**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?
2. 0.3875
3. 0.2676
4. 0.5
5. 0.6987

**Answer:**

import scipy.stats as stats

#Given Data

pop\_mean = 45 #mean of the distribution

std\_dev = 8 # standard deviation

start\_time = 10 #the transmission of a customer’s car 10 minutes

time\_limit = 60 #the car will be ready within 1 hour(60 minutes) from drop-off

X = time\_limit - start\_time

X #random variable (time in this case)

O/p - 50

#finding z-score

Z = (X - pop\_mean) / std\_dev

Z

O/p - 0.625

#we need to find the probability that a randomly selected time exceeds 50 minutes

p\_value = 1 - stats.norm.cdf(Z)

p\_value

O/p - 0.26598552904870054

Therefore, the answer is 0.2676

1. 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.
2. More employees at the processing center are older than 44 than between 38 and 44.

**Answer:**

import scipy.stats as stats

#Given Data

pop\_mean = 38

std\_dev = 6

X\_38 = 38

X\_44 = 44

#Calculate Z-score for 38

Z\_38 = (X\_38 - pop\_mean) / std\_dev

Z\_38

O/p - 0.0

#Calculate Z-score for 44

Z\_44 = (X\_44 - pop\_mean) / std\_dev

Z\_44

O/p - 1.0

#Calculte probabilites

p\_value38 = stats.norm.cdf(Z\_38)

p\_value38

O/p - 0.5

p\_value44 = stats.norm.cdf(Z\_44)

p\_value44

O/p - 0.8413447460685429

#Calculate the difference in probabilites

diff\_pvalue = p\_value44 - p\_value38

diff\_pvalue

O/p - 0.3413447460685429

#Check if more employees are older than 44 than between 38 and 44

result = p\_value44 > p\_value38

result

O/p - True

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

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

**Answer:**

import scipy.stats as stats

#Given Data

pop\_mean = 38

std\_dev = 6

age = 30

#Calculate Z-score for 30

Z = (age - pop\_mean) / std\_dev

Z

O/p - -1.3333333333333333

#Calculate the probability of being under 30

p\_value = stats.norm.cdf(Z)

p\_value

O/p - 0.09121121972586788

#assuming a total 400 employees

total\_emp = 400

#calculate the expected count under the target age

excepted\_count = p\_value \* total\_emp

excepted\_count

O/p - 36.484487890347154

Therefore, the expected count of 36 employees under the age of 30 is likely to be too high

Hence, the statement is false

1. If *X1* ~ *N*(μ, σ2) and *X*2 ~ *N*(μ, σ2) 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.

**Answer:**

1. The Normal Distribution has its link with the Central Limit Theorem, which states that ‘Any large sum of independent identically distribution random variables are approximately Normal then (X1 + X2) and (2X1) tends to have Normal distribution only If X1 and X2 are i.i.d and n is Large.
2. The Difference between 2X1 and (X1 + X2) is the magnitude they hold of two different sample subsets (X1 and X2) from the same source (population).
3. X1 and X2 can be a different subset of a sample from a similar source (population) but If X1 ~ N(μ, σ2) then, 2 X1 ~ N(2 μ, 4 σ2 ) If X1 ~ N(μ, σ2) and X2 ~ N(μ, σ2) are iid normal random variables then (X1 + X2)(2 μ, 2 σ2)
4. Hence, 2X1 – (X1+X2) ~(2 μ – 2 μ, 4 σ2 + 2σ2 )
5. The distribution remains the same for every sample subset of similar source, it tends to fall under Normal distribution and slight deviations in parameters.
6. The Normal distribution has two parameters, the mean, µ, and the variance, σ2. µ And σ2satisfy −∞ < µ < ∞, σ2> 0. We write X ∼ Normal (µ, σ2) or X ~ N(µ, σ2 ).
7. Let X ~ N(100, 202). 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.
8. 90.5, 105.9
9. 80.2, 119.8
10. 22, 78
11. 48.5, 151.5
12. 90.1, 109.9

**Answer:**

import scipy.stats as stats

#Given Data

pop\_mean = 100

std\_dev = 20

#For a normal distribution with mean and standard deviation,about 99% of the values lie within approximately

#mean +\_ 2.57\*std\_dev

#finding the values of a & b

a = pop\_mean - 2.57 \* std\_dev

b = pop\_mean + 2.57 \* std\_dev

print('The two values of a and b, symmetric about the mean, are such that the probability of the random variable taking a value between them is 0.99:', a,'&',b)

O/p - The two values of a and b, symmetric about the mean, are such that the probability of the random variable taking a value between them is 0.99: 48.6 & 151.4

1. Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions Profit1 ~ N(5, 32) and Profit2 ~ N(7, 42) 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
2. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
3. Specify the 5th percentile of profit (in Rupees) for the company
4. Which of the two divisions has a larger probability of making a loss in a given year?

**Answer:**

import scipy.stats as stats

import math

import numpy as np

#Given Value:

#Profit1 as X and Profit2 as Y

mean\_x = 5

var\_x = 3\*\*2

mean\_y = 7

var\_y = 4\*\*2

#$1 = RS. 45

conv\_rate = 45

#A. To find the Rupee range containing 95% probability for the annual profit of the company, we can use the properties of the normal distribution.

mean\_z = mean\_x + mean\_y #finding total mean

mean\_z #in $

stddev\_z = math.sqrt(var\_x + var\_y)

stddev\_z #in $

#interval = 95% = 0.95

CI1, CI2 = np.round(stats.norm.interval(0.95,mean\_z,stddev\_z),2)

CI1, CI2

range\_rs = ((CI[0] \* conv\_rate) , (CI[1] \* conv\_rate))

range\_rs

print('Rupee Ranges from', range\_rs[0], 'to' , range\_rs[1], 'million rupees in Annual profit of the company 95% of the time.')

O/p -

Rupee Ranges from 99.00000000000001 to 981.0 million rupees in Annual profit of the company 95% of the time.

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

z\_score = stats.norm.ppf(0.05)

z\_score

mean\_z\_RS = mean\_z \* 45 # converting into RS

stddev\_z\_RS = stddev\_z \* 45

percentile5 = mean\_z\_RS + (z\_score \* stddev\_z\_RS)

np.round(percentile5,2)

print('The 5th percentile of Profit for the company is',percentile5, 'million rupees.' )

O/p -

The 5th percentile of Profit for the company is 169.9079339359186 million rupees.

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

# The probability of Division1 making a loss

div1 = stats.norm.cdf(0, 5,3)

div1\*100

# The probability of Division2 making a loss

div2 = stats.norm.cdf(0,7,4)

div2\*100

if div1>div2:

print('The Division 1 has a larger Probability of making a loss')

else:

print('The Division 2 has a larger Probability of making a loss')

O/p -

The Division 1 has a larger Probability of making a loss