Statistics Assignment 7.1

Problem Statement 1: [50 marks]

The marks awarded for an assignment set for a Year 8 class of 20 students were as follows: 6 7 5 7 7 8 7 6 9 7 4 10 6 8 8 9 5 6 4 8

Mean is given by: $\bar{X} = \frac{1}{20}\sum X_i = \frac{1}{20}(6+7+5+7+7+\cdots+8) = \frac{137}{20} = \mathbf{6.85}$

Median is given by: arrange the data in order of size and average the middle two number i.e.

4 4 5 5 6 6 6 6 7 7 7 7 7 8 8 8 8 9 9 10

$$median = \frac{1}{2}(7+7) = \mathbf{7}$$

Mode is the number which appears the most (highest frequency)

$$mode = \mathbf{7}$$

Standard Deviation is given by the square root of variance

Variance is given by

$$s^2 = \frac{1}{20-1}\sum(X_i - \bar{X})^2 = \frac{1}{19}\left[(6-6.85)^2 + \cdots + (8-6.85)^2\right] = \mathbf{2.660526}$$

Hence the standard deviation is given by:

$$s = \sqrt{s^2} = \sqrt{2.660526} = \mathbf{1.631112}$$

Problem Statement 2: [50 marks]

The number of calls from motorists per day for roadside service was recorded for a particular month: 28, 122, 217, 130, 120, 86, 80, 90, 140, 120, 70, 40, 145, 113, 90, 68, 174, 194, 170, 100, 75, 104, 97, 75, 123, 100, 75, 104, 97, 75, 123, 100, 89, 120, 109

Calculate the mean, median, mode and standard deviation for the problem statements 1 & 2

Median is given by the number in the middle of the ordered data set below

28 40 68 70 75 75 75 75 80 86 89 90 90 97 97 100 100 100 104 104 109 113 120 120 120 122 123 123 130 140 145 170 174 194 217

Hence the median is **100**

Mode is the number which appears the most (highest frequency)

28	40	68	70	75	80	86	89	90	97	100	104	109	113	120	122	123	130	140	145	170	174	194	217
1	1	1	1	4	1	1	1	2	2	3	2	1	1	3	1	2	1	1	1	1	1	1	1

$$mode = \mathbf{75}$$

Standard Deviation is given by the square root of Variance

Variance is given by

$$s^2 = \frac{1}{35-1}\sum(X_i - \bar{X})^2 = \frac{1}{34}\left[(28-107.51)^2 + \cdots + (109-107.51)^2\right] = \mathbf{1547.55}$$

Hence the standard deviation is given by:

$$s = \sqrt{s^2} = \sqrt{1547.55} = \mathbf{39.34}$$

Problem Statement 3: [100 marks]

The number of times I go to the gym in weekdays, are given below along with its associated probability:

x = 0, 1, 2, 3, 4, 5 and f(x) = 0.09, 0.15, 0.40, 0.25, 0.10, 0.01 Calculate the mean no. of workouts in a week. Also evaluate the variance involved in it.

Mean is given by:

$$E[X] = \sum X_i f(X) = 0(0.09) + 1(0.15) + 2(0.40) + 3(0.25) + 4(0.10) + 5(0.01)$$

$$= 0 + 0.15 + 0.80 + 0.75 + 0.40 + 0.05$$

$$= \mathbf{2.15}$$

Variance is given by:

$$Var[X] = E[X^2] - E^2[X]$$

$$= [0^2(0.09) + 1^2(0.15) + 2^2(0.40) + 3^2(0.25) + 4^2(0.10) + 5^2(0.01)] - 2.15^2$$

$$= [0(0.09) + 1(0.15) + 4(0.40) + 9(0.25) + 16(0.10) + 25(0.01)] - 4.6225$$

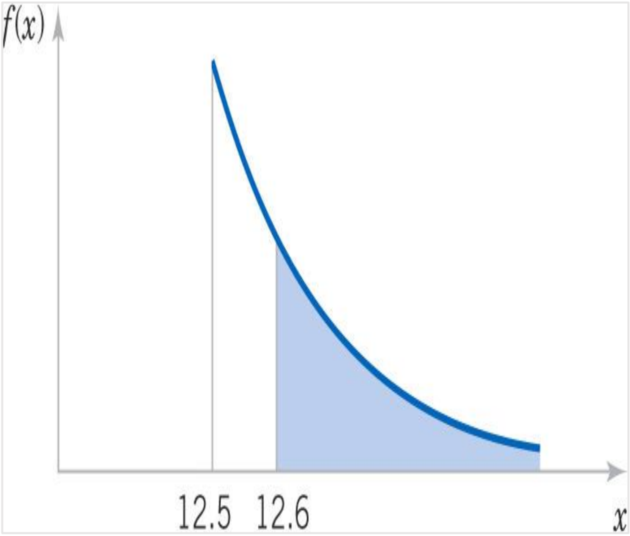
$$= [0 + 0.15 + 1.60 + 2.25 + 1.6 + 0.25] - 4.6225$$

$$= 5.85 - 4.6225 = \mathbf{1.2275}$$

Problem Statement 4: [100 marks]

Let the continuous random variable D denote the diameter of the hole drilled in an aluminium sheet. The target diameter to be achieved is 12.5mm. Random disturbances in the process often result in inaccuracy. Historical data shows that the distribution of D can be modelled by the PDF $f(d) = 20e^{-20(d-12.5)}, d \geq 12.5$. If a part with diameter > 12.6 mm needs to be scrapped, what is the proportion of those parts? What is the CDF when the diameter is of 11 mm? What is the conclusion of this experiment?

If the PDF is given by $f(d) = 20e^{-20(d-12.5)}, d \geq 12.5$, then if a part of diameter > 12.6 needs to be scrapped, it means we are looking for the area under the graph of $f(d)$



The proportion of parts greater than 12.6mm is given by:

$$P(D > 12.6) = \int_{12.6}^{\infty} 20e^{-20(D-12.5)}\,dD$$
$$= 20\int_{12.6}^{\infty} e^{-20(D-12.5)}\,dD$$
$$= \frac{20\left[e^{-20(D-12.5)}\right]}{-20}\Bigg|_{12.6}^{\infty}$$
$$= -\left[e^{-20(D-12.5)}\right]\Big|_{12.6}^{\infty} = -0 + e^{-2}$$
$$= \mathbf{0.135}$$

The proportion of parts between 12.5mm and 12.6mm is given by:

$$P(12.5 < D < 12.6) = 1 - 0.135 = \mathbf{0.865}$$

The CDF is given by

$$F(x) = 0, \qquad for\ x < 12.5$$

$$F(x) = \int_{12.6}^{x} 20e^{-20(D-12.5)}\,dD$$

$$= 1 - 20e^{-20(x-12.5)}, \quad for\ x \geq 12.5$$

Hence the CDF is given as $F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - 20e^{-20(x-12.5)}, & x \geq 12.5 \end{cases}$

When the diameter is 11mm, $F(x) = 0$