Kyle Nabors
Dr. Ura
ECN 140
10th April 2022
Homework 1

1. Textbook
    a. To test the effect of training, I would break the factories into two random groups. One group that would receive the training and one that would not. This is to get a large enough sample size to see the effect of training and allow us to randomly assign the training to make sure other changes in the factory were not affecting the results.
    b. An observational cross-sectional data set to study the effect of training would involve two random groups of factories. One of the groups would receive the training and one would not. Then at the end of the training measure the efficiency of all groups and see if the ones that received the training are more efficient in output per worker per hour.
    c. A time series data set that would measure the effect would be collecting data on factories performance would be to take the data from a factory before the training, give the training, and then measure again to see if there were any changes in output per worker per hour.
    d. An observational data set for panel data would be to do a mix of both the cross-sectional and time series tests. The test would involve two random groups of factories where one is given the training and one is not. Then you would measure the efficiency both before and after giving the training to see the change in output per worker per hour over time of both the treated and non-treated group.
2. Stata
    a. The following answers are based on the WAGE1 dataset provided by Dr. Ura which contains 526 observations.
    b. Using the summarize command, we calculated the mean and standard deviation for hourly wage (wage) and years of education (educ). For wage, the mean was 5.90 and the standard deviation was 3.69. For education, the mean was 12.56 and standard deviation was 2.77.
    c. Using the hist command we drew histograms for wage and educ which are shown in figures 1 and 2 respectively.
    d. Using the scatter command, we drew a scatter plot with wage as the horizontal axis and educ as the vertical axis which can be seen in figure 3.
    e. Stringing together the mean command and an if statement we can calculate the mean of wage if education if less than or equal to 12. This calculation contains 314 observations and is 4.89 with a standard error of .16.
    f. Using the ttest command we can test whether the mean of educ is equal to 11. Based on our results we reject the null hypothesis that the mean is equal to 11 with 95% confidence.
    g. Using the log function, we can calculate the natural of wage. Then we can subtract this from lwage to find the difference. Using the summarize command we can then find the mean and standard deviation of this variable witch are both 0. This is because both variables are identical as they are both the natural log of wage.
3. Stata esttab

    a. Using estabb commands we can generate tables based on datasets. Using this command, we can generate a table of mpg of different cars seen in figure 4.
    b. We can also set specific sizes and variables to make custom tables. Using this we can make a table showing the mean and standard deviation of price, car origin, and mpg as shown in figure 5.
4. Various Distributions
    a. Using the set obs function we can simulate observations and use them for data analysis. Using this command, we can generate 100,000 observations and use them to calculate distributions.
    b. Using the rnormal command we can generate a normal distribution of the generated data. Using the hist command we can then draw a histogram of the distribution which can be seen in figure 6.
    c. Using the rchi2 function we can generate a chi^2 distribution of the generated data wit with 5 degrees of freedom. Using the hist command we can then draw a histogram of the distribution which can be seen in figure 7.
    d. Using the rt command we can generate a t distribution from the generated data with 3 and 100 degrees of freedom. Using the hist command we can then draw a histogram of the distribution witch can be seen in figure 8.

5)

	Y=0	Y=1	
X=0	8/24	4/24	12/24
X=1	6/24	6/24	12/24
	14/24	10/24	

13. $P(X=1 \mid Y=1) = 6/24$

14. $P(Y=0) = 58\%$
$P(Y=1) = 42\%$

15. $E(Y) = 0 \cdot .58 + 1 \cdot .42 = .42$

16. $Var(Y) = E(Y^2) - E(Y)^2$
$= (0^2 \cdot .58 + 1^2 \cdot .42) - (.42^2)$
$= .42 - .1764$
$var(Y) = .2436$

