Dear all, this is a close book exam. Please turn in by 9:50am. Calculator is NOT allowed. Please scan your exam in one PDF file and submit to the class Brightspace site. Academic dishonesty will result in severe consequence.

1. Let $(x_1, y_1), \cdots, (x_n, y_n)$ be n points observed. Please derive the least squares regression line for simple linear regression through the origin:
   $y_i = \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \cdots, n$

Solution:

To minimize

$$SS = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2:$$

$$\frac{\partial SS}{\partial \beta_1} = 0 \Longrightarrow \sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_1 x_i\right) = 0 \Longrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$\hat{y} = \hat{\beta}_1 x$ is the least squares regression line through the origin.

2. Let $(x_1, y_1), \cdots, (x_n, y_n)$ be n points observed (i.e., the same setting as Q1). Our goal is to fit the simple linear regression through the origin model using the maximum likelihood estimator (MLE). The model and assumptions are as follows:

$$y_i = \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n.$$

We assume that the x_i's are given (condition on the x_i's, so that the x_i's are not considered as random variables here), and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

Solution:

Based on the assumptions that

$$y_i = \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n, x_i \text{ is given, and } \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2),$$

we can derive the p.d.f. for y_i that

$$y_i \underset{but\ not\ iid}{\overset{independent}{\sim}} N(\beta_1 x_i, \sigma^2) \Leftrightarrow f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y_i - \beta_1 x_i]^2}{2\sigma^2}}.$$

Thus, the likelihood and log-likelihood function can be derived (assuming the $x_i's$ are not random) as follows:

$$likelihood: \mathcal{L} = f(y_1, y_2, \cdots, y_n)$$

$$= \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y_i - \beta_1 x_i]^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \beta_1 x_i]^2}$$

Note: Alternatively, one can derive the likelihood function in a conditional approach as follows:

$$likelihood: \mathcal{L} = f(y_1, y_2, \cdots, y_n | x_1, x_2, \cdots, x_n)$$
$$= \prod_{i=1}^{n} f(y_i|x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y_i - \beta_1 x_i]^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \beta_1 x_i]^2}$$

In either approach, the log likelihood is:

$$log - likelihood: l = \ln \mathcal{L} = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \beta_1 x_i]^2.$$

In order to maximize the likelihood function, which in turn is equivalent to maximizing the log-likelihood function, we take the first derivatives of involved parameters in the as-derived log-likelihood function and set them to 0 as follows:

$$\begin{cases} \dfrac{\partial l}{\partial \beta_1} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n} x_i[y_i - \beta_1 x_i] = 0 \\ \dfrac{\partial l}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}[y_i - \beta_1 x_i]^2 = 0 \end{cases}.$$

By solving the three equations above, we can derive the MLEs as shown below:

$$\begin{cases} \hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \\ \widehat{\sigma^2} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}[y_i - \widehat{\beta_1} x_i]^2 \end{cases}$$

Now we have the fitted regression model by plugging the above MLEs of the model parameters:

$$\widehat{y_i} = \widehat{\beta_1} x_i, i = 1, 2, \dots, n.$$

Please note that the fitted regression model based on the MLE approach here assuming the regressors ($x_i's$) are not random (or equivalently, in a conditional approach), is the same as the fitted regression model using the OLS approach.