本試題是否可以使用計算機： ☑可使用 ，☐不可使用 （請命題老師勾選）

1. (10%) Suppose a random variable X has pdf $f_X(x)$, CDF $F_X(x)$, and mean $E(X)$, and a random variable Y has pdf $f_Y(y)$, CDF $F_Y(y)$, and mean $E(Y)$. Both random variables are distributed over the interval bounded by a and b, where $a<b$. (No credit will be given if the answer is partially correct.)
   a. (5%) Which of the following statements would be true? (A) $F_X(a)=F_Y(a)$ (B) $F_X(b)=F_Y(b)$ (C) $f_X(x)$ could be less than zero for some $x, a\le x\le b$. (D) If $f_X(x_1)<f_X(x_2)$ for all $a\le x_1<x_2\le b$, then $F_X(x_1)<F_X(x_2)$ for all $a\le x_1<x_2\le b$. (E) If $f_X(t)\le f_Y(t)$ for all $t\in[c,d], a\le c<d\le b$, then $F_X(t)<F_Y(t)$ for all $t\in[c,d]$.
   b. (5%) Which of the following statements would be true? (A) $\int_a^b (x-E(X))f_X(x)dx$ could be less than zero. (B) If $E(X)\le E(Y)$, $f_X(x)\le f_Y(y)$ for all $x\le y$. (C) If the median, m_X, of X is less than or equal to the median, m_Y, of Y, then $\int_{m_Y}^b f_X(x)dx\le\int_{m_X}^b f_Y(y)dy$. (D) If X and Y are (statistically) independent, then the joint density of X and Y is $f_X(x)f_Y(y)$. (E) If X is exponential, then the variance of X increases as $E(X)$ increases.
2. (40%) The instructor is aware that the students were not paying attention to the homework assignments (HAs) that can help students understand not only the concepts covered in the course but also the process required for solving problems. The instructor thus decides to include several homework assignment questions (HAQs), totaling a points, in the midterm examination. The instructor believes that in the midterm examination the students who really did the HAs will answer these HAPs well and obtain X points with pdf $f_X(x)=(x-c)/200$ over the interval bounded by c and a, where $c<a$, and answer the remainder of the questions, totaling b points, relatively well and obtain Y points with pdf $f_Y(y)=1/20$ over the interval bounded by d and b, where $d<b$. The instructor also believes that those who did not do HAs themselves will answer these HAQs poorly and obtain U points with pdf $f_U(u)=(2/a)(1-u/a)$ over the interval bounded by g and a, where $g<a$, and will answer the remainder of the questions relatively poorly and obtain V points with pdf $f_V(v)=1/(b-15)$ over the interval bounded by h and $b-15$, where $h<b-15$. Note that $a+b=100$.
   a. (10%) From the instructor's viewpoint, what is the expected score of the midterm examination for a student who really did the HAs?
   b. (15%) Suppose that the instructor sets $a=60$ and that she has the feeling that 40% of the students did the HAs themselves and 60% of the students did not. If a student obtains no more than 45 points for the HAQs in the midterm examination, what is the probability that this student really did the HAs himself?
   c. (15%) Suppose that the instructor sets $a=60$. What would the midterm examination scores of the students who did not do the HAs themselves look like? If a student did not do the HAs herself, what is the probability that her midterm examination score is less than or equal to 60?

（背面仍有題目，請繼續作答）

3. (20%)Multiple choice problem---ONLY ONE answer for each of the following problems.

a. (4%)Which one is NOT the assumptions of the regression:
   A) The error terms ε_i are normally distributed $N(0, \sigma^2)$.
   B) $cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j,\ i, j = 1,2,\ldots,n$
   C) The independent variables and dependent variable are independent.
   D) The independent variables and dependent variable are observed by pairs.

b. (4%)Which one is NOT true about the ANOVA table:
   A) One does not need the information on when the data are obtained.
   B) The degree of freedom is determined once the number of observation is chosen.
   C) The test statistic, F_0, does need the information on how many factors are used.
   D) The ANOVA table can be obtained from the computer output if one is familiar with the statistical package.

c. (4%)Which one is NOT true about the degree of freedom:
   A) The degree of freedom is the actual amount of variability in the distribution
   B) The mean sum of square error is equal to the sum of square error divided by the degree of freedom
   C) In the F distribution with α=0.3, The F_α is getting larger as the degree of freedom of numerator is increased when the degree of freedom of denominator is unchanged.
   D) The degrees of freedom in both numerator and denominator of the F distribution are not interchangeable.

d. (4%)Which one is NOT true about the design of experiment(DOE):
   A) We use ANOVA to compare the variances when there are more than two levels of a single factor.
   B) The role of randomization in the DOE is very important.
   C) An interaction effect is the case that the difference in response between levels of one factor is not the same at all levels of the other factors.
   D) Factorial experiments are the ways to discover interactions between variables.

e. (4%)Which one is NOT true about the nonparametric statistics:
   A) Nonparametric statistics are the distribution free methods.
   B) Generally, we will use the parametric methods if some distributions can be fitted.
   C) "rank" and "sign" are the popular ways to process data in nonparametric statistics.
   D) We can easily test the hypothesis without using any table in Wilcox Rank-Sum test.

本試題是否可以使用計算機： ☑可使用 ， □不可使用 （請命題老師勾選）

4. (16%)NCKU convenient store wants to use regression analysis to build a model for purchasing amount, y, of a customer. Two variables thought to effect the purchasing amount are customer's length of time in the store and time section. These variables are described below:

   y= purchasing amount (NT$)
   x_1= Length of time in store(minutes)
   x_2= 1 if morning section(0600-1200), 0 if not
   x_3= 1 if afternoon section(1200-1800), 0 if not (Base level = evening section(1800-2400))

   Data for 12 customers on the NCKU convenient store were collected and used to fit the interaction model:

   $$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$$

   a. (3%)What is the experimental unit?
   b. (4%)What null hypothesis would you test to determine whether the slope of the linear relationship between the purchasing amount (y) and the time in the store (x_1) varies according to section time?
   c. (6%)Sketch the theoretical relationship between the variables hypothesized by the model?
   d. (3%)In term of β's in the model, give the change in the purchasing amount (y) for every 1 minute increased in time in store (x_1) when coming in the evening section.

5. (14%)The table shows partially completed ANOVA table for a two-factor factorial experiment.

Source	df	SS	MS	F
A	3	2.6	?	?
B	5	9.2	?	?
AXB	?	?	3.1	?
Error	?	18.7		
Total	47	?		

   a. (9%)Complete the ANOVA table.
   b. (5%) Please give the appropriate inference if F_α is given.