17. $sd(Y) = \sqrt{var(Y)}$
$sd(X) = \sqrt{.2436}$

18. $corr(Y,X) = \frac{cov(Y,X)}{sd(Y) \cdot sd(X)}$
$E(X) = 0 \cdot .5 + 1 \cdot .5 = .5$
$E(YX) = 0 \cdot 18/24 + 1 \cdot 6/24 = 6/24 = .25$
$cov(YX) = .25 - (.42 \cdot .5)$
$= .25 - .21 = .04$
$sd(X) = \sqrt{var(X)}$
$var(X) = E(X^2) - E(X)^2$
$E(X^2) = 0^2 \cdot .5 + 1^2 \cdot .5 = .5$
$E(X)^2 = (0 \cdot .5 + 1 \cdot .5)^2 = .25$

$$Var(x) = .5 - .25 = .25$$

$$Sd(x) = \sqrt{.25}$$

$$Corr(Y,X) = \frac{.04}{\sqrt{.2436} \cdot \sqrt{.25}}$$

19 No, since $Corr(Y,X) \neq 0$ X and Y are on some level dependent.

20. $E(Y|X=1) = 0 \cdot 6/24 + 1 \cdot 6/24 = .25$

Appendix

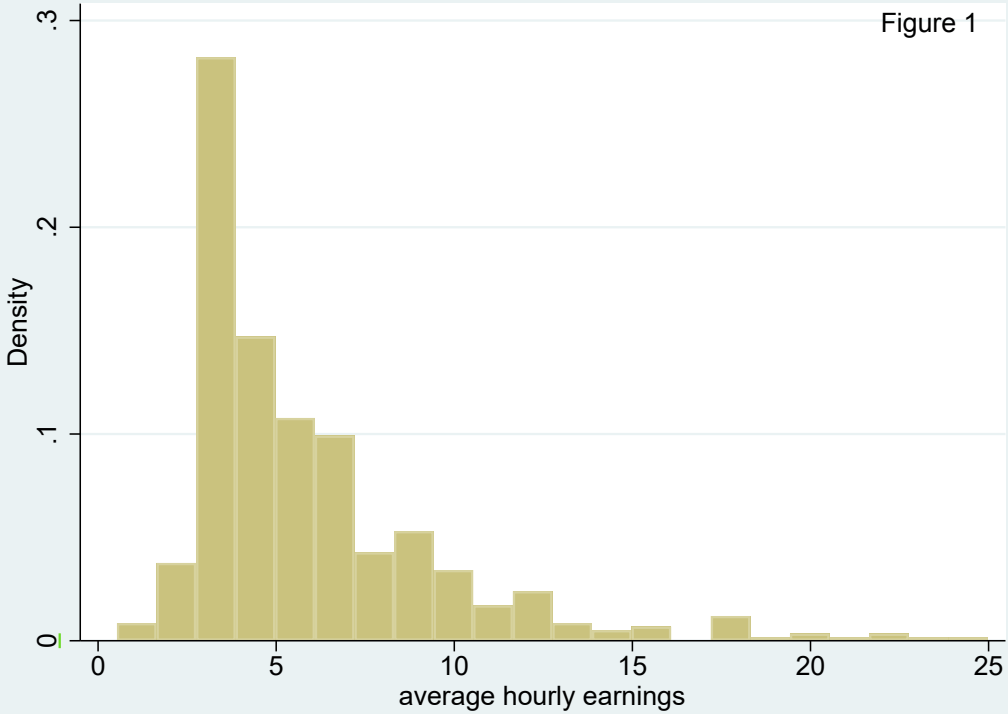
Figure 1
Density
0
.1
.2
.3
0
5
10
15
20
25
average hourly earnings

