Data analytics and visualisation Assignment II

Dr. T. Ansah-Narh

April 5, 2024

1. In a cybersecurity scenario, if two security breaches A and B are dependent, then

   (a) $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

   (b) $P(A \cap B) = P(A) + P(B)$

   (c) $P(A \cap B) = P(A) \times P(B \mid A)$

   (d) $P(A \cap B) = P(B) \times P(A \mid \bar{B})$

2. Which of the following statements are correct?

   (a) If the data privacy laws are independent then they are also mutually exclusive.

   (b) If the data privacy laws are mutually exclusive then they are dependent.

   (c) If the data privacy laws are independent, then they cannot be mutually exclusive.

   (d) If the data privacy laws are mutually exclusive then they are independent.

3. In an IT company, the number of software bugs reported hourly follows the probability distribution in Table 1.

Table 1: Hourly Occurrence of Software Bugs

Number of bugs (X)	0	1	2	3	4	5	6
Probability [P(X)]	0.15	0.20	0.25	0.20	0.15	0.04	0.01

Calculate

(a) the expected number of bugs X reported per hour,

(b) the expected number of bugs $4X - 2$ reported per hour

(c) the variance of bugs X and $4X - 2$,

(d) the standard deviation for the discrete random variable.

4. According to a survey on cybercrime activities, the time taken by IT experts to resolve a security breach follows a normal distribution with a mean of 45 minutes and a standard deviation of 8 minutes. Assume normality.

   (a) What proportion of IT experts take more than 60 minutes to resolve a security breach?

   (b) What proportion of security breaches are resolved in less than 30 minutes?

   (c) What proportion of security breaches are resolved between 40 minutes and 50 minutes?

5. Estimated value of a parameter from a sample is called ...

   (a) Sample statistic
   (b) Population parameter
   (c) Unbiased estimator
   (d) Sampling

6. In a random sampling, the sample subjects are selected from the dataset using ...

   (a) Normal distribution
   (b) Uniform distribution
   (c) Binomial distribution
   (d) Random distribution

7. An estimate of population parameter is unbiased when ...

   (a) It has minimum variance
   (b) It has minimum error
   (c) Its expected value is same as the population parameter
   (d) It has minimum standard deviation

8. In an IT security audit, the time taken by employees to log into the system (in seconds) is given in Table 2. Calculate the 95% confidence interval for the average login time.

Table 2: Login Times (in seconds)

12	16	20	9	25	18	22	27	14	8
32	28	21	17	19	23	26	30	30	24
11	10	40	9	12	38	8	21	25	26

9. The data shown below describes the processing times for legal documents between 2010 and 2020. The times measured in hours were obtained as follows:

   Assume that the standard deviation is known to be $\sigma = 0.7$.

Table 3: Document Processing Times (in hours)

3.5	4.2	5.1	6.2	4.8	5.3	4.9	4.7	4.6	5.5	6.1

(a) Construct a 99% two-sided confidence interval on the mean processing time.

(b) Construct a 95% lower-confidence bound on the mean processing time.

(c) Suppose that we wanted to be 95% confident that the error in estimating the mean processing time is less than 1 hour. What sample size should be used?

10. In a legal case, the number of hours spent in court by lawyers follows a Poisson distribution with a mean of 10 hours per day. Calculate the probability of:

    (a) Exactly 8 hours spent in court by lawyers in a day.

    (b) At most 5 hours spent in court by lawyers in a day.

    (c) More than 12 hours spent in court by lawyers in a day.

11. In an IT department, the number of successful login attempts per hour follows a binomial distribution with $n = 20$ and $p = 0.8$. Calculate:

    (a) The probability of exactly 15 successful login attempts in an hour.

    (b) The probability of at least 18 successful login attempts in an hour.

    (c) The expected number of successful login attempts in an hour.

    (d) The variance and standard deviation of successful login attempts in an hour.

12. According to IT regulations, the proportion of websites compliant with security standards is estimated to be 0.75. In a random sample of 100 websites, calculate:

    (a) The probability that exactly 80 websites are compliant with security standards.

    (b) The probability that at most 70 websites are compliant with security standards.

    (c) The average number of compliant websites in the sample.

    (d) The variance and standard deviation of the number of compliant websites in the sample.

13. In an e-discovery process, the time taken to review documents follows a normal distribution with a mean of 120 minutes and a standard deviation of 25 minutes. Assuming normality:

(a) What proportion of reviewers spend more than 150 minutes on document review?

(b) What proportion of reviewers spend less than 90 minutes on document review?

(c) What proportion of reviewers spend between 100 minutes and 140 minutes on document review?

14. In a litigation case, the time taken by different parties to submit evidence (in days) is given in Table 4. Calculate the 95% confidence interval for the average time taken to submit evidence.

Table 4: Time Taken to Submit Evidence (in days)

14	20	22	9	25	18	26	30	16	21
32	28	21	17	19	23	26	30	30	24
11	10	40	9	12	38	8	21	25	26