

# Summary Report: Credit Data

---

CAPP 30254

April 20, 2015



THE UNIVERSITY OF  
CHICAGO



**PRIMARY INVESTIGATOR:**

**JOSHUA MAUSOLF**



**1-i. Summary Data for Serious Delinquencies in the Last Two Years:**

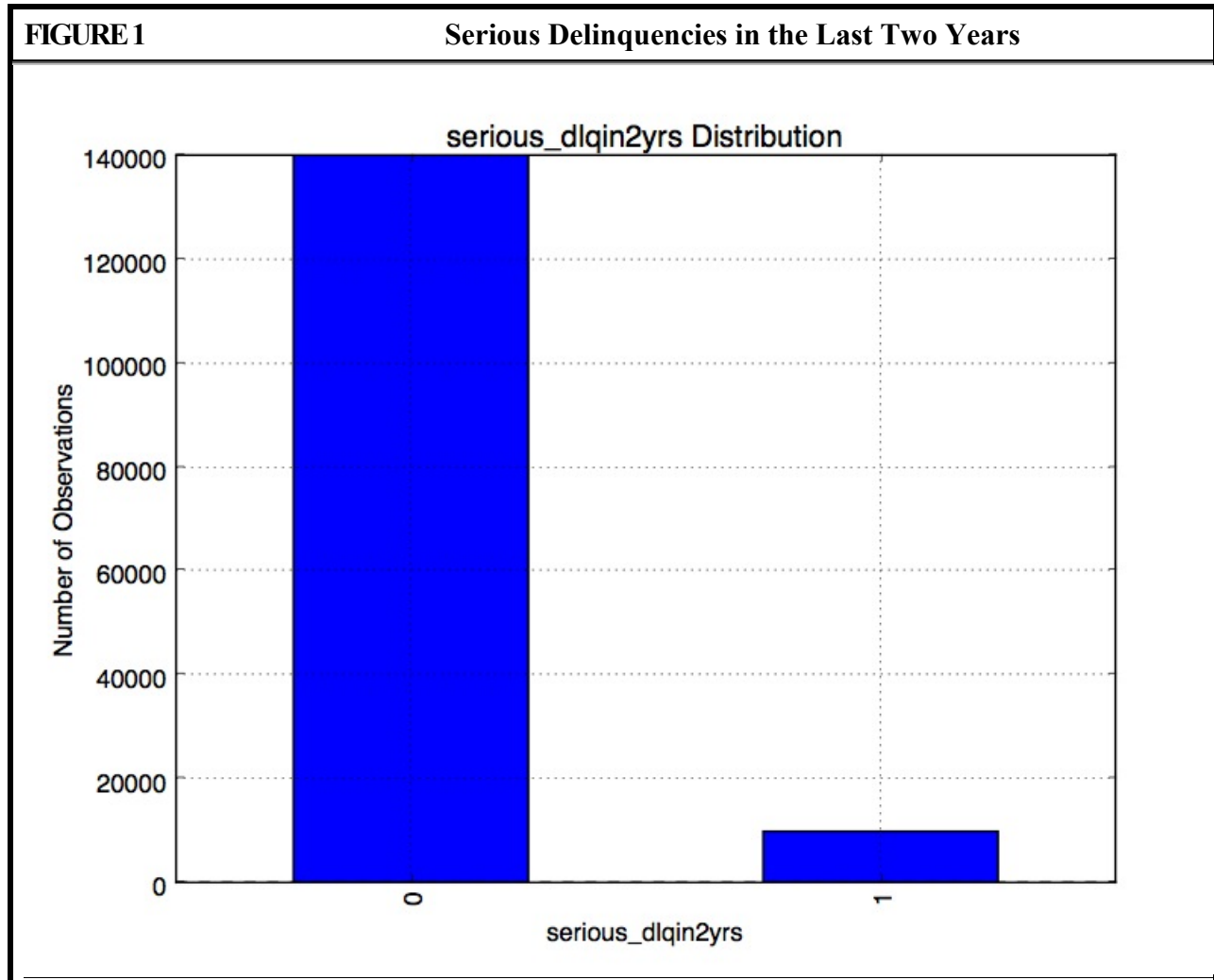


TABLE 1		Serious Delinquencies in the Last Two Years	
	N =	150,000	
	Missing =	0	
Percent Delinquent =		6.68%	
	Mean =	0.067	
	Median =	0.00	
	Min, Max =	(0, 1)	
Source: cs-training.csv			

