



UNIVERSITY OF
MARYLAND

ROBERT H. SMITH
SCHOOL OF BUSINESS

Health Insurance Coverage: Comparing the Insured and Uninsured in the United States

By

Jason Davis

Clarette Kim

Hari Kosaraju

Scott Wood

BUDT 733, Shmueli

DC, Fall 2007

Agenda

- I. Key Questions**
- II. Data Sources**
- III. Data Exploration**
- IV. Modeling**
- V. Recommendations**

Key Questions

1. What demographic or socioeconomic groups in the United States, if any, are more likely *not* to have healthcare coverage?
2. What factors differentiate the uninsured from the insured?
3. What recommendations can be developed to help policy makers and philanthropic organizations more effectively target the uninsured?

Data Source

- Demographic data obtained from US Census Bureau's 2007 Current Population Survey (CPS)
- CPS published annually from a nationally representative survey of U.S. Households
- Sample contained random cross-section of 30,000 individuals
- Variables selected using domain knowledge included:

Health Insurance Coverage

Household Income, \$

Work Experience

Education Level

Marital Status

Gender

Age, in years

Geographic Region Indicator

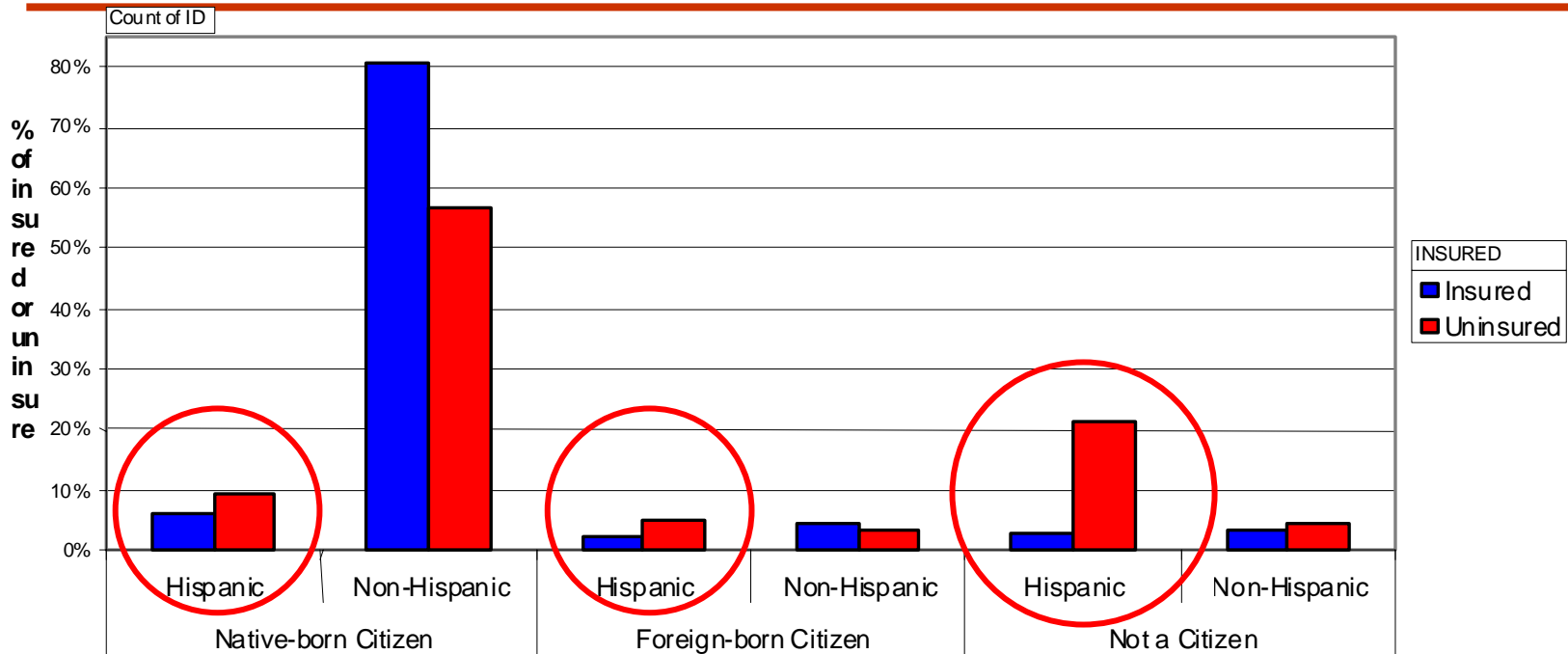
Metropolitan Region Indicator

Health Condition

US Citizenship

Race and Ethnicity Indicators

Data Exploration



CITIZEN | HISPANIC

Categorical variable pair		p-value	Chi2-stat	Df
Hispanic	Citizen	0.00E+000	6971.96	2
Educ	Citizen	0.00E+000	5038.83	30
Hispanic	Educ	0.00E+000	4620.36	15
Marital	Race	0.00E+000	1867.56	12
Educ	Marital	2.02E-295	1700.18	90
Educ	Health	5.17E-247	1370.49	60
Hispanic	Region	2.15E-282	1303.85	3
Educ	Race	3.80E-141	762.78	30

- **Strong interdependency between Hispanic, Citizenship and Education**

- **Strong interdependency between Marital Status, Education and Race**

