3. More compact display of distribution summary
   a. Task: Make a plot that can display multiple samples or variables at the same time
      i. Ex. midterm 1 and 2 test scores and homework total
      ii. Overlyting histograms is not so informative
   b. Solution:
      i. Draw axis representing range of data values.
      ii. Draw short line segments perpendicular to axis at lower and upper quartiles and median
      iii. Draw lines connecting either ends to make a box
      iv. Draw line parallel to axis through midpoints of segments representing quartiles out to most extreme points
         • Except omit the part inside the box, to form two line segments
         • These are called whiskers.
         • Refinement: extend segments only to most extreme values within 1.5 IQR of median, and represent remainder by stars or circles or something.
   c. Procedure is called boxplot.

VIII. Task: Calculate from the data some quantity $T$ to estimate some aspect $\mu$ of the population that they are drawn from.

A. In order for this to make sense, be some aspect of the distribution of each of the variables must be the same for all of them, in order to have something to measure.

   1. For the present we will assume that these variables are identically distributed.

B. Call something that is calculated from data a statistic.

C. What makes a statistic a good estimator?

   1. Does the estimator in the long run center around the true value?
      a. Does $E[T] = \mu$?
      b. If so, estimator is called unbiased.
      c. Otherwise, estimator is called biased.

   2. Is estimator very variable, or do you get similar results from various runs of the experiment?
      a. Is $\text{Var}[T]$ small?

   3. Is estimator typically close to the target?
      a. True of last two properties are both satisfied.

D. Potential estimators of location:

   1. $T_1 =$ sample mean.
   2. $T_2 =$ sample mean of first half of observations.
   3. $T_3 =$ sample median.
   4. $T_4 =$ sample trimmed mean.

E. It is always true that sample mean is an unbiased estimator of expectation.

   1. Examples:
      a. $X_1, \ldots, X_n \sim N(\mu, 1)$
         i. Clearly $T_1$ and $T_2$ are unbiased for $\mu$, by fact above.
         ii. Distributions of $T_3$ and $T_4$ are also unbiased for $\mu$, since their distributions are symmetric about $\mu$.
         iii. If $X_1, \ldots, X_n$ are independent, then $\text{Var}[T_1] = 1/n$ and $\text{Var}[T_2] = 1/(n/2) = 2/n$, and so $T_1$ is better.
      b. $X_1, \ldots, X_n \sim \text{Expon}(\lambda)$