PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-2 - Random Variables

QB SOLVED

Normal Distribution and Chebyshev's inequality

Exercises for Section 4.5

1. **If X ~ N(2, 9), compute**
   a) **$P(X \geq 2)$**
   b) **$P(1 \leq X < 7)$**
   c) **$P(-2.5 \leq X < -1)$**
   d) **$P(-3 \leq X - 2 < 3)$**

[Text Book Exercise – Section 4.5 – Q. No. 4 – Pg. No. 252]

Solution

From the given information of mean and variance,

$$\mu_X = 2 \qquad \sigma_X^2 = 9 \qquad \sigma = \sqrt{9} = 3$$

a) **$P(X \geq 2)$**

Compute z-score for X = 2 using the formula,

$$z = \frac{x - \mu}{\sigma} = \frac{2-2}{3} = 0$$

The area under the normal curve to the right of z = 0 is 0.5.

b) **$P(1 \leq X < 7)$**

Compute z-score for X = 1 and X = 7

$$z_{X=1} = \frac{1-2}{3} = -0.33$$

$$z_{X=7} = \frac{7-2}{3} = 1.67$$

The area between z = -0.33 is 0.3707 and z = 1.67 is 0.9525.

$P\ (1\ \leq X\ \leq 7) = P\ (-0.33\ \leq z\ \leq 1.67)$

$= 0.9525 - 0.3707$

$= 0.5818$

The area between z = -0.33 and z = 1.67 is = 0.5818

c) P(−2.5 ≤ X < −1)

Compute z-score for X = -2.5 and X = -1

$$z_{X=-2.5} = \frac{-2.5 - 2}{3} = -1.5$$

$$z_{X=-1} = \frac{-1 - 2}{3} = -1$$

The area between z = -1.5 is 0.0668 and z = -1 is 0.1587.

$P\ (-2.5\ \leq X < -1) = P\ (-1.5\ \leq z\ \leq -1)$

$= 0.1587 - 0.0668$

$= 0.0919$

The area between z = -1.5 and z = -1 is = 0.1587.

d) P(−3 ≤ X − 2 < 3)

Compute z-score for X - 2 = -3 and X - 2 = 3

Now, X – 2 = -3 => X = -1

$$z_{X=-1} = \frac{-1 - 2}{3} = -1$$

Now, X – 2 = 3 => X = 5

$$z_{X=5} = \frac{5 - 2}{3} = 1$$

The area between z = -1 is 0.1587 and z = -1 is 0.8413.

$P\,(-3\ \leq X - 2 < 3) = P\,(-1\ \leq z\ \leq 1)$

$= 0.8413 - 0.1587$

$= 0.6826$

The area between z = -1 and z = 1 is = 0.6826.

2. **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?**

[Text Book Exercise – Section 4.5 – Q. No. 8 – Pg. No. 253]

Solution

a) **What proportion of female cats have weights between 3.7 and 4.4 kg?**

Let X be the weight of the female cat. $X \sim N(4.1, 0.6)$

$\mu_X = 4.1$ $\qquad\qquad$ $\sigma = 0.6$

To compute cats having weights between 3.7 and 4.4.

$$z_{X=3.7} = \frac{3.7 - 4.1}{0.6} = -0.67$$

$$z_{X=4.4} = \frac{4.4 - 4.1}{0.6} = 0.5$$

The area between z = -0.67 in z-table is 0.2514 and z = 0.5 is 0.6915

To find, $P(3.7 < X < 4.4) = P(-0.67 < Z < 0.5)$

$$= 0.6915 - 0.2514 = 0.4401$$

The proportion of female cats weigh between 3.7 kg and 4.4 kg is 0.4401.

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?

Here, it is given that weight of certain female cats in 0.5 standard deviation above the mean.

$X > \mu + 0.5\sigma$

To find, the proportion of female cats are heavier than this, $P(X > \mu + 0.5\sigma)$

$P(X > \mu + 0.5\sigma) = P(X - \mu > 0.5\sigma)$

$$= P\left(\left(\frac{X-\mu}{\sigma}\right) > 0.5\right)$$

$$= P\ (Z > 0.5)$$

The area to the right of z = 0.50 is 0.6915 (1 – 0.6915 = 0.3085)

$$= 1 - P\ (Z < 0.5)$$

$$= 1 - 0.3915 = 0.3085$$

Therefore approximately 30.85% of cats are heavier than this one.

c) How heavy is a female cat whose weight is on the 80th percentile?

From z-table the area closest to 0.8 (80th percentile) is 0.7995 and the corresponding z-score is 0.84. Using the formula of z-score;

$$0.84 = \frac{X - 4.1}{0.6}$$

$X = 4.604$

d) A female cat is chosen at random. What is the probability that she weighs more than 4.5 kg?

To compute P(X > 4.5)

$$z_{X=4.5} = \frac{4.5 - 4.1}{0.6} = 0.67$$

From z-table, z = 0.67 is 0.7486

$P\ (X > 4.5) = 1 - \ P\ (Z < 0.67)$

$= 1 - \ 0.7486 = 0.2514$

e) Six female cats are chosen at random. What is the probability that exactly one of them weighs more than 4.5 kg?

Let X be the number of cats that weigh more than 4.5 kg. Using part (d), the probability that a cat weighs more than 4.5 kg is 0.2514.

Therefore $X \sim$ Bin (6, 0.2514).

$$P\ (X = 1) = \frac{6!}{1!\,(6-1)!}\ (0.2514)^1\ (1 - 0.2514)^{6-1}$$

$$= \ 6 * 0.2514 * 0.2350$$

$$= \ 0.3544$$

Therefore, the probability that exactly one of the chosen 6 cats weighs more than 0.3544.

3. 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 − μ| ≥ kσ) ≤ $1/k^2$. Let X ~ N (μ, σ^2). Compute P (|X −μ| ≥ kσ) 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?

[Text Book Exercise – Section 4.5 – Q. No. 26 – Pg. No. 256]

<u>**Solution**</u>

Case: 1

Consider, k = 1

About 98% of population is in the interval, $\mu \pm \sigma$

Using Chebyshev's Inequality $P\left(|X - \mu_X| > k_{\sigma_X}\right) \leq \frac{1}{k^2}$

$P(|X - \mu_X| > \sigma) = 1 - 0.68$

$$= 0.32$$
$$\leq \frac{1}{k^2} = \frac{1}{1^2}$$
$$= 1$$

Case: 2

Consider, k = 2

About 95% of population is in the interval, $\mu \pm 2\sigma$

Using Chebyshev's Inequality $P\left(|X - \mu_X| > k_{\sigma_X}\right) \leq \frac{1}{k^2}$

$P(|X - \mu_X| > 2\sigma) = 1 - 0.95$

$$= 0.05$$
$$\leq \frac{1}{k^2} = \frac{1}{2^2} = \frac{1}{4}$$
$$= 0.25$$

Case: 3

Consider, k = 2

About 99.7% of population is in the interval, $\mu \pm 3\sigma$

Using Chebyshev's Inequality $P\left(|X - \mu_X| > k_{\sigma_X}\right) \leq \frac{1}{k^2}$

$P(|X - \mu_X| > 2\sigma) = 1 - 0.997$

$$= 0.003$$
$$\leq \frac{1}{k^2} = \frac{1}{3^2} = \frac{1}{9}$$
$$= 0.111$$

k	$P\left(\lvert X - \mu_X\rvert > k_{\sigma_X}\right)$	$\frac{1}{k^2}$
1	0.32	1
2	0.05	0.25
3	0.003	0.111

The actual probabilities are much smaller than the Chebyshev bounds.