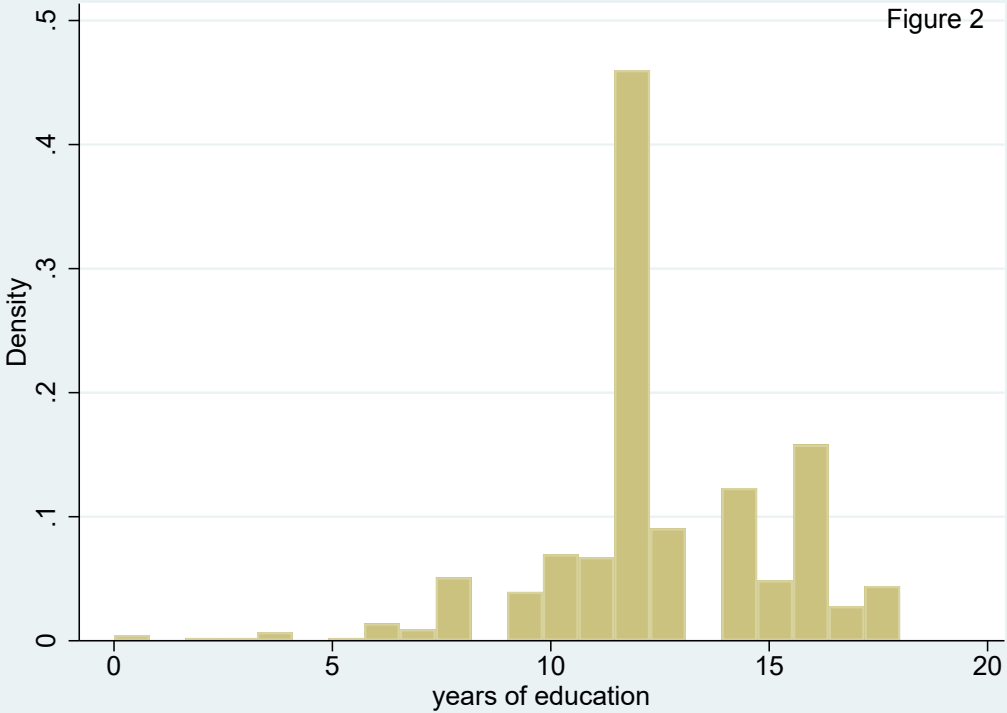
Figure 2
Density
0
.1
.2
.3
.4
.5
0
5
10
15
20
years of education

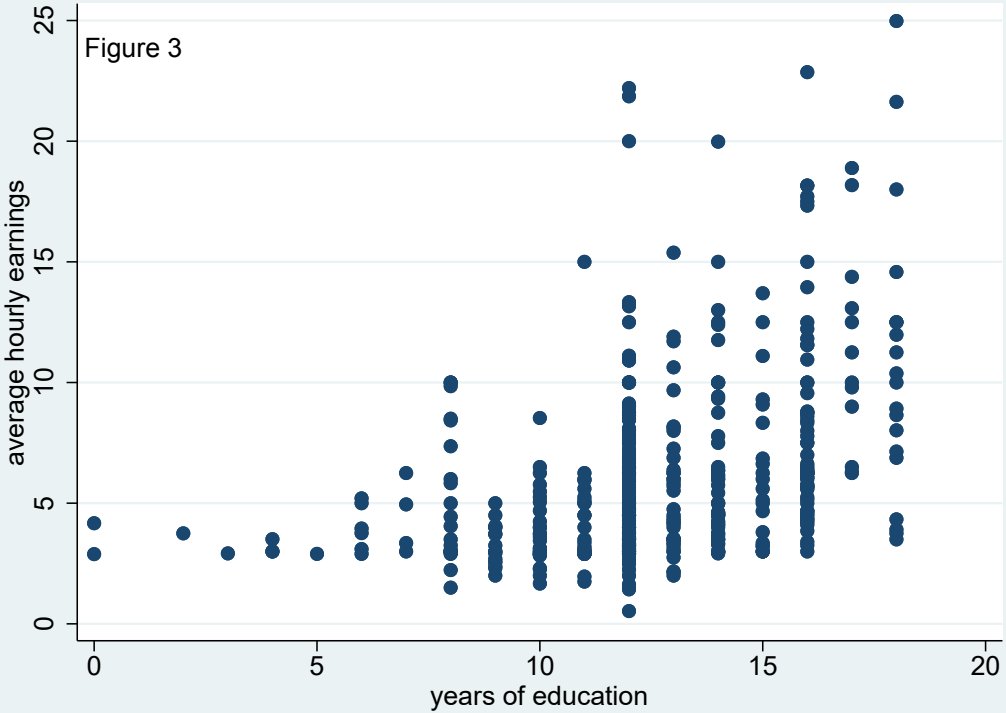
Figure 3
average hourly earnings
0
5
10
15
20
25
0
5
10
15
20
years of education

