Lecturer:	*Date*	**Approved by:**	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM **Faculty of Applied Science**	**Final Exam**	Academic year	2023-2024	Semester	1
		Exam date	December 18th, 2023		
	Course title	Probability and Statistics		**Score**	
	Course ID	MT2013	Sheet code	**3131**	
	Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 15 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.67 points for each correct answer and 0.13 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.75 and the probability that it is incorrect is 0.25. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ None of the others

   Ⓑ All the statements are true.

   Ⓒ Only (2) and (3) are true.

   Ⓓ Only (3) is true.

   Ⓔ Only (1) is true.

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.6625 Ⓑ 0.7625 Ⓒ None of the others. Ⓓ 0.5625 Ⓔ 0.1625

3. Find the probability that a student skips 3 questions.
   Ⓐ 0.2252 Ⓑ 0.1252 Ⓒ 0.0252 Ⓓ None of the others. Ⓔ 0.4252

4. If a student skips 3 questions, find the expected score of the student.
   Ⓐ 6.14 Ⓑ None of the others. Ⓒ 6.64 Ⓓ 5.64 Ⓔ 4.64

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of more than 738.5 hours. A random sample of 10 bulbs is selected and their lifespans are recorded as follows: 686.3, 806.3, 630.1, 827.5, 868.5, 862.8, 755.7, 697.9, 739.5, 790.1. Based on this random sample, the company wants to test whether its claim is supported at the 0.05 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.
   Ⓐ H0: $\mu \neq 738.5$, H1: $\mu = 738.5$.
   Ⓑ H0: $\mu > 738.5$, H1: $\mu \leq 738.5$.
   Ⓒ H0: $\mu = 738.5$, H1: $\mu \neq 738.5$.
   Ⓓ H0: $\mu = 738.5$, H1: $\mu > 738.5$.
   Ⓔ H0: $\mu \geq 738.5$, H1: $\mu < 738.5$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?
   Ⓐ Normal distribution with the mean 766.47 and unknown variance.
   Ⓑ None of the others.
   Ⓒ Normal distribution with unknown variance.
   Ⓓ Normal distribution with with the standard deviation of 24.9729.
   Ⓔ Student distribution with degree of free dom 9.

7. Calculate the statistic for the hypothesis test.
   Ⓐ 40.008 Ⓑ 23.3415 Ⓒ 51.119 Ⓓ 1.12 Ⓔ None of the others.

8. Find a 95% confidence interval for the mean lifespan of the light bulbs produced by the company.
   Ⓐ [720.6946 , 812.2454] Ⓑ [709.9812 , 822.9588] Ⓒ [717.523 , 815.417] Ⓓ [697.5444 , 779.4556] Ⓔ None of the others.

9. Suppose further that the standard deviation of the bulb lifespan is 57.97 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 95% confidence interval for the expected lifespan μ is no more than 7 (hours)?
   Ⓐ 309 Ⓑ 244 Ⓒ 224 Ⓓ 264 Ⓔ 279

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 20$, $\sum_{i=1}^{n} x_i = 719$, $\sum_{i=1}^{n} y_i = 335.407$, $\sum_{i=1}^{n} x_i^2 = 31501$,
$\sum_{i=1}^{n} y_i^2 = 6788.1211$ and $\sum_{i=1}^{n} x_i y_i = 14620.353$.

10. Find the sample correlation coefficient for this sample.
    Ⓐ None of the others. Ⓑ 0.6564 Ⓒ 0.7689 Ⓓ 0.9993 Ⓔ 0.8395

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to
    - Ⓐ increase approximately by 0.4533 $\times 10^9$ particles.
    - Ⓑ decrease approximately by 0.4533 $\times 10^9$ particles.
    - Ⓒ None of the others.
    - Ⓓ decrease approximately by 0.4743 $\times 10^9$ particles.
    - Ⓔ increase approximately by 0.4743 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.
    Ⓐ 0.4171 Ⓑ None of the others. Ⓒ 0.068 Ⓓ 0.004 Ⓔ 0.2911

13. Find a 95% confidence interval for the slope β_1 of the linear regression line.
    Ⓐ [0.4454,0.4612] Ⓑ [0.4463,0.4603] Ⓒ None of the others. Ⓓ [0.4448,0.4618] Ⓔ [-0.5431,1.4497]

14. Estimate the y-intercept of the linear regression line.
    Ⓐ 0.292 Ⓑ 0.6537 Ⓒ None of the others. Ⓓ 0.4743 Ⓔ 0.7078

15. Using the linear regression model, approximate the residual for the predicted value $y = 5.13$ at $x = 10$. Ⓐ -0.0641 Ⓑ None of the others. Ⓒ -0.0698 Ⓓ 0.1227 Ⓔ -0.2501

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 228 products made by printer 1 and 244 products made by printer 2. Among these chosen products, there are 5 defective products made by printer 1 and 7 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 1 at the significance level $\alpha = 0.05$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.
    Ⓐ None of the others. Ⓑ 0.5144 Ⓒ 0.07 Ⓓ 0.0145 Ⓔ 0.2922

17. Determine the appropriate hypothesis test.
    - Ⓐ z-test since the sample proportion follows a normal distribution.
    - Ⓑ t-test since the population distribution is unknown.
    - Ⓒ z-test since the sample proportion approximately follows a normal distribution.
    - Ⓓ z-test since we want to compare two mean values.
    - Ⓔ t-test since the population variance is unknown.

18. Calculate the statistic of the hypothesis test.
    Ⓐ -1.3552 Ⓑ -1.9107 Ⓒ -0.4662 Ⓓ -0.2442 Ⓔ None of the others.

19. Which of the following statements is true?
    - Ⓐ Printer 2 is better than printer 1.
    - Ⓑ None of the others.
    - Ⓒ Printer 1 is better than printer 2.
    - Ⓓ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$
    - Ⓔ Choose printer 1 since $\hat{p}_1 < \hat{p}_2$.

20. Construct a 95 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ [-0.0352 , 0.0217] Ⓑ [-0.035 , 0.0215] Ⓒ [-0.0304 , 0.0169] Ⓓ [-0.0352 , 0.0217] Ⓔ None of the others.

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	**Approved by:**	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM **Faculty of Applied Science**	**Final Exam**	Academic year	2023-2024	Semester	1
		Exam date	December 18th, 2023		
	Course title	Probability and Statistics		**Score**	
	Course ID	MT2013	Sheet code	**3132**	
	Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 15 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.67 points for each correct answer and 0.13 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.7 and the probability that it is incorrect is 0.3. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ Only (3) and (4) are true.
   Ⓑ Only (1) is true.
   Ⓒ Only (1) and (3) are true.
   Ⓓ All the statements are true.
   Ⓔ None of the others

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.925 Ⓑ 0.725 Ⓒ None of the others. Ⓓ 0.525 Ⓔ 0.225

3. Find the probability that a student skips 3 questions.

   Ⓐ 0.4252 Ⓑ 0.2252 Ⓒ None of the others. Ⓓ 0.0252 Ⓔ 0.5252

4. If a student skips 3 questions, find the expected score of the student.

   Ⓐ 5.16 Ⓑ 6.16 Ⓒ 5.66 Ⓓ None of the others. Ⓔ 3.66

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of less than 726.5 hours. A random sample of 7 bulbs is selected and their lifespans are recorded as follows: 742.6, 631.7, 683.2, 645.8, 737.7, 626.4, 805.4. Based on this random sample, the company wants to test whether its claim is supported at the 0.01 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

   Ⓐ H0: $\mu \leq 726.5$, H1: $\mu > 726.5$.

   Ⓑ H0: $\mu > 726.5$, H1: $\mu \leq 726.5$.

   Ⓒ H0: $\mu = 726.5$, H1: $\mu \neq 726.5$.

   Ⓓ H0: $\mu \neq 726.5$, H1: $\mu = 726.5$.

   Ⓔ H0: $\mu = 726.5$, H1: $\mu < 726.5$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

   Ⓐ None of the others.

