bureau Score WOE (R)
1	0	757	716 – 765	1.0914
2	1	NA	Missing	-0.6972
3	0	626	605 – 629	-0.9586
4	0	693	665 – 716	0.1776
5	0	706	665 – 716	0.1776
6	0	673	665 – 716	0.1776
7	0	730	716 – 765	1.0914

Different Inputs – Traditional Approach

```
for (i in 1:nrow(train)) {

    bin_name <- "bureau_score_bin"

    bin <- substr(train[[bin_name]][i], 2, 2)
    woe_name <- "bureau_score_WOE"

    if(bin == 0) {

      bin <- dim(result_all_sig$bureau_score$ivtable)[1] - 1
      train[[woe_name]][i] <- result_all_sig$bureau_score$ivtable[bin, "WoE"]

    } else {
      train[[woe_name]][i] <- result_all_sig$bureau_score$ivtable[bin, "WoE"]
    }
}
```

Different Inputs – Traditional Approach

Observation	Target	Bureau Score	Bureau Score Bin (R)	Bureau Score WOE (R)
1	0	757	716 – 765	1.0914
2	1	NA	Missing	-0.6972
3	0	626	605 – 629	-0.9586
4	0	693	665 – 716	0.1776
5	0	706	665 – 716	0.1776
6	0	673	665 – 716	0.1776
7	0	730	716 – 765	1.0914

Initial Scorecard – Traditional Approach

```
initial_score <- glm(data = train, bad ~ tot_derog_WOE +
                                         tot_tr_WOE +
                                         age_oldest_tr_WOE +
                                         tot_rev_line_WOE +
                                         rev_util_WOE +
                                         bureau_score_WOE +
                                         down_pyt_WOE +
                                         ltv_WOE,
                     weights = train$weight, family = "binomial")
summary(initial_score)
```

Initial Scorecard – Traditional Approach

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6969  -0.7432  -0.4273  -0.1679   3.3704

Coefficients:
                  Estimate Std. Error z value              Pr(>|z|)
(Intercept)       -2.98190    0.04101 -72.706 < 0.0000000000000002 ***
tot_derog_WOE     -0.14478    0.08285  -1.747              0.08055 .
tot_tr_WOE        -0.04041    0.12726  -0.318              0.75084
age_oldest_tr_WOE -0.28207    0.09501  -2.969              0.00299 **
tot_rev_line_WOE  -0.38840    0.07963  -4.878           0.00000107 ***
bureau_score_WOE  -0.77495    0.05833 -13.286 < 0.0000000000000002 ***
rev_util_WOE      -0.23923    0.07643  -3.130              0.00175 **
down_pyt_WOE      -0.39379    0.14828  -2.656              0.00791 **
ltv_WOE           -0.86395    0.10116  -8.541 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6910.4  on 4376  degrees of freedom
Residual deviance: 6080.4  on 4368  degrees of freedom
AIC: 6185.1
```

Different Inputs

- Instead of using the original variables for the model, scorecard models have the binned variables as their foundation.

Another Approach

Observation	Target	Bureau Score	Bureau Score Bin (R)	Bureau Score WOE (R)
1	0	757	716 – 765	1.0914
2	1	NA	Missing	-0.6972
3	0	626	605 – 629	-0.9586
4	0	693	665 – 716	0.1776
5	0	706	665 – 716	0.1776
6	0	673	665 – 716	0.1776
7	0	730	716 – 765	1.0914

Different Inputs

- Instead of using the original variables for the model, scorecard models have the binned variables as their foundation.

- Inputs are still treated as **categorical**.
- Need LRT to evaluate p-values.
- Model **coefficients** are desired output for the scorecard.
- Larger number of variables (tons of categorical variables).
- Scorecard creation preprogrammed into a lot of packages.

Bureau Score Bin (R)
716 – 765
Missing
605 – 629
665 – 716
665 – 716
665 – 716
716 – 765

Different Inputs

```
smbinning.gen(df = train, ivout = result_all_sig$bureau_score,
              chrname = “bureau_score_bin")
```

Different Inputs

```
smbinning.gen(df = train, ivout = result_all_sig$bureau_score,
              chrname = “bureau_score_bin")
```

Observation	Target	Bureau Score	Bureau Score Bin (R)	Bureau Score WOE (R)
1	0	757	716 – 765	1.0914
2	1	NA	Missing	-0.6972
3	0	626	605 – 629	-0.9586
4	0	693	665 – 716	0.1776
5	0	706	665 – 716	0.1776
6	0	673	665 – 716	0.1776
7	0	730	716 – 765	1.0914

Initial Scorecard – Another Approach

```
initial_score2 <- glm(data = train, bad ~ tot_derog_bin +
                                          tot_tr_bin +
                                          age_oldest_tr_bin +
                                          tot_rev_line_bin +
                                          rev_util_bin +
                                          bureau_score_bin +
                                          down_pyt_bin +
                                          ltv_bin,
                      weights = train$weight, family = "binomial")
summary(initial_score2)
```

Initial Scorecard – Another Approach

