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 : option B.

We have a normal distribution with = 45 and = 8.0. Let X be the amount of time it takes to complete the repair on a customer's car. To finish in one hour you must have X ≤ 50 so the question is to find Pr(X > 50).
Pr(X > 50) = 1 - Pr(X ≤ 50).
Z = (X -)/ = (X - 45)/8.0
Thus the question can be answered by using the normal table to find
Pr(X ≤ 50) = Pr(Z ≤ (50 - 45)/8.0) = Pr(Z ≤ 0.625)=73.4%
Probability that the service manager will not meet his demand will be = 100-73.4 = 26.6% or 0.2676

The work begin after 10 min, so the average time increase from 45min to 55min.

for normal distribution :-

z = (X-µ)/б

= (60-55)/8

= 0.625

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.

Ans : False

Around 70% of the data falls within one standard deviation of the mean

(µ+s= 38+6=44)

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

Ans : True

Z=(X-µ)/ s

P(X<=30) = p(Z<=(30-38)/6) = p(Z<= - 1.33) = 0.0918 (using z table)

Expected count = 0.0918*400 = 36.72

```
In [41]: from scipy import stats
         from scipy.stats import norm

In [45]: #A.More employees at the processing center are older than 44 than between 38 and 44.

         # p(X>44); Employees older than 44 yrs of age
         1-stats.norm.cdf(44,loc=38,scale=6)

Out[45]: 0.15865525393145707

In [46]: # p(38<X<44); Employees between 38 to 44 yrs of age
         stats.norm.cdf(44,38,6)-stats.norm.cdf(38,38,6)

Out[46]: 0.3413447460685429
```

Statement A is False . As Probability of employees aged between 38 to 44 is more.

```
In [49]: #B.A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

         # P(X<30); Employees under 30 yrs of age
         stats.norm.cdf(30,38,6)

Out[49]: 0.09121121972586788

In [52]: # No. of employees attending training program from 400 nos. is N*P(X<30)
         400*stats.norm.cdf(30,38,6)

Out[52]: 36.484487890347154
```

Statement B is True as no. of employees aged below 33 yrs attending training is 36

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.

Ans : As we know that if $X \sim N(\mu1, \sigma1^2)$, and $Y \sim N(\mu2, \sigma2^2)$ are two independent random variables then $X + Y \sim N(\mu1 + \mu2, \sigma1^2 + \sigma2^2)$, and $X - Y \sim N(\mu1 - \mu2, \sigma1^2 + \sigma2^2)$.

Similarly if $Z = aX + bY$, where X and Y are as defined above, i.e Z is linear combination of X and Y, then $Z \sim N(a\mu1 + b\mu2, a^2\sigma1^2 + b^2\sigma2^2)$.

Therefore in the question

$2X1 \sim N(2u, 4\sigma^2)$ and

$X1+X2 \sim N(\mu + \mu, \sigma^2 + \sigma^2) \sim N(2u, 2\sigma^2)$

$2X1-(X1+X2) = N(4\mu, 6\sigma^2)$

2 X1 will be greater scale version than X1 + X2 . If X1 and X2 are normally distributed then the sum of the random sample will be exactly same.

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

Ans : option D.

The Probability of getting value between a and b should be 0.99.

the Probability outside the a and b area is 0.01 (ie. 1-0.99).

The Probability towards left from a = -0.005 (0.01/2).

The Probability towards right from b = +0.005 (0.01/2).

So since we have the probabilities of a and b, we need to calculate X, the random variable at a and b which has got these probabilities.

By finding the Standard Normal Variable Z (Z Value), we can calculate the X values.

Z=(X- pop.mean) / sd

For Probability 0.005 the Z Value is -2.57 (from Z Table).

Z * pop.mean + sd = X

Z(-0.005)*20+100 = -(-2.57)*20+100 = 151.4

Z(+0.005)*20+100 = (-2.57)*20+100 = 48.6

qnorm(0.995,100,20)

qnorm(0.005,100,20)

```
In [53]: from scipy import stats
         from scipy.stats import norm

In [58]: stats.norm.interval(0.99,100,20)

Out[58]: (48.48341392902199, 151.516586070978)

In [59]: #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 $Profit_1 \sim N(5, 3^2)$ and $Profit_2 \sim N(7, 4^2)$ 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 5^{th} 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?

Ans :

A) qnorm(0.025,45*5,3) # 219.1201

qnorm(0.975,45*5,3) # 230.8799

qnorm(0.025,45*7,3) # 309.1201

qnorm(0.975,45*7,3) # 320.8799

The Rupee Range will be [219.12, 230.87] + [309.12, 320.87] = [528.24, 551.74]

B) qnorm(0.05,45*7,3) # 310.0654

qnorm(0.05,45*5,3) # 220.0654

5th percentile of profit (in Rupees) = 310.0654+ 220.0654 = 530.1308

C) 2nd Division

jupyter ExcelR Assignment no. 2 (Basic Level Statistics II) Last Checkpoint: 4 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python

In [60]:
```python
from scipy import stats
from scipy.stats import norm
```

In [61]:
```python
# 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
```

In [62]:
```python
# 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 225.0 Million
```

In [63]:
```python
# 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 (99.00810347848784, 980.9918965215122) in Millions
```

In [64]:
```python
# B. Specify the 5th percentile of profit (in Rupees) for the company
# To compute 5th Percentile, we use the formula X=μ + Zσ; 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 170.0
```

In [65]:
```python
# 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)
```

Out[65]:
```
0.0477903522728147
```

In [66]:
```python
# Probability of Division 2 making a loss P(X<0)
stats.norm.cdf(0,7,4)
```

Out[66]:
```
0.040059156863817086
```

In [67]:
```python
# because probability of Division 1 making a loss in a given year is more than Division
```