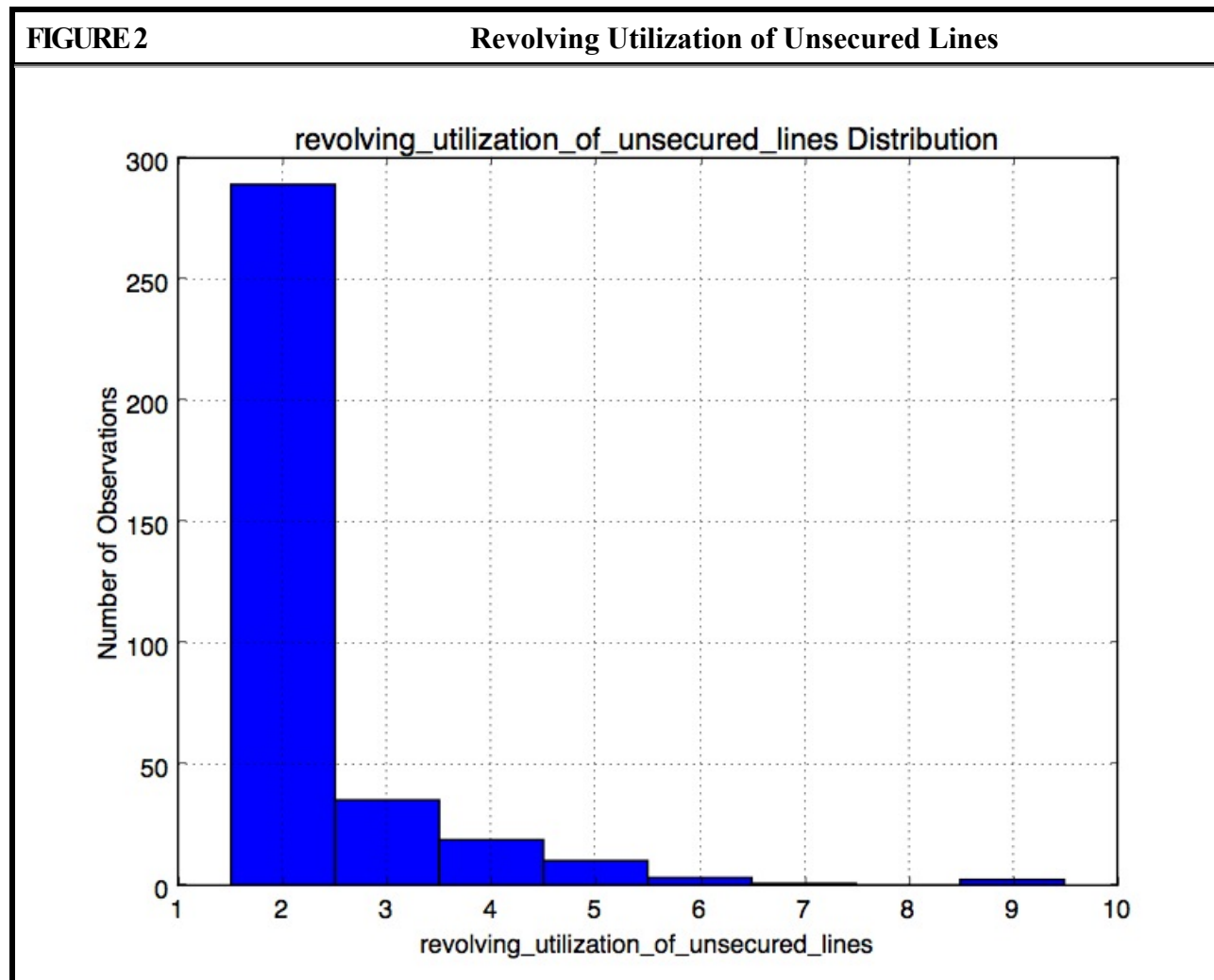
**1-ii. Summary Data for Revolving Utilization of Unsecured Lines:**

TABLE 2		Revolving Utilization of Unsecured Lines
	N =	150,000
	Missing =	0
	Mean =	6.05
	Median =	0.15
	Standard Deviation =	249.76
	Min, Max =	(0, 50,708)
Source: cs-training.csv		

