14.03/14.003 Problem Set 3
Fall 2014
SOLUTIONS

Professor David Autor

Due Thursday, 10/23/2014 at 5pm EST v

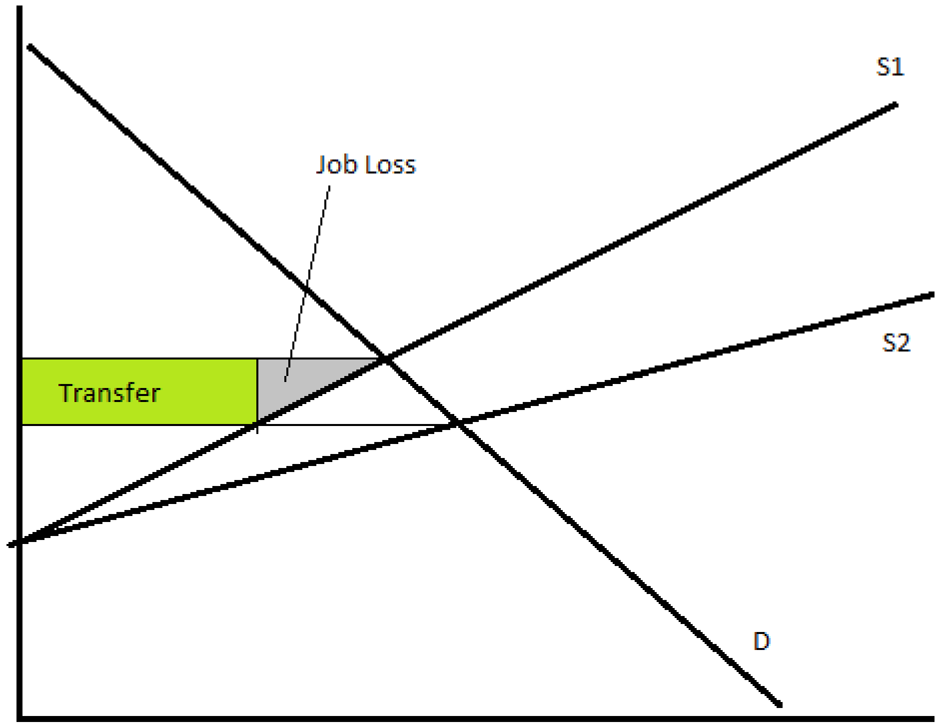
1 Subsidies and Incidence (30 points)

1. Consider the labor market for high skill workers characterized by the following supply and demand:

   $S(w) = -50 + 5w$

   $D(w) = 1000 - 10w$

   (a) Calculate the equilibrium wage in this market. Draw the supply and demand curves on a graph, and calculate worker and employer surplus. [**4 points**]
   In equilibrium, $S(w) = D(w)$. Solving the equation gives $w = 70$. At these wages, quantity is $S(70) = -50 + 5(70) = 300$. Worker surplus is all of the area above the supply curve below w=70. This is a triangle with height 60 and width 300, area 9000. $WS = 9000$. Employer surplus is the area between the demand curve and the w=70 line. This is a triangle with height 30 and width 300, area 4500. $ES = 4500$



(b) There is another group of workers who may be high skilled, but have not been tested. The government institutes a free testing service which verifies workers as being skilled. The result is that a second, identical group of high skilled workers join the workforce, doubling the supply of high skilled labor. What is the new supply equation that the employer faces? Draw the new supply curve on your previous graph (and verify that at each level of wage the supply of labor doubled). What is the new equilibrium wage? [**4 points**]

The new supply equation is $S(w) = -100 + 10w$. **When** $S(w) = D(w)$, $w = 55$ **and quantity is 450. As before, employer surplus is equal to the area below demand and above w.** $ES = 10125$, **an increase. This is expected because supply increased, and as a a result quantity increased and wage decreased. Worker surplus for each population is derived from their individual supply functions (**$S'(w) = -50+5w$**). Using these supply functions, we can find the area above supply curve and below** $w = 55$. **For each group, the area is a triangle with height and width of** 45×225 **and area** 5062.5. **Total worker surplus increased but the original population is worse off.**

(c) At the new equilibrium, calculate the employer's surplus, the surplus for the original group of high skilled workers and the surplus for the new group of high skilled workers. How would each group be willing to individually pay to implement this government testing service? [**4 points**]

Each group is willing to pay an amount equal to their change in surplus. The employer would pay $10125-4500 = 5625$ **for this. The original employees are worse off and would not pay. The new workers would pay up to** 5062.5.

(d) The original group of workers rightfully views this new development as a wage decrease for themselves. Associated with this wage decrease is a transfer of wealth to the employer (since each working employee is now getting paid less) as well as a loss of welfare from workers who were productive but no longer have a job. On your graph label the areas associated with the transfer and job loss. What is the sum of the two areas? [**3 points**]

The transfer is a loss of 15 for each worker still employed, an area of 3375. The job loss resulted in 75 workers losing their jobs, with the loss in welfare ranging from 0 to 15, with an area of 562.5. The sum of the areas is equal to 3937.5, their loss of utility.

(e) If the program costs K to implement, for what values of K does this service increase total welfare? If $K = 0$, what are the deadweight losses associated with this policy change? [**3 points**]
Total change is welfare for this program is 6749-K. For K less than 6749, the service increases welfare. For K=0, there is no deadweight loss because total welfare increases.