   Ⓑ Normal distribution with the mean 696.1143 and unknown variance.

   Ⓒ Student distribution with degree of free dom 6.

   Ⓓ Normal distribution with with the standard deviation of 25.6075.

   Ⓔ Normal distribution with unknown variance.

7. Calculate the statistic for the hypothesis test.

   Ⓐ -28.9646 Ⓑ -17.8536 Ⓒ -1.1866 Ⓓ -12.2981 Ⓔ None of the others.

8. Find a 99% confidence interval for the mean lifespan of the light bulbs produced by the company.

   Ⓐ [630.0469 , 762.1816] Ⓑ [615.6299 , 776.5987] Ⓒ None of the others. Ⓓ [666.8345 , 786.1655] Ⓔ [601.1873 , 791.0413]

9. Suppose further that the standard deviation of the bulb lifespan is 76.3 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 99% confidence interval for the expected lifespan μ is no more than 5 (hours)?

   Ⓐ 1581 Ⓑ 1551 Ⓒ 1591 Ⓓ 1596 Ⓔ 1586

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 16$, $\sum_{i=1}^{n} x_i = 459$, $\sum_{i=1}^{n} y_i = 107.175$, $\sum_{i=1}^{n} x_i^2 = 16061$,
$\sum_{i=1}^{n} y_i^2 = 866.1281$ and $\sum_{i=1}^{n} x_i y_i = 3728.325$.

10. Find the sample correlation coefficient for this sample.

    Ⓐ 0.8229 Ⓑ 0.6471 Ⓒ 0.9983 Ⓓ None of the others. Ⓔ 0.7973

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to

    Ⓐ increase approximately by 0.2259 $\times 10^9$ particles.

    Ⓑ decrease approximately by 0.2259 $\times 10^9$ particles.

    Ⓒ increase approximately by 0.2168 $\times 10^9$ particles.

    Ⓓ None of the others.

    Ⓔ decrease approximately by 0.2168 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.

    Ⓐ 0.3529 Ⓑ None of the others. Ⓒ 0.3323 Ⓓ 0.0911 Ⓔ 0.0036

13. Find a 90% confidence interval for the slope β_1 of the linear regression line.

    Ⓐ [0.2201,0.2318] Ⓑ [0.2196,0.2322] Ⓒ [0.2211,0.2307] Ⓓ None of the others. Ⓔ [-0.1559,0.6078]

14. Estimate the y-intercept of the linear regression line.

    Ⓐ 0.4734 Ⓑ -0.0869 Ⓒ None of the others. Ⓓ 0.198 Ⓔ 0.2168

15. Using the linear regression model, approximate the residual for the predicted value $y = 6.48$ at $x = 27$. Ⓐ -0.2709 Ⓑ 0.595 Ⓒ 0.1628 Ⓓ 0.1252 Ⓔ None of the others.

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 156 products made by printer 1 and 236 products made by printer 2. Among these chosen products, there are 2 defective products made by printer 1 and 7 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 1 at the significance level $\alpha = 0.01$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.

    Ⓐ 0.2932 Ⓑ 0.1821 Ⓒ 0.071 Ⓓ 0.0155 Ⓔ None of the others.

17. Determine the appropriate hypothesis test.

    Ⓐ t-test since the population distribution is unknown.

    Ⓑ z-test since we want to compare two mean values.

    Ⓒ z-test since the sample proportion follows a normal distribution.

    Ⓓ t-test since the population variance is unknown.

    Ⓔ z-test since the sample proportion approximately follows a normal distribution.

18. Calculate the statistic of the hypothesis test.

    Ⓐ -1.0897 Ⓑ -3.0897 Ⓒ -1.9787 Ⓓ None of the others. Ⓔ -0.8677

19. Which of the following statements is true?

    Ⓐ Choose printer 1 since $\hat{p}_1 < \hat{p}_2$.

    Ⓑ Printer 2 is better than printer 1.

    Ⓒ Printer 1 is better than printer 2.

    Ⓓ None of the others.

    Ⓔ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$

20. Construct a 99 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ [-0.0569 , 0.0232] Ⓑ [-0.0536 , 0.0199] Ⓒ None of the others. Ⓓ [-0.0567 , 0.023] Ⓔ [-0.05 , 0.0164]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	**Approved by:**	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM **Faculty of Applied Science**	**Final Exam**	Academic year	2023-2024	Semester	1
		Exam date	December 18th, 2023		
	Course title	Probability and Statistics		**Score**	
	Course ID	MT2013	Sheet code	**3133**	
	Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 15 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.67 points for each correct answer and 0.13 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.85 and the probability that it is incorrect is 0.15. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ Only (3) and (4) are true.

   Ⓑ All the statements are true.

   Ⓒ Only (1) is true.

   Ⓓ Only (2) and (3) are true.

   Ⓔ None of the others

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.3375 Ⓑ 0.2375 Ⓒ None of the others. Ⓓ 0.7375 Ⓔ 0.6375

3. Find the probability that a student skips 4 questions.

   Ⓐ None of the others. Ⓑ 0.0252 Ⓒ 0.2252 Ⓓ 0.5252 Ⓔ 0.6252

4. If a student skips 4 questions, find the expected score of the student.

   Ⓐ 4.55 Ⓑ 6.05 Ⓒ 5.55 Ⓓ None of the others. Ⓔ 6.55

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of là 958.5 hours. A random sample of 8 bulbs is selected and their lifespans are recorded as follows: 976.2, 903.1, 832.6, 1022.1, 909.6, 984.6, 879.1, 866.3. Based on this random sample, the company wants to test whether its claim is supported at the 0.05 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

   Ⓐ H0: $\mu \leq 958.5$, H1: $\mu > 958.5$.

   Ⓑ H0: $\mu = 958.5$, H1: $\mu \neq 958.5$.

   Ⓒ H0: $\mu \geq 958.5$, H1: $\mu < 958.5$.

   Ⓓ H0: $\mu > 958.5$, H1: $\mu \leq 958.5$.

   Ⓔ H0: $\mu \neq 958.5$, H1: $\mu = 958.5$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

   Ⓐ Student distribution with degree of free dom 7.

   Ⓑ Normal distribution with the mean 921.7 and unknown variance.

   Ⓒ None of the others.

   Ⓓ Normal distribution with unknown variance.

   Ⓔ Normal distribution with with the standard deviation of 23.2694.

7. Calculate the statistic for the hypothesis test.

   Ⓐ -12.693 Ⓑ -46.026 Ⓒ None of the others. Ⓓ -1.5815 Ⓔ -29.3595

8. Find a 95% confidence interval for the mean lifespan of the light bulbs produced by the company.

   Ⓐ [876.0921 , 967.3079] Ⓑ [920.3383 , 996.6617] Ⓒ [866.668 , 976.732] Ⓓ [877.6046 , 965.7954] Ⓔ None of the others.

9. Suppose further that the standard deviation of the bulb lifespan is 73.77 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 95% confidence interval for the expected lifespan μ is no more than 5 (hours)?

   Ⓐ 812 Ⓑ 792 Ⓒ 807 Ⓓ 837 Ⓔ 797

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 20$, $\sum_{i=1}^{n} x_i = 782$, $\sum_{i=1}^{n} y_i = 34.78$, $\sum_{i=1}^{n} x_i^2 = 34380$,
$\sum_{i=1}^{n} y_i^2 = 69.338$ and $\sum_{i=1}^{n} x_i y_i = 1524.2$.

10. Find the sample correlation coefficient for this sample.

    Ⓐ 0.7952 Ⓑ 0.6105 Ⓒ 0.881 Ⓓ 0.8952 Ⓔ None of the others.

