PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – DATA SCIENCE

Unit-2 - Question Bank

Exercises for Section 2.4

1. **Computer chips often contain surface imperfections. For a certain type of computer chip, the probability mass function of the number of defects *X* is presented in the following table.**

x	0	1	2	3	4
$p(x)$	0.4	0.3	0.15	0.10	0.05

a) **Find P(X ≤ 2).**
b) **Find P(X > 1).**
c) **Find μ_X**
d) **Find σ_X^2**

[Text Book Exercise – Section 2.4 – Q. No.2 – Pg. No. 112]

Solution:

a) **Compute P(X ≤ 2).**
Sum the probabilities of the values of X that are less than or equal to 2, namely 0, 1, 2.

$= \mathrm{P}\,(\mathrm{X} \leq 2) = \mathrm{P}(\mathrm{X} = 0) + \mathrm{P}\,(\mathrm{X} = 1) + \mathrm{P}\,(\mathrm{X} = 2)$
$= 0.4 + 0.3 + 0.15 = 0.85$

b) **Compute P(X > 1).**
Sum the probabilities of the values of X that are greater than 1, namely 2, 3, 4.

$= \mathrm{P}\,(\mathrm{X} > 1) = \mathrm{P}(\mathrm{X} = 2) + \mathrm{P}\,(\mathrm{X} = 3) + \mathrm{P}\,(\mathrm{X} = 4)$
$= 0.15 + 0.10 + 0.05 = 0.3$

c) **Compute $\boldsymbol{\mu_X}$**

The formula to compute mean of X,
$\mu_X = \sum_x x\, P(X = x)$

$\mu_X = 0(0.4) + 1(0.3) + 2(0.15) + 3(0.1) + 4(0.05)$

$\mu_X = 1.1$

d) **Compute $\boldsymbol{\sigma_X^2}$.**

The formula to compute variance of X,

$\sigma_X^2 = \sum_x x^2\, P(X = x) - {\mu_X}^2$

$\sigma_X^2 = 0^2(0.4) + 1^2\,(0.3) + 2^2(0.15) + 3^2(0.1) + 4^2(0.05) - (1.1)^2$

$\sigma_X^2 = 1.39$

2. **Elongation (in percent) of steel plates treated with aluminum are random with probability density function**

$$\boldsymbol{f(x)} = \begin{cases} \boldsymbol{\dfrac{x}{250}} & 20 < x < 30 \\ \boldsymbol{0} & otherwise \end{cases}$$

a) **What proportion of steel plates has elongations greater than 25%?**
b) **Find the mean elongation.**
c) **Find the variance of the elongations.**
d) **Find the standard deviation of the elongations.**
e) **Find the cumulative distribution function of the elongations.**
f) **A particular plate elongates 28%. What proportion of plates elongate more than this?**

[Text Book Exercise – Section 2.4 – Q. No.14 – Pg. No. 114]

<u>Solution:</u>

a) **What proportion of steel plates has elongations greater than 25%?**

Compute P (X > 25)

$P\,(X > 25) = \int_{25}^{\infty} f(x)\ dx$

$$= \int_{25}^{30} 250\, dx + \int_{30}^{\infty} 0\ dx$$

$$= \frac{1}{250}\left(\frac{x^2}{2}\right) \mid \begin{matrix} 30 \\ 25 \end{matrix}$$

$$= 0.55$$

b) Compute mean elongation.

The formula to compute mean elongation is,

$$\mu_X = \int_{-\infty}^{\infty} x\, f(x) dx$$

$$= \int_{-\infty}^{20} 0\, dx + \int_{20}^{30} x\left(\frac{x}{250}\right) dx + \int_{30}^{\infty} 0\, dx$$

$$= 0 + \frac{1}{250}\left(\frac{x^3}{3}\right) \mid \begin{matrix} 30 \\ 20 \end{matrix} + 0$$

$$= \left(\frac{x^3}{750}\right)\Bigg| \begin{matrix} 30 \\ 20 \end{matrix}$$

$$= 25.33$$

c) Compute the variance of the elongations.