2. Return to our original market for labor

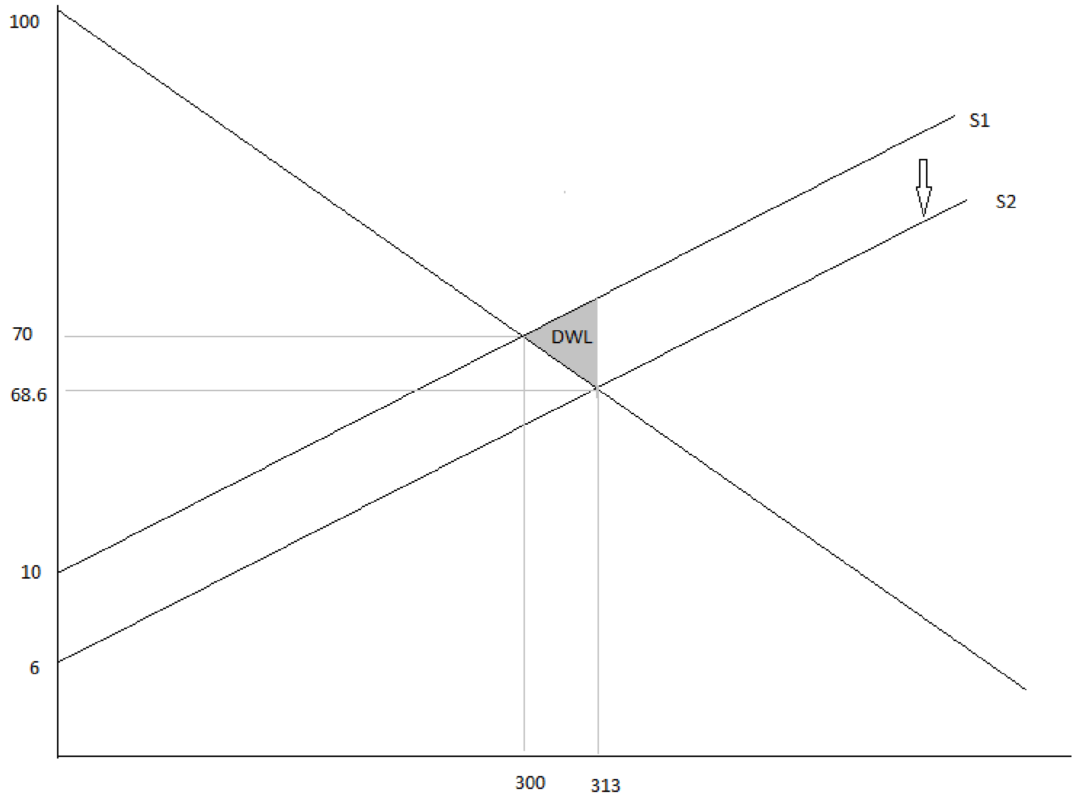
$S(w) = -50 + 5w$

$D(w) = 1000 - 10w$

Suppose that the government implements an earned income tax credit. Any worker in the market that is employed gets a government subsidy of 4, so our new labor supply is $S(w) = -50 + 5(w + 4)$.

(a) For subsidies, we are often interested in the incidence, which describes how much each party gains from the subsidy. To calculate this, find the the new equilibrium wage. What is the difference in the amount the worker gets paid? What is the difference in the amount the employer pays? The ratio of the difference for the worker and the employer is the incidence. This also the ratio of the elasticities of supply and demand. What is this value? [**4 points**]
The equilibrium is where the new $S(w) = D(w)$. At that point, $w = 68.6$, and quantity is equal to 313.4 employees. At these wages, employers pay 1.33 dollars less per employee. Employee gets paid $68.6 + 4$, or 2.66 more per person. This tells us that the incidence is 2. Supply elasticities are equal to $\frac{dQ/Q}{dP/P} = 5\frac{68.8}{313.4} = 1.09$. The demand elasticity is $-10\frac{68.8}{313.4} = 2.18$. The ratio of elasticities is 2, equal to the incidence.

i. Is there deadweight loss associated with this program? On a new graph draw the shift in supply and demand and label on the graph the deadweight loss associated with the subsidy, and calculate the change in total welfare and the cost of the subsidy to determine the deadweight loss. [**4 points**]
13.4 more employees are hired than efficient. The loss for these employees are getting paid more than they are producing, with the first extra employee costing 0 and the last one costing 4, with an average of 2. The total deadweight loss is26.8



(b) Imagine instead that the labor supply curve was $S(w) = -400 + 10w$. Recalculate the incidence as in part (a) and compare to your answer above. Explain intuitively why the incidence of the subsidy on equilibrium wages is higher or lower in this case. [**4 points**]
The equilibrium without the subsidy is $w = 70$. With the subsidy, $S(w) = -400+10(w+4)$, the new equilibrium at $S(w) = D(w)$ is $w = 68$. Employers pay 68 and employees get 72, so each group benefits by 2, exactly splitting the subsidy. Incidence is 1. Employees are much more elastic in this scenario, and as a result they benefit less from the subsidy.

