

Part I: Multiple-Choice Questions. For each question, circle only one of the proposed choices. Each multiple-choice question has 3 marks

- In order to test $H_0 : \mu = 50$ vs $H_a : \mu \neq 50$, a random sample of 9 observations (from a normally distributed population) is obtained, yielding $\bar{x} = 61$ and $s = 21$. What is the P-value of the test?
(a) greater than 0.1 (b) between 0.05 and 0.10
(c) between 0.01 and 0.05 (d) less than 0.01.
- Which of the following is a consequence of the Central Limit Theorem?
(a) A large population will be normally distributed with mean μ and variance σ^2/n
(b) A large sample will be normally distributed with mean μ and variance σ^2/n
(c) The sample mean \bar{X} is approximately normally distributed with mean μ and variance σ^2/n
(d) The sample mean \bar{X} is approximately binomially distributed with mean μ and variance σ^2/n
- Suppose we repeat an experiment identically and independently 100 times. Each time we construct a 99% confidence interval for μ via the t-distribution. Let X = the number of times the confidence interval fails to contain the true value of μ . The distribution of X is
(a) Normal with $\mu=99$ and $\sigma^2=0.99$ (b) Normal with $\mu=0$ and $\sigma^2=1$
(c) Binomial with $n=100$ and $p=0.99$ (d) Binomial with $n=100$ and $p=0.01$
- Under which of the following conditions is the t-distribution used in statistical inference?
(a) The population standard deviation is not known and the sample is small.
(b) The population standard deviation is known.
(c) The sample is not random or the population distribution is strongly skewed.
(d) The sample size is at least 30.
- Which of the following randomly selected measurements, X , might be considered a potential outlier if it was selected from the given population?
(a) $X=0$ from a population with $\mu = 0$ and $\sigma = 2$
(b) $X=-5$ from a population with $\mu = 1$ and $\sigma = 4$
(c) $X=7$ from a population with $\mu = 4$ and $\sigma = 2$
(d) $X=4$ from a population with $\mu = 0$ and $\sigma = 1$
- After taking 90 observations, you construct a 90% confidence interval for μ . You are told that your interval is 3 times too wide (i.e., your interval is 3 times wider than what was required). Your sample size should have been
(a) 30 (b) 270 (c) 810 (d) 10
- In a histogram, the proportion of the total area which must be to the right of the mean is
(a) Less than 0.50 if the histogram is skewed to the left
(b) Exactly 0.50 (always)
(c) Exactly 0.50 if the histogram is symmetric and unimodal
(d) More than 0.50 if the histogram is skewed to the right.
- Which of the following is an example of a binomial experiment?
(a) A shopping mall is interested in the income levels of its customers and is taking a survey to gather information.

- (b) A business firm introducing a new product wants to know how many purchases its clients will make each year.
- (c) A sociologist is researching an area in an effort to determine the proportion of households with a male head of household.
- (d) A study is concerned with the average number of hours worked by high school students.
9. Let X be a Poisson random variable with mean 2.5. Find $P(X = 0)$
- (a) 2.5 (b) 0.0821
(c) 1.5811 (d) 0.40
10. The time it takes Jessica to bicycle to school is normally distributed with mean 15 minutes and variance 4. Jessica has to be at school at 8:00 am. What time should she leave her house so she will be late only 4% of the time?
- (a) 8:00 (b) 11.5 minutes before 8:00
(c) 22 minutes before 8:00 (d) 18.5 minutes before 8:00
11. A student took a chemistry exam where the exam scores were mound-shaped with a mean score of 90 and a standard deviation of 64. She also took a statistics exam where the scores were mound-shaped, the mean score was 70 and standard deviation 16. If the student's grades were 102 on the chemistry exam and 77 on the statistics exam, then
- (a) the student did relatively better on chemistry exam than on the statistics exam, compared to the other students in each class
- (b) the student did relatively better on statistics exam than on the chemistry exam, compared to the other students in the two classes
- (c) the student's scores on both exams are comparable, when accounting for the scores of the other students in the two classes.
- (d) it is impossible to say which of the student's exam scores indicates the better performance.
12. From a sample of size $n = 100$, the following descriptive measures were calculated: median=23, mean=20, standard deviation=5, range=35; 75 sample values are between 10 and 30; and 99 sample values are between 5 and 35. If you knew the sample mean, median, and standard deviation were correct, which of the following conclusions might you draw?
- (a) The distribution is skewed to the right because the median exceeds the mean.
- (b) The range must have been calculated incorrectly because it should not be seven times the standard deviation's value.
- (c) The number of sample values between 10 and 30 was miscounted.
- (d) The number of sample values between 5 and 35 must have been miscounted because all 100 values must be in this interval.
13. Which of the following is always true for two events A and B with,
- $$P(A) = 0.5, \quad P(B) = 0.7 \quad P(A \cap B) = 0.3?$$
- (a) A and B are independent (b) $P(A \cup B) = 0.8$
(c) A and B are mutually exclusive events (d) $P(A|B) = 0.6$
14. If $P(A) = 0.3$, $P(B) = 0.4$, and $P(A|B) = 0.6$ then $P(A \cup B)$ is
- (a) 0.7 (b) 0.1 (c) 0.8 (d) 0.46

15. We have three identical Boxes I, II and III. Each box has two drawers. Each drawer of Box I contains a gold coin. One drawer of Box II has a gold coin and the other one contains a silver coin and each drawer of Box III contains a silver coin. We choose one of the boxes in random and we open one of its drawers. If that drawer contains a gold coin then what is the probability that the second drawer contains a gold coin too?

