Worksheet-Set 1: Statistics Assignment

Answer keys for questions from Q1 to Q11:

Question Number	Answer keys
Q1 :	a) 2 Only
Q2 :	d) 1, 2 and 4
Q3 :	a) True
Q4 :	a) 1 only
Q5 :	b) 1
Q6 :	b) No
Q7 :	a) Yes
Q8 :	d) All of the above
Q9 :	a) K-means clustering algorithm
Q10 :	d) All of the above
Q11 :	d) All of the above

Q12. Is K sensitive to outliers?

Yes. K means is sensitive to outliers.
For e.g. Data set point are 1 2 3 7 8 80
Here 80 is outlier.

For two clusters: K=2.
Initial centroid points will be: C1=1 C2=7.

After first iteration
New centroid points will be: C1=2 C2=31.67

As 80 data point which is outlier comes in cluster 2.
Cluster 2 centroid changes to accommodate 80.
Therefore K means is sensitive to outliers.

Q13. Why is K-means better?

- Relatively simple to implement.
- Scales to large data sets.
- Guarantees convergence.
- Can warm-start the positions of centroids.
- Easily adapts to new examples.
- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

Q14. Is K means a deterministic algorithm?

No. k-Means is a non-Deterministic model. k-means is a partitioning-based clustering algorithm. k-means method for clustering is an iterative process in which an initial partition of given k clusters is then improved by applying a search algorithm to the data. Simplifying, given a pre-defined number (k) of clusters, the algorithm:
-begins with an initial set of k cluster centers (i.e. the centroids)
-(re)assigns objects to the closest centroids
-recalculates centroids according to new memberships of the data points.
-repeats the last two steps until a consistent result is found or until the maximum number of iterations is reached.
This means that running the algorithm several times on the same data, could give different results.