The Basics of a Hypothesis Test

Dr Tom Ilvento
Department of Food and Resource Economics
University of Delaware

Overview

- Alternative way to make inferences from a sample to the Population is via a **Hypothesis Test**
- A hypothesis test is based upon
  - A **point estimate** from a sample
  - Knowledge of a **sampling distribution**
  - A **Null hypothesis**
  - A probability level of being wrong that we are comfortable with, called **α**
- All via a **rare event approach** – we want to see how rare it is that we drew a random sample and got our estimate if the true value was really that specified in the Null Hypothesis

Newspaper Print Problem (BLACKNESS.xls)

- The production department of the Springfield Herald newspaper has embarked on a quality improvement effort and the first project is the blackness of the newspaper print.
- Blackness is measured on a standard scale in which the target value is 1.0.
- **A desirable level is that the average is at least .97.**
- A random **sample of 50** newspapers have been selected
- We want to determine i**f the average blackness is different from .97, use alpha = .05**

We can use software to help us analyze the data

- Most often, we use software to give us basic statistical data for our analyses
- I want you to start your analysis with basic descriptive statistics
- **What do you see?**
  - The distribution looks curiously bi-modal
  - The mean is .959
  - Median is .985
  - Standard deviation is .156
  - CV = .156/.959*100 = 16.3

Stem-and-Leaf Display for times 10
Stem unit: 1

```
 6|2 8
 7|0 2 3 6 8
 8|0 0 1 1 1 2 5 6 7 7 8 9
 9|0 3 3 5 7 8 9
10|0 1 1 1 2 2 3 4 5 5 6 9 9
11|1 1 1 1 2 3 4 6
12|2 2 9
```

95% C.I. = .959 ± .044
= .915 to 1.003

Hypothesis Test for the Blackness of Newspaper Print

- If the sampling distribution for n=50 has a mean level of .970, **μ or the Hypothesized value**
- I could find out how rare an event it was to take a sample of 50 and get a mean value of .959
- I can test to see how my sample compares to the hypothesized population
  - different from **Two-tailed**
  - Greater than **One-Tailed Upper**
  - or Less than **One-tailed Lower**

Hypothesis Test

- Ho: $\mu = .970$
- Ha: $\mu \neq .970$ I will check if it is different
- I calculate a z or t-score to see how far away it is from the hypothesized value
  - $t^* = (.959 - .970)/.022$ **$SE = .156/(50)^{.5} = .022$**
  - $t^* = -.011/.022$
  - $t^* = -.500$
- Our sample estimate is .5 standard deviations below the mean in relation to the sampling distribution.
- This is not so unusual.

What is a Hypothesis Test?

- We are going to use a rare-event approach to make an inference from our sample to a population
- We define two hypotheses
  - **Null hypothesis** – the hypothesis that will be accepted unless we have convincing evidence to the contrary - **we seek to reject the Null Hypothesis**
  - **Alternative Hypothesis** – aka as the Research Hypothesis. We look to see if the data provide convincing evidence of its truth. **We hope for the alternative, but we are reluctant to accept it outright.**

Here is our Strategy: Null and Alternative Hypothesis

- The **Null Hypothesis** is based on expectations of no change, nothing happening, no difference, the same old same old
  - In many ways it is a straw man (or person) and in contrast to the rare event
  - In most cases **we want to reject the null hypothesis**
- The **Alternative Hypothesis** is more in line with our true expectations for the experiment or research
  - The sample value is not equal to the hypothesized value, or
  - It is greater than the hypothesized value, or
  - It is less than the hypothesized value

The logic of a Hypothesis Test

- We **assume** our **sample estimate** comes from a population with a **parameter** as stated under the **null hypothesis**
- We can **compare** our **sample estimate** to this **population parameter** to see **how likely** it is that our **sample** comes from a **sampling distribution from this population**
- The comparison is made via a **Test Statistic**

The logic of a Hypothesis Test

- To do this we calculate a **Test Statistic**
- This is a z-score (or a t-score) based on:
  - Ho: μ = *some value* or *P* = *some value*
  - The sample estimate of the standard deviation, **s**
  - The Standard Error of the sampling distribution for our estimator, **s/(n)$^{.5}$**

$$z^* = \frac{(\bar{x} - \mu)}{\sigma_{\bar{X}}} \qquad z^* = \frac{(p - P)}{\sigma_P} \qquad t^* = \frac{(\bar{x} - \mu)}{s_{\bar{X}}}$$

