Hypothesis Testing of Means

Dr. Tom Ilvento
FREC 408

Water Quality Problem

- Suppose I am concerned about the quality of drinking water for people who use wells in a particular geographic area
- I will test for nitrogen, as Nitrate+Nitrite
- The U.S. EPA sets a MCL of 10 mg/l of Nitrate/Nitrite (MCL=Maximum contaminant level)
  - Below the threshold is considered safe

Water Quality Example

- A priori, I decide to test to see if my sample will result in a mean level that is above or below the EPA threshold
- Once again I decide to take a sample to save time and money
- **I want to see if I have evidence to believe my sample is different from a population with a mean value of 10 mg/l**

Water Quality Example

- My sample
  - n = 50
  - Mean = 7 mg/l
  - s = 3 mg/l
- The sample provides an estimate, but I know other possible samples would have yielded slightly different mean levels, so I think of my sample as part of a sampling distribution
  - Standard error = $3/(50)^{.5}$ = .424

Water Quality Example

- If the sampling distribution of n=50 has a mean level of 10 mg/l, I could find out how rare an event it was to take a sample of 50 wells and get a mean value of 7 mg/l
  - I can test to see if my sample is different from the hypothesized population
  - Greater than
  - Less than

Water quality example

- Ho: μ=10
- Ha: $\mu \neq 10$ I will check if it is different
- I calculate a **z-score** to see how far away if is from the hypothesized value
  - $z^* = (7 – 10)/.424$
  - $z^* = -3/.424$
  - $z^* = -7.08$
- Our sample estimate is more than 7 standard deviations away from the mean in relation to the sampling distribution!

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 (Def8.3 p391)
  - **Alternative Hypothesis** – aka as the Research Hypothesis. We look to see if the data provide convincing evidence of its truth. (Def8.2 p391)

Here's our strategy

- **We set up a null hypothesis**
  - 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**

Hypothesis strategy

- Then we try to state an **Alternative Hypothesis** which is more in line with our true expectations for the experiment or research
  - The sample value is not equal to the hypothesized value
  - It is greater than the hypothesized value
  - It is less than the hypothesized value

How a hypothesis test works

- If we **assume our sample estimate** comes **from a population with a parameter** as stated in 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

How It Works

- To do this we calculate a ***TEST STATISTIC*** (Def8.9 p397)
  - z-score (or a t-score) based on:
  - H_0: μ = *some value*
  - The sample estimate of the standard deviation
  - The Standard Error of the sampling distribution for our estimator

$$Z* = \frac{(\overline{x} - \mu)}{\sigma_{\overline{X}}}$$

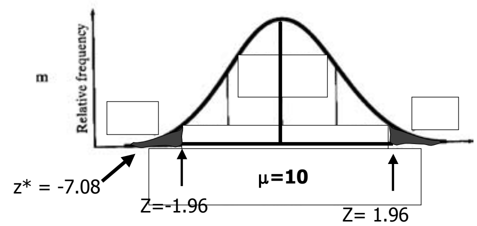
Basic Elements of a Hypothesis Test

- **H_0:**
- **H_a:**
- **Assumptions**
- **Test Statistic**
- **Rejection Region**
- **Calculation:**
- **Conclusion:**

Water Quality Hypothesis Test

- **H_0:** Ho: μ=10
- **H_a:** Ha: $\mu \neq 10$ 2-tailed
- **Assumptions** n= 50, large sample
- **Test Statistic** z* = (7 – 10)/.424
- **Rejection Region** α = .05, .025 in each tail, so I use z=1.96
- **Calculation:** z* = -7.08
- **Conclusion:** Reject Ho: μ=10

Here's how it looks in pictures



Example problem

- Humerus bones from the same species have the same length to width ratio, so they are often used as a means to identify bones by archeologists
- It is known that a **Species A** exhibits a **mean ratio of 8.5**
- Suppose 41 fossils of humerus bones were unearthed in East Africa
- **Test whether the population mean ratio from this sample <u>differs</u> from 8.5.**
- Use α = .01

Example Problem

- The length to width ratio was calculated for the sample and resulted in the following univariate statistics
  - n= 41
  - Mean = 9.26
  - s= 1.20
  - Min value = 6.23
  - Max value=12.00

Bone data

Stem and Leaf Plot of Bone Data

```
 6  | 23
 6* | 66 85
 7  |
 7* |
 8  | 07 17 30 37 38 48
 8* | 52 71 80 86 87 89 91 93
 9  | 07 17 17 20 29 35 38 39 41
 9* | 57 59 84 89 93 94 98
10  | 02 33 39 48
10* | 73
11  |
11* | 67 77
12  | 00
```

Stems are whole numbers, leafs are 2 decimal places;
* represents a breakpoint at .5

Descriptive Statistics from Excel

Descriptives	
Mean	9.26
Standard Error	0.19
Median	9.20
Mode	9.17
Standard Deviation	1.20
Sample Variance	1.45
Kurtosis	1.07
Skewness	-0.08
Range	5.77
Minimum	6.23
Maximum	12.00
Sum	379.56
Count	41.00
Confidence Level(99.0	0.51

What do we do?

