1. D
2. A
3. A
4. C
5. A
6. A
7. C
8. B
9. B
10. Histograms and box plots are graphical representations for the frequency of numeric data values. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets.
11. There are three ways to select the metrics:

    Use standards. I prefer metrics that have been tested by others;

    Measure yourself the way your customer measures you

    Only measure metrics that have an owner

12. Statistical significance can be accessed using hypothesis testing:
    – Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)
    – Then, we choose a suitable statistical test and statistics used to reject the null hypothesis
    – Also, we choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value)
    – We calculate the observed test statistics from the data and check whether it lies in the critical region
    Common tests:
    – One sample Z test
    – Two-sample Z test
    – One sample t-test
13. Examples of data does not have a Gaussian Distribution-
    - Allocation of wealth among individuals
    - Values of oil reserves among oil fields (many small ones, a small number of large ones)
14. distribution of salaries for residents in a certain city
15. Likelihood refers to how well a sample provides support for particular values of a parameter in a model.