The formula to compute variance of the elongation is,

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2\ f(x) dx - \mu_X^2$$

$$= \int_{-\infty}^{20} 0x^2\, dx + \int_{20}^{30} x^2\left(\frac{1}{250}\right) dx + \int_{30}^{\infty} 0\, x^2 dx - (25.33)^2$$

$$= \frac{1}{250}\left(\frac{x^4}{4}\right) \mid \begin{matrix} 30 \\ 20 \end{matrix} - (25.33)^2$$

$$= 8.3911$$

d) Find the standard deviation of the elongations.

Compute standard deviation.

$$\sigma_X = \sqrt{\sigma_X^2}$$

$$= \sqrt{8.3911} = 2.9$$

e) Find the cumulative distribution function of the elongations.

The cumulative distribution function is given by,

$$F\,(x) = \int_{-\infty}^{x} f(x)\,dt$$

The computation of the cumulative distribution function involves the following cases.

(i) If X ≤ 20,

$$F\,(x) = \int_{-\infty}^{20} 0\,dt = 0$$

(ii) If 20 < x < 30,

$$F(x) = \int_{-\infty}^{20} 0\,dt + \int_{20}^{x} \frac{t}{250}\,dt$$

$$= 0 + \frac{1}{250}\left(\frac{t^2}{2}\right) \mid \begin{matrix} x \\ 20 \end{matrix}$$

$$= \frac{1}{500}\,(x^2 - 400)$$

(iii) If X ≥ 30,

$$F(x) = \int_{-\infty}^{20} 0\,dt + \int_{20}^{30} \frac{t}{250}\,dt + \int_{30}^{x} 0\ dt$$

$$= 0 + \frac{1}{250}\left(\frac{t^2}{2}\right) \mid \begin{matrix} x \\ 20 \end{matrix} + 0$$

$$= 1$$

f) A particular plate elongates 28%. What proportion of plates elongate more than this?

Compute P (X > 28)

$$P\ (X > 28) = \int_{25}^{30} \frac{x}{250}\, dx$$

$$= \frac{1}{250}\left(\frac{x^2}{2}\right) \Big|_{28}^{30}$$

$$= 0.232$$

Exercises for Section 2.5

1. **The oxygen equivalence number of a weld is a number that can be used to predict properties such as hardness, strength, and ductility. The article "Advances in Oxygen Equivalence Equations for Predicting the Properties of Titanium Welds" (D. Harwig, W. Ittiwattana, and H. Castner, The Welding Journal, 2001:126s–136s) presents several equations for computing the oxygen equivalence number of a weld. One equation, designed to predict the hardness of a weld, is X = O + 2N + (2/3) C, where X is the oxygen equivalence, and O, N, and C are the amounts of oxygen, nitrogen, and carbon, respectively, in weight percent, in the weld. Suppose that for welds of a certain type, μ_O= 0.1668, μ_N = 0.0255, μ_C = 0.0247, σ_O = 0.0340, σ_N = 0.0194, and σ_C = 0.0131.**

    a) **Find μ_X.**
    b) **Suppose the weight percents of O, N, and C are independent. Find σ_X.**

[Text Book Exercise – Section 2.5 – Q. No.14 – Pg. No. 126]

Solution

a) Find μ_X.

We will use the formula, $\mu_{C_1X_1+C_2X_2+\ldots\ldots+C_nX_n} = C_1\mu_{X_1} + C_2\mu_{X_2} + \ldots\ldots + C_n\mu_{X_n}$

Therefore the mean of the oxygen equivalence number of a weld is given by,

$$\mu_X = \mu_{O+2N+2C/3} = \mu_O + 2\mu_N + \left(\frac{2}{3}\right)\mu_C$$

$$= 0.1668 + (2)(0.0255) + \left(\frac{2}{3}\right)(0.0247)$$

$$= 0.1668 + (2)(0.0255) + \left(\frac{2}{3}\right)(0.0247)$$

$$= 0.2342$$

b) Suppose the weight percents of O, N, and C are independent. Find σ_X.

