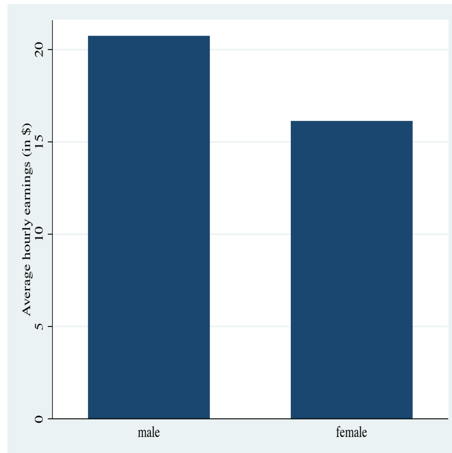
Elisabeth Sadoulet

Is there gender discrimination in wage? Using dummy variables and interactions

Source: These are data from the 2006 Current Population Survey. 2000 working adults

```
wage            float  %9.0g          Average hourly earnings (in $)
educ            byte   %8.0g          Years of education
exper           byte   %8.0g          Potential years of experience
female          byte   %8.0g          Female
union           byte   %8.0g          Union member
profocc         byte   %8.0g          Occupation is professional
services        byte   %8.0g          Works in service sector
cateduc         float  %9.0g          Educ: incomplete high, high sch., some college
```

1. Estimating difference in means between male and female:



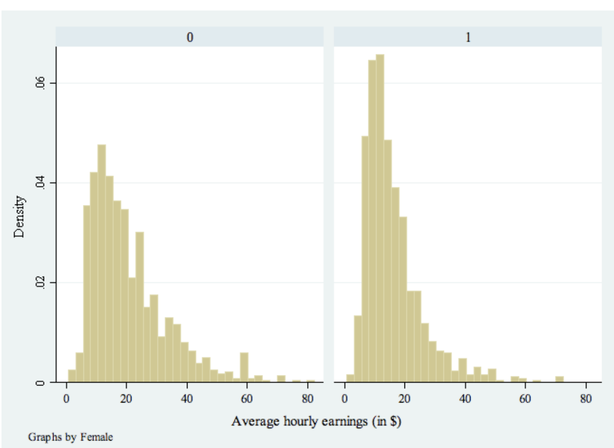

```
. sum female wage if female==1;
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      female |      1033           1           0          1          1
        wage |      1033    16.12258    9.715608      2.125     72.125

. sum female wage if female==0;
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      female |       967           0           0          0          0
        wage |       967    20.72326    12.71402         .7   82.42857
```

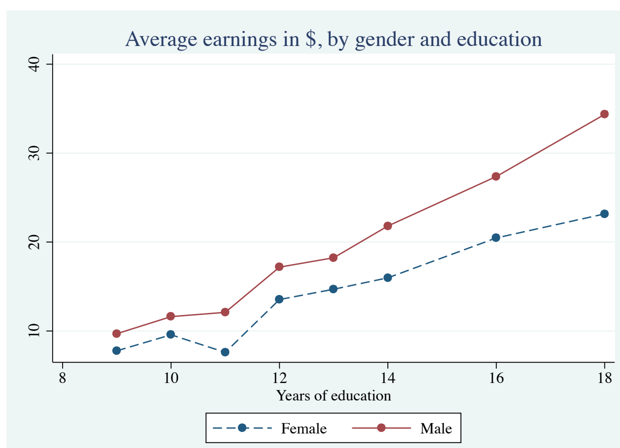
Test?

```
. ttest wage, by(female);

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     967    20.72326    .4088552    12.71402    19.92091    21.52561
       1 |    1033    16.12258    .3022872    9.715608    15.52942    16.71575
---------+--------------------------------------------------------------------
combined |    2000    18.34701    .2570348    11.49495    17.84293    18.85109
---------+--------------------------------------------------------------------
    diff |            4.600677    .5040778                3.612104     5.58925
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   9.1269
Ho: diff = 0                                     degrees of freedom =     1998

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

2. Whole distribution of wage (hourly earnings in dollars)

```
. histogram wage, by(female)
```





3. Discrimination, even after controlling for difference in characteristics: Additive female effect.