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to
    - Ⓐ increase approximately by 0.0501 $\times 10^9$ particles.
    - Ⓑ increase approximately by 0.0432 $\times 10^9$ particles.
    - Ⓒ decrease approximately by 0.0501 $\times 10^9$ particles.
    - Ⓓ None of the others.
    - Ⓔ decrease approximately by 0.0432 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.
    Ⓐ None of the others. Ⓑ 0.4273 Ⓒ 0.2021 Ⓓ 0.0655 Ⓔ 0.0051

13. Find a 90% confidence interval for the slope β_1 of the linear regression line.
    Ⓐ [0.0344,0.052] Ⓑ [0.0349,0.0515] Ⓒ [0.0365,0.0499] Ⓓ None of the others. Ⓔ [-0.0437,0.1301]

14. Estimate the y-intercept of the linear regression line.
    Ⓐ -0.029 Ⓑ 0.2767 Ⓒ None of the others. Ⓓ 0.0501 Ⓔ 0.0165

15. Using the linear regression model, approximate the residual for the predicted value $y = 0.8$ at $x = 15$. Ⓐ 0.454 Ⓑ 0.1938 Ⓒ 0.3027 Ⓓ None of the others. Ⓔ 0.102

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 304 products made by printer 1 and 182 products made by printer 2. Among these chosen products, there are 3 defective products made by printer 1 and 16 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 1 at the significance level $\alpha = 0.05$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.
    Ⓐ 0.0182 Ⓑ 0.2959 Ⓒ 0.1848 Ⓓ None of the others. Ⓔ 0.5181

17. Determine the appropriate hypothesis test.
    - Ⓐ z-test since the sample proportion approximately follows a normal distribution.
    - Ⓑ z-test since we want to compare two mean values.
    - Ⓒ t-test since the population variance is unknown.
    - Ⓓ z-test since the sample proportion follows a normal distribution.
    - Ⓔ t-test since the population distribution is unknown.

18. Calculate the statistic of the hypothesis test.
    Ⓐ -4.0743 Ⓑ -2.9633 Ⓒ -4.2963 Ⓓ -5.1853 Ⓔ None of the others.

19. Which of the following statements is true?
    - Ⓐ Choose printer 2 since H0 could be rejected.
    - Ⓑ None of the others.
    - Ⓒ Choose printer 1 since H0 could be rejected.
    - Ⓓ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$
    - Ⓔ Choose printer 1 since $\hat{p}_1 < \hat{p}_2$

20. Construct a 95 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ [-0.1137 , -0.0423] Ⓑ [-0.1137 , -0.0424] Ⓒ None of the others. Ⓓ [-0.1207 , -0.0354] Ⓔ [-0.1136 , -0.0424]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	**Approved by:**	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM **Faculty of Applied Science**	**Final Exam**	Academic year	2023-2024	Semester	1
		Exam date	December 18th, 2023		
	Course title	Probability and Statistics		**Score**	
	Course ID	MT2013	Sheet code	**3134**	
	Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 20 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.5 points for each correct answer and 0.1 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.75 and the probability that it is incorrect is 0.25. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ Only (2) and (4) are true.

   Ⓑ Only (1), (2), and (3) are true.

   Ⓒ None of the others

   Ⓓ Only (2) and (3) are true.

   Ⓔ All the statements are true.

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.3625 Ⓑ 0.6625 Ⓒ 0.4625 Ⓓ 0.5625 Ⓔ None of the others.

3. Find the probability that a student skips 4 questions.

   Ⓐ 0.1897 Ⓑ 0.0897 Ⓒ 0.5897 Ⓓ None of the others. Ⓔ 0.6897

4. If a student skips 4 questions, find the expected score of the student.

   Ⓐ 6.6 Ⓑ 4.1 Ⓒ None of the others. Ⓓ 7.1 Ⓔ 5.6

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of more than 701.9 hours. A random sample of 9 bulbs is selected and their lifespans are recorded as follows: 742.3, 697.6, 704, 737, 635.7, 684.7, 744.7, 699.5, 676.7. Based on this random sample, the company wants to test whether its claim is supported at the 0.01 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

   Ⓐ H0: $\mu \geq 701.9$, H1: $\mu < 701.9$.

   Ⓑ H0: $\mu = 701.9$, H1: $\mu > 701.9$.

   Ⓒ H0: $\mu = 701.9$, H1: $\mu \neq 701.9$.

   Ⓓ H0: $\mu > 701.9$, H1: $\mu \leq 701.9$.

   Ⓔ H0: $\mu \neq 701.9$, H1: $\mu = 701.9$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

   Ⓐ Student distribution with degree of free dom 8.

   Ⓑ Normal distribution with the mean 702.4667 and unknown variance.

   Ⓒ None of the others.

   Ⓓ Normal distribution with unknown variance.

   Ⓔ Normal distribution with with the standard deviation of 11.8058.

7. Calculate the statistic for the hypothesis test.

   Ⓐ 16.714 Ⓑ 0.048 Ⓒ 27.825 Ⓓ 50.047 Ⓔ None of the others.

8. Find a 99% confidence interval for the mean lifespan of the light bulbs produced by the company.

   Ⓐ [672.0078 , 732.9256] Ⓑ None of the others. Ⓒ [668.2653 , 736.668] Ⓓ [674.3925 , 729.4075] Ⓔ [662.8583 , 742.0751]

9. Suppose further that the standard deviation of the bulb lifespan is 42.33 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 99% confidence interval for the expected lifespan μ is no more than 7 (hours)?

   Ⓐ 259 Ⓑ 249 Ⓒ 294 Ⓓ 244 Ⓔ 219

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 17$, $\sum_{i=1}^{n} x_i = 658$, $\sum_{i=1}^{n} y_i = 233.81$, $\sum_{i=1}^{n} x_i^2 = 28988$,
$\sum_{i=1}^{n} y_i^2 = 3637.9267$ and $\sum_{i=1}^{n} x_i y_i = 10267.86$.

10. Find the sample correlation coefficient for this sample.

    Ⓐ 0.5872 Ⓑ 0.8517 Ⓒ None of the others. Ⓓ 0.6664 Ⓔ 0.9992

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to

    Ⓐ increase approximately by 0.3461 $\times 10^9$ particles.

    Ⓑ None of the others.

    Ⓒ increase approximately by 0.3582 $\times 10^9$ particles.

    Ⓓ decrease approximately by 0.3461 $\times 10^9$ particles.

    Ⓔ decrease approximately by 0.3582 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.

    Ⓐ 0.0036 Ⓑ 0.3331 Ⓒ None of the others. Ⓓ 0.3676 Ⓔ 0.0788

13. Find a 99% confidence interval for the slope β_1 of the linear regression line.

    Ⓐ None of the others. Ⓑ [0.3355,0.3566] Ⓒ [0.3368,0.3553] Ⓓ [0.3368,0.3554] Ⓔ [-0.7095,1.4016]

14. Estimate the y-intercept of the linear regression line.

    Ⓐ 0.7656 Ⓑ 0.3582 Ⓒ None of the others. Ⓓ -0.0696 Ⓔ -0.0257

15. Using the linear regression model, approximate the residual for the predicted value $y = 6.66$ at $x = 19$. Ⓐ -0.5252 Ⓑ -0.2737 Ⓒ -0.5789 Ⓓ None of the others. Ⓔ -0.6146

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 356 products made by printer 1 and 326 products made by printer 2. Among these chosen products, there are 28 defective products made by printer 1 and 20 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 2 at the significance level $\alpha = 0.05$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.

    Ⓐ None of the others. Ⓑ 0.5195 Ⓒ 0.0196 Ⓓ 0.4084 Ⓔ 0.2973

17. Determine the appropriate hypothesis test.

    Ⓐ t-test since the population variance is unknown.

    Ⓑ t-test since the population distribution is unknown.

    Ⓒ z-test since the sample proportion approximately follows a normal distribution.

    Ⓓ z-test since the sample proportion follows a normal distribution.

    Ⓔ z-test since we want to compare two mean values.

18. Calculate the statistic of the hypothesis test.

    Ⓐ 1.6599 Ⓑ None of the others. Ⓒ 2.7709 Ⓓ 0.8824 Ⓔ 1.1044

19. Which of the following statements is true?

    Ⓐ None of the others.

    Ⓑ Printer 1 is better than printer 2.

    Ⓒ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$

    Ⓓ Choose printer 1 since $\hat{p}_1 < \hat{p}_2$.