(a) $\frac{1}{2}$

(b) $\frac{2}{3}$

(c) $\frac{1}{3}$

(d) $\frac{1}{6}$

Part II Long-answer questions. For these questions clearly show all your work; otherwise only partial or NO marks may be awarded. Marks for each question are given in []

1. Suppose men's heights are normally distributed with mean of 174 cm and a standard deviation of 6 cm.

(a) [4 marks] What proportion of men are between 170 cm and 179 cm tall?

(b) [4 marks] Find the minimum ceiling of an airplane such that at most 5% of the men walking down the aisle will have to duck their heads.

(c) [4 marks] Find the probability that the average height of a random sample of 49 men is greater than 176.

2. We randomly selected 50 billing statements from the computer databases of two hotel chains: Marriott and Radisson, and recorded the nightly room rates. A summary of this study is given below.

	Marriott	Radisson
Sample average	\$170	\$145
Sample standard deviation	\$15	\$10

- (a) [4 marks] Do this data indicate a difference between the average room rates for these two hotels? Test at $\alpha = .01$.

- (b) [4 marks] Find P -value of this test. Does this value confirm your findings in (a)? Why?

3. [4 marks] Five soft drink bottling companies have agreed to implement a time management program in hopes of increasing productivity (measured in cases of soft drinks bottled per hour). The number of cases of soft drinks bottled per hour before and after implementation of the program are listed below.

	Company				
	1	2	3	4	5
Before	500	475	525	490	530
After	510	480	525	495	533

Test at $\alpha = .05$ if the time management program is efficient in increasing the productivity.

4. A food processor wants to compare two preservatives for their effects on retarding spoilage. Suppose 10 cuts of fresh meat are treated with preservative A and 10 cuts are treated with preservative B, and the time (in hours) until spoilage begins is recorded for each of the 20 cuts. The results are summarized in the following table.

	Preservative A	Preservative B
Sample mean	108.7 hours	98.7 hours
Sample standard deviation	9.5 hours	11.5 hours

- (a) [4 marks] State clearly the null and alternative hypotheses to determine if the average time until spoilage begins differs for preservatives A and B.

- (b) [4 marks] Compute the pooled variance and the value of the test statistic.

- (c) [4 marks] Determine the rejection region at $\alpha = .05$, and write down a proper conclusion.

5. [4 marks] An Internet server conducted a study on 225 of its customers and found that the average amount spent online was 12.5 hours per week with a standard deviation of 5.4 hours. Construct a 95% confidence interval for the average online time for all users of this particular Internet server.
6. [4 marks] A salesperson has found that the probability of making a sale on a particular product manufactured by him or her company is .05. If the salesperson contacts 140 potential customers, what is the probability he or she will sell at least 2 of these products? Use and justify Poisson approximation to Binomial.
7. [4 marks] A firm with many traveling salespersons decides to check the salespersons' travel expenses to see if they are correctly reported. An auditor for the firm selects 200 expense reports at random to audit. What is the probability that more than 40 of these 200 sampled reports will be incorrect when, in fact, only 10% of the firm's reports are improperly documented. Use and justify normal approximation to binomial.

8. The number of household members, x , and the amount spent on groceries per week, y , rounded to the nearest dollar are measured for 8 households in a suburb of Ottawa. the data are shown below.

x	5	2	2	1	4	3	5	3
y	140	50	55	35	95	70	130	65

- (a) [4 marks] Compute \bar{y} , S_x , S_y , and S_{xy}

- (b) [4 marks] Compute the correlation coefficient r between x and y . What would you estimate a household of 6 to spend on groceries per week.

Formulae

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1}$$

$$\mathbb{E}(X) = \mu = \sum_x xp(x), \quad \sigma^2 = \sum_x (x - \mu)^2 p(x) = \sum_x x^2 p(x) - \mu^2$$

$$S_{xy} = \frac{1}{n-1} \left(\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} \right), \quad r = \frac{S_{xy}}{S_x S_y}, \quad b = r \frac{S_y}{S_x}, \quad a = \bar{y} - b\bar{x}, \quad \hat{y} = bx + a$$

S_x and S_y are standard deviations of X and Y .

- Binomial distribution with parameters n, p :

$$P(X = k) = C_k^n p^k (1-p)^{n-k} \quad \mu = np, \quad \sigma^2 = np(1-p)$$

$$C_k^n = \frac{n!}{k!(n-k)!}$$

- Hypergeometric distribution with parameters N, M, n :

$$P(X = k) = \frac{C_k^M C_{n-k}^{N-M}}{C_n^N}, \quad \mu = n \frac{M}{N}, \quad \sigma^2 = n \left(\frac{M}{N} \right) \left(1 - \frac{M}{N} \right) \left(\frac{N-n}{N-1} \right)$$

- Poisson distribution with parameter μ : $P(X = k) = e^{-\mu} \frac{\mu^k}{k!}$, $\mathbb{E}(X) = \mu = \text{Var}(X)$

Approximations

- If $n/N \leq .05$, then hypergeometric distribution may be approximated by the Binomial.
- If $np < 7$ and $n \geq 50$ then the binomial may be approximated by Poisson distribution.
- If $np \geq 5$ and $n(1-p) \geq 5$, then binomial may be approximated by normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$.

Important Test Statistics	Confidence Intervals
$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
$\frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}}$	$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}\hat{q}/n}$
$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$	$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$
$\frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$
$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$	$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$
$\frac{\bar{d}}{s_d/\sqrt{n}}$	$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$

$$\text{pooled variance } s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad \text{pooled proportion } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

For One-sided test and One-sided confidence interval (lower and upper confidence bounds), use α instead of $\alpha/2$.

