

# ABA Regulatory Compliance Conference

## Statistical Analysis for Community Banks

Rick Preiss  
Phil Preiss

# Presentation Overview

- Pre-Analysis Risk
  - Data
  - Proxies
- Approaches to Statistical Analysis
  - Levels
    - Data Visualization
    - Raw Data (Non-Regression) Statistics
    - Regression Statistics
- Not So “Statistical” Analyses
  - Redlining
  - REMA

# I. Pre-Analysis Risk – Data

- Data Accuracy and Integrity
  - 20-40% of Project Time
  - Accuracy
    - Extreme Values
    - Pre-Determined Out of Range Values
    - Missing Values
    - Duplicates
    - Specific Values
  - Integrity
    - Consistency Across Fields
    - Invalid/Misspelled values

# I. Pre-Analysis Risk – Proxies

- Analysis May Require Proxies
  - Many Proxy Methodologies
    - Surname – Using Applicant/Co-Applicant Last Name to Determine Race/Ethnicity
    - Combined Surname and Geography (BISG)
    - First Names (Gender)
    - Geography – Using Census Tract Demographics to Determine Maj-Min or Low/Mod Designation
    - Race Composition ( $\geq 80\%$ , etc.)
  - Used Where Race/Gender/Ethnicity not Provided
    - Indirect/Direct Auto, Boat, Overdraft Fees, etc.

# II. Approaches to Statistical Analysis – Summary

- **Goal: Identify a Process That Community Banks of Varying Sizes Can Use to Test for Differential Treatment.**
- **Three Levels of Statistical Analysis**
  - Level 1: Data Visualization/Re-Organizing
  - Level 2: Raw Data Statistical Testing, i.e. Non-Regression Analysis
  - Level 3: Regression Analysis
- **Presenters' Notes**
  - Based on Preiss&Associates Experience, Not Regulator Dictated
  - Each Level Builds on the Last
  - Analysis Level Volume/Asset Criteria Are Estimates, Not Hard Rules

## II. Approaches to Statistical Analysis – Level 2

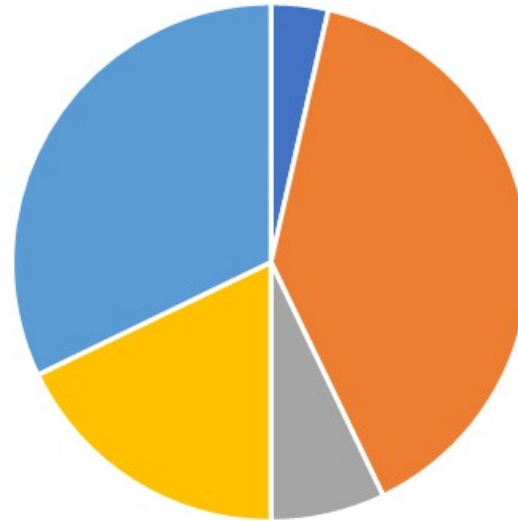
- Analysis Datasets Typically Contain More Records Than Level 1 Datasets
- Builds on Level 1 Approach by Using Summary Statistics
- Summary Statistics Used in two areas:
  - Data Summary
    - Calculating Frequencies, Means, Standard Deviations, Minimums, and Maximums for Fields of Interest.
  - Statistical T-Tests
    - Determine if Differential Treatment is Occurring.
    - Calculates Magnitude of Difference, Does Not Suggest Why there is a Difference
    - “5/25 Rule”

# II. Approaches to Statistical Analysis – Level 1

- Analysis Dataset Typically Contains a Small Number of Records, Less Than 25 Loans
- Data Visualization/Re-Organization Tools
  - Tables/Charts
  - Mapping
- Data Segmentations
  - Prohibited Basis Groups (i.e. Gender, Age, Race, and Ethnicity)
  - Demographics (Majority-Minority and Low/Mod Income Census Tracts)
  - Applications vs Originations
  - Geographies (MSA, Counties)
- Data Review (i.e. Count of Abnormal, Missing, or Zero Values)
- Matching Reports

