UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: **ECON4130 – Statistics 2**

Date of exam: Friday, December 4, 2015 **Grades are given: January 4, 2016**

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

The problem set covers 6 pages

Resources allowed:
- Open book exam, where all written and printed resources are allowed

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

Problem 1

A. If Y is exponentially distributed with parameter 1 (in short $Y \sim \exp(1)$), the cumulative distribution function (*cdf*) of Y is given by

$$F_Y(y) = P(Y \le y) = \begin{cases} 1 - e^{-y} & \text{if } y > 0 \\ 0 & \text{if } y \le 0 \end{cases}$$

i) Explain why the function $F_Y(y)$ is a proper cdf (i.e., satisfies the necessary properties any cdf should fulfil).

ii) Explain why any of the two functions $f_1(y) = \begin{cases} e^{-y} & \text{for } y > 0 \\ 0 & \text{for } y \le 0 \end{cases}$, and $f_2(y) = \begin{cases} e^{-y} & \text{for } y \ge 0 \\ 0 & \text{for } y < 0 \end{cases}$, may serve as the probability density function (*pdf*) of Y.

iii) Explain why $E\left(Y^r\right) = \Gamma(1+r)$ for any real number $r > -1$, where $\Gamma(t)$ denotes the gamma function.

B. Let X be the length in minutes of an arbitrary telephone call during work hours in a company. We assume X is a continuous random variable (*rv*) with cdf

$$F_X(x) = \begin{cases} 1 - e^{-\sqrt{\alpha x}} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

and pdf

$$f_X(x) = \begin{cases} \dfrac{\sqrt{\alpha}}{2} x^{-\frac{1}{2}} e^{-\sqrt{\alpha x}} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

where $\alpha > 0$ is a parameter in the distribution.

i) Show that $Y = \sqrt{\alpha X} \sim \exp(1)$ distributed.

ii) Show that $E\left(X^r\right) = \dfrac{\Gamma(1+2r)}{\alpha^r}$ for any real number $r > -\dfrac{1}{2}$.

C. We want to estimate α based on an *iid* (independent and identically distributed) sample, $X_1, X_2, \ldots, X_n$, where X_i is distributed as X in section **B**, $i = 1, 2, \ldots, n$.

i) Find the moment method estimator (*mme*) for α.

ii) Find the maximum likelihood estimator (*mle*) for α.

iii) Calculate both estimates (mme and mle) when $n = 55$ and the data resulted in the observed values $\sum_{i=1}^{n} X_i = 274.6$ and $\sum_{i=1}^{n} \sqrt{X_i} = 89.2570$.

iv) Why is the mme a consistent estimator for α?

D. i) Show that the Fisher information for one observation is $I(\alpha) = \dfrac{1}{4\alpha^2}$.

ii) Develop an approximate 95% confidence interval for α based on the mle, $\hat{\alpha} = \dfrac{1}{\left(\dfrac{1}{n}\sum_{i=1}^{n}\sqrt{X_i}\right)^2}$, mle theory , and Slutsky's lemma.

iii) Calculate the observed value of the confidence interval in **ii)**.

E. We want to use the data in section **C** to test $H_0 : E(X) \le 5$ against $H_1 : E(X) > 5$ (minutes).

i) Set up a test criterion based on the mle $\hat{\alpha}$ in section **D**.

ii) Perform the test by calculating the p-value approximately and state your conclusion.