```
Coefficients:
                            Estimate Std. Error z value              Pr(>|z|)
(Intercept)                  2.46437    0.16306  15.113 < 0.0000000000000002 ***
age_oldest_tr_bin01 <= 73    0.22710    0.21363   1.063             0.287764
age_oldest_tr_bin02 <= 181   0.43125    0.22467   1.920             0.054919 .
age_oldest_tr_bin03 > 181    0.80804    0.24045   3.361             0.000778 ***
tot_rev_line_bin01 <= 1900   0.06311    0.18986   0.332             0.739605
tot_rev_line_bin02 <= 11519  0.35959    0.20546   1.750             0.080085 .
tot_rev_line_bin03 <= 22084  0.49078    0.22921   2.141             0.032263 *
tot_rev_line_bin04 > 22084   0.85830    0.23751   3.614             0.000302 ***
...
ltv_bin03 <= 110            -0.54225    0.10383  -5.222         0.00000017674 ***
ltv_bin04 > 110             -1.03094    0.10994  -9.377 < 0.0000000000000002 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6910.4  on 4376  degrees of freedom
Residual deviance: 6080.4  on 4368  degrees of freedom
AIC: 6185.1
```

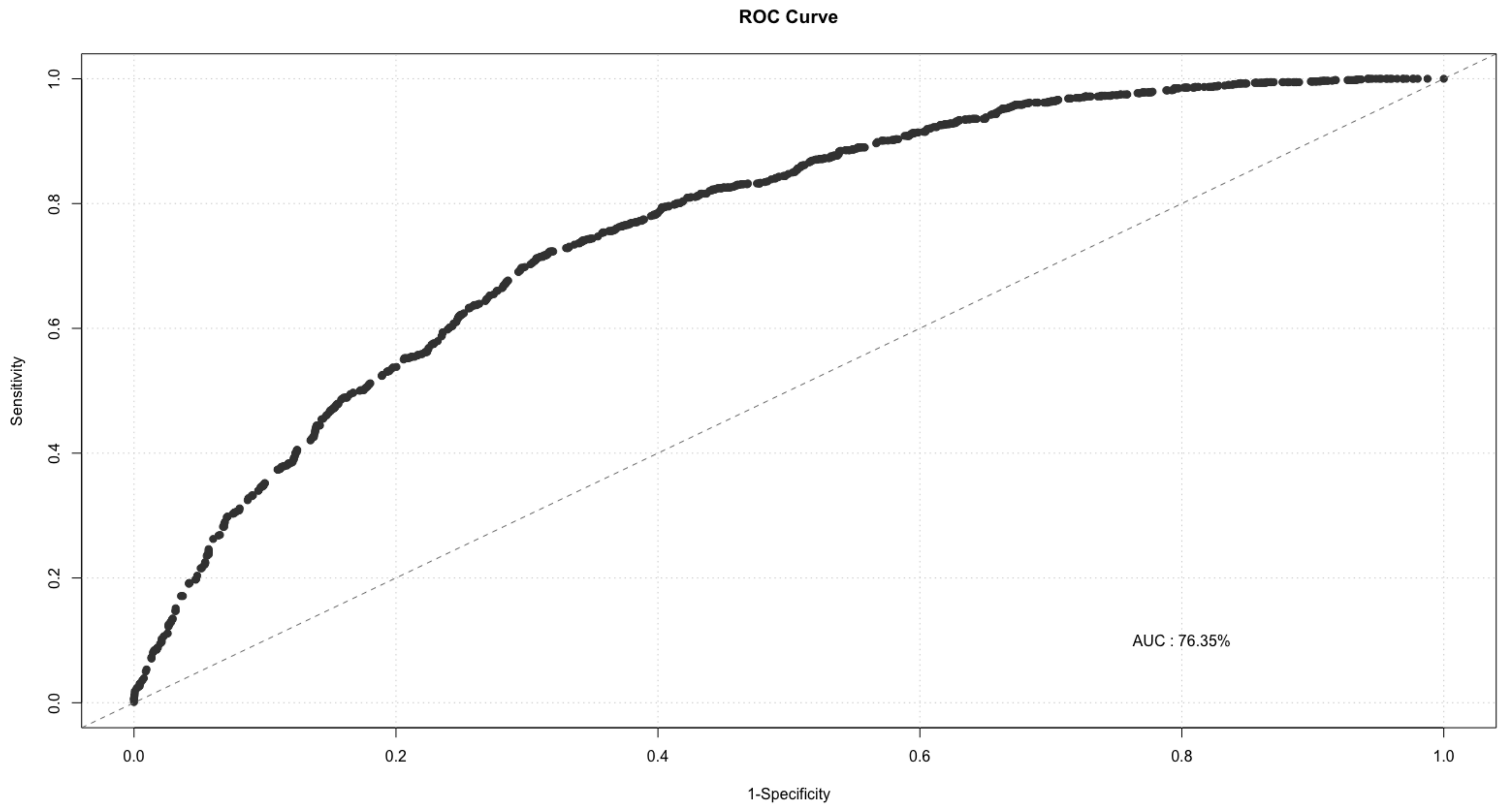
Model Evaluation

- Variable significance – review using “standard” output of logistic regression, but don’t forget **business logic**.
- Overall performance of model – AUC (area under ROC curve, also called c) is the most popular criterion.
- This is only a **preliminary scorecard**.
- Final scorecard is created after reject inference is performed.

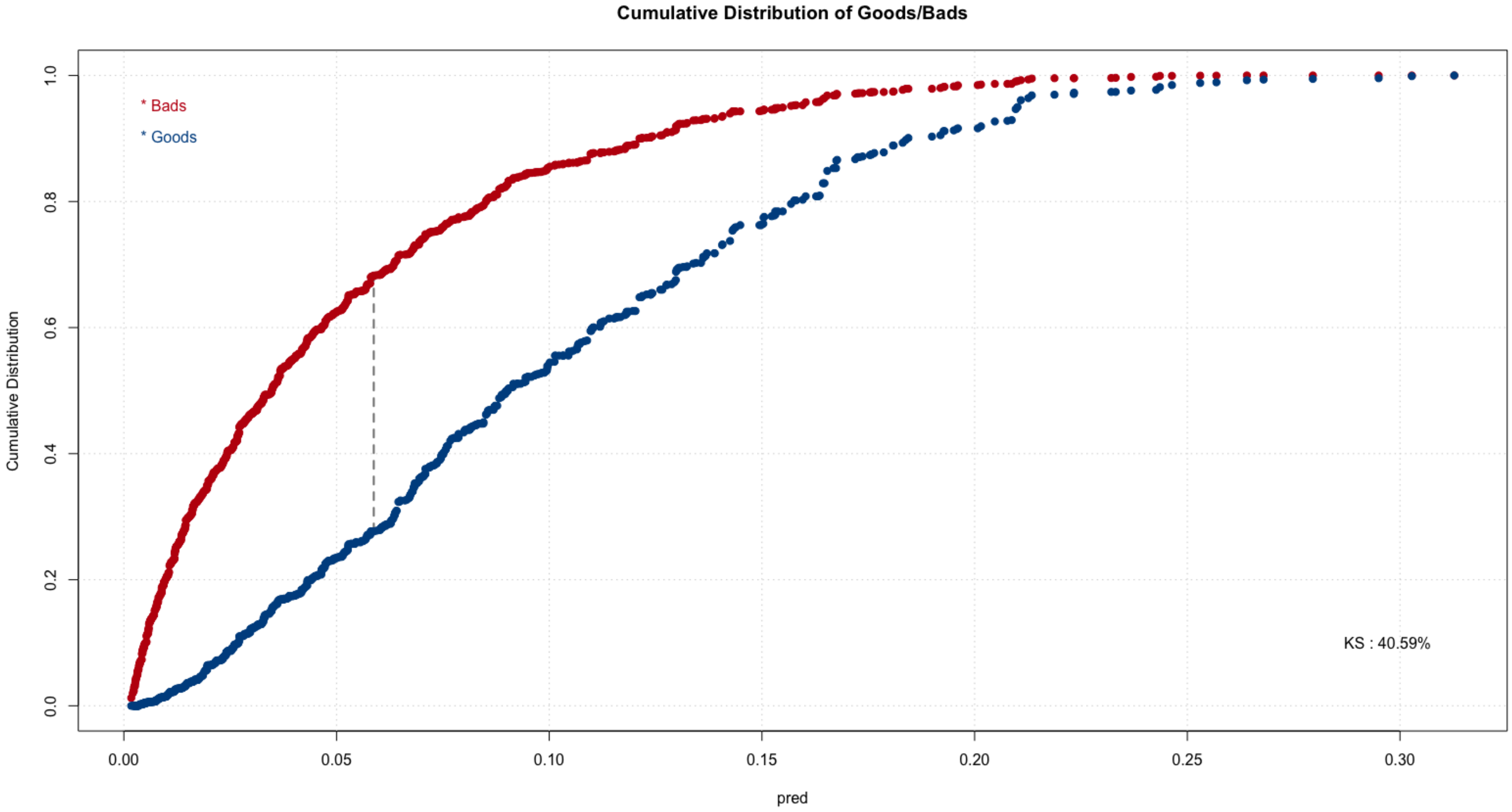
Model Evaluation – Traditional Approach

```
smbinning.metrics(dataset = train, prediction = "pred",
                  actualclass = "bad", report = 1)
smbinning.metrics(dataset = train, prediction = "pred",
                  actualclass = "bad", plot = "ks")
smbinning.metrics(dataset = train, prediction = "pred",
                  actualclass = "bad", plot = "auc")
```

Model Evaluation – Traditional Approach



Model Evaluation – Traditional Approach



Model Evaluation – Comparison

Metric	Traditional Approach (WOE Variables)	Another Approach (Binned Variables)
AUC	0.7635	0.7667
KS	0.4059	0.4057

SCALING THE SCORECARD

Scaling the Scorecard

- The relationship between odds and scores is represented by a linear function:

$$Score = Offset + Factor \times \log(odds)$$

- If the scorecard is developed using "odds at a certain score" and "points to double the odds" (PDO), $Factor$ and $Offset$ can be calculated using the simultaneous equations:

$$Score = Offset + Factor \times \log(odds)$$

$$Score + PDO = Offset + Factor \times \log(2 \times odds)$$

Scaling the Scorecard

- Solving the equations for PDO, you get the following results:

$$PDO = Factor \times \log(2)$$

- Therefore,

$$Factor = \frac{PDO}{\log(2)}$$

$$Offset = Score - Factor \times \log(odds)$$

Scaling the Scorecard – Example

- If a scorecard were scaled where the developer wanted odds of 50:1 at 600 points and wanted the odds to double every 20 points (PDO = 20), *Factor* and *Offset* would be:

$$Factor = \frac{20}{\log(2)} = 28.8539$$

$$Offset = 600 - (28.8539 \times \log(50)) = 487.123$$

- Therefore, each score corresponding to each set of odds can be calculated as follows:

$$Score = 487.123 + 28.8539 \times \log(odds)$$

Scaling the Scorecard – Example

- If a scorecard were scaled where the developer wanted odds of 50:1 at 600 points and wanted the odds to double every 20 points (PDO = 20), *Factor* and *Offset* would be:

$$Factor = \frac{20}{\log(2)} = 28.8539$$

$$Offset = 600 - (28.8539 \times \log(50)) = 487.123$$

- Therefore, each score corresponding to each set of odds can be calculated as follows:

$$Score = 487.123 + 28.8539 \boxed{\times \log(odds)}$$

Why people still love logistic regression!

Scaling the Scorecard – Example

- If a scorecard were scaled where the developer wanted odds of 50:1 at 600 points and wanted the odds to double every 20 points (PDO = 20), *Factor* and *Offset* would be:

$$Factor = \frac{20}{\log(2)} = 28.8539$$

$$Offset = 600 - (28.8539 \times \log(50)) = 487.123$$

- Therefore, each score corresponding to each set of odds can be calculated as follows:

$$Score = 487.123 + 28.8539 \boxed{\times \log(odds)}$$

This is predicted value from logit function.

Scaling the Scorecard – Example

Score	Odds
600	50.0
601	51.8
604	57.4
.	
.	
.	
.	
620	100.0

Points Allocation – Traditional Approach

- The points allocated to attribute *i* of characteristic *j* are computed as follows:

$$Points_{i,j} = -\left(WOE_{i,j} \times \hat{\beta}_j + \frac{\hat{\beta}_0}{L}\right) \times Factor + \frac{Offset}{L}$$

- $WOE_{i,j}$: Weight of evidence for attribute *i* of characteristic *j*
- $\hat{\beta}_j$: Regression coefficient for characteristic *j*
- $\hat{\beta}_0$: Intercept term from model
- L: Total number of characteristics
- Points typically rounded to nearest integer.

Points Allocation – Traditional Approach

