

STAT 481 Bonus Project

Due on 05/03/2016

Background:

Groundwater normally contains impurities in such micro-quantities as to be harmless. For one region with clean groundwater, for example, the arsenic contents of individually collected vials of groundwater are normally distributed with a mean of 25 micrograms per liter ($\mu g/l$) and a standard deviation $5 \mu g/l$. This particular distribution of arsenic levels will be called *standard levels* throughout this project. Standard levels are okay, because not until the arsenic content reaches $50 \mu g/l$ is it considered unsafe.

We will say that the arsenic is at *contaminated levels* if the arsenic levels in individually collected vials of groundwater are normally distributed with a standard deviation of $5 \mu g/l$, but with a mean of $35 \mu g/l$.

People living in a region near a power plant are concerned because power plants are known to permit seepage of arsenic into groundwater. The residents decided to check the groundwater for safety periodically themselves.

One way of ensuring the safety of the water supply, each week or each month, is to collect five individual vials of water (one sample with size $n = 5$) and compute the mean of their arsenic levels. If the mean exceeds a certain pre-determined amount, then the water is declared contaminated. (The pre-determined amount is one that you yourself specify.)

Another way of ensuring safety is to collect five individual vials of water and determine the largest (max) of their arsenic levels. If the max exceeds a certain pre-determined amount, then the water is declared to be contaminated. (Again the pre-determined amount is one that you yourself specify. It need not be the same amount that you would have specified had you been using the sample mean.)

In either case, the five vials are samples from different parts of the region and are assumed to be independent of each other.

In this project you will use simulations to determine which of these two methods is better for deciding whether arsenic levels are at standard levels or at contaminated levels.

The Project:

You ultimately want to be able to take a single sample of five vials of groundwater from the region, and make a judgement about whether the arsenic levels in that region's ground water are at standard levels or contaminated levels. Using simulations (R/SAS/Matlab/SPSS...) to decide which of the two sample statistics - mean or max - will be better toll for distinguish a region where arsenic in the ground water is at standard levels from one that arsenic in the groundwater at contaminated levels.

The Report:

Write up a report of your results in no more than six pages, including all graphs (limited to two pages) and discussions. Attach your code as an appendix, no page limit for code. The report should include a brief introduction to the problem, a discussion how you studied the problem, a clear statement explaining how you decided which statistic was better, with appropriate model justification, and a short conclusion. You may include descriptive statistics and distribution comparison with appropriate graphs. When writing your report, be sure to present your graphs in such a way that they are maximally useful to the reader for your intended purpose.

Please send your report (.pdf) and code (.txt) to me via email. A well-written analysis report will count 2% (bonus points) in the final grades.

Deadline for the project is at 12:00 PM on Tuesday, May 3.