1. (10%) Let X_1 and X_2 have the joint probability density function given by

$$f(x_1,x_2)=\begin{cases} Kx_1x_2 & 0\le x_1\le 1, 0\le x_2\le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

   (a) Find the value of K that makes this a probability density function.
   (b) Find the marginal densities of X_1 and X_2.
   (c) Find the joint distribution for X_1 and X_2.
   (d) Find the probability $P(X_1 < 1/2, X_2 < 3/4)$.
   (e) Find the probability $P(X_1 \le 1/2 \mid X_2 > 3/4)$.

2. (10%) Suppose X takes on the values 0, 1, 2, 3, 4, 5 with probabilities P_i, $i = 0, \dots, 5$, respectively. If $Y = g(X) = (X-2)^2$, what is the distribution of Y?

3. (10%) Find the expected values of the random variables X and Y if $P(X = 0) = 1/2$, $P(X = 1) = 1/2$, $P(Y = 1) = 1/4$, and $P(Y = 2) = 3/4$. Compare $E(X) + E(Y)$ with $E(X+Y)$, if $P(X=x, Y=y) = P(X=x)P(Y=y)$.

4. (10%) Let X_1 and X_2 be the random independent samples from normal distribution $N(0, 1)$. What is the distribution of $(X_2 - X_1)/\sqrt{2}$?

5. (10%) A college officer wants to estimate the proportion of students who favor a new policy in a college. Suppose there are only two possible answers, i.e., "yes" or "no", for the survey. The estimate is to be within 0.05 of the true proportion favoring the new policy, with a 90% confidence coefficient. Determine the largest sample size necessary to ensure the desired results. $(z_{0.05} = 1.645, z_{0.025} = 1.96)$

（背面仍有題目,請繼續作答）

6. (21%) Explain the following terms.
   (a) (3%) Contingency table
   (b) (3%) p-value
   (c) (3%) F_{α} in F test
   (d) (3%) Randomized block design
   (e) (3%) Coefficient of determination
   (f) (3%) Autocorrelation
   (g) (3%) Kruskal-Wallis test
7. (29%) Problem set. Suppose we want to evaluate whether four automatic assembly machines A, B, C, and D have different production quantities per day by deploying ANOVA. A set of hypotheses to be completed is as follows:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_1 : \_\_\_\_\_\_\_\_$$

   (a) (2%) What are the meanings of $\mu_A, \mu_B, \mu_C, \mu_D$ in the above incomplete set?
   (b) (2%) What shall the alternative hypothesis be in the above incomplete set?

Under a completely randomized design with each treatment having 5 observations, we construct the ANOVA table in order to test the hypotheses.

   (c) (3%) What are the degrees of freedom for all sources of variation in the ANOVA table?

Alternatively, we can use multiple regression to test the above hypotheses. The following three questions are related to this approach.

   (d) (8%) Write down the multiple regression equation in a form similar to the following

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_i x_i + \cdots + \beta_p x_p$$

   Note that to answer (d), you shall specify the number of the independent variables, the meaning and characteristics of each independent variable x_i and the dependent variable y.

   (e) (6%) Which test shall be used? And what are the hypotheses for this test?
   (f) (8%) Reason that the approach adopted in (e) achieves the same result as the ANOVA does.