UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam: **ECON3150/4150 – Introductory econometrics**

Date of exam: Tuesday, August 9, 2011

Time for exam: 09:00 a.m. – 12:00 noon

The problem set covers 5 pages (incl. cover sheet)

Resources allowed:

- All written and printed resources, as well as calculator, is allowed.

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

ECON 3150/4150: INTRODUCTORY ECONOMETRICS

PROBLEM SET, POSTPONED EXAM SPRING 2011

PROBLEM 1

We want to investigate how the hourly wage rate of individuals depends on education, work experience and some other factors. Data from the U.S.A. for 4733 persons are available for the following variables, with summary statistics and correlation coefficients given below the definitions:

```
wage = earnings per hour, U.S.Dollars
educ = years of education
exper = years of experience
age = age in years
female = 1 if female
married = 1 if married
union = 1 if member of trade union
fulltime = 1 if full time worker
```

```
-------------------------------------------------------------------------------
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------------
       wage  |      4733     10.1872    6.213761       1.05      78.71
logw=log(wage)|      4733    2.166535    .5497659   .0487902    4.36577
        educ |      4733    13.30235    2.356101          1         18
       exper |      4733    19.03803    11.39808          0         52
         age |      4733    38.33446    11.30146         18         64
      female |      4733    .4851046    .4998309          0          1
     married |      4733    .6044792    .4890139          0          1
       union |      4733    .1633214    .3696977          0          1
    fulltime |      4733    .8814705    .3232682          0          1
*******************************************************************************
-------------------------------------------------------------------------------
             |     logw     educ    exper      age   female  married    union fulltime
-------------+-----------------------------------------------------------------------
        logw |   1.0000
        educ |   0.4498   1.0000
       exper |   0.1779  -0.1480   1.0000
         age |   0.2736   0.0590   0.9785   1.0000
      female |  -0.2109   0.0317  -0.0016   0.0051   1.0000
     married |   0.2099   0.0418   0.2915   0.3034  -0.0604   1.0000
       union |   0.1826   0.0452   0.1271   0.1377  -0.0640   0.0418   1.0000
    fulltime |   0.2048   0.0349   0.0887   0.0973  -0.2208   0.0242   0.0612   1.0000
*******************************************************************************
```

Question (1-A): Consider `REGRESSION 1-1` below. Construct a 95 % confidence interval for the partial effect of one year longer education, `educ`, on the logarithm of the wage rate `logw`. Make the necessary addition assumptions you need, and explain precisely, in words, the interpretation of this interval.

```
REGRESSION 1-1
                                  ------------------------------------------------------
   Number of obs =    4733            logw |      Coef.   Std. Err.      t    P>|t|
   F(  2,  4730) =  845.77         -------------+----------------------------------------
   Prob > F      =  0.0000            educ |   .1135814   .0029442    38.58   0.000
   R-squared     =  0.2634           exper |   .0120574   .0006086    19.81   0.000
   Root MSE      =  .47193           _cons |   .4260865   .0430064     9.91   0.000
                                  ------------------------------------------------------
*******************************************************************************
```

Question (1-B): In `REGRESSION 1-2` below, `age` is also included as a regressor. Can you explain why this changes the results substantially: the point estimate for the effect of age is positive and the effect of experience becomes negative?

Question (1-C): In `REGRESSION 1-3`, four dummy variables have been added to the regressor set specified in `REGRESSION 1-1`. Interpret the coefficient estimate -0.2078 for the `female` regressor. Can you reject the hypothesis that there is no gender differences in the wage rate? Explain briefly.

```
REGRESSION 1-2
                                  ------------------------------------------------------
   Number of obs =    4733        logw |      Coef.   Std. Err.      t    P>|t|
   F(  3,  4729) =  565.79     -------------+----------------------------------------
   Prob > F      =  0.0000          educ |    .009577   .0488553     0.20   0.845
   R-squared     =  0.2641         exper |  -.0920859   .0488355    -1.89   0.059
   Root MSE      =  .47176           age |   .1040662   .0487956     2.13   0.033
                                   _cons |  -.1970469   .2953267    -0.67   0.505
                                  ------------------------------------------------------
*******************************************************************************

REGRESSION 1-3
                                  ------------------------------------------------------
   Number of obs =    4733        logw |      Coef.   Std. Err.      t    P>|t|
   F(  6,  4726) =  430.49     -------------+----------------------------------------
   Prob > F      =  0.0000          educ |   .1094801   .0027837    39.33   0.000
   R-squared     =  0.3534         exper |   .0090616   .0006064    14.94   0.000
   Root MSE      =  .44236        female |  -.2027842   .0132597   -15.29   0.000
                                  married |   .1312241   .0138364     9.48   0.000
                                    union |   .1686208   .0176231     9.57   0.000
                                 fulltime |   .2063088   .0205301    10.05   0.000
                                    _cons |   .3473335   .0438642     7.92   0.000
                                  ------------------------------------------------------
*******************************************************************************
```

