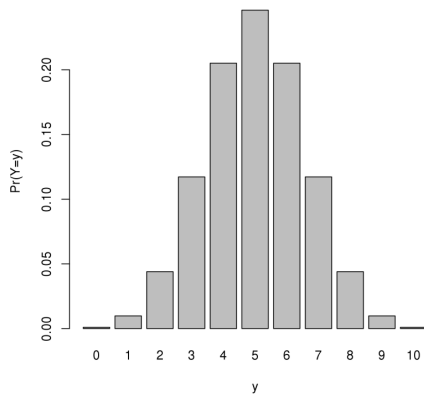**Multiple choice questions:**
Each question carries 1 mark

1. The researcher sequenced the nucleotide sequence from *Aspergillus niger* (Fungus), and further wants to study the protein sequence of the same. Which alignment tool will be best suited for this analysis?
   a. BLASTP
   b. BLASTN
   c. **BLASTX**
   d. TBLASTN

2. What is this distribution called?

   Ans: Bernouli/normal/gaussian/binomial

3. To use molecular data to reconstruct evolutionary history requires making a number of reasonable assumptions. Which of the following is incorrect about it?
   a. The molecular sequences used in phylogenetic construction are homologous
   b. The molecular sequences used in phylogenetic construction share a common origin
   c. **Phylogenetic divergence cannot be bifurcating**
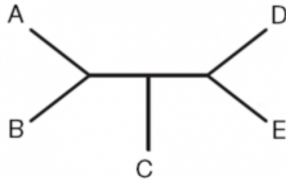   d. Parent branch splits into two daughter branches at any given point

4.
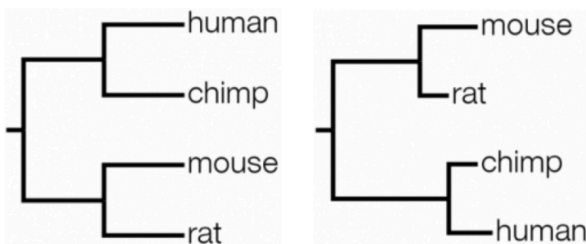
   These trees have same topology
   a. **True**
   b. False

5.

The tree above is unrooted, where could you place the root so that A and B are not each other's closest relatives
   a. **At the node where A joins B**
   b. At the not where D joins E
   c. At the midpoint of the tree
   d. Both a and b

6. Progenitor sequences represented by the _____ branches of the tree are derived by alignment of the _____ sequences.
   a. Outer, outermost
   b. **Inner, outermost**
   c. Inner, innermost
   d. Outer, innermost

   **clue:consider progenitor as reference sequence or root sequence

7. In contrast to microarray methods, RNA sequencing-based approaches have the following advantages:
   a. It offers single-nucleotide resolution
   b. A priori knowledge of genome or genomic regions are not required
   c. Higher dynamic range of expression levels
   d. **All of the above**

8. Which statistical test is commonly used to identify differentially expressed genes by comparing two groups?
   a. **t-test**
   b. Pearson correlation
   c. ANOVA
   d. Chi-square test

9. Which statistical measure is used to quantify the spread of a distribution?
   a. Mean
   b. Median
   c. **Variance**
   d. Mode

10. In a box plot, what does the length of the box represent?
    a. **Interquartile range (IQR)**
    b. Mean
    c. Standard deviation
    d. Variance

Close-end:
1. PSI BLAST uses the top ranked matches from first search to build a profile and uses this profile to re-search the database for matches. (TRUE or FALSE).
   Ans:**TRUE**

2. In PSI-BLAST first search is done using BLASTP, and second search is done using PSSM profile and significance of match is tested using p-value instead of e-value. (TRUE or FALSE).
   Ans:**FALSE**
3. Please provide an example when we need to use the Chromosome BLAST.
   Ans:**When comparing two chromosomes, such as human chromosome 20 vs mouse chromosome 2**
4. The scoring of gaps in a MSA (Multiple Sequence Alignment) has to be performed in a different manner from scoring gaps in a pairwise alignment. Is it true or false? Give an explanation.
   **Ans: True. As more sequences are added to a profile of an existing MSA, gaps accumulate and influence the alignment of further sequences. CLUSTALW calculates gaps in a novel way designed to place them between conserved domains.**
5. From the bootstrap value given below, what value supports the grouping of c, d and e as common ancestral roots?
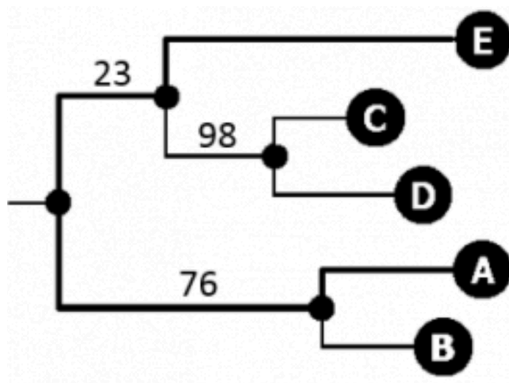
   Ans: 23