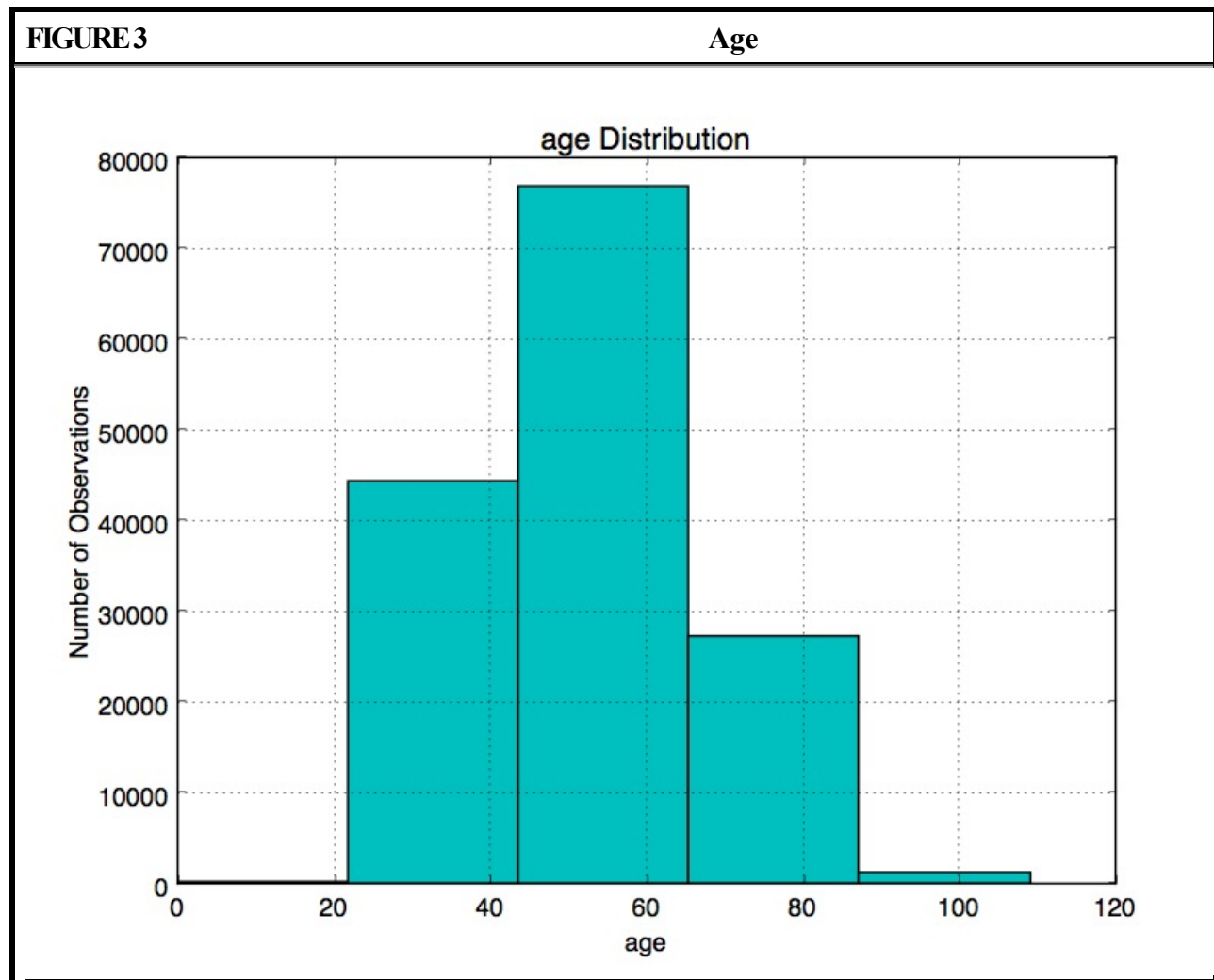
**1-iii. Summary Data for Age:**

TABLE3		Age
N =		150,000
Missing =		0
Mean =		52
Median =		52
Mode =		49
Standard Deviation =		15
Min, Max =		(0, 109)
