

MATH 203: PRINCIPLES OF STATISTICS 1
FALL 2024

ASSIGNMENT 3 – SOLUTIONS

1. A survey of Canadian teenagers reported that 72% of this population used a fee-based website to download music. You decide to interview 25 people in this age group; their responses can be treated as independent events.
- (a) What is the distribution of the number, Y , in the sample who have used a fee-based website to download music? Explain your answer. 2 MARKS
 - (b) What is the probability that at least 20 (that is, 20 or more) people in the sample have used a fee-based website to download music? 3 MARKS
 - (c) What is the expected number of people in the sample who have used a fee-based website to download music? What is the population standard deviation? Explain your calculation. 3 MARKS

Show your method of calculation in each case.

Solutions:

- (a) This is the $Binomial(25, 0.72)$ distribution, as the sample are a collection of independent binary draws from the population. 2

- (b) Using R function `pbinom`, this is

$$\begin{aligned} P(Y \geq 20) &= 1 - P(Y \leq 19) \\ &= 1 - \text{pbinom}(19, 25, 0.72) = 1 - 0.7414612 = 0.2585388. \end{aligned}$$

3

- (c) The expectation of the $Binomial(n, p)$ is np , so here the expectation is $25 \times 0.72 = 18$. The standard deviation is $\sqrt{npq} = \sqrt{25 \times 0.72 \times 0.28} = 2.2450$. 3

2. A missile defence system uses satellite imagery to detect and track missiles. Suppose that the probability of a missile will be detected is 0.74. Consider a sample of 15 simulated missile tracks. Let random variable Y be the number of these tracks on which the system detects the object. A suitable probability model for Y is determined to be $Binomial(n, p)$.

- (a) Explain briefly why the Binomial model for Y is appropriate. 1 MARK
- (b) Give the value of p and n for the Binomial distribution. 2 MARKS
- (c) Find $P(Y = 8)$; show your working. 2 MARKS
- (d) Find $P(Y \geq 8)$; show your working. 2 MARKS
- (e) Find $E(Y)$. Explain the calculation. 1 MARK

Solutions:

(a) The Binomial is appropriate because we have a series of identical success/failure trials where the probability of success is the same, and the outcomes are independent.

1

(b) $n = 15$ and $p = 0.74$

2

(c) We can compute $P(Y = 8)$ directly

```
choose(n, 8) * p^8 * (1-p)^7  
## [1] 0.04647461
```

numerically using the `dbinom`

$$P(Y = 8) = \binom{15}{8} (0.74)^8 (1 - 0.74)^6 = \text{dbinom}(8, 15, 0.74) = 0.0464746$$

2

(d) We can compute $P(Y \geq 8)$ numerically using the `dbinom`

$$P(Y \geq 8) = \sum_{j=8}^{15} P(Y = j) = 0.9780988$$

or using the `pbinom`

$$P(Y \geq 8) = 1 - P(Y \leq 7) = 1 - \text{pbinom}(7, 15, 0.74) = 0.9780988$$

2

(e) From lectures $E(Y) = np = 15 \times 0.74 = 11.1$

1

3. In a clinical study, a new treatment is given to a sequence of trial participants. The goal of the treatment is provide pain relief over a 12 hour period after the treatment is administered. The probability that a participant reports pain relief of this duration is 0.92, and it may be assumed that reports from different participants are independent. The study is halted when a patient in the sequence does not report pain relief of the 12 hour duration. What is the expected number of patients that will be treated (including the final patient who reports inadequate pain relief) before the study is halted? Justify your answer and show your calculation. 4 MARKS

Solutions: Here we need to use the Geometric distribution, and the number of patients up until the trial is halted is $Y \sim \text{Geometric}(0.08)$. Therefore, using the formula from lectures,

$$E(Y) = \frac{1}{p} = \frac{1}{0.08} = 12.5.$$

4

4. When printing an article, the probability that a character is misprinted is 0.0002, independently of what happens to other characters. A given article comprises 25000 characters. Compute the probability that the total number of misprints exceeds five

(a) using an exact Binomial calculation

4 MARKS

(b) using a Poisson approximation

4 MARKS

Explain your working in each case.

