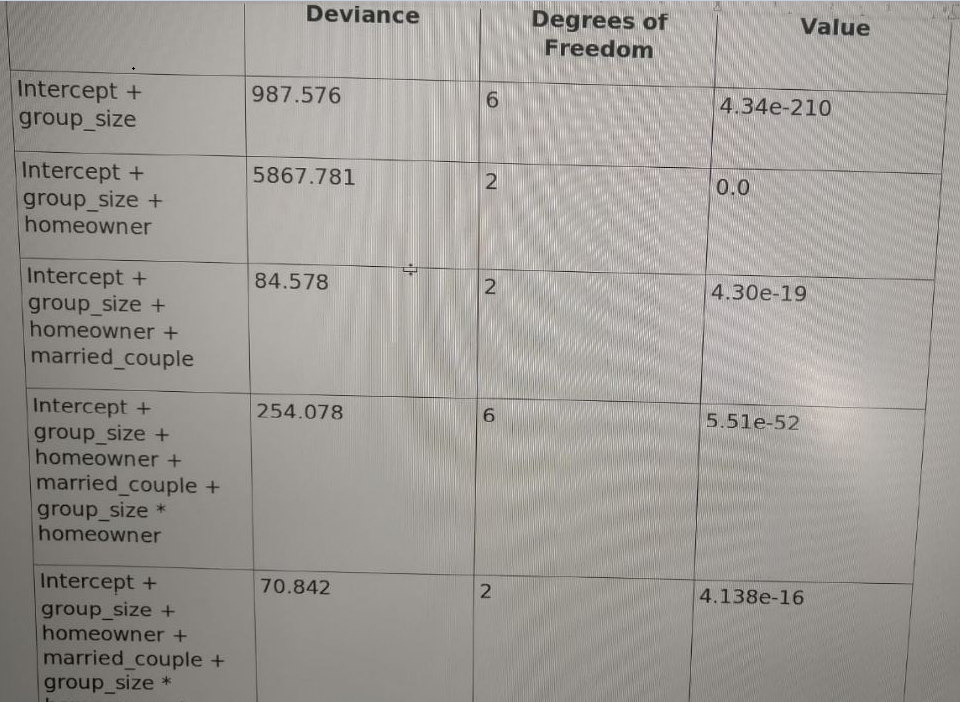
CS 584-04: Machine Learning

**Fall 2019: Assignment 4**

Amitdeb Prasad Bhattacharya A20402789

**Question 1**

1. (5 points) List the aliased parameters that you found in your model.



1. (5 points) How many degrees of freedom do you have in your model?

Degrees of Freedom = 20

1. (10 points) After entering a model effect, calculate the Deviance test statistic, its degrees of freedom, and its significance value between the current model and the previous model. List your Deviance test results by the model effects in a table.

Column Numbers of the Non-redundant Columns:

(0, 1, 2, 3, 5, 7, 9, 11, 13, 17)

Optimization terminated successfully.

Current function value: 0.889553

Iterations 5

MNLogit Regression Results

==============================================================================

Dep. Variable: A No. Observations: 665249

Model: MNLogit Df Residuals: 665229

Method: MLE Df Model: 18

Date: Sat, 02 Nov 2019 Pseudo R-squ.: 0.006101

Time: 02:13:18 Log-Likelihood: -5.9177e+05

converged: True LL-Null: -5.9541e+05

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================================

A=1 coef std err z P>|z| [0.025 0.975]

--------------------------------------------------------------------------------------------------

const 0.4396 0.091 4.822 0.000 0.261 0.618

group\_size\_1 1.0885 0.093 11.763 0.000 0.907 1.270

group\_size\_2 0.9573 0.092 10.454 0.000 0.778 1.137

group\_size\_3 0.3439 0.095 3.610 0.000 0.157 0.531

homeowner\_0 0.8002 0.259 3.093 0.002 0.293 1.307

married\_couple\_0 -0.2157 0.017 -12.873 0.000 -0.249 -0.183

group\_size\_1 \* homeowner\_0 -1.5056 0.260 -5.793 0.000 -2.015 -0.996

group\_size\_2 \* homeowner\_0 -1.1646 0.259 -4.493 0.000 -1.673 -0.657

group\_size\_3 \* homeowner\_0 -0.6546 0.267 -2.450 0.014 -1.178 -0.131

homeowner\_0 \* married\_couple\_0 0.2125 0.026 8.224 0.000 0.162 0.263

--------------------------------------------------------------------------------------------------

