Assignment #15 (part a): 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 $\alpha$ and $\beta$ 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 $\alpha$ and $\beta$ 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 $\alpha=0.05$. 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 $\mu=100$ and $\sigma=16$. 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.

Assignment #15 (part b): One-sample mean tests and confidence intervals

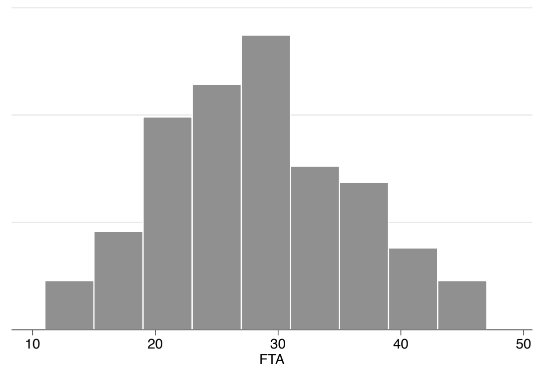
**Scenario: A**

During the 2010-11 NBA season, each team averaged just under 25 free throw attempts per game.

The Miami Heat averaged 27.902 free throw attempts per game that season:

Free Throw Attempts in each game

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 25 | 25 | 23 | 17 | 18 | 28 | 30 | 39 | 24 |
| 31 | 27 | 18 | 20 | 20 | 28 | 20 | 28 | 14 |
| 27 | 12 | 25 | 34 | 29 | 29 | 23 | 34 | |
| 34 | 38 | 21 | 37 | 26 | 23 | 19 | 22 | |
| 31 | 29 | 24 | 47 | 33 | 26 | 18 | 39 | |
| 24 | 29 | 38 | 19 | 22 | 37 | 30 | 30 | |
| 29 | 35 | 41 | 37 | 32 | 11 | 24 | 32 | |
| 47 | 42 | 18 | 32 | 28 | 27 | 28 | 21 | |
| 41 | 21 | 32 | 16 | 36 | 45 | 26 | 35 | |
| 36 | 21 | 25 | 30 | 19 | 22 | 29 | 26 | |




Summary: $n = 82$
average = 27.902
std. dev. = 7.881

We're going to conduct a test to determine if the Miami Heat attempted a significantly higher number of free throws than the rest of the league.

1) State the null and alternate hypotheses. Express the consequences of an alpha error in this study.

Null hypothesis: ______________________________

Alternate hypothesis: ______________________________

Alpha error consequence: ____________________________________________________________

2) Why can't we simply look at the sample average, see it's higher than 25, and conclude the Miami Heat attempted more free throws per game than other teams?

3) What would be the more appropriate test in this situation: a z-test or a t-test? Explain. Conduct this test on your calculator and record the p-value.

4) Consider the assumptions necessary for us to conduct a t-test. Are these assumptions reasonably satisfied in this situation?

5) We're interested in μ (the population average number of free throws attempted by the Miami Heat). We'll never know the value of this parameter, but our best estimate is 27.902.

Suppose we could go back in time and play this 2010-11 season again. We'd expect the average free throws attempted by the Miami Heat would be different number (it wouldn't be exactly 27.902). Now suppose we go back in time again and again, each time recording the average number of free throws attempted by the Miami Heat. What would the sampling distribution of those averages look like? Sketch it below and label its center and spread.

6) Let's use $\alpha = 0.05$ in this study. Find the critical value of the t-statistic and shade-in the critical region of the distribution you sketched above. Convert this t-statistic to an average number of free-throws attempted.

Now, locate your observed average (27.902) on this distribution. From this, what's your decision regarding the null hypothesis? Do you reject or retain the null hypothesis? What's your conclusion regarding the number of free throws attempted by the Miami Heat? Estimate the p-value.

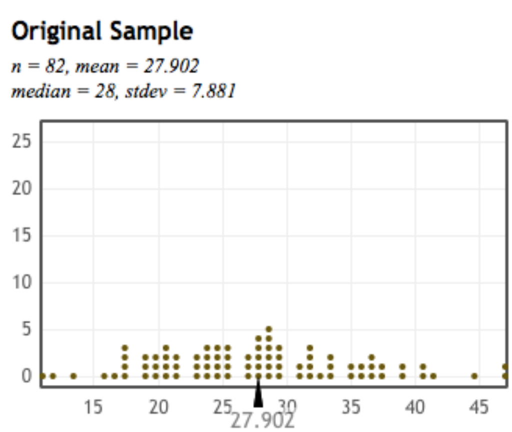
7) Suppose the true population average number of free throws attempted by the Miami Heat was 26. Estimate the power of this study.