The components of a Hypothesis Test

Ho:	**Ho: μ =.970**
Ha:	**Ha: μ ≠ .970 2-tailed**
Assumptions	**n= 50, σ unknown, use t**
Test Statistic	**t* = (.959 – .970)/.022**
Rejection Region	**α = .05, .05/2, 49 d.f., t = ± 2.010**
Calculation:	**t* = -.50**
Conclusion:	**t* > $t_{.05/2, 49 df}$**
	-.50 > -2.010
	Cannot Reject Ho: =.970

The components of a Hypothesis Test

- **Set up the Null Hypothesis**
  - Ho: μ = ???
  - **Ho: μ = .970**
  - some use ≥ or ≤ with a one-tailed test
- **Set up the Alternative Hypothesis**
  - It takes up one of three forms
  - Ha: ≠ .970 Two-tailed
  - Ha: > .970 One-tailed, upper tail
  - Ha: < .970 One-tailed, lower tail
  - **Ha: ≠ .970 Two-tailed**

The Alternative Hypothesis

- A **one-tailed test** of hypothesis is one in which the alternative hypothesis is directional, and includes either the **<** symbol or the **>** symbol.
- A **two-tailed test** of hypothesis is one in which the alternative hypothesis does not specify a particular direction; it is “different.” It will be written with the **≠** symbol.
- **We gain by specifying a one-tailed test - more knowledge is usually better in statistics**

The Assumptions of the Test

- **Proportion**
  - Is the sample size large?
    - Yes — If Yes, use the normal approximation to the binomial

$$z^* = \frac{(p - P)}{\sigma_P}$$

    - No — If No, use the binomial distribution

The Assumptions of the Test

- **Mean**
  - Is sigma known?
    - Yes — If Yes, use the known σ for the standard error and use z
    - No — If No, use the sample estimate, s, for the standard error, and use t
  - Is sample size small?
    - Yes — If n is small, especially if σ is unknown, we have to worry about whether the population is distributed as normal
    - No

The Test Statistic

$$t^* = \frac{(\bar{x} - \mu)}{s_{\bar{X}}}$$

- Let **t*** equal the calculated **test statistic**
- The value for μ will come from the Null Hypothesis
- If we don’t know σ we use the sample estimate of s to calculate the standard error as **($s/n^{.5}$)**
- This t* represents a “z-score” of our sample compared to the sampling distribution of the mean, **if the null hypothesis is true**
- We want to see how far away our sample estimate is from the null hypothesis population mean

$$t^* = \frac{(.959 - .970)}{.156 / \sqrt{50}} = -.50$$

The Critical Value and the Rejection Region

- The rejection region is the set of possible values of the test statistic for which the null hypothesis will be rejected.
- So if the calculated test statistic falls within the rejection region, we reject Ho
- Alpha (α) is the probability level at which you are willing to be wrong when rejecting Ho.
- The value at the boundary of the rejection region is called the **Critical Value**
- **We establish the Critical Value linked to α and the type of test (1 or 2-tailed) and compare our Test Statistic to this value**

The Critical Value and the Rejection Region

- We look to see where t* falls in relation to the t-distribution
- If it lies in the Rejection Region, in one of the tails, it is possible that it came from that hypothesized distribution -- but not very likely
- Because it is unlikely, I am willing to reject the Null Hypothesis
- We need to set the level of alpha and the type of test (1 or 2-tailed) **a priori**

More on the Rejection Region

- I want to pick an alpha value far out in the tail of the sampling distribution
- So if I find a difference, I can be reasonably sure that my sample is not part of the sampling distribution specified under the null hypothesis
- α represents the probability that my sample mean actually comes from the hypothesized sampling distribution, **but I'm going to say it doesn't**
- This is called a **Type I Error**
- **The probability of committing a Type I Error is the probability of rejecting Ho when Ho is true**

