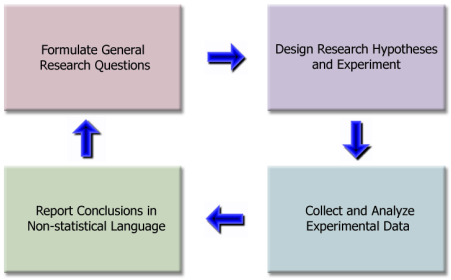
Inferential Statistics and Probability
a Holistic Approach

Chapter 9
One Population Hypothesis Testing

1

1

Procedures of Hypotheses Testing



2

2

Hypotheses Testing – Procedure 1

Formulate General Research Questions

3

3

General Research Question

- Decide on a topic or phenomena that you want to research.
- Formulate general research questions based on the topic.
- Example:
  - Topic: Health Care Reform
  - Some General Questions:
    - Would a Single Payer Plan be less expensive than Private Insurance?
    - Do HMOs provide the same quality care as PPOs?
    - Would the public support mandated health coverage?

4

4

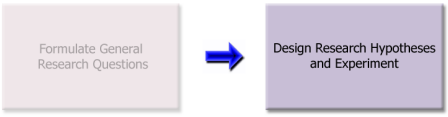
EXAMPLE – General Question

- A food company has a policy that the stated contents of a product match the actual results.
- A General Question might be "Does the stated net weight of a food product match (on average) the actual weight?"
- The quality control statistician could then decide to test various food products for accuracy.

5

5

Hypotheses Testing – Procedure 2



6

6

Hypothesis Testing Design

State Your Hypotheses	
Null Hypothesis	Alternative Hypothesis

Determine Appropriate Model	
Test Statistic	One or Two Tailed

Determine Decision Criteria	
α – Significance Level	β and Power Analysis

7

7

What is a Hypothesis?

- Hypothesis: A statement about the value of a population parameter developed for the purpose of testing.
- Examples of hypotheses made about a population parameter are:
  - The mean monthly income for programmers is $9,000.
  - At least twenty percent of all juvenile offenders are caught and sentenced to prison.
  - The standard deviation for an investment portfolio is no more than 10 percent per month.

8

8

What is Hypothesis Testing?

- Hypothesis testing: A procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and should be rejected.

9

9

Hypothesis Testing Design

State Your Hypotheses	
Null Hypothesis	Alternative Hypothesis

Determine Appropriate Model	
Test Statistic	One or Two Tailed

Determine Decision Criteria	
α – Significance Level	β and Power Analysis

10

10

Definitions

- Null Hypothesis H_0: A statement about the value of a population parameter that is assumed to be true for the purpose of testing.
- Alternative Hypothesis H_a: A statement about the value of a population parameter that is assumed to be true if the Null Hypothesis is rejected during testing.

11

11

Hypotheses written in words and population parameters

- Ho: The mean monthly income for programmers is $9,000.
- Ha: The mean monthly income for programmers is not $9,000.

$$Ho: \mu = 9000 \quad Ha: \mu \neq 9000$$

- Ho: At least 20% of all juvenile offenders sentenced to prison.
- Ha: Less than 20% of all juvenile offenders sentenced to prison.

$$Ho: p \geq 0.20 \quad Ha: p < 0.20$$

- Ho: The standard deviation for an investment portfolio is no more than 10 percent per month.
- Ha: The standard deviation for an investment portfolio is more than 10 percent per month.

$$Ho: \sigma \leq 10 \quad Ha: \sigma > 10$$

12

12

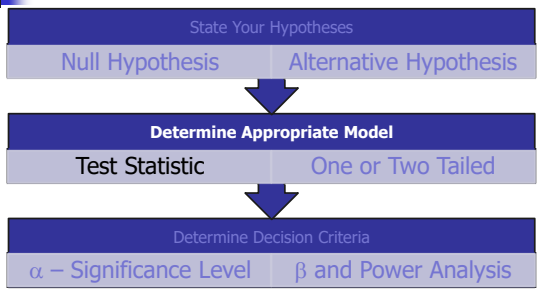
EXAMPLE – Stating Hypotheses

- A food company has a policy that the stated contents of a product match the actual results.
- The quality control statistician decides to test the claim that a 16 ounce bottle of Soy sauce contains on average 16 ounces.
- Ho: The mean amount of Soy Sauce is 16 ounces
- Ha: The mean amount of Soy Sauce is not 16 ounces.
- Ho: $\mu=16$ Ha: $\mu \neq 16$

13

13

Hypothesis Testing Design



14

14

Definitions

- Statistical Model: A mathematical model that describes the behavior of the data being tested.
- Normal Family = the Standard Normal Distribution (Z) and functions of independent Standard Normal Distributions (eg: t, χ^2, F).
  - Most Statistical Models will be from the Normal Family due to the Central Limit Theorem.
- Model Assumptions: Criteria which must be satisfied to appropriately use a chosen Statistical Model.
- Test statistic: A value, determined from sample information, used to determine whether or not to reject the null hypothesis.

