Tutorial 2

1. **Airfreight breakage** A substance used in biological and medical research is shipped by airfreight to users in cartons of 1000 ampules. In the **(data)**, X is the number of times the carton was transferred from one aircraft to another over the shipment route, and Y the number of ampules found to be broken upon arrival. A linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$ is used.

   (a) Write down the estimated regression model. Plot the estimated regression function and the data. Does the linear regression function appear to give a good fit here?

   (b) Obtain a point estimator of the expected number of broken ampules when $X = 1$.

   (c) Estimate β_1 with 95% confidence interval. Interpret your interval estimate

   (d) conduct a t-test to decide whether or not there is a linear association between number of times a carton is transferred X and the number of broker ampules Y with significant level $\alpha = 0.05$? what is the P-value of the test?

   (e) β_0 here is the mean number of ampuls broken when no transfer of shipment are made. Obtain 95% confidence interval for β_0 and interpret it.

   (f) Based on the past experience, Hypothesis $\beta_1 = 2$ is made. Please test it based on the data with $\alpha = 0.025$

2. Suppose we have the following output of a regression

$$\hat{Y} = 350.7 - 1.4X$$

the p-value for the slop is 0.91

   A student concludes that "the message I get here is that the more X is, the fewer Y will be". Comment.

3. refer to the **Grade Point average** problem in tutorial 1,

   (a) Obtain 99% confidence interval for β_1. Does it include 0? why are we interested in whether the interval include 0?

   (b) Test, using t-statistic, whether there is linear association between ACT and GPA when $\alpha = 0.01$ is used?

(c) what is the p-value for part (b)? How does it support the conclusion in part (b)?

4. For linear regression model $Y_i = \beta_1 X_i + \varepsilon_i, i = 1, ..., n$

   (a) Find the least square estimator of β_1?

   (b) denote the estimator by b_1 then the estimated model is $\hat{Y}_i = b_1 X_i$. Let $e_i = Y_i - \hat{Y}_i$. Can you conclude

   $$\sum_{i=1}^{n} e_i = 0?$$

   (c) Assume that the error term are IID $N(0, \sigma^2)$ with σ^2 unknown. Find the Standard Error for the estimator of β_1

   (d) Design a procedure to test

   $$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

5. (Please ignore this question if you find it is too difficult) In the linear regression model,

   $$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, ..., n$$

   ASSUME that $X_i, i = 1, ..., n$ are RANDOM with $EX = x$ and independent of $\varepsilon_i, i = 1, ..., n$, and that the n observations are independent. If the least square estmation method is used, are b_0 and b_1 are unbiased?