Activity #19b: Hypothesis testing (one-sample t-test) practice

Scenario: During World War II, information about Germany's war potential was essential to the Allied Forces in order to schedule the time of invasions and to carry out the allied strategic bombing program. Methods for estimating German production used during the early phases of the war proved to be inadequate. In order to obtain more reliable estimates of German war production, American and British experts began analyzing markings and serial numbers obtained from captured German equipment.

Allied forces eventually deciphered the serial numbers from captured German tanks and gave those serial numbers to statisticians. The statisticians believed the Germans, being Germans, had logically numbered their tanks in the order in which they were produced (1, 2, 3, and so on). It turns out the statisticians were correct.

Suppose the Allies captured five German tanks with the serial numbers: 16, 19, 24, 56, 61.

1) From our sample of 5 tanks, all we can say is that Germany produced at least 61 tanks. Suppose the Allied intelligence agencies estimated that Germany produced 350 tanks. Since our data possibly indicate fewer than 350 tanks were produced, we can develop the following null and alternative hypotheses:

H_0: N = 350 (we'll assume the estimates are correct...)
H_1: N < 350 (unless we get data that are highly unlikely under those estimates)

2) With these hypotheses and this scenario, briefly explain the consequences of making α and β errors. Which error, in your opinion, is more costly?

3) We now need to calculate a summary statistic (an estimate of the parameter we really want) from our data. We're interested in the total number of tanks produced by Germany. Since the tanks were given serial numbers in order (starting with 1, 2, 3, ...), we're interested in **N = the (unknown) maximum serial number in the population**. Thus, N = the total number of tanks produced by Germany.

The five serial numbers we collected only represent a sample from (a subset of) all the tank serial numbers. In fact, we can think of our data as representing **n=5 selections drawn without replacement** from all the tank serial numbers. Note that we do **not** have a random sample, as the serial numbers are dependent (no two tanks can have the same serial number).

To estimate N from n, we first need to compute the expected value of a maximum (which we'll call M). Since we know the Germans produced at least 61 tanks and we estimated that they produced 350 tanks, let's calculate the probability that M = k, where k = 61, 62, 63, ..., 350.

If we have N tanks in the population, there are $\binom{N}{n}$ ways of selecting n tanks without replacement. Thus, each of those ways has a 1/(N choose n) probability.

In order to have M = k, we must have one number equal k and choose the other n-1 numbers out of the remaining serial numbers (1, 2, 3, ..., k-1). There are $\binom{k-1}{n-1}$ ways to do this. Therefore, for all possible values k = n, n+1, n+2, ..., N,

$$P(M=k)=\frac{\binom{k-1}{n-1}}{\binom{N}{n}}=\frac{(k-1)!}{(k-n)!(n-1)!}\bullet\frac{(N-n)!n!}{N!}=n\bullet\frac{(k-1)!}{(k-n)!}\frac{(N-n)!}{N!}$$

We can then use our formula for expected values to calculate the expectation of M:

$$E[M]=\sum_{k=n}^{N}kP(M=k)=\sum_{k=n}^{N}kn\bullet\frac{(k-1)!}{(k-n)!}\frac{(N-n)!}{N!}=\sum_{k=n}^{N}n\bullet\frac{(k)!}{(k-n)!}\frac{(N-n)!}{N!}=n\bullet\frac{(N-n)!}{N!}\sum_{k=n}^{N}\frac{(k)!}{(k-n)!}$$

We can then use a trick and rearrange terms:

$$1=\sum_{j=n}^{N}P(M=j)=\sum_{j=n}^{N}n\bullet\frac{(j-1)!}{(j-n)!}\frac{(N-n)!}{N!}$$

finding that:

$$\sum_{j=n}^{N}\frac{(j-1)!}{(j-n)!}=\frac{(N)!}{n(N-n)!}$$

This holds for any N and any n ≤ N, so we can replace N with N+1 and replace n with n+1:

$$\sum_{j=n+1}^{N+1}\frac{(j-1)!}{(j-n-1)!}=\frac{(N+1)!}{(n+1)(N-n)!}$$

Changing the summation variable to k = j – 1, we obtain:

$$\sum_{k=n}^{N}\frac{k!}{(k-n)!}=\frac{(N+1)!}{(n+1)(N-n)!}$$

We can now substitute that into the equation at the top-right of this page:

$$E[M]=n\bullet\frac{(N-n)!}{N!}\bullet\frac{(N+1)!}{(n+1)(N-n)!}=n\bullet\frac{N+1}{n+1}$$

What did we just do? We just demonstrated that if we sample n values from a series of numbers (1, 2, 3, ...) that go up to an unknown maximum, the expected value for the maximum value in our sample is

$E[M]=n\bullet\frac{N+1}{n+1}$ where N = the unknown maximum value and n = the number of values in our sample.

So what is the expected maximum value of our sample of 5 tank serial numbers? We don't know. We know it would be:

$E[M]=5\bullet\frac{N+1}{6}$ but we don't know N (because it's the population maximum that we're interested in finding)

Recall that in hypothesis testing, we assume the null hypothesis is true. Assuming the null hypothesis is true (and Germany produced exactly 350 tanks), calculate the expected maximum value we should have obtained in our sample of 5 serial numbers:

4) You just calculated the maximum value we expected to obtain in our sample (assuming the null hypothesis were true). The actual maximum value we obtained in our sample was 61. Based on this, what would you conclude about the null hypothesis? What does this mean about the number of German tanks produced during World War II?

