HYPOTHESIS TESTING (HT): GENERALIZED STEPS

rev. Oct. 2022

Hypothesis testing questions the ____

Progression: Suspicion that parameter is incorrect …. collect a sample & gather statistics …

determine probability of getting that sample if the parameter **is** correct …decide about the parameter

STEPS: Short List

Formulate Problem Identify the

1. population,
2. sample,
3. parameter, and
4. state the null (H_o) and alternative hypotheses (H_a).

Review Conditions Either (i) variable of parent population is normally distributed or (ii) $n \geq 30$.

Execute Calculation Find the

1. test statistic (t),
2. degrees of freedom (df),
3. p-value, and
4. make decision using α (in terms of H_o)

Draw Conclusion As with Cis, restate parameter in Formulate, but also state whether or not there is statistically significant evidence to reject H_o (in terms of H_a).

STEPS: Explanations

F The new part of Formulate Problem is to state the null and alternative hypotheses. Here is what they are:

- H_o: parameter = null value *Remember:* *the null ALWAYS carries the "=" sign**

  H_a: parameter $\boxed{}$ null value *the alternative uses anything but the "=" sign*

  >, <, ≠ $\boxed{\text{A number}}$

  ** In practice, H_o does not always have a strict = sign; can be ≤ or ≥. However, for simplicity, I will only use =.*

- Parameter can be μ, μ_d, $\mu_1 - \mu_2$, or β_1.

- Example: A consumer advocacy group suspected that a particular brand of soda was underfilling their bottles. 64 bottles of the soda marked "20 oz" from a particular factory were tested for the real number of ounces. It is known that number of ounces in these bottles is normally distributed. The result was an average of 18.2 ounces with a standard deviation of 5.6 ounces. Is there sufficient evidence to indicate that the soda companies are underfilling their bottles?

  The null and alternative hypotheses are:

  H_o: $\mu = 20$

  H_a: $\mu < 20$

E Almost everything in Execute Calculation is new.

<u>Test statistic</u>: Formulas are on the formula sheet and they vary depending on which parameter you are testing from the list in the F section, above. Here is one of the test statistic formulas for HT of μ:

Using the example on the previous page about the bottles of soda,

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$ with df = n-1

n = 64   μ = 20 $\bar{x}$ = 18.2   s = 5.6

The test statistic is $t = \frac{18.2 - 20}{\frac{5.6}{\sqrt{64}}} = \frac{-1.8}{0.7} = -2.5714$ with df = 63

<u>p-value</u>: The *p*-value is merely the probability of getting the test statistic or anything more extreme. The process is exactly like getting probabilities for the normal distribution except now we use the t-distribution. (The *p* in *p*-value stands for *probability*.)

The p-value is the heart of HT!!

(i) Draw a curve of the sampling distribution,
(ii) shade the area that corresponds with the probability that you want,
(iii) jot down the probability statements,
(iv) then put the values in the calculator and get the answer. Be sensitive to one-tailed or two-tailed.

<u>Make decision</u>: Need to decide about the parameter; either to REJECT H_O or to FAIL TO REJECT H_O.

The idea is that you're using the p-value (which comes from the statistics) as "evidence" to determine if the parameter is accurate or not. If the parameter is found to be inaccurate, then it is rejected, or rather, the hypothesis with the claim of the parameter's value is rejected. If the parameter is likely to be correct, then we say that the claim of the parameter's value is not rejected. We never say "accept H_o" since another sample may bring stronger evidence against the null.

- If p-value is between 0 and 0.01, (really small), REJECT H_O (*think: p-value < 1%, reject H_o*)
- If p-value is greater than 0.1, (really big), FAIL TO REJECT H_O (*think: p-value > 10%, fail to reject H_o*)
- If p-value is between 0.01 and 0.1, need level of significance, α, to help decide. If no α is given, use 0.05.
  - If the *p*-value is less than α, REJECT H_O
  - If the *p*-value is greater than α, FAIL TO REJECT H_O

Your decision should be of the form, "Since [insert p-value] is [< or >] [stated α or 0.05], then [state conclusion]."