# II. Approaches to Statistical Analysis – Level 1

Row Labels	Volume
Approved, not accepted	1
Denied	11
Withdrawn	2
Incomplete	5
Originated	9
Total	28



- Approved, not accepted
- Denied
- Withdrawn
- Incomplete
- Originated



## II. Approaches to Statistical Analysis – Level 2

### *Pricing Analysis – Rate Spread Frequency*

Race Category	Origination	Rate Spread		Results	
	Volume	Frequency	Percentage	T-Value	Statistically Significant
American Indian/Alaska Native	169	89	52.63%	-0.60	No
Asian	1,041	530	50.91%	-0.80	No
African American	876	319	36.36%	-2.77	No
Hawaiian/Pacific Islander	80	35	43.48%	-2.64	No
Joint-Race	258	145	56.38%	0.13	No
Multi-Race	33	14	42.76%	-2.40	No
White	31,458	17,308	55.02%	NA	NA

# II. Approaches to Statistical Analysis – Level 2

## *Underwriting Analysis*

Race Category	Applications			Results		
	Originated	Denied	Total	Denial Rate	T-Value*	Statistical Significance
American Indian/Alaska Native	101	37	138	26.81%	0.56	No
Asian	600	216	816	26.47%	1.14	No
African American	412	182	594	30.64%	3.11	Yes
Hawaiian/Pacific Islander	45	19	64	29.69%	0.88	No
Joint-Race	215	87	302	28.81%	1.57	No
Multi-Race	16	8	24	33.33%	0.90	No
White	15,861	5,198	21,059	24.68%		

\*T-Values greater than 1.96 or less than -1.96 are considered statistically significant

## II. Approaches to Statistical Analysis – Level 2

### *Pricing Analysis – Average Rate*

Race Category	Origination Volume	Average Interest Rate	Results	
			T-Value*	Statistical Significance
American Indian/Alaska Native	169	5.46%	3.19	Yes
Asian	1,041	5.12%	-1.50	No
African American	876	5.23%	1.34	No
Hawaiian/Pacific Islander	80	5.36%	1.69	No
Joint-Race	258	5.35%	2.24	Yes
Multi-Race	33	5.71%	2.63	Yes
White	31,458	5.17%	NA	NA

\*T-Values greater than 1.96 or less than -1.96 are considered statistically significant