- **Set up the Null Hypothesis**
  - H_0: μ = ???
  - H_0: $\mu = 8.5$

Alternative Hypothesis

- **Set up the Alternative Hypothesis**
- It takes up one of three forms
  - H_a: $\mu > 8.5$ One-tailed, upper tail
  - H_a: $\mu < 8.5$ One-tailed, lower tail
  - H_a: $\mu \neq 8.5$ Two-tailed

Alternative Hypothesis

- **Set up 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. (Def8.4 p392)
  - 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. (Def8.5 p392)

Alternative Hypothesis for Humerus Bones

- What do we want?
  - H_a: $\mu \neq 8.5$ Two-tailed test

The Assumptions of the Test

- You want to ask the following questions:
- Is the variable under study a proportion or a mean?
  - If it is a proportion:
    - Is the sample size sufficiently large so that I can use the normal approximation to the binomial?

The assumptions of the test

- Is this a large sample or small sample test?
  - Large sample – use z-score
  - Small sample – ask more questions!
- If small sample:
  - Is the variable distributed normally? If yes, use the t-distribution
  - If no, you can't use this test

Next the Test Statistic

- Let **z*** equal the calculated test statistic
  - **$z^* = (\bar{x} - \mu)/\sigma_{\bar{x}}$**
  - The value for μ will come from the Null Hypothesis
  - If we don't know σ we use the sample estimate of σ (s) to calculate the standard error
  - **$z^* = (\bar{x} - \mu)/(s/\sqrt{n})$** (p397)

$$z^* = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}$$

The test statistic

- This z* represents a z-score of our sample compared to the sampling distribution of the mean, if **the null hypothesis was true**
- We want to see how far away our sample estimate is from the null hypothesis population mean
- **$Z^* = (9.26-8.5)/(1.20/\sqrt{41})$**

Specify a rejection region

- The rejection region is tied to a level of probability called α. The **rejection region** is the set of possible values of the test statistic for which the null hypothesis will be rejected.
  (Def8.10 p397)
- So if the calculated test statistic falls within rejection region, we reject H_0

Specify a rejection region (Cont.)

- **Alpha (α)** is the probability level at which you are willing to be wrong when rejecting H_0, also called level of significance or significance level.
  (Def8.8 p394)
- The value at the boundary of the rejection region is called the **critical value**. (Def8.11 p399)

Specify a Rejection Region

- We look to see where z* falls in relation to the standard normal 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 **a priori**

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

Rejection Region

- **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 H_0 when H_0 is true (Def8.6 p393)

Rejection Region

- For this problem, α=.01
- Next I want to find a z-value that will correspond with α=.01
- This is a two-tailed test as specified in the alternative hypothesis
  - H_a: μ ≠ 8.5
- So if α=.01, we need to split this probability level in either tail of the sampling distribution

Rejection Region

- So I look for a ±z-score that corresponds to a probability of .005
- In the table this is a value relating to
  - .5 - .005 = .495
  - The z value from the table is 2.575
- **So our Rejection Region begins at**
  - **< –2.575 or > 2.575**

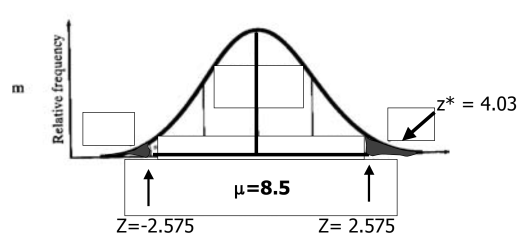
Here is how it looks for the Humerus Bones example

H_0:	μ = 8.5
H_a:	μ ≠ 8.5 two-tailed test
Assumptions	large sample, use normal distribution and z-scores
Test Statistic	z*=(9.258-8.5)/(1.204/√41)
Rejection Region	α = .01/2 z=±2.575 ← Critical values
Calculation:	z*=4.03
Conclusion:	compare z* to the Rejection Region and make a conclusion

