

Risk Evaluations for the Patient Characteristics Survey

Cameron Flores & Leslie Placencia

November 2021

Matching-Based Approach

Known variables: Age, Gender, Race

Synthesized variables: Alcohol-related disorder, Drug substance disorder, Cannabis use, SSDI Cash Assistance, Region

Results on confidential data

- Expected match risk: 0
- True match rate: 0
- False match rate: 1
- Unique matches: 132

Results on synthetic data

- Average expected match risk: 0.04087355
- Average true match rate: 0
- Average false match rate: 1
- Average unique matches: 126.2

Record Linkage Approach

Available variables: Age, Gender, Race

Unavailable variables: EDUC, INSUR, EMPLOY, REGION, ALC, DRUG, CANN, CASH

Results on confidential data

	greedy	
true	FALSE	TRUE
FALSE	3341692	3780
TRUE	3780	1220

Results on synthetic data

The average true linkage percentage is $52.6/5000=1.052\%$.

The average false linkage percentage is therefore $4947.4/5000=98.948\%$.

Known variables: Age, Gender, Race

Target variables: Alcohol-related disorder

Results on confidential data

Average CAP is 0.8199345.

Results on synthetic data

Average CAP, across all $m=5$ synthetic data sets, is 0.8170616.

Classification-based Measure

We can first compare the mean-square error (MSE) of the predictions from the k-NN classifier trained on the synthetic and the confidential datasets

The larger the MSE the larger the error:

$$\text{knn_MSE_con} = 3.514667$$

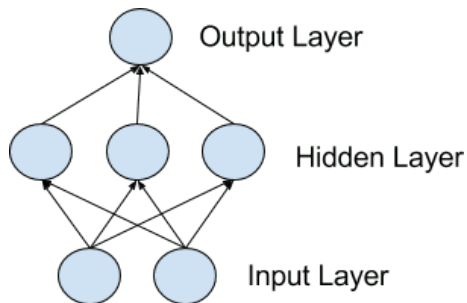
$$\text{knn_MSE_syn} = 3.5712$$

Classification-based Measure

Now we return record-level prediction. We return the proportion of observations that will have a less accurate prediction with the synthetic data compared to the confidential data

```
mean(relative_error_con >  
relative_error_syn) = 0.1744
```

Multi-Layer Perceptron



For comparison, the weights on the neurons are very much like the coefficients used in a regression equation. For every observation we input the weights update so that the next time that observation or one like it comes we can output the correct answer.