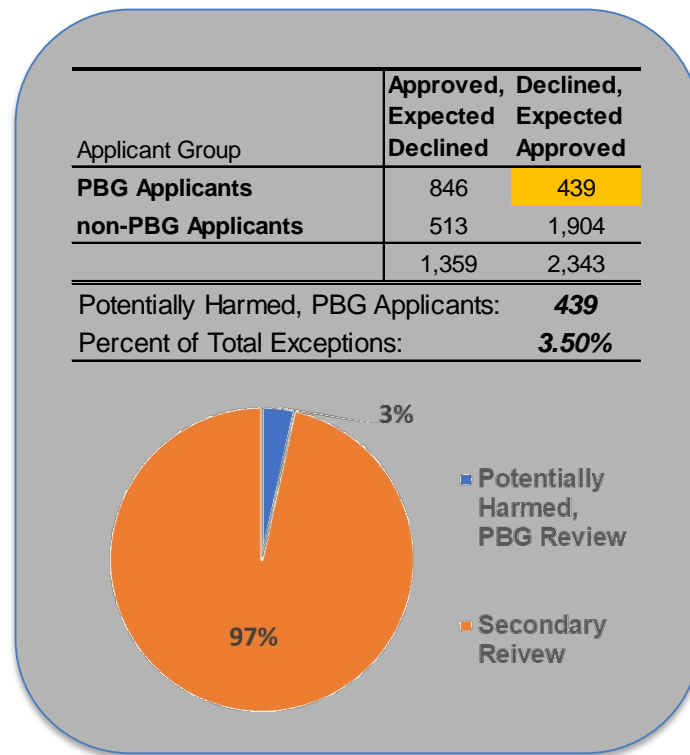
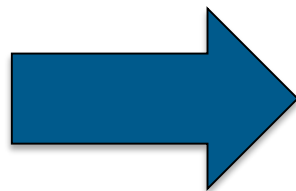
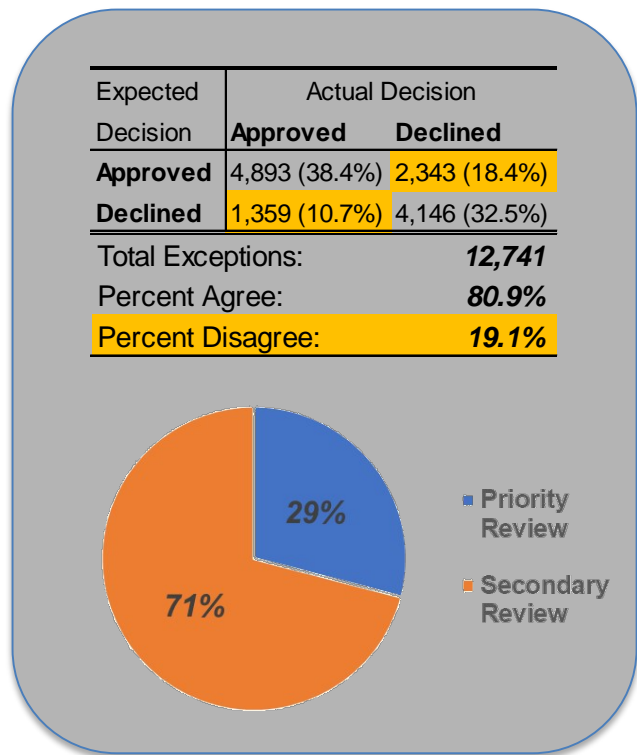
## II. Approaches to Statistical Analysis – Level 3

- Advances Analysis Results Produced in Level 2 by Estimating a Regression Equation.
- Regression Analysis Attempts to Explain Any Differential Treatment Found in Level 2.
  - “30/30/100 Rule”
  - Criteria in Your Institution’s Underwriting and Pricing guidelines Are the Variables Used in a Regression Analysis
  - Non-guideline Criteria, i.e. Pricing Promotion, May be Used As Well
- Popular For Underwriting and Pricing Monitoring.

# II. Approaches to Statistical Analysis – Level 3

- What Are the Elements of a Regression Model
  - Dependent Variables (Accept/Deny, Interest Rates)
  - Independent Variables (DTI, LTV, Credit Score, Loan Products, etc.)
- The Regression Process
  - Review Underwriting or Pricing Guidelines to Determine Regression Variables to be Used
  - Customize Regression Equation for Your Institution
  - Run Regression and Review Results
    - Ensure Underwriting/Pricing Risk Factors are Reasonable
  - Edit and Re-Estimate if necessary

