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

**Topics: Sampling Distributions and Central Limit Theorem**

1. The normal quantile plots: Before looking at the actual plots, recall that a normal quantile plot plots the theoretical values on the X-axis and the actual (data) values on the Y-axis. For each point in the data, it calculates the expected theoretical value (or its Z-score) based on the number of points and the properties of the normal distribution. This is the X-coordinate of the point. The Y-coordinate is the actual data point observed. If these two are equal, the point will lie along the diagonal. If a point is above the diagonal, this means that its actual value is higher than the one that would be expected based on the normal distribution (with the mean and SD given by the data). If it lies lower than the diagonal line, it is the opposite. Now we can examine each of the plots.
2. **Skewed**. Explanation: The normal quantile plot has a crescent shape with the points lying above the diagonal both at the lower and upper ends. Further, the points are a lot closer together in the left tail than in the right tail. This implies that the data is right skewed i.e. has a lot more points closer to the mean on the left tail, and has a long tail on the right side. Both of these result in data values that are higher than those expected.
3. **Outliers**. Explanation: There are a couple of points which are quite far away from the rest of the data and have values quite different from those expected from a normal distribution (especially in the left tail. It is not quite so pronounced on the right tail).
4. **Normal**. Explanation: All of the points track the diagonal quite closely and lie within the bounds.
5. **Bimodal**. Explanation: Notice that the density of points near the median is much sparser than on either side of the median. This means significant probability masses on either side of the median with a low probability mass near the median, a shape characteristic of a bimodal distribution.
6. Warehouse shipments.
7. **False**. While normality of the weights of individual packages guarantees that the mean weights will be normally distributed, it is not necessary that the weights of individual packets will be normally distributed. Based on the Central Limit Theorem, the mean weights are normally distributed even if the individual package weights don’t have a normal distribution so long as the sample size conditions necessary for normality are satisfied. So those conditions should be checked.
8. **True.**
9. **D**.

Here, μ = 50, σ = 40, n = 100. Note that the probability of investigation is the same as probability that the sample mean is beyond $45 and $55.

From Central Limit Theorem, we know that .

We first calculate the standard error of the mean, which is

Then, we calculate the z-values corresponding to 45 and 55, which are (-1.25) and (+1.25). Therefore,

Consequently, probability that the sample mean is beyond the limits (and hence there is an investigation) is (1 - 0.79) = **0.21**.

1. **D**.

In this case, we know that . Thus, the z-values corresponding to 45 and 55 are (-1.96) and (+1.96) respectively.

Using the definition of z-value, we get

Thus,

Finally, using the definition of , we obtain which gives n = 245.86.

We choose 250 as the closest integer greater than this number. Choosing anything smaller than 245.86 will yield a probability of investigation higher than 0.05.

1. **D**.

Again, from Central Limit Theorem, . Hence, the average of many sample means will tend to the population mean, which is 720.