**CBA: Solutions to Practice Problem Set 2**

**Topics: Normal distribution, Functions of Random Variables**

1. **B**.

Let T be the time it takes to work on a transmission. It is given that T ~ N(45, 82). Since the work begins 10 minutes after drop-off, we need the work to be completed in *t* ≤ 50 minutes. Since T has a normal distribution, Z = (T – μ)/σ has a standard normal distribution.

Therefore:

P(T ≤ 50) = P(Z ≤ (50 – 45)/8) = P(Z ≤ 0.62) = 0.7324 (from the Z-table)

Therefore P(T > 50) = 1 – P(T ≤ 50) = 1 – 0.7324 = 0.2676

Alternatively we can use Excel to get the same value using the function =1 – NORMSDIST(0.62) or =1 – NORMDIST(50, 45, 8, 1)

Or we can use R-function [1-pnorm(50,45,8)]

1. Employee ages:
2. **False**. The probability of the region between 38 and 44 corresponds to *P*(0 < *Z* < 1) ≈1/3 whereas the region above 44 corresponds to *P*(*Z* > 1) ≈ 1/6. Both of these probabilities were calculated based on the empirical rule, but you can also look them up in the table if you wish.
3. **True**. If *X* ~ *N*(38,6), then *P*(*X* < 30) = P(Z ≤ (30 – 38)/6) = P(Z ≤ - 1.33) ≈ 0.0912. This multiplied by 400 gives us the expected count.
4. 2*X1* is simply a scaled version of the random variable *X1*. So the distribution of 2*X1* will have a shape identical to that of *X1* i.e. it will have a normal distribution. Further we know that E[*aX*] = *a*E[*X*] and SD[*aX*] = |*a*|SD[*X*]. Therefore 2*X1* ~ N(2μ, (2σ)2)

The sum of two independent normally distributed random variables also has a normal distribution. Therefore *X*1 + *X*2 has a normal distribution. Further, E(*X*1 + *X*2) = E(*X*1) +E(*X*2) and Var(*X*1 + *X*2) = Var(*X*1) + Var(*X*2) when *X*1 and *X*2 are independent. Therefore (*X*1 + *X*2) ~ N(2μ, 2σ2).

Thus both will have normal distributions, but the second one has lower variance. Think about why.

1. **D**.

Based on the empirical rule, in order to include 99% of the probability, we need to go 2.58 standard deviations on either side of the mean. Therefore, the required range here is 100 ± 2.58\*20 = [48.4, 151.6]

1. Let X be the profit of the entire company in Rupees. Before answering the specific questions, let us specify the distribution of X. Since X is the sum of two normal random variables it also has a normal distribution. Further applying the formulae related to the means and SDs of functions of random variables (see solution to problem 3 above), we have:

E[X] = E[45\*(Profit1 + Profit2)] = 45\*(5 + 7) = 540 Million Rupees

SD[X] = SD[Profit1 + Profit2) = 45\*(sqrt(Var(Profit1) + Var(Profit2))) = 45\*sqrt(9 + 16) = 45\*5 = 225 Million Rupees

Therefore, X ~ N(540, 2252)

1. Based on the empirical rule, we need to go 1.96 standard deviations on either side of the mean to get 95% probability. Therefore, the required range is 540 ± 1.96\*225 = [99,981] Million Rupees.
2. We need the 5th percentile of X i.e. the point on the distribution of X, such that there is only 5% of the area to the left (i.e. 5% cumulative probability). We know from the standard normal table that this cumulative probability is associated with Z = -1.645. Therefore, the required value of X is 540 – 1.645\*225 = 169.875 Million Rupees.

Note that another way of thinking about this is to consider the 90% range based on the empirical rule. The required value of X here is the lower end of that range since a 90% area centered on the mean implies that there is 5% of the area on either side of that range.

1. This question concerns the original profit distributions. Let us calculate the Z-scores associated with zero for each of the divisions.

For Division 1: Z-score for a profit of zero = (0 – 5)/3 = (- 1.67)

For Division 2: Z-score for a profit of zero = (0 – 7)/4 = (- 1.75)

The probability of loss for division 1 is the area under the standard normal distribution pdf to the left of (-1.67), and that for division 2 is the area under the standard normal distribution pdf to the left of (-1.75). Therefore the probability associated with the second one will be lower (as we are going farther into the tail). *Note that we can compare probabilities by comparing the associated Z-scores without having to compute actual probabilities*.