Source: cs-training.csv		

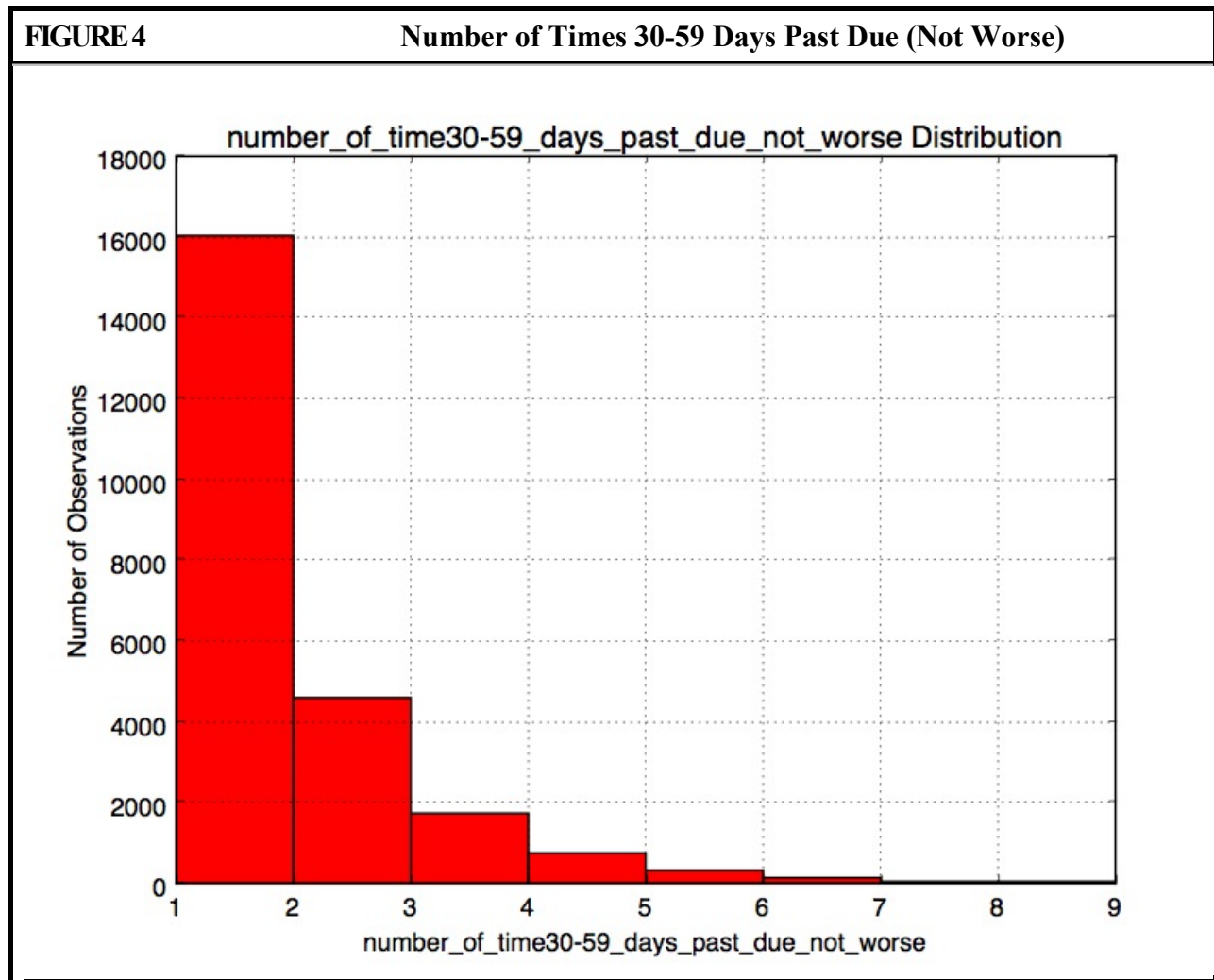
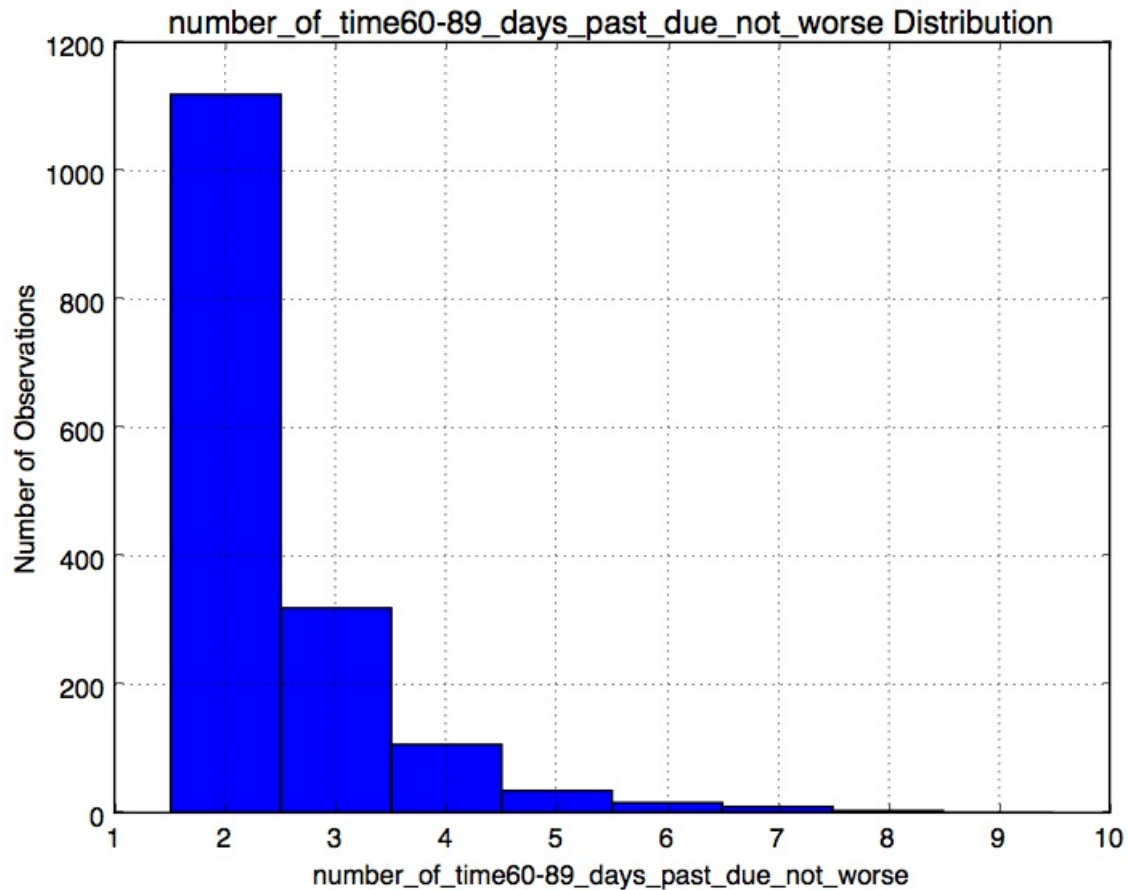
**1-iv. Summary Data for Number of Times 30-59 Days Past Due (Not Worse):**

