**Part II: Comparison of Fish Species Frequently Consumed**

**In the real world, no one actually has complete sampling distributions.**  Instead, the sampling distributions for quantitative data are **modeled** using t distributions.  These distributions vary slightly based on degrees of freedom, which is one less than the sample size.   P-values are typically found using either tables or technology.   For this lesson, we will use Excel, but you may have access to other software or graphing calculators, to find the p-value.



Figure 6: T-distribution for mercury in fish assuming the null hypothesis of 0.50 mg/kg is true

The center of the t distribution is equal to the **mean** of the population as stated in the null hypothesis, which is 0.5 mg/kg for our fish population (Figure 6).   The amount the distribution is spread out is based on the standard error.  The **standard error** is the standard deviation of the statistics (e.g., sample means) assuming we could find every possible sample mean that could be drawn from a population.

Since this can’t be done, we use the formula ,  where **SEx** is the estimated standard error, **s** is the standard deviation of the sample, and **n** is the sample size.

1. What is the effect on SEx if the sample size increases?

To find the p-value, it is necessary to determine the number of standard errors our data is from the hypothesized mean.  Doing so will give us a *t*-value, which we will use to find a p-value.

The *t* formula is: 

1. What is the effect on *t* if the sample size increases?

The numerator of this formula finds the distance between the sample mean and the hypothesized mean of the population.  Dividing by the standard error results in *t* being the number of standard errors that the sample mean is from the hypothesized mean of the population.

We will use the Separation Lake data to demonstrate how to test the hypothesis using the t-test.

H0:  μ = 0.5,   H1:  μ > 0.5, Level of Significance:  0.05

Using the actual data:  Bass in Separation Lake

Sample Mean:   

Sample Standard Deviation:  s = 0.3,

Sample size:  n = 22   Degrees of Freedom (df = n-1) = 21

t-value 

Since the direction of extreme is to the right, we can use the following formula in Excel to find the probability a t-value would be greater than 0.469, assuming the null hypothesis was true.

Excel Formula for finding the p-value if the direction of extreme is to the right:  =T.DIST.RT(t value,df)

Excel Formula for finding the p-value if the direction of extreme is to the left:  =T.DIST(t value,df)

For this problem, =T.DIST.RT(0.469,21) = 0.322



Figure 7: T-distribution for mercury in bass in Separation Lake

The p-value of 0.322 is greater than the level of significance of 0.05 (Figure 7).  Therefore, the data supports the null hypothesis, so we would conclude that the average mercury concentration in the population of fish is not significantly greater than 0.5 mg/kg (t = 0.469, p = 0.322, n = 22).

1. How would the p-value differ if the sample size was 500 instead of 22 fish, but kept the same mean and standard deviation?  What do you conclude about the effect of sample size?

1. How would the p-value differ if the standard deviation was 0.03 instead of 0.3, but kept the same mean and a sample size of 22?  What do you conclude about the effect of variation in the sample as determined by standard deviation?

Large-mouth Bass is just one of the fish species most frequently consumed by Grassy Narrows and Wabaseemong First Nation community members.  Walleye (*Stizostedion vitreum vitreum*), Northern Pike (*Esox lucuis*) and Whitefish (*Coregonus clupeaformis*) are three other commonly harvested fish species in the English-Wabigoon river system.

1. Would you expect the methylmercury levels to differ among these fish species and Large-mouth Bass?  Why or why not?