```
. tabstat wage educ exper union service profocc , by(female)

Summary statistics: mean
  by categories of: female (Female)

  female |      wage      educ     exper     union  services   profocc
---------+------------------------------------------------------------
       0 |  20.72326   13.5274   20.3061  .1323681  .1323681  .1664943
       1 |  16.12258  13.73185  20.84608  .1452081  .1703775  .2545983
---------+------------------------------------------------------------
   Total |  18.34701    13.633    20.585      .139      .152      .212
----------------------------------------------------------------------
```

And yes, the wages are lower in the service sector (average of $12.9 compared to $19.3) and higher for professional occupation (average of $22.9 compared to $17.1)

```
Use:   tabstat wage, by(service)
```

Note that the female wages are lower even within the same sector, or within the same professional category

```
. table female services, c(mean wage)

------------------------------
          |  Works in service
          |       sector
   Female |       0         1
----------+-------------------
        0 | 21.62085  14.83982
        1 |  17.0615  11.55071
------------------------------

. reg wage female educ exper union service profocc

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  6,  1993) =  118.29
       Model |  69360.4741     6   11560.079           Prob > F      =  0.0000
    Residual |  194775.003  1993  97.7295551           R-squared     =  0.2626
-------------+------------------------------           Adj R-squared =  0.2604
       Total |  264135.478  1999  132.133806           Root MSE      =  9.8858

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.161436   .4466179   -11.56   0.000    -6.037323   -4.285549
        educ |   2.000723   .1182173    16.92   0.000      1.76888    2.232565
       exper |   .1711896   .0175183     9.77   0.000     .1368336    .2055457
       union |   2.229349   .6460632     3.45   0.001      .962319    3.496379
    services |  -2.729558   .6422974    -4.25   0.000    -3.989203   -1.469914
     profocc |   1.531844   .6118991     2.50   0.012     .3318149    2.731873
       _cons |  -10.00724   1.645929    -6.08   0.000    -13.23516   -6.779315
------------------------------------------------------------------------------
```

Test? Interpretation?

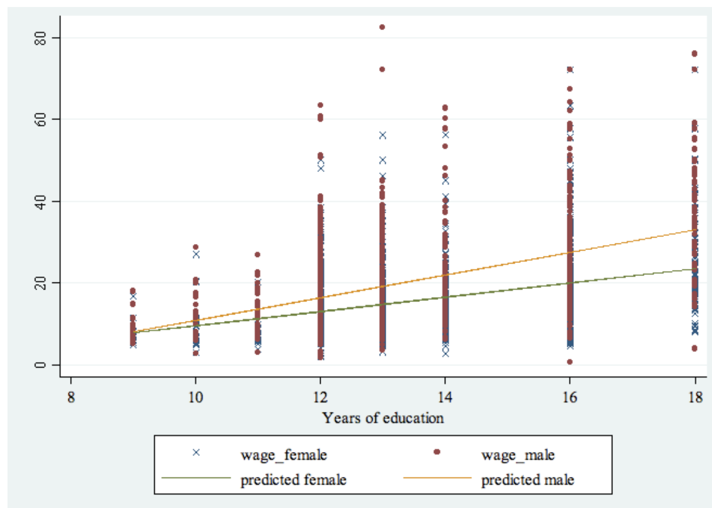
Estimating difference in means with a simple regression, not controlling for characteristics:

```
. reg wage female;

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  1,  1998) =   83.30
       Model |   10571.589     1   10571.589           Prob > F      =  0.0000
    Residual |  253563.889  1998  126.908853           R-squared     =  0.0400
-------------+------------------------------           Adj R-squared =  0.0395
       Total |  264135.478  1999  132.133806           Root MSE      =  11.265

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -4.600677   .5040778    -9.13   0.000     -5.58925   -3.612104
       _cons |   20.72326   .3622703    57.20   0.000     20.01279    21.43373
------------------------------------------------------------------------------
```

4. Do females have differential return to some characteristics?

Is there a differential return to education for male and female? Interaction between dummy and continuous variables

```
. g femeduc=female*educ
. reg wage female educ femeduc
      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  3,  1996) =  182.86
       Model |  56945.4372     3  18981.8124           Prob > F      =  0.0000
    Residual |   207190.04  1996  103.802625           R-squared     =  0.2156
-------------+------------------------------           Adj R-squared =  0.2144
       Total |  264135.478  1999  132.133806           Root MSE      =  10.188

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   8.838706   3.013838     2.93   0.003     2.928108     14.7493
        educ |   2.772576   .1560434    17.77   0.000     2.466551    3.078601
     femeduc |  -1.019981   .2186047    -4.67   0.000    -1.448698   -.5912633
       _cons |   -16.7825   2.136137    -7.86   0.000    -20.97179   -12.59321
------------------------------------------------------------------------------
Female effect on wage = (8.8 – 1.02 educ)
Education effect on wage = (2.77 - 1.02 female)
```




