PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-2 - Random Variables

QUESTION BANK

Normal Distribution and Chebyshev's inequality

Exercises for Section 4.5
[Text Book Exercise – Section 4.5 – Q. No. [1 – 26] – Pg. No. [252 - 256]]

1. Find the area under the normal curve

   a) To the right of $z = -0.85$.
   b) Between $z = 0.40$ and $z = 1.30$.
   c) Between $z = -0.30$ and $z = 0.90$.
   d) Outside $z = -1.50$ to $z = -0.45$.

2. Find the area under the normal curve

   a) To the left of $z = 0.56$.
   b) Between $z = -2.93$ and $z = -2.06$.
   c) Between $z = -1.08$ and $z = 0.70$.
   d) Outside $z = 0.96$ to $z = 1.62$.

3. Let $Z \sim N(0, 1)$. Find a constant c for which

   a) $P(Z \geq c) = 0.1587$
   b) $P(c \leq Z \leq 0) = 0.4772$
   c) $P(-c \leq Z \leq c) = 0.8664$
   d) $P(0 \leq Z \leq c) = 0.2967$
   e) $P(|Z| \geq c) = 0.1470$

4. If $X \sim N(2, 9)$, compute

   a) $P(X \geq 2)$
   b) $P(1 \leq X < 7)$

c) $P(-2.5 \leq X < -1)$
d) $P(-3 \leq X - 2 < 3)$

5. The lifetime of a battery in a certain application is normally distributed with mean $\mu = 16$ hours and standard deviation $\sigma = 2$ hours.

   a) What is the probability that a battery will last more than 19 hours?
   b) Find the 10th percentile of the lifetimes.
   c) A particular battery lasts 14.5 hours. What percentile is its lifetime on?
   d) What is the probability that the lifetime of a battery is between 14.5 and 17 hours?

6. The temperature recorded by a certain thermometer when placed in boilingwater (true temperature 100◦C) is normally distributed with mean $\mu = 99.8$◦C and standard deviation 0.1◦C.

   a) What is the probability that the thermometer reading is greater than 100◦C?
   b) What is the probability that the thermometer reading is within ±0.05◦C of the true temperature?

7. Scores on a standardized test are approximately normally distributed with a mean of 480 and a standard deviation of 90.

   a) What proportion of the scores are above 700?
   b) What is the 25th percentile of the scores?
   c) If someone's score is 600, what percentile is she on?
   d) What proportion of the scores are between 420 and 520?

8. Weights of female cats of a certain breed are normally distributed with mean 4.1 kg and standard deviation 0.6 kg.

   a) What proportion of female cats have weights between 3.7 and 4.4 kg?
   b) A certain female cat has a weight that is 0.5 standard deviations above the mean. What proportion of female cats are heavier than this one?
   c) How heavy is a female cat whose weight is on the 80th percentile?
   d) A female cat is chosen at random. What is the probability that she weighs more than 4.5 kg?
   e) Six female cats are chosen at random. What is the probability that exactly one of them weighs more than 4.5 kg?

9. The lifetime of a lightbulb in a certain application is normally distributed with mean μ = 1400 hours and standard deviation σ = 200 hours.

    a) What is the probability that a lightbulb will last more than 1800 hours?
    b) Find the 10th percentile of the lifetimes.
    c) A particular lightbulb lasts 1645 hours. What percentile is its lifetime on?
    d) What is the probability that the lifetime of a lightbulb is between 1350 and 1550 hours?

10. In a certain university, math SAT scores for the entering freshman class averaged 650 and had a standard deviation of 100. The maximum possible score is 800. Is it possible that the scores of these freshmen are normally distributed? Explain.

11. Penicillin is produced by the Penicillium fungus, which is grown in a broth whose sugar content must be carefully controlled. The optimum sugar concentration is 4.9 mg/mL. If the concentration exceeds 6.0 mg/mL, the fungus dies and the process must be shut down for the day.

    a) If sugar concentration in batches of broth is normally distributed with mean 4.9 mg/mL and standard deviation 0.6 mg/mL, on what proportion of days will the process shut down?
    b) The supplier offers to sell broth with a sugar content that is normally distributed with mean 5.2 mg/mL and standard deviation 0.4 mg/mL. Will this broth result in fewer days of production lost? Explain.