15

15

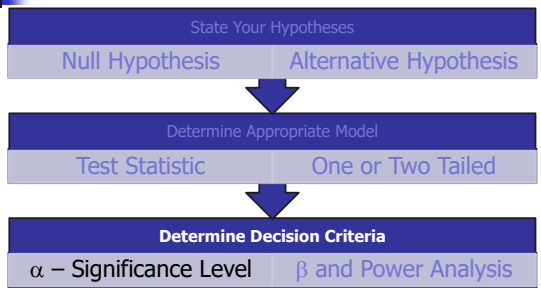
EXAMPLE – Choosing Model

- The quality control statistician decides to test the claim that a 16 ounce bottle of Soy sauce contains on average 16 ounces. We will assume the population standard is known
- Ho: $\mu=16$ Ha: $\mu \neq 16$
- Model: One sample Z test of mean
- Test Statistic: $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

16

16

Hypothesis Testing Design



17

17

Definitions

- Level of Significance: The probability of rejecting the null hypothesis when it is actually true. (signified by α)
- Type I Error: Rejecting the null hypothesis when it is actually true.
- Type II Error: Failing to reject the null hypothesis when it is actually false.

18

18

Outcomes of Hypothesis Testing

	Fail to Reject Ho	Reject Ho
Ho is true	Correct Decision	Type I error
Ho is False	Type II error	Correct Decision

19

19

EXAMPLE – Type I and Type II Errors

- Ho: The mean amount of Soy Sauce is 16 ounces
- Ha: The mean amount of Soy Sauce is not 16 ounces.
- Type I Error: The researcher **supports** the claim that the mean amount of soy sauce is not 16 ounces when the actual mean is 16 ounces. The company needlessly "fixes" a machine that is operating properly.
- Type II Error: The researcher **fails to support** the claim that the mean amount of soy sauce is not 16 ounces when the actual mean is not 16 ounces. The company fails to fix a machine that is not operating properly.

20

20

Hypothesis Testing Design

State Your Hypotheses

Null Hypothesis | Alternative Hypothesis

Determine Appropriate Model

Test Statistic | One or Two Tailed

Determine Decision Criteria

α – Significance Level | β and Power Analysis

21

21

Definitions

- Critical value(s): The dividing point(s) between the region where the null hypothesis is rejected and the region where it is not rejected. The critical value determines the decision rule.
- Rejection Region: Region(s) of the Statistical Model which contain the values of the Test Statistic where the Null Hypothesis will be rejected. The area of the Rejection Region = α

22

22

One-Tailed Tests of Significance

- A test is one-tailed when the alternate hypothesis, H_a , states a direction, such as:
  - H_0 : The mean income of females is less than or equal to the mean income of males.
  - H_a : The mean income of females is greater than males.
- Equality is part of H_0
- H_a determines which tail to test
  - H_a: $\mu > \mu_0$ means test upper tail.
  - H_a: $\mu < \mu_0$ means test lower tail.

23

23

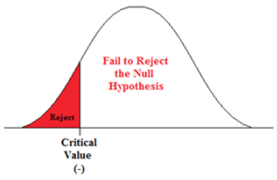
Left-tailed test

$$H_0 : \mu \geq \mu_0$$

$$H_a : \mu < \mu_0$$

$$\alpha = .05$$

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$



24

24

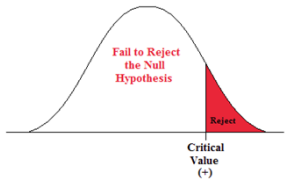
Right-tailed test

$$H_0 : \mu \le \mu_0$$

$$H_a : \mu > \mu_0$$

$$\alpha = .05$$

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$



25

25

Two-Tailed Tests of Significance

- A test is two-tailed when no direction is specified in the alternate hypothesis H_a , such as:
  - H_0 : The mean income of females is equal to the mean income of males.
  - H_a : The mean income of females is not equal to the mean income of the males.
- Equality is part of H_0
- H_a determines which tail to test
  - H_a: $\mu \neq \mu_0$ means test both tails.

