Week 3 Extension Questions

① A simple linear regression between two variables x_i and y_i can be written:

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

where ε_i is iid white noise, ie. $\varepsilon_i \sim iid\,(0, \sigma^2)$. $i = 1, ..., n$. By considering how Ordinary Least Squares minimises the sum of squared residuals (RSS), show that:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \text{and} \quad \hat{\beta} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

② The parameters, or estimated parameters, in OLS are often said to be the Best Linear Unbiased Estimators (BLUE). Show that both estimates are unbiased, ie

$$E[\hat{\beta}] = \beta, \quad E[\hat{\alpha}] = \alpha$$

③ Assuming that the variance of the residuals/model is known, ie. $\varepsilon_i \sim iid\,(0, \sigma^2)$, show that

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

④ In some cases, we assume the residuals are Gaussian white noise, ie $\varepsilon_i \sim N(0, \sigma^2)$ and iid. Assuming this is true and we don't know the variance σ^2 so have to estimate it, show that the test statistic for $\hat{\beta}$ exactly follows a t-distribution.

Hint: Consider the relationship between $N(\cdot)$, χ^2 and the t-distribution.

$$T = \frac{\hat{\beta} - \beta_{H_0}}{S_{\hat{\beta}}} \sim t_{n-2}$$

(5) Show that if we don't make the assumption of normality, ie $\varepsilon_i \sim iid(0, \sigma^2)$, then our test statistic is approximately normally distributed under the Central Limit Theorem

$$T = \frac{\hat{\beta} - \beta_{H_0}}{S_{\hat{\beta}}} \overset{a}{\sim} N(0,1)$$