A=2 coef std err z P>|z| [0.025 0.975]

--------------------------------------------------------------------------------------------------

const -0.9255 0.134 -6.927 0.000 -1.187 -0.664

group\_size\_1 0.8015 0.135 5.923 0.000 0.536 1.067

group\_size\_2 0.7281 0.134 5.429 0.000 0.465 0.991

group\_size\_3 0.5275 0.138 3.810 0.000 0.256 0.799

homeowner\_0 0.5423 0.361 1.504 0.133 -0.164 1.249

married\_couple\_0 -0.1882 0.023 -8.327 0.000 -0.232 -0.144

group\_size\_1 \* homeowner\_0 -0.9834 0.362 -2.716 0.007 -1.693 -0.274

group\_size\_2 \* homeowner\_0 -0.7156 0.361 -1.981 0.048 -1.423 -0.008

group\_size\_3 \* homeowner\_0 -0.5987 0.372 -1.611 0.107 -1.327 0.130

homeowner\_0 \* married\_couple\_0 0.2124 0.035 6.065 0.000 0.144 0.281

==================================================================================================

Model Parameter Estimates:

0 1

const 0.439563 -0.925506

group\_size\_1 1.088485 0.801493

group\_size\_2 0.957293 0.728103

group\_size\_3 0.343931 0.527471

homeowner\_0 0.800157 0.542297

married\_couple\_0 -0.215748 -0.188178

group\_size\_1 \* homeowner\_0 -1.505554 -0.983441

group\_size\_2 \* homeowner\_0 -1.164638 -0.715556

group\_size\_3 \* homeowner\_0 -0.654639 -0.598700

homeowner\_0 \* married\_couple\_0 0.212483 0.212433

Model Log-Likelihood Value = -591774.333631724

Number of Free Parameters = 20

Deviance Chi=Square Test

Chi-Square Statistic = 70.84227676969022

Degrees of Freedom = 2

Significance = 4.138043547449837e-16

1. (5 points) Calculate the Feature Importance Index as the negative base-10 logarithm of the significance value. List your indices by the model effects.

Feature Importance Index for (Intercept + group\_size) = 209.36172341080683

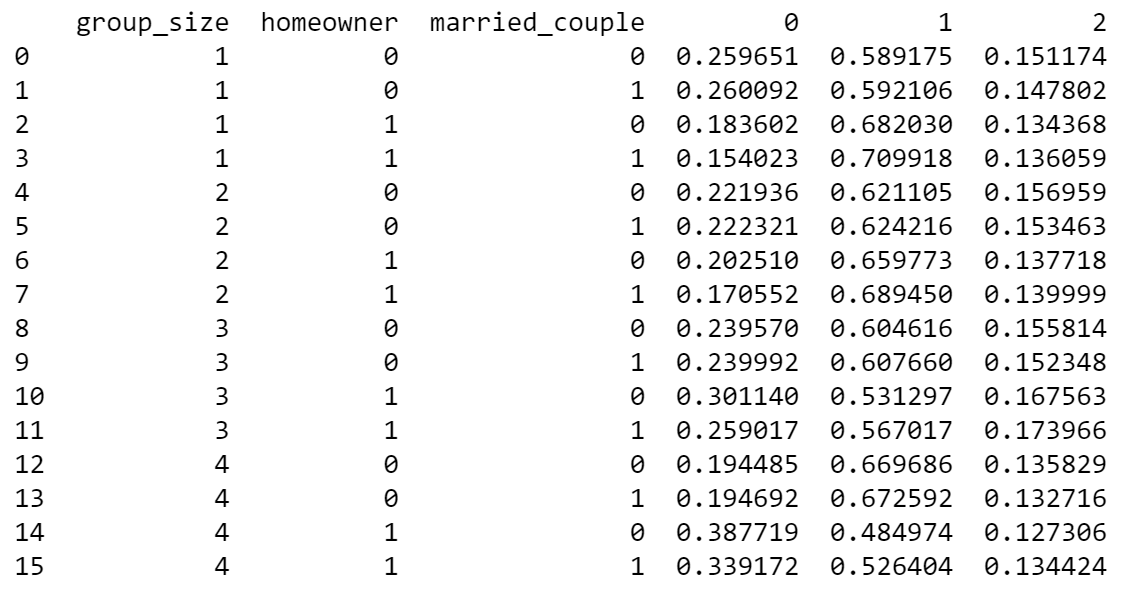
Feature Importance Index for (Intercept + group\_size + homeowner) = inf