```r
pdo <- 20
score <- 600
odds <- 50
fact <- pdo/log(2)
os <- score - fact*log(odds)
var_names <- names(initial_score$coefficients[-1])

for(i in var_names) {

    beta <- initial_score$coefficients[i]
    beta0 <- initial_score$coefficients["(Intercept)"]
    nvar <- length(var_names)
    WOE_var <- train[[i]]
    points_name <- paste(str_sub(i, end = -4), "points", sep = "")

    train[[points_name]] <- -(WOE_var*(beta) + (beta0/nvar))*fact + os/nvar
}
```

Points Allocation – Traditional Approach

Observation	Target	Variables…	Observation Score
1	1	…	625
2	0	…	597
3	0	…	615
4	0	…	601
5	0	…	589
6	0	…	571
7	0	…	584

Points Allocation – Traditional Approach

WOE for Bureau Score					
Group	Values	Event Count	Non-event Count	WOE	**Scorecard Points**
1	< 603	111	112	-1.32	**50.4**
2	604 – 662	378	678	-0.74	**64.0**
3	663 – 699	185	754	0.08	**83.4**
4	700 – 717	74	440	0.46	**92.4**
5	718 – 765	75	824	1.07	**106.9**
6	> 765	15	498	2.18	**133.0**
7	MISSING	80	153	-0.68	**65.6**
Total		**918**	**3,459**		

Points Allocation – Another Approach

```
bin_model <- smbinning.scaling(initial_score2, pdo = 20, score = 600, odds = 50)

train_bin <- smbinning.scoring.gen(bin_model, dataset = train)
```

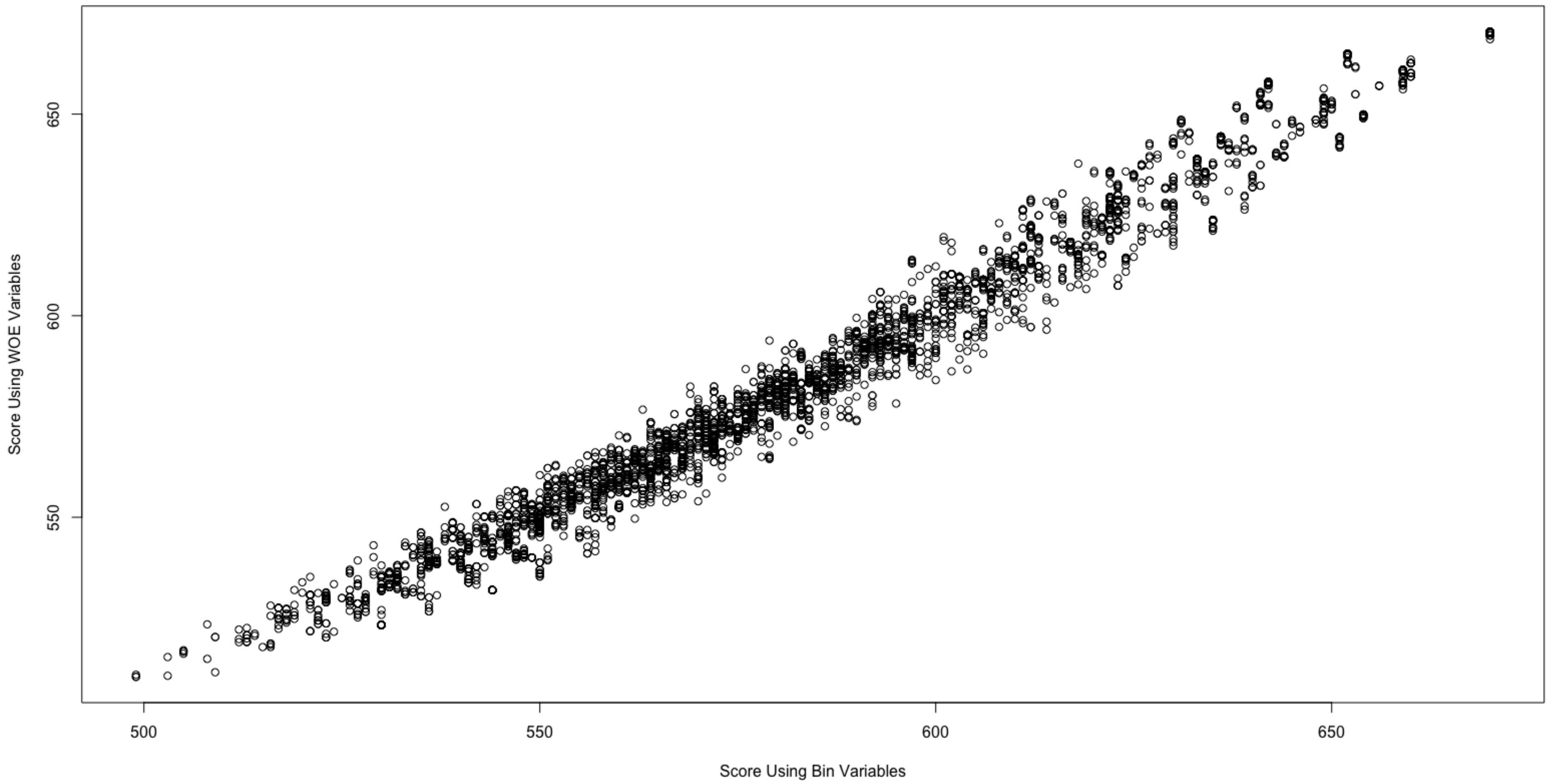
Points Allocation – Another Approach

Observation	Target	Variables…	Observation Score (T)	Observation Score (A)
1	1	…	625	611
2	0	…	597	612
3	0	…	615	618
4	0	…	601	593
5	0	…	589	597
6	0	…	571	574
7	0	…	584	584

Points Allocation – Another Approach

WOE for Bureau Score					
Group	Values	Event Count	Non-event Count	WOE	**Scorecard Points**
1	< 603	111	112	-1.32	**85**
2	604 – 662	378	678	-0.74	**103**
3	663 – 699	185	754	0.08	**125**
4	700 – 717	74	440	0.46	**134**
5	718 – 765	75	824	1.07	**146**
6	> 765	15	498	2.18	**165**
7	MISSING	80	153	-0.68	**112**
Total		**918**	**3,459**		

Score Comparison



REJECT INFERENCE

Process Flow

Data Collection
- Variable Selection
- Sample Size
- Sample / Performance Window

Data Cleaning
- Eliminate Duplicates
- Examine / Remove Outliers

Variable Grouping and Selection
- Weights of Evidence (WOE)
- Information Value (IV)
- Gini Criterion

Initial Scorecard Creation
- Logistic Regression
- Accuracy
- Threshold
- Assessment

Reject Inference
- Remove bias resulting from exclusion of rejects

Final Scorecard Creation
- Final Model Assessment

Reject Inference

- **Reject inference** is the process of inferring the status of the rejected applicants based on the accepted applicants model in an attempt to use their information to build a scorecard that is representative of the entire applicant population.
- Reject inference is about solving sample bias so that the development sample is similar to the population to which the scorecard will be applied.

Rejected Inference

- Can we develop a scorecard without rejected applications? **YES!**
- Is it **legally permissible** to develop a scorecard without rejected applications? **YES!**
- If yes, then how **biased** would the scorecard model be? **DEPENDS!**

- *“My suggestion is to develop the scorecard using what data you have, but start saving rejected applications ASAP.”*
  Raymond Anderson, Head of Scoring at Standard Bank Africa, South Africa

Why Reject Inference?

- Initial scorecard used only **known** good and bad loans (accepted applicants only) – also called “behavioral scoring”
- Reduce bias in model and provide risk estimates for the “through-the-door” population – also called “application scoring”
- Comply with regulatory requirements (FDIC, Basel)
- Provide a scorecard that is able to generalize better to the entire credit application population.

Reject Inference Techniques

- Three common techniques for reject inference:
  1. Hard Cutoff Augmentation
  2. Parceling Augmentation
  3. Fuzzy Augmentation (DEFAULT in SAS)

Hard Cutoff Augmentation

1. Build a scorecard model using the known good/bad population (accepted applications)
2. Score the rejected applications with this model to obtain each rejected applicant's probability of default and their score on the scorecard model.
3. Create weighted cases for the rejected applicants – weight applied is the "rejection rate" which adjusts the number of sampled rejects to accurately reflect the number of rejects from population.

Hard Cutoff Augmentation

4. Set a cut-off score level above which applicant is deemed good and below applicants deemed bad.
5. Add inferred goods and bads with known goods and bads and rebuild scorecard.

Hard Cutoff Augmentation

