

SURESH LOKU RATHOD

DATA SCIENCE BATCH 1ST FEBRUARY 2023

ASSIGNMENT NO 2-STATISTIC BASIC 2 (SET .2.)

Topics: Normal distribution, Functions of Random Variables

- 1. The time required for servicing transmissions is normally distributed with $\mu = 45$ minutes and $\sigma = 8$ 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



Solution:

Given: $\mu = 45$ $\sigma = 8$

$P(x > 50)$

Import numpy as np

from scipy import stats

1-stats.norm.cdf (50,loc=45,scale=8)

=0.26598

Probability that the service manager cannot meet his commitment is 0.26598

- 2. The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean $\mu = 38$ and Standard**

deviation $\sigma=6$. 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.

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

Solution:

Mean = 38

SD = 6

Z score = (Value - Mean)/SD

Z score for 44 = $(44 - 38)/6 = 1 \Rightarrow 84.13 \%$

People above 44 age = $100 - 84.13 = 15.87\% \approx 63$ out of 400

Z score for 38 = $(38 - 38)/6 = 0 \Rightarrow 50\%$

Hence People between 38 & 44 age = $84.13 - 50 = 34.13 \% \approx 137$ out of 400

Hence More employees at the processing center are older than 44 than between 38 and 44. is **FALSE**

Z score for 30 = $(30 - 38)/6 = -1.33 = 9.15 \% \approx 36$ out of 400

Hence A training program for employees under the age of 30 at the centre would be expected to attract about 36 employees - **TRUE**

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 $2X_1$ and $X_1 + X_2$? Discuss both their distributions and parameters.



Solution:

$$\begin{aligned}2X_1 &\sim N(2\mu, \sigma^2) \\ X_1 + X_2 &\sim N(\mu + \mu, \sigma^2 + \sigma^2) \\ &\sim N(2\mu, 2\sigma^2)\end{aligned}$$

For both distribution location parameter is same only scale parameter is different.

4. Let $X \sim N(100, 20^2)$. 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**



Solution:

Given: $p(a < x < b) = 0.99$

Mean = 100

SD = 20

For identifying the standard normal distribution such that area enclosed is 0.99 we have to exclude area of 0.005 in each of the left and right tails.

Hence we have to find z scores of 0.5th and 99.5th percentile

`stats.norm.ppf(0.005)`

-2.575829303548901

```
stats.norm.ppf(0.995)
```

```
2.5758293035489004
```

```
z= (x-100)/20
```

```
x=20z + 100
```

```
a=20*(-2.576) +100
```

```
=48.5
```

```
b=20*(2.576) +100
```

```
=151.5
```

D option is correct

5. Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions $\text{Profit}_1 \sim N(5, 3^2)$ and $\text{Profit}_2 \sim N(7, 4^2)$ 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.**
- B. Specify the 5th percentile of profit (in Rupees) for the company**
- C. Which of the two divisions has a larger probability of making a loss in a given year?**



Solution:

```
import pandas as pd
```

```
import numpy as np
```

```
from scipy import stats
```

```
from scipy.stats import norm
```

```
# Mean profits from two different divisions of a company = Mean1 + Mean2
```

```
Mean = 5+7
```

```
print('Mean Profit is Rs', Mean*45,'Million')
```

```
Mean Profit is Rs 540 Million
```

```
# Variance of profits from two different divisions of a company =  $SD^2 = SD1^2 + SD2^2$ 
```

```
SD = np.sqrt((9)+(16))
```

```
print('Standard Deviation is Rs', SD*45, 'Million')
```

```
Standard Deviation is Rs 170 Million
```

```
# A. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
```

```
print('Range is Rs',(stats.norm.interval(0.95,540,225)), 'in Millions')
```

```
Range is Rs (90,990) in Millions
```

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

```
# To compute 5th Percentile, we use the formula  $X = \mu + Z\sigma$ ; wherein from z table, 5 percentile = -1.645
```

```
X= 540+(-1.645)*(225)
```

```
print('5th percentile of profit (in Million Rupees) is',np.round(X,))
```

```
5th percentile of profit (in Million Rupees) is 202.05Millions
```

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

```
# Probability of Division 1 making a loss  $P(X < 0)$ 
```

```
stats.norm.cdf(0,5,3)
```

```
0.04779
```

```
# Probability of Division 2 making a loss  $P(X < 0)$ 
```

```
stats.norm.cdf(0,7,4)
```

```
0.04005
```