    Ⓔ Printer 2 is better than printer 1.

20. Construct a 95 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ [-0.0212 , 0.0558] Ⓑ None of the others. Ⓒ [-0.0209 , 0.0555] Ⓓ [-0.0211 , 0.0557] Ⓔ [-0.0147 , 0.0493]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	**Approved by:**	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM **Faculty of Applied Science**	**Final Exam**	Academic year	2023-2024	Semester	1
		Exam date	December 18th, 2023		
	Course title	Probability and Statistics		**Score**	
	Course ID	MT2013	Sheet code	**3135**	
	Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 10 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 1 points for each correct answer and 0.2 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.8 and the probability that it is incorrect is 0.2. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ Only (2) and (4) are true.

   Ⓑ All the statements are true.

   Ⓒ None of the others

   Ⓓ Only (4) is true.

   Ⓔ Only (3) and (4) are true.

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.3 Ⓑ 0.6 Ⓒ 0.2 Ⓓ 0.4 Ⓔ None of the others.

3. Find the probability that a student skips 2 questions.

   Ⓐ 0.0816 Ⓑ 0.2816 Ⓒ 0.7816 Ⓓ 0.3816 Ⓔ None of the others.

4. If a student skips 2 questions, find the expected score of the student.

   Ⓐ 5.58 Ⓑ 6.08 Ⓒ 8.08 Ⓓ None of the others. Ⓔ 7.58

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of less than 560.6 hours. A random sample of 7 bulbs is selected and their lifespans are recorded as follows: 551.8, 553.4, 568.9, 498.9, 555.2, 584.8, 579.2. Based on this random sample, the company wants to test whether its claim is supported at the 0.01 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

   Ⓐ H0: $\mu \neq 560.6$, H1: $\mu = 560.6$.

   Ⓑ H0: $\mu > 560.6$, H1: $\mu \leq 560.6$.

   Ⓒ H0: $\mu = 560.6$, H1: $\mu \neq 560.6$.

   Ⓓ H0: $\mu \leq 560.6$, H1: $\mu > 560.6$.

   Ⓔ H0: $\mu = 560.6$, H1: $\mu < 560.6$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

   Ⓐ None of the others.

   Ⓑ Normal distribution with with the standard deviation of 10.7122.

   Ⓒ Student distribution with degree of free dom 6.

   Ⓓ Normal distribution with unknown variance.

   Ⓔ Normal distribution with the mean 556.0286 and unknown variance.

7. Calculate the statistic for the hypothesis test.

   Ⓐ -28.2048 Ⓑ None of the others. Ⓒ -0.4268 Ⓓ -17.0938 Ⓔ -33.7603

8. Find a 99% confidence interval for the mean lifespan of the light bulbs produced by the company.

   Ⓐ [535.6406 , 585.5594] Ⓑ [522.3602 , 589.697] Ⓒ None of the others. Ⓓ [516.3185 , 595.7386] Ⓔ [528.3911 , 583.666]

9. Suppose further that the standard deviation of the bulb lifespan is 28.9 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 99% confidence interval for the expected lifespan μ is no more than 8 (hours)?

   Ⓐ 87 Ⓑ 122 Ⓒ 67 Ⓓ 137 Ⓔ 72

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 17$, $\sum_{i=1}^{n} x_i = 608$, $\sum_{i=1}^{n} y_i = 258.164$, $\sum_{i=1}^{n} x_i^2 = 24426$,
$\sum_{i=1}^{n} y_i^2 = 4377.7093$ and $\sum_{i=1}^{n} x_i y_i = 10338.908$.

10. Find the sample correlation coefficient for this sample.

    Ⓐ 0.7089 Ⓑ 0.7378 Ⓒ 0.6339 Ⓓ 0.9987 Ⓔ None of the others.

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to

    Ⓐ increase approximately by 0.4357 $\times 10^9$ particles.

    Ⓑ None of the others.

    Ⓒ increase approximately by 0.4124 $\times 10^9$ particles.

    Ⓓ decrease approximately by 0.4357 $\times 10^9$ particles.

    Ⓔ decrease approximately by 0.4124 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.

    Ⓐ 0.212 Ⓑ None of the others. Ⓒ 0.2647 Ⓓ 0.3557 Ⓔ 0.0054

13. Find a 90% confidence interval for the slope β_1 of the linear regression line.

    Ⓐ [0.4036,0.4212] Ⓑ None of the others. Ⓒ [0.4052,0.4196] Ⓓ [-0.3513,1.1762] Ⓔ [0.403,0.4218]

14. Estimate the y-intercept of the linear regression line.

    Ⓐ 0.5299 Ⓑ 0.8044 Ⓒ None of the others. Ⓓ 0.6591 Ⓔ 0.4357

15. Using the linear regression model, approximate the residual for the predicted value $y = 18.65$ at $x = 44$. Ⓐ None of the others. Ⓑ 0.0674 Ⓒ -0.117 Ⓓ 0.2522 Ⓔ -0.3005

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 112 products made by printer 1 and 322 products made by printer 2. Among these chosen products, there are 10 defective products made by printer 1 and 6 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 2 at the significance level $\alpha = 0.01$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.

    Ⓐ 0.0762 Ⓑ None of the others. Ⓒ 0.5206 Ⓓ 0.0207 Ⓔ 0.2984

17. Determine the appropriate hypothesis test.

    Ⓐ z-test since the sample proportion follows a normal distribution.

    Ⓑ t-test since the population variance is unknown.

    Ⓒ z-test since we want to compare two mean values.

    Ⓓ t-test since the population distribution is unknown.

    Ⓔ z-test since the sample proportion approximately follows a normal distribution.

18. Calculate the statistic of the hypothesis test.

    Ⓐ None of the others. Ⓑ 4.1954 Ⓒ 3.4179 Ⓓ 5.3064 Ⓔ 1.4179

19. Which of the following statements is true?

    Ⓐ Choose printer 1 since H0 could be rejected.

    Ⓑ None of the others.

    Ⓒ Choose printer 1 since H0 could NOT be rejected

    Ⓓ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$

    Ⓔ Choose printer 2 since H0 could be rejected.

20. Construct a 99 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ [0.0055 , 0.1358] Ⓑ [-0.0015 , 0.1428] Ⓒ [0.0172 , 0.1241] Ⓓ None of the others. Ⓔ [0.0173 , 0.124]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	**Approved by:**	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM **Faculty of Applied Science**	**Final Exam**	Academic year	2023-2024	Semester	1
		Exam date	December 18th, 2023		
	Course title	Probability and Statistics		**Score**	
	Course ID	MT2013	Sheet code	**3136**	
	Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 20 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.5 points for each correct answer and 0.1 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.9 and the probability that it is incorrect is 0.1. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ Only (2) and (3) are true.

   Ⓑ All the statements are true.

   Ⓒ Only (2) and (4) are true.

   Ⓓ Only (1) and (2) are true.

   Ⓔ None of the others

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.175 Ⓑ 0.775 Ⓒ 0.675 Ⓓ 0.875 Ⓔ None of the others.

3. Find the probability that a student skips 5 questions.

   Ⓐ 0.3023 Ⓑ 0.0023 Ⓒ None of the others. Ⓓ 0.2023 Ⓔ 0.5023

4. If a student skips 5 questions, find the expected score of the student.

   Ⓐ 8.1 Ⓑ None of the others. Ⓒ 5.1 Ⓓ 6.6 Ⓔ 7.1

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of less than 916.1 hours. A random sample of 8 bulbs is selected and their lifespans are recorded as follows: 1048, 856.4, 871.4, 825.8, 736.8, 918.5, 985.9, 866.9. Based on this random sample, the company wants to test whether its claim is supported at the 0.01 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

   Ⓐ H0: $\mu = 916.1$, H1: $\mu < 916.1$.

   Ⓑ H0: $\mu \leq 916.1$, H1: $\mu > 916.1$.

   Ⓒ H0: $\mu \neq 916.1$, H1: $\mu = 916.1$.

