The dataset ‘names.dta’ contains Curriculum Vitae (CV), call-back and employer information for 4,870 fictitious CVs sent in response to employment advertisements in Chicago

in 2001. The CVs contained information concerning the race of the applicant. Because

race is not typically included on a CV, CVs were differentiated on the basis of so-called

‘white sounding names’ (such as Emily Walsh or Gregory Baker) and ‘African American

sounding names’ (such as Lakisha Washington or Jamal Jones). A large collection of fictitious CVs were created and the presupposed ‘race’ (based on the ‘sound’ of the name)

was randomly assigned to each CV. These CVs were sent to prospective employers to see

which CVs generated a phone call (a ‘call back’) from the prospective employer. You can

find variable descriptions in Table 1 below.

Today is your first day in your new job as an Economist at a US research organisation.

You get assigned a project that focuses on racial bias in employment. More specifically,

the project aims to explore if the race of the applicant has an impact on whether or not

the applicant receives a call back from a prospective employer. You are given the dataset

‘names.dta’ and asked to answer the following questions.

1. What is the call back rate in the sample? What is the call back rate for whites and

for African Americans?

. tab call\_back

call\_back | Freq. Percent Cum.

------------+-----------------------------------

0 | 4,478 91.95 91.95

1 | 392 8.05 100.00

------------+-----------------------------------

Total | 4,870 100.00

. tab call\_back black ,column row

+-------------------+

| Key |

|-------------------|

| frequency |

| row percentage |

| column percentage |

+-------------------+

| black

call\_back | 0 1 | Total

-----------+----------------------+----------

0 | 2,200 2,278 | 4,478

| 49.13 50.87 | 100.00

| 90.35 93.55 | 91.95

-----------+----------------------+----------

1 | 235 157 | 392

| 59.95 40.05 | 100.00

| 9.65 6.45 | 8.05

-----------+----------------------+----------

Total | 2,435 2,435 | 4,870

| 50.00 50.00 | 100.00

| 100.00 100.00 | 100.00

2. Use a regression model to test the effect of being black on call back rate. Make

sure to use a regression model that appropriately accounts for the binary nature of

the dependent variable (this applies for the entire assignment and only one model

is needed for each question). Do call back probabilities differ significantly by race

(black)? If yes, by how much?

. logit call\_back black

Iteration 0: log likelihood = -1363.4605

Iteration 1: log likelihood = -1355.0392

Iteration 2: log likelihood = -1354.9689

Iteration 3: log likelihood = -1354.9689

Logistic regression Number of obs = 4,870

LR chi2(1) = 16.98

Prob > chi2 = 0.0000

Log likelihood = -1354.9689 Pseudo R2 = 0.0062

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

call\_back | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

black | -.4381802 .1073232 -4.08 0.000 -.6485298 -.2278306

\_cons | -2.236627 .0686285 -32.59 0.000 -2.371136 -2.102118

Do call back probabilities differ significantly by race

No. The Pseudo R2 is least statistically I insignificant.

3. Using call back as the dependent variable, does an African American/white name on

the CV have a different effect for men compared to women?

. melogit call\_back black female

Iteration 0: log likelihood = -1390.0942

Iteration 1: log likelihood = -1354.698

Iteration 2: log likelihood = -1354.4728

Iteration 3: log likelihood = -1354.4727

Logistic regression Number of obs = 4,870

Wald chi2(2) = 17.64

Log likelihood = -1354.4727 Prob > chi2 = 0.0001

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

call\_back | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

black | -.4395834 .1073436 -4.10 0.000 -.6499729 -.2291939

female | .1270304 .1288154 0.99 0.324 -.1254432 .3795039

\_cons | -2.334819 .1218919 -19.15 0.000 -2.573723 -2.095915

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

4. Using call-back as the dependent variable, is there a significant difference in high-

quality/low-quality CVs for whites versus African Americans?

Please provide an explanation for your finding (tip: read through the data description above when answering this question).

. melogit call\_back high black

Iteration 0: log likelihood = -1389.478

Iteration 1: log likelihood = -1353.5848

Iteration 2: log likelihood = -1353.3358

Iteration 3: log likelihood = -1353.3356

Iteration 4: log likelihood = -1353.3356

Logistic regression Number of obs = 4,870

Wald chi2(2) = 19.89

Log likelihood = -1353.3356 Prob > chi2 = 0.0000

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

call\_back | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

high | .1911067 .1059381 1.80 0.071 -.0165282 .3987416

black | -.4384713 .1073585 -4.08 0.000 -.64889 -.2280526

\_cons | -2.336293 .0894162 -26.13 0.000 -2.511545 -2.16104

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

The P>|z| value of call\_back versus high is statistically signant i.e 0.071 while that of call\_back versus black is statistically insignificant i.e 0.000

5. Estimate a regression model that tests for racial bias in call back rates and controls for

all relevant variables. What is the effect of race (black) for a sample average applicant?

. mprobit call\_back black

Iteration 0: log likelihood = -1357.5375

Iteration 1: log likelihood = -1354.9702

Iteration 2: log likelihood = -1354.9689

Iteration 3: log likelihood = -1354.9689

Multinomial probit regression Number of obs = 4,870

Wald chi2(1) = 16.84

Log likelihood = -1354.9689 Prob > chi2 = 0.0000

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

call\_back | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

0 | (base outcome)

-------------+----------------------------------------------------------------

1 |

black | -.3062382 .0746336 -4.10 0.000 -.4525174 -.159959

\_cons | -1.840884 .0494928 -37.20 0.000 -1.937888 -1.74388

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

. What is the effect of race (black) for a sample average applicant?

Analysing the P>|z| value there is zero to almost negligible effect of race for a sample average applicant keeping controls on the other variables.

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

Table 1: Variable Descriptions

call back 1= applicant was called back for interview; 0 otherwise

female 1= female; 0 otherwise

black 1= black, African American (indicator for race); 0 otherwise

high 1= high quality CV; 0 otherwise

ofjobs = number of jobs listed on CV

yearsexp = number of years of work experience on the CV

honors 1= CV mentions some honors; 0 otherwise

volunteer 1= CV mentions some volunteering experience; 0 otherwise

college 1= applicant has college degree or more; 0 otherwise

specialskills 1= CV mentions some special skills; 0 otherwise

dog owner 1= applicant owns a dog; 0 otherwise