2 General Equilibrium in an Exchange Economy (30 points)

Romeo is an MIT undergraduate who lives in East Campus. Juliet is a Wellesley student. Suppose Romeo has utility

$$u_r(R, C) = R \quad C^{1-}$$

where R are roses and C are chocolates. Juliet has utility

$$u_j(R, C) = R \quad C^{1-}$$

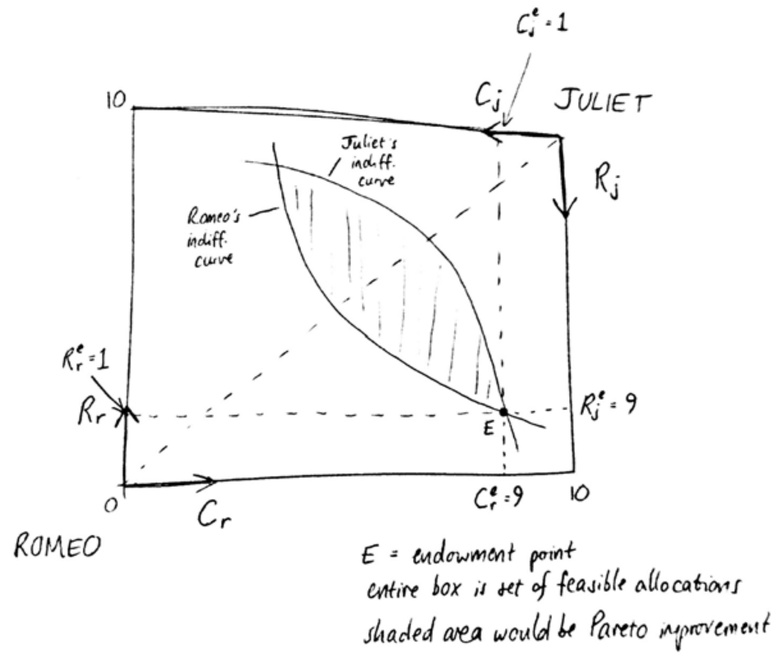
with $\in (0,1)$ and $\in (0,1)$. Initially Romeo and Juliet do not know each other, and so are unable to trade the roses and chocolates they own, only consuming their own endowment. Romeo begins with 1 rose and 9 chocolates ($\bar{R}_r = 1$, $\bar{C}_r = 9$), Juliet begins with 1 chocolate and 9 roses ($\bar{C}_j = 1$, $\bar{R}_j = 9$).

1. What are the initial utility levels of Romeo and Juliet? [**2 points**]

*[**Solution:** Romeo has $u_r = 1 \quad 9^{1-} \quad = 9^{1-}$ and Juliet has $u_j = 9 \quad 1^{1-} \quad = 9$]*

2. Suppose that Romeo bumps into Juliet at Toscanini's and after talking about life over ice cream, they decide to enter into a trading agreement. Draw an Edgeworth box and show clearly all feasible allocations of chocolates and roses for Romeo and Juliet. Draw Romeo and Juliet's indifference curves within the box [you can assume = = $\frac{1}{2}$ for this part if you would like, but the indifference curves need not be completely accurate] and mark the endowment allocation. Shade in the allocations which would be Pareto improvements over the endowment allocation. **[5 points]**

[***Solution:*** *See Figure:*



]

3. Suppose that Romeo and Juliet sell/buy roses at price p_R and sell/buy chocolates at price p_C. We can normalise and set $p_R = 1$. What relative price $\frac{p_C}{p_R} = p_C$ clears these mini-markets for roses and chocolates? [Note that chocolates and roses are divisible for the purposes of this question, that is, you can buy a fraction of a rose or take a small bite of a chocolate.] **[6 points]**

[***Solution:*** *For this question we want to find the relative price p_C such that total demand for roses equals total supply (it will follow automatically that this price also clears the market for chocolates – i.e. we only need to check that one market clears. This is due to something called 'Walras' Law'). We know total supply is just equal to the sum of Romeo and Juliet's endowments of roses. For total demand we find Marshallian demand of both people. Romeo maximises u_r subject to constraint*

$$p_R R + p_C C = p_R \bar{R}_r + p_C \bar{C}_r = p_R + 9p_C$$

where $\bar{R}_r$ is Romeo's endowment of roses and $\bar{C}_r$ is his endowment of chocolates. Since utility is of the Cobb-Douglas form, Romeo has demands (taking prices as given) such that fraction of income is *spent on roses and the remaining fraction* $1-\quad$ *on chocolates. I.e. we have*

$$p_R R_r^* = \quad (p_R + 9p_C) \Rightarrow R_r^* = \quad \frac{p_R + 9p_C}{p_R}$$

$$p_C C_r^* = (1- \quad)(p_R + 9p_C) \Rightarrow C_r^* = (1- \quad)\frac{p_R + 9p_C}{p_C}$$

and taking into account Juliet's different endowment, Juliet has demands

$$p_R R_j^* = \quad (9p_R + p_C) \Rightarrow R_j^* = \quad \frac{9p_R + p_C}{p_R}$$

$$p_C C_j^* = (1- \quad)(9p_R + p_C) \Rightarrow C_j^* = (1- \quad)\frac{9p_R + p_C}{p_C}$$

and we now have enough to find the relative price that clears the market. First note that total demand for roses is $R_r^* + R_j^*$ *and total supply is 10 (again, we could instead look at the market for chocolates, but we would find the same result). Setting supply equal demand we have*

$$\frac{p_R + 9p_C}{p_R} + \quad \frac{9p_R + p_C}{p_R} = 10$$

$$\Rightarrow \quad (p_R + 9p_C) + \quad (9p_R + p_C) = 10p_R$$

$$\Rightarrow (9 \quad + \quad) p_C + (9 \quad + \quad) p_R = 10p_R$$

$$\Rightarrow \frac{p_C}{p_R} = p_C = \left(\frac{10 - 9 \quad - }{9 \quad + }\right)$$

which pins down the relative price of chocolates to roses.]