```
* graph;
qui reg wage female educ femeduc ;
predict wagehat;
gen wage_female=wage if female==1;
gen wage_male=wage if female==0;
gen trfem=wagehat if female==1;
gen trmale=wagehat if female==0;
label variable trfem "predicted female";
label variable trmale "predicted male";
twoway scatter wage_female  wage_male trfem
        trmale educ, ms(X o i i) c(i i l l);
```

Is there a differential effect of union on women and men's wage: interaction between dummy variables

```
. gen femunion=female*union;
. reg wage female union femunion educ exper;

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  5,  1994) =  134.42
       Model |  66586.3651     5   13317.273           Prob > F      =  0.0000
    Residual |  197549.112  1994  99.0717715           R-squared     =  0.2521
-------------+------------------------------           Adj R-squared =  0.2502
       Total |  264135.478  1999  132.133806           Root MSE      =  9.9535

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.094864   .4801934   -10.61   0.000    -6.036597    -4.15313
       union |   2.577524   .9480136     2.72   0.007     .7183235    4.436725
    femunion |   -.612306    1.29469    -0.47   0.636    -3.151394    1.926782
        educ |   2.228979   .1073513    20.76   0.000     2.018447    2.439512
       exper |   .1756898   .0175311    10.02   0.000     .1413085    .2100711
       _cons |  -13.33871   1.533331    -8.70   0.000    -16.3458   -10.33161
------------------------------------------------------------------------------
```

General case of interaction terms
Does the marginal effect of experience depend on education?

```
. gen expeduc=exper*educ
. reg wage female educ exper expeduc;

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  4,  1995) =  164.13
       Model |  65400.0183     4  16350.0046           Prob > F      =  0.0000
    Residual |  198735.459  1995  99.6167715           R-squared     =  0.2476
-------------+------------------------------           Adj R-squared =  0.2461
       Total |  264135.478  1999  132.133806           Root MSE      =  9.9808

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.145376   .4478427   -11.49   0.000    -6.023664   -4.267087
        educ |   2.158549   .1985612    10.87   0.000      1.76914    2.547958
       exper |   .1162292   .1165983     1.00   0.319    -.112438     .3448965
     expeduc |   .0049071   .0087011     0.56   0.573    -.0121571    .0219713
       _cons |  -12.19337   2.675967    -4.56   0.000    -17.44135   -6.945383
------------------------------------------------------------------------------
```

5. Use of ordinal variables

If education is given in 3 levels: cateduc =1 for high school dropout, =2 for high school, and =3 for some college education.

```
. gen HSDrop = cateduc ==1
. gen HS = cateduc ==2
. gen Col = cateduc ==3
. reg wage female HS Col exper

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  4,  1995) =   99.55
       Model |  43948.4454     4  10987.1113           Prob > F      =  0.0000
    Residual |  220187.032  1995   110.36944           R-squared     =  0.1664
-------------+------------------------------           Adj R-squared =  0.1647
       Total |  264135.478  1999  132.133806           Root MSE      =  10.506

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.144792   .4713192   -10.92   0.000    -6.069121   -4.220462
          HS |   3.851033   1.011044     3.81   0.000     1.868221    5.833845
         Col |   9.865773   .9566261    10.31   0.000     7.989682    11.74186
       exper |   .1773909   .0187487     9.46   0.000     .1406219    .2141599
       _cons |   9.996611   .9495125    10.53   0.000     8.134471    11.85875
------------------------------------------------------------------------------
```

Would it make sense to treat cateduc as if it was a real number?

```
. reg wage female cateduc exper;

. . . . .
------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.154772   .4714936   -10.93   0.000    -6.079443   -4.230101
     cateduc |   5.417826   .3805958    14.24   0.000     4.671419    6.164233
. . . .
```

6. Are the wage equations for male and female the same?