We will use the formula, $\sigma_{C_1X_1+C_2X_2+\ldots\ldots+C_nX_n} = C_1\sigma_{X_1} + C_2\sigma_{X_2} + \ldots\ldots + C_n\sigma_{X_n}$

Therefore the standard deviation of the oxygen equivalence number of a weld is given by,

$$\sigma_X = \sigma_{O+2N+2C/3} = \sqrt{\sigma_O^2 + (2)^2\sigma_N^2 + \left(\frac{2}{3}\right)\sigma_C^2}$$

$$\sqrt{(0.0340)^2 + 4(0.0194)^2 + \left(\frac{2}{3}\right)^2 (0.0131)^2} = 0.05232$$

2. The thickness X of a wooden shim (in mm) has probability density function

$$\boldsymbol{f(x)} = \begin{cases} \boldsymbol{\frac{3}{4} - \frac{3(x-5)^2}{4}} & 4 < x < 6 \\ \boldsymbol{0} & otherwise \end{cases}$$

a) **Find μ_X**
b) **Find σ_X^2**
c) **Let *Y* denote the thickness of a shim in inches (1 mm = 0.0394 inches). Find μ_Y and σ_Y^2**
d) **If three shims are selected independently and stacked one atop another, find the mean and variance of the total thickness.**

[Text Book Exercise – Section 2.5 – Q. No.16 – Pg. No. 126]

Solution

a) **Find μ_X**

The formula to compute mean is,

$$\mu_X = \int_{-\infty}^{\infty} x\, f(x) dx$$

$$= \int_{-\infty}^{4} 0\ dx + \int_{4}^{6} \left(\frac{3}{4} - \frac{3(x-5)^2}{4}\right) x\ dx + \int_{6}^{\infty} 0\ dx$$

$$= 0 + \int_{4}^{6} \left(\frac{3}{4} - \frac{3(x-5)^2}{4}\right) x\ dx + 0$$

$$= \int_{4}^{6} \left(-\frac{3}{4}x^3 + \frac{15}{2}x^2 - 18x\right) dx$$

$$= \left(-\frac{3}{16}x^4 + \frac{5}{2}x^3 - 9x^2\right) \mid \begin{matrix} 6 \\ 4 \end{matrix}$$

$$= 5$$

b) Find σ_X^2

The formula to compute variance is,

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2\ f(x)dx - \mu_X^2$$

$$= \int_{-\infty}^{4} 0x^2\ dx + \int_{4}^{6} x^2 \left(\frac{3}{4} - \frac{3(x-5)^2}{4}\right) dx + \int_{6}^{\infty} 0x^2\ dx - (5)^2$$

$$= \int_{4}^{6} \left(-\frac{3}{4}x^4 + \frac{15}{2}x^3 - 18x^2\right) dx - (5)^2$$

$$= \left(-\frac{3}{20}x^5 + \frac{15}{8}x^4 - 6x^3\right) \Big|_{4}^{6} - 25$$

$$= \frac{126}{5} - 25 = \frac{1}{5}$$

$$= 0.2$$

c) Let *Y* denote the thickness of a shim in inches (1 mm = 0.0394 inches). Find μ_Y and σ_Y^2

The thickness of a shim in inches is Y = 0.0394 inches.

$$\mu_Y = 0.0394\ \mu_X$$

$$= 0.0394 * 5$$

$$= 0.197$$

$$\sigma_Y^2 = (0.0394)\sigma_X^2$$

$$= 0.0394 * 0.2$$

$$= 0.00031$$

d) If three shims are selected independently and stacked one atop another, find the mean and variance of the total thickness.

