NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF STATISTICS AND DATA SCIENCE

ST2334 PROBABILITY AND STATISTICS

FINAL EXAMINATION SAMPLE PAPER 1

(SEMESTER II, AY 2024/2025)

TIME ALLOWED: 120 MINUTES

Suggested Solutions

INSTRUCTIONS TO STUDENTS

1. Please write your student number only. **Do not write your name.**
2. This assessment contains 30 questions and comprises 1 printed pages.
3. The total marks is 60; marks are equal distributed for all questions.
4. Please answer ALL questions.
5. Calculators may be used.
6. This is an **OPEN BOOK** assessment. Only **HARD COPIES** of materials are allowed.

1. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
   A random variable X has the following probability function.

$$f_X(x) = \frac{x}{4}, \quad \text{for } x = 0.4, 0.6, 0.9, 2.1;$$

   and $f_X(x) = 0$ elsewhere.

   What type of random variable is X?

   (a) Continuous
   (b) Discrete
   (c) Unable to determine
   (d) None of the given options

   SOLUTION
   (b)

2. **TRUE/FALSE**
   Suppose $X \sim N(1, \sigma_1^2)$ and $Y \sim N(0, \sigma_2^2)$. $P(X < 1)$ is larger than $P(Y < 1)$ when $\sigma_1^2 > \sigma_2^2 > 0$.

   - TRUE
   - FALSE

   SOLUTION
   False
   No matter what are the values for σ_1 and σ_2, we must have $P(X < 1) = 0.5$ and $P(Y < 1) > 0.5$.

3. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
   A professor receives, on average, 21.7 emails from students the day before the midterm exam. To compute the probability of receiving at least 10 emails on such day, what type of probability distribution will he use?

   (a) Binomial distribution.
   (b) Poisson distribution.
   (c) Normal distribution.
   (d) Negative Binomial distribution.

   SOLUTION
   (b)

4. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
   Which of the following would be an appropriate null hypothesis?

   (a) The mean of a population is equal to 55.
   (b) The mean of a sample is equal to 55.
   (c) The mean of a population is greater than 55.
   (d) None of the given options

   SOLUTION
   (a)

5. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
   Let A, B be events in sample space S. Which of the following may **NOT** be true?

(a) $A \cap A' = \emptyset$

(b) $A \cup A' = S$.

(c) $(A \cup B)' = A' \cup B'$

(d) $A \cup B = A \cup (B \cap A')$

SOLUTION
(c)

6. **FILL IN THE BLANK**
Let X and Y be independent random variables such that $E(X) = 1$, $E(Y) = 2$, $V(X) = 3$, $V(Y) = 4$. Compute $V(2X - Y)$.

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION
16

7. **FILL IN THE BLANK**
We toss a fair die until the outcome "6" appears twice. Find the probability that it takes 5 tosses.

Answer: ________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

SOLUTION
$X \sim NB\left(2, \frac{1}{6}\right)$. $P(X = 5) = \binom{4}{1}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = 0.064.$

8. **TRUE/FALSE**
Under the usual random sampling setup, we can halve the standard deviation of the sample mean by doubling the sample size.

- TRUE
- FALSE

SOLUTION
False

9. **TRUE/FALSE**
Let (X, Y) be a random vector, then for any real numbers x and y, we must have

$$P(X \leq x, Y \leq y) = 1 - P(X > x, Y > y).$$

- TRUE
- FALSE

SOLUTION
FALSE.

10. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Which of the following is a valid cumulative distribution function?