Figure 4

	(1) mpg	(2) mpg	(3) mpg
foreign	4.946**	-1.650	-2.246+
	(0.001)	(0.130)	(0.074)
weight		-0.00659**	-0.00675**
		(0.000)	(0.000)
displacement			0.00825
			(0.472)
gear_ratio			2.058
			(0.245)
_cons	19.83**	41.68**	34.52**
	(0.000)	(0.000)	(0.000)
N	74	74	74

p-values in parentheses
+ p<0.1, * p<0.05, ** p<0.01

Figure 5

	Mean	Standard Deviation
Price	6165.257	(2949.496)
Car origin	.2972973	(.4601885)
Mileage (mpg)	21.2973	(5.785503)
Observations	74	

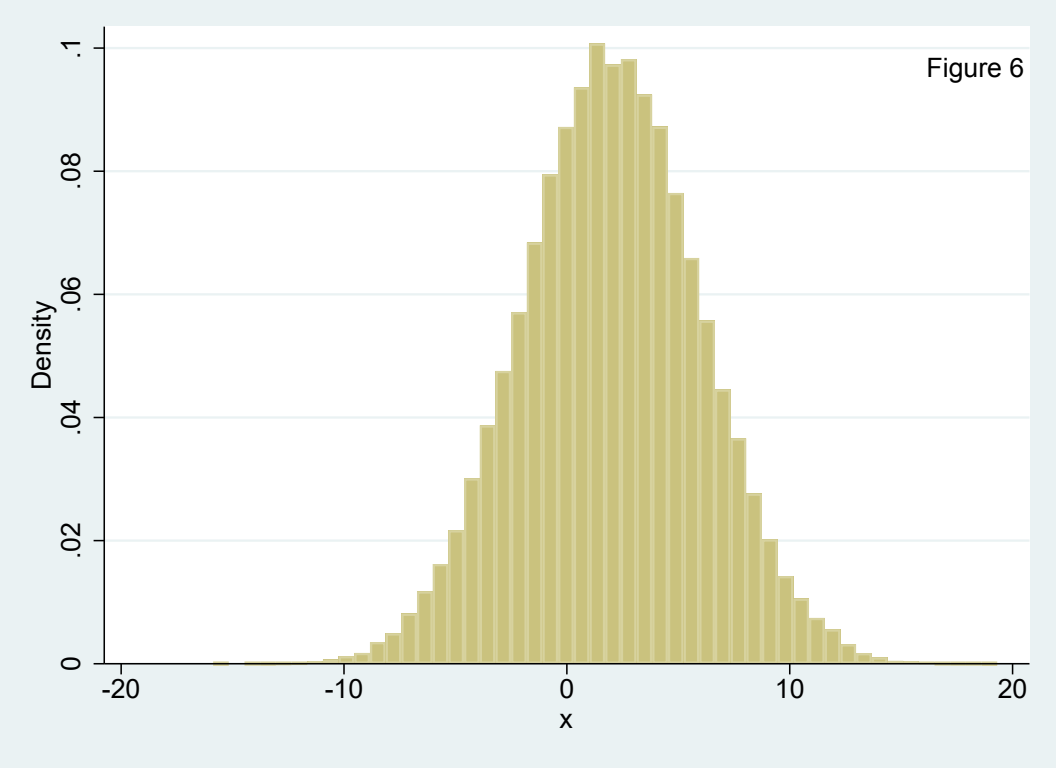
Figure 6
Density
0
.02
.04
.06
.08
.1
-20
-10
0
10
20
x

