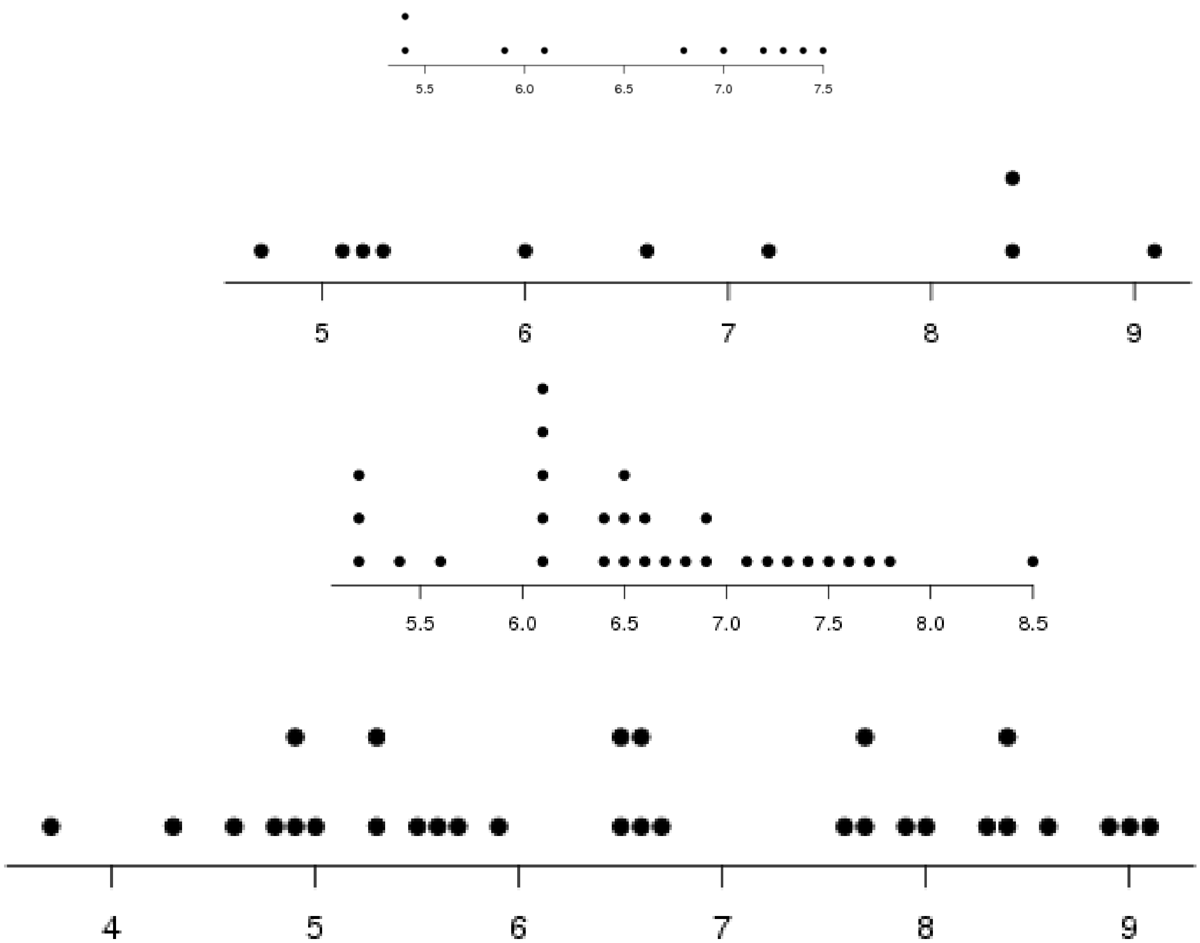
MATH 143: Introduction to Probability and Statistics

Worksheet 7 for Thurs., Nov. 5: Inference for a Single Population Mean

Suppose the 4 dotplots below represent the number of hours slept last night for 4 different groups of students, Class 1 on top down to Class 4 on bottom.



1. A statistical software package reports the means and standard deviations for the sleep times of the four classes as

Class	$\bar{x}$	s	Class	$\bar{x}$	s
1	6.6	0.8246	3	6.6	0.8246
2	6.6	1.5972	4	6.6	1.5968

   (a) What is the standard error for the mean $SE_{\bar{x}}$ for each class?

   (b) If you were going to employ the data from one class to construct a 95% confidence interval for the mean number of hours slept last night by Calvin students, which class would produce the narrowest interval?

2. (a) Construct a 95% confidence interval for μ (the mean number of hours slept by Calvin students last night) using the hypothetical data from Class 4.

   (b) One might say she "is 95% confident the true mean μ lies in the confidence interval" (from part (a)). Describe what this actually means.

   (c) Based on this confidence interval alone (i.e., not doing any other work), what can be said about the result of an hypothesis test concerning μ whose null and alternative hypotheses are

$$\mathbf{H}_0: \ \mu = 7\,, \qquad \mathbf{H}_a: \ \mu \neq 7\ ?$$

Like the construction of a *level C confidence interval for the mean μ of a single population*, the manner in which we conduct an *hypothesis test for the mean of a single population* is not much changed from the z-procedures we learned in Chapter 15 when the standard error for the mean $\mathrm{SE}_{\bar{x}}$ is used in place of the unknown standard deviation $\sigma/\sqrt{n}$ of the sampling distribution for $\bar{x}$. The two main differences are:

- When one looks up a P-value in Table C for the corresponding t-distribution, one selects the row where the degrees of freedom is 1 less than the sample size (i.e., $df = n - 1$). (This is no surprise, as the same requirement comes up when finding the critical value t^* for constructing a CI.)

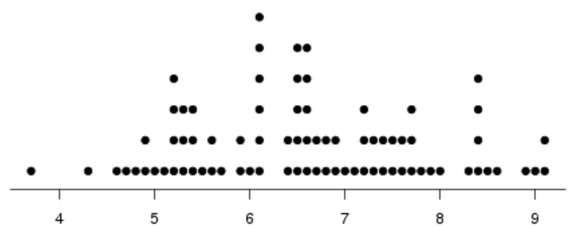
- The values in the interior of Table C are T-values, not probabilities as they are in Table A. The corresponding one- and two-sided P-values are given in rows at the bottom of Table C.

As with the z-procedures, a P-value still indicates the probability, assuming the null hypothesis to be true, of having obtained a sample result at least as extreme as the one actually attained. (The smaller the P-value, the stronger the evidence *against* the null hypothesis.)

3. Suppose we are told that Calvin students, on average, sleep 7 hours per night. Friday's are often days on which tests are given or assignments come due.

   (a) State appropriate null and alternative hypotheses about the average number of hours sleep for Calvin students last night.

   (b) When all four classes are combined, the mean number of hours slept is still 6.6 while the standard deviation for these 80 students is 1.24645. (See the resulting distribution at right.) Using this data, compute a test statistic (call it t instead of z).



   (c) Determine an approximate P-value which corresponds to the test statistic from part (b) and the alternative hypothesis you stated in part (a). Does this represent strong evidence against the null hypothesis? Explain why or why not.

   (d) In Problem 4 of yesterday's worksheet (Worksheet 6), you were asked to compute certain probabilities stated in the form $P(T >$ some number$)$. Use notation such as this to provide some meaning to the P-value from part (c).

   (e) Review the assumptions we make about our data when carrying out our inference procedures. Do these assumptions seem to hold for *this* data? Would your answer change if you knew that all 4 classes were 8-o'clocks?