12. Specifications for an aircraft bolt require that the ultimate tensile strength be at least 18 kN. It is known that 10% of the bolts have strengths less than 18.3 kN and that 5% of the bolts have strengths greater than 19.76 kN. It is also known that the strengths of these bolts are normally distributed.

    a) Find the mean and standard deviation of the strengths.
    b) What proportion of the bolts meet the strength specification?

13. A cylindrical hole is drilled in a block, and a cylindrical piston is placed in the hole. The clearance is equal to one-half the difference between the diameters of the hole and the piston. The diameter of the hole is normally distributed with mean 15 cm and standard deviation 0.025 cm, and the diameter of the piston is normally distributed with mean 14.88 cm and standard deviation 0.015 cm.

a) Find the mean clearance.
b) Find the standard deviation of the clearance.
c) What is the probability that the clearance is less than 0.05 cm?
d) Find the 25th percentile of the clearance.
e) Specifications call for the clearance to be between 0.05 and 0.09 cm. What is the probability that the clearance meets the specification?
f) It is possible to adjust the mean hole diameter. To what value should it be adjusted so as to maximize the probability that the clearance will be between 0.05 and 0.09 cm?

14. Shafts manufactured for use in optical storage devices have diameters that are normally distributed with mean μ = 0.652 cm and standard deviation σ = 0.003 cm. The specification for the shaft diameter is 0.650 ± 0.005 cm.

a) What proportion of the shafts manufactured by this process meet the specifications?
b) The process mean can be adjusted through calibration. If the mean is set to 0.650 cm, what proportion of the shafts will meet specifications?
c) If the mean is set to 0.650 cm, what must the standard deviation be so that 99% of the shafts will meet specifications?

15. The fill volume of cans filled by a certain machine is normally distributed with mean 12.05 oz and standard deviation 0.03 oz.

a) What proportion of cans contain less than 12 oz?
b) The process mean can be adjusted through calibration. To what value should the mean be set so that 99% of the cans will contain 12 oz or more?
c) If the process mean remains at 12.05 oz, what must the standard deviation be so that 99% of the cans will contain 12 oz or more?

16. The amount of paint required to paint a surface with an area of 50 m2 is normally distributed with mean 6 L and standard deviation 0.3 L.

a) If 6.2 L of paint are available, what is the probability that the entire surface can be painted?
b) How much paint is needed so that the probability is 0.9 that the entire surface can be painted?
c) What must the standard deviation be so that the probability is 0.9 that 6.2 L of paint will be sufficient to paint the entire surface?

17. A fiber-spinning process currently produces a fiber whose strength is normally distributed with a mean of 75 N/m^2. The minimum acceptable strength is 65 N/m^2.

    a) Ten percent of the fiber produced by the current method fails to meet the minimum specification. What is the standard deviation of fiber strengths in the current process?
    b) If the mean remains at 75 N/m^2, what must the standard deviation be so that only 1% of the fiber will fail to meet the specification?
    c) If the standard deviation is 5 N/m^2, to what value must the mean be set so that only 1% of the fiber will fail to meet the specification?

18. The area covered by 1 L of a certain stain is normally distributed with mean $10m^2$ and standard deviation $0.2m^2$.

    a) What is the probability that 1 L of stain will be enough to cover $10.3m^2$?
    b) What is the probability that 2 L of stain will be enough to cover $19.9m^2$?

19. Let $X \sim N(\mu, \sigma^2)$, and let $Z = (X - \mu)/\sigma$. Use Equation (4.25) to show that $Z \sim N(0, 1)$.

20. The quality-assurance program for a certain adhesive formulation process involves measuring how well the adhesive sticks a piece of plastic to a glass surface. When the process is functioning correctly, the adhesive strength X is normally distributed with a mean of 200 N and a standard deviation of 10 N. Each hour, you make one measurement of the adhesive strength. You are supposed to inform your supervisor if your measurement indicates that the process has strayed from its target distribution.

    a) Find $P(X \leq 160)$, under the assumption that the process is functioning correctly.
    b) Based on your answer to part (a), if the process is functioning correctly, would a strength of 160 N be unusually small? Explain.
    c) If you observed an adhesive strength of 160 N, would this be convincing evidence that the process was no longer functioning correctly? Explain.
    d) Find $P(X \geq 203)$, under the assumption that the process is functioning correctly.
    e) Based on your answer to part (d), if the process is functioning correctly, would a strength of 203 N be unusually large? Explain.
    f) If you observed an adhesive strength of 203 N, would this be convincing evidence that the process was no longer functioning correctly? Explain.
    g) Find $P(X \leq 195)$, under the assumption that the process is functioning correctly.

h) Based on your answer to part (g), if the process is functioning correctly, would a strength of 195 N be unusually small? Explain.
i) If you observed an adhesive strength of 195 N, would this be convincing evidence that the process was no longer functioning correctly? Explain.