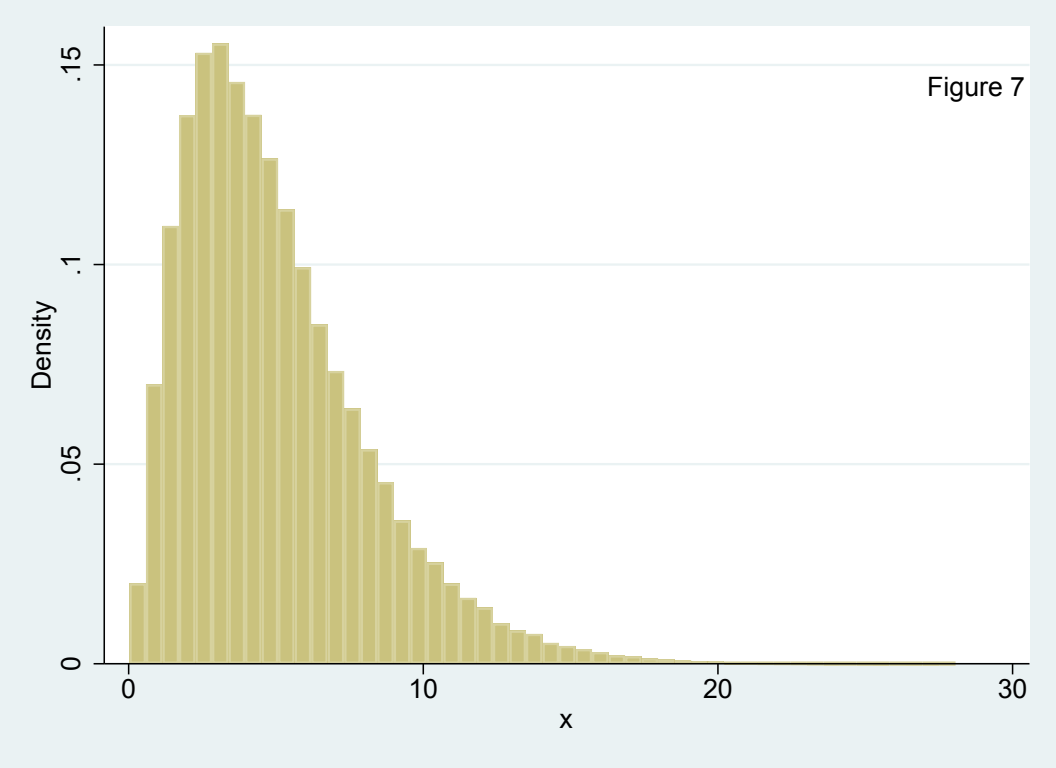
Figure 7
Density
0
.05
.1
.15
0
10
20
30
x

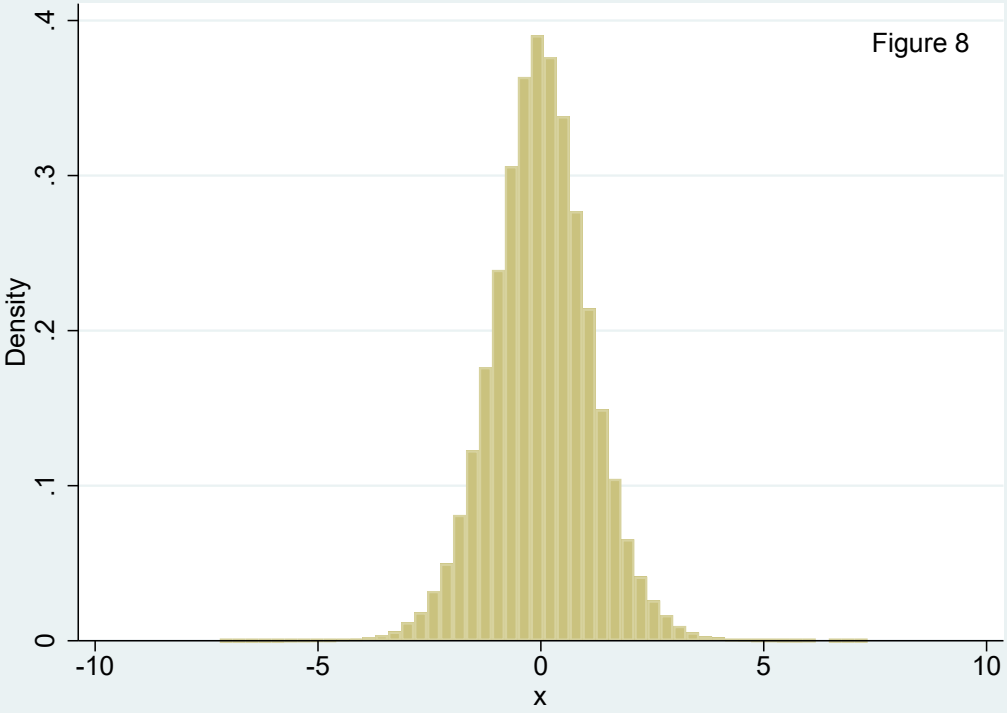
Figure 8
Density
0
.1
.2
.3
.4
-10
-5
0
5
10
x