Dept of AIML, RVCE Bangalore

Statistics for Data Science

Module 2: Sampling, Sampling Distributions and Normal Distribution

1. We know that the PDF of a normally distributed random variable with mean μ and variance σ_2 is given by
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty \leq x \leq \infty$$
Why do we do standardization to get a standard normal variable?

2. Use standard normal table to evaluate the following probabilities
A. $z \leq 1.26$ B. $z \leq -0.58$ C. -1.45 $\leq z \leq 0.42$ D. $z > 0.68$

3. Find the area under the standard normal curve that lies outside the interval between

   (a) Outside the interval between $z = -1.11$ and $z = 3.21$
   (b) Outside the interval $z = 0.46$ and $z = 1.75$
   (c) Outside the interval $z = -2.73$ and $z = -1.39$

4. Find the z-score for which the area
   (a) to its left is 0.54
   (b) to its left is 0.93
   (c) to its left is 0.14
   (d) to its right is 0.35
   (e) to its right is 0.92
   (f) to its right is 0.14
   (g) that bounds the middle 70% of the area under the standard normal curve
   (h) that bounds the middle 80% of the area under the standard normal curve.

5. Mensa is an organization whose membership is limited to people whose IQ is in the top 2% of the population. Assume that scores on an IQ test are normally distributed with mean μ = 100 and standard deviation σ = 15. What is the minimum score needed to qualify for membership in Mensa?

6. Which normal distribution has a wider spread: the one with mean 1 and standard deviation 2 or the one with mean 2 and standard deviation 1? Explain your answer.

7. The daily sales total (excepting Saturday) at a small restaurant has a probability distribution that is approximately normal, with a mean m equal to $1230 per day and a standard deviation s equal to $120.

(a) What is the probability that the sales will exceed \$1400 for a given day?

(b) The restaurant must have at least \$1000 in sales per day to break even. What is the probability that on a given day the restaurant will not break even?

8. State Central Limit Theorem.

9. The duration of Alzheimerâs disease from the onset of symptoms until death ranges from 3 to 20 years; the average is 8 years with a standard deviation of 4 years. The administrator of a large medical center randomly selects the medical records of 30 deceased Alzheimerâs patients from the medical centerâs database, and records the average duration. Find the approximate probabilities for these events:

   (a) The average duration is less than 7 years.

   (b) The average duration exceeds 7 years.

   (c) The average duration lies within 1 year of the population mean $\mu = 8$.

10. A certain type of automobile battery is known to last an average of 1110 days with a standard deviation of 80 days. If 400 of these batteries are selected, find the following probabilities for the average length of life of the selected batteries:

    (a) The average is between 1100 and 1110.

    (b) The average is greater than 1120.

    (c) The average is less than 900.

11. A normal random variable X has an unknown mean and standard deviation $\sigma = 2$. If the probability that X exceeds 7.5 is 0.8023, find μ.

12. Most graduate schools of business require applicants for admission to take the Graduate Management Admission Councilâs GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 527 and a standard deviation of 112.

    (a) What is the probability of an individual scoring above 500 on the GMAT?

    (b) How high must an individual on the GMAT in order to score in the highest 5%?

13. The weight of male reindeer of the R. Santaclausus subspecies is normally distributed with mean 102.4 kg and standard deviation 13.9 kg.

    (a) What proportion of these reindeer would weigh more than 118.0 kg?

    (b) Rudolph is a fairly small reindeer. If exactly 10% of the reindeer weigh less than Rudolph, how much does Rudolph weigh?

    (c) If 36 reindeer are randomly selected and their average weight calculated, what is the probability that the mean weight is less than 100.0 kg?

14. The time it takes for high school boys to run 1 mile is normally distributed with mean 460 seconds and standard deviation 40 seconds. All those falling in the slowest 20 percent are deemed to need additional training. What is the critical time above which one is deemed to need additional training?

15. The bolts produced by a manufacturer are specified to be between 1.09 and 1.11 inches in diameter. If the production process results in the diameter of bolts being a normal random variable with mean 1.10 inches and standard deviation 0.005 inch, what percentage of bolts do not meet the specifications?