Solutions: Let Y be the number of misprints.

- (a) We have $Y \sim \text{Binomial}(n, p)$ where $n = 25000$ and $p = 0.0002$. Here we can compute using `pbinom`

$$\begin{aligned} P(Y > 5) &= 1 - P(Y \leq 5) \\ &= 1 - \text{pbinom}(5, 25000, 0.0002) \\ &= 1 - 0.6159607 = 0.3840393. \end{aligned}$$

4

- (b) By the Poisson approximation we have that approximately $Y \sim \text{Poisson}(\lambda)$ where $\lambda = np = 5$. Here we can compute using `ppois`

$$\begin{aligned} P(Y > 5) &= 1 - P(Y \leq 5) \\ &\approx 1 - \text{ppois}(5, 5) \\ &= 1 - 0.6159607 = 0.3840393. \end{aligned}$$

4

5. The physical fitness of a patient is often measured by their maximum oxygen uptake (recorded in millilitres per kilogram, ml/kg). The population mean maximum oxygen uptake for cardiac patients who regularly participate in sports or exercise programs was found to be 24.1, with a population standard deviation of 6.30 (Adapted Physical Activity Quarterly, Oct. 1997).

Assuming that measurements of the maximum oxygen uptake for cardiac patients who regularly participate in sports or exercise programs can be represented by the Normal distribution:

- (a) Compute the probability that a patient has a maximum oxygen uptake of at least 20 ml/kg?

2 MARKS

- (b) Compute the probability that a patient has a maximum oxygen uptake of 10.5 ml/kg or lower?

2 MARKS

- (c) A patient has a maximum oxygen uptake of 10.5 ml/kg. Is it probable that this patient participates regularly in sports or exercise programs? Justify your answer.

4 MARKS

Show your working in each case. Use probability tables for the Normal distribution provided for the course.

Solutions: Let Y be the maximum oxygen uptake, so that $Y \sim \text{Normal}(\mu, \sigma^2)$ with $\mu = 24.1$ and $\sigma = 6.3$.

(a) Here we need $P(Y > 20)$. We use the z-transform approach

$$\begin{aligned} P(Y > 20) &= P(Z > (20 - 24.1)/6.3) = P(Z > -0.6507937) \\ &= P(Z < 0.6507937) && \text{symmetry} \\ &\approx P(Z < 0.65) && \text{to use tables} \\ &= 0.7422 \end{aligned}$$

The exact value is 0.7424102. 2

(b) Here we need $P(Y \leq 10.5)$. We use the z-transform approach

$$\begin{aligned} P(Y \leq 10.5) &= P(Z \leq (10.5 - 24.1)/6.3) = P(Z \leq -2.1587302) \\ &= 1 - P(Z \geq 2.1587302) && \text{symmetry} \\ &\approx 1 - P(Z \leq 2.16) && \text{to use tables} \\ &= 1 - 0.9846 \\ &= 0.0154. \end{aligned}$$

The exact value is 0.0154356. 2

(c) If the patient has a maximum oxygen uptake of 10.5 ml/kg, then, as observing a value as low as this is a very low probability event, we can interpret this as an indication that the patient does **not** participate in sports or exercise programs. That is, the probability that a patient has a value lower than this is given by (b), and this probability is low. 4

Here the question said to use tables, so the answers need to demonstrate that that is done (i.e. rounding z to two d.p. to use tables) , although can verify using exact calculation.

6. A study suggests that the birth weight of a full-term baby is distributed as a Normal random variable with mean 3500 grams (g) and standard deviation 600g. What is the probability that the birth weight of a randomly selected full-term baby

- (a) exceeds 4000g? 2 MARKS
- (b) lies between 3000g and 4000g? 2 MARKS
- (c) is either less than 2000g or greater than 5000g ? 2 MARKS

Show your working in each case. Use probability tables for the Normal distribution provided for the course.

Solutions: Let Y be the birthweight, so that $Y \sim \text{Normal}(\mu, \sigma^2)$, $\mu = 3500$ and $\sigma = 600$.