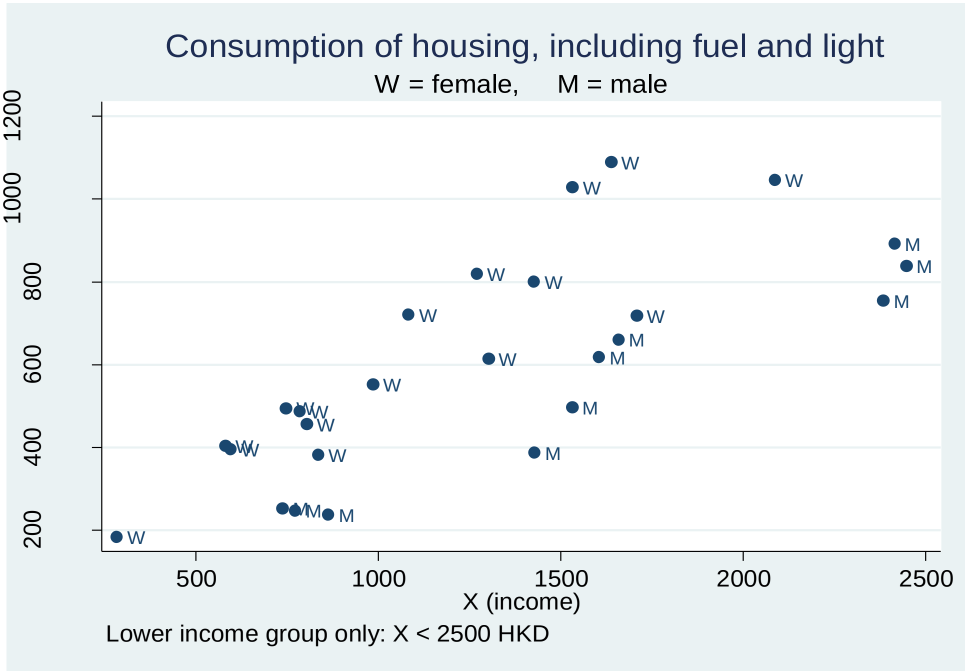
Problem 2

Introduction. In this problem we will look at the effect of gender on the consumption of housing, including fuel and light, for lower income (< 2500 HKD) consumers in Hong Kong. The original data (40 consumers) are given in table 1 while the lower income data from the sample (26 consumers) are plotted in figure 1.

Table 1 Consumption of housing, including fuel and light, and income for a sample of 40 Hong Kong consumers.

Women	Consumer no.	1	2	3	4	5	6	7	8	9	10
	Consumption	820	184	921	488	721	614	801	396	864	845
	Income	1271	284	3128	786	1084	1303	1428	596	2899	3258
	Consumer no.	11	12	13	14	15	16	17	18	19	20
	Consumption	404	781	457	1029	1047	552	718	495	382	1090
	Income	581	3186	804	1533	2088	986	1709	748	836	1639
Men	Consumer no.	21	22	23	24	25	26	27	28	29	30
	Consumption	497	839	798	892	1585	755	388	617	248	1641
	Income	1532	2448	3358	2416	6582	2385	1429	2972	773	10615
	Consumer no.	31	32	33	34	35	36	37	38	39	40
	Consumption	1180	619	253	661	1981	1746	1865	238	1199	1524
	Income	4004	1606	738	1659	5371	6748	9731	864	2899	5637

Figure 1 Consumption of housing, including fuel and light, vs. income for the lower income group (26 consumers) of the sample.



For a randomly selected consumer we define *Y* as the consumption (in HKD) of housing, including fuel and light, for the period in question, *X* the income (in HKD) for the same period, and *M* a dummy variable for gender (*M* = 0 for female and *M* = 1 for male).

The population of interest consists of consumers in Hong Kong with income $X < 2500$ HKD.

Model. Assume that the conditional distribution of *Y* given fixed values $M = m$ and $X = x$, is normal with expectation

(1) $E(Y \mid x, m) = \beta_0 + \beta_1 x + \beta_2 m + \beta_3 m \cdot x$

and constant variance

(2) $\text{var}(Y \mid x, m) = \sigma^2$

Questions.

A. **i)** The ceteris paribus (cet. par.) effect of gender is defined as the expected difference in consumption between males and females for a given income being the same for both genders. Explain why the cet. par. effect of gender is $\beta_2 + \beta_3 x$ based on the model assumption (1), where the common income for both genders is *x*.

ii) Find the cet. par. effect of a unit change in income *x* on the expected consumption. In what way does this effect depend on the gender?

B. **Introduction.** The corresponding model for the random mechanism behind the data is specified as

(3) $Y_i = E(Y_i \mid x_i, m_i) + e_i = \beta_0 + \beta_1 x_i + \beta_2 m_i + \beta_3 m_i \cdot x_i + e_i, \quad i = 1, 2, \ldots, n \quad (n = 26)$

where the regressors, $x_i, m_i, \quad i = 1, 2, \ldots, n$, are considered fixed numbers due to their exogeneity, and where the error terms, $e_1, e_2, \ldots, e_n$, are assumed to be iid and normally distributed random variables, $e_i \sim N(0, \sigma^2)$.
The model (3) reduces to two simple regressions, one for women (16 observation units) and one for men (10 units). Model (3) also assumes that the error variances of the two regressions are the same, i.e., $\sigma_W^2 = \sigma_M^2 = \sigma^2$, where σ_W^2, σ_M^2 are the error variances of the two regressions respectively. We should check if there is any evidence in the data against this assumption.