Data Modeling

Goal was explanatory, so our models included:

1. Classification Tree

Result: Household income less than \$54,652 and Hispanic and Non-Citizens

2. Logistic Regression

Result: Hispanic, non-citizen, living in southern U.S. and lower household income

3. Discriminant Analysis

Result: Hispanic, non-citizen, living in southern U.S. and not college educated

Data Modeling

Training Data scoring - Summary Report (Using Full Tree)

Cut off Prob.Val. for Success (Updatable) **0.5**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	2
1	8253	204
2	1194	348

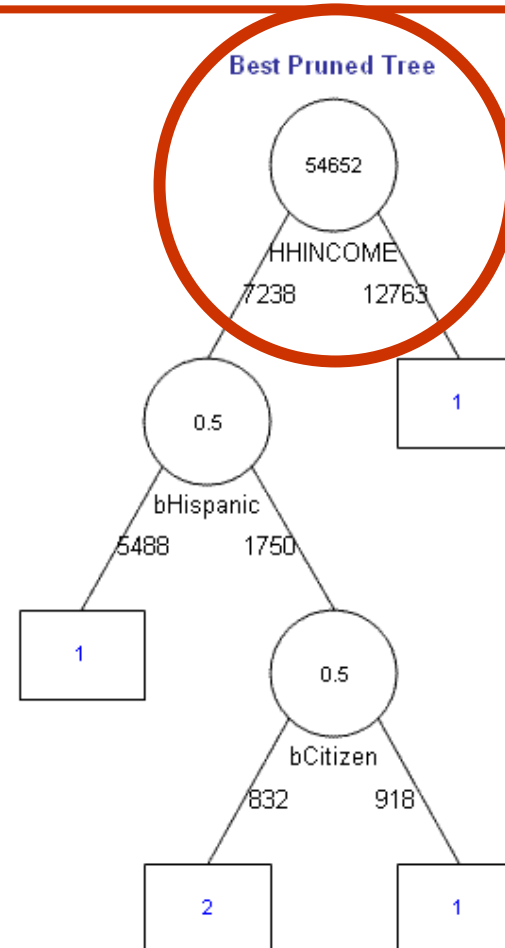
Error Report			
Class	# Cases	# Errors	% Error
1	8457	204	2.41
2	1542	1194	77.43
Overall	9999	1398	13.98

Validation Data scoring - Summary Report (Using Best Pruned Tree)

Cut off Prob.Val. for Success (Updatable) **0.5**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	2
1	16603	313
2	2566	519

Error Report			
Class	# Cases	# Errors	% Error
1	16916	313	1.85
2	3085	2566	83.18
Overall	20001	2879	14.39



Evaluation of Models

Model	Significant Predictors	Overall Error	Error in Classifying Uninsured	Other Performance Measures
Classification Tree	Household Income, Hispanic status, Citizenship	14.39%	83.18%	
Logistic Regression	Household Income, Hispanic status, Citizenship, Income, Age, Education, Metro, Marital Status, Black, Asian	<u>12.02%</u>	48.27%	Multiple R-Square: 37.69%
Discriminant Analysis	Household Income, Hispanic status, Citizenship, Income, Age, Education, Marital Status, Black, Asian, South, and West	18.09%	<u>17.14%</u>	

Recommendations

1. Target *Hispanic* populations in *Southern* and *Western* states
2. Target the *Working, Rural Poor*
3. Target *Young, Unmarried* individuals and those with *High School Education (or less)*

Questions?