Let X_1, X_2, X_3 be three thickness in millimeters.
Then $S = X_1 + X_2 + X_3$ is the total thickness.

Where, $\mu = 5$, $\sigma^2 = 0.2$

Then mean and variance of total thickness includes,

$$\mu_{X_1+X_2+X_3} = \mu_{X_1} + \mu_{X_2} + \mu_{X_3}$$

$$= 3\mu = (3)(5)$$

$$= 15$$

$$\sigma^2_{X_1+X_2+X_3} = \sigma^2_{X_1} + \sigma^2_{X_2} + \sigma^2_{X_3}$$

$$= 3\sigma^2 = (3)(0.2)$$

$$= 0.6$$

Exercises for Section 4.1

1. **Let X and Y be Bernoulli random variables. Let Z = X + Y.**

    a) **Show that if X and Y cannot both be equal to 1, then Z is a Bernoulli random variable.**
    b) **Show that if X and Y cannot both be equal to 1, then $p_Z = p_X + p_Y$.**
    c) **Show that if X and Y can both be equal to 1, then Z is not a Bernoulli random variable.**

[Text Book Exercise – Section 4.1 – Q. No.4 – Pg. No. 203]

Solution

a) **Show that if X and Y cannot both be equal to 1, then Z is a Bernoulli random variable.**

It is given that $X \sim Bernoulli\ (p_X)$.

$Y \sim Bernoulli\ (p_Y)$.

So, the possible values of Z are 0 and 1 (when X = 0 or 1 and Y = 0 or 1).

X and Y both cannot be equal to 1. Because the possible values of Z can be 0 or 1.

Therefore, Z is a Bernoulli random variable.

b) **Show that if X and Y cannot both be equal to 1, then $p_Z = p_X + p_Y$.**

If X and Y cannot be both equal to 1, then $p_Z = p_X + p_Y$

$P_Z = P(X = 1\ or\ Y = 1)$

$= P(X = 1) + P(\ Y = 1) - P(X = 1\ and\ Y = 1)$

$= P(X = 1) + P(\ Y = 1)$

$= P_X + P_Y$

c) Show that if X and Y can both be equal to 1, then Z is not a Bernoulli random variable.

If X and Y both equal to 1, then Z = 2.

So, the possible values of Z are 0, 1, 2.

$P(X = 1\ and\ Y = 1) \neq 0$ (Z = 2)

Z can take only 0 and 1. So, Z is not a Bernoulli random variable.

Exercises for Section 4.2

1. **Let X ~ Bin(9, 0.4). Find**

   a) **P(X > 6)**
   b) **P(X ≥ 2)**
   c) **P(2 ≤ X < 5)**
   d) **P(2 < X ≤ 5)**
   e) **P(X = 0)**
   f) **P(X = 7)**
   g) $\boldsymbol{\mu_X}$
   h) $\boldsymbol{\sigma_X^2}$

[Text Book Exercise – Section 4.2 – Q. No.2 – Pg. No. 212]

Solution

a) **P(X > 6)**

$X \sim Bin\ (9, 0.4)$, n = 9, p = 0.4.

To find **P(X > 6)**

The formula to be used,

$$P\ (X = x) = \begin{cases} \dfrac{n!}{x!\,(n-x)!}\ p^x\ (1-p)^{n-x} & x = 0,1,\dots,n \\ 0 & otherwise \end{cases}$$

$$P\ (X > 6) = P(X = 7) + P\ (X = 8) + P\ (X = 9)$$

$$= \frac{9!}{7!\,(9-7)!}\ (0.4)^7\ (1-0.4)^{9-7} + \frac{9!}{8!\,(9-8)!}\ (0.4)^8\ (1-0.4)^{9-8} + \frac{9!}{9!\,(9-9)!}\ (0.4)^9\ (1-0.4)^{9-9}$$

$$= 0.0212 + 0.0035 + 0.0003$$

$$= 0.0250$$

b) P(X ≥ 2)

$X \sim Bin\ (9, 0.4)$, n = 9, p = 0.4.