21. Two resistors, with resistances R_1 and R_2, are connected in series. R_1 is normally distributed with mean 100Ω and standard deviation 5Ω, and R_2 is normally distributed with mean 120Ω and standard deviation 10Ω.

a) What is the probability that $R_2 > R_1$?
b) What is the probability that R_2 exceeds R_1 by more than 30Ω?

22. The molarity of a solute in solution is defined to be the number of moles of solute per liter of solution (1 mole = 6.02×10^{23} molecules). If X is the molarity of a solution of sodium chloride (NaCl), and Y is the molarity of a solution of sodium carbonate (Na_2CO_3), the molarity of sodium ion (Na^+) in a solution made of equal parts NaCl and Na_2CO_3 is given by M = 0.5X + Y . Assume X and Y are independent and normally distributed, and that X has mean 0.450 and standard deviation 0.050, and Y has mean 0.250 and standard deviation 0.025.

a) What is the distribution of M?
b) Find P(M > 0.5).

23. A binary message m, where m is equal either to 0 or to 1, is sent over an information channel. Because of noise in the channel, the message received is X, where X = m + E, and E is a random variable representing the channel noise. Assume that if X ≤ 0.5 then the receiver concludes that m = 0 and that if X > 0.5 then the receiver concludes that m = 1. Assume that E ~ N (0, 0.25).

a) If the true message is $m = 0$, what is the probability of an error, that is, what is the probability that the receiver concludes that $m = 1$?
b) Let σ^2 denote the variance of E. What must be the value of σ^2 so that the probability of error when $m = 0$ is 0.01?

24. Refer to Exercise 23. Assume that if m = 0, the value s = −1.5 is sent, and if m = 1, the value s = 1.5 is sent. The value received is X, where X = s + E, and E ~ N (0, 0.64). If X ≤ 0.5, then the receiver concludes that m = 0, and if X > 0.5, then the receiver concludes that m = 1.

a) If the true message is m = 0, what is the probability of an error, that is, what is the probability that the receiver concludes that m = 1?
b) If the true message is m = 1, what is the probability of an error, that is, what is the probability that the receiver concludes that m = 0?
c) A string consisting of 60 1s and 40 0s will be sent. A bit is chosen at random from this string. What is the probability that it will be received correctly?
d) Refer to part (c). A bit is chosen at random from the received string. Given that this bit is 1, what is the probability that the bit sent was 0?
e) Refer to part (c). A bit is chosen at random from the received string. Given that this bit is 0, what is the probability that the bit sent was 1?

25. A company receives a large shipment of bolts. The bolts will be used in an application that requires a torque of 100 J. Before the shipment is accepted, a quality engineer will sample 12 bolts and measure the torque needed to break each of them. The shipment will be accepted if the engineer concludes that fewer than 1% of the bolts in the shipment have a breaking torque of less than 100 J.

a) If the 12 values are 107, 109, 111, 113, 113, 114, 114, 115, 117, 119, 122, 124, compute the sample mean and sample standard deviation.
b) Assume the 12 values are sampled from a normal population, and assume the the sample mean and standard deviation calculated in part (a) are actually the population mean and standard deviation. Compute the proportion of bolts whose breaking torque is less than 100 J. Will the shipment be accepted?
c) What if the 12 values had been 108, 110, 112, 114, 114, 115, 115, 116, 118, 120, 123, 140? Use the method outlined in parts (a) and (b) to determine whether the shipment would have been accepted.
d) Compare the sets of 12 values in parts (a) and (c). In which sample are the bolts stronger?
e) Is the method valid for both samples? Why or why not?

26. Chebyshev's inequality (Section 2.4) states that for any random variable X with mean μ and variance σ^2, and for any positive number k, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$. Let $X \sim N(\mu, \sigma^2)$. Compute $P(|X - \mu| \geq k\sigma)$ for the values k = 1, 2, and 3. Are the actual probabilities close to the Chebyshev bound of $1/k^2$, or are they much smaller?