Home assignment # 1
(Suggested solutions)

1. Imagine you work in a bulb factory as a supervisory technician, who is asked to check the lifespan of light bulbs. You know that the lifespan of light bulbs produced in your factory is distributed $N(200, 400)$, that is, normally distributed with a mean of 200 hours and a variance of 400 hours. What is the probability that you find a light bulb with a life span longer than 245 hours? (*Hint*: convert the distribution to a standard normal one and then refer to statistical tables.)

   Solution:

   We have $X \sim N(200, 400)$, and we are asked to compute $P(X > 245)$.

   We need to first convert X to a standard normal variable, which will be done by substracting the mean (200) and dividing it by its standard deviation ($\sqrt{400}$):

$$Z = \frac{X - 200}{\sqrt{400}} \sim N(0, 1) \quad .$$

   We have

$$\begin{aligned} P(X > 245) &= 1 - P(X \leq 245) \\ &= 1 - P\left(\frac{X - 200}{\sqrt{400}} \leq \frac{245 - 200}{\sqrt{400}}\right) = 1 - P\left(Z \leq \frac{45}{20}\right) \\ &= 1 - F_Z(2.25) = 1 - \Phi(2.25) \end{aligned}$$

   ,

   where $\Phi()$ is the cdf of standard normal distribution, whose values can be found in statistical tables. We find in the statistical tables that

$$\Phi(2.25) = 0.9878$$

   and so we obtain

$$P(X > 245) = 1 - 0.9878 = 0.0122$$

   Hence, the probability that you find a light bulb with a life span longer than 245 hours is slightly over 1%.

2. You receive a unique dataset that includes wages of all citizens of Ostrava as well as their experience (number of years spent working). Obviously, you are very curious about what is the effect of experience on wages. You run an OLS regression of monthly wage in CZK on the number of years of experiences and obtain the following results:

$$\widehat{wage}_i = 14450 + 1135 \cdot exper_i$$

(a) Interpret the meaning of the coefficient of $exper_i$.

(b) Use the estimates to determine the average wage of a person with 1, 5, 20, and 40 years of experience. Do they seem as realistic?

(c) Use your answer to part (b) and explain what is the problem with your estimation. Draw the theoretical diagram you'd expect for the relationship between $wage_i$ and $exper_i$, then plot the estimated equation onto the same diagram. Suggest a way to improve you estimation equation.

Solution:

(a) The coefficient is positive and the interpretation is that the increase of the experience by 1 year increases the monthly wage by CZK 1135 per month.

(b) Average wages of individuals with 1, 5, 20, and 40 years of experience is the following:
$\widehat{wage}_i(1 \text{ year}) = 14450 + 1135 \cdot 1 = 15585$
$\widehat{wage}_i(5 \text{ years}) = 14450 + 1135 \cdot 5 = 20125$
$\widehat{wage}_i(20 \text{ years}) = 14450 + 1135 \cdot 20 = 37150$
$\widehat{wage}_i(40 \text{ years}) = 14450 + 1135 \cdot 40 = 59850$

The predicted wages see to be realistic for people with low experience, but seems to be way too high for individuals with 20 and 40 years of experience.

(c) The problem is, that we estimated this relationship as linear which obviously is not true. The more probable relationship is non-linear, for example with functional form $wage_i = \beta_0 + \beta_1 \cdot exper_i + \beta_2 \cdot exper_i^2$. We would then expect the β_1 to be positive and β_2 to be negative, which would ensure decreasing returns to experience. The difference is plotted in Figure 1.

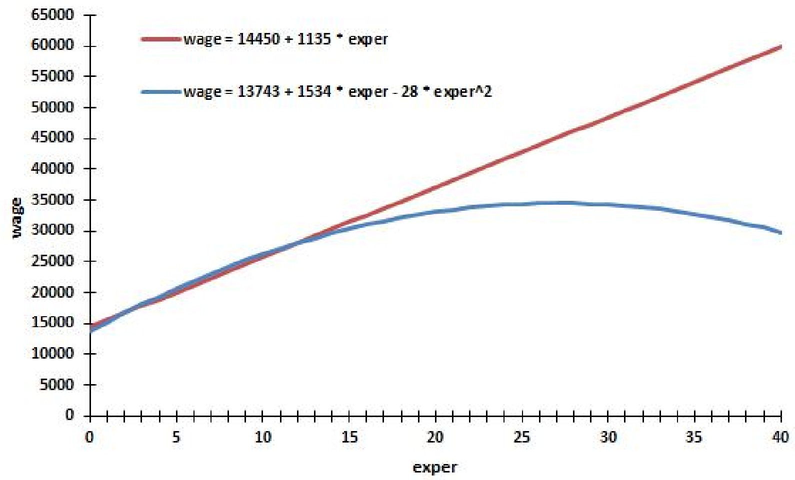

Figure 1: Relation between *wage* and *exper*

3. Let us investigate the results of an experiment in broiler (poultry meat) production. The average weight of an experimental lot of broilers and their corresponding level of average feed consumption was tabulated over the time period in which they changed from baby chickens to mature broilers ready for market. At time t, we denote the average weight of the experimental group of broilers as the output y_t, the total feed consumed by the input x_t. The observations are summarized in Table 1.

