Name - Niranjan Nevase
Batch - 5th June Data Science

Topics: Normal distribution, Functions of Random Variables

1. The time required for servicing transmissions is normally distributed with μ = 45 minutes and σ = 8 minutes. The service manager plans to have work begin on the transmission of a customer's car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

    A. 0.3875
    B. 0.2676
    C. 0.5
    D. 0.6987

Ans- Given:
μ (mean) = 45 minutes
σ (standard deviation) = 8 minutes

Since work begins 10 minutes after the car is dropped off, we need to adjust the mean and standard deviation:

Adjusted mean = μ + 10 minutes = 45 minutes + 10 minutes = 55 minutes
Adjusted standard deviation remains the same = σ = 8 minutes
Probability that the service time exceeds 60 minutes. To do this, we'll convert this into a standard normal distribution (Z-score) using the z-score formula:

$Z = (X - \mu) / \sigma = (60 - 55) / 8 = 5 / 8 = 0.625$
$P(Z > 0.625) \approx 1 - P(Z < 0.625) \approx$ **0.268**

2. The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean μ = 38 and Standard deviation σ=6. For each statement below, please specify True/False. If false, briefly explain why.
    A. More employees at the processing center are older than 44 than between 38 and 44.
    False
    Probability that an employee is older than 44, Using Excel, we can use the NORM.DIST =NORM.DIST(44, 38, 6, TRUE) = 0.841
    Probability that an employee is older than 44 = 1 – 0.841 = 0.159
    This means that about **15.9%** of the employees are older than 44.
    Now, probability that an employee is between 38 and 44 :
    NORM.DIST(44, 38, 6, TRUE) - NORM.DIST(38, 38, 6, TRUE) = 0.341
    This means that about 34.1% of the employees are between 38 and 44.

Therefore, more employees are between 38 and 44 than older than 44.

B. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

True

Probability that an employee is under age of 30 , using Excel function
=NORM.DIST(30, 38, 6, TRUE) = 0.091 = 9.1%

To find the expected number of employees under the age of 30 who would attend the training program, we need to multiply this probability by 400:
0.091 * 400 = **36.4**
This means that about 36 employees under the age of 30 would be expected to attend the training program.

3. If $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim N(\mu, \sigma^2)$ are *iid* normal random variables, then what is the difference between 2 X_1 and $X_1 + X_2$? Discuss both their distributions and parameters.

- **Distribution of $2X_1$:**
  When we multiply a random variable by a constant (in this case, 2), it affects both the mean (μ) and the variance (σ^2) of the resulting random variable.
  For $2X_1$:
  - Mean (μ): Since we multiply X_1 by 2, the mean of $2X_1$ is 2μ.
  - Variance (σ^2): When we multiply a random variable by a constant, we square that constant in the variance. So, the variance of $2X_1$ is $4\sigma^2$.

Therefore, $2X_1$ follows a normal distribution with parameters:
Mean: 2μ & Variance: $4\sigma^2$

- **Distribution of $X_1 + X_2$:**
  When we add two independent random variables, the mean of the sum is the sum of their individual means, and the variance of the sum is the sum of their individual variances.
  For X1 + X2:
  - Mean (μ): The mean of the sum is μ + μ, which is 2μ.
  - Variance (σ^2): The variance of the sum is $\sigma^2 + \sigma^2$, which is $2\sigma^2$.

  Therefore, X1 + X2 follows a normal distribution with parameters:
  Mean: 2μ & Variance: $2\sigma^2$

  The difference between $2X_1$ and $X_1 + X_2$ lies in their variances. $2X_1$ has a larger variance ($4\sigma^2$) compared to X1 + X2 ($2\sigma^2$), while both have the same mean (2μ).

Name - Niranjan Nevase
Batch - 5th June Data Science

4. Let X ~ $N(100, 20^2)$. Find two values, *a* and *b*, symmetric about the mean, such that the probability of the random variable taking a value between them is 0.99.

    A. 90.5, 105.9
    B. 80.2, 119.8
    C. 22, 78
    D. 48.5, 151.5
    E. 90.1, 109.9

For X ~ $N(100, 20^2)$:

Mean (μ) = 100
Standard Deviation (σ) = 20
Now, we want to find the z-scores for the 0.005th and 0.995th percentiles of the standard normal distribution:

Using a standard normal distribution table or calculator, you can find that the z-score for the 0.005th percentile is approximately -2.576 and the z-score for the 0.995th percentile is approximately 2.576.

Now, we can find the corresponding values a and b for our original normal distribution:

a = μ + z1 * σ
a = 100 + (-2.576) * 20 ≈ 100 - 51.52 ≈ 48.48

b = μ + z2 * σ
b = 100 + 2.576 * 20 ≈ 100 + 51.52 ≈ 151.52

So, the values a and b, symmetric about the mean, such that the probability of X falling between them is approximately 0.99, are **a ≈ 48.48 and b ≈ 151.52.**

5. Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions $Profit_1$ ~ $N(5, 3^2)$ and $Profit_2$ ~ $N(7, 4^2)$ respectively. Both the profits are in $ Million. Answer the following questions about the total profit of the company in Rupees. Assume that $1 = Rs. 45
    A. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.

    - To convert from dollars to rupees, we need to multiply by 45. Therefore, in rupees, we have TotalProfit ~ $Profit_1 + Profit_2 = N(540, 1125)$.

Name - Niranjan Nevase
Batch - 5th June Data Science

To specify a rupee range that contains 95% probability for the annual profit of the company, we need to find the values that are 1.96 standard deviations away from the mean. This is because 95% of a normal distribution lies within 1.96 standard deviations from the mean. The standard deviation is the square root of the variance, so in this case it is sqrt(1125) = 33.54. Therefore, the rupee range is:

540 – (1.96 * 33.54) < Total Profit < 540 + (1.96 * 33.54)
474.28 < Total Profit < 605.72
This means that there is a 95% chance that the total profit of the company will be between Rs. 474.28 million and Rs. 605.72 million.

B. Specify the 5th percentile of profit (in Rupees) for the company

- Using Excel, we can use the NORM.INV function with the following syntax:

=NORM.INV(probability, mean, standard_dev)
where probability is the cumulative probability up to the value we want to find, mean is the mean of the normal distribution, and standard_dev is the standard deviation of the normal distribution.
In this case, we want to find the value that is below 5% of the distribution, so we enter:
=NORM.INV(0.05, 540, 33.54) = **484.83**
This means that there is a 5% chance that the total profit of the company will be below Rs. 484.83 million.

C. Which of the two divisions has a larger probability of making a loss in a given year?

- Using Excel, we can use the NORM.DIST function with the following syntax:

=NORM.DIST(x, mean, standard dev, cumulative)

where x is the profit value we want to find the probability for, mean is the mean of the normal distribution, standard dev is the standard deviation of the normal distribution, and cumulative is a logical value that indicates whether we want the cumulative distribution function (TRUE) or the probability density function (FALSE).
In this case, we want to find the cumulative probability up to zero for each division, so we enter:

For division 1: =NORM.DIST(0, 5, 3, TRUE) = 0.04779 = 4.78%
For division 2: =NORM.DIST(0, 7, 4, TRUE) = 0.04006 = 4.01%

Name - Niranjan Nevase
Batch - 5th June Data Science

This means that division 1 has a larger probability of making a loss in a given year than division 2 (4.78% vs 4.01%).