   Ⓓ H0: $\mu = 916.1$, H1: $\mu \neq 916.1$.

   Ⓔ H0: $\mu > 916.1$, H1: $\mu \leq 916.1$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

   Ⓐ Normal distribution with with the standard deviation of 33.9538.

   Ⓑ Normal distribution with unknown variance.

   Ⓒ Student distribution with degree of free dom 7.

   Ⓓ None of the others.

   Ⓔ Normal distribution with the mean 888.7125 and unknown variance.

7. Calculate the statistic for the hypothesis test.

   Ⓐ None of the others. Ⓑ -34.1401 Ⓒ -17.4736 Ⓓ -6.3626 Ⓔ -0.8066

8. Find a 99% confidence interval for the mean lifespan of the light bulbs produced by the company.

   Ⓐ None of the others. Ⓑ [836.9876 , 995.2124] Ⓒ [786.919 , 990.506] Ⓓ [801.1117 , 976.3133] Ⓔ [769.8742 , 1007.5508]

9. Suppose further that the standard deviation of the bulb lifespan is 91.35 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 99% confidence interval for the expected lifespan μ is no more than 6 (hours)?

   Ⓐ 1538 Ⓑ 1533 Ⓒ 1558 Ⓓ 1543 Ⓔ 1578

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 16$, $\sum_{i=1}^{n} x_i = 567$, $\sum_{i=1}^{n} y_i = 12.771$, $\sum_{i=1}^{n} x_i^2 = 23165$,
$\sum_{i=1}^{n} y_i^2 = 12.1637$ and $\sum_{i=1}^{n} x_i y_i = 459.945$.

10. Find the sample correlation coefficient for this sample.

    Ⓐ 0.0302 Ⓑ -0.0815 Ⓒ None of the others. Ⓓ 0.2787 Ⓔ 0.0948

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to
    - Ⓐ decrease approximately by 0.0024 $\times 10^9$ particles.
    - Ⓑ None of the others.
    - Ⓒ increase approximately by 0.7131 $\times 10^9$ particles.
    - Ⓓ decrease approximately by 0.7131 $\times 10^9$ particles.
    - Ⓔ increase approximately by 0.0024 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.
    Ⓐ 0.4422 Ⓑ 0.2714 Ⓒ 0.4251 Ⓓ 0.0067 Ⓔ None of the others.

13. Find a 90% confidence interval for the slope β_1 of the linear regression line.
    Ⓐ None of the others. Ⓑ [-0.0067,0.0115] Ⓒ [-0.0095,0.0143] Ⓓ [-0.0086,0.0134] Ⓔ [-1.2536,1.2584]

14. Estimate the y-intercept of the linear regression line.
    Ⓐ 0.4784 Ⓑ 0.27 Ⓒ 1.0033 Ⓓ 0.7131 Ⓔ None of the others.

15. Using the linear regression model, approximate the residual for the predicted value $y = 0.7$ at $x = 15$. Ⓐ -0.0491 Ⓑ -0.4699 Ⓒ -0.1506 Ⓓ 0.2583 Ⓔ None of the others.

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 352 products made by printer 1 and 130 products made by printer 2. Among these chosen products, there are 28 defective products made by printer 1 and 7 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 2 at the significance level $\alpha = 0.01$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.
    Ⓐ 0.0821 Ⓑ 0.4154 Ⓒ 0.0266 Ⓓ None of the others. Ⓔ 0.5265

17. Determine the appropriate hypothesis test.
    - Ⓐ t-test since the population distribution is unknown.
    - Ⓑ z-test since the sample proportion follows a normal distribution.
    - Ⓒ t-test since the population variance is unknown.
    - Ⓓ z-test since the sample proportion approximately follows a normal distribution.
    - Ⓔ z-test since we want to compare two mean values.

18. Calculate the statistic of the hypothesis test.
    Ⓐ None of the others. Ⓑ 1.7424 Ⓒ 1.1869 Ⓓ 2.2979 Ⓔ 0.9649

19. Which of the following statements is true?
    - Ⓐ Printer 2 is better than printer 1.
    - Ⓑ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$
    - Ⓒ None of the others.
    - Ⓓ Choose printer 1 since $\hat{p}_1 < \hat{p}_2$.
    - Ⓔ Printer 1 is better than printer 2.

20. Construct a 99 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ None of the others. Ⓑ [-0.0432 , 0.0946] Ⓒ [-0.043 , 0.0944] Ⓓ [-0.0375 , 0.0889] Ⓔ [-0.0314 , 0.0828]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	Approved by:	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM
Faculty of Applied Science

Final Exam	Academic year	2023-2024	Semester	1
	Exam date	December 18th, 2023		
Course title	Probability and Statistics		**Score**	
Course ID	MT2013	Sheet code	**3137**	
Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 20 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.5 points for each correct answer and 0.1 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.8 and the probability that it is incorrect is 0.2. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ Only (1) and (4) are true.
   Ⓑ Only (2) and (3) are true.
   Ⓒ None of the others
   Ⓓ All the statements are true.
   Ⓔ Only (1) and (2) are true.

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ 0.4 Ⓑ 0.1 Ⓒ None of the others. Ⓓ 0.9 Ⓔ 0.6

3. Find the probability that a student skips 4 questions.

Ⓐ 0.4897 Ⓑ 0.0897 Ⓒ None of the others. Ⓓ 0.1897 Ⓔ 0.2897

4. If a student skips 4 questions, find the expected score of the student.

Ⓐ None of the others. Ⓑ 5.58 Ⓒ 4.58 Ⓓ 7.08 Ⓔ 6.08

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of less than 809.5 hours. A random sample of 10 bulbs is selected and their lifespans are recorded as follows: 746.8, 807, 730.2, 776.9, 827.3, 832.2, 814.3, 783.9, 834.2, 875. Based on this random sample, the company wants to test whether its claim is supported at the 0.05 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

Ⓐ H0: $\mu \neq 809.5$, H1: $\mu = 809.5$.

Ⓑ H0: $\mu \leq 809.5$, H1: $\mu > 809.5$.

Ⓒ H0: $\mu = 809.5$, H1: $\mu \neq 809.5$.

Ⓓ H0: $\mu > 809.5$, H1: $\mu \leq 809.5$.

Ⓔ H0: $\mu = 809.5$, H1: $\mu < 809.5$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

Ⓐ Student distribution with degree of free dom 9.

Ⓑ Normal distribution with unknown variance.

Ⓒ Normal distribution with with the standard deviation of 13.8436.

Ⓓ None of the others.

Ⓔ Normal distribution with the mean 802.78 and unknown variance.

7. Calculate the statistic for the hypothesis test.

Ⓐ None of the others. Ⓑ -39.3744 Ⓒ -0.4854 Ⓓ -22.7079 Ⓔ -50.4854

8. Find a 95% confidence interval for the mean lifespan of the light bulbs produced by the company.

Ⓐ None of the others. Ⓑ [775.6465 , 829.9135] Ⓒ [786.7965 , 832.2035] Ⓓ [771.4658 , 834.0942] Ⓔ [777.4047 , 828.1553]

9. Suppose further that the standard deviation of the bulb lifespan is 41.81 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 95% confidence interval for the expected lifespan μ is no more than 5 (hours)?

Ⓐ 289 Ⓑ 234 Ⓒ 219 Ⓓ 254 Ⓔ 269

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 15$, $\sum_{i=1}^{n} x_i = 542$, $\sum_{i=1}^{n} y_i = 185.386$, $\sum_{i=1}^{n} x_i^2 = 23350$,
$\sum_{i=1}^{n} y_i^2 = 2705.5276$ and $\sum_{i=1}^{n} x_i y_i = 7946.45$.