   Let us suppose that the underlying model is

   $$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t \quad .$$

   Answer the following questions. Do not use statistical software (Stata etc.) for the solution!!! You can use Excel or some mathematical software (e.g. Matlab) for multiplication and inversion of the matrices in question[1], but otherwise you are asked to do all estimation and computation "by hand" - meaning according to the formulae we introduced on the lecture. Make sure you show in your solution all the matrices you construct and compute.

   (a) Using the Least Squares formula, find the coefficients β_0, β_1 and β_2.
   (b) Find and list the residuals of the model.
   (c) What is the sign of $\widehat{\beta}_2$? How do you interpret this coefficient?
   (d) Based on your estimation result, find $\frac{\partial y_t}{\partial x_t}$, the marginal productivity of the feed input. Interpret your your result.
   (e) We denote the price of broilers by p_b and the price of feed by p_f. Explain why $\frac{\partial y_t}{\partial x_t} = \frac{p_f}{p_b}$. [Hint: Solve for the firm's profit maximization problem.]

[1]For Excel, you may want to check out the commands =MMULT() and =MINVERSE()

End of Time Period	Average Weight of Broiler in Pounds	Average Cumulative Feed Inputs in Pounds
t	y_t	x_t
1	0.57	1.00
2	1.01	2.00
3	1.20	3.00
4	1.27	4.00
5	1.91	5.00
6	2.52	6.00
7	2.55	7.00
8	2.92	8.00
9	3.38	9.00
10	3.43	10.00
11	3.53	11.00
12	4.11	12.00
13	4.26	13.00
14	4.33	14.00
15	4.41	15.00

Table 1: Data on poultry production

Solution:

(a) Let us rewrite the model and the data in matrix form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad ,$$

where:

$$\mathbf{y} = \begin{pmatrix} 0.57 \\ 1.01 \\ 1.20 \\ 1.27 \\ 1.91 \\ 2.52 \\ 2.55 \\ 2.92 \\ 3.38 \\ 3.43 \\ 3.53 \\ 4.11 \\ 4.26 \\ 4.33 \\ 4.41 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 8 & 64 \\ 1 & 9 & 81 \\ 1 & 10 & 100 \\ 1 & 11 & 121 \\ 1 & 12 & 144 \\ 1 & 13 & 169 \\ 1 & 14 & 196 \\ 1 & 15 & 225 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad .$$

To find the OLS estimate, we compute the following matrices:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 15 & 120 & 1240 \\ 120 & 1240 & 14400 \\ 1240 & 14400 & 178312 \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.7934 & -0.2044 & 0.0110 \\ -0.2044 & 0.0656 & -0.0039 \\ 0.0110 & -0.0039 & 0.0002 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 41.40 \\ 412.17 \\ 4682.65 \end{pmatrix} \quad ,$$

which gives us

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0.059 \\ 0.426 \\ -0.009 \end{pmatrix} \quad ,$$

and so the estimated model is

$$\widehat{y} = 0.059 + 0.426x - 0.009x^2 \quad .$$

(b) The residuals ($e_i = y_i - 0.059 - 0.426x + 0.009x^2$) are:

$$\mathbf{e} = \begin{pmatrix} 0.094 \\ 0.133 \\ -0.060 \\ -0.356 \\ -0.065 \\ 0.213 \\ -0.072 \\ 0.000 \\ 0.180 \\ -0.034 \\ -0.180 \\ 0.170 \\ 0.108 \\ -0.017 \\ -0.115 \end{pmatrix}$$

(c) $\widehat{\beta}_2 = -0.009 < 0$, which means that the model exhibits decreasing returns to scale. This means that broilers do not gain in weight at constant pace and that the weight gain becomes slower over time. Hence, when a broiler achieves certain weight, it does not pay to feed him any more (if our interest is purely economic, i.e. if we just want to sell our broilers for meat).

(d) Since

$$\widehat{y} = 0.059 + 0.426x - 0.009x^2 \quad ,$$

we can say that

$$\frac{\partial y_t}{\partial x_t} = 0.426 - 0.009 \cdot 2x_t = 0.426 - 0.018x_t \quad .$$

(e) Let us consider the firm's maximization problem, where we subtract the costs (price of feed) from the revenues (price of broilers sold), remembering that the weight of broilers (y) is a function of the feed (x) and so the feed (x) is the variable over which we maximize:

$$\max_{x_t} \pi = \max_{x_t} p_b y_t(x_t) - p_f x_t \quad .$$

First order condition gives us

$$\begin{aligned} p_b \frac{\partial y_t}{\partial x_t} - p_f &= 0 \\ \frac{\partial y_t}{\partial x_t} &= \frac{p_f}{p_b} \quad , \end{aligned}$$

which is what we had to show.