Feature Importance Index for (Intercept + group\_size + homeowner + married\_couple) = 18.365879862820417

Feature Importance Index for (Intercept + group\_size + homeowner + married\_couple + group\_size \* homeowner) = 51.25868244189017

Feature Importance Index for (Intercept + group\_size + homeowner + married\_couple + group\_size \* homeowner + homeowner \* married\_couple) = 15.383204943269693

1. (10 points) For each of the sixteen possible value combinations of the three features, calculate the predicted probabilities for A = 0, 1, 2 based on the multinomial logistic model. List your answers in a table with proper labelling.



1. (5 points) Based on your model, what values of group\_size, homeowner, and married\_couple will maximize the odds value Prob(A=1) / Prob(A = 0)? What is that maximum odd value?
2. (5 points) Based on your model, what is the odds ratio for group\_size = 3 versus group\_size = 1, and A = 2 versus A = 0? Mathematically, the odds ratio is (Prob(A=2)/Prob(A=0) | group\_size = 3) / ((Prob(A=2)/Prob(A=0) | group\_size = 1).

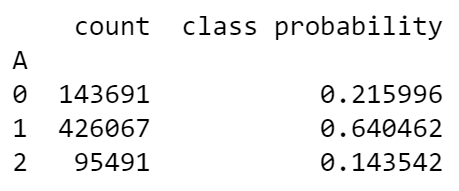
1.0249543364157785

1. (5 points) Based on your model, what is the odds ratio for homeowner = 1 versus homeowner = 0, and A = 0 versus A = 1? Mathematically, the odds ratio is (Prob(A=0)/Prob(A=1) | homeowner = 1) / ((Prob(A=0)/Prob(A=1) | homeowner = 0).

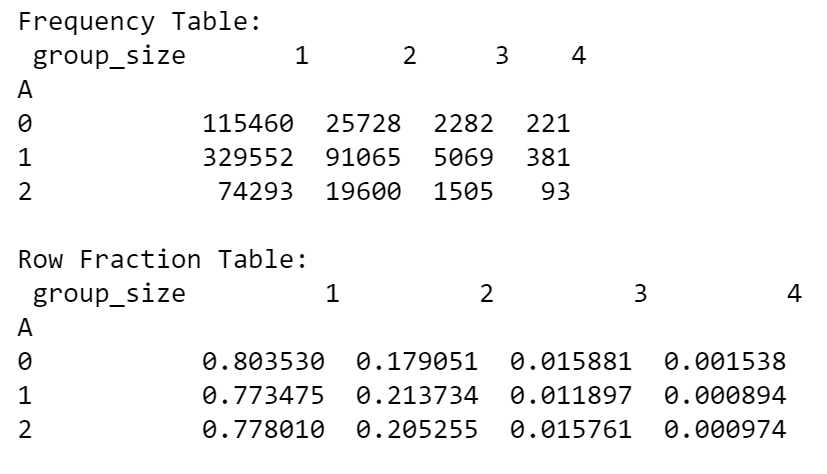
0.6232245044401726

**Question 2**

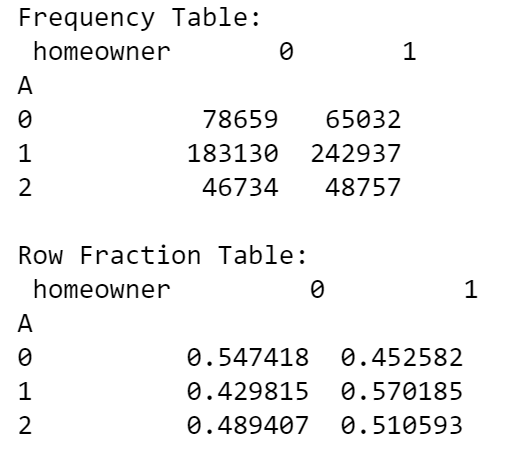
1. (5 points) Show in a table the frequency counts and the Class Probabilities of the target variable.



