Statistics I, Fall 2012
Midterm Exam

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

Name: ____________________ Student ID: ____________________

1. (10 points; 1 point each) For the following statements, CIRCLE or BOX "T" if one is true or "F" if it is false. DO NOT provide any explanation.

   (a) T F The height of a student is of the ordinal level.

   (b) T F The income of a household is of the quantitative level.

   (c) T F In studying the average lifetime of digital cameras produced by a manufacturer, 100 cameras produced by that manufacturer form a sample and the average lifetime of the 100 cameras is a parameter.

   (d) T F $\mathbb{E}\big[\sum_{i=1}^{n} X_i\big] = \sum_{i=1}^{n} \mathbb{E}[X_i]$ if and only if X_is are independent.

   (e) T F $\mathbb{E}\big[\prod_{i=1}^{n} X_i\big] = \prod_{i=1}^{n} \mathbb{E}[X_i]$ if and only if X_is are independent.

   (f) T F A uniformly distributed random variable may take negative value.

   (g) T F The sum of two binomial random variables is a binomial random variable.

   (h) T F A conditional probability is a ratio of two probabilities. Therefore, it is possible for a conditional probability to be greater than 1.

   (i) T F For two random variables, zero covariance implies independence.

   (j) T F For two random variables, independence implies zero covariance.

2. (5 points) Let $X \sim \text{ND}(\mu, \sigma)$ be a normal random variable with mean μ and standard deviation σ. Consider

$$Y = \frac{|X - \mu|}{\sigma}.$$

   What is the probability density function of Y?

3. (12 points) Consider two sets of population data $\{x_i\}_{i=1,\ldots,8}$ and $\{y_i\}_{i=1,\ldots,8}$ whose values are listed in the table below:

i	1	2	3	4	5	6	7	8
x_i	10	12	7	9	16	2	8	16
y_i	14	11	15	11	10	5	7	14

(a) (4 points) Draw a scatter plot for $\{(x_i, y_i)\}_{i=1,\ldots,8}$.

(b) (4 points) Calculate the standard deviations for $\{x_i\}_{i=1,\ldots,8}$ and $\{y_i\}_{i=1,\ldots,8}$, respectively.

(c) (4 points) Calculate the covariace and correlation coefficient for $\{(x_i, y_i)\}_{i=1,\ldots,8}$. Then make a conclusion on the correlation between the two sets of data.

4. (5 points) In a survey, each of the 1000 consumers is asked which color she/he would choose for her/his laptop. The survey data are summarized in the table below:

Color	Black	White	Red	Yellow	Blue	Green
Frequency	270	217	86	125	176	126

(a) (1 points) Are these data nominal, ordinal, interval, or ratio?

(b) (2 points) Draw a bar chart for this set of data.

(c) (2 points) Draw a Pareto chart for this set of data.

5. (13 points) A set of 100 values has been sampled from a population. These values are given below:

0.83	0.72	1.12	5.07	2.35	1.35	0.13	1.40	4.01	1.39
0.49	4.84	2.43	2.05	0.05	0.18	1.36	3.20	0.66	0.23
0.91	7.69	2.52	0.92	3.08	1.77	0.76	0.06	4.46	0.44
1.01	4.99	0.29	1.26	1.73	1.79	3.43	3.42	1.52	0.67
1.11	2.52	2.63	2.82	0.68	1.98	3.22	0.41	0.65	0.86
3.01	0.11	0.16	1.20	1.40	0.27	1.94	3.82	0.60	2.40
0.24	0.14	0.50	1.45	0.39	3.50	0.56	5.11	1.00	0.90
0.14	1.41	0.49	0.62	2.58	1.90	1.78	1.17	0.31	1.81
3.76	2.09	2.77	0.11	1.89	1.37	1.65	2.26	0.92	5.26
2.92	4.00	4.53	0.01	0.94	0.02	1.39	0.58	1.14	4.89

(a) (3 points) Fill in blanks in the following frequency distribution of these data:

Class	Frequency	Relative frequency	Cumulative frequency
$[0, 1)$			
$[1, 2)$			
$[2, 3)$			
$[3, 4)$			
$[4, 5)$			
$[5, 6)$			
$[6, 7)$			
$[7, 8)$			

(b) (3 points) Draw a histogram based on the first two columns of the table in Part (a).

(c) (3 points) Calculate the grouped mean and variance based on the table in Part (a).

(d) (1 point) Are these data skewed to the left, symmetric, or skewed to the right?

(e) (3 points) It seems that these data follow an exponential distribution. To verify this, I calculated the the ungrouped mean of the 100 values, which is 1.77. Compare the ungrouped mean (1.77) and the grouped mean you calculated in Part (d). Then explain whether the relationship between them supports the conjecture that the data follow an exponential distribution.

