İnsan TUNALI, Kuzey YILMAZ October 7, 2009
Econ 311: Econometrics I

ANSWERS to ASSIGNMENT 1

Part I.

1. Sample cmf of price (column 3) given age (column 2):

Subsample	Age	Mean price	No. of obs.
1	**1**	**10,894.0**	2
2	**2**	**13,800.0**	3
3	**3**	**9,900.0**	2
4	**4**	**7,000.0**	1
5	**5**	**7,883.0**	3
6	**6**	**5,739.0**	5
7	**7**	**4,950.0**	1
8	**8**	**3,544.0**	2
9	**9**	**1,800.0**	1
10	**18**	**4,500.0**	2
11	**23**	**4,500.0**	1
12	**25**	**5,216.7**	3
13	**26**	**8,250.0**	1

2. Summary statistics (see Excel document):

	Price	Age
Mean	**7,169.3**	**9.4**
Variance	**12,834,325.9**	**69.7**
Std. dev.	**3,582.5**	**8.3**

Cov(price,age)	Corr(price,age)
-13,590.9	**-0.5**

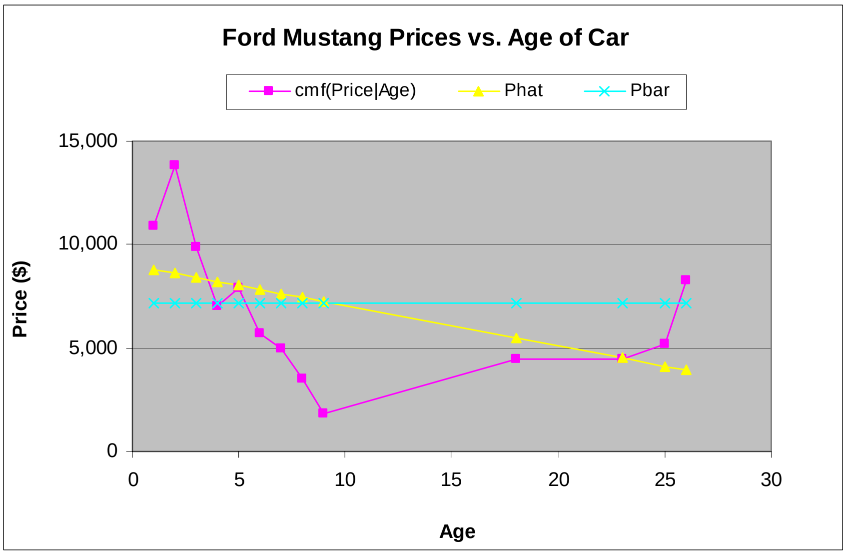
3. LS linear regression of price (Y) on age (X):

$$b = S_{XY}/S_{XX} = -13{,}590.9/69.7 = -195.1$$

$$a = \bar{Y} - b\bar{X} = 7{,}169.3-(-195.119)\ 9.4 = 9012.1.$$

4. (i) See 1; (ii) $\bar{Y} =$ 7,169.3; (iii) $\hat{Y} = 9012.1 - 195.1\ X$ (or Phat = 9012.1 −195.1*Age).

The graphs are on the next page.



5. It makes sense to use the sample cmf of price given age.

Sample mean of price is a constant (it does not capture how price changes with age).
The regression line does, but it seems to misrepresent how the price changes with age. In particular it misses the sharp decline in price as a function of age early on, and the higher prices of really old cars.
The sample cmf appears to do a more accurate job of showing how price depreciates, and how valuable really old Mustangs (which became classics) are valuable.

Part II. Problems from Goldberger

1. G1.1. Why features of a distribution other than the sample mean might be more relevant in summarizing the information in the subsamples (different levels of schooling):

Median earnings: Mean is influenced by outliers (extreme observations), median is not. Thus median is a more robust measure of central tendency.

Chances of being a low earner (= probability earning less than $5 an hour): Not everyone can earn the mean or the median wage -- some earn less, some more (in fact some earn much more). For example, new graduates typically start with low wages. If one's objective is to learn about the low end of the distribution, it would make sense to focus on the probability of falling below that threshold. That's why we examine percentiles of a distribution.

2. To learn how wages of men differed from the wages of women, one could calculate two sample cmf's, one for men, a second one for women. That is, we could do the work in Table 1.1 separately for men and women.

3. G2.9. Let $Z = 1$ if $Y < 5$, $Z = 0$ otherwise. Thus $P(Z = 1| X) = P(Y < 5| X)$. This probability can be approximated as the proportion of observations for which $Y < 5$, for each value of X. Armed with this information, the individual could compare the cost of an additional year of education with the benefit in terms of a reduction in the probability of being a low earner.