Tutorial Worksheet 2 - Probability Distributions (with solutions)

Objective Questions

1. If the values taken by a random variable are negative, the negative values will have

   (a) Positive probability (b) Negative probability (c) May have negative or positive probabilities (d) Insufficient data

   Solution: Positive probability

2. The expected value of a random variable is its

   (a) Mean (b) Standard Deviation (c) Mean Deviation (d) Variance

   Solution: Mean

3. In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean and variance is given by

   (a) $np, np(1-p)$ (b) $np(1-p), np$ (c) n, np^2 (d) p, np

   Solution: $np, np(1-p)$

4. If μ and σ denote the mean and standard deviation of a population, then the standard normal distribution is better described as (select the correct option from the list of options given below):

   (a) $f(x;\mu,\sigma) = \begin{cases} \frac{1}{\mu-\sigma}, & \mu \le x \le \sigma \\ 0, & \text{Otherwise} \end{cases}$ (b) $f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$

   (c) $f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$ (d) $f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, & x \ge 0 x < 0 \\ 0, & \text{otherwise} \end{cases}$

   Solution: $f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$

5. If $f(x)$ is a probability density function of any continuous random variable, then which of the follow statement(s) is (are) correct?

   (a) $0 \le f(x) \le 1$ (b) $P(a \le X \le b) = \int_a^b f(x)dx < 1$ (c) $y = \int_{-\alpha}^{\alpha} x f(x)dx$ there exist $y \in R$ (d) $z = \int_{-\alpha}^{\alpha} (x-\mu)^2 f(x)dx$ there exist $\mu \in R$ and $z \in R$

   Solution: $0 \le f(x) \le 1$ and $P(a \le X \le b) = \int_a^b f(x)dx < 1$

Subjective Questions

Problem 1. A quiz test for a course Data Analytics was conducted for a total score of 100 where 600 students took the test. From the result of the test it was found that mean score $\mu = 90$ and standard deviation $\sigma = 20$. Students are randomly distributed among six sections and each section includes 100 students.

(a) What is the standard error rate?

(b) If you select any section at random, what is the probability of getting a mean score is 86 or lower?

Solution:

(a) As per the Central Limit Theorem, the standard error is $\varepsilon = \frac{\sigma}{\sqrt{n}} = \frac{20}{10} = 2.0$.

(b) The sample distribution statistics can be obtained with the z-distribution. For the sample, The probability of getting 86 or lower is $P(Z < -2.0)$. From the standard normal distribution table it is found that P(Z $< -2.0) = 0.0228$.

Problem 2. A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

Solution:

(a) Denote by X the number of defective devices among the 20. Then X follows a $b(x; 20, 0.03)$ distribution. Hence, $P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20, 0.03) = 1 - (0.03)^0(1 - 0.03)^{20-0} = 0.4562$.

(b) In this case, each shipment can either contain at least one defective item or not. Hence, testing of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item, Y follows another binomial distribution $b(y; 10, 0.4562)$. Therefore,

$$P(Y = 3) = \left(\begin{array}{c} 10 \\ 3 \end{array} \right) 0.4562^3(1 - 0.4562)^7 = 0.1602$$

Problem 3. A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

(a) What fraction of the cups will contain more than 224 milliliters?

(b) What is the probability that a cup contains between 191 and 209 milliliters?

(c) How many cups will probably overflow if 230- milliliter cups are used for the next 1000 drinks?

Solution:

(a) $z = \frac{224-200}{15} = 1.6$ Fraction of the cups containing more than 224 millimeters is $P(Z > 1.6) = 0.0548$

(b) $z_1 = \frac{191-200}{15} = -0.6, z_2 = \frac{209-200}{15} = 0.6;$

$$P(191 < X < 209) = P(-0.6 < Z < 0.6) = 0.7257 - 0.2743 = 0.4514$$

(c) $z = \frac{230-200}{15} = 2.0; P(X > 230) = P(Z > 2.0) = 0.0228$. Therefore, $(1000)(0.0228) = 22.8$ or approximately 23 cups will overflow.