Example of 64 soda bottles that should be 20 ounces: draw curve with shading, then go to tcdf on the TI calculator to get p-value….

p- value = P($\bar{x}$ < 18.2) = P(t < -2.57) = 0.00625   (TI: tcdf; df = 63)

<u>Decision</u>: Since *p* – value of 0.00625 < 0.01, REJECT H_O.

D The draw Conclusion step is very similar to CIs, but we need to include the decision.

- Rejecting H_o means that the parameter is probably not correct, and that decision requires lots of statistical evidence. That must be stated as "statistically significant evidence" or "sufficient evidence" that the alternative hypothesis is true.

- Failing to reject H_o means that the parameter is probably OK, and though we brought statistical evidence against the null, it was not enough to decide that it may be wrong. In that case, the lack of evidence must be stated as "do not have statistically significant evidence" or "insufficient evidence" that the alternative hypothesis is true.

- Notice that regardless of the decision, both conclusions are expressed in terms of H_a.

This is the template:

"There is {sufficient, insufficient} evidence to indicate that the true {state parameter} {state context} is {restate symbol in H_a} {restate μ}."

Let's finish the 20-oz soda bottle example:

There is sufficient evidence to indicate that the mean amount of soda being poured into 20 oz bottles at the factory is less than 20 oz.

Here is the whole example without interruption:

64 bottles of soda marked "20 oz" from a particular factory were tested for the real number of ounces. It is known that number of ounces in these bottles is normally distributed. The result was an average of 18.2 ounces with a standard deviation of 5.6 ounces. Is there sufficient evidence to indicate that the soda companies are underfilling their bottles? Test at $\alpha = 0.01$.

F

<u>Population</u>: all of the bottles of soda from this factory
<u>Sample</u>: 64 randomly selected 20 oz. bottles from this factory
<u>Parameter</u>: True mean number of ounces in bottles marked 20 ounces

H_o: $\mu = 20$
H_a: $\mu < 20$

R

Since the number of ounces in the population of 20-oz. bottles is known to be normally distributed, the condition is met.

E

$$t = \frac{18.2 - 20}{\frac{5.6}{\sqrt{64}}} = \frac{-1.8}{0.7} = -2.5714 \text{ with df} = 63$$

Draw curve:

p- value = $P(\bar{x} < 18.2) = P(t < -2.57) = 0.00625$ (TI: tcdf; df = 63)

Decision: Since *p* – value of 0.00625 < 0.01, REJECT H_O.

D

There is sufficient evidence to indicate that the mean amount of soda being poured into 20 oz bottles at the factory is less than 20 oz.

“P” IS FOR “PROBABILITY,” A DISCUSSION ABOUT P-VALUES

“P-value” is the probability of getting the sample results or anything more extreme. It is calculated assuming that the null hypothesis is true. Start by stating the four facts that you need:

(i) the null hypothesis, (ii) the statistic that results from the study, (iii) the sample size, and (iv) the p-value.

Ex: In order to interpret the p-value from the soda bottle example, start by writing down the four facts ::

(i) (ii) (iii) (iv)

Here is how to put the four facts in words:

> If the null hypothesis is true, that is, if the true mean fill in the bottles of soda from this factory really is 20 oz., then the probability of obtaining a sample mean of 18.2 ounces from 64 randomly sampled bottles, or anything more extreme, **is 0.63%.**

Our “magic level of significance” is 0.01. So if you think about it, that means that if something is so unlikely to happen that its probability of occurrence is less than 1%, then the null hypothesis needs to be rejected. Smaller p-values are regarded as stronger evidence. In fact, **THE SMALLER THE P-VALUE, THE MORE LIKELY WILL BE REJECTION OF THE NULL**.

Rejecting the null hypothesis is a pretty big deal. You really need strong evidence to do it. It is such a big deal that we call it “statistically significant.” That’s why rejection is determined based on “Level of **Significance**,” or α.