26

26

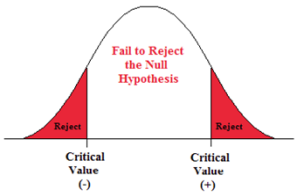
Two-tailed test

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$

$$\alpha = .05 \quad \alpha/2 = .025$$

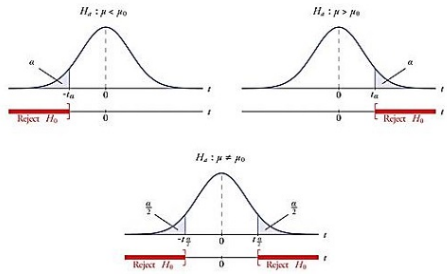
$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$



27

27

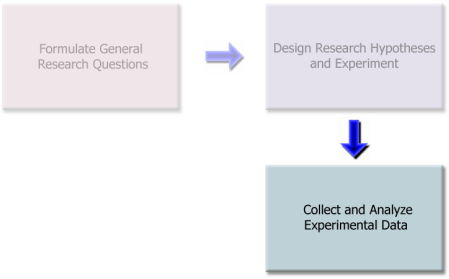
Left, Right and Two-Tailed Tests



28

28

Hypotheses Testing – Procedure 3



29

29

Collect and Analyze Experimental Data

Collect and Verify Data	
Conduct Experiment	Check for Outliers
Determine Test Statistic and/or p-value	
Compare to Critical Value	Compare to α
Make a Decision about Ho	
Reject Ho and support Ha	Fail to Reject Ho

30

30

Collect and Analyze Experimental Data

Collect and Verify Data

Conduct Experiment	Check for Outliers

Determine Test Statistic and/or p-value

Compare to Critical Value	Compare to α

Make a Decision about Ho

Reject Ho and support Ha	Fail to Reject Ho

31

31

Outliers

- An outlier is data point that is far removed from the other entries in the data set.
- Outliers could be
  - Mistakes made in recording data
  - Data that don't belong in population
  - True rare events

32

32

Outliers have a dramatic effect on some statistics

- Example quarterly home sales for 10 realtors:

2 2 3 4 5 5 6 6 7 50

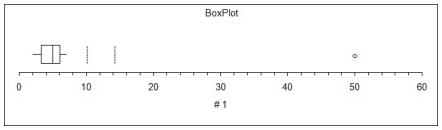
	with outlier	without outlier
Mean	9.00	4.44
Median	5.00	5.00
Std Dev	14.51	1.81
IQR	3.00	3.50

33

33

Using Box Plot to find outliers

- The "box" is the region between the 1st and 3rd quartiles.
- Possible outliers are more than 1.5 IQR's from the box (inner fence)
- Probable outliers are more than 3 IQR's from the box (outer fence)
- In the box plot below, the dotted lines represent the "fences" that are 1.5 and 3 IQR's from the box. See how the data point 50 is well outside the outer fence and therefore an almost certain outlier.



34

34

Using Z-score to detect outliers

- Calculate the mean and standard deviation without the suspected outlier.
- Calculate the Z-score of the suspected outlier.
- If the Z-score is more than 3 or less than -3, that data point is a probable outlier.

$$Z = \frac{50 - 4.4}{1.81} = 25.2$$

35

35

Outliers – what to do

- Remove or not remove, there is no clear answer.
- For some populations, outliers don't dramatically change the overall statistical analysis. Example: the tallest person in the world will not dramatically change the mean height of 10000 people.
- However, for some populations, a single outlier will have a dramatic effect on statistical analysis (called "**Black Swan**" by Nicholas Taleb) and inferential statistics may be invalid in analyzing these populations. Example: the richest person in the world will dramatically change the mean wealth of 10000 people.

