Statistics I, Fall 2012
Final 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 A Pareto chart is a bar chart but a bar chart may not be a Pareto chart.

   (b) T F The Poisson distribution is discrete. Because its mean and variance are identical, it is symmetric.

   (c) T F The linear combination of two independent normal random variables is normal.

   (d) T F If the population is normal, the sample size is small, and the population variance is unknown, we may use the t test to test the population mean.

   (e) T F If the population is normal, the sample size is large, and the population variance is known, we may use the t test to test the population mean.

   (f) T F In estimating the population variance, the sample variance locates at the center of the confidence interval.

   (g) T F In testing the population mean, if the p-value becomes smaller, the difference between the true population mean and its hypothesized value becomes more significant.

   (h) T F Increasing the sample size will always decrease the two probabilities of making the Type I and Tpye II errors at the same time in expectation.

   (i) T F When we do not reject H_0 in a test whose significance level is α, the probability that H_a is true is α.

   (j) T F When we reject H_0 in a test whose significance level is α, the probability that H_0 is true is α.

2. (10 points) Let $X_1 \sim \text{Bi}(8, 0.2)$ be a binomial random variable with size 8 and probability 0.2 and $X_2 \sim \text{Bi}(6, 0.1)$ be a binomial random variable with size 6 and probability 0.1. 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 $\text{Var}(X_1 + X_2)$.

   (d) (4 points) Calculate $\Pr(X_1 + X_2 = 3)$.

3. (10 points) As a retail store owner, you order products and sell them to consumers. When you order product from the manufacturer, the delivery time is random. If the shipping company is strong, the delivery time X (in days) will satisfy $\Pr(X = 1) = 0.5$, $\Pr(X = 2) = 0.3$, $\Pr(X = 3) = 0.1$, and $\Pr(X = 4) = 0.1$. If the shipping company is weak, however, X will satisfy $\Pr(X = 1) = 0.2$, $\Pr(X = 2) = 0.2$, $\Pr(X = 3) = 0.3$, and $\Pr(X = 4) = 0.3$. Unfortunately, you do not know whether the shipping company is strong or weak. At this moment, you believe that it is strong with probability 40% and weak with probability 60%.

   (a) (3 points) Suppose after you place your first order, the products arrive two days later. How would you adjust your belief on whether the shipping company is strong or weak?

   (b) (3 points) Continue from Part (a). Suppose after you place your second order, the products again arrive two days later. How would you adjust your belief on whether the shipping company is strong or weak?

   (c) (4 points) Ignore Parts (a) and (b). Suppose in a month in which another retail store owner places six orders, four orders are fulfilled in one day but two orders are fulfilled in more than one day. For the two orders not fulfilled in one day, you do not know whether they are fulfilled in two, three, or four days. Use the given information to update your belief on whether the shipping company is strong or weak.

4. (10 points) Suppose a random variable X has the following probability mass function:

$$\Pr(X = 1) = 0.2, \quad \Pr(X = 2) = 0.3, \quad \text{and } \Pr(X = 3) = 0.5.$$

Consider a random sample of sample size $n = 3$ with replacement.

(a) (2 points) Let $\overline{X}$ be the sample mean. Find $\mathbb{E}\big[\overline{X}\big]$.

(b) (2 points) Let $\widehat{P}$ be the sample proportion of 1. Find $\mathrm{Var}\big(\widehat{P}\big)$.[1]

(c) (5 points) Let M be the sample median. Find the sampling distribution of M.

(d) (1 points) Continue from Part (c) and find $\mathbb{E}[M]$.

[1]For example, if the sample is $(1, 1, 3)$, the sample proportion of 1 is $\frac{2}{3}$.

5. (10 points) Consider the experiment of rolling one fair dice. Let X be the outcome.

   (a) (1 point) What is the distribution of X?

   (b) (2 points) What is the moment generating function of X?

   (c) (2 points) Use the moment generating function to find $\mathbb{E}[X]$.

   (d) (3 points) Use the moment generating function to find $\mathrm{Var}(X)$.

   (e) (2 points) Use the moment generating function to find $\mathbb{E}[X^n]$ for $n \in \mathbb{N}$. Express your answer as a function of n.

6. (16 points) Jack Cheng from IBM was invited to give a talk to our class on December 26th. In the questionnaire distributed by him, there were questions "What is your general impression of X as a whole?", where X is Microsoft, Google, Oracle, or IBM. Among the 55 students in our class, 34 responded to the survey though not all of them answered all the four questions. Below are some statistics of the survey.

Company	Microsoft	Google	Oracle	IBM
Sample size	32	33	32	33
Sample mean	3.125	4.455	3.031	3.394
Proportion of 4 or above	0.406	0.970	0.375	0.485
Sample variance	0.887	0.318	0.934	1.059

Please use hypothesis testing to answer the following questions based on the above results and a 10% significance level.[2] Treat our class (the 55 students) as the population.

Note. For all questions in this problem, when you do hypothesis testing, you need to write down all the relevant elements, including the notations, hypothesis, selection of test, the conditions for using the test, calculations, conclusions, and interpretations.

(a) (4 points) Before Jack did the survey, one thing he wanted to know is whether the average rating for IBM is above or below 3.2. If above, he will try to write a proposal to his boss and suggest to aggressively promote an IBM's software into the IM department in NTU. If this is not the case, he will do nothing. Doing an aggressive promotion definitely costs a lot and is effective only if people in the IM department sufficiently like IBM (i.e., the average rating is above 3.2). Determine whether Jack should write the proposal.

[2] All the stories in these questions are artificially made by the instructor. They are made only for testing your Knowledge.