# II. Approaches to Statistical Analysis – Level 3



# II. Approaches to Statistical Analysis – Level 3

## Exception Report

**Appendix**  
**Bank of Preiss&Associates**  
**HMDA Mortgage Credit Decision Exception Report**

LOAN NUMBER	OPEN DATE	LOAN TERM	LOAN AMT	CREDIT SCORE	DTI	LTV	LOAN TYPE	LOAN PURP	OCCUP TYPE	GENEDR DESC	AGE DESC	RACE/ETHN DESC	RESIDUAL
<b>Approved, Expected Declined</b>													
xxxx	1/12/2019	360	238,140	665	62	81	Conv.	Refinance	Owner Occ.	Female	LT62	White	6.14
xxxx	8/30/2019	360	415,000	658	59	104	Conv.	Refinance	Owner Occ.	Joint	LT62	White	6.14
xxxx	4/1/2019	360	217,950	755	56	109	Conv.	Refinance	Owner Occ.	Joint	GTE62	White	5.29
<b>Declined, Expected Approved</b>													
xxxx	6/20/2019	360	225,000	741	41	63	Conv.	Refinance	Non-Owner Occ.	Joint	LT62	White	-4.80
xxxx	6/1/2019	360	299,000	785	43	60	Conv.	Purchase	Owner Occ.	Joint	GTE62	White	-3.71
xxxx	12/19/2019	360	144,000	694	30	80	Conv.	Purchase	Owner Occ.	Male	LT62	White	-3.71
xxxx	3/31/2019	360	70,000	692	38	70	Conv.	Purchase	Owner Occ.	Unknown	LT62	Unknown	-3.71

**Proprietary & Confidential**  
**Preiss & Associates, LLC. 1551 Tara Lane Lake Forest IL 60045**

# II. Approaches to Statistical Analysis – Level 3

## Matching Report

Appendix Bank of Preiss&Associates HMDA Mortgage Credit Decision Matching Report												
LOAN NUMBER	OPEN DATE	LOAN TERM	LOAN AMT	CREDIT SCORE	DTI	LTV	LOAN TYPE	LOAN PURP	OCCUP TYPE	GENEDR DESC	AGE DESC	RACE/ETHN DESC
<b>Exception Applicant</b>												
xxxx	2/23/2019	360	142,000	688	31	57	Conv.	Refinance	Owner Occ.	Female	LT62	White
<b>Similarly-Situated Applicants</b>												
xxxx	2/11/2019	360	156,000	637	27	43	Conv.	Refinance	Owner Occ.	Joint	LT62	White
xxxx	2/24/2019	360	186,400	701	33	80	Conv.	Refinance	Owner Occ.	Joint	LT62	White
<b>Exception Applicant</b>												
xxxx	8/27/2019	360	283,000	776	45	79	Conv.	Refinance	Owner Occ.	Female	LT62	White
<b>Similarly-Situated Applicants</b>												
xxxx	8/17/2019	360	268,000	688	35	80	Conv.	Refinance	Owner Occ.	Joint	GTE62	White
xxxx	8/23/2019	360	266,400	758	32	80	Conv.	Refinance	Owner Occ.	Joint	LT62	White
xxxx	8/26/2019	360	360,000	722	37	80	Conv.	Refinance	Owner Occ.	Male	LT62	White
<p align="center"><b>Proprietary &amp; Confidential</b> Preiss &amp; Associates, LLC. 1551 Tara Lane Lake Forest IL 60045</p>												