More on the Rejection Region

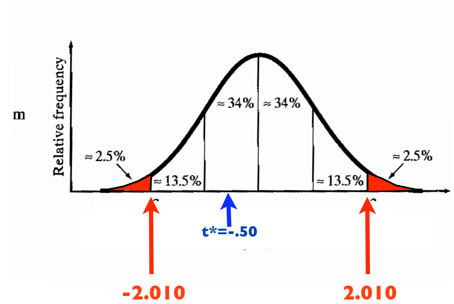
- For this problem, α=.05
- Next I want to find a t-value that will correspond with an overall α =.05, with n-1 degrees of freedom
- This is a two-tailed test as specified in the alternative hypothesis
  - Ha: $\mu \neq .970$
  - So if α = .05, we need to split this probability level in the upper and lower tail of the sampling distribution, much like we do for a confidence interval
- If it were a one-tailed test, we would put all of alpha into one tail

Rejection Region

- So I look for a t-score that corresponds to an overall probability of .05
- In the t-table this is a value relating to
  - α = .05 for two-tails, or .025 in one tail
  - d.f. = 49
  - The t-value from the table is **2.010**
- The values of **-2.010 and 2.010** are the **critical values** which **marks** the **rejection region**
  - So our Rejection Region is:
  - Any test statistic value < -2.010 or > 2.010

Look at it in pictures

- Our **test statistic is t* = -.50**
- The **Critical Values** that mark the rejection region are **-2.010** and **2.010**
- If our test statistic is in the **Rejection Region**, we would consider it far enough away from the Null Value
- To reject the Null Hypothesis
- Which makes the Alternative Hypothesis more plausible
- Otherwise, we **Fail to Reject**



My test statistic does not fall in the rejection region, therefore I cannot reject the Null Hypothesis

So what do I conclude?

- I don't have any evidence that the average blackness of the print is different from .97.
- It was not so unusual to draw a random sample of 50 papers and get an average of .959 given the true value is .97
- So I can't conclude I have a problem from this sample…
- **….but I could be wrong!!!**

The components of a Hypothesis Test

Ho:	**Ho: μ =.970**
Ha:	**Ha: $\mu \neq$ .970 2-tailed**
Assumptions	**n= 50, σ unknown, use t**
Test Statistic	**t* = (.959 – .970)/.022**
Rejection Region	**α = .05, .05/2, 49 d.f., t = ± 2.010**
Calculation:	**t* = -.50**
Conclusion:	**$t^* > t_{.05/2,\ 49\ df}$**
	-.50 > -2.010
	Cannot Reject Ho: =.970

Hypothesis Test from JMP

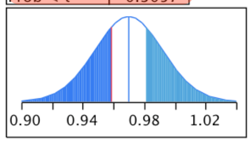
- Here is the output from JMP
  - You would be asked to specify a Hypothesized Value, Ho
  - An alpha level, .05
  - Then it gives the results of a one or two-tailed test
- **What if we multiplied blackness by 100?**

Blackness

Test Mean=value

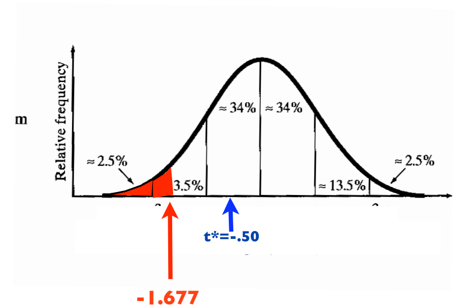
Hypothesized Value	0.97
Actual Estimate	0.95872
df	49
Std Dev	0.15597

	t Test
Test Statistic	-0.5114
Prob > |t|	0.6114
Prob > t	0.6943
Prob < t	0.3057



What if this were a one-tailed test?

- Let's say that the Alternative Hypothesis was:
  - Ha: $\mu < .97$
  - $\alpha = .05$
- The test statistic remains the same **$t^* = -.50$**
- Now, all of alpha would fall in just one tail
  - $t_{.05,\ 49\ df} = -1.677$
- This shifted the Rejection Region to the right and made it easier to reject the Null Hypothesis



My test statistic does not fall in the rejection region, therefore I cannot reject the Null Hypothesis

Summary

- We established the basis for a Hypothesis Test
- We base it on a **Rare Event Approach**
  - We ask how rare an event it was to draw a sample of size **n**
  - and observe our sample estimate - **mean or *p***
  - If the the true value were really that specified in the Null Hypothesis
- To do this we need
  - A **Null** and **Alternative Hypothesis**
  - A Test Statistic, **z^*** or **t^***
  - A criterion for decision making - how far away is enough to reject the Null Hypothesis - based on **α**