**Part I: Are the fish in the rivers and lakes by Grassy Narrows First Nation safe to eat now?**

Nearly all fish and shellfish contain traces of mercury. But some contain high levels. Eating large amounts of these fish and shellfish can result in high levels of mercury in the human body. In a fetus or young child, this can damage the brain and nerves. Because of the mercury found in fish, Health Canada has set a standard (maximum limit) of 0.5 ***parts per million (ppm)***total mercury in retail fish. Standards are enforced by the Canadian Food Inspection Agency (Health Canada, 2021).

Let’s conduct a hypothetical investigation of the fish in the lake.  To determine if the fish are too contaminated to consume, we have to address a few problems.

* **We cannot test all the fish**, therefore we cannot know the absolute truth about the average contamination concentration.
* The concentration in the fish is *heterogeneous*.
* We want to base our decision on evidence, which means we need to take some fish ***samples***.

There are two types of mathematical problems that you may be familiar with, deterministic and stochastic.  **Deterministic** problems use math, such as arithmetic or algebra, and in which everyone should get the **same answer**.

Conversion Example: Convert 0.5 ppm to mg/kg and convert 0.5 mg/kg to ppm. Your answers should be the same.

**Stochastic** problems are ones in which the inputs or variables are described by probability distributions rather than a single value.  For example, the mercury concentration in the fish varies between individuals.  We can describe the fish population’s mercury concentration by using statistics, such as the mean and standard deviation.  Statistics provide tools to understand variation in the world and make decisions when dealing with partial evidence.  This is our current problem.  *How do we decide if the mercury concentration in the fish is too high for safe consumption when we can only get evidence from the fish we* ***sample****?*

The process begins with asking an appropriate question such as:  **Is the average mercury concentration in the fish greater than 0.5 mg/kg, which is the Canadian standard for safety?**

Next, we define the population:  The **population** is all

the Large-mouth Bass (*Micropterus salmoides*) in

 Separation Lake, which lies between two First Nation

Communities (Grassy Narrows and Wabaseemong).

Figure 2: Map of Separation Lake in the Wabigoon River system (Neff et al. 2012)

The average mercury concentration was an estimate based on the **sample of fish** that was caught, and may differ considerably from an estimate derived from another sample of fish.  One of the fundamental elements of hypothesis testing is a sampling distribution.  A sampling distribution is a collection of statistics, such as sample means, that could possibly be drawn from the population if we were to repeatedly draw samples from the population. We always start hypothesis testing with the assumption that the null hypothesis is true.  The table below is an example of a population that conforms to the null hypothesis with a mean of 0.5 mg/kg.

**Table 1: Mercury Concentration in a fictitious population of Bass in Separation Lake with a mean of 0.5 mg/kg.**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 9 | 0.48 | 0.68 | 0.41 | 0.33 | 0.19 | 0.65 | 0.40 | 0.43 | 0.47 | 0.31 |
| 8 | 0.63 | 0.01 | 0.68 | 0.55 | 0.25 | 0.69 | 0.21 | 0.40 | 0.60 | 0.12 |
| 7 | 0.41 | 0.42 | 0.30 | 0.39 | 0.49 | 0.70 | 0.24 | 0.18 | 0.89 | 0.57 |
| 6 | 0.24 | 0.43 | 0.64 | 0.47 | 0.70 | 0.45 | 0.73 | 0.74 | 0.30 | 0.62 |
| 5 | 0.55 | 0.51 | 0.84 | 0.32 | 0.39 | 0.66 | 0.44 | 0.84 | 0.58 | 0.61 |
| 4 | 0.49 | 0.56 | 0.60 | 0.56 | 0.44 | 0.71 | 0.75 | 0.60 | 0.70 | 0.42 |
| 3 | 0.41 | 0.58 | 0.47 | 0.41 | 0.60 | 0.51 | 0.29 | 0.68 | 0.29 | 0.29 |
| 2 | 0.52 | 0.70 | 0.53 | 0.37 | 0.70 | 0.71 | 0.31 | 0.68 | 0.44 | 0.41 |
| 1 | 0.49 | 0.28 | 0.47 | 0.35 | 0.46 | 0.33 | 0.42 | 0.51 | 0.41 | 0.45 |
| 0 | 0.84 | 0.63 | 0.75 | 0.65 | 0.52 | 0.35 | 0.51 | 0.58 | 0.55 | 0.58 |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

To help you understand a sampling distribution, you will work with your classmates to generate a small one.  You will need to use a random process to select the data for the sampling distribution.  Follow the sampling process described by your instructor. Every student will conduct this process so the class generates at least 30 sample means.

