DESCRIPTIVE STATISTICS

Before undertaking a complex analysis of the determinants of wages and investigating whether race or gender inequities exist, one needs to get a sense of what the data are saying. In particular, one needs to inspect the distribution of the primary response or dependent variable, hourly wages, and the explanatory variables, such as education level, gender, and occupational status, to learn about the characteristics of the population being studied. Typically, one looks at such features of the distribution as the mean, median, standard deviation, range and percentiles, in order to assess where the average or typical values lie and how spread out they are, in addition to whether the distribution is skewed and whether there are atypical values.

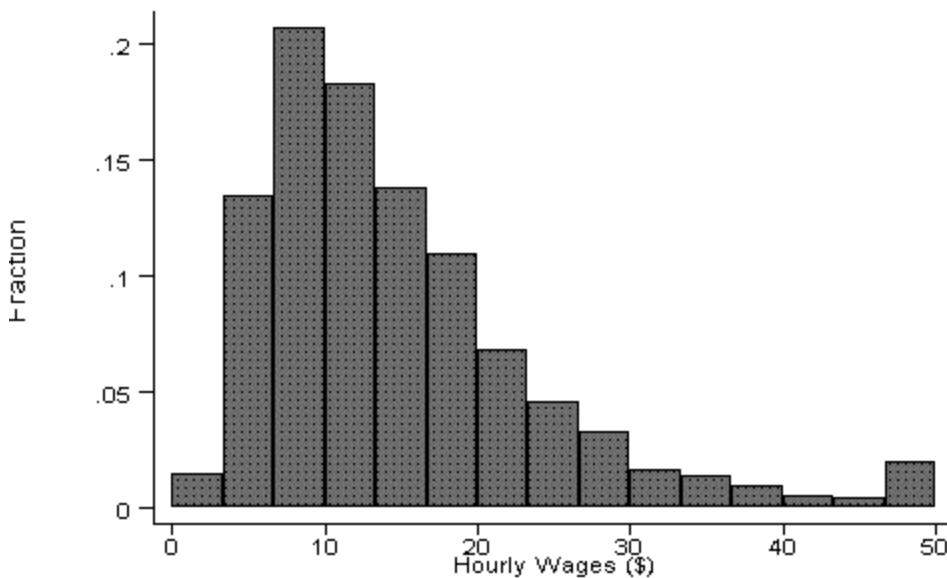<u>Description of Hourly Wages</u>

Figure 2.1. Histogram of hourly wages, in dollars, for this population.

The summary descriptive statistics in Table 2.1 indicate that average (mean) hourly wages are \$14.77 and median (50th percentile) wages are \$12.50. This means that half the people earn less than \$12.50 per hour and half earn more. But mean wages are significantly higher than median wages and most people are earning less than the average. In fact, this is a characteristic of right skewed distributions and Figure 2.1 clearly shows this pattern. The "blip" at \$48 per hour is due to "top coding" of wages to protect the confidentiality of the highest wage earners; most of this group earn more than \$48 per hour.

The standard deviation of hourly wages is \$9.27, the range is \$48.08 – \$1.92 = \$46.16 and the interquartile range (75th percentile – 25th percentile) is \$19.23 – \$8.17 = \$11.06. Variability in wages are high: although median wages are \$12.50, the central half of the sample earns between \$8.17 and \$19.23, the interquartile range; the lowest quarter of the sample earns less than \$8.17 and the top quarter, more than \$19.23. By all accounts, wages vary considerably from person to person. The aim of the analysis will be to study personal and job-related factors that determine this variation and how much of the variation can be explained by these factors.