```
rejects_scored$pred <- predict(initial_score, newdata = rejects_scored,
                               type = 'response')
rejects$bad <- as.numeric(rejects_scored$pred > 0.0545)
rejects$weight <- ifelse(rejects$bad == 1, 2.80, 0.59)
rejects$good <- abs(rejects$bad - 1)

comb_hard <- rbind(accepts, rejects)
```

Hard Cutoff Augmentation

```
rejects_scored$pred <- predict(initial_score, newdata = rejects_scored,
                               type = 'response')
rejects$bad <- as.numeric(rejects_scored$pred > 0.0545)
rejects$weight <- ifelse(rejects$bad == 1, 2.80, 0.59)
rejects$good <- abs(rejects$bad - 1)

comb_hard <- rbind(accepts, rejects)
```

Hard Cutoff Augmentation

```
rejects_scored$pred <- predict(initial_score, newdata = rejects_scored,
                               type = 'response')
rejects$bad <- as.numeric(rejects_scored$pred > 0.0545)
rejects$weight <- ifelse(rejects$bad == 1, 2.80, 0.59)
rejects$good <- abs(rejects$bad - 1)

comb_hard <- rbind(accepts, rejects)
```

Parceling Augmentation

1. Build a scorecard model using the known good/bad population (accepted applications)
2. Score the rejected applications with this model to obtain each rejected applicant's probability of default and their score on the scorecard model.
3. Create weighted cases for the rejected applicants – weight applied is the "rejection rate" which adjusts the number of sampled rejects to accurately reflect the number of rejects from population.

Parceling Augmentation

4. Define score ranges manually or automatically with simple bucketing.
5. The inferred good/bad status of the rejected applicants will be assigned **randomly** and proportional to the number of goods and bads in the accepted population within each score range.
6. If desired, apply the event rate increase factor to P(bad) to increase the proportion of bads among the rejects (oversampling with the rejects)
7. Add the inferred goods and bads back in with the known goods and bads and rebuild the scorecard.

Parceling Augmentation – Example

	Accepted Applicants				Rejected Applicants		
Score Range	# Bad	# Good	% Bad	%Good	Rejects	# Inferred Bad	# Inferred Good
< 655	0	0	0%	0%	5	?	?

Parceling Augmentation – Example

	Accepted Applicants				Rejected Applicants		
Score Range	# Bad	# Good	% Bad	%Good	Rejects	# Inferred Bad	# Inferred Good
< 655	0	0	0%	0%	5	5	0

Assume bad if no information to prove otherwise

Parceling Augmentation – Example

	Accepted Applicants				Rejected Applicants		
Score Range	# Bad	# Good	% Bad	%Good	Rejects	# Inferred Bad	# Inferred Good
< 655	0	0	0%	0%	5	5	0
655 – 665	300	360	45.5%	54.5%	190	?	?

Parceling Augmentation – Example

	Accepted Applicants				Rejected Applicants		
Score Range	# Bad	# Good	% Bad	%Good	Rejects	# Inferred Bad	# Inferred Good
< 655	0	0	0%	0%	5	5	0
655 – 665	300	360	45.5%	54.5%	190	86	?

$$\mathbf{0.455 \times 190 \approx 86}$$

Randomly assign!

Parceling Augmentation – Example

	Accepted Applicants				Rejected Applicants		
Score Range	# Bad	# Good	% Bad	%Good	Rejects	# Inferred Bad	# Inferred Good
< 655	0	0	0%	0%	5	5	0
655 – 665	300	360	45.5%	54.5%	190	86	114

$$190 - 86 = 114$$

Parceling Augmentation – Example

	Accepted Applicants				Rejected Applicants		
Score Range	# Bad	# Good	% Bad	%Good	Rejects	# Inferred Bad	# Inferred Good
< 655	0	0	0%	0%	5	5	0
655 – 665	300	360	45.5%	54.5%	190	86	114
665 – 675	450	700	39.1%	60.9%	250	98	152
…	…	…	…	…	…	…	…

```
parc <- seq(500, 725, 25)

accepts_scored$Score_parc <- cut(accepts_scored$Score, breaks = parc)
rejects_scored$Score_parc <- cut(rejects_scored$Score, breaks = parc)

table(accepts_scored$Score_parc, accepts_scored$bad)

parc_perc <- table(accepts_scored$Score_parc, accepts_scored$bad)[,2] /
                   rowSums(table(accepts_scored$Score_parc, accepts_scored$bad))

rejects$bad <- 0

rej_bump <- 1.25

for(i in 1:(length(parc) - 1)) {
    for(j in 1:length(rejects_scored$Score)) {
        if((rejects_scored$Score[j] > parc[i]) &
           (rejects_scored$Score[j] <= parc[i+1]) &
           (runif(n = 1, min = 0, max = 1) < (rej_bump*parc_perc[i]))) {
            rejects$bad[j] <- 1
        }
    }
}

table(rejects_scored$Score_parc, rejects$bad)
rejects$weight <- ifelse(rejects$bad == 1, 2.80, 0.59)
rejects$good <- abs(rejects$bad - 1)

comb_parc <- rbind(accepts, rejects)
```

```
parc <- seq(500, 725, 25)

accepts_scored$Score_parc <- cut(accepts_scored$Score, breaks = parc)
rejects_scored$Score_parc <- cut(rejects_scored$Score, breaks = parc)

table(accepts_scored$Score_parc, accepts_scored$bad)

parc_perc <- table(accepts_scored$Score_parc, accepts_scored$bad)[,2] /
                   rowSums(table(accepts_scored$Score_parc, accepts_scored$bad))

rejects$bad <- 0

rej_bump <- 1.25

for(i in 1:(length(parc) - 1)) {
    for(j in 1:length(rejects_scored$Score)) {
        if((rejects_scored$Score[j] > parc[i]) &
           (rejects_scored$Score[j] <= parc[i+1]) &
           (runif(n = 1, min = 0, max = 1) < (rej_bump*parc_perc[i]))) {
            rejects$bad[j] <- 1
        }
    }
}

table(rejects_scored$Score_parc, rejects$bad)
rejects$weight <- ifelse(rejects$bad == 1, 2.80, 0.59)
rejects$good <- abs(rejects$bad - 1)

comb_parc <- rbind(accepts, rejects)
```

Fuzzy Augmentation

1. Build a scorecard model using the known good/bad population (accepted applications)
2. Score the rejected applications with this model to obtain each rejected applicant's probability of being good, P(Good), and probability of being bad, P(Bad).
3. **Do not assign a reject to a good/bad class** – create **two weighted cases** for each rejected applicant using P(Good) and P(Bad).

Fuzzy Augmentation

4. Multiply P(Good) and P(Bad) by the user-specific rejection rate to form frequency variables.
5. For each rejected applicant, create **two observations** – one observation has a frequency variable (rejection weight × P(Good)) and a target variable of 0; other observation has a frequency variable (rejection weight × P(Bad)) and a target variable of 1.
6. Add inferred goods and bads back in with the known goods and bads and rebuild the scorecard.

Fuzzy Augmentation

```
rejects_scored$pred <- predict(initial_score, newdata = rejects_scored,
                               type = 'response')

rejects_g <- rejects
rejects_b <- rejects

rejects_g$bad <- 0
rejects_g$weight <- (1 - rejects_scored$pred)*2.80
rejects_g$good <- 1

rejects_b$bad <- 1
rejects_b$weight <- (rejects_scored$pred)*0.59
rejects_b$good <- 0