6. (5 points) Consider an arrival process for which the interarrival time follows an exponential distribution with mean 10 minutes per arrival.

   (a) (2 points) What is the probability that in an hour there are at least 10 arrivals?

   (b) (2 points) What is the probability that in an hour there are no more than 5 arrivals?

   (c) (1 point) What is the expected number of arrivals in 30 minutes?

7. (10 points) Let $X_1 \sim \mathrm{Bi}(10, 0.5)$ be a binomial random variable with size 10 and probability 0.5 and $X_2 \sim \mathrm{Bi}(10, 0.4)$ be a binomial random variable with size 10 and probability 0.4. Suppose X_1 and X_2 are independent.

   (a) (2 points) Calculate $\Pr(X_1 + X_2 = 0)$.

   (b) (3 points) Calculate $\Pr(X_1 X_2 = 0)$.

   (c) (1 points) Calculate $\mathrm{Var}(X_1 + X_2)$.

   (d) (4 points) Calculate $\Pr(X_1 + X_2 = 4)$.

8. (5 points) Let $X \sim \text{Exp}(\lambda)$ be an exponential random variable with rate λ. Prove that

$$\Pr(X > t + s | X > t) = \Pr(X > s).$$

9. (3 points) A set of 30 values has been sampled from a population. These values are ordered and given below:

5	7	8	15	15	18	25	34	35	38
42	47	47	51	55	61	62	62	64	66
70	82	83	87	87	89	95	96	96	99

(a) (1 point) Find the median of these 30 values.

(b) (2 points) Find the first and third quartiles of these 30 values.

10. (7 points) Consider an experiment whose outcome may be 1, 2, or 3. Let $\{X_i\}_{i=1,...,10}$ be ten such experiments and assume X_i and X_j are independent for all $i \neq j$. For X_i, $i = 1, ..., 10$, we have $\Pr(X_i = 1) = 0.5$, $\Pr(X_i = 2) = 0.1$, and $\Pr(X_i = 3) = 0.4$. Let Y_j be the number of outcome j observed in the ten trials, $j \in \{1, 2, 3\}$. Naturally $Y_1 + Y_2 + Y_3 = 10$.

(a) (2 point) Calculate $\Pr(Y_1 = 3)$.

(b) (2 point) Calculate $\Pr(Y_2 = 2)$.

(c) (3 points) Calculate $\Pr(Y_3 = 2|Y_1 = 2)$.

11. (8 points) At an intersection, the number of car accident in a week follows a Poisson distribution with rate λ accidents per week. As the rate λ is unknown, our prior belief is $\Pr(\lambda = 1) = \frac{1}{3}$, $\Pr(\lambda = 2) = \frac{1}{3}$, and $\Pr(\lambda = 3) = \frac{1}{3}$.

    Note. You may use $e^{-1} = 0.368$, $e^{-2} = 0.135$, and $e^{-3} = 0.05$.

    (a) (3 points) Suppose we have observed two accidents in the first week. Calculate the posterior belief on λ (i.e., find the posterior probabilities of $\lambda = 1$, 2, and 3, respectively).

    (b) (5 points) Suppose we have observed two accidents in the first week and no accident in the second week. Calculate the posterior belief on λ.

12. (12 points) The US national average door-to-doctor waiting time for patients to see a doctor is now 25 minutes. Suppose such waiting times are normally distributed with a standard deviation of 7 minutes.

    Note. For Parts (a) to (c), you MUST specify units of measurements.

    (a) (1 point) What is the coefficient of variation of the waiting times?

    (b) (1 point) What is the median of the waiting times?

    (c) (1 point) What is the variance of the waiting times?

    (d) (2 points) What is the probability that a patient must wait for less than 20 minutes?

    (e) (2 points) What is the proportion of patients whose waiting time is between 23 and 33 minutes?

    (f) (2 points) Suppose 3% of patients must wait for more than x minutes. Find x.

(g) (3 points) Use Chebyshev's theorem to find the lower bound of proportion of patients whose waiting times are between 15 minutes and 35 minutes.

13. (5 points) Consider five events A, B, C, D, and E and the joint probability table

	C	D	E
A	0.2	0.1	0.2
B	0.4	0.1	0

(a) (2 points) Calculate $\Pr(C \cup A)$.

(b) (2 points) Calculate $\Pr(A|E)$.

(c) (1 point) Calculate $\Pr(A)$.

14. (10 points) Consider a random variable X with the following probability density function

$$f(x) = \begin{cases} x^2 & \text{if } 0 \le x \le 1 \\ \frac{2}{3} & \text{if } 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}.$$

(a) (3 points) Calculate the cumulative distribution function $\Pr(X \le x)$.

(b) (3 points) Calculate the expectation $\mathbb{E}[X]$.

(c) (4 points) Calculate $\mathbb{E}\left[\left|X - \frac{3}{2}\right|\right]$.