Problem 4. The Los Angeles Times (December 13, 1992) reported that what airline passengers like to do most on long flights is rest or sleep; in a survey of 3697 passengers, almost 80% did so. Suppose that for a particular route the actual percentage is exactly 80%, and consider randomly selecting six passengers.

(a) Calculate $p(4)$, and interpret this probability.

(b) Calculate $p(6)$, the probability that all six selected passengers rested or slept.

(c) Determine $P(X \geq 4)$.

Solution: Suppose that the actual percentage of passengers who rested or slept for a particular route is exactly 80%. Let x denote the number of six randomly selected passengers who rested or slept, so x is a binomial random variable with $n = 6$ and $p = 0.8$. Then the probability mass function of x is given by:

$$\begin{aligned} p(x) = P(x \text{ successes among } n \text{ trials}) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n \\ &= \frac{6!}{x!(6-x)!} (0.8)^x (1-0.8)^{6-x}, \quad x = 0, 1, 2, \ldots, 6 \end{aligned}$$

(a) Now find $p(4)$ using the expression of probability mass function given above.

$$\begin{aligned} p(4) &= \frac{6!}{4!(6-4)!} \times (0.8)^4 \times (1-0.8)^{6-4} = \left(\frac{720}{24 \times 2}\right) \times (0.8)^4 \times (0.2)^2 \\ &= (15) \times (0.4096) \times (0.04) = 0.2458 \end{aligned}$$

If a group of 6 passengers is examined, the long-run percentage of exactly four of those passengers rested or slept will be 24.58%.

(b) Now find the probability that all six selected passengers rested or slept:

$$p(6) = \frac{6!}{6!(6-6)!} \times (0.8)^6 \times (1-0.8)^{6-6} = \left(\frac{720}{720 \times 1}\right) \cdot (0.8)^6 \cdot (0.2)^0 = 0.2621$$

If group after group of 6 passengers is examined, the long-run percentage of all six of those passengers rested or slept will be 26.21%.

(c) Now find $P(x \geq 4)$, the probability that at least 4 of six selected passengers rested or slept:

$$\begin{aligned} P(x \geq 4) &= p(4) + p(5) + p(6) \\ &= \frac{6!}{4!2!} \times (0.8)^4 \times (1-0.8)^2 + \frac{6!}{5!1!} \times (0.8)^5 \times (1-0.8) + \frac{6!}{6!0!} \times (0.8)^6 \times (1-0.8)^0 \\ &= 0.2458 + 0.3932 + 0.2621 = 0.9011 \end{aligned}$$

If group after group of 6 passengers is examined, the long-run percentage of at least four of those passengers rested or slept will be 90.11%.

Problem 5. Emissions of nitrogen oxides, which are major constituents of smog, can be modeled using a normal distribution. Let X denote the amount of this pollutant emitted by a randomly selected vehicle (in parts per billion). The distribution of X can be described by a normal distribution with mean of 1.6 and standard deviation of 0.4. Suppose that the EPA wants to offer some sort of incentive to get the worst polluters off the road. What emission levels constitute the worst 10% of the vehicles?

Solution:

Given that, mean $\mu = 1.6$; standard deviation $\sigma = 0.4$, using Standard Normal Table we have, $P(Z > z) = 0.10 \Rightarrow 1 - P(Z < z) = 0.10 \Rightarrow P(Z < z) = 0.90 \Rightarrow z = 1.282$.

Using z-score formula we have, $x = z \times \sigma + \mu \Rightarrow x = 1.282 \times 0.4 + 1.6 \Rightarrow x = 2.1128$. The worst 10% of vehicles are those with emission levels greater than 2.1128 parts per billion.

Problem 6. A symmetric die is thrown 600 times. How many times can you expect to get a six?