10. Find the sample correlation coefficient for this sample.

Ⓐ 0.7522 Ⓑ 0.999 Ⓒ None of the others. Ⓓ 0.9073 Ⓔ 0.6739

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to
    - Ⓐ decrease approximately by 0.3314 $\times 10^9$ particles.
    - Ⓑ increase approximately by 0.3857 $\times 10^9$ particles.
    - Ⓒ increase approximately by 0.3314 $\times 10^9$ particles.
    - Ⓓ decrease approximately by 0.3857 $\times 10^9$ particles.
    - Ⓔ None of the others.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.
    Ⓐ None of the others. Ⓑ 0.344 Ⓒ 0.0648 Ⓓ 0.0041 Ⓔ 0.0071

13. Find a 99% confidence interval for the slope β_1 of the linear regression line.
    Ⓐ [-0.8305,1.4932] Ⓑ [0.3189,0.3438] Ⓒ [0.3207,0.342] Ⓓ [0.3204,0.3423] Ⓔ None of the others.

14. Estimate the y-intercept of the linear regression line.
    Ⓐ 0.4277 Ⓑ 0.7234 Ⓒ 0.6791 Ⓓ 0.3857 Ⓔ None of the others.

15. Using the linear regression model, approximate the residual for the predicted value $y = 6.29$ at $x = 18$. Ⓐ -0.0603 Ⓑ -0.4944 Ⓒ -0.0566 Ⓓ -0.1906 Ⓔ None of the others.

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 438 products made by printer 1 and 198 products made by printer 2. Among these chosen products, there are 26 defective products made by printer 1 and 4 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 2 at the significance level $\alpha = 0.05$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.
    Ⓐ None of the others. Ⓑ 0.0182 Ⓒ 0.407 Ⓓ 0.1848 Ⓔ 0.2959

17. Determine the appropriate hypothesis test.
    - Ⓐ z-test since the sample proportion follows a normal distribution.
    - Ⓑ t-test since the population distribution is unknown.
    - Ⓒ z-test since the sample proportion approximately follows a normal distribution.
    - Ⓓ z-test since we want to compare two mean values.
    - Ⓔ t-test since the population variance is unknown.

18. Calculate the statistic of the hypothesis test.
    Ⓐ 2.1569 Ⓑ 2.3789 Ⓒ 3.4899 Ⓓ None of the others. Ⓔ 4.0454

19. Which of the following statements is true?
    - Ⓐ Choose printer 1 since H0 could be rejected.
    - Ⓑ None of the others.
    - Ⓒ Choose printer 2 since H0 could be rejected.
    - Ⓓ Choose printer 1 since H0 could NOT be rejected
    - Ⓔ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$

20. Construct a 95 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ None of the others. Ⓑ [0.0096 , 0.0687] Ⓒ [0.0035 , 0.0748] Ⓓ [0.0036 , 0.0747] Ⓔ [0.0144 , 0.0639]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Lecturer:	*Date*	Approved by:	*Date*
Dr. Phan Thi Huong	*29/11/2023*	Dr. Nguyen Tien Dung	*29/11/2023*

. . . . . .

UNIVERSITY OF TECHNOLOGY - VNUHCM
Faculty of Applied Science

Final Exam	Academic year	2023-2024	Semester	1
	Exam date	December 18th, 2023		
Course title	Probability and Statistics		**Score**	
Course ID	MT2013	Sheet code	**3138**	
Duration	100 minutes	Shift	13:00	

Instructions to students:

- *At the beginning of the working time, you MUST fill in your full name and student ID on this question sheet. There are 22 questions on 4 pages.*
- *You are allowed to use your OWN materials and calculator. Total available score: 10.*
- *Do not round between steps. Round your final answers to 4 decimal places.*
- *For multiple-choice questions, choose the closest answers, mark your selection on both your multiple-choice answer sheet and question sheet. Your overall score will be deducted 20% of one multiple-choice question score for each wrong multiple-choice answer. For essay questions, show your work with all essential steps of calculations, analyses, and justifications on the question sheet. Submit your question sheet and your multiple-choice answer sheet.*

Student's full name:	Invigilator 1: . . . . .
Student Id: . . . . . Group:	Invigilator 2: . . . . .

Part I: Multiple Choice (7 points)

Questions 1 through 4 (L.O.1.1, L.O.2.1, L.O.4). A student takes a test consisting of 20 multiple-choice questions. In each question, only one choice is correct and students can only choose one answer for a question. Knowing that students receive 0.5 points for each correct answer and 0.1 points for each incorrect answer. In addition, students neither earn nor lose points for questions skipped and suppose that the probability of skipping a question is 0.25. If an answer is chosen, the probability that it is correct is 0.85 and the probability that it is incorrect is 0.15. The answers are independently chosen.

1. Which of the following statements is (are) always true?

   (1) The number of skipped questions is a binomial random variable.

   (2) The number of answered questions is a binomial random variable.

   (3) The number of questions answered correctly among unskipped questions is a binomial random variable.

   (4) The number of questions answered incorrectly among unskipped questions is a binomial random variable.

   Ⓐ None of the others

   Ⓑ All the statements are true.

   Ⓒ Only (1), (2), and (3) are true.

   Ⓓ Only (1) and (3) are true.

   Ⓔ Only (1) and (2) are true.

2. Select a question at random, what is the probability that it is correctly answered?

   Ⓐ None of the others. Ⓑ 0.1375 Ⓒ 0.4375 Ⓓ 0.7375 Ⓔ 0.6375

3. Find the probability that a student skips 4 questions.

Ⓐ 0.1897 Ⓑ 0.0897 Ⓒ 0.6897 Ⓓ None of the others. Ⓔ 0.4897

4. If a student skips 4 questions, find the expected score of the student.

Ⓐ 6.56 Ⓑ 8.56 Ⓒ 7.06 Ⓓ 5.56 Ⓔ None of the others.

Questions 5 through 9 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). An electronics company claims that the light bulbs it produces have an average lifespan of là 835.7 hours. A random sample of 10 bulbs is selected and their lifespans are recorded as follows: 812.3, 898.1, 745.1, 851.7, 816.3, 811.4, 929.1, 888, 815.9, 882.6. Based on this random sample, the company wants to test whether its claim is supported at the 0.05 level of significance. Assume that the lifespan of each light bulb is normally distributed.

5. Determine the null hypothesis and the alternative hypothesis.

Ⓐ H0: $\mu \neq 835.7$, H1: $\mu = 835.7$.

Ⓑ H0: $\mu = 835.7$, H1: $\mu \neq 835.7$.

Ⓒ H0: $\mu \geq 835.7$, H1: $\mu < 835.7$.

Ⓓ H0: $\mu > 835.7$, H1: $\mu \leq 835.7$.

Ⓔ H0: $\mu \leq 835.7$, H1: $\mu > 835.7$.

6. What is the distribution of the sample mean for the lifespan of the chosen bulbs?

Ⓐ Student distribution with degree of free dom 9.

Ⓑ Normal distribution with the mean 845.05 and unknown variance.

Ⓒ Normal distribution with unknown variance.

Ⓓ None of the others.

Ⓔ Normal distribution with with the standard deviation of 17.3362.

7. Calculate the statistic for the hypothesis test.

Ⓐ 28.3163 Ⓑ 0.5393 Ⓒ None of the others. Ⓓ 50.5383 Ⓔ 33.8718

8. Find a 95% confidence interval for the mean lifespan of the light bulbs produced by the company.

Ⓐ [805.8356 , 884.2644] Ⓑ [813.2728 , 876.8272] Ⓒ [811.0711 , 879.0289] Ⓓ None of the others. Ⓔ [807.2687 , 864.1313]

9. Suppose further that the standard deviation of the bulb lifespan is 67.19 hours. At least how many bulbs should be inspected to guarantee that the estimated error of the 95% confidence interval for the expected lifespan μ is no more than 6 (hours)?

Ⓐ 482 Ⓑ 442 Ⓒ 462 Ⓓ 457 Ⓔ 517

Questions 10 through 15 (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. In a study, the number of Cesium atoms (y, unit: 10^9 particles) cooled by lasers of various powers (x, unit: mW) were counted. Use the following information to answer the following questions:

$n = 15$, $\sum_{i=1}^{n} x_i = 580$, $\sum_{i=1}^{n} y_i = 18.26$, $\sum_{i=1}^{n} x_i^2 = 25232$,
$\sum_{i=1}^{n} y_i^2 = 26.3148$ and $\sum_{i=1}^{n} x_i y_i = 783.144$.