comb_fuzz <- rbind(accepts, rejects_g, rejects_b)
```

Reject Inference Techniques

- Three common techniques for reject inference:
  1. Hard Cutoff Augmentation
  2. Parceling Augmentation
  3. Fuzzy Augmentation (DEFAULT in SAS EM)

- There are other techniques as well, but are not as highly recommended.

Other Reject Inference Techniques

1. Assign all rejects to bads.
2. Assign rejects in the same proportion of goods to bads as reflected in the accepted data set.
3. Similar in-house model on different data.
4. Approve all applicants for certain period of time.
5. Clustering
6. Memory based reasoning

Other Reject Inference Techniques

1. Assign all rejects to bads
    - Appropriate only if approval rate is very high (ex. 97%) and there is a high degree of confidence in adjudication process.
2. Assign rejects in the same proportion of goods to bads as reflected in the accepted data set.
3. Similar in-house model on different data.
4. Approve all applicants for certain period of time.
5. Clustering
6. Memory based reasoning

Other Reject Inference Techniques

1. Assign all rejects to bads.
2. Assign rejects in the same proportion of goods to bads as reflected in the accepted data set.
   - Assignment done completely at random!
   - Valid only if current system has no consistency.
3. Similar in-house model on different data.
4. Approve all applicants for certain period of time.
5. Clustering
6. Memory based reasoning

Other Reject Inference Techniques

1. Assign all rejects to bads.
