SDSC2102 Statistical Methods and Data Analysis
Semester B 2021/22

Final Exam
Part II
(5 problems, 50 points)

- Write your answers on blank A4 papers. On the first page write your name and student ID.
- Responses should be clearly labelled and in order.
- Show all your work in detail to get credits. Answers without supporting work may be severely downgraded. Answers should be clear and easy to follow. No effort will be made to decipher unclear steps.

1. (10 points) The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable *X* that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \leq x < 2 \\ 0, & otherwise \end{cases}$$

(a) (6 points) What is the CDF of *X*?

(b) (4 points) What is the probability that a family runs their vacuum cleaner less than 120 hours?

2. (8 points) Let *A* and *B* be two independent events. Please prove that A^C and *B* are independent.

3. (10 points) It is given that *X* is a positive-valued random variable, which means that *X* does not take on negative values. It is also known that

$$Var[X] = 24$$
$$E[(X-2)^2] = 40$$

Calculate $E[-5X + 30]$.

4. (10 points) A well-known sprinter ran 22 races in the years 1972-1974 and 20 races in the years 1975-1977. Between the 1974 and 1975 track seasons, the sprinter contracted a rare tropical disease and lost his luxurious head of hair. The sprinter's coach contends that due to decreased wind resistance, the sprinter ran significantly faster after the disease than before. The sample mean for his run times in the years 1972-1974 is 9.83 seconds. For the years 1975-1977, the sample mean is 9.67 seconds. The pooled sample standard deviation for the run times from all years is 0.3 seconds.

(a) (2 points) State the appropriate hypotheses to test the contention of the sprinter's coach.

(b) (8 points) Conduct the test at the 0.01 level of significance. What do you conclude?

5. (12 points) In a certain type of metal test, the normal stress (X) on a specimen is known to be functionally related to the shear resistance (Y). Suppose you have collected an experimental dataset of n samples on the two variables, and want to build a linear regression model using the data.

(a) (8 points) Let the data be $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, and the slope and intercept of the linear regression model be m and b. Please derive the least squares estimates of m and b.

(b) (4 points) Given the dataset below

Sample	Normal Stress (X)	Shear Resistance (Y)
1	2.8	5.5
2	1.3	2.5
3	3.5	7.1
4	2.5	4.8
5	3.0	6.2

Find the regression line using the formulas obtained in (a).

-End-