10. Find the sample correlation coefficient for this sample.

Ⓐ 0.6234 Ⓑ 0.72 Ⓒ 0.9122 Ⓓ 0.3147 Ⓔ None of the others.

11. If the power is increased by 1 unit, the number of Cesium atoms is expected to
    - Ⓐ decrease approximately by 0.1548 $\times 10^9$ particles.
    - Ⓑ decrease approximately by 0.0275 $\times 10^9$ particles.
    - Ⓒ increase approximately by 0.1548 $\times 10^9$ particles.
    - Ⓓ None of the others.
    - Ⓔ increase approximately by 0.0275 $\times 10^9$ particles.

12. Find an estimated standard deviation (estimated standard error) for $\hat{\beta}_1$.
    Ⓐ 0.4065 Ⓑ 0.0073 Ⓒ None of the others. Ⓓ 0.0155 Ⓔ 0.3306

13. Find a 95% confidence interval for the slope β_1 of the linear regression line.
    Ⓐ [0.0145,0.0405] Ⓑ [-0.3069,0.3618] Ⓒ [0.0116,0.0433] Ⓓ [0.0131,0.0419] Ⓔ None of the others.

14. Estimate the y-intercept of the linear regression line.
    Ⓐ 0.0828 Ⓑ 0.0375 Ⓒ 0.1548 Ⓓ None of the others. Ⓔ 0.5486

15. Using the linear regression model, approximate the residual for the predicted value $y = 0.61$ at $x = 24$. Ⓐ -0.2043 Ⓑ None of the others. Ⓒ -0.3522 Ⓓ -0.3276 Ⓔ -0.3418

Questions 16 through 20(L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). Two 3D printers were examined for quality. A printer is considered better if it produces fewer defective products and will be chosen for the company. The company randomly selects 436 products made by printer 1 and 426 products made by printer 2. Among these chosen products, there are 4 defective products made by printer 1 and 4 defective products made by printer 2. Does the data provide enough evidence for the company to choose printer 1 at the significance level $\alpha = 0.05$?

16. Under the null hypothesis H_0, estimate the standard deviation (standard error) of the difference in proportion of defective products made by the two printers.
    Ⓐ 0.062 Ⓑ None of the others. Ⓒ 0.5064 Ⓓ 0.0065 Ⓔ 0.2842

17. Determine the appropriate hypothesis test.
    - Ⓐ z-test since the sample proportion approximately follows a normal distribution.
    - Ⓑ t-test since the population variance is unknown.
    - Ⓒ z-test since we want to compare two mean values.
    - Ⓓ t-test since the population distribution is unknown.
    - Ⓔ z-test since the sample proportion follows a normal distribution.

18. Calculate the statistic of the hypothesis test.
    Ⓐ -2.033 Ⓑ -0.033 Ⓒ None of the others. Ⓓ -0.922 Ⓔ -1.4775

19. Which of the following statements is true?
    - Ⓐ Printer 2 is better than printer 1.
    - Ⓑ Printer 1 is better than printer 2.
    - Ⓒ Choose printer 1 since $\hat{p}_1 < \hat{p}_2$.
    - Ⓓ None of the others.
    - Ⓔ Choose printer 2 since $\hat{p}_2 < \hat{p}_1$

20. Construct a 95 % confidence interval for the true difference between the proportions of defective products made by these printers.

Ⓐ [-0.0109 , 0.0105] Ⓑ [-0.013 , 0.0126] Ⓒ None of the others. Ⓓ [-0.013 , 0.0126] Ⓔ [-0.013 , 0.0126]

Part II: Essay (3 points)

21. (L.O.1.1, L.O.2.1, L.O.4) A worker is responsible for maintaining a production line. Suppose that the time (in minutes) to fix a failure of the production line is a random variable following a uniform distribution on [1, 5] and failures are independent of each other.

(a) Find the probability that, among 5 random failures, there are exactly 2 failures that take the worker no more than 2 minutes to fix each of them.

(b) Suppose that there are 50 failures in one month. Find the probability that it takes the workers more than 180 minutes in total to fix all of these failures.

22. (L.O.1.2, L.O.2.1, L.O.2.2, L.O.4). The below table shows the data on the level of fine particles (PM2.5) within Ho Chi Minh City collected biweekly in the first quarter of each of three years, from 2020 to 2022. Assume further that the data satisfies the Anova model.

Year	Level of PM2.5					
2022	74	96	63	74	121	84
2021	78	92	68	71	99	86
2020	73	56	64	74	62	43

(a) Use the Anova method to compare the level of PM2.5 in Ho Chi Minh City in the first quarter of three years at the significance level of 5%.

(b) Construct a 95% confidence interval for the average level of PM2.5 in the first quarter of 2022.

(c) Construct a 95% confidence interval for the difference in mean level of PM2.5 in the first quarter between two years 2022 and 2020. What could we conclude about the comparison between these two mean values?

Answers Sheet

Question sheet code 3131:

1 B.	2 D.	3 A.	4 D.	5 D.	6 C.	7 D.	8 B.	9 D.	10 D.	11 A.
12 D.	13 D.	14 D.	15 D.	16 D.	17 C.	18 C.	19 B.	20 B.		

Question sheet code 3132:

1 D.	2 D.	3 B.	4 A.	5 E.	6 E.	7 C.	8 E.	9 B.	10 C.	11 A.
12 E.	13 B.	14 E.	15 C.	16 D.	17 E.	18 A.	19 D.	20 B.		

Question sheet code 3133:

1 B.	2 E.	3 C.	4 B.	5 B.	6 D.	7 D.	8 C.	9 D.	10 D.	11 B.
12 E.	13 A.	14 D.	15 E.	16 A.	17 A.	18 C.	19 C.	20 D.		

Question sheet code 3134:

1 E.	2 D.	3 A.	4 E.	5 B.	6 D.	7 B.	8 E.	9 D.	10 E.	11 A.
12 A.	13 B.	14 B.	15 B.	16 C.	17 C.	18 D.	19 A.	20 C.		

Question sheet code 3135:

1 B.	2 B.	3 B.	4 B.	5 E.	6 D.	7 C.	8 D.	9 A.	10 D.	11 C.
12 E.	13 E.	14 E.	15 B.	16 D.	17 E.	18 C.	19 E.	20 B.		

Question sheet code 3136:

1 B.	2 C.	3 D.	4 D.	5 A.	6 B.	7 E.	8 E.	9 D.	10 E.	11 E.
12 D.	13 C.	14 D.	15 A.	16 C.	17 D.	18 E.	19 C.	20 D.		

Question sheet code 3137:

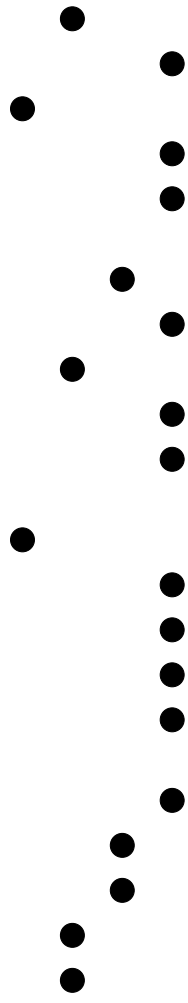
1 D.	2 E.	3 D.	4 E.	5 E.	6 B.	7 C.	8 D.	9 E.	10 B.	11 C.
12 D.	13 B.	14 D.	15 A.	16 B.	17 C.	18 A.	19 C.	20 B.		