2. Assign rejects in the same proportion of goods to bads as reflected in the accepted data set.
3. Similar in-house model on different data.
   - Performance on similar products used as proxy.
   - Hard to pass by regulators.
4. Approve all applicants for certain period of time.
5. Clustering
6. Memory based reasoning

Other Reject Inference Techniques

1. Assign all rejects to bads.
2. Assign rejects in the same proportion of goods to bads as reflected in the accepted data set.
3. Similar in-house model on different data.
4. Approve all applicants for certain period of time.
   - Provides actual performance of rejects instead of inferred.
   - Might be “legal” problems…
5. Clustering
6. Memory based reasoning

Other Reject Inference Techniques

1. Assign all rejects to bads.
2. Assign rejects in the same proportion of goods to bads as reflected in the accepted data set.
3. Similar in-house model on different data.
4. Approve all applicants for certain period of time.
5. Clustering
6. Memory based reasoning

FINAL SCORECARD CREATION

Process Flow

Data Collection
- Variable Selection
- Sample Size
- Sample / Performance Window

Data Cleaning
- Eliminate Duplicates
- Examine / Remove Outliers

Variable Grouping and Selection
- Weights of Evidence (WOE)
- Information Value (IV)
- Gini Criterion

Initial Scorecard Creation
- Logistic Regression
- Accuracy
- Threshold
- Assessment

Reject Inference
- Remove bias resulting from exclusion of rejects

Final Scorecard Creation
- Final Model Assessment