TABLE4                      Number of Times 30-59 Days Past Due (Not Worse)	
N = 150,000	
Missing = 0	
Mean = 0.42	
Median = 0.00	
Standard Deviation = 4.19	
Min, Max = (0, 98)	
Source: cs-training.csv	

**1-v. Summary Data for Number of Times 60-89 Days Past Due (Not Worse):****FIGURE 5****Number of Times 60-89 Days Past Due (Not Worse)****TABLE 5****Number of Times 60-89 Days Past Due (Not Worse)**

N =	150,000
Missing =	0
Mean =	0.24
Median =	0.00
Standard Deviation =	4.16
Min, Max =	(0, 98)
Source: cs-training.csv	

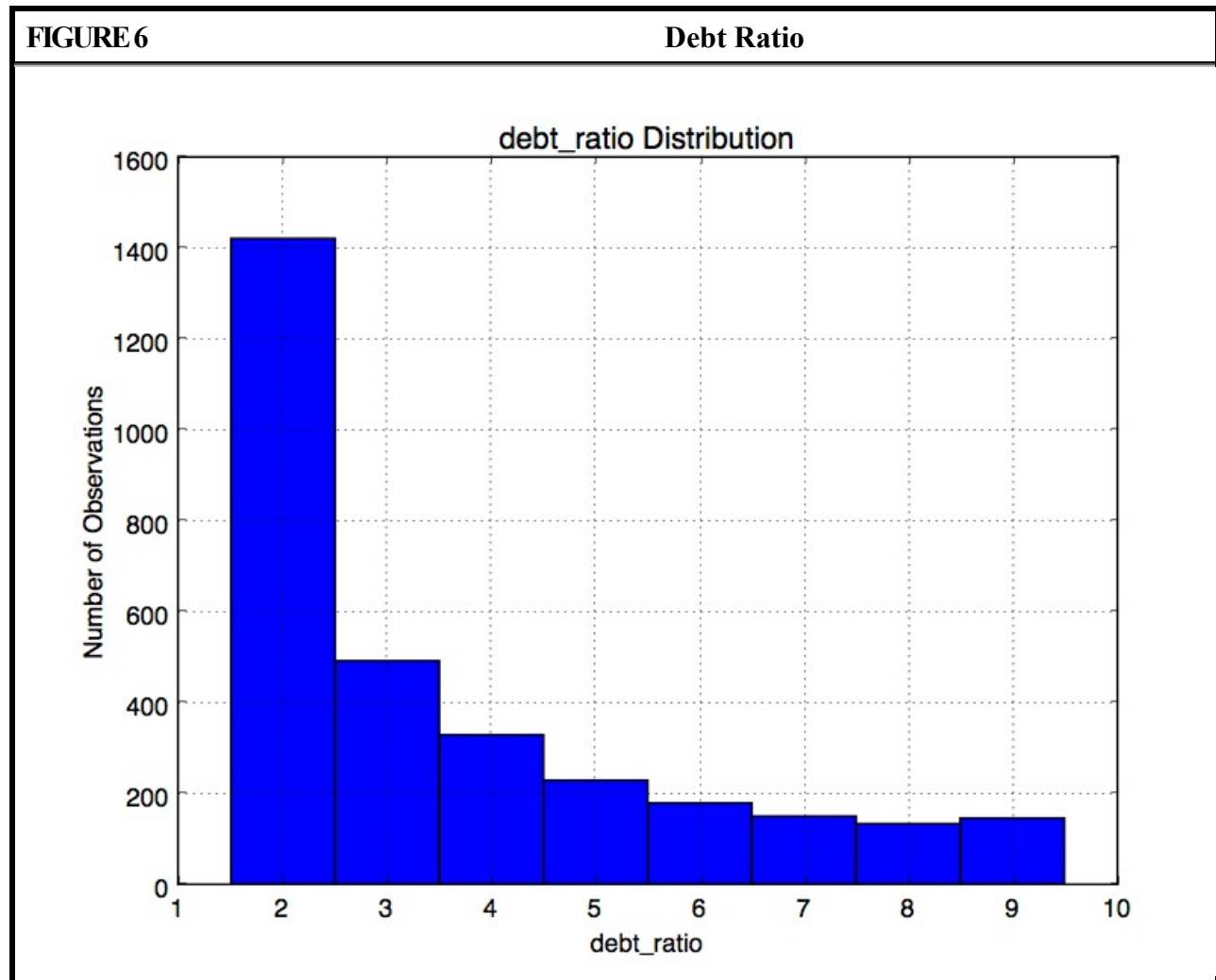
**1-vi. Summary Data for Debt Ratio:**

TABLE 6		Debt Ratio	
N =		150,000	
Missing =		0	
Mean =		353.01	
Median =		0.37	
Standard Deviation =		2037.82	
Min, Max =		(0, 329,664)	
Source: cs-training.csv			

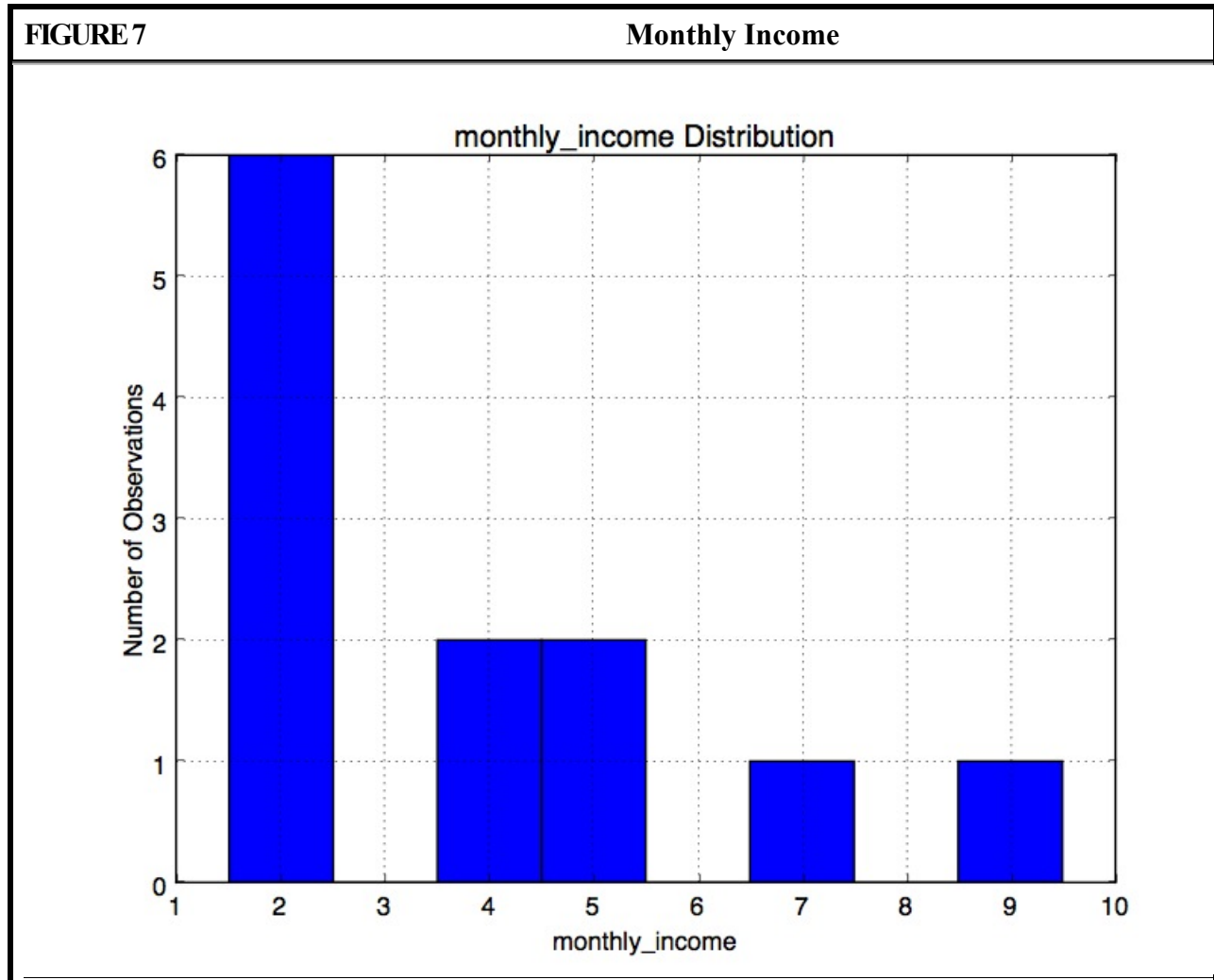
**1-vii. Summary Data for Monthly Income:**

TABLE 7		Monthly Income
N =	120,269	
Missing =	29,271	
Mean =	6,670.22	
Median =	5,400.00	
Standard Deviation =	14,384.67	
Min, Max =	(0, 3,008,750)	
Source: cs-training.csv		



