

MATH 143: Introduction to Probability and Statistics

Worksheet 6 for Wed., Nov. 4: t Distributions

The distinctions between *population* and *sample* and between *parameter* and *statistic* are at the heart of statistical inference. The whole point is to learn something about unknown population parameters on the basis of observed sample statistics.

1. For each of the following,

- identify whether the quantity described is a parameter or a statistic,
- what you consider the underlying population to be, and
- the symbol used to denote the value.

(a) the mean time spent sleeping last night by the students in your class

(b) the standard deviation of sleeping times (for just last night) for all students at Calvin College

(c) the mean number of runs scored in all Major League baseball games played in 2007

(d) the standard deviation of *amount of change* being carried by shoppers in a certain store one Monday morning

Up until now we have carried out two inference procedures—the construction of a confidence interval for a population mean, and an hypothesis test concerning hypotheses over the same quantity—under the unreasonable assumption that, though we did not know μ for the population, we knew σ . These inference procedures used the only available estimate for μ , the sample mean \bar{x} . There is a similar estimate for σ available—namely, s , the sample standard deviation—but we have not made use of it. We will now learn the right way to carry out these inference procedures, doing so with the available quantity s instead of σ (nearly always unavailable). There is not much that changes in the procedures, really. For instance, a confidence interval is still determined by the general principle (with s in place of where σ was formerly used)

$$\bar{x} \pm (\text{margin of error}) \quad \text{where} \quad (\text{margin of error}) = (\text{critical value}) \frac{s}{\sqrt{n}} .$$

The main difference is in finding the critical value. We know that when our sample is an SRS from a large (a good deal larger than the sample size n) population, then the sampling distribution for

\bar{x} is $N(\mu, \sigma/\sqrt{n})$. But, we would obtain inaccurate probabilities if we tried to use, in place of this, the normal distribution $N(\mu, s/\sqrt{n})$. Instead, we must get our probabilities (and the corresponding critical values) from the t -**distribution** with $n - 1$ degrees of freedom, abbreviated $t(n - 1)$.

2. Make a sketch of the t -distribution with 8 degrees of freedom. Identify approximate locations for $(-t^*)$ and t^* so that the region between the two under the curve would capture approximately 90% of the area under the curve. Shade that region. What percent of the area lies to the *right* of t^* ? To the right of $(-t^*)$?

3. Use Table C, p. 687 to determine the missing values in the table below:

Pct. of values captured between $(-t^*)$ and t^*	Number of degrees of freedom	t^*
70%	1	1.963
90%	8	
95%	14	
99%	30	
99.5%	65	
95%	525	

4. Use Table C to determine, approximately, the values of

- (a) $P(T > 1.415)$, assuming $df = 7$.
- (b) $P(T < -2.956)$, assuming $df = 7$.
- (c) $P(T < -2.68 \text{ or } T > 0.9)$, assuming $df = 7$.
- (d) $P(T > 1.70)$, assuming $df = 24$.