To find **P(X ≥ 2)**

The formula to be used,

$$P\ (X = x) = \begin{cases} \frac{n!}{x!\,(n-x)!}\ p^x\ (1-p)^{n-x} & x = 0,1, \dots, n \\ 0 & otherwise \end{cases}$$

$$P\ (X \geq 2) = 1 - P(X \leq 1)$$

$$= 1 - P(X = 0) - P(X = 1)$$

$$= 1 - \frac{9!}{0!\,(9-0)!}\ (0.4)^0\ (1-0.4)^{9-0} - \frac{9!}{1!\,(9-1)!}\ (0.4)^1\ (1-0.4)^{9-1}$$

$$= 1 - 0.0101 - 0.0605$$

$$= 0.9295$$

c) P(2 ≤ X < 5)

$$P\ (2 \leq X < 5) = P(X = 2) + P\ (X = 3) + P\ (X = 4)$$

$$= \frac{9!}{2!\,(9-2)!}\ (0.4)^2\ (1-0.4)^{9-2} + \frac{9!}{3!\,(9-3)!}\ (0.4)^3\ (1-0.4)^{9-3} + \frac{9!}{4!\,(9-4)!}\ (0.4)^4\ (1-0.4)^{9-4}$$

$$= 0.1612 + 0.2508 + 0.2508$$

$$= 0.6659$$

d) P(2 < X ≤ 5)

$$P\ (2 < X \leq 5) = P(X = 3) + P\ (X = 4) + P\ (X = 5)$$

$$= \frac{9!}{3!\,(9-3)!}\,(0.4)^3\,(1-0.4)^{9-3} + \frac{9!}{4!\,(9-4)!}\,(0.4)^4\,(1-0.4)^{9-4} + \frac{9!}{5!\,(9-5)!}\,(0.4)^5\,(1-0.4)^{9-5}$$

$$= 0.2508 + 0.2508 + 0.1672$$

$$= 0.6689$$

e) P(X = 0)

$$P\,(X = 0) = \frac{9!}{0!\,(9-0)!}\,(0.4)^0\,(1-0.4)^{9-0}$$

$$= 0.0101$$

f) P(X = 7)

$$P\,(X = 7) = \frac{9!}{7!\,(9-7)!}\,(0.4)^7\,(1-0.4)^{9-7}$$

$$= 0.0212$$

g) $\boldsymbol{\mu_X}$

The mean can be found by the formula,

$$\mu = np$$

$$\mu = (9)\,(0.4)$$

$$\mu = 3.6$$

h) $\boldsymbol{\sigma_X^2}$

The variance can be found by the formula,

$$\sigma_X^2 = np(1-p)$$

$$\mu = (9)(0.4)(1-0.4)$$

$$\sigma_X^2 = 2.16$$

2. **A quality engineer takes a random sample of 100 steel rods from a day's production, and finds that 92 of them meet specifications.**

    a) **Estimate the proportion of that day's production that meets specifications, and find the uncertainty in the estimate.**
    b) **Estimate the number of rods that must be sampled to reduce the uncertainty to 1%.**

[Text Book Exercise – Section 4.2 – Q. No.10 – Pg. No. 213]

Solution

Let X be the number of rods that meets the specifications.

Let p be the proportion of the day's production that meets specifications.

n = 100, The observed value of X = 92.