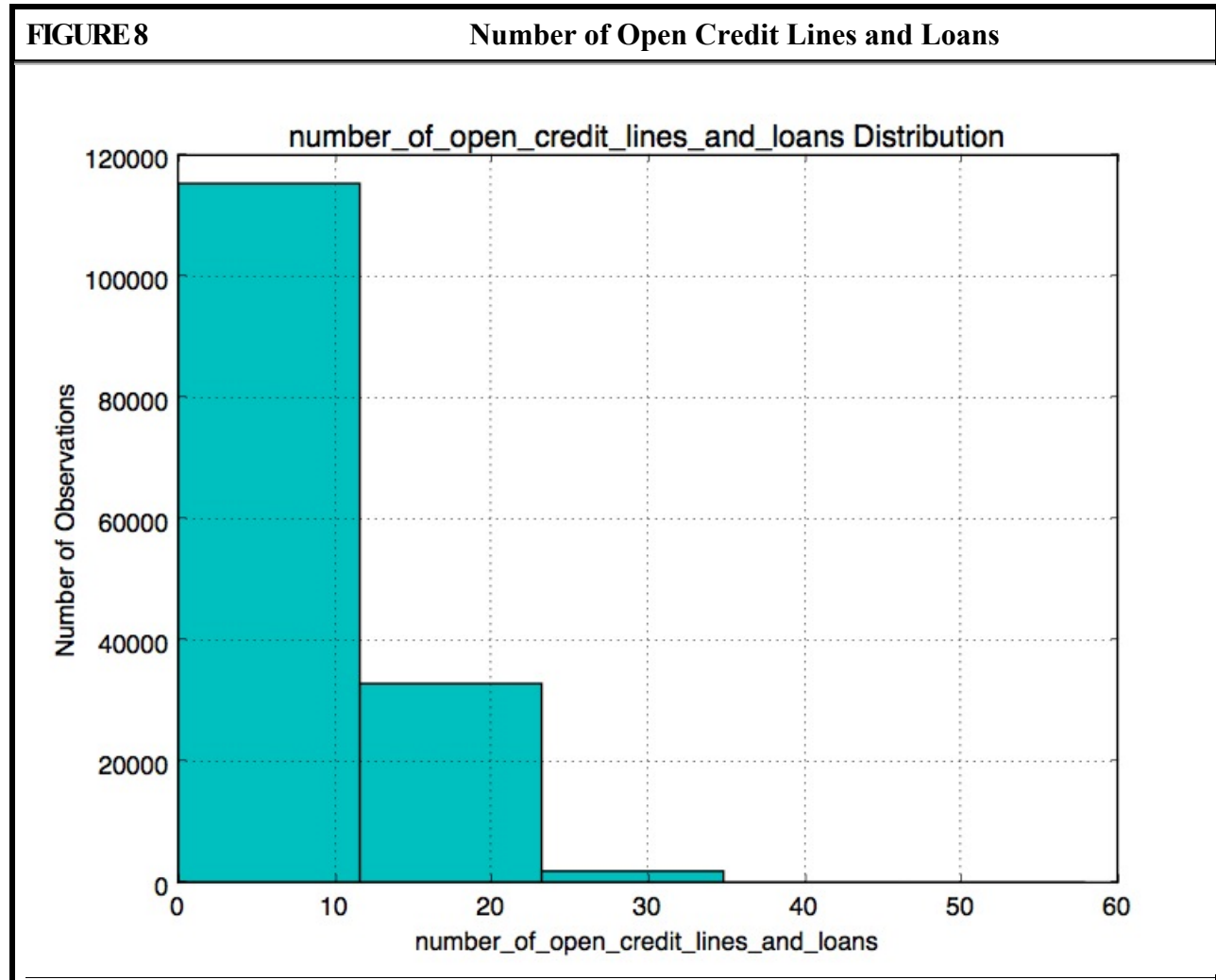
**1-viii. Summary Data for Number of Open Credit Lines and Loans:**

TABLE8		Number of Open Credit Lines and Loans	
		N =	150,000
		Missing =	0
		Mean =	8.45
		Median =	8.00
		Mode =	6.00
		Standard Deviation =	5.15
		Min, Max =	(0, 58)
Source: cs-training.csv			

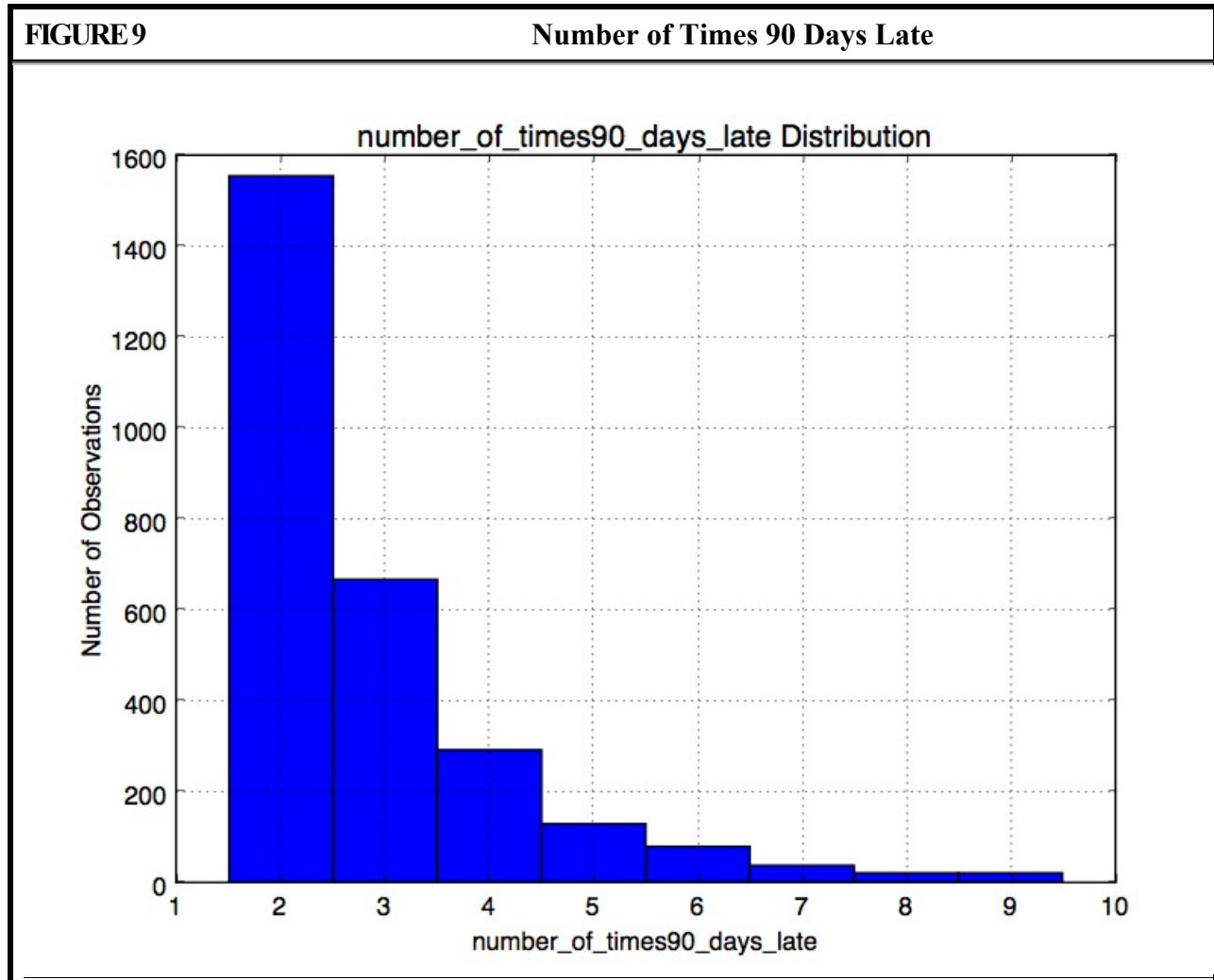
**1-ix. Summary Data for Number of Times 90 Days Late:**

