API-202 B: Empirical Analysis for Policy
Stock

John F. Kennedy School of Government
API-202B, Spring 1994

Packet API-202 B.f

Assignment #1
Returns to Education and the "Gender Gap"

Due: Monday February 7, 1994

It is well known that on average women earn less than men; a commonly quoted statistic is that women's hourly earnings are only 59% of men's. One possible explanation is that there historically has been intentional or unintentional sex discrimination in compensation. However, there are other explanations of the wage gap which do not directly involve discrimination in the workplace. For example women might have lower levels of education, or fewer years of on-the-job experience, than men of the same age.

One approach to examining these "third factors" is through the use of chi-squared statistics. However, this approach is cumbersome and, more importantly, does not work well when there are multiple explanations, because the number of observations in each "cell" can be too small. In contrast, multiple regression provides a simple way to control for these multiple effects. In so doing, it also becomes possible to answer another interesting question: what is the return to a year of education, that is, on average, by what percent does an additional year of education increase hourly earnings?

This exercise examines the gender gap using a sample of individuals in the data set "CPS83," which is a sample of 1000 individuals drawn at random from the May 1983 Current Population Survey (CPS). A brief description of the data set and variable definitions is attached. You will focus on the wage gap for full time workers. Also, the analysis will focus on percentage gaps in hourly wages (more precisely, the gap in log wages). This focus on percentages (logarithms) makes sense: across a wide range of wages and jobs, one might expect to see wage gaps of a certain percent rather than a fixed dollar value. This common-sense leaning towards logarithms is supported by a very large body of research which suggests that wage equations are most appropriately specified in logarithms.

Part A: Regressions

General instructions. Define the variable fem, where fem= 0 if a man and = 1 if a woman. Let hwage = hourly wage, lhwage = log(hwage). Throughout, compute statsitics for restricting the data to full time workers only (workers with weekly hours of 35 hours or greater).

Econometric Analysis. Using SST, perform the following calculations.

A1. Preliminary data analysis.
(a) Compute the sample mean and standard deviation of hourly earnings (hwage) and log hourly earnings (lhwage) for men and then for women. What is your estimate of the wage gap, in logs?
(b) Compute the t-statistic testing the hypothesis that log hourly earnings are equal for men and women.
(c) Produce a scatterplot of lhwage vs. experience, highlighting the points which represent data for women.

A2. Run a regression of lhwage on a constant and fem.

A3. To control for the education of the worker, and to find out the effect of education on earnings, run a regression of lhwage on: a constant, fem, and yrseduc.

Susan Dynarski
API 202B

February 6, 1994
Assignment 1

17.5/20

A1

a)

HWAGE	*N*	*Mean*	*Standard Deviation*
Men	494	9.87232	4.75637
Women	307	7.03894	3.25414

LHWAGE	*N*	*Mean*	*Standard Deviation*
Men	494	2.16907	0.51378
Women	307	1.86219	0.41510

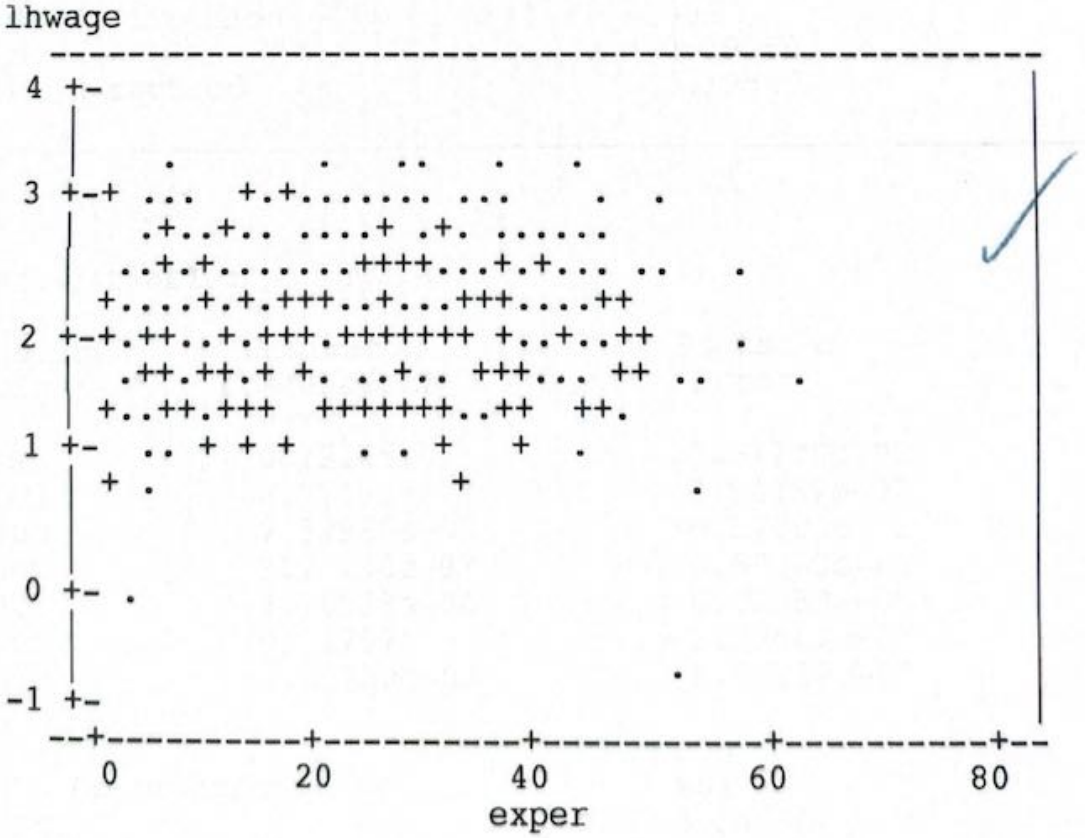
The gender gap, in logs, is 0.30688. (2.16907 − 1.86219)

b) We can test the hypothesis that log hourly earnings are the same for women and men using a difference in means test.

$$z = \text{xbar}_W - \text{xbar}_m \;/\; \sqrt{[(s_W^2 / n_W) + (s_m^2 / n_m)]} = -9.27$$

We reject H_0 that log hourly earnings for men and women are the same, using a two-tailed test, at a < 1% level of significance.

c)



A2

Dependent Variable: lhwage

Independent Variable	*Estimated Coefficient*	*Standard Error*	*t-Statistic*
const	2.16907	2.15243e-02	1.00773e+02
fem	-0.30688	3.47677e-02	-8.82651

Number of Observations	801
R-squared	8.88433e-02
Corrected R-squared	8.77029e-02

A3

Dependent Variable: lhwage

Independent Variable	*Estimated Coefficient*	*Standard Error*	*t-Statistic*
const	1.40819	7.59378e-02	18.54392
fem	-0.30050	3.26550e-02	-9.20213
yrseduc	5.83751e-02	5.61578e-03	10.39483

Number of Observations	801
R-squared	0.19750
Corrected R-squared	0.19549

A4

Dependent Variable: lhwage

Independent Variable	*Estimated Coefficient*	*Standard Error*	*t-Statistic*
const	0.79146	8.61770e-02	9.18408
fem	-0.21496	3.84467e-02	-5.59120
yrseduc	7.59980e-02	5.57092e-03	13.64190
exper	3.71138e-02	3.82663e-03	9.69881
exper2	-5.70538e-04	8.07353e-05	-7.06678
union	0.11707	3.25449e-02	3.59708
pcfemale	-1.86950e-03	5.96212e-04	-3.13563

Number of Observations	801
R-squared	0.34390
Corrected R-squared	0.33895