Using the logarithm of wages produces a more symmetric, bell-shaped distribution, with mean (2.51) and median (2.53) closer together. While the scale of this transformed variable is not immediately interpretable (would you like to be paid in "log dollars"?), it is often the preferred scale for analysis since many estimators have statistical properties that apply well to bell-shaped, symmetric distributions, and not as well to skew distributions. Note that this distribution is now slightly left skew, possibly because of truncation of the highest wages.
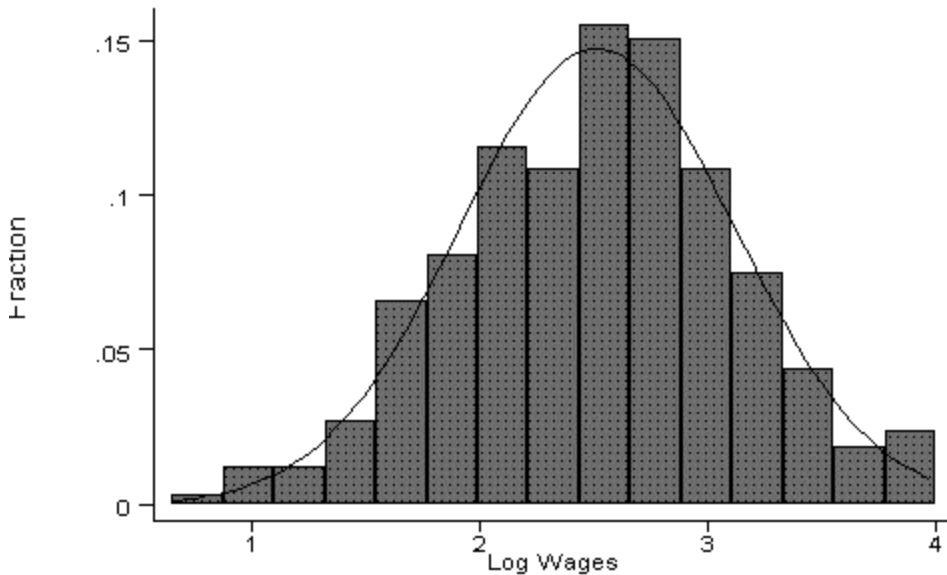

Figure 2.2. Histogram of the logarithm of hourly wages for this population.

Description of the Sample Population

Table 3.2 gives descriptive statistics for other characteristics of the population that are related to their wage earning potential. The sample consists of people between 18 and 68 years of age, with an average age of 40 years, having between 0 and 50 years of work experience, and between 2.5 and 20 years of education, with an average of 13.5 years. They are 46% female, 10% black and 6% Hispanic, 45% have a high school diploma or less and 10% have post-graduate degrees, 22% work in manufacturing, 28% in blue collar occupations, 35% in professional or management positions, 63% are married, 18% are union members, and they are uniformly distributed across the regions of the U.S. All in all, a typical cross-section of the American working population!

Table 2.2 also reports the mean hourly wages, classified by subgroups of each of the explanatory variables. For example, males earn a mean hourly wage of $17.05 and females, a mean hourly wage of $12.14, 29% less than men. Whites ($15.50) earn more than either blacks ($12.47) or Hispanics ($9.83), on average. These observations will be the basis of an extensive evaluation of gender and racial bias to see if there are other factors that explain this discrepancy.

Hourly wages increase dramatically with years of education, especially for those with post-graduate education ($\geq$ 16 years of schooling). Wages generally increase with age and with work experience, but the increase tapers off after age 40 and after 10 years of experience. Union members earn slightly more than non-union, married people earn more than either divorced or unmarried. Wages vary by industry and occupation, with management and professional positions having the highest earning power, jobs in manufacturing paying the highest, and jobs in retail paying the least. Wages do not vary by region across most of the country, but they are higher in the northeast. Some of these patterns of variability were expected, but the magnitude of variability across gender and racial categories were perhaps unexpected.

Finally, wages are moderately correlated with years of education (r=0.41), and age and years of experience are almost perfectly correlated (r=0.97) (Table 3.3). This latter relationship is not surprising since years of work experience was not asked in the survey, but since people start school at age 6, it was derived as,
years of experience = (current age – 6) – years of education.