Here is how it looks for the Humerus Bones example

H_0:	μ = 8.5
H_a:	μ ≠ 8.5 two-tailed test
Assumptions	large sample, use normal distribution and z-scores
Test Statistic	z*=(9.258-8.5)/(1.204/√41)
Rejection Region	α = .01/2 z=±2.575 ← Critical values
Calculation:	z*=4.03
Conclusion:	**$z^* > z_{\alpha/2=.005}$** **4.03 > 2.575** **So we reject H_0: μ = 8.5**

Here's how it looks in pictures



Basic Elements of a Hypothesis Test

- H_0:
- H_a:
- **Assumptions**
- **Test Statistic**
- **Rejection Region**
- **Calculation:**
- **Conclusion:**

The Elements of a Hypothesis Test

- **Null Hypothesis**
- **Alternative Hypothesis**
- **Assumptions** – clear statement about assumptions of the population or the sampling distribution of the estimator
- **Test Statistic** – the z or t score we calculate to determine if we can reject the null hypothesis

The Elements of a Hypothesis Test

- **Rejection Region** – the numerical value of the Test Statistic at which the null hypothesis will be rejected
  - Based on a probability level α set a priori
  - Tied to the alternative hypothesis
- **Calculate the Test Statistic**
- **Draw a conclusion**

Using Excel for Hypothesis Tests and Confidence Intervals

- Data for speeders on a federal highway. The data are a random sample of 63 motorists on a Federal highway where the speed limit is 65 mph.
  - Use Excel to give the descriptive statistics. Briefly (one paragraph) describe the data using the summary statistics
  - Calculate a 99% Confidence Interval for the data
  - Conduct a Hypothesis Test where the null hypothesis is that the average speed is 65 mph. Use an α level of .01.

Descriptive Statistics

MPH		
Mean	69.222	← **Mean**
Standard Error	1.016	← **Standard Error**
Median	67.000	
Mode	65.000	
Standard Deviation	8.065	
Sample Variance	65.047	
Kurtosis	1.454	
Skewness	1.235	
Range	39	
Minimum	56	
Maximum	95	
Sum	4361	
Count	63	← **Sample Size (n)**
Confidence Level(99.0%	2.700	← **Confidence Interval**

99% Confidence Interval

$$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$$

α= 1.0 -.99 = .01

$\alpha/2$ = .01/2 = .005

$z_{.01/2}$ = 2.575

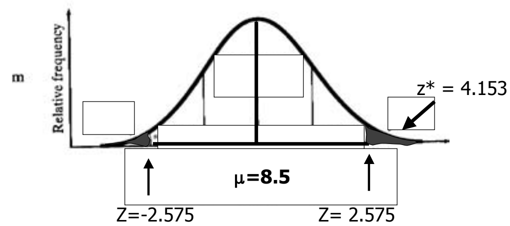
69.22 $\pm$ 2.575(1.016) 1.016 = 8.065/(63)$^{.5}$

69.22 $\pm$ 2.616 **Why is the BOE less than Excel's value of 2.7?**

Hypothesis Test

H_0:	$\mu = 65$
H_a:	$\mu \neq 65$ two-tailed test
Assumptions	large sample, use normal distribution and z-scores
Test Statistic	$z^* = (69.22 - 65)/(8.065/\sqrt{63})$
Rejection Region	$\alpha = .01/2$ $z = \pm 2.575$
Calculation:	$z^* = 4.153$
Conclusion:	**$z^* > z_{\alpha/2=.005}$** **4.153 > 2.575** **So we reject H_0: $\mu = 65$**

Here's how it looks in pictures



Golf Course Problem

- Golf course designers are worried that the new equipment is making old courses obsolete.
- One designer says that courses need to be built with the expectation that players will be able to drive the ball an average of 250 yards or more.

Golf Course Problem

- A sample of 135 golfers is taken and they measured their driving distance
  - $\bar{x} = 256.3$ yards
  - $s = 43.4$ yards
- Does the sample provide enough evidence to suggest that golfers are already hitting it farther than the 250 mark? Use $\alpha = .05$
- Also, calculate the p-value for this problem

Golf Course Problem

Null hypothesis	H_0: $\mu = 250$
Alternative	H_a: $\mu > 250$ one-tailed test,upper
Assumptions	Large sample, normal
Test Statistic	$z^* = (256.3 - 250)/(43.4/\sqrt{135})$
Rejection Region	$z_{\alpha=.05} = 1.645$
Calculation	$z^* = 1.69$
Conclusion	$z^* > z_{\alpha=.05}$ 1.69 > 1.645 Reject H_0: $\mu = 250$ **p= .0455**