Then, X ~ Bin (100, p).

a) **Estimate the proportion of that day's production that meets specifications, and find the uncertainty in the estimate.**

$$Sample\ Proportion \quad \hat{p} = \frac{92}{100} = 0.92$$

To find the uncertainty, substitute $\hat{p}$ in the formula,

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.92(1-0.92)}{100}} = 0.27$$

b) **Estimate the number of rods that must be sampled to reduce the uncertainty to 1%.**

To find the value of n,

$$\sigma_{\hat{p}} = \sqrt{\frac{0.92(1-0.92)}{n}} = 0.01$$

$$= 0.92 * 0.08 = 0.0001 * n$$

$$n = \frac{0.92 * 0.08}{0.0001} = 736$$

n = 736

Exercises for Section 4.3

1. **The number of flaws in a given area of aluminum foil follows a Poisson distribution with a mean of 3 per m^2. Let X represent the number of flaws in a 1 m^2 sample of foil.**

   a) **P(X = 5)**
   b) **P(X = 0)**
   c) **P(X < 2)**
   d) **P(X > 1)**
   e) $\boldsymbol{\mu_X}$
   f) $\boldsymbol{\sigma_X}$

   [Text Book Exercise – Section 4.3 – Q. No. 2 – Pg. No. 227]

   Solution

   Based on the given data, X ~ Poisson (3)

   a) **P(X = 5)**

   Using the formula for Poisson probability function,

   $$P\ (X = x) = \ e^{-\lambda}\frac{\lambda^x}{x!}$$

   $$\mathrm{P}\ (\mathrm{X} = 5) = \ \mathrm{e}^{-3}\ \frac{3^5}{5!} = 0.1008$$

   b) **P(X = 0)**

   $$\mathrm{P}\ (\mathrm{X} = 0) = \ \mathrm{e}^{-3}\ \frac{3^0}{0!} = 0.0498$$

   c) **P(X < 2)**

   $$\mathrm{P}\ (\mathrm{X} < 2) = \ \mathrm{P}\ (\mathrm{X} = 0) + \ \mathrm{P}\ (\mathrm{X} = 1)$$

   $$= \ e^{-3}\ \frac{3^0}{0!} + \ e^{-3}\ \frac{3^1}{1!}$$

   $$= 0.0497 + 0.1494$$

   $$= 0.1991$$

d) P(X > 1)

$$P\ (X > 1) = 1 - P\ (X = 0) - P\ (X = 1)$$

$$= 1 - e^{-3}\frac{3^0}{0!} - e^{-3}\frac{3^1}{1!}$$

$$= 1 - 0.0497 - 0.1494$$

$$= 0.8009$$

e) $\boldsymbol{\mu_X}$

Based on the given data, X ~ Poisson (3)

$$\mu_X = 3$$

f) $\boldsymbol{\sigma_X}$

The standard deviation is given by,

$$\sigma_X = \sqrt{\lambda} = \sqrt{3} = 1.732$$

2. A chemist wishes to estimate the concentration of particles in a certain suspension. She withdraws 3mL of the suspension and counts 48 particles. Estimate the concentration in particles per mL and find the uncertainty in the estimate.

[Text Book Exercise – Section 4.3 – Q. No. 10 – Pg. No. 228]

<u>Solution</u>

Let X represents the number of particles observed in 3 ml.

λ denotes the mean particles per ml. It can be estimated as

$$\hat{\lambda} = \frac{X}{t}$$

Here X = 48, t= 3, the concentration in particles per mL,

$$\hat{\lambda} = \frac{X}{t} = \frac{48}{3} = 16$$

To find uncertainty,

$$\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda}{t}} = \frac{16}{3} = 2.3094$$

Exercises for Section 4.5

1. **If X $\sim$ N(2, 9), compute**
   a) **P(X ≥ 2)**
   b) **P(1 ≤ X < 7)**
   c) **P(−2.5 ≤ X < −1)**
   d) **P(−3 ≤ 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 ≥ 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 ≤ 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.

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

$$= 0.9525 - 0.3707$$

$$= 0.5818$$

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. $X \sim N(4.1, 0.6)$

$\mu_X = 4.1$ $\quad\quad\quad\quad\quad$ $\sigma = 0.6$

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

The actual probabilities are much smaller than the Chebyshev bounds.