To do this, it might be helpful to sketch (once again) the sampling distribution you sketched in question #5. Then, just to the right of that curve, sketch the alternate (true) distribution centered at 26. Thinking about what power represents, shade-in the power of this study and then calculate that power.

8) Let's construct a 90% confidence interval for the true population average number of free throws attempted by the Miami Heat. Before you construct this interval, can you predict whether it will contain 25?

9) If you completed the previous assignment (18b), you received solutions explaining a *bootstrap method* for constructing confidence intervals. The bootstrap method doesn't require a normality assumption like our standard (parametric) confidence interval method.

I'll demonstrate this bootstrap method using the following website:
[http://lock5stat.com/statkey/bootstrap_1_quant/bootstrap_1_quant.html](http://lock5stat.com/statkey/bootstrap_1_quant/bootstrap_1_quant.html)




First, I enter the data (our 82 free throw attempt values from page 1). You can see the data and summary statistics to the right.

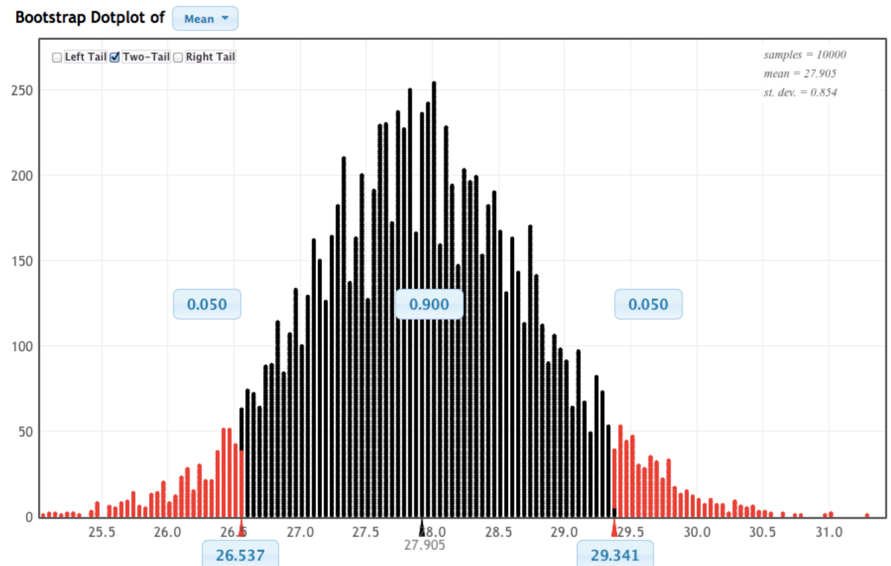
We'll now have the computer pretend these 82 values are the entire population. We'll instruct the computer to randomly select 82 values <u>with replacement</u> from this dataset and calculate and average.

We'll then have the computer repeat this process 10,000 times to give us an idea of the averages we could have obtained from our population.

On the next page, I've pasted the distribution of those possible averages.

10) Keep in mind what this graph represents. It shows possible averages we could have obtained from sample sizes of n = 82. It's centered at our observed sample average, since that's our best estimate of the population average.

To estimate our 90% confidence interval, I simply have the computer find the top and bottom 5% of our possible averages. As you can see below, the computer estimates this confidence interval to be (26.537, 29.341). How does this compare to the confidence interval you computed in question #8?




Remember our original question in this scenario: we're interested in determining if the Miami Heat attempted more free throws than the average NBA team (with 25 free throw attempts). We can use our bootstrap distribution to estimate just how unlikely it would be for the Miami Heat to average 25 (or fewer) free throw attempts.

Just by looking at the bootstrap distribution shown above, you can see it is extremely unlikely for the Miami Heat to average 25 or fewer free throw attempts per game.

11) Let's finish this example by running a sign test. Looking at sample of 82 free throw attempts (our original data on the first page), I see:

29 observations were less than 25 (so I will call these "–")
4 observations were exactly 25 (so I will ignore these)
49 observations were more than 25 (so I will call these "+")

So, in summary, out of 78 observations, we had 29 –'s and 49 +'s. Under a null hypothesis, we'd expect an equal number of +'s and –'s. Use the binomial distribution to estimate the likelihood of observing 29 or fewer –'s if, in fact, we'd expect an equal number.

| | |
|---|---|
| **Scenario: B** | According to the International Diabetes Research Foundation, an individual has diabetes if their blood glucose concentration is at or above 200 mg/dl. Over the course of 3 months, you sample your blood 25 times and find the average glucose concentration is 201.38 with a standard deviation of 7.348 mg/dl. Test the hypothesis that you have diabetes based on these 25 measurements. |

12) State the null and alternate hypotheses. Express the consequences of an alpha error in this study.