TABLE9	Number of Times 90 Days Late
N =	150,000
Missing =	0
Mean =	0.27
Median =	0.00
Standard Deviation =	4.17
Min, Max =	(0, 98)
Source: cs-training.csv	

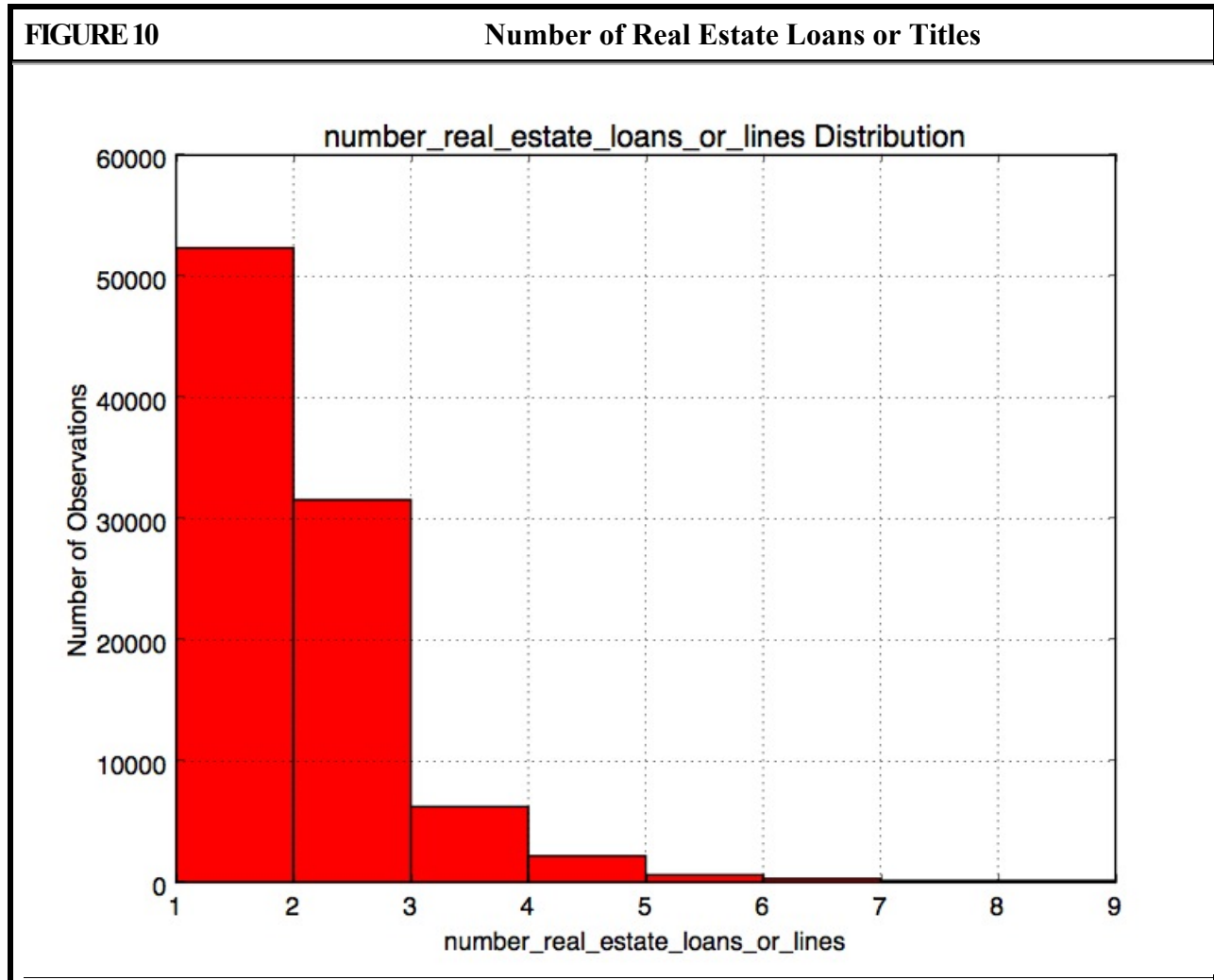
**1-x. Summary Data for Number of Real Estate Loans or Titles:**