The relationship between wages and education is more clearly demonstrated using a plot than a table. Figure 2.3 is a plot of hourly wages against education, superimposing mean wages at each education level. The relationship between wages and years of education would not be discernible from the plot of observed values alone. However, focusing on mean wages filters out the random variability within each education level and clearly shows a trend of increasing hourly wages with education level, especially beyond high school. In fact, this is what linear regression analysis attempts to do: fit a straight line to the curve of mean wages against education.
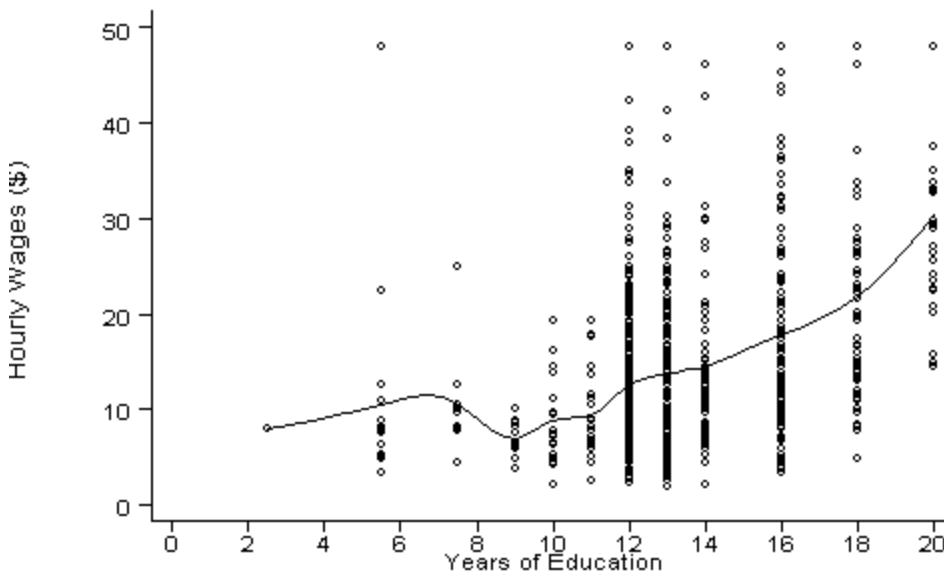
Figure 2.3. Plot of hourly wages and mean hourly wages against years of education.

Figure 2.3 also shows that wages are not symmetrically distributed within each education level. You can see this more clearly if you tilt the plot sideways! The distributions tend to be right skew, so that a few people have very high wages but most are lower than the average for their education level.

This preliminary descriptive analysis gives a good sense about the univariate relationships between hourly wages and the factors in the survey that cause or contribute to variations in income. Chapter 4 will demonstrate how linear regression is performed and how log transforming the data takes care of the asymmetry in wages to produce a valid estimate of the relationship between wages and these factors.

Table 2.1 Descriptive Statistics for Hourly Wages

Hourly Wages (Dollars)

       Percentiles   Smallest
 1%   2.884615   1.923077
 5%   4.625   2.076923
10%   5.769231   2.115385   Obs   1003
25%   8.173077   2.403846   Sum of Wgt.   1003

50%   12.5   Mean   14.7697
              Largest   Std. Dev.   9.257249
75%   19.23077   48.07644
90%   26.92308   48.07644   Variance   85.69666
95%   33.65385   48.07644   Skewness   1.514996
99%   48.07644   48.07644   Kurtosis   5.599439

Table 2.2 Mean Hourly Wages According to Categories of Demographic and Labor Factors in the 1995 Current Population Survey.