Solution:

Let S be the total number of successes (getting a six). Then,

$$E(S) = n \times p = 600 \times \frac{1}{6} = 100 \text{ and } V(S) = n \times p \times q = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{500}{6}$$

Problem 7. Let X be a continuous random variable with pdf $f(x) = \begin{cases} \frac{2}{3}x, & 1 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$ Compute the expectation and variance for the random variable X and find $P(\frac{10}{9} < X < 2)$.

Solution:

The mean and the variance are

$$E(X) = \int_1^2 x\frac{2}{3}xdx = \frac{14}{9}, \quad \text{Var}(X) = \int_1^2 x^2\frac{2}{3}xdx - E(X)^2 = 0.080247.$$

$P\left(\frac{10}{9} \leq X \leq 2\right) = \int_{\frac{10}{9}}^{2} \frac{2}{3}xdx = 0.92181.$

Problem 8. A student took two national aptitude tests. The national average and standard deviation were 475 and 100, respectively, for the first test and 30 and 8, respectively, for the second test. The student scored 625 on the first test and 45 on the second test. Use **z** scores to determine on which exam the student performed better relative to the other test takers.

Solution:

The formula for z score is $z = \frac{x-\text{mean}}{\text{standard deviation}}$.
For the first exam $z = \frac{625-475}{100} = 1.5$ and for the second exam $z = \frac{45-30}{8} = 1.875$

By looking at the probabilities in a z score table we have $\text{P}(x < z) = \begin{cases} 0.93319, & z = 1.5 \\ 0.9696, & z = 1.875. \end{cases}$

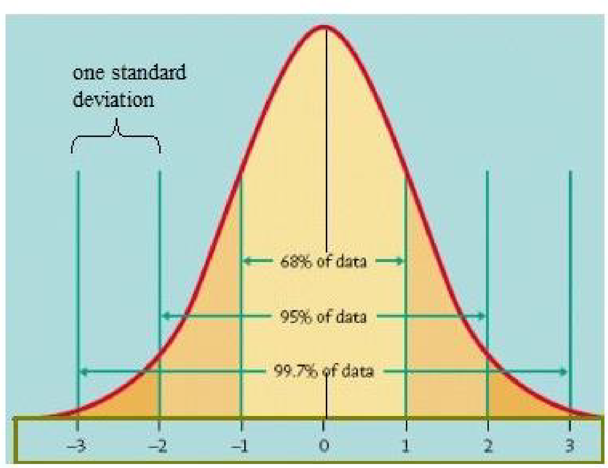
From the above probabilities we can say that the student performed better than 93.32% of others on the first test and better than 96.96% of others on the second.
Hence, the student performed better relative to others on the second exam.

Problem 9. A sample of concrete specimens of a certain type is selected, and the compressive strength of each specimen is determined. The mean and standard deviation are calculated as $\bar{x} = 3000$ and $s = 500$, and the sample histogram is found to be well approximated by a normal curve.

(a) Approximately what percentage of the sample observations are between 2500 and 3500?

(b) Approximately what percentage of sample observations are outside the interval from 2000 to 4000 ?

(c) What can be said about the approximate percentage of observations between 2000 and 2500 ?



Solution:

(a) By the empirical rule, we know that 68% lies within 1 standard deviation from the mean and thus between 2500 and 3500.

(b) By the empirical rule, we know that 95% lies within 2 standard deviations from the mean and thus between 2000 and 4000, thus 5% lies outside the given interval.

(c) Since the normal distribution is symmetric, we know that

$$\frac{95\% - 68\%}{2} = 13.5\%$$

has to lie between 2000 and 2500.

Problem 10. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Solution:

Given that, $\mu = 800, \sigma = 40, n = 16,$

$$\Rightarrow P(x < 775) = P\left(Z < \frac{775 - 800}{\frac{40}{\sqrt{16}}}\right)$$
$$\Rightarrow P(Z < -2.5) = 0.5 - P(0 < Z < 2.5) = 0.5 - 0.4938 = 0.0062.$$