

考試科目	統計學	所(組)別	統計學系	考試時間	103 年 11 月 2 日 星期日 10:00-11:40
------	-----	-------	------	------	-----------------------------------

請注意，每一大題的作答分別寫在一份答案本上。答案紙上請清楚標示各大題及小題的題號，無法辨識者，以零分計算。每一大題皆為 **25 分**，各小題配分以 **pts** 表之。

第一大題

- (6 pts)** For a data set with 100 observations, mean=60, standard deviation=10,
  - If the histogram of the data set is bell shaped, approximately how many observations will fall in the interval (45, 80).
  - If the histogram of the data set is unknown, approximately how many observations will fall outside the interval (44, 76).
- (15 pts)** For the following data set:  
9 28 15 21 12 22 29 20 23 31 11 36 19 24 16 13
  - Compute the interquartile range (IQR), coefficient of variation, and the 85<sup>th</sup> percentile.
  - Draw the box plot and determine if there are any outliers.
- (4 pts)** State the condition when it is appropriate to use (or draw)
  - Coefficient of variation
  - Line chart

第二大題

- Consider independent trials of rolling a pair of fair dice and denote the outcome to be the sum of two dice.
  - (5 pts)** Let  $E_n$  denote the event that no 4 or 5 appears on the first  $(n-1)$  trials and a 4 appears on the  $n$ th trial.  
Then  $P(E_n) = ?$
  - (5 pts)** What is the probability of the event that “an outcome of 4 appears before an outcome of 5”?  
(Hint: You can write down the event in terms of  $E_n$ .)

考試科目	統計學	所(組)別	統計學系	考試時間	103 年 11 月 2 日 星期日 10:00 - 11:40
------	-----	-------	------	------	-------------------------------------

2. Consider a series of independent Bernoulli trials of which the probability of getting a “success” is  $p$  and probability of getting a “failure” is  $1 - p$ . Answer the following questions.
- (a) (5 pts) Let  $X$  denote the number of trials needed for getting the first failure, what is the probability mass function of  $X$ ? (also specify the correct range of  $X$ )
- (b) (5 pts) Let  $X$  denote the number of successes over  $n$  trials, what is the probability mass function of  $X$ ? (also specify the correct range of  $X$ )
- (c) (5 pts) Let  $X$  denote the number of trials needed for getting  $r$  successes, what is the probability mass function of  $X$ ? (also specify the correct range of  $X$ )

### 第三大題

Suppose that a grocery store sells two brands of rice, Brand A and Brand B, and the rice is packaged in 2 kg bags for both brands. Suppose that a customer purchased 45 bags of rice from the store, out of which 25 are of Brand A and 20 are of Brand B. For the 25 Brand A bags, the the sample mean and standard deviation of the bag weights are 1.98 kgs and 1.10 kgs respectively. For the 20 Brand B bags, the the sample mean and standard deviation of the bag weights are 2.32 kgs and 1.20 kgs respectively. Suppose that for Brand A, rice bag weights (in kilograms) are normally distributed with mean  $\mu_1$  and standard deviation  $\sigma_1$ , and for Brand B, rice bag weights (in kilograms) are normally distributed with mean  $\mu_2$  and standard deviation  $\sigma_2$ .

- (a) (6 pts) Suppose that it is known that  $\sigma_1 = 0.001$ . Based on the sample of bag weights, can we conclude that  $\mu_1 < 2$  at level 0.05? Show your work.
- (b) (6 pts) Suppose that  $\sigma_1$  is unknown. Based on the sample of bag weights, can we conclude that  $\mu_1 < 2$  at level 0.05? Show your work.
- (c) (6 pts) Construct a 95% confidence interval for  $\mu_1$  assuming  $\sigma_1$  is unknown.
- (d) (7 pts) Suppose that both  $\sigma_1$  and  $\sigma_2$  are unknown and  $\sigma_1 = \sigma_2$ . Can we conclude that  $\mu_1 \neq \mu_2$  at level 0.05? Show your work.

### 第四大題

The objective of a study was to find out whether there is an association between instructor's appearance and the scores of course evaluations. Given the data below:

- (a) (1 pt) State the null hypothesis in symbols.
- (b) (10 pts) Complete the given ANOVA summary table.
- (c) (1 pt) What is your conclusion?

考試科目	統計學	所(組)別	統計學系	考試時間	103 年 11 月 2 日 星期日 10:00 - 11:40
------	-----	-------	------	------	-------------------------------------

Total number of subjects 448

Standard deviation for the 448 observations = 20

Instructor's appearance	Number of students	Mean scores of evaluations
Very pretty/handsome	169	41
Pretty/handsome	209	40
Normal	44	30
Ugly	26	35
Total	448	39

**ANOVA summary table**

Source of variation	Sum of Squares	DF	Mean Square	F	Approximate P value
Between					
Within					
Total					

- (d) (1 pt) Suppose that the study showed that several factors appear to be related to evaluation scores, including beauty, age, gender, tenure track and ethnicity. We constructed a correlation table for the 5 variables, and found out that the correlation between “beauty” and “age” and between “gender” and “beauty” are very strong. What kinds of problem should you consider when fitting a linear regression?
- (e) (2 pts) Sample size here is larger than 30, would we need to assume that the population values of dependent variable are normally distributed? Explain.
- (f) (9 pts) Suppose we decided to fit a simple linear model based on 10 results to study the relationship between beauty and evaluation score. A sample of 10 instructors had an average evaluation score of 15.2, with a standard deviation of 4.638. The average beauty result was 1.625, with a standard deviation of 0.757. In addition,  $\sum_{i=1}^n e_i^2 = 41.552$ , and the sample correlation coefficient was 0.886. Construct a 95% confident interval for the slope (5 pts), and interpret your interval in terms of the problem (2 pts). How much variance in the beauty is unexplained by the linear relationship with evaluation scores? (2 pts)
- (g) (1 pt) If we reject the null hypothesis that  $H_0: \beta_1 = 0$ , in general we can conclude there is a causal relationship between scores and beauty (i.e beauty cause higher evaluation scores) in linear regression. (True/False)