TABLE 10		Number of Real Estate Loans or Titles	
	N =	150,000	
	Missing =	0	
	Mean =	1.02	
	Median =	1.00	
	Standard Deviation =	1.13	
	Min, Max =	(0, 54)	
Source: cs-training.csv			

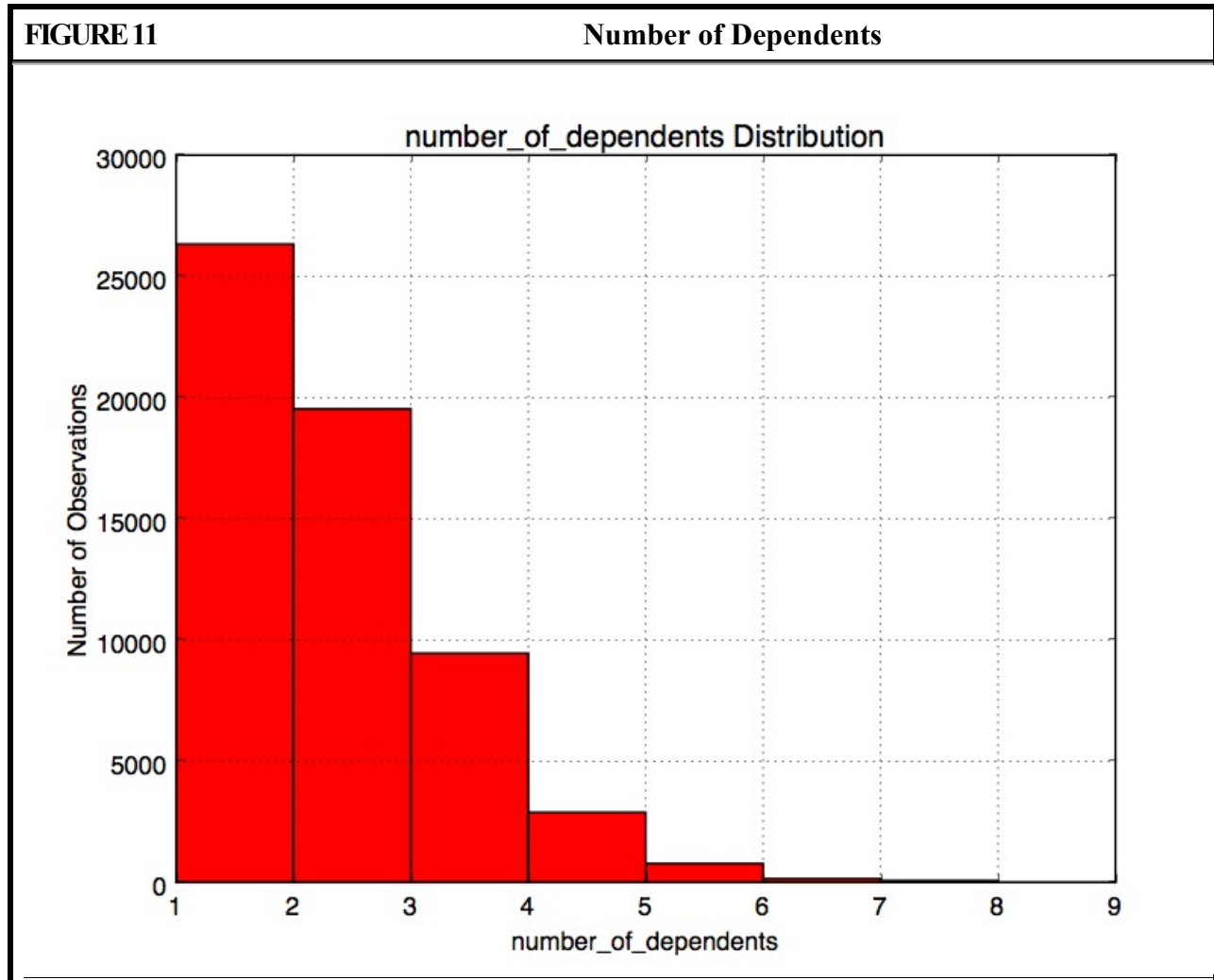
**1-xi. Summary Data for Number of Dependents:**

TABLE 11		Number of Dependents	
		N =	146,076
		Missing =	3,924
		Mean =	0.76
		Median =	0.00
		Standard Deviation =	1.12
		Min, Max =	(0, 20)
Source: cs-training.csv			

**5-i. Summary Data for Logistic Prediction Classifier for Serious Delinquency:**

TABLE 12			Logistic Regression Classifier Predicting Likelihood of Serious Delinquency Last Two Years		
Serious Delinquency Last Two Years					
Variable		Odds Ratio		P> z	
Age		0.951***		0.000	
Debt Ratio		0.999*		0.022	
Monthly Income		0.999***		0.000	
Number of Dependent		1.033***		0.000	
Number of Open Credit Lines and Loans		0.979***		0.000	
Number of Times 30-59 Days Past Due Not Worse		1.634***		0.000	
Number of Times 60-89 Days Past Due Not Worse		0.413***		0.000	
Number of Times 90 Days Late		1.523***		0.000	
Number of Real Estate Loans or Lines		1.106***		0.000	
Revolving Utilization of Unsecured Lines		0.999		0.373	
Psuedo R <sup>2</sup> = 0.069					
Log-Likelihood = -34,276.00					
Mean Classifier Accuracy = 93.38%					
(10-Fold Cross Validation)					
N = 150,000					
Source : cs-training.csv					
Note: +p<0.1, *p<.05, **p<.01, ***p<.001. T-test values. P> Z					
Logistic Model with MLE using 10-Fold Cross Validation for Mean Classifier Accuracy					