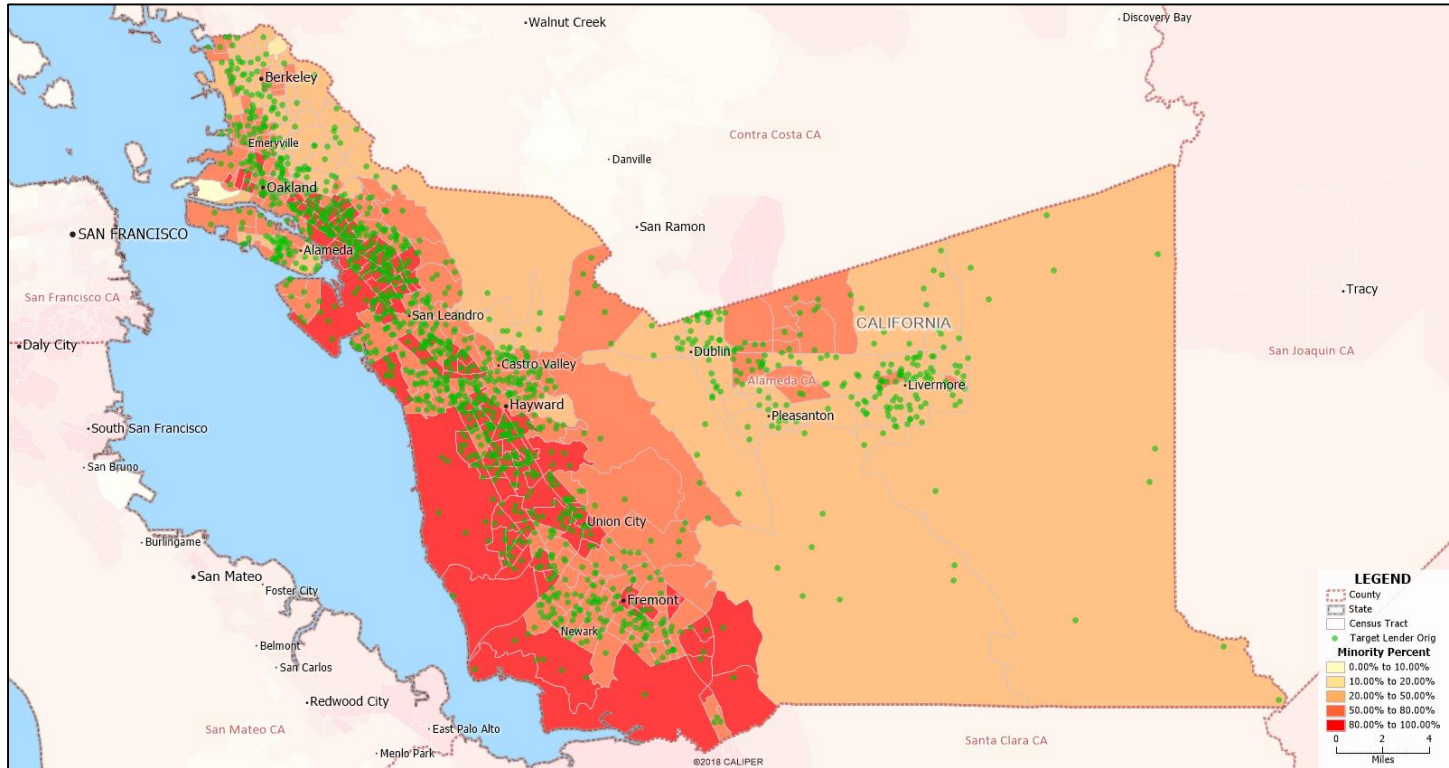
# III. Not So “Statistical” Analyses – Redlining

- What is Redlining?
  - Redlining is form of illegal disparate treatment in which a lender provides unequal access to credit, or unequal terms of credit because of race, color, national origin or other prohibited basis characteristic(s) of the area in which the credit seeker resides or will reside or in which the residential property to be mortgaged is located.
- In addition to Mortgages, Redlining May Occur in With Respect to Different Credit Products and Pricing Based on Geography
  - Brokers, Auto Dealers, Credit Cards
  - New: Digital Redlining (Differential Access to Online and Mobile Services, i.e., social media, other online activities, digital deserts)
- In Most Cases Redlining Analyses Combine Statistical and Mapping Analysis