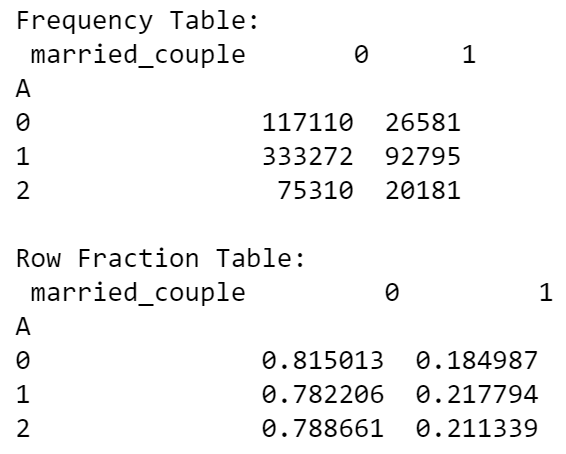
1. (5 points) Show the crosstabulation table of the target variable by the feature group\_size. The table contains the frequency counts.



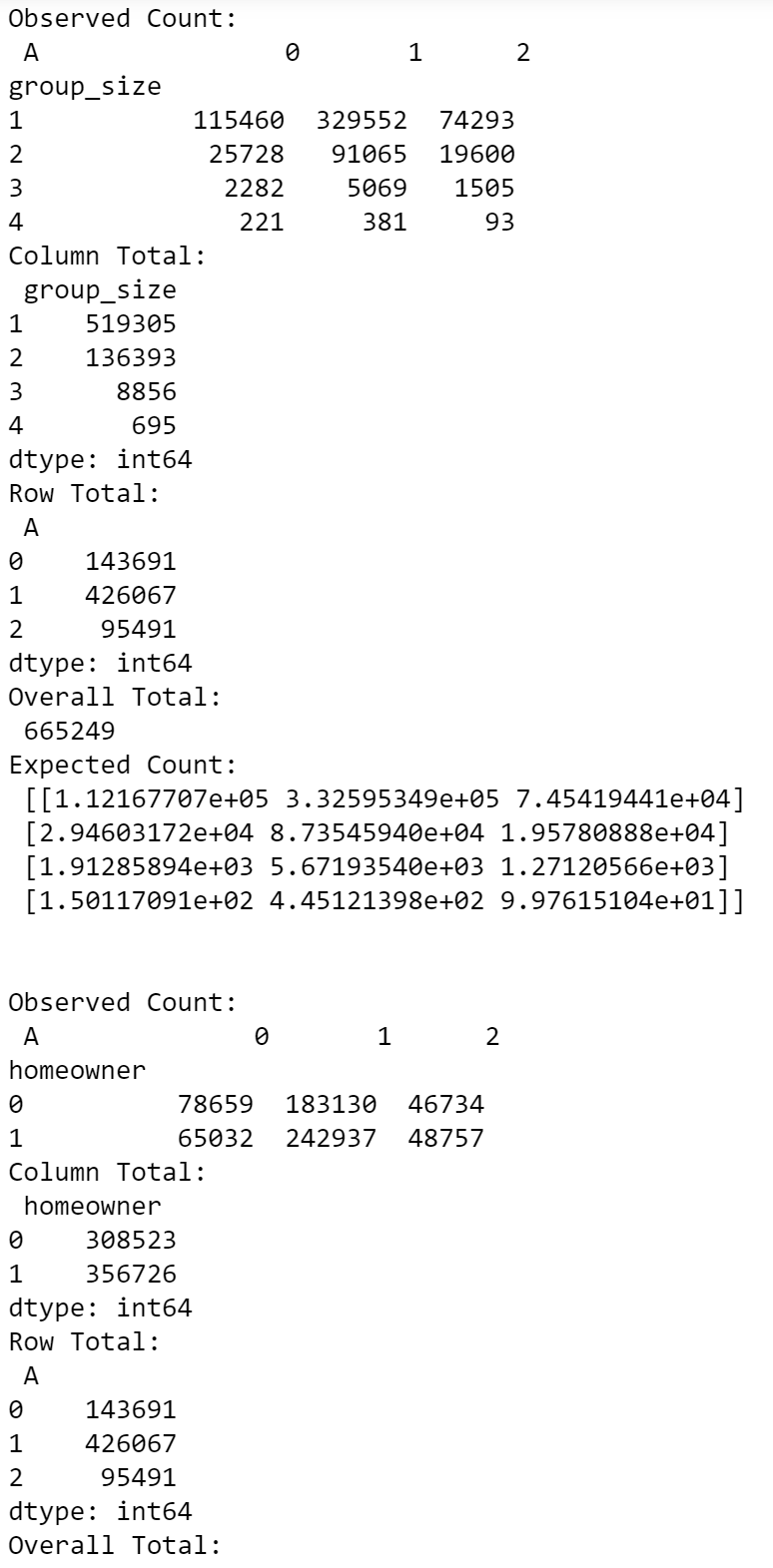
1. (5 points) Show the crosstabulation table of the target variable by the feature homeowner. The table contains the frequency counts.