(a) $F(x) = \begin{cases} 0 & x \le -1 \\ 0.3 & -1 < x \le 1 \\ 0.7 & 1 < x \le 10 \\ 1 & \text{elsewhere} \end{cases}$

(b) $F(x) = \begin{cases} 0 & x < -1 \\ 0.5 & -1 < x < 1 \\ 0.7 & 1 < x < 10 \\ 1 & \text{elsewhere} \end{cases}$

(c) $F(x) = \begin{cases} 0 & x < -1 \\ 0.6 & 1 < x \le 10 \\ 0.7 & -1 < x \le 1 \\ 1 & \text{elsewhere} \end{cases}$

(d) $F(x) = \begin{cases} 0 & x < -1 \\ 0.6 & -1 \le x < 1 \\ 0.7 & 1 \le x < 10 \\ 1 & \text{elsewhere} \end{cases}$

SOLUTION
(d)

11. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
The life time (in years) of a certain brand of light bulb follows an exponential distribution with the probability density function: $\frac{1}{2}\exp(-x/2)$. What is the probability that the bulb will last for more than 5 years, given that it has been working for 3 years?

(a) $\frac{1}{2}\exp(-5/2)$

(b) $\exp(-5/2)$

(c) $\frac{1}{2}\exp(-1)$

(d) $\exp(-1)$

SOLUTION
(d)

12. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
The Central Limit Theorem is important in statistics because

(a) for a large sample size, n, it says the population is approximately normal.

(b) for any population, it says the sampling distribution of the sample mean is approximately normal, regardless of the sample size.

(c) for a large sample size, n, it says the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.

(d) for any sized sample, it says the sampling distribution of the sample mean is approximately normal.

SOLUTION
(c)

13. **FILL IN THE BLANK**
John rolls a fair die 6 times independently. What is the probability that he will get numbers more than 2 at least twice?

Answer: ________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

SOLUTION
Let X = number of times to get numbers more than 2. $X \sim \text{Bin}(6, 2/3)$.
$P(X \ge 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - (0.001372 + 0.016461) = 0.9822$.

14. **FILL IN THE BLANK**

Jill's bowling scores are approximately normally distributed with mean 170 and standard deviation 20, while Jack's bowling scores are approximately normally distributed with mean 160 and standard deviation 15. If Jack and Jill each bowl one game, then assuming that their scores are independent random variables, the probability that the sum of the scores is higher than 340 is approximately equal to $\Phi(c)$. Find the value of c.

Answer: $c =$________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

Note: $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0,1)$.

SOLUTION

-0.4

Denote $Y = X_1 + X_2 \sim N(170+160, 20^2+15^2) = N(330, 25^2)$.

$$P(Y > 340) = P\left(\frac{Y-330}{25} > 0.4\right) = P(Z < -0.4).$$

15. **FILL IN THE BLANK**

An experiment was carried out to test whether mean weight gain for pigs fed ration A is higher than those fed ration B. Eight pairs of pigs were used. The rations were assigned at random to the two animals within each pair. The gain (in kilograms) after 45 days, assuming normally distributed, are given as follows.

Pairs	1	2	3	4	5	6	7	8	mean	sd
Ration A	30	17	18	21	22	30	24	27	23.625	5.0409
Ration B	26	18	15	20	21	25	27	23	21.875	4.1555
Difference, A - B	4	-1	3	1	1	5	-3	4	1.75	2.7646

Suppose that the pigs within each pair were littermates. What is the observed value of the test statistic in testing the alternative hypothesis that ration A is better, in terms of mean weight gain, than ration B at a 5% significance level?

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION

We have the paired data. $T = \dfrac{\overline{D}}{S_D/\sqrt{n}} = 1.75/(2.7646/\sqrt{8}) = 1.7904.$

16. **FILL IN THE BLANK**

The mean lifetime of 100 randomly selected pumps made by a particular factory was 200 days. Assuming it is known that the population standard deviation $\sigma = 40$, find a 95% confidence interval for the mean lifetime of pumps made by the factory.

Answer: (_________, _________).

Note: $z_{0.025} = 1.96$; $z_{0.05} = 1.64$.

(Provide your answers in decimal form and round them to two decimal places if necessary.)