4. Assume now that $\quad = \quad$. Suppose that Romeo and Juliet trade at the relative price p_C that you just found. Can you show that the utility of each person is higher than it was in part (1) (i.e. that we have a Pareto improvement, a.k.a. gains from trade)? **[7 points]**

*[**Solution:** Romeo's indirect utility is*

$$v_r = (R_r^*) \quad (C_r^*)^{1-} \quad = \left(\quad \frac{p_R + 9p_C}{p_R}\right) \quad \left((1- \quad)\frac{p_R + 9p_C}{p_C}\right)^{1-}$$

and using that $p_R = 1$ *we have*

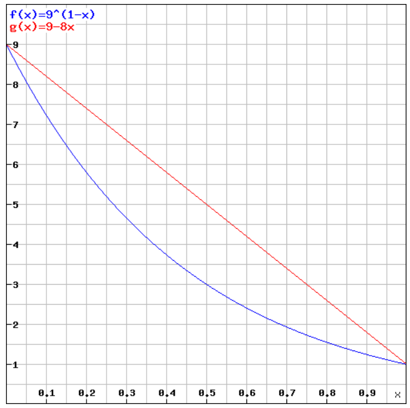
$$v_r = \quad (1- \quad)^{1-} \quad (1 + 9p_C)(p_C) \quad ^{-1}$$

and substituting in for p_C *and using* $\quad = \quad$ *we get*

$$v_r = \quad (1- \quad)^{1-} \quad \left(1 + 9\left(\frac{1-}{\quad}\right)\right)\left(\frac{1-}{\quad}\right) \quad ^{-1}$$

$$= \quad \left(1 + 9\left(\frac{1-}{\quad}\right)\right) = 9 - 8$$

which is always greater than Romeo's pre-trade utility 9^{1-} *for* $\quad \in (0,1)$*. See this graph to convince yourself of this:*



Now we check Juliet's indirect utility:

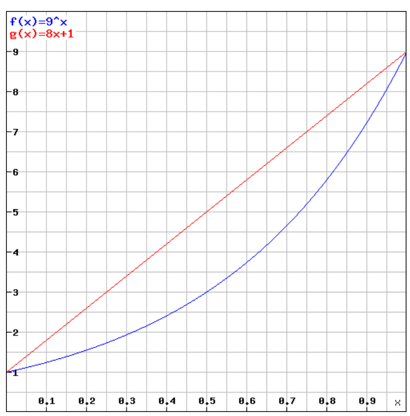
$$v_j = \left(R_j^*\right) \left(C_j^*\right)^{1-} = \left(\frac{9p_R + p_C}{p_R} \right) \left((1- \ ) \frac{9p_R + p_C}{p_C} \right)^{1-}$$

and simplifying as before, including that $\ = \ $*, we get*

$$v_j = \quad (1- \ )^{1-} \left(9 + \frac{1-}{\ } \right) \left(\frac{1-}{\ } \right)^{\ -1}$$

$$= \left(9 + \frac{1-}{\ } \right) = 8 \ +1$$

which is always greater than Juliet's pre-trade utility 9 *for* $\ \in (0,1)$*. Again, the graph makes this clear:*



This shows that trade yields a Pareto improvement. Beyond this, is shows that Toscanini's may serve a useful market matching role.]

5. Still assuming that $\ = \ $, what happens to the relative price of chocolates p_C and equilibrium consumption levels as $\ $ increases (i.e. as preferences of both shift toward roses)? Can you give any intuition for these effects? [**5 points**]

*[**Solution:** we have that $p_C = \frac{1-}{\ }$ so it is straightforward to see that as increases the relative price of chocolate falls (equivalently the relative price of roses increases). This is because at the current relative price there would be excess demand for roses – the demand from both Romeo and Juliet increases as increases, but the total supply stays fixed at the initial endowment. To clear the market, the relative price of roses must rise.*

Now let's consider equilibrium consumption levels. Recall that

$$R_r^* = \frac{p_R + 9p_C}{p_R} = \left(1 + 9\left(\frac{1-}{\ }\right)\right) = 9 - 8$$

$$C_r^* = (1- \ )\frac{p_R + 9p_C}{p_C} = (1- \ )\frac{1 + 9\left(\frac{1-}{\ }\right)}{\frac{1-}{\ }} = 9 - 8$$

So Romeo has $R_r^* = C_r^*$ and both are falling in – as preferences shift towards roses, his equilibrium demand of both roses and chocolates falls. Intuitively this is because roses are in greater demand, but his endowment is mostly chocolates. In some sense the economic power has shifted toward Juliet. Now consider Juliet's consumption levels:

$$R_j^* = \frac{9p_R + p_C}{p_R} = \left(9 + \frac{1-}{\ }\right) = 8 \quad + 1$$

$$C_j^* = (1- \ )\frac{9p_R + p_C}{p_C} = (1- \ )\frac{9 + \left(\frac{1-}{\ }\right)}{\frac{1-}{\ }} = 8 \quad + 1$$

which shows as suspected that Juliet's equilibrium consumption levels are both increasing in .]