Hourly Wages (dollars)

	Mean Wages	Std. Dev.	N	Percent
GENERER				

GENDER
------------+--
    Male | 17.04845   10.240742   537   53.54
  Female | 12.14377   7.132306   466   46.46

RACE
------------+--
   White | 15.508646   9.5208053   788   78.56
   Black | 12.467623   7.5165665   102   10.17
 Hispanic | 9.832409   6.9178947   61   6.08
   Other | 13.879308   8.2283586   52   5.18

EDUCATION LEVEL ACHIEVED
------------+--
 0-12 yrs | 9.310918   6.038696   86   8.57
 HS Diplm | 12.522641   6.970573   362   36.09
   Assoc | 13.961516   8.086241   280   27.92
 BA/AB/BS | 17.896819   10.012366   175   17.45
 Masters | 21.820186   11.431568   70   6.98
     MD | 32.114526   13.480527   14   1.40
    PHD | 27.993269   7.697482   16   1.60

YEARS OF EDUCATION
------------+--
 Educ<=12 | 9.310918   6.038696   86   8.57
 Educ=12 | 12.522641   6.970573   362   36.09
 Educ=13 | 13.730448   8.074917   202   20.14
13<Educ<=16 | 16.868053   9.582990   253   25.22

Educ>16	24.389087	11.893388	100	9.97

AGE

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

Age<=25	8.582077	4.118150	100	9.97
25<Age<=30	11.964276	5.768639	150	14.96
30<Age<=40	15.150492	10.072461	295	29.41
40<Age<=50	17.374961	9.283117	268	26.72
Age>50	15.975147	10.091382	190	18.94

YEARS OF WORK EXPERIENCE

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

Exper<=5	10.274019	5.5195622	97	9.67
5<Exper<=10	12.073586	7.2312446	135	13.46
10<Expr<=20	15.587707	10.452645	304	30.31
20<Expr<=30	16.724456	8.8239055	268	26.72
Exper>30	14.907941	9.4999288	199	19.84

INDUSTRY

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

Constr	13.856577	8.243378	53	5.28
Manufg	16.661392	10.541315	216	21.54
Transp/Comm	17.310800	8.651183	95	9.47
Retail	10.390213	6.726743	137	13.66
Finance	15.479026	10.520977	75	7.48
Medical	14.941064	9.666591	97	9.67
Educat	14.641352	6.792929	80	7.98
Publ Admin	16.647628	8.767678	60	5.98
Other	13.854691	9.038844	190	18.94

OCCUPATION

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

Managemt	20.567792	11.695310	153	15.25
Profess	18.814216	9.879802	197	19.64
Sales	14.144551	9.611419	111	11.07
Clerical	10.711768	4.651472	170	16.95
Service	10.409413	6.515944	95	9.47
Bluecollar	12.927072	7.168035	277	27.62

MARITAL STATUS

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

Nevermarr	11.651915	7.525551	185	18.44
Wid/Divor	13.747714	8.128243	184	18.34
Married	15.976069	9.771783	634	63.21

UNION MEMBER

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

No	14.581807	9.651701	824	82.15
Yes	15.634650	7.125761	179	17.85

REGION OF RESIDENCE

```
------------+-------------------------------------------
 Northeast |  16.475080  10.044439      247      24.63
  Midwest |  14.092022   8.120782      244      24.33
   South |  14.066194   9.110295      317 31.61
    West |  14.601179   9.579763      195 19.44
```

Table 2.3 Relationships Among Hourly Wages and the Continuous Variables, Age, Work Experience and Years of Education.

```
              |  Mean   Std. Dev.   N    Min    Max
-------------------------------------------------------------
  Age  | 39.59721  10.98771   1003   18    68
Experience    | 20.14506  11.16894   1003    0     50
Education    | 13.45214   2.60904   1003   2.5    20
```

corr wage age exper educ
(obs=1003)

```
      |  wage    age    exper    educ
--------+-----------------------------------
  wage|  1.0000
   age|  0.2385   1.0000
 exper|  0.1381   0.9724   1.0000
  educ|  0.4133   0.0487  -0.1857   1.0000
```