(b) (4 points) Jack also wanted to know whether the average rating for Microsoft is above or below 3.2. He wanted to know this because one of his friends is currently marketing Microsoft products. If he has a strong evidence that people in the IM department do not like Microsoft enough, he will be able to discourage his friend from doing promotion. Determine whether Jack can really discourage his friend.

(c) (4 points) Jack observed that 48.5% of people responding to the survey rate IBM as 4 or above. He wanted to know whether the proportion of people in our class rating IBM as 4 or above is below one half. Is this true or not?

(d) (4 points) Suppose Jack wanted to check whether the population variance of the ratings for Google is 0.3. Based on the given information in this problem, how should he conclude?

7. (8 points) Tim is given two sets of independent random sample data from two populations and he needs to test whether the mean of population 1, μ_1, is higher than that of population 2, μ_2. Though some methods in Chapter 10 of the textbook may be applied to test whether $\mu_1 > \mu_2$, Tim has no knowledge on them. For the two observed sample means, $\bar{x}_1$ for sample 1 and $\bar{x}_2$ for sample 2, there are some observations given below. Please help Tim make conclusions.

   Note. For all questions in this problem, if you want to write something that can be guaranteed, that must be relevant to the two population means and under the given significance level.

   (a) (2 points) Suppose $\bar{x}_1 > \bar{x}_2$, can he be 100% sure that $\mu_1 > \mu_2$? Why or why not?

   (b) (3 points) Suppose he applies hypothesis testing on μ_1 and shows that, with a 5% significance level, there is a strong evidence showing that $\mu_1 > \bar{x}_2$. In other words, he uses the observed sample mean for population 2, $\bar{x}_2$, as the hypothesized population mean. Can he be 95% sure that $\mu_1 > \mu_2$? If so, why? If not, what can he guarantee with a 95% confidence?

   (c) (3 points) Suppose he applies hypothesis testing and shows that, with a 5% significance level, there is a strong evidence showing that $\mu_1 > 100$. He then does another test on sample 2 and shows that, with a 5% significance level, there is a strong evidence showing that $\mu_2 < 100$. Can he be 95% sure that $\mu_1 > \mu_2$? If so, why? If not, what can he guarantee and how confident he can be?

8. (17 points) Consider a uniformly distributed population with lower bound a and upper bound $a+6$. You know the population is uniformly distributed but you do not know the value of a. Let μ be the population mean. Let $\overline{X}$ be the mean of a random sample whose sample size is $n = 2$.

   (a) (1 point) Express μ as a function of a.

   (b) (2 points) Write down the sampling distribution of the sample mean $\overline{X}$ as a function of a. For this part only, you are allowed to write down your answer with no calculation.
   Hint. The curve looks like a triangle. Use your intuition!

   (c) (4 points) Suppose the observed sample mean is $\bar{x} = 8$. Construct a two-tailed 90% confidence interval of μ centered at $\bar{x}$.

   (d) (3 points) Write down an unbiased estimator of a. Prove that it is indeed unbiased.

(e) (3 points) Let $\overline{Y}$ be the mean of a random sample whose sample size is 100. Write down the (approximated) sampling distribution of $\overline{Y}$ as a function of a.

(f) (4 points) Continue from Part (e). Suppose the observed sample mean is $\bar{y} = 8$. Construct a two-tailed 90% confidence interval of μ centered at $\bar{y}$. Is this interval smaller or larger than that found in Part (c)? Explain it by comparing the sample sizes.

9. (4 points) Consider the following hypothesis regard a population mean μ:

$$H_0 : \mu = 10$$
$$H_a : \mu > 10.$$

(a) (1 point) Suppose with a 5% significance level, we do not reject H_0. Clearly write down the complete concluding statement.

(b) (3 points) Suppose with a 5% significance level, we reject H_0. Based on this, one claims that there is a 95% probability for μ to be greater than 10. Is this claim true? If so, express this claim mathematically with a conditional probability. If not, explain why.

10. (6 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.

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

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

    (c) (2 points) Suppose 10% of patients must wait for more than x minutes. Find x.

11. (10 points) In a country, 10000 1-dollar coins have been used on the market for many years. One day a group of people making counterfeit (fake) money were caught. There were enough evidence indicating that they have made 3200 counterfeit 1-dollar coins. However, while 3000 counterfeit coins were proved to be distributed to the market, these criminals claimed that they have destroyed the other 200. It was known that if these 200 counterfeit coins were not destroyed, they must all be on the market. In other word, there were either 3000 or 3200 counterfeit 1-dollar coins currently on market. Let p be the proportion of 1-dollar coins on the market that were counterfeit. It was then clear that p is either $\frac{3000}{13000} = 0.2308$ or $\frac{3200}{13200} = 0.2424$. To discover whether the lost 200 counterfeit coins were really destroyed, the police conducted a statistical survey by randomly collecting 50 1-dollar coins from the market.

    (a) (2 points) One police claimed that the statistical study can be completed under each of the following two hypotheses

$$\begin{aligned} H_0 &: p = 0.2308 \\ H_a &: p \neq 0.2308 \end{aligned} \qquad \text{and} \qquad \begin{aligned} H_0 &: p \neq 0.2308 \\ H_a &: p = 0.2308. \end{aligned}$$

    Is he correct? Why or why not?

(b) (4 points) Suppose the police adopted the hypothesis

$$H_0 : p = 0.2308$$
$$H_a : p > 0.2308.$$

Among the 50 coins they collected, there were 12 counterfeit ones. With a 1% significance level, what conclusion should the police make through hypothesis testing?

(c) (4 points) For this test, what is the probability of making the Type II error?