5) How likely were we to obtain a maximum value of 61 if the Germans actually had produced 350 tanks? In other words, if the null hypothesis were true, what was the likelihood of obtaining data as or more extreme than what we actually observed?

To calculate this, we find:

P(M ≤ 61 | true null hypothesis) = P(the maximum value in our sample of 5 values ≤ 61 | 350 is the overall maximum) =

$$\frac{61}{350} \bullet \frac{60}{350} \bullet \frac{59}{350} \bullet \frac{58}{350} \bullet \frac{57}{350} = 0.00014$$

This is the p-value in this study. What does this p-value represent in this study?

6) Suppose the Allied Intelligence agencies adjusted their methods and estimated only 80 German tanks. Using this new estimate, write out the new null and alternative hypotheses. Calculate the expected maximum value in a sample of 5 values (like you did in problem #3). Calculate a p-value (like was done in problem #5). Based on your calculations, what would you conclude about the null hypothesis and German tank production?

Scenario: In 2002, 483 Canadian citizens reported seeing a UFO (Canadian Press; 2/12/2003). Many of these UFO sightings were later explained (the objects were identified to be airplanes, weather balloons, reflections, etc.), while some of the objects remain unidentified. Some believe these UFOs are actually alien ships visiting Earth. Skeptics believe all the UFO sightings can be explained by natural phenomena. These skeptics also tend to believe that individuals reporting UFO sightings are of low intelligence (because more intelligent individuals are able to identify objects they see). Believers think that UFO observers are of high intelligence, since they are open to all possible explanations

To study the intelligence of UFO observers, the United States UFO Information and Research Center conducted a study. After receiving hundreds of responses to an ad on their website, they randomly sampled 25 Canadian citizens who had officially reported a UFO sighting and administered an IQ test to these subjects. Our task is to formally conduct a hypothesis test for this study and decide whether or not UFO observers have lower intelligence than the general public.

7) The researchers in this study sampled 25 Canadian citizens who responded to an online advertisement. Will this sampling procedure introduce any bias into this study? Briefly explain.

8) Write out hypotheses for this study. Remember that we're interested in determining if UFO observers have lower IQs than the general public. The null hypothesis should be that the average IQ of UFO observers is 100 (since that's the average for the entire population of adults).

H_0: $\mu = 100$ (we'll assume the UFO observers are like all other adults)

H_1: ____________________

9) Briefly explain the consequences of making α and β errors in this study.

10) We know IQ scores (if the null hypothesis is true) follow a normal distribution with a mean of 100 and a standard deviation of 16. Sketch this distribution below and label the mean and standard deviation.

11) Researchers randomly selected 25 UFO observers in this study. Assuming the null hypothesis is true, sketch the sampling distribution of sample mean IQs we would get if we could repeatedly sample 25 UFO observers. Label the mean and standard error of this distribution:

12) Suppose researchers set α=0.05. Find the critical value that cuts-off 5% of your sampling distribution (to the left, since our alternate hypothesis is $\mu < 100$). Label that critical value on the sampling distribution you sketched above. Shade in the critical region (everything more extreme than your critical value).

Remember that if we assume the null hypothesis is true, all we've done is sampled 25 people from a population with μ=100 and σ=16. Thus, if the null hypothesis is true, the average we calculate from this sample must come from the sampling distribution you sketched above.

If our (observed) sample average is near the center of our hypothesized sampling distribution, we will conclude that it likely came from this sampling distribution (and, therefore, the null hypothesis is true). If our observed mean comes from the critical region of our sampling distribution, then we'll conclude the null hypothesis is false.

13) The actual average IQ of our 25 UFO observers was found to be 97.2. Locate this point on your sampling distribution. What do you conclude about the null hypothesis? What do you conclude about the average IQ of UFO observers?

14) Remember that a p-value is the probability of observing something as or more extreme than what we actually observed (assuming the null hypothesis is true). Calculate the p-value for this study:

$$P\left(\bar{X} < 97.2\right)$$

Extra Credit Scenario: (Complete this and turn-in for extra credit)

Recall the following coal data from Activity 18b:

30.990	31.030	31.060	30.921	30.920	30.990	31.024	30.929	31.050	30.991	31.208
31.330	30.830	30.810	31.060	30.800	31.091	31.170	31.026	31.020	30.880	31.125

Source: A.M.H. van der Veen and A.J.M. Broos. Interlaboratory study programme "ILS coal characterization" – reported data. Technical report, NMi Van Swinden Laboratorium B.V., The Netherlands, 1996.

These 22 values represent ISO 1928 measurements. The mean is 31.012 and the standard deviation is 0.1294.

15) Assume the ISO 1928 measurements follow a normal distribution. We've decided that we will buy a shipment of coal if the ISO 1928 measurements are greater than 30.9. Because of this, we're going to to test the null hypothesis that the average measurement is equal to 30.9. Write out the alternative hypothesis, sketch the sampling distribution of sample means (we would get if we repeatedly sampled 22 values), and conduct a hypothesis test. Write out your conclusion.