Question sheet code 3138:

1 B.	2 E.	3 A.	4 A.	5 B.	6 C.	7 B.	8 A.	9 A.	10 B.	11 E.
12 B.	13 C.	14 C.	15 A.	16 D.	17 A.	18 B.	19 D.	20 B.		

3 1 3 1

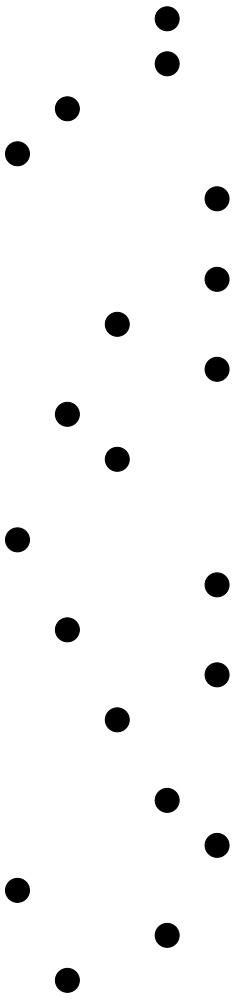
Probability and Statistics
OISP



3 1 3 2

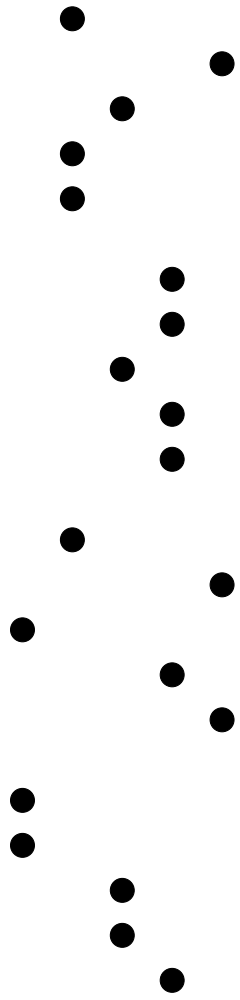
Probability and Statistics

OISP



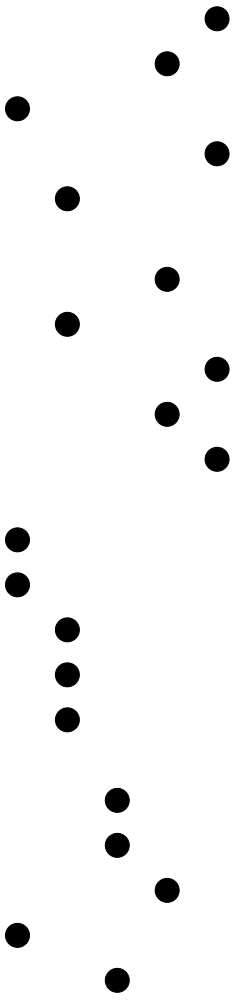
3 1 3 3

Probability and Statistics
OISP



3 1 3 4

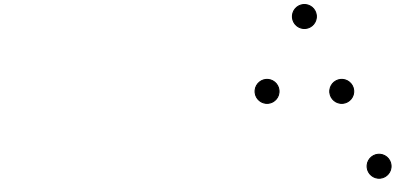
Probability and Statistics
OISP

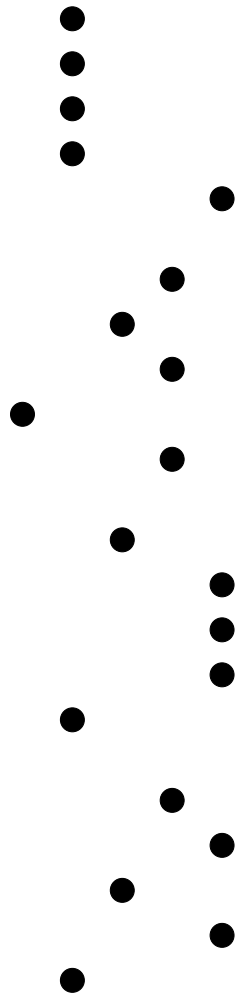


Probability and Statistics

OISP

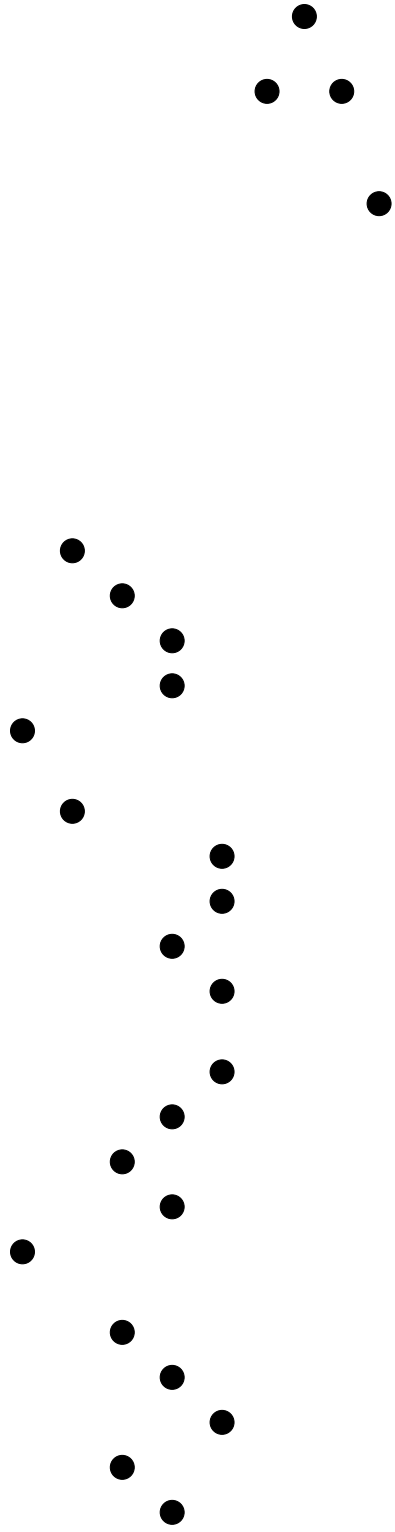
3 1 3 5





3 1 3 6

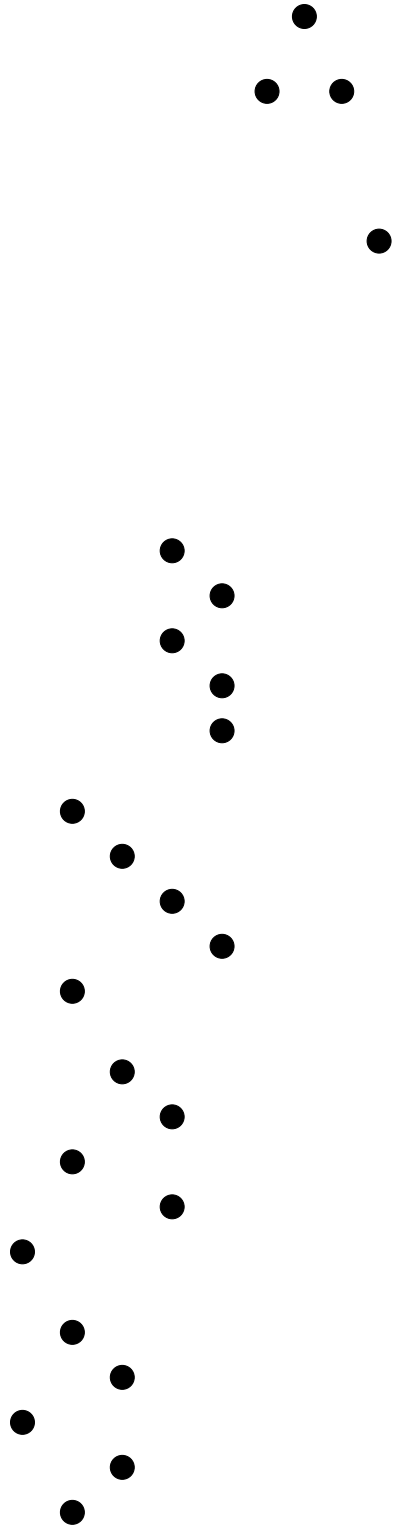
Probability and Statistics
OISP



Probability and Statistics

OISP

3 1 3 7



18/12/2023 13:00

3 1 3 8

PS MT2013

Probability and Statistics

OISP