6. Now suppose we start with a scenario where $\neq$ and Romeo has 9 roses and 9 chocolates ($\bar{R}_r = 9$, $\bar{C}_r = 9$), Juliet has 1 rose and 1 chocolate ($\bar{R}_j = 1$, $\bar{C}_j = 1$). The endowments are very unequal. Are there still gains from trade? What does this say about the possibility of true love between the rich and poor? [**5 points**]

   *[**Solution:** even though the endowments are very unequal, there will still be gains from trade provided that the initial endowment is not on the contract curve (the set of Pareto efficient allocations). The contract curve is the set of allocations where the indifference curves of Romeo and Juliet are tangent (same MRS), i.e. where*

$$\frac{MU_R^r}{MU_C^r} = \frac{MU_R^j}{MU_C^j}$$

   so we can just check that this condition does not hold at the initial endowment. For Romeo we have

$$\frac{MU_R^r}{MU_C^r} = \frac{\left(\frac{C}{R}\right)^{1-}}{(1- \ )\left(\frac{R}{C}\right)} = \frac{\ }{1-}\frac{C}{R} = \frac{\ }{1-}$$

   substituting in that $C = R = 9$ at the initial endowment. For Juliet we have

$$\frac{MU_R^j}{MU_C^j} = \frac{\left(\frac{C}{R}\right)^{1-}}{(1- \ )\left(\frac{R}{C}\right)} = \frac{\ }{1-}\frac{C}{R} = \frac{\ }{1-}$$

   substituting in $C = R = 1$. Since $\neq$ we have $\frac{MU_R^r}{MU_C^r} \neq \frac{MU_R^j}{MU_C^j}$ – so the indifference curves are not tangent at the initial endowment. It follows that there are still gains from trade (Pareto-improving allocations) despite the very unequal endowment. This says that Romeo and Juliet would still want to trade goods even though one is very 'rich' and one is very 'poor'. But probably true love is not affected that much by being able to trade goods, so this result tells us little about that.]

3 Subsidizing or Cost sharing (40 points)

In this exercise, you will replicate some of the results from the paper "Free distribution or cost sharing: Evidence from a randomized Malaria prevention experiment" by Pascaline Dupas and Jessica Cohen. You might find it helpful to read the paper to familiarize yourself with the context.

For this part of the problem set, you will need to submit the stata code (the do-file) you used to generate the results. Please also include the relevant portion of the stata results in the content of your problem set solutions. (We don't want to read the do files line by line, but we do want to see them.)

In Kenya, Insecticide Treated Bednets (ITNs) are sold through clinics where pregnant women come for their routine prenatal check ups and vaccinations. Bednets are sold in the market at 50 Ksh per piece. The authors designed an experiment to find out how the demand of bednets change when price changes. The authors have 20 clinics which they randomly assigned into one of five groups:

- Group 1; 4 clinics: control group
- Group 2; 5 clinics: bednets to be provided free for anyone visiting the clinic
- Group 3; 5 clinics: bednets to be sold at 10 Ksh
- Group 4; 3 clinics: bednets to be sold at 20 Ksh
- Group 5; 3 clinics: bednets to be sold at 40 Ksh

For this problem, you will use four stata datasets: **baseline.dta**, **buypatient.dta**, **sales.dta**, and **salesMarket.dta**. baseline.dta contains clinic level data collected during baseline. buypatient.dta contains patient (all women visiting the clinic) level information collected after the program is implemented. sales.dta contains weekly sales data for the 6 weeks after the program is implemented. **salesMarket.dta contains data on bed net prices and bed net sales at 100 clinics across Kenya collected *prior to* the experiment (this is non-experimental, AKA observational, data).**

The relevant variables in each dataset are described below:

Dataset	Variables	Description
baseline.dta		
	`clinicid`	clinic ID
	`district`	District in which clinic is located
	`treatment`	=1 if clinic received any kind of subsidy, 0 otherwise
	`netprice`	Price at which bednets are sold (randomized), for the control group this variable is Ksh 50
	`group`	Group assignment (described as above)
	`anc_fee`	Fee charged by clinic for ANC services
buypatient.dta		
	`respid`	ID of patients in the clinic
	`clinicid`	clinic ID
	`district`	District in which clinic is located
	`treatment`	=1 if clinic received bednet distribution program
	`netprice`	Price at which bednets are sold (randomized)
	`group`	Group assignment
	`buy`	=1 if Purchased net during prenatal care visit
sales.dta		
	`clinicid`	Clinic ID
	`week`	Week since net distribution program start
	`weeklynetsales`	Total number of nets sold in the week
	`district`	District in which clinic is located
	`treatment`	=1 if clinic received bednet distribution program
	`netprice`	Price at which bednets are sold (randomized)
	`group`	Group assignment
salesMarket.dta		
	`clinicid`	clinicid
	`p`	Bed net price on day of observation
	`salesMarket`	Number of bed nets sold on day of observation

1. [**10 points**] In Table 1, Cohen and Dupas show that baseline characteristics are balanced even in their small sample of clinics. You will replicate results from row 3. For this part, load **baseline.dta** into stata

   (a) [**2 points**] Find the average fees charged by clinics for ANC (Ante-natal care) services for each of the five groups of clinics (the group information is given by variable `group`). Find the standard deviations as well.

