Midterm *2020. 11. 10.*

1. (50%)

(a) (40%) Consider a sample with data values of

7	5	5	9	4

Compute the following.

(i) The variance and the coefficient of variation.

(ii) The 30th percentile and 65th percentile.

(iii) The box plot.

(iv) Can the empirical rule be applied to this data? Explain.

(b) (10%) Suppose the data have a bell-shaped distribution with a mean of 7 and a standard deviation of 2.

(i) At least what percentage of data will have a value falling $[1, 13]$?

(ii) Determine the range within which contains approximately 68% of data.

2. (20%) Suppose we have the following data:

30	78	59	65	40	64	52	53	57
39	61	47	50	60	48	50	58	67

Suppose the number of non-overlapping classes is determined to be 4.

(a) Construct a frequency distribution and cumulative percent frequency distribution.

(b) Based on the result of (a), compute the grouped mean.

3. (30%)

(a) (10%) The following data are for 20 observations on two qualitative variables: Majors (A: Accounting; B: Statistics; C: Management) and whether taking some test (Y: taking the test; N: not taking the test).

Observation	1	2	3	4	5	6	7	8	9	10
Majors	C	B	C	A	B	A	B	C	C	C
Taking Test	Y	Y	Y	N	Y	N	Y	N	Y	Y
Observation	11	12	13	14	15	16	17	18	19	20
Majors	C	B	C	A	B	C	C	A	B	B
Taking Test	N	Y	Y	Y	Y	N	Y	N	Y	Y

Develop a cross-tabulation for the data.

(b) (20%) For the following two samples of data, $(x_i, y_i), i = 1, \cdots 5$,

Sample 1: x_i	1	0	-1	-2	-3
Sample 2: y_i	1	1	2	2	4

(i) Give the scatter diagram. (5%)

(ii) Compute and interpret the sample correlation coefficient. (15%)

4. (25%) You are given the following information on Events *A, B,* and *C.*

$$P(A) = 0.4, P(A \cap B) = 0.16, P(B) = 0.4,$$
$$P(A|C) = 0.2, P(C) = 0.65.$$

(a) Compute $P(A \cup B^c)$.

(b) Compute $P(A \cap C)$.

(c) Compute $P(B^c|A)$.

(d) Are *A* and *B* independent? Explain your answer.

(e) Are *B* and *C* mutually exclusive? Explain your answer.

5. (10%) You travel from country A to country B, then country B to country C. The probabilities that your luggage is lost at different countries are 20% at country A and 25% at country B. Given that your luggage is lost as reaching country C, which country or countries it was most likely lost at? (Hint: Using conditional probabilities).

Equations:

$$s_{XZ} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(z_i - \overline{z})}{n-1}, r_{XZ} = \frac{s_{XZ}}{s_X s_Z},$$

$$\overline{x}_g = \frac{\sum_{k=1}^{m} f_k M_k}{n}, s_g^2 = \frac{\sum_{k=1}^{m} f_k \left(M_k - \overline{x}_g\right)^2}{n-1}.$$