```
. reg wage educ exper;

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  2,  1997) =  245.43
       Model |  52115.2387     2  26057.6194           Prob > F      =  0.0000
    Residual |  212020.239  1997  106.169373           R-squared     =  0.1973
-------------+------------------------------           Adj R-squared =  0.1965
       Total |  264135.478  1999  132.133806           Root MSE      =  10.304

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   2.192429   .1103889    19.86   0.000      1.97594    2.408919
       exper |   .1770016   .0180643     9.80   0.000     .1415748    .2124285
       _cons |  -15.18658   1.566867    -9.69   0.000    -18.25944   -12.11371
------------------------------------------------------------------------------

. reg wage educ exper if female==1;

      Source |       SS       df       MS              Number of obs =    1033
-------------+------------------------------           F(  2,  1030) =  101.42
       Model |  16027.8434     2  8013.92168           Prob > F      =  0.0000
    Residual |  81385.7713  1030   79.015312           R-squared     =  0.1645
-------------+------------------------------           Adj R-squared =  0.1629
       Total |  97413.6147  1032  94.3930375           Root MSE      =  8.8891

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.795799   .1337998    13.42   0.000     1.533248     2.05835
       exper |   .1210147    .021846     5.54   0.000     .0781469    .1638824
       _cons |  -11.05997   1.938354    -5.71   0.000    -14.86355   -7.256399
------------------------------------------------------------------------------

. reg wage educ exper if female==0;

      Source |       SS       df       MS              Number of obs =     967
-------------+------------------------------           F(  2,   964) =  174.73
       Model |  41545.5802     2  20772.7901           Prob > F      =  0.0000
    Residual |  114604.694   964  118.884537           R-squared     =  0.2661
-------------+------------------------------           Adj R-squared =  0.2645
       Total |  156150.274   966  161.646246           Root MSE      =  10.903

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |     2.6853   .1673038    16.05   0.000     2.356978    3.013621
       exper |   .2352962   .0273877     8.59   0.000     .1815499    .2890426
       _cons |  -20.38104   2.324115    -8.77   0.000    -24.94195   -15.82013
------------------------------------------------------------------------------

. g femexper=female*exper
. reg wage educ exper female femeduc femexper;

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  5,  1994) =  138.66
       Model |  68145.0125     5  13629.0025           Prob > F      =  0.0000
    Residual |  195990.465  1994  98.2901028           R-squared     =  0.2580
-------------+------------------------------           Adj R-squared =  0.2561
       Total |  264135.478  1999  132.133806           Root MSE      =  9.9141

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |     2.6853   .1521241    17.65   0.000     2.386961    2.983638
       exper |   .2352962   .0249028     9.45   0.000     .1864581    .2841344
      female |   9.321065   3.023169     3.08   0.002     3.392165    15.24996
     femeduc |  -.8895005    .213099    -4.17   0.000    -1.307421   -.4715804
    femexper |  -.1142815   .0348398    -3.28   0.001    -.1826078   -.0459553
       _cons |  -20.38104   2.113245    -9.64   0.000    -24.52544   -16.23664
------------------------------------------------------------------------------
```

```
. test femeduc femexper;
 ( 1)  femeduc = 0
 ( 2)  femexper = 0

       F(  2,  1994) =   14.12
       Prob > F =    0.0000
```

```
. test female femeduc femexper;
( 1)  female = 0
( 2)  femeduc = 0
( 3)  femexper = 0
       F(  3,  1994) =   54.36
       Prob > F =    0.0000
```

7. Pooled Cross Sections – Looking at the evolution of gender wage gap over time.

Source: Compiled from the 1996 and 2006 Current Population Surveys

```
union           float  %8.0g                 =1 if respondent is union member
wage            float  %9.0g                 average hourly earnings (in $)
female          float  %9.0g                 1=female, 0=male
educ            float  %9.0g                 years of education
exper           float  %9.0g                 years potential experience
year            float  %9.0g                 1996 or 2006

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       union |      4000      .98975    .9251264          0          2
        wage |      4000    15.59488    10.43896        .05     144.25
      female |      4000      .50125    .5000609          0          1
        educ |      4000     13.5125    2.113637          9         18
       exper |      4000     19.7475    12.54303          0         64
        year |      4000        2001    5.000625       1996       2006
```

a. To compare wages across different years, need to use real wages

Use the Consumer Price Index (CPI) 1996: 156.9, 2006: 201.6, i.e., inflation has been 28.5% over 10 years, which is in average 2.53% per year

```
. g rwage=wage
. replace rwage=wage*201.6/156.9) if year==1996
. label variable rwage "average real hourly earnings (in 2006$)"
. bysort year: sum wage rwage;

------------------------------------------------------------------------------------------
-> year = 1996

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |      2000    12.69577    7.907586       .125       62.5
       rwage |      2000    16.31273    10.16042   .1606119   80.30593

------------------------------------------------------------------------------------------
-> year = 2006

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |      2000    18.49399    11.77498        .05     144.25
       rwage |      2000    18.49399    11.77498        .05     144.25