Group	mean(anc_fee)	sd(anc_fee)
1	10	8.164966
2	12	8.3666
3	4	8.944272
4	13.33333	11.54701
5	10	10

(b) [**1 point**] Test whether fees charged by clinics the control group (Group 1) is different from fees charged by clinics that receive full subsidy (Group 2). What is the **t-statistic**? Is the difference statistically significant at the 95% significance level?

(Hint: you can use the `ttest` command in stata, but be sure to *restrict your data only to the two groups that you are comparing*)

[See SOLUTION to part c]

(c) [**3 points**] Repeat the previous exercise for the following groups: Group 1 vs Group 3; Group 1 vs Group 4; Group 1 vs Group 5. Are any of the differences statistically significant at the 95% significance level

As you can see from the table below, the differences are never significant at the 95% significance level. You can check this by looking at the p-value

	Diff	SE	t-stat	p-value
G1vG2	2	5.5549206	-.36004115	.72943431
G1vG3	-6	5.7817447	1.037749	.33389914
G1vG4	3.3333333	7.3786479	-.45175395	.67036758
G1vG5	0	6.8313005	0	1

(d) [**3 points**] A more succinct way to do this test is using a regression. Generate indicator variables for each of the groups (for example, `GR1` should be 1 if the clinic is in group 1 and 0 otherwise). Regress fees variable against the four indicator variables (omit `GR1`, the control group). Are any of the coefficients significantly different from zero at 95% significance?

The results are very similar to part c). None of the coefficients are significant at 95% significance level. The p-values from these approaches are quite similar.

Source	SS	df	MS
Model	228.333333	4	57.0833333
Residual	1266.66667	15	84.4444444
Total	1495	19	78.6842105

Number of obs	= 20
F(4, 15)	= 0.68
Prob > F	= 0.6190
R-squared	= 0.1527
Adj R-squared	= -0.0732
Root MSE	= 9.1894

anc_fee	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
GR2	2	6.164414	0.32	0.750	-11.13914	15.13914
GR3	-6	6.164414	-0.97	0.346	-19.13914	7.139137
GR4	3.333333	7.018494	0.47	0.642	-11.62623	18.2929
GR5	3.72e-15	7.018494	0.00	1.000	-14.95957	14.95957
_cons	10	4.594683	2.18	0.046	.2066652	19.79333

(e) [**1 points**] Now do a joint-test that all the coefficients are zero. Do we reject the null that all the coefficients are jointly zero? Compare your p-value (`Prob > F`) with the corresponding value reported in Table 1.

(Hint: use `test` command in stata after the `regress` command)

The hypothesis that you are testing is that all of the coefficients are jointly zero. That is, none of the differences are statistically different from zero. Note that this is a much stronger test (easier to reject) than testing each coefficient individually.

*The F-statistic is **0.68** and* `Prob > F`*=**0.619**. This is identical to the value reported in last column of Table 1.*

Hence, we fail to reject the null that all the coefficients are jointly zero. This shows that randomization was done properly and that ANC fees are statistically identical across groups.

2. **[14 points]** Now, lets compute the effect of the subsidy on whether a patient that visits the clinic buys a bednet or not. The researchers arrived at clinics on random, non-announced days and asked all the visitors present in that day about whether they purchased bednets or not. For this part, load **buypatient.dta** into Stata.

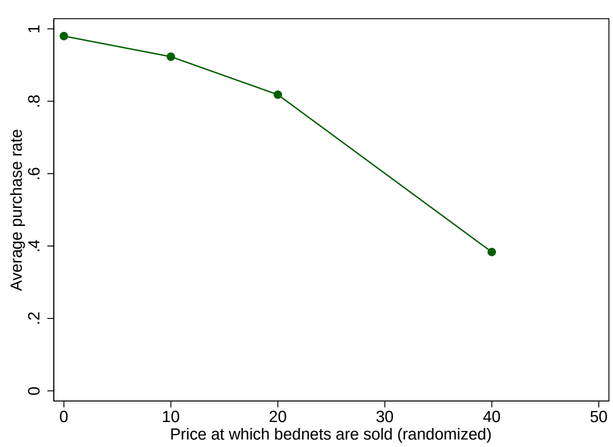
   (a) **[4 points]** Find the average probability that a patient in Group 2 (free distribution) buys ITNs. Find the same for groups 3, 4 and 5. Plot these probabilities in a graph. Compare your graph with Figure 1 in the paper.

   [Hint: In stata you can use the `collapse (mean)` command to collapse your data to an aggregated level that you specify (in the `by(varname)` option) and plot the data using `twoway scatter` command. Or you can simply get the averages into excel and plot it in excel]

   The means are:

Group	mean(buy)
1	
2	.98
3	.923077
4	.818182
5	.383562

The figure looks identical to Figure 1 from the paper. Demand (purchase rate) drops to about 40 percent when price increases from 0 to 40 Ksh.



(b) **[3 points]** This data is at the patient level whereas the experiment was randomized at the clinic level. So, it makes sense to aggregate things to the clinic level. Use `collapse` command to collapse individual level purchase data to clinic level averages (your collapsed dataset should have 20 observations, 1 for each clinic). Now test whether purchase rate in Group 2 (free distribution) clinics were different from purchase rates in Group 3 (10 Ksh). Give the t-statistic and indicate whether the difference is significant at 95% significance level.

