TUTORIAL 3

1. Refer to the **Grade point average** problem (see tutorial 1)

   (a) obtain a 95% percent interval estimate of the mean freshman GPA for students whose ACT test score is 28. Interpret your confidence interval.

   (b) Mary Jones obtained a score of 28 on the entrance test. Predict her freshman GPA using a 95% prediction interval. Interpret your prediction interval.

   (c) Is the prediction interval in part (b) wider than the confidence interval in part (a)? Should it be.

2. Refer to the **Airfreight breakage** problem (see tutorial 2)

   (a) When $X = 2$ and 4, find 99% confidence intervals for each of X, interpret your results

   (b) with $X = 2$ find the 99% prediction interval. interpret your interval

3. An analyst fitted normal error regression model and conducted an F-test of $\beta_1 = 0$ versus $\beta_1 \neq 0$. The P-value of the test was 0.033 and the analyst concluded $H_a : \beta_1 \neq 0$. Was the α level used by the analyst greater than or smaller than 0.033? If the α level had been 0.01, what would be the appropriate conclusion?

4. For conducting statistical tests concerning the parameter β_1, why is the t test more versatile than the F-test?

5. Refer to the **Grade point average** problem (see tutorial 1)

   (a) Set up the ANOVA table

   (b) conduct an F-test for $H_0 : \beta_1 = 0$ with $\alpha = 0.01$

   (c) what is the absolute magnitude of the reduction in the variation of Y when X is introduced into the model? what is the relative reduction? what is the name for the later measure?

   (d) obtain r_{XY} and attach the appropriate sign

   (e) which measure R^2 or r has more clear-cut operational interpretation, explain.

6. Refer to the **Airfreight breakage** problem (see tutorial 2)

(a) set up the ANOVA table. which elements are additive?

(b) conduct F-test to check the linear association between X and Y with $\alpha = 0.05$

(c) conduct t-test to check the linear association between X and Y with $\alpha = 0.05$. Compare with the F-test and show the equivalence.

(d) Calculate R^2 and r_{XY}. What proportion of the variation in Y is accounted for by introducing X into the model?

7. **Muscle mass** A person's muscle mass is expected to decrease with age. X : age and Y : muscle mass. 15 women were selected and **(data)** are recorded.

   (a) conduct a test to decide whether or not there is a negative linear association between amount of muscle mass and age with $\alpha = 0.05$. What is the P-value of the test?

   (b) The two-sided P-value for the test whether $\beta_0 = 0$ is 0.000001. Can it now be concluded that b_0 provides relevant information on the amount of muscle mass at birth for a females child?

   (c) Estimate a 95% confidence interval for the difference in expected muscle mass for women whose ages differ by one year.

   (d) Obtain a 95% confidence interval for the mean muscle mass for women of age 60. Interpret your confidence interval.

   (e) set up the ANOVA table

   (f) what proportion of the total variation in the muscle mass remain "unexplained" after age is introduced into the model? is that proportion relatively small or large?

8. Find $Var(b_0 + 2b_1)$

9. Suppose we have observations $(X_1, Y_1), ..., (X_n, Y_n)$ and consider two models

$$Y_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$$

and

$$X_i = \beta_0 + \beta_1 Y_i + \epsilon_i$$

(a) Find the LSE estimators for α_0, α_1 and β_0, β_1, denoted by a_0, a_1 and b_0, b_1 respectively

(b) check whether $a_1 = 1/b_1$ as in a mathematical model (i.e. $\varepsilon_i \equiv 0$ or $\epsilon_i \equiv 0$? What is the relation a_1 and b_1?

(c) what if the data satisfies (after standardization) $\sum_{i=1}^n (X_i - \bar{X})^2 = 1$ and $\sum_{i=1}^n (Y_i - \bar{Y})^2 = 1$?