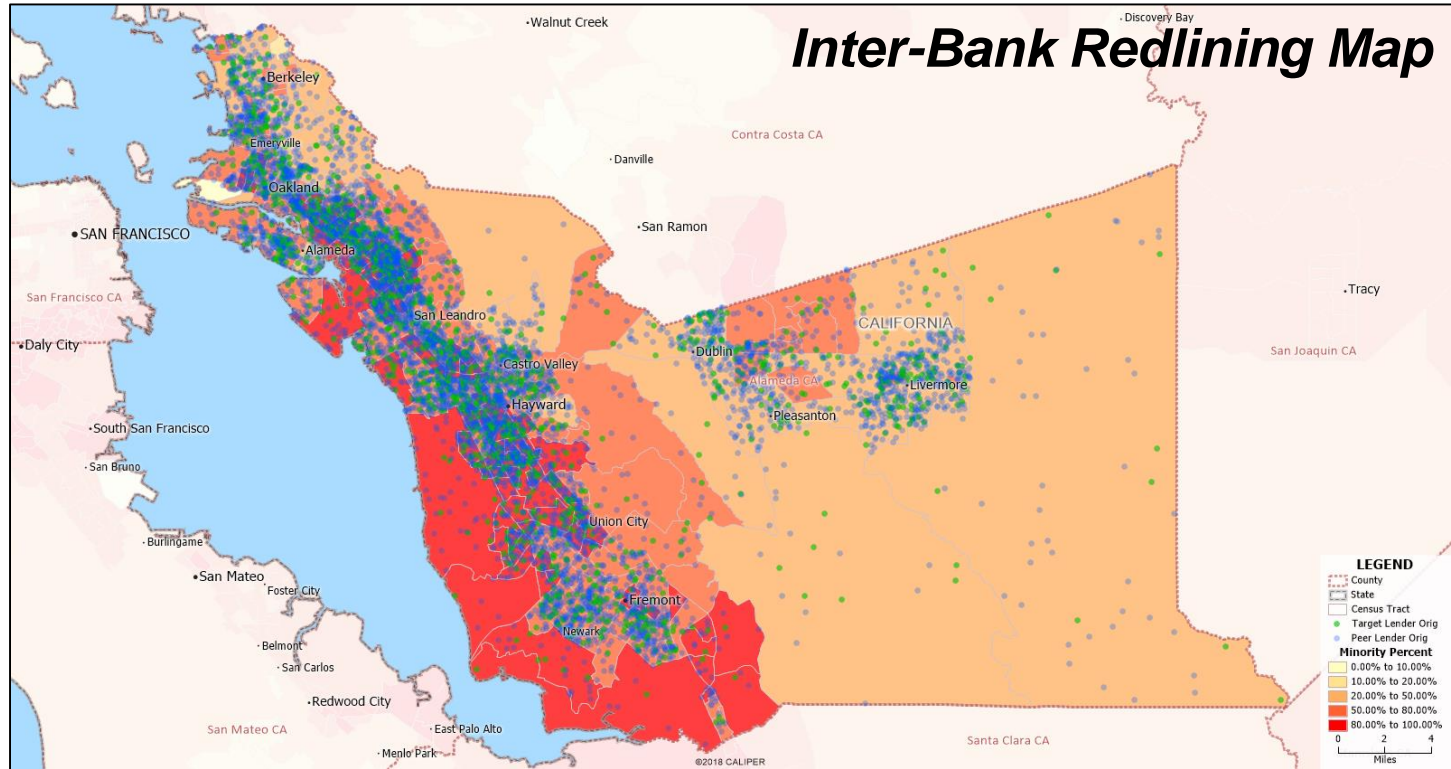
# III. Not So “Statistical” Analyses – Redlining

- Types of Redlining
  - Intra-Bank: Compares Institution Data to Target Geography Demographics (Population, Housing Stock)
  - Inter-Bank: Compares Institution Data to Peer Institution Data
  - REMA: Compares Institution Data to Geographical Area a Regulator Believes a Bank Can Serve
- Types of Analysis Comparisons
  - Applications, Originations
  - Geography Level (i.e. MSA or County)
  - Underlying Comparison
    - Majority-Minority, Hispanic, or African American Census Tract Categories
    - Tract/MSA MFI Percent Categories
- REMA Comments
  - Develop **Your** Suggested REMA Based Institution’s Application/Loan Distribution and Marketing/Outreach Efforts
  - Avoid Letting Your Regulator Tell You Your REMA if Possible