[See solutions to part c)]

(c) **[1 point]** Now, compare purchase rates for Group 2 vs Group 4 and for Group 2 vs Group 5. Can you reject the null that free distribution has no effect on purchase rate compared to clinic that charge higher prices.

The test results are shown below. At 95% significance level, we can reject that each difference is statistically different from zero. Particularly at higher prices (groups 4 and 5), the differences are much larger, with very large t-statistic.

	Diff	SE	t-stat	p-value
G2vG3	-.06722893	.02640898	2.5456841	.03440501
G2vG4	-.17296418	.01938434	8.9228834	.0001105
G2vG5	-.58310548	.05662436	10.297785	.00004898

(d) **[3 points]** As in part d in the previous question, use a regression to perform all three comparison in one statistical test. (exclude group 2 in this case). Do your conclusions differ from part b and c? Explain.

See the result below. Note that the magnitudes are the same as the test in part c). The differences are statistically different for groups 4 and 5, but not for group 3. According to the regression, purchase rates in the group that receives free distribution of bednets is no different from the rates in the group that sells at 10 Ksh (at 95% significance level). But for higher prices, the differences are statistically different. This shows that as price increases, purchase rates drop.

Source	SS	df	MS
Model	.708468357	3	.236156119
Residual	.049022491	12	.004085208
Total	.757490848	15	.05049939

Number of obs =	16
F(3, 12) =	57.81
Prob > F =	0.0000
R-squared =	0.9353
Adj R-squared =	0.9191
Root MSE =	.06392

buy	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
GR3	-.0672289	.0404238	-1.66	0.122	-.1553048	.020847
GR4	-.1729642	.0466774	-3.71	0.003	-.2746654	-.0712629
GR5	-.5831055	.0466774	-12.49	0.000	-.6848067	-.4814042
_cons	.9885714	.0285839	34.58	0.000	.9262924	1.05085

(e) **[3 points]** Instead of averages at the clinic level where you worked with the share of patients that bought the bednets, one could do this exercise with the number of patients that bought the bednets. Go back to using the original **buypatients.dta** but instead of collapsing using means, collapse the sum at the clinic level. That is, use `collapse (sum)` instead of just `collapse`. You will still have 20 observations (one per each clinic), but the variable will now be a count of individuals who bought the bednet rather than share. In this dataset, regress number of patients who bought the bednets on the price which clinics sell ITNs. Report your coefficients as well as standard errors for the slope and the constant terms.

The regression result is below

```
      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  1,    18) =   11.85
       Model |  1339.60114     1  1339.60114           Prob > F      =  0.0029
    Residual |  2034.14886    18  113.00827            R-squared     =  0.3971
-------------+------------------------------           Adj R-squared =  0.3636
       Total |     3373.75    19  177.565789           Root MSE      =  10.631

------------------------------------------------------------------------------
         buy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    netprice |  -.4297037   .1248063    -3.44   0.003    -.6919119   -.1674954
       _cons |   26.48863   3.584787     7.39   0.000     18.95727    34.01999
------------------------------------------------------------------------------
```

3. [**10 points**] The researchers also collect weekly sales data from the log-books of the clinics. They do so to provide another measure of the demand of bednets. Lets now look at this data and see if we find the same (the drastic reduction in demand when price is increased). For this part, load **sales.dta** into stata.

   (a) [**2 points**] The log-book data records all the sales done by the clinic. In an ideal setting, this measures the sales data perfectly. Why do you think the researchers went on to the trouble of going to the clinics unannounced and collecting the data from those who were present?

   If the sales data were perfect, then it would not make sense to collect the data as described in part 2. The researchers were possibly worried that the sales data from the clinics could be tampered with or incomplete. Clinics may fake the sales data so as to show the researchers that the program is doing really well. The staff at the clinic might also forget to maintain the log-book properly (these kinds of issues are quite common in developing countries like Kenya). So, the researchers wanted to have an independent way to measure demand that had nothing to do with how the clinics kept the log-books.

   (b) [**2 points**] Regress weekly sales on price at which clinics sell ITNs. Compare your coefficient with Table 3, column 1.

   The regression result is reported below. The coefficient is almost identical to Table 3, column 1.

```
      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  1,    88) =   11.74
       Model |  9770.59492     1  9770.59492           Prob > F      =  0.0009
    Residual |  73253.9051    88   832.43074           R-squared     =  0.1177
-------------+------------------------------           Adj R-squared =  0.1077
       Total |     83024.5    89  932.859551           Root MSE      =  28.852

------------------------------------------------------------------------------
weeklynets~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    netprice |  -.8081551   .2358893    -3.43   0.001    -1.276936   -.3393746
       _cons |   43.06996   4.263445    10.10   0.000     34.59726    51.54267
------------------------------------------------------------------------------
```

(c) [**2 points**] Interpret the slope coefficient. What does it tell you? Interpret the intercept.

The slope coefficient is -0.808. It means that when price of benets goes up by 10 Ksh, a clinic sells, on average, 8.08 fewer bednets in a week.

The intercept tells you the level of sales when netprices are zero. That is, when the nets are distributed for free (netprice=0), a clinic will sell, on average, 43 bednets.

(d) [**1 point**] Based on your coefficients, how many bednets would a clinic expect to sell if the price of the bednet is sold at its prevailing price of 50 Ksh?