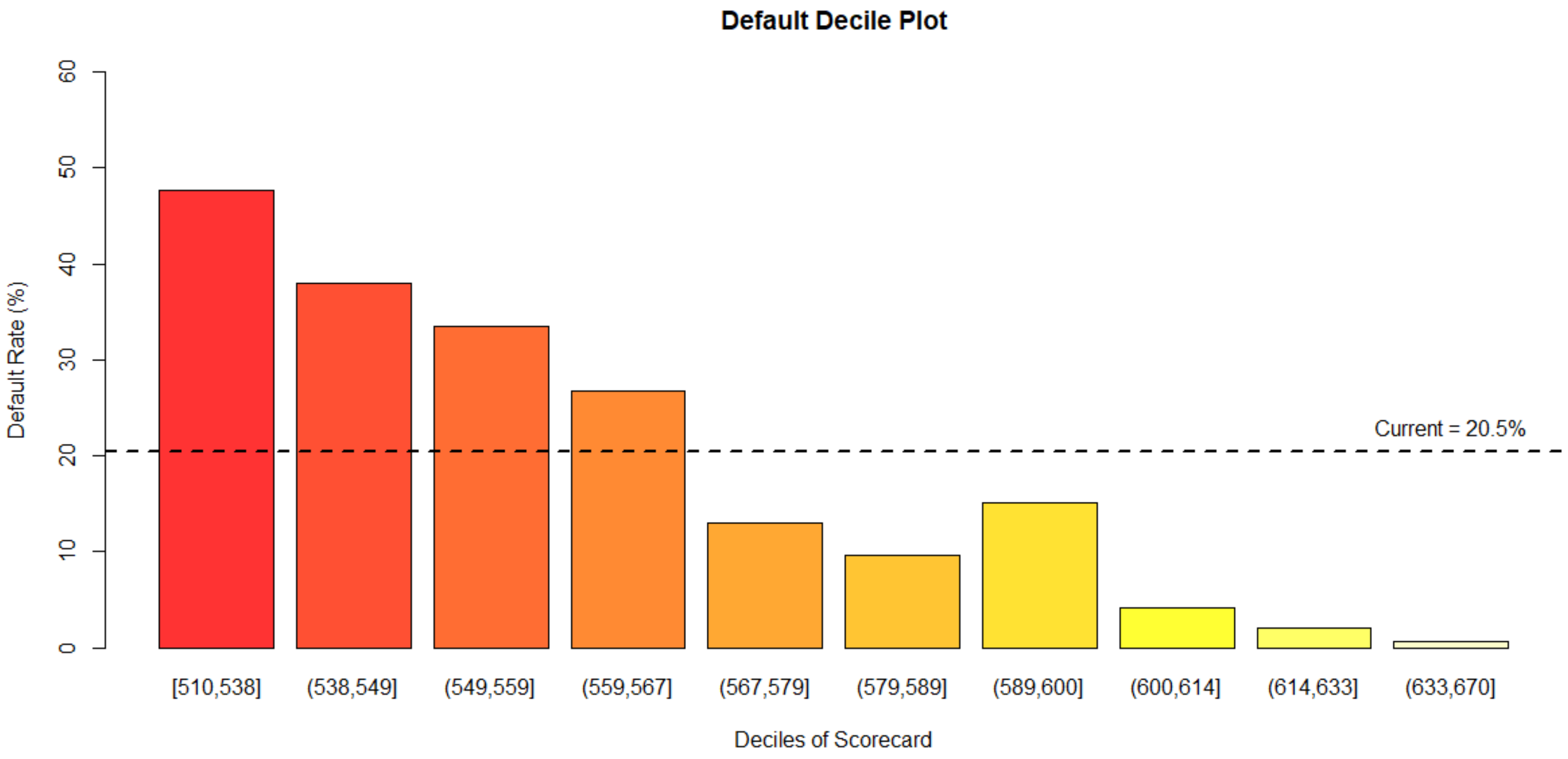
Final Scorecard Creation

- The mechanics of building the final scorecard model are identical with the initial scorecard creation except that analysis is performed **after reject inference**.
- Accuracy Measurements:
  - Repeat review of the logistic model estimated parameters, life, KS, ROC, etc.

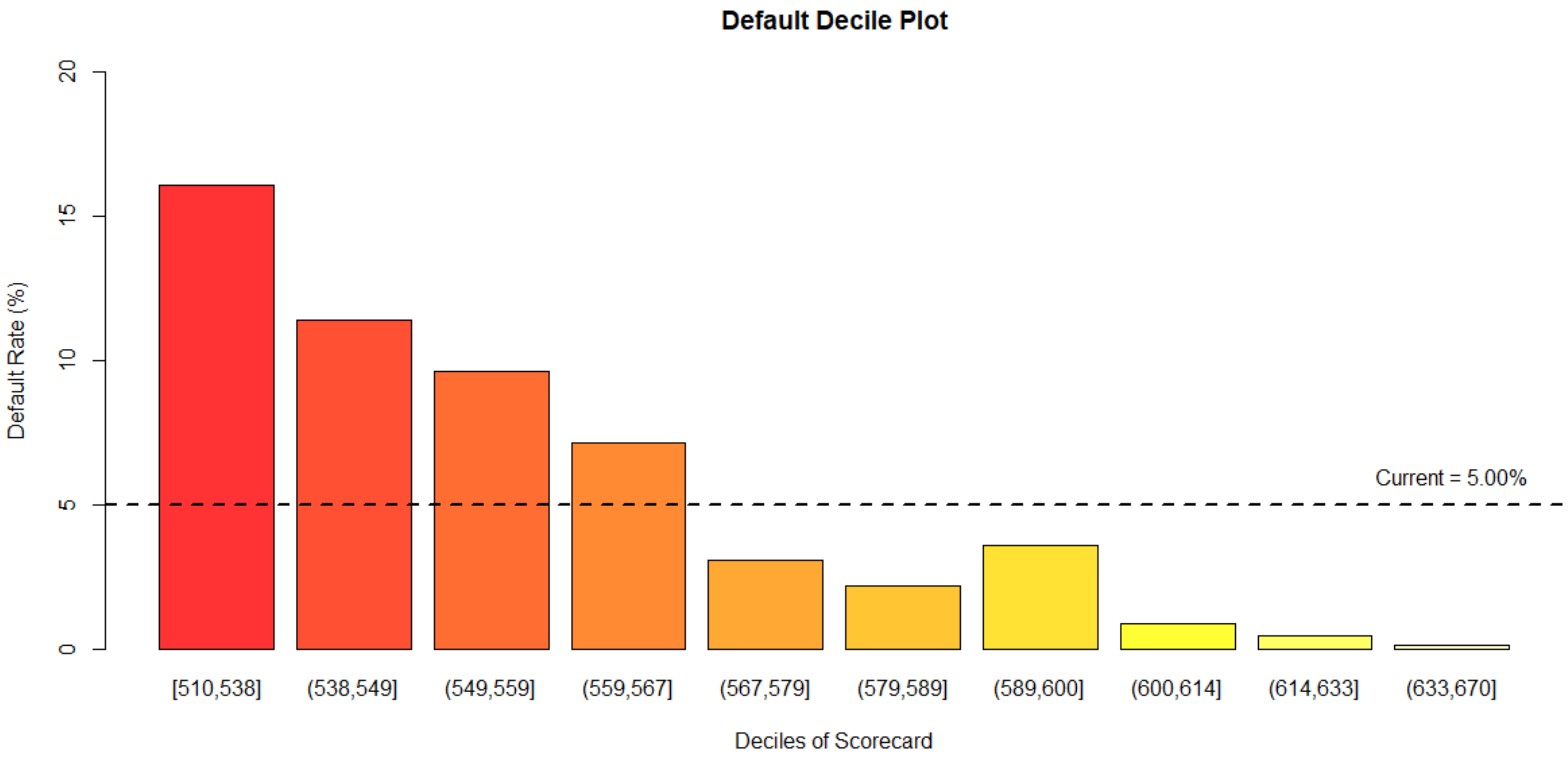
Defining Cut-off

- A new scored should be better than the last in terms of one of the following:
  - Lower bad rate for the same approval rate.
  - Higher approval rate while holding the bad rate constant.

Default Decile Plot



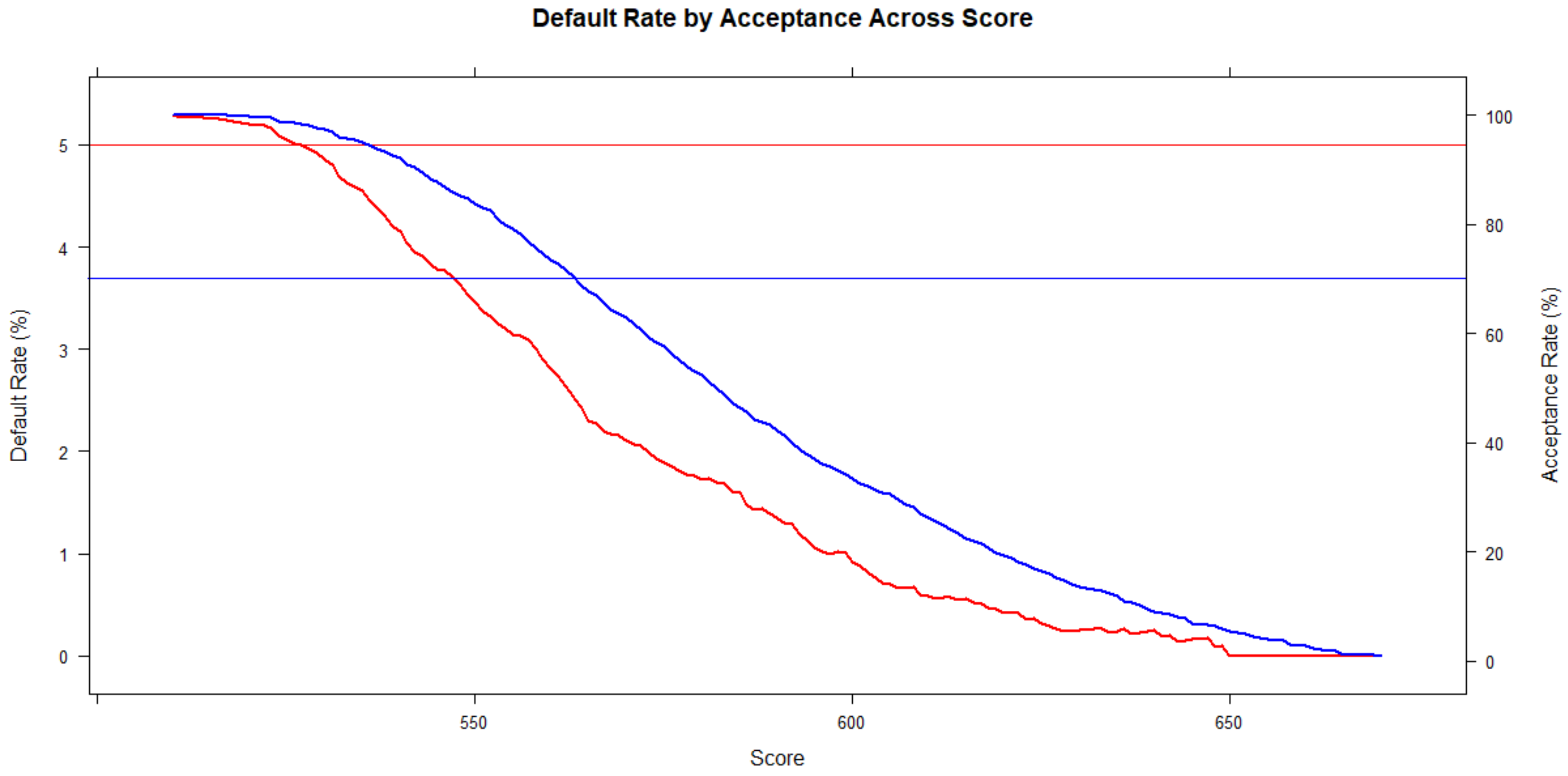
Default Decile Plot



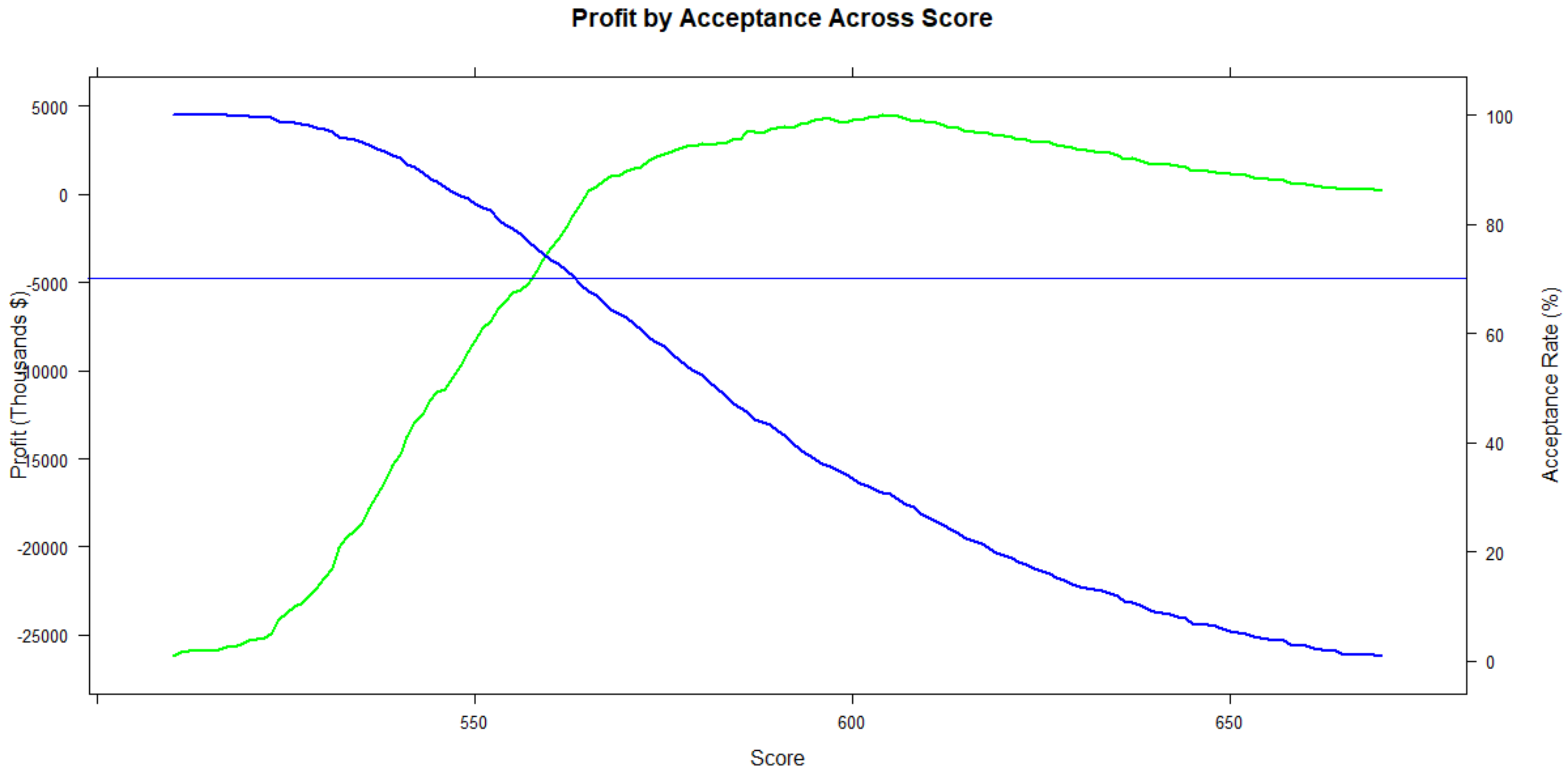
Defining Cut-off

- Trade-off Plots:
  - The reference lines of approval rate and event (bad) rate are predefined by analyst.
  - How much risk are you willing to take on?

Defining Cut-off



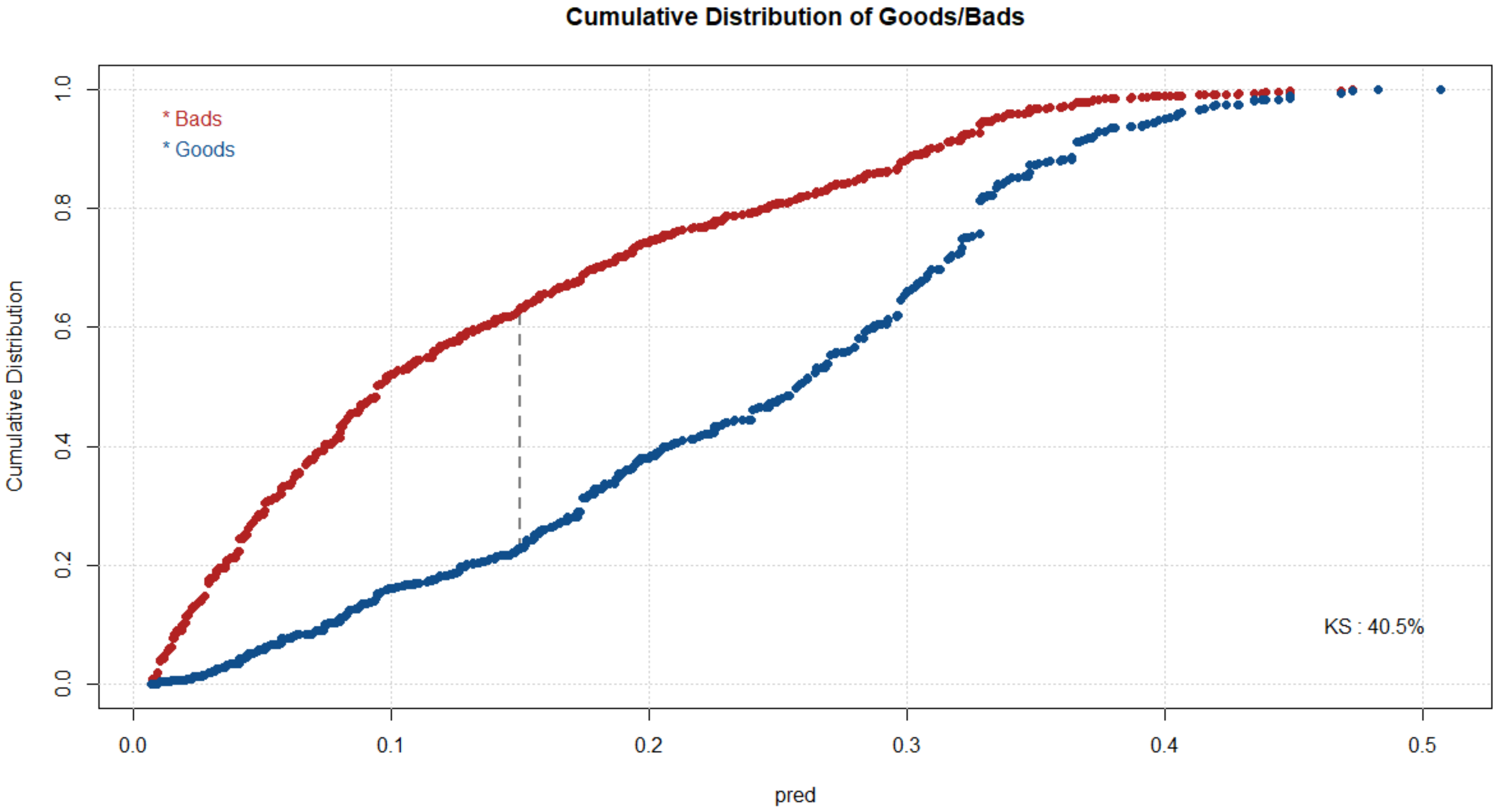
Defining Cut-off



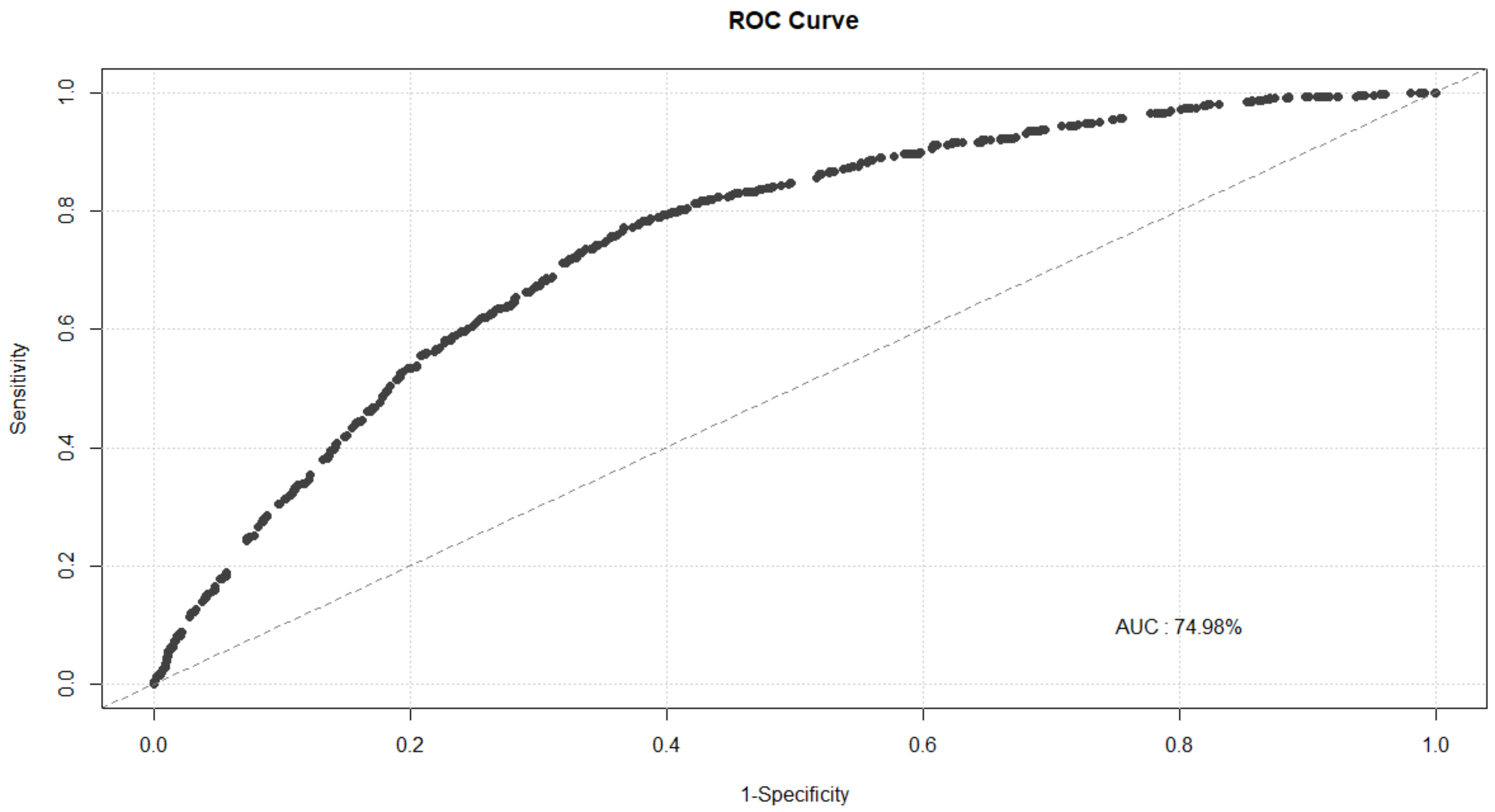
Defining Cut-off

- Setting Multiple Cut-offs Example:
  - Anyone who scores above 210 points is accepted automatically.
  - Anyone who scores below 190 is declined.
  - Any scores in between 190 and 210 are referred to manual adjudication.

Final Scorecard – Example



Final Scorecard – Example



CREDIT SCORING MODEL EXTENSIONS

Lack of Interactions

- Benefits of tree based algorithms are inherent interactions of every split of the tree.
    - Also a detriment to interpretation.

Multi-stage Model