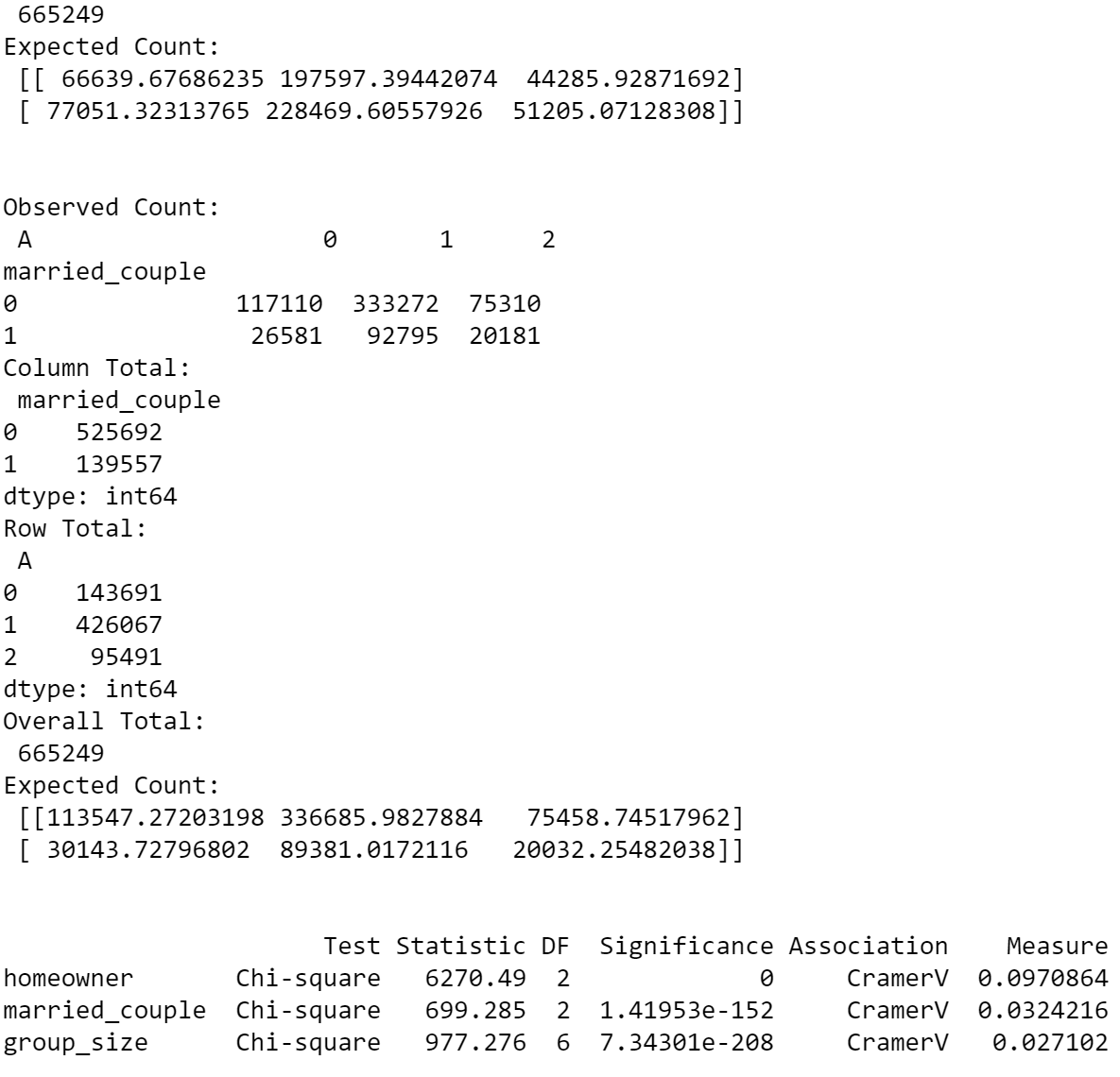


1. (5 points) Show the crosstabulation table of the target variable by the feature married\_couple. The table contains the frequency counts.



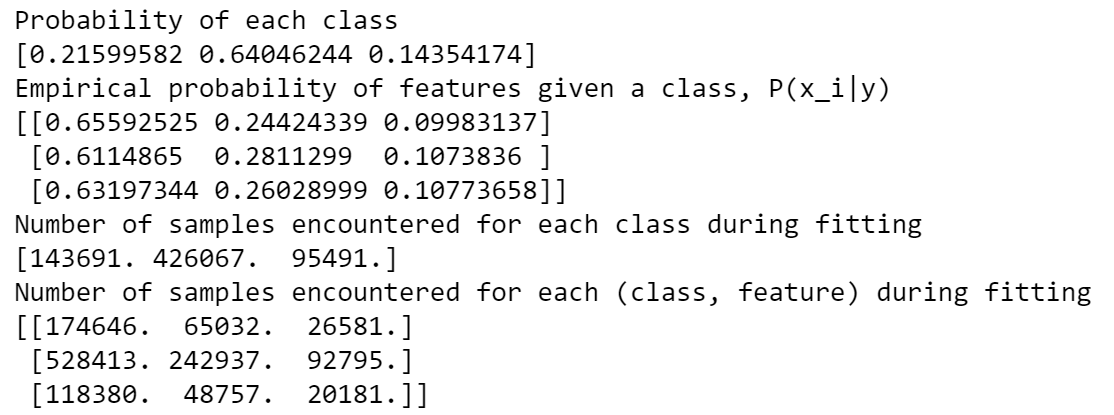
1. (10 points) Calculate the Cramer’s V statistics for the above three crosstabulations tables. Based on these Cramer’s V statistics, which feature has the largest association with the target A?