SOLUTION
$\bar{x} = 200$; $n = 100$ is large; $\sigma = 40$ is known. Thus the 95% confidence interval is

$$\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 200 \pm 1.96\frac{40}{\sqrt{100}} = (192.16,\ \ 207.84).$$

17. **FILL IN THE BLANK**
We roll a fair die 3 times. Find the probability that the sum is equal to 5.

Answer: ________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

SOLUTION
The possible three numbers are $\{1,1,3\}$ or $\{1,2,2\}$; this leads to six possibilities:

$$(1,1,3),(1,3,1),(3,1,1);(1,2,2),(2,1,2),(2,2,1).$$

In total there are $6^3 = 216$ possibilities. Therefore the probability is $6/216 = 1/36 = 0.0278$.

18. **TRUE/FALSE**
Consider the Z-test for $H_0 : \mu = 0$ based on $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$. It turns out that $\bar{X} = 2.3$. The p-value for the one-sided test ($H_1 : \mu > 0$) is half of that for the two-sided test ($H_1 : \mu \neq 0$).

- TRUE
- FALSE

SOLUTION
True

19. **FILL IN THE BLANK**
A new COVID rapid test is able to correctly diagnose that you do not have the virus 90% of the time. However, if you do have the virus, it fails to detect it 25% of the time. Given that the overall COVID infection rate at a particular worker dorm is 20%, what is the probability of a worker being infected if his rapid test does not detect the virus?

Answer: ________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

SOLUTION

$$P(T'|D') = 0.9, \quad P(T'|D) = 0.25, \quad P(D) = 0.2$$

$$P(D|T') = \frac{P(T'|D)P(D)}{P(T'|D)P(D) + P(T'|D')P(D')} = \frac{(0.25)(0.2)}{(0.25)(0.2) + (0.9)(0.8)} = 0.0649.$$

20. **MULTIPLE RESPONSE: CHOOSE ALL ANSWERS THAT APPLY**
Which of the following can happen if the hypothesis is rejected.

(a) p-value$> \alpha$;

(b) test statistic falls in the rejection region;

(c) type I error occurs;

(d) type II error occurs.

SOLUTION
(b), (c).

21. **FILL IN THE BLANK**
Let $X_1, X_2, \ldots, X_{100}$ be independent and identically distributed continuous random variables with $E(X_i) = 5$ and $V(X_i) = 4$. Compute approximately $P\left(\sum_{i=1}^{100} X_i > 510\right)$.

Answer: ________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

Note: $\Phi(0.5) = 0.6915$, $\Phi(1) = 0.8413$, $\Phi(1.5) = 0.9332$; where $\Phi(\cdot)$ denotes the cumulative distribution function for $N(0,1)$.

SOLUTION
0.3085

By CLT, $\sum_{i=1}^{100} X_i \approx N(500, 400)$.

$$P\left(\sum_{i=1}^{100} X_i > 510\right) = P\left(Z > \frac{510-500}{\sqrt{400}}\right) = P(Z > 0.5) = 1 - 0.6915 = 0.3085.$$

22. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Which of the following statements about probability is **INCORRECT**?

(a) If A and B are two events and $P(A \cap B) = P(A)$, then $P(A \cap B') = 0$.

(b) Let S be the sample space and let A be an event. If there exists an $x \in S$ but $x \notin A$, then $P(A) < 1$.

(c) Let A and B be two events; then $P(A \cup B) \leq P(A) + P(B)$.

(d) Let A and B be independent events. If $P(A) > 0$ and $P(B) > 0$, then A and B are not mutually exclusive.

SOLUTION
(b)

23. **FILL IN THE BLANK**
Consider the following game:

- First round: the gamer flips a fair coin. If he gets a head, he loses; otherwise he wins the round.
- Second round: the gamer flips two fair coins independently. If he gets two heads, he loses; otherwise he wins the round.
- Third round: the gamer flips three fair coins independently. If he gets three heads, he loses; otherwise he wins the round.
- And so on.

What is the probability that the gamer will make his first win in the 4th round?

