

Utility Evaluations for the Patient Characteristics Survey

Cameron Flores & Leslie Placencia

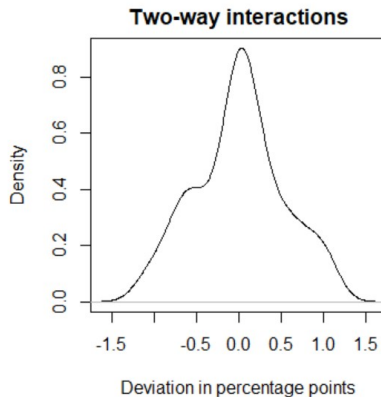
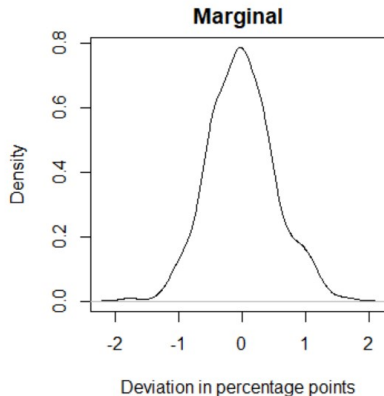
October 2021

pMSE of m Synthetic Data Sets using Logistic Regression

m	pMSE
m=1	0.006459464
m=2	0.006065811
m=3	0.00649827
m=4	0.007124899
m=5	0.005567177
Overall	0.006343124

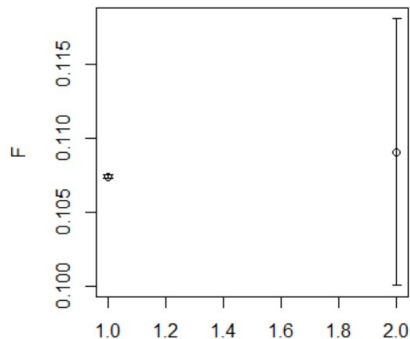
pMSE of m Synthetic Data Sets using Multi-Layer Perceptron AI	
m	pMSE
m=1	0.04537986
m=2	0.04484598
m=3	0.04271045
m=4	0.0443121
m=5	0.04324433
Overall	0.04409854

Distributions of Relative Frequencies

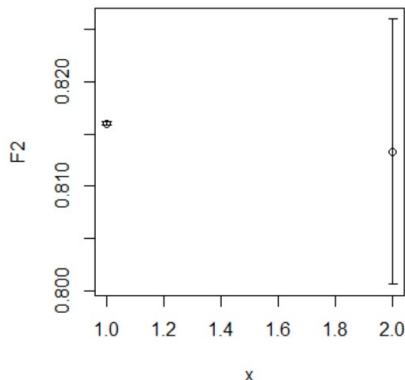


CI's for Population Proportion

Proportion of people with an alcohol-related disorder



Proportion of people who use cannabis



Interval Overlap Utility	
Variable	Interval Overlap Value
alcohol-related disorder	0.506734
drug-related disorder	0.5069527
cannabis use	0.5059634
cash assistance	0.5062462
region (Central NY)	0.5064357

Regression Coefficients

Confidential		
Variable	Coefficient	CI
age	3.37224696	[2.18205709, 5.1971431]
gender	-0.68504559	[-0.91120005, 0.4614502] -
race2	-0.02013342	[-0.77250124, 0.6731797]
Synthetic		
Variable	Coefficient	CI
age	2.850201	[1.670207, 4.030194]
gender	-0.20647201	[-0.44809353, 0.03514951]
race2	0.1218977	[-0.5567310, 0.8005265]

Regression Coefficients

Interval Overlap Utility	
Variable	Interval Overlap Value
age	-0.6661861
gender	-0.02866883
race2	0.8784609