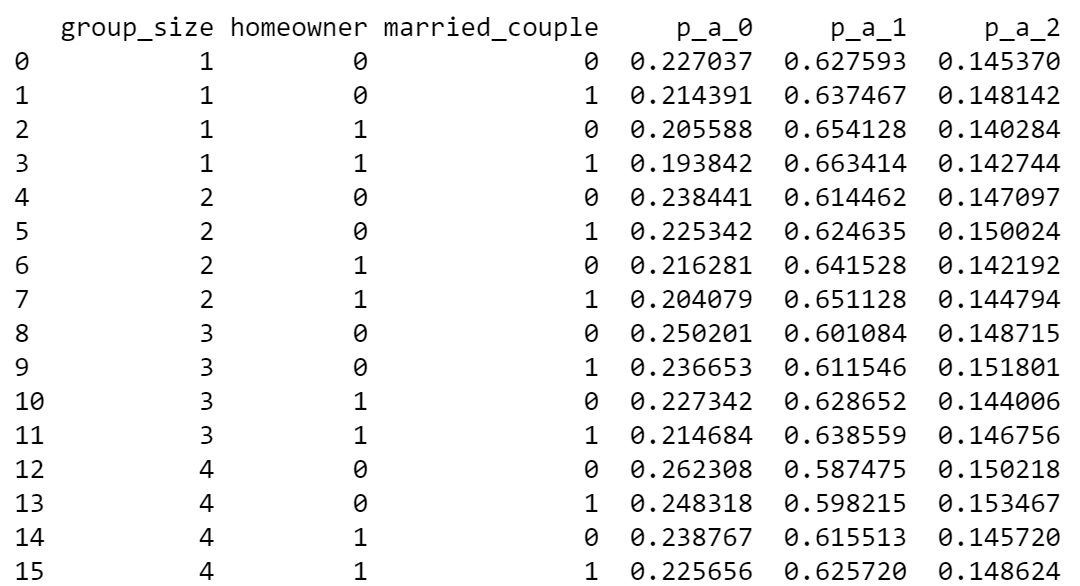


homeowner

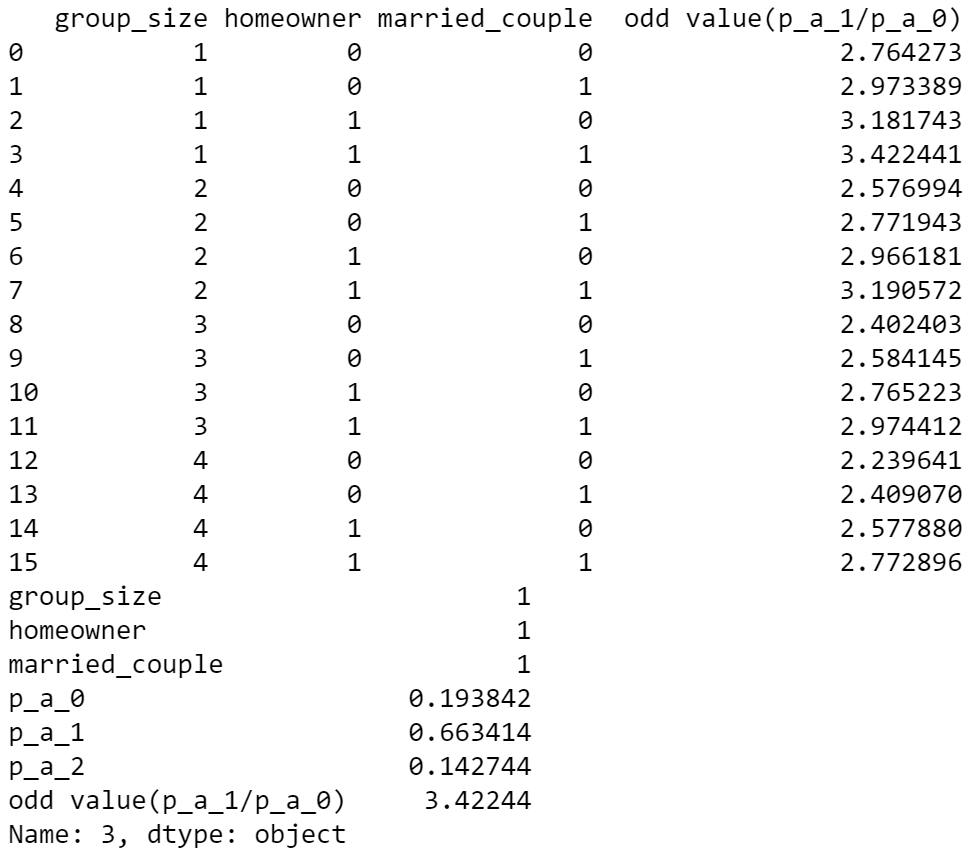
1. (5 points) Based on the assumptions of the Naïve Bayes model, express the joint probability Prob(A = a, group\_size = g, homeowner = h, married\_couple = m) as a product of the appropriate probabilities.



1. (10 points) For each of the sixteen possible value combinations of the three features, calculate the predicted probabilities for A = 0, 1, 2 based on the Naïve Bayes model. List your answers in a table with proper labelling.



1. (5 points) Based on your model, what values of group\_size, homeowner, and married\_couple will maximize the odds value Prob(A=1) / Prob(A = 0)? What is that maximum odd value?



3.42244