Question (1-D): In `REGRESSION 1-4` another two variables have been added to those in `REGRESSION 1-3`. (a) Estimate the effect of a one year longer education, `educ`, on the log of the wage rate for males and for females. (b) Can you reject the hypothesis that a one year longer experience, `exper`, has the same effect on wage for males and females?

```
REGRESSION 1-4
                                  ------------------------------------------------------
   Number of obs =    4733        logw |      Coef.   Std. Err.      t    P>|t|
   F(  8,  4724) =  325.17     -------------+----------------------------------------
   Prob > F      =  0.0000          educ |   .1037604   .0036861    28.15   0.000
   R-squared     =  0.3551         exper |   .0104339   .0008268    12.62   0.000
   Root MSE      =  .44186        female |  -.3276217   .0818733    -4.00   0.000
                             educ*female |   .0130892   .0055954     2.34   0.019
                            exper*female |  -.0026325   .0011458    -2.30   0.022
                                 married |   .1279937   .0138719     9.23   0.000
                                   union |   .1627573   .0177129     9.19   0.000
                                fulltime |   .2047946   .0205121     9.98   0.000
                                   _cons |   .4014313   .0561129     7.15   0.000
                                  ------------------------------------------------------
*******************************************************************************
```

PROBLEM 2

We have observations on 4682 sales of dwellings in a city in the U.S.A. in the years 1991–1996, with, inter alia, the three variables below recorded. Again, summary statistics and correlation coefficients are given below the definitions:

```
   price = house price, in U.S. Dollars
   area = total living area, in 10 square feet
   age = age, in years
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	4682	109561.4	35411.8	20000	375000
area	4682	144.8934	34.12112	60	220
age	4682	24.95323	20.14404	0	96

	price	area	age
price	1.0000		
area	0.7282	1.0000	
age	-0.2979	-0.2510	1.0000

Question (2-A): We have regressed the dwelling `price` on only the year dummy for 1994, `d94`, without including intercept. The result is given below. Show that the estimated coefficient of this dummy equals the mean price of the dwellings sold in 1994.

price	Coef.	Std. Err.	t
d94	108199.8	3571.863	30.29

Question (2-B): `REGRESSION 2-1` reports the results of regressing the `price` of the dwelling sold on its `area` and its `age`. `REGRESSION 2-2` and `REGRESSION 2-3` show the results when one of the two regressors are omitted. What do you conclude from this? The estimated coefficient of `age` and the `R-squared` when `area` is omitted are more strongly affected than are the estimated coefficient of `area` and the `R-squared` when `age` is omitted. Have you any comments to this result?

REGRESSION 2-1

Number of obs =	4682	price	Coef.	Std. Err.	t	P>|t|
F(2, 4679) =	2795.55	area	723.7403	10.57953	68.41	0.000
Prob > F =	0.0000	age	-215.9494	17.9202	-12.05	0.000
R-squared =	0.5444	_cons	10084.83	1736.638	5.81	0.000
Root MSE =	23907					

REGRESSION 2-2

Number of obs =	4682	price	Coef.	Std. Err.	t	P>|t|
F(1, 4680) =	5283.09	area	755.7386	10.39747	72.68	0.000
Prob > F =	0.0000	_cons	59.86211	1547.726	0.04	0.969
R-squared =	0.5303					
Root MSE =	24273					

REGRESSION 2-3

Number of obs =	4682	price	Coef.	Std. Err.	t	P>|t|
F(1, 4680) =	455.68	age	-523.6375	24.53029	-21.35	0.000
Prob > F =	0.0000	_cons	122627.8	786.6383	155.89	0.000
R-squared =	0.0887					
Root MSE =	33808					

Question (2-C): In `REGRESSION 2-4` squared terms in `are` and `area` and their product have been added to the regressors in `REGRESSION 2-1`. Next test by an F-test, using the values of the two `R-squared` reported, the hypothesis that the relationship is linear in both variables. State briefly your conclusions.

```
REGRESSION 2-4
```

Number of obs = 4682	price	Coef.	Std. Err.	t	P>|t|
F(5, 4676) = 1138.62	area	618.2097	87.3101	7.08	0.000
Prob > F = 0.0000	age	330.6214	94.77704	3.49	0.000
R-squared = 0.5490	area^2	.5885961	.2774832	2.12	0.034
Root MSE = 23793	age^2	-2.801938	.7133769	-3.93	0.000
	area*age	-2.589901	.5287026	-4.90	0.000
	_cons	10493.29	6872.28	1.53	0.127

****

Question (2-D): `REGRESSIONS 2-1 & 2-4` have been rerun after having replaced `area` and `age` by `areademean=area-mean(area)` and `agedemean=age-mean(age)`, their values measured from their respective sample means. The results are given in `REGRESSIONS 2-5 & 2-6`. All slope coefficients in `REGRESSION 2-1` and `REGRESSION 2-5` are the same, but only the three last slope coefficients in `REGRESSION 2-4` and `REGRESSION 2-6` are unchanged. Can you explain this?

```
REGRESSION 2-5
```

Number of obs = 4682	price	Coef.	Std. Err.	t	P>|t|
F(2, 4679) = 2795.55	areademean	723.7403	10.57953	68.41	0.000
Prob > F = 0.0000	agedemean	-215.9494	17.9202	-12.05	0.000
R-squared = 0.5444	_cons	109561.4	349.393	313.58	0.000
Root MSE = 23907					

****

```
REGRESSION 2-6
```

Number of obs = 4682	price	Coef.	Std. Err.	t	P>|t|
F(5, 4676) = 1138.62	areademean	724.1507	10.67903	67.81	0.000
Prob > F = 0.0000	agedemean	-184.4728	22.27279	-8.28	0.000
R-squared = 0.5490	areademean^2	.5885962	.2774832	2.12	0.034
Root MSE = 23793	agedemean^2	-2.801938	.7133769	-3.93	0.000
	area*agedem	-2.589901	.5287026	-4.90	0.000
	_cons	109566.3	547.0994	200.27	0.000