Note: you can assume that from rounds to rounds, the flips are independently conducted.

Answer: ________

(Provide your answer in decimal form and round it to four decimal places if necessary.)

SOLUTION

Let $A_i = \{$the gamer wins the ith round$\}$. The question is asking

$$P(A_1' \cap A_2' \cap A_3' \cap A_4) = P(A_1')P(A_2')P(A_3')P(A_4) = \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{8}\cdot\frac{15}{16} = 0.01464844.$$

24. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**

Let $\{X_1, X_2, \ldots, X_{25}\}$ be a random sample from the $N(\mu, 2^2)$ distribution. Consider the hypotheses: $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$, and the test statistic $Z = \dfrac{\bar{X}}{2/\sqrt{25}}$. Suppose that we reject H_0 if $|z_{obs}| > 2$; and do not reject H_0 otherwise. What is the probability that we will not reject H_0 given that the true value of μ is equal to 2?

Note: $\Phi(z)$ denotes the c.d.f. of the standard normal distribution: $\Phi(z) = Pr(\widetilde{Z} \leq z)$ with $\widetilde{Z} \sim N(0,1)$.

(a) $\Phi(-3) - \Phi(-7)$

(b) $\Phi(2) - \Phi(-2)$

(c) $\Phi(-2) - \Phi(-6)$

(d) $\Phi(-3) - \Phi(-8)$

SOLUTION

(a)

We are to compute the type II error probability:

$$\begin{aligned} Pr(|Z| \leq 2|\mu = 2) &= Pr\left(-2 \leq \frac{\bar{X}}{2/5} \leq 2 \middle| \mu = 2\right) \\ &= Pr\left(-2 - \frac{2}{2/5} \leq \frac{\bar{X}-2}{2/5} \leq 2 - \frac{2}{2/5} \middle| \mu = 2\right) \\ &= Pr(-7 \leq \widetilde{Z} \leq -3) = \Phi(-3) - \Phi(-7). \end{aligned}$$

25. **TRUE/FALSE**

For any $\theta \in \mathbb{R}$, the function

$$f(x) = \begin{cases} \theta - x & \text{if } \theta - 1 \leq x < \theta \\ x - \theta & \text{if } \theta \leq x \leq \theta + 1 \\ 0 & \text{elsewhere} \end{cases}$$

can serve as a probability density function of some distribution, whose population mean is equal to θ.

- TRUE
- FALSE

SOLUTION
True

26. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
In a course, students are graded based on a "normal curve". For example, students within 0.5 standard deviation from the mean receive a C; between 0.5 and 1.0 standard deviation above the mean receive a $C+$; between 1.0 and 1.5 standard deviation above the mean receive a B; between 1.5 and 2.0 standard deviation receive a $B+$, etc. The class average in an exam was 60 with a standard deviation of 10. What are the bounds for a B grade and the percentage of students who will receive a B grade?

(a) (65, 75), 24.17%
(b) (65, 75), 12.08%
(c) (70, 75), 18.38%
(d) (70, 75), 9.19%

Note: $\Phi(1) = 0.8413$, $\Phi(1.5) = 0.9332$, $\Phi(2) = 0.9772$, where $\Phi(\cdot)$ denotes the cumulative distribution of $N(0,1)$.

SOLUTION
(d)

27. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**

Flip an unfair coin. If a head shows, roll a fair die and report the number; otherwise, roll a fair die twice and report the summation minus 1. Then $P(6 \text{ is reported}) = ?$

(a) 1/6
(b) 1/4
(c) 1/2
(d) can not tell

SOLUTION
(a)
Let $X = 1$ if coin shows head; $Y =$ number obtained from rolling the die (once, or summation of twice).

$$\begin{aligned} P(6 \text{ is reported}) &= P(\{X=1, Y=6\} \text{ or } \{X=0, Y=7\}) \\ &= P(X=1)P(Y=6|X=1) + P(X=0)P(Y=7|X=0) \\ &= p(1/6) + (1-p)(1/6) = 1/6. \end{aligned}$$