Table 2 shows randomly generated X and Y values.  The data comes from Table 1.  Make a similar table in your notes for the numbers you randomly selected.

**Table 2: Random x and y values from Table 1.** *Write your values to the side of this and find the mean.*

|  |  |  |
| --- | --- | --- |
| X | Y | Data |
| 2 | 0 | 0.75 |
| 8 | 0 | 0.55 |
| 0 | 0 | 0.84 |
| 9 | 2 | 0.41 |
| 1 | 6 | 0.43 |
|  | Mean: | 2.98/5 =0.596 |

Write down your sample mean on the Post-it note provided and then place it on the consensogram on the main board at the front of the classroom. Use it to complete the figure below.

Draw a histogram of the class data for mercury contamination in Figure 3.



Figure 3: Class Data for Mercury Contamination in a fictitious population of fish in Separation Lake.

This plot shows the number of observations for each interval class with bars that are ~0.05 mg/kg wide (E.g., range from 0.2-0.249, 0.25-0.299).  This is a sampling distribution that is based on a simulated population where the mean is 0.5 mg/kg.



Figure 4: Sampling distribution of the mean mercury concentration in fish, assuming the null hypothesis is true.

Figure 4 is a sampling distribution constructed from this data but with 500 sample means.  The numbers at the top of the bars indicate the proportion of the sample means in that class (bar).  They also indicate the probability that a sample mean would fall in that class.  Notice the shape of the distribution and that 0.5 is in the center.

**Complete Table 3 below to demonstrate your understanding of the probabilities in the sampling distribution. Check your answer with a peer.**

Table 3: Probabilities in a sampling distribution (Figure 4)

|  |  |
| --- | --- |
| Interval | Probability |
| [0.55,0.60) | 0.16 |
| [0.30,0.35) |  |
| [0.25,0.35) |  |
| **Less than or equal to 0.35** |  |
| [0.6,0.75) |  |
| **Greater than or equal to 0.6** |  |
| **Greater than or equal to 0.75** |  |
| **Less than or equal to 0.25** |  |
| **Greater than 0.25** |  |

In hypothesis testing, we write hypotheses about the average we think we would find if we could actually test all the fish (i.e., the population).  This average for the population is called a **parameter** (a number that summarizes a population).  The symbol that we use for the average of a population is the Greek letter mu, **μ**.

When we write hypotheses, we actually write two of them, the **null hypothesis** (H0), which we will begin by assuming is true, and the **alternative hypothesis** (H1), which we will accept if the evidence does not support the null.  Our choice of parameter should be meaningful, leading to different courses of action if the evidence supports the null or alternative hypothesis.  For this problem, the hypotheses would be:

H0:  μ = 0.5 mg/kg

H1:  μ > 0.5 mg/kg

1. Which of these hypotheses indicates the fish are **safe to eat**?
2. Which of these hypotheses indicates the fish are **not safe to eat**?
3. Do you want the average mercury concentration of the fish population to be at the maximal limit, or do we want to be more conservative and have it lower?
4. How might this be affected by consumption levels?

Before exploring this problem with some evidence, it can be helpful to develop an intuitive sense of how partial evidence leads to drawing conclusions.  We will do this using what-if questions.

For homework (before next class), answer the following questions on Socrative.

1.  What if we took some Bass samples and found the average concentration of mercury was 3 mg/kg, which hypothesis would be supported, and would it be wise to eat the fish?

2.  What if we took some Bass samples and found the average concentration of mercury was 0.1 mg/kg, which hypothesis would be supported, and would it be wise to eat the fish?

3.  What if we took some Bass samples and found the average concentration of mercury was 0.57 mg/kg, which hypothesis would be supported, and would it be wise to eat the fish?