6. Draw a morphology of box-plot and label the necessary pointers.
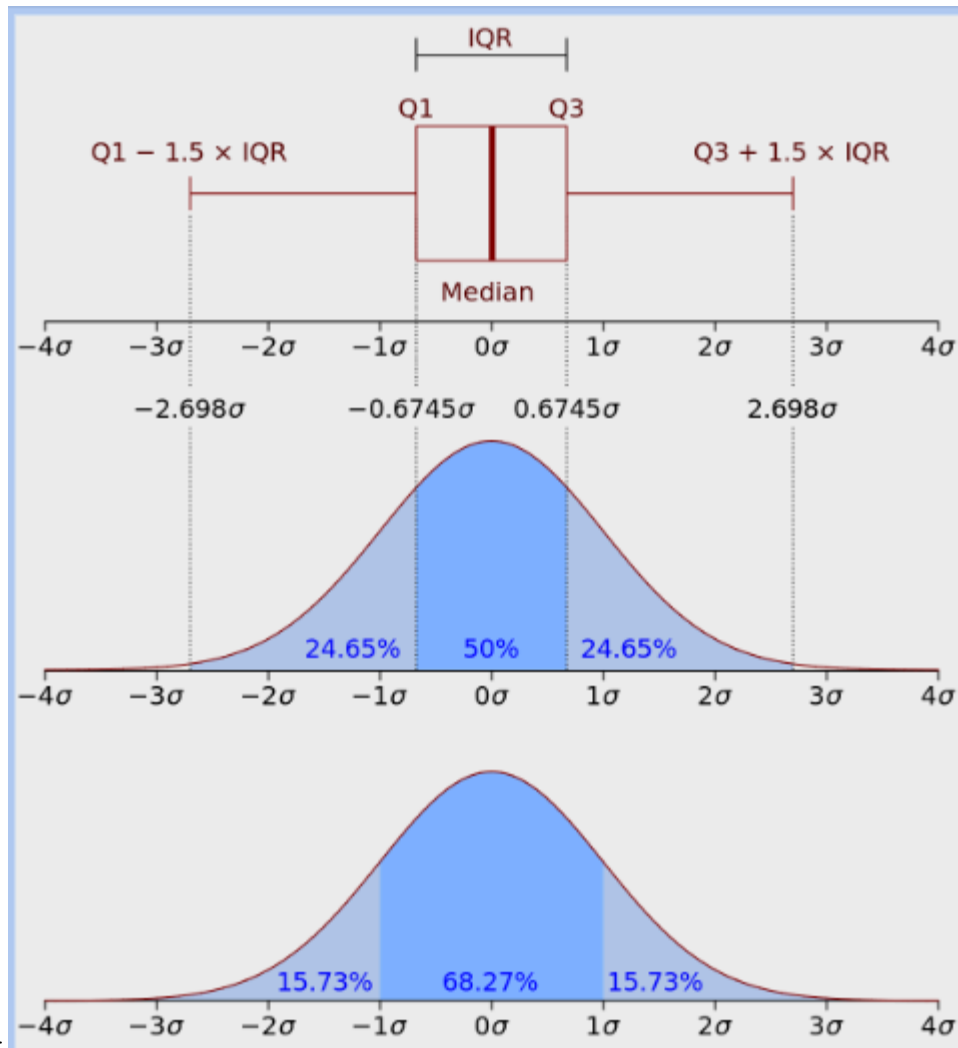
Ans:

7. Imagine you are designing an experiment to measure gene expression in the blood of patients who have been treated with a new drug. You want to measure expression of all genes in the genome and may want to use the data to identify novel transcripts (e.g. Exon-retention in Alternative splicing) in the future. Which method would you choose and why?

      a. RNA-Sequencing

      b. Microarray

Ans: RNA-SEQ. Microarray analysis can only be used to measure gene expression of known transcripts that are printed on the array. RNAseq can also be used to measure gene expression across the genome. It captures information about all RNA molecules in a sample (regardless of whether their sequence is known) and can be used to identify novel transcripts.

8. Suppose you're conducting RNA expression analysis on rheumatoid arthritis. Which database would you choose to access raw data, and for investigating the disease association of four specific genes, which database would be most appropriate?

Ans: GEO and OMIM.

9. What unique advantage does box plots provide in data visualization?

Ans:They allow for easy comparison of multiple distributions side by side

10. If the median line in a box plot is not centered within the box. It indicates —-------- in the distribution of the data.

     Ans: skewness