# III. Not So “Statistical” Analyses – Redlining



# III. Not So “Statistical” Analyses – Redlining



# III. Not So “Statistical” Analyses – Redlining

Minority Category	Originations		Population		GTE-50%	Binomial		
	Count	Percent	Count	Percent		Count	Percent	
0-10%	0	0.0%	4,418	0.1%		Originations	349	57.1%
10-20%	44	7.2%	188,486	6.0%		Population	1,700,262	54.6%
20-50%	218	35.7%	1,222,903	39.2%			<b>T-Value</b>	<b>Result</b>
50-80%	191	31.3%	974,938	31.3%		T-Test	<b>0.96</b>	<b>No</b>
80-100%	158	25.9%	725,324	23.3%				
Total	611		3,116,069					

## ***INTRA-BANK ANALYSIS***

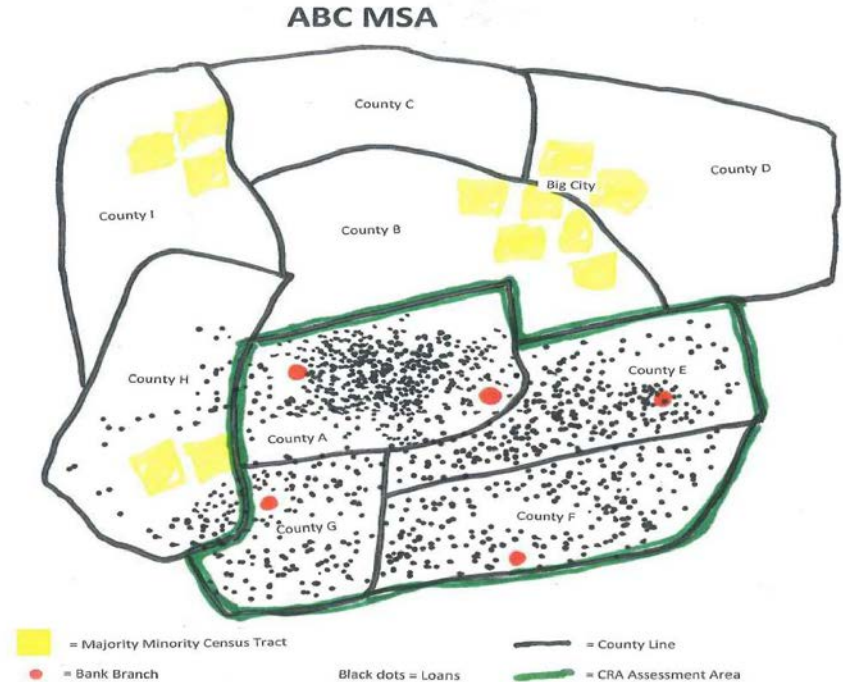
## ***INTER-BANK ANALYSIS***

Minority Category	Preiss&Assoc.		Peers		GTE-50%	Binomial		
	Count	Percent	Count	Percent		Count	Percent	
0-10%	0	0.0%	0	0.0%		Preiss&Assoc.	349	57.1%
10-20%	44	7.2%	116	5.0%		Peers	1,282	54.6%
20-50%	218	35.7%	935	40.1%			<b>T-Value</b>	<b>Result</b>
50-80%	191	31.3%	925	39.6%		T-Test	<b>1.13</b>	<b>No</b>
80-100%	158	25.9%	357	15.3%				
Total	611		2,333					

# III. Not So “Statistical” Analyses – Redlining

## ■ REMA Example

- CRA Assessment Area: Counties A, E, F, and G
- REMA: Counties A, E, F, G, and H
- Basis for REMA: Significant Number of Loans Originated in County H



# Session Takeaways – Almost Done

- Fair Lending Statistics Can Benefit the Bank
- Requires Flexibility and Judgement
- Must Understand:
  - The Analysis
  - The Results
  - What to do With the Results
- There Are Some Tasks All Lenders Can Perform
  - Data Accuracy and Integrity
  - Data Visualization
  - Mapping
- *Perform the Analysis That is Right for Your Institution*
  - No One Knows Your Institution Better Than You
  - If You Don't Perform Your Institution's Analysis Your Regulator May Do It for You
    - Which Analysis Would You Rather Live With?