(a) Here using the z-transform approach

$$\begin{aligned} P(Y > 4000) &= P(Z > (4000 - 3500)/600) \\ &\approx P(Z > 0.83) \\ &= 1 - P(Z \leq 0.83) \\ &= 1 - 0.7967 = 0.2033 \end{aligned}$$

The exact value is 0.2023284.

2

(b) Here using the z-transform approach

$$\begin{aligned} P(3000 < Y < 4000) &= P((3000 - 3500)/600 < Z < (4000 - 3500)/600) \\ &= P(-0.8333333 < Z < 0.8333333) \\ &\approx P(Z < 0.83) - P(Z < -0.83) \\ &= P(Z \leq 0.83) - (1 - P(Z \leq 0.83)) \\ &= 0.7967 - (1 - 0.7967) \\ &= 0.7967 - 0.2033 \\ &= 0.5934 \end{aligned}$$

The exact value is 0.5953432.

2

(c) Here using the z-transform approach, realizing that

$$P(Y < 2000) + P(Y > 5000) = 1 - P(2000 < Y < 5000)$$

and

$$\begin{aligned} P(2000 < Y < 5000) &= P((2000 - 3500)/600 < Z < (5000 - 3500)/600) \\ &= P(-2.5 < Z < 2.5) \\ &\approx P(Z < 2.5) - P(Z < -2.5) \\ &= P(Z \leq 2.5) - (1 - P(Z \leq 2.5)) \\ &= 2P(Z \leq 2.5) - 1 \\ &= 2 \times 0.9938 - 1 \\ &= 0.9876 \end{aligned}$$

so therefore the required probability is 0.0124. The exact value is 0.0124193.

2

Here the question said to use tables, so the answers need to demonstrate that that is done (i.e. rounding z to two d.p. to use tables), although can verify using exact calculation.

7. A machine that cuts corks for wine bottles operates in such a way that the distribution of the diameter of corks is represented by a Normal distribution with mean 3cm and standard deviation 0.15 cm. The specifications require corks with diameters between 2.85cm and 3.15cm. A cork that does not meet the specifications is considered defective (a cork that is too small leaks and causes the wine to deteriorate; a cork that is too large does not fit in the bottle).

(a) What proportion of corks produced by this machine are defective? 2 MARKS

(b) If you choose 15 corks at random produced by this machine, what is the probability that at most 5 of them (that is, 5 or fewer) will be defective? 3 MARKS

(c) What machine standard deviation would be required if the manufacturer insists that the probability that a cork is defective must be no more than 0.02? 3 MARKS

Justify your answer and show your working in each case.

Solutions: Let Y be the cork size for one cork, so that $Y \sim \text{Normal}(\mu, \sigma^2)$, $\mu = 3$ and $\sigma = 0.15$.

(a) We want $P(Y < 2.85) + P(Y > 3.15)$ and by symmetry, these two component probabilities are equal. We have by the z-transform

$$P(Y > 3.15) = 1 - P(Y < 3.15) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587$$

so the required probability is 0.3174. 2

(b) If we define X as the random variable that counts the number of defective corks, we have that $X \sim \text{Binomial}(15, 0.3174)$. Therefore using `pbinom`

$$P(X \leq 5) = \text{pbinom}(5, 15, 0.3174) = 0.6688436$$

3

(c) We need to find a σ such that, by the z-transform method

$$2P(Z > 0.15/\sigma) \leq 0.02 \quad \text{i.e.} \quad P(Z > 0.15/\sigma) \leq 0.01$$

that is

$$P(Z < 0.15/\sigma) \geq 0.99$$

Using the tables in reverse, we look up the value of z that provides a cumulative probability closest to this value but above it. We see that if $z = 2.33$, the probability is 0.9901. Therefore we require

$$\frac{0.15}{\sigma} = 2.33 \quad \text{i.e.} \quad \sigma = \frac{0.15}{2.33} = 0.0643777$$

With this choice, the probability of being defective is 0.0198062. 3