. gen yr2006=(year==2006);

. reg rwage educ exper union female yr2006;

      Source |       SS       df       MS              Number of obs =    4000
-------------+------------------------------           F(  5,  3994) =  293.43
       Model |  131177.324     5  26235.4647           Prob > F      =  0.0000
    Residual |  357107.093  3994  89.4108896           R-squared     =  0.2686
-------------+------------------------------           Adj R-squared =  0.2677
       Total |  488284.417  3999   122.10163           Root MSE      =  9.4557

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
        educ |   2.327535    .071202    32.69   0.000     2.187939    2.467131
       exper |   .1504879   .0119518    12.59   0.000     .1270557    .1739202
       union |  -.7340047   .4264605    -1.72   0.085    -1.570105     .102096
      female |  -4.748954    .299762   -15.84   0.000    -5.336655   -4.161253
      yr2006 |  -.2317524   .7904708    -0.29   0.769    -1.781516    1.318012
       _cons |  -13.79644   1.249616   -11.04   0.000    -16.24639    -11.3465
------------------------------------------------------------------------------
```

b. What happened to the gender gap?

```
. gen yr2006female=yr2006*female;

. reg rwage educ exper union female yr2006 yr2006female;

      Source |       SS       df       MS              Number of obs =    4000
-------------+------------------------------           F(  6,  3993) =  245.60
       Model |   131623.93     6  21937.3217           Prob > F      =  0.0000
    Residual |  356660.487  3993  89.3214342           R-squared     =  0.2696
-------------+------------------------------           Adj R-squared =  0.2685
       Total |  488284.417  3999   122.10163           Root MSE      =   9.451

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   2.332599   .0712024    32.76   0.000     2.193003    2.472196
       exper |   .1507326   .0119463    12.62   0.000     .1273111    .1741541
       union |  -.7856355   .4268721    -1.84   0.066    -1.622543    .0512721
      female |  -4.078936   .4237362    -9.63   0.000    -4.909695   -3.248176
      yr2006 |   .3487309   .8316317     0.42   0.675    -1.281731    1.979193
yr2006female |  -1.339191   .5989048    -2.24   0.025    -2.513378    -.165003
       _cons |  -14.10789   1.256733   -11.23   0.000    -16.57178   -11.64399
------------------------------------------------------------------------------
```

The New York Times
December 12, 2002, Thursday, Late Edition - Final
Economic Scene; Sticks and stones can break bones, but the wrong name can make a job hard to find. By Alan B. Krueger
WHAT'S in a name? Evidently plenty if you are looking for a job.

To test whether employers discriminate against black job applicants, Marianne Bertrand of the University of Chicago and Sendhil Mullainathan of M.I.T. conducted an unusual experiment. They selected 1,300 help-wanted ads from newspapers in Boston and Chicago and submitted multiple resumes from phantom job seekers. The researchers randomly assigned the first names on the resumes, choosing from one set that is particularly common among blacks and from another that is common among whites.

So Kristen and Tamika, and Brad and Tyrone, applied for jobs from the same pool of want ads and had equivalent resumes. Nine names were selected to represent each category: black women, white women, black men and white men. Last names common to the racial group were also assigned. Four resumes were typically submitted for each job opening, drawn from a reservoir of 160. Nearly 5,000 applications were submitted from mid-2001 to mid-2002. Professors Bertrand and Mullainathan kept track of which candidates were invited for job interviews.

No single employer was sent two identical resumes, and the names on the resumes were randomly assigned, so applicants with black- and white-sounding names applied for the same set of jobs with the same set of resumes.

Apart from their names, applicants had the same experience, education and skills, so employers had no reason to distinguish among them.

The results are disturbing. Applicants with white-sounding names were 50 percent more likely to be called for interviews than were those with black-sounding names. Interviews were requested for 10.1 percent of applicants with white-sounding names and only 6.7 percent of those with black-sounding names.

Within racial groups, applications with men's or women's names were equally likely to result in calls for interviews, providing little evidence of discrimination based on sex in these entry-level jobs.

Their most alarming finding is that the likelihood of being called for an interview rises sharply with an applicant's credentials -- like experience and honors -- for those with white-sounding names, but much less for those with black-sounding names. A grave concern is that this phenomenon may be damping the incentives for blacks to acquire job skills, producing a self-fulfilling prophecy that perpetuates prejudice and misallocates resources.

(Source: "Are Emily and Brendan More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination" *The American Economic Review*, 2004)