Null hypothesis: ______________________________

Alternate hypothesis: ______________________________

Alpha error consequence: ____________________________________________________________

13) Why can't we simply look at the sample average, see it's higher than 200, and conclude we have diabetes?

14) What would be the more appropriate test in this situation: a z-test or a t-test? Explain. Conduct this test on your calculator and record the p-value.

15) Consider the assumptions necessary for us to conduct a t-test. Are these assumptions reasonably satisfied in this situation?

16) We're interested in μ (the population average blood glucose concentration). We'll never know the value of this parameter, but our best estimate is 201.38.

Suppose we could go back in time and get another sample of 25 measurements. We'd expect the average blood glucose level would be different number (it wouldn't be exactly 201.38). Now suppose we go back in time again and again, each time recording the average blood glucose level. What would the sampling distribution of those averages look like? Sketch it below and label its center and spread.

17) Let's use $\alpha = 0.05$ in this study. Find the critical value of the t-statistic and shade-in the critical region of the distribution you sketched above. Convert this t-statistic to an average blood glucose concentration.

Now, locate your observed average (201.38) on this distribution. From this, what's your decision regarding the null hypothesis? Do you reject or retain the null hypothesis? What's your conclusion regarding whether you have diabetes? Estimate the p-value.

18) Suppose your true population average blood glucose concentration was 202. Estimate the power of this study.

To do this, it might be helpful to sketch (once again) the sampling distribution you sketched in question #16. Then, just to the right of that curve, sketch the alternate (true) distribution centered at 202. Thinking about what power represents, shade-in the power of this study and then calculate that power.

19) Let's construct a 90% confidence interval for the true population blood glucose concentration. Before you construct this interval, can you predict whether it will contain 200?

**Scenario C:** In a discussion of SAT scores, someone comments:

*Because only a minority of high school students take the SAT, the scores overestimate the ability of typical high school students. If all students took the SAT, the mean score would be no more than 450.*

Suppose you give the SAT to 50 randomly sampled students in a high school. This sample may have included students who were or were not planning to take the SAT. From these 50 students, you find an average ACT score of 475 with a standard deviation of 97.

Conduct a t-test to determine if the average SAT score of all high school students is greater than 450.

a) Write your null and alternate hypotheses
b) Explain the consequences of an alpha error and choose the alpha level for your study
c) Briefly explain if you are concerned about any of the assumptions needed to conduct a t-test.
d) Sketch the sampling distribution, shade-in the critical region, and locate your observed data
e) Estimate the p-value and state your conclusion(s).
f) Assuming the true average SAT score for all high school students is 470, estimate power.

**Scenario D:** Do middle-aged male executives have high average blood pressure than the general population? The National Center for Health Statistics reports the average systolic blood pressure for males 35-44 years of age is 128. You get a random sample of medical records from 72 middle-aged male executives and calculate an average systolic blood pressure of 130 with a standard deviation of 15.

Conduct a t-test to determine if male executives have higher blood pressure than the general population.

a) Write your null and alternate hypotheses
b) Explain the consequences of an alpha error and choose the alpha level for your study
c) Briefly explain if you are concerned about any of the assumptions needed to conduct a t-test.
d) Sketch the sampling distribution, shade-in the critical region, and locate your observed data
e) Estimate the p-value and state your conclusion(s).
f) Assuming the true average SAT score for all high school students is 470, estimate power.

**Scenario E:** Your friend declares himself to be the world's greatest rock-paper-scissors player. You decide to challenge him for that title, so you play 510 times. Of those 510 trials, you lost 183 times (and either tied or won the other 327 times). Construct a 95% confidence interval to determine if your friend is better than average.

**Scenario F:** Evidence suggests that the drug Lipitor reduces total cholesterol. 1.9% of patients who used previously available cholesterol medications reported flu-like symptoms when taking the drug. In clinical trials, 11 out of 863 patients taking Lipitor complained of flu-like symptoms. Construct a 95% confidence interval to determine if Lipitor reduces the prevalence of flu-like symptoms.