- Benefits of tree based algorithms are inherent interactions of every split of the tree.
  - Also a detriment to interpretation.

- Multi-stage model:
  1. Decision Tree to initially get a couple of layers of splits.
  2. Build logistic regression based scorecard in each of the splits.
  3. Interpretation is now **within** a split (sub-group) of the data.

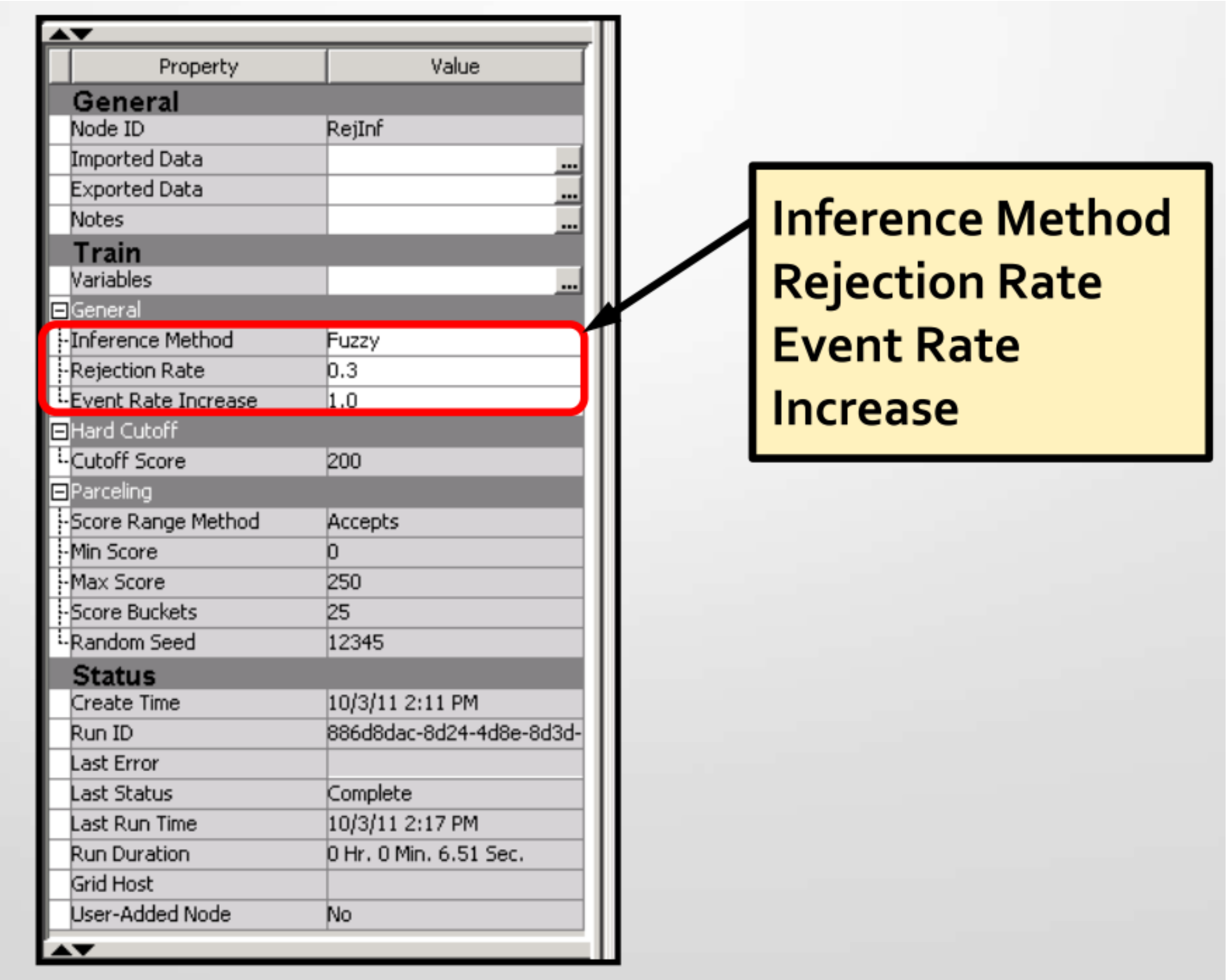
Machine Learning

- Model interpretation is **KEY** in the world of credit scoring.
- Scorecard layer may help drive interpretation of machine learning algorithms, but regulators are still hesitant.

- Great for internal comparison and variable selection.
  - Build a neural network, tree based algorithm, etc. to see if model is statistically different than logistic regression scorecard.
  - Empirical examples have shown WOE based logistic regressions perform very well in comparison to more complicated approaches.

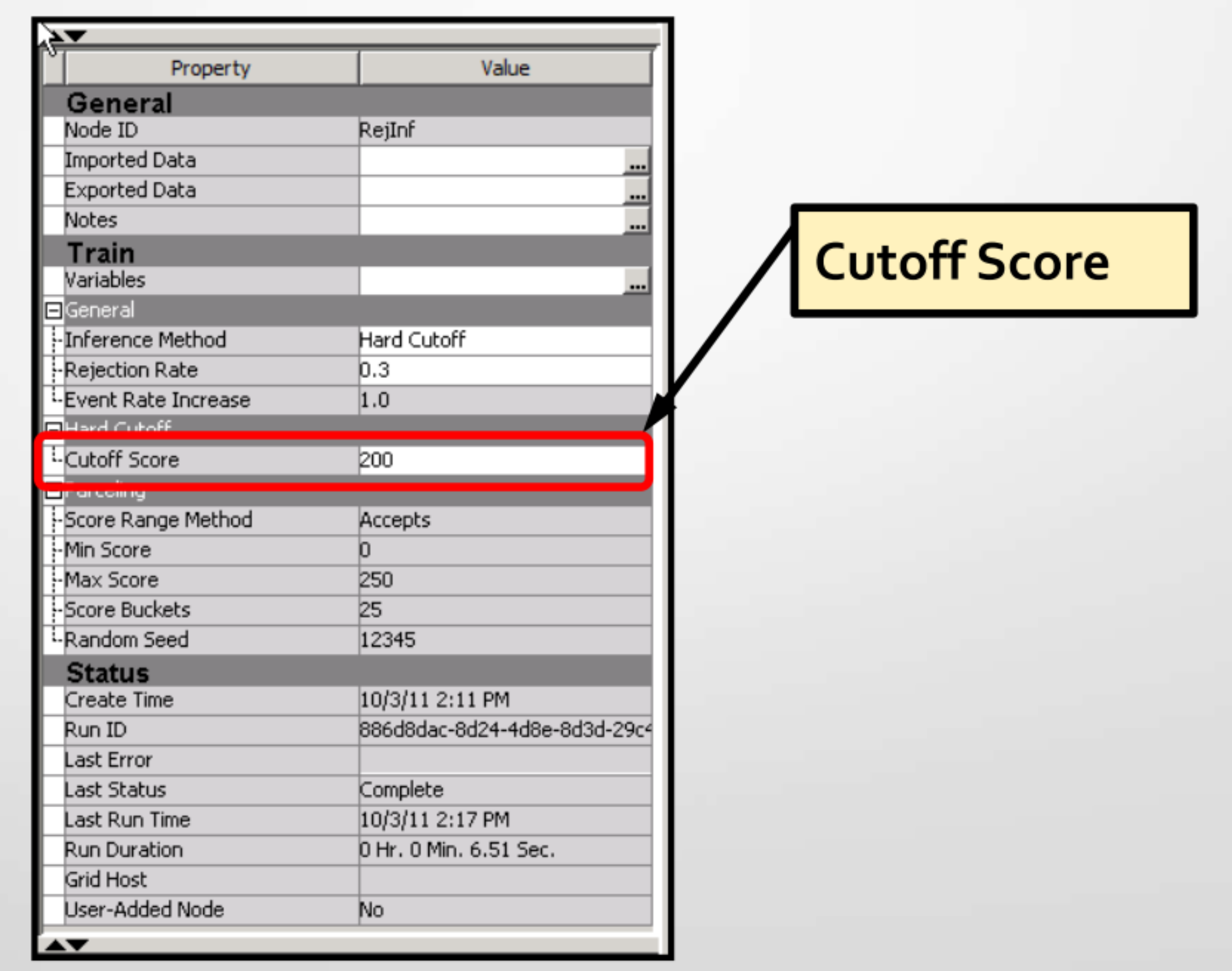
REJECT INFERENCE NODE IN SAS EM

General Options



Property	Value
General	
Node ID	RejInf
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
General	
Inference Method	Fuzzy
Rejection Rate	0.3
Event Rate Increase	1.0
Hard Cutoff	
Cutoff Score	200
Parceling	
Score Range Method	Accepts
Min Score	0
Max Score	250
Score Buckets	25
Random Seed	12345
Status	
Create Time	10/3/11 2:11 PM
Run ID	886d8dac-8d24-4d8e-8d3d-
Last Error	
Last Status	Complete
Last Run Time	10/3/11 2:17 PM
Run Duration	0 Hr. 0 Min. 6.51 Sec.
Grid Host	
User-Added Node	No

Hard Cut-off Options



Parceling Options