Simply plug in in the regression result

$$
\begin{aligned}
E\left[y|p\right] =& 43.07 - 0.808 \times p \\
E\left[y|p = 50\right] =& 43.07 - 0.808 \times 50 \\
=& 2.66
\end{aligned}
$$

So, in absence of any subsidy, a clinic will sell, on average, only 2.66 bednets in a week (compared to free distribution, this is a reduction of 94 percent!)

(e) [**3 points**] Compare the estimate in 3b with 2e. You would not expect the coefficients to be similar because your dependent variable is different in each case. In part 2e, it measured total number of patients who bought the nets in the few random days the researchers went and collected data. Whereas, in part 3b, it measured the total number of bednets sold by the clinics in a week. To compare the coefficients, you can perform this simple exercise for each of the estimates:

- Assume that $p = 0$, what is the sales?
- Now suppose the price increased to $p = 1$. What is the sales?
- What is the proportional increase in sales/demand?

Perform this exercise for both estimates, and explain whether you think the two measures of bednet demand are giving you the same result.

When price is 0, the intercept term gives the sales/demand. When price is 1, the slope term gives the increase in sales/demand. The proportional fall in demand in case of 2e *was 1.6% whereas for the estimate in* 3b *it was 1.9%. The effect of price on demand is very similar from these two measures though they measured demand differently in each case.*

4. [**6 points**] Now, suppose another researcher had collected data from 100 clinics in the same region before the experiment was conducted. This researcher collected data on weekly sales of ITNs and the price at which the clinic sold the bednets. The data is present in **salesMarket.dta**. (Note that this dataset has nothing to do with the experiment described above)

(a) [**2 points**] Regress the sale of bednets on the price of bednets. Is the slope coefficient significantly different from zero?

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =    4.35
       Model |  5.27533466     1  5.27533466           Prob > F      =  0.0396
    Residual |  118.817571    98  1.21242419           R-squared     =  0.0425
-------------+------------------------------           Adj R-squared =  0.0327
       Total |  124.092906    99  1.25346369           Root MSE      =  1.1011

------------------------------------------------------------------------------
       sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |   .0752836   .0360913     2.09   0.040     .0036616    .1469056
       _cons |   -1.37716   1.826106    -0.75   0.453    -5.001008    2.246687
------------------------------------------------------------------------------
```

(b) [**1 point**] Interpret the slope coefficient.

The slope coefficient indicates that increase in price of 1 Ksh leads to an increase in bednet sales by 0.075 units

(c) [**3 points**] Propose an economic hypothesis on why your the estimate of the slope from 4a differs from that of 2e and 3b.

In the market data, price is endogenous and could be affected by demand for bednets. It could be that clinics that operate in more malaria-intense places would have higher demand for bednets, leading to a higher price (and vice versa). This would lead to a positive correlation between price and quantity sold.

In the experimental data, price is exogenous and is chosen by the researchers. Therefore, the relationship is causal and the regression actually measure the demand curve. (Essentially, the experiment traces out the demand curve by varying supply curve.)

Sample do-file for the problem

Solution do-file

```
/*
Solution do-file for 14.03 Problem Set 3
Written by Maheshwor Shrestha
You can ignore the lines that begin with "*" and are stata commands
- These are just conveniences to output the results
*/
**boilerplates
clear
set more 1
set scheme s1color
*set working directory
cd "C:\Users\mahesh\Dropbox (MIT)\14.03-Fall-2014\p-sets\ps3"
*mat drop _all
**Problem 2
use baseline, clear
```

```
**Means and SD by group
table group, c(mean anc_fee sd anc_fee)
** t-test comparing group 1 with the others
foreach i in 2 3 4 5{
ttest anc_fee if inlist(group, 1, `i'), by(group)
* mat A2 = nullmat(A2) \ (r(mu_2)-r(mu_1), r(se), r(t), r(p))
}
*mat colnames A2 = Diff SE t-stat p-value
*mat rownames A2 = G1vG2 G1vG3 G1vG4 G1vG5
*mat list A2
** do the test at once using a regression
qui tab group, gen(GR)
reg anc_fee GR2 GR3 GR4 GR5
** test all the coefficients are jointly zero
testparm GR*
**Problem 3
use buypatient, clear
** Mean purchase rate by group
table group, c(mean buy)
**Plot the grapch
collapse (mean) buy, by(netprice group)
sc buy netprice, c(l) ytitle(Average purchase rate) ylab(0(.2)1)
graph export "fig2a.pdf", replace
**Since collapse changes the dataset in memory, reload it
use buypatient, clear
** collapse it to clinic level averages (also include netprice and group)
collapse (mean) buy, by(clinicid netprice group)
*t-test comparing group 2 with 3,4, and 5
foreach i in 3 4 5{
ttest buy if inlist(group, 2,`i'), by(group)
* mat B2 = nullmat(B2) \ (r(mu_2)-r(mu_1), r(se), r(t), r(p))
}
*mat colnames B2 = Diff SE t-stat p-value
*mat rownames B2 = G2vG3 G2vG4 G2vG5
*mat list B2
** doing the same tests through a regression
qui tab group, gen(GR)
reg buy GR3 GR4 GR5
**Problem 2e
use buypatient, clear
** collapse the data by sum so that it counts the number of people who bought
collapse (sum) buy, by(clinicid netprice group)
reg buy netprice
*** Problem 3
use sales, clear
reg weeklynetsales netprice
*** Problem 4
use salesMarket, clear
reg sales p
```