Note that

$$\begin{aligned} P(Y=7|X=0) &= P(\text{rolling two dice and get sum } = 7) \\ &= P((1,6) \text{ or } (2,5) \text{ or } (3,4) \text{ or } (4,3) \text{ or } (5,2) \text{ or } (6,1)) \\ &= 6/36 = 1/6. \end{aligned}$$

28. **FILL IN THE BLANK**
An urn contains 3 red balls and 5 white balls. 4 balls are drawn uniformly at random without replacement from the urn. Let X be the random variable for the number of red balls drawn. If $X \leq 1$, you win \$$X$. If $X > 1$, you flip a fair coin. If the coin comes up heads, you double your winnings and win a total of \$$2X$. If the coin comes up tails, you

still win $\$X$. Let W be the random variable for your winnings. Find the probability that W is odd.

Answer: ________

(Provide your answer in decimal form and round it to four decimal places if necessary.)

SOLUTION
0.4642
X can take on values from $\{0,1,2,3\}$, so W can only take on values from $\{0,1,2,3,4,6\}$.

$$\begin{aligned} P(W \text{ is odd}) &= P(W=1)+P(W=3) \\ &= P(X=1)+P(X=3 \text{ and coin flips tail}) \\ &= \frac{\binom{3}{1}\binom{5}{3}}{\binom{8}{4}} + \frac{\binom{3}{3}\binom{5}{1}}{\binom{8}{4}} \cdot \frac{1}{2} \\ &= \frac{3}{7} + \frac{1}{14} \cdot \frac{1}{2} \\ &= \frac{13}{28} = 0.4642. \end{aligned}$$

29. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Let X be a random variable with density function

$$f(x) = \begin{cases} k\sqrt{x} & \text{, for } 0 \le x \le 1 \\ ke^{\frac{1-x}{2}} & \text{, for } x > 1 \\ 0 & \text{, otherwise} \end{cases}$$

where k is a constant. What is the value of k?

(a) 1/2
(b) 2/5
(c) 3/8
(d) 4/9

SOLUTION
(c)

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\,dx &= \int_0^1 k\sqrt{x}\,dx + \int_1^{\infty} ke^{\frac{1-x}{2}} \\ &= \left[\frac{2}{3}kx^{\frac{3}{2}}\right]_0^1 + \left[-2ke^{\frac{1-x}{2}}\right]_1^{\infty} \\ &= \frac{2}{3}k - (-2k) = \frac{8}{3}k \\ &\Rightarrow \frac{8}{3}k = 1 \Rightarrow k = \frac{3}{8}. \end{aligned}$$

30. **MULTIPLE RESPONSE: CHOOSE ALL ANSWERS THAT APPLY**
Let $X_1, X_2, \ldots, X_{n_1}$ be independent and identically distributed (i.i.d.) random variables with population mean μ_1; let $Y_1, Y_2, \ldots, Y_{n_2}$ be i.i.d. random variables with population mean μ_2; let $U_1, U_2, \ldots, U_{n_3}$ be i.i.d. random variables with population mean 4. All these random variables have the unknown but common variance σ^2. Which of the following is/are unbiased estimator(s) for σ^2?

(a) $$\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2+\sum_{k=1}^{n_3}(U_k-4)^2}{n_1+n_2+n_3-2}.$$

(b) $$\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2+\sum_{k=1}^{n_3}(U_k-4)^2}{n_1+n_2+n_3-3}.$$

(c) $$\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2+\sum_{k=1}^{n_3}(U_k-\bar{U})^2}{n_1+n_2+n_3-2}.$$

(d) $$\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2+\sum_{k=1}^{n_3}(U_k-\bar{U})^2}{n_1+n_2+n_3-3}.$$

SOLUTION
(a), (d).

END OF PAPER