Questions. The two simple regressions are estimated by Stata in the appendix, see A1 and A2.

i) Use the outputs in A1 and A2 to set up unbiased estimates for σ_W^2 and σ_M^2.

ii) **Specification test**: Use the outputs in A1 and A2 to test $H_0 : \sigma_W^2 = \sigma_M^2$ against $H_1 : \sigma_W^2 \neq \sigma_M^2$ at the 5% level of significance.

(**Hint:** If you don't find the right critical level in the Rice table you use, e.g., if the degrees of freedom needed are not represented in the table, you can guess roughly the critical value from the nearest values in the table.)

C. We want to test if there is evidence in the data to claim that gender has an effect on the consumption in question, i.e., if the cet. par. effect of gender derived in section **A i)** is different from zero. The full (in (3)) and reduced model that can be used to test this, have been estimated in appendix **A3** and **A4**. Set up a proper null-hypothesis, perform a test at 1% level of significance, and state a conclusion.

D. Let the population mean income in the lower income group be μ_0 and μ_1 for women and men respectively, or, in other words,

$$\mu_m = E(X \mid m) = \begin{cases} \mu_0 & \text{for women} \\ \mu_1 & \text{for men} \end{cases}$$

Explain why the model assumption (1) implies that

$$E(Y \mid m) = \beta_0 + \beta_1\mu_m + (\beta_2 + \beta_3\mu_m)\cdot m = \begin{cases} \beta_0 + \beta_1\mu_0 & \text{for women} \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_3)\mu_1 & \text{for men} \end{cases}$$

Hint: Use the law of total expectation on the relation (1), where the outer expectation is referring to the conditional distribution of *X* given $M = m$ fixed.

Appendix: Stata Outputs for Problem 2

A1. Simple regression Y on X for 16 lower income WOMEN

```
      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  1,    14) =   70.86
       Model |  890213.453     1  890213.453           Prob > F      =  0.0000
    Residual |  175882.297    14  12563.0212           R-squared     =  0.8350
-------------+------------------------------           Adj R-squared =  0.8232
       Total |  1066095.75    15   71073.05            Root MSE      =  112.08

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |    .498313   .0591973     8.42   0.000     .3713473    .6252786
       _cons |   86.86374   71.14858     1.22   0.242    -65.73479    239.4623
------------------------------------------------------------------------------
```

A2. Simple regression Y on X for 10 lower income MEN

```
      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =  124.45
       Model |  530115.571     1  530115.571           Prob > F      =  0.0000
    Residual |  34076.4291     8  4259.55364           R-squared     =  0.9396
-------------+------------------------------           Adj R-squared =  0.9321
       Total |      564192     9       62688           Root MSE      =  65.265

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |   .3638185   .0326123    11.16   0.000     .2886144    .4390226
       _cons |  -37.65232   55.65846    -0.68   0.518     -166.001    90.69631
------------------------------------------------------------------------------
```

A3. Full model regression for problem 2C

```
      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  3,    22) =   51.69
       Model |  1479883.74     3  493294.579           Prob > F      =  0.0000
    Residual |  209958.726    22  9543.57845           R-squared     =  0.8758
-------------+------------------------------           Adj R-squared =  0.8588
       Total |  1689842.46    25  67593.6985           Root MSE      =  97.691

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |    .498313   .0515954     9.66   0.000     .3913107    .6053152
           M |  -124.5161    103.857    -1.20   0.243    -339.9023    90.87012
          XM |  -.1344945   .0710282    -1.89   0.072     -.281798     .012809
       _cons |   86.86374   62.01187     1.40   0.175    -41.74101    215.4685
------------------------------------------------------------------------------
```

A4. Reduced model regression for problem 2C

```
      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  1,    24) =   32.17
       Model |  967783.563     1  967783.563           Prob > F      =  0.0000
    Residual |  722058.898    24  30085.7874           R-squared     =  0.5727
-------------+------------------------------           Adj R-squared =  0.5549
       Total |  1689842.46    25  67593.6985           Root MSE      =  173.45

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |   .3277504   .0577876     5.67   0.000     .2084826    .4470182
       _cons |   176.9169   81.91226     2.16   0.041     7.858315    345.9755
```