Sample questions:

1) You wish to estimate the mean of a population. Although you do not know it, the population has mean 1, and standard deviation 2. You draw 100 samples randomly from the population.
   1) How can you estimate the population mean from these samples?
   2) How accurate is your estimate? (i.e. what is its standard deviation)
   3) How can you make your estimate twice as accurate? (i.e. reduce its standard deviation by a factor of two).

2) I have a population, and I state that its mean is 0, and its standard deviation is 1. I give you 100 samples randomly drawn from this population. The mean of these samples is 2.
   1) The sample mean of a set of samples is a random variable. Write an expression for its standard deviation.
   2) Give an expression for the probability of drawing a sample of 100 items from a population of mean 0 and standard deviation 1, where the sample mean is 2.
   3) Suggest a conclusion about my assertion that the population mean is 0

2) I have two normal random variables. X has zero mean, and unit variance. Y has zero mean and unit variance.
   1) what is mean(2X)?
   2) what is the variance of (2X-4Y)?
   3) what is the standard deviation of (X-Y)

3) I have two populations. I draw a random sample from population A, which has 100 items in it. The mean of this sample is 0. The standard deviation is 1. I now draw a random sample from population B, which has 100 items in it. The mean of this sample is 0.1, and the standard deviation is 0.9.
   1) What can you say about the hypothesis that the two populations have the same mean?

4) I observe Bevande cafe, to see how many times a day a customer uses a $50 bill. I make observations on 10 days, and see: 1, 2, 3, 2, 1, 1, 2, 2, 1, 3 uses per day.
   1) I want to fit a Poisson model to this data. How should I choose the intensity \( \lambda \)?
   2) What is the best value of \( \lambda \)?
   3) Using this model, what is the probability that I see 5 uses in one day?

5) I wish to choose a normal distribution to model a 1D dataset.
   1) Name a good method to choose the mean and standard deviation
   2) Show that this method chooses a mean that is the average of the observations
6) I flip a coin ten times. I do not know \( p \), the probability it comes up heads
   1) I observe \( h \) heads and \( 10-h \) tails. Name a good method to estimate \( p \)
   2) Show that, using this method, the estimate of \( p \) is \( h/(t+h)=h/10 \)

7) I flip a coin 3 times. I do not know \( p \), the probability it comes up heads. I estimate \( p \) using maximum likelihood. I observe \( h=0 \) heads and \( t=3 \) tails.
   1) What value does maximum likelihood report for \( p \)?
   2) Why is it not safe to assume that \( p=0 \)?
   3) Describe (briefly) another estimation procedure that might help

8) I flip a coin repeatedly until it comes up heads.
   1) What distribution describes this experiment?
   2) I flip the coin \( n \) times (i.e. \( n-1 \) tails, then the final head). How can I estimate \( p(\text{head}) \) from this?
   3) Use the estimation procedure of 2 to estimate \( p(\text{head}) \)

9) Why do we usually use separate training, validation and test sets for classification problems?

10) I wish to classify a high dimensional dataset.
    1) Why can I not build a histogram based classifier?
    2) How does Naive Bayes help?
    3) What assumption does Naive Bayes require?

11) A logistic regression classifier uses the model \( \log \left( \frac{p(1|x)}{p(-1|x)} \right) = a^T x \)
    1) Show that this means the decision boundary is a hyperplane, with equation \( a^T x = \text{constant} \)

12) How can I build a multi-class classifier out of two-class classifiers (briefly)?

13) Why is \( [[1, 0];[0, -1]] \) not a covariance matrix?
14) \( [[1, 0, 0];[0, 4, 0];[0, 0, 9]] \) is the covariance matrix of a dataset. This data forms a “blob” in 3D.
    1) Describe the blob qualitatively
    2) What transformation would make this blob a sphere?
    3) What are the first two principal components of this dataset?

15) Explain (briefly) two uses of principal components analysis
16) I have a set of \( R \) movie renters and a set of \( M \) movies. Each renter gives me a score of how much they liked a movie they have rented; nobody has seen all the movies. I build an \( RxM \) table \( D \) which contains either this score, or an \( x \) (if there isn’t a score for that renter-movie pair).
    1) What is an SVD (Singular Value Decomposition)?
    2) Assume that my table is full (i.e. there are no \( x \)’s). What would the singular values of that table be like? why?
3) I can compute a useful factorization of D without knowing all the entries. I will choose A, B so that AxB is similar to D in the entries I know.
   1) Why do I expect that this would work in a practical situation?
   2) How would I obtain A, B?

17) I have a set of d-dimensional vectors, uniformly distributed in the cube with edge length 2 whose center is the origin.
   1) Write an expression for the fraction of vectors that have at least one component with absolute value $(1-\epsilon)$
   2) What do you conclude?