STAT 210
Applied Statistics and Data Analysis
Group Session 3

Joaquin Ortega

Fall 2020

Exercises

The Quality Control Engineer at a soft drinks plant wants to test a bottling machine that fills bottles with 500 ml of soda. From previous experience, the engineer knows that the normal distribution is a good approximation to the distribution of the content of the bottles.

The engineer draws a sample of size 10 from the machine's production and obtains the values

```
bottles <- c(494.04, 499.07, 497.03, 502.79, 495.57, 498.09, 500.40,
             491.80, 494.34, 498.65)
```

Part 1

1) Using this sample, estimate mean, variance, and standard deviation.
2) Which is the correct sampling distribution for the sample mean in this situation?
3) Using this distribution, find a two-sided confidence interval at the 98% level for the mean.

Part 2

4) Using this sample, the engineer wants to test whether the machine fills the bottles correctly. What hypothesis test should he carry out? What is the test statistic? Calculate the value for this statistic using the sample above.
5) The engineer always sets a level of confidence of 98% for these tests. Find a rejection region for the test at this level. Is the value of the test statistic inside or outside the rejection region? What is your conclusion?

Part 3

6) Carry this test out using a command in R and look at the p-value that you obtain. What is your conclusion?
7) Since the p-value is close to the confidence level that she set, the engineer decides to take a new sample of size 20 and obtains the following values

```
bottles2 <- c(497.24, 497.43, 500.64, 490.59, 496.24, 497.44, 501.69,
              489.98, 493.83, 495.60, 504.33, 495.11, 497.94, 495.03,
              490.75, 498.16, 491.87, 492.94, 494.12, 500.08)
```

Using this new sample, repeat the test you carried out in (6) and comment on the results you obtain.

8) Suppose now that the company is not worried about underfilling the bottles since management argues that consumers do not notice a small difference in the amount of soda in the bottle, but it is concerned about overfilling the bottles, as this would mean less profit. What would be a reasonable test of hypothesis in this context? Carry out this test and comment on the results.

Part 4

9) In the lectures, we worked out an example about the power of a hypothesis test for the normal distribution. Assuming that the variance for the population was known, we calculated and plotted the power function. However, when the variance of the population is not known, calculating the power is not so easy since the sampling distribution under the alternative distribution changes. Fortunately, there is a function in `R` for doing this. It is `power.t.test`. Look at the help for this function to get familiar with the required arguments.

   Using `power.t.test`, calculate the power of the test

$$H_0 : \mu = 500 \qquad \text{vs.} \qquad \mu = 502$$

   using a confidence level of 98% for the two sample sizes that we have considered before.

10) We can use the `power.t.test` function to determine the sample size to obtain a given power for a fixed alternative and significance level for the test. Suppose the engineer wants to detect when the machine is overfilling the bottles by 2.5 ml with a probability of at least 0.7. Calculate the sample size.

Part 5

11) All the calculations we have made rely on the fact that the engineer 'knows' by experience that the production of the machine follows a normal distribution. How would you verify this assumption for the two samples considered?

12) Use a non-parametric test as an alternative to the tests carried out in (6) and (7) and compare your results.