36

36

Example – Analyze Data

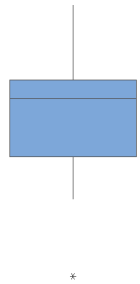
- In the Soy Sauce Example, a 36 bottles were measured, volume is in fluid ounces

14.51	15.16	15.28	15.33	15.36	15.42
15.43	15.45	15.49	15.59	15.60	15.61
15.62	15.63	15.71	15.81	15.87	16.00
16.01	16.02	16.05	16.06	16.06	16.09
16.09	16.11	16.16	16.16	16.27	16.31
16.35	16.36	16.45	16.72	16.75	16.79

37

37

Example – Analyze Data

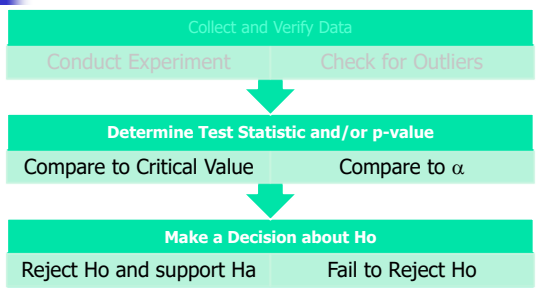


- Although 14.51 might be a possible outlier and the data seems negatively skewed, the Central Limit Theorem assures that the sample mean will have a normal distribution

38

38

Collect and Analyze Experimental Data



39

39

The logic of Hypothesis Testing

- This is a "Proof" by contradiction.
  - We assume H_o is true before observing data and design H_a to be the complement of H_o.
  - Observe the data (evidence). How unusual are these data under H_o?
  - If the data are too unusual, we have "proven" H_o is false: Reject H_o and go with H_a (Strong Statement)
  - If the data are not too unusual, we fail to reject H_o. This "proves" nothing and we say data are inconclusive. (Weak Statement)
  - We can never "prove" H_o, only "disprove" it.
  - "Prove" in statistics means support with the Alternative Hypothesis.
  - Note: It is **never correct** to say (1-α)100% certain of our decision. (example: if α=.05, then we are not 95% certain if we Reject Ho.)

40

40

Test Statistic

- Test Statistic: A value calculated from the Data under the appropriate Statistical Model from the Data that can be compared to the Critical Value of the Hypothesis test
- If the Test Statistic fall in the Rejection Region, Ho is rejected.
- The Test Statistic will also be used to calculate the p-value as will be defined next.

41

41

Example - Testing for the Population Mean
Large Sample, Population Standard Deviation Known

- When testing for the population mean from a large sample and the population standard deviation is known, the test statistic is given by:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

42

42

p-value in Hypothesis Testing

- **p-Value**: the probability, assuming that the null hypothesis is true, of getting a value of the test statistic at least as extreme as the computed value for the test.
- If the p-value is smaller than the significance level, H_0 is rejected.
- If the p-value is larger than the significance level, H_0 is not rejected.

43

43

Comparing p-value to α

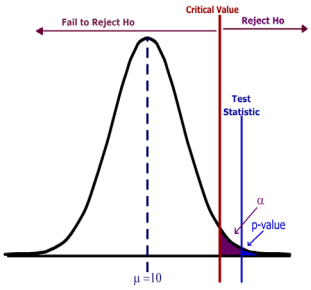
- Both **p-value** and **α** are probabilities.
- The **p-value** is determined by the **data**, and is the probability of getting results as extreme as the data assuming H_0 is true. Small values make one more likely to reject H_0.
- **α** is determined by **design**, and is the maximum probability the experimenter is willing to accept of rejecting a true H_0.
- Reject H_0 if p-value < α for ALL MODELS.

44

44

Graphic where decision is to Reject Ho

- Ho: μ = 10
  Ha: μ > 10
- Design: Critical Value is determined by significance level α.
- Data Analysis: p-value is determined by Test Statistic
- Test Statistic falls in Rejection Region.
- p-value (blue) < α (purple)
- Reject Ho.
- Strong statement: Data supports Alternative Hypothesis.

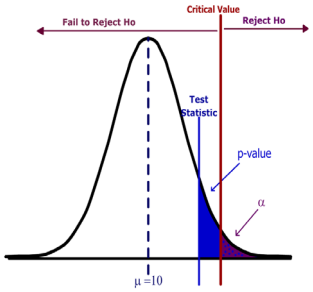


45

45

Graphic where decision is Fail to Reject Ho

- Ho: μ = 10
  Ha: μ > 10
- Design: Critical Value is determined by significance level α.
- Data Analysis: p-value is determined by Test Statistic
- Test Statistic falls in Non-rejection Region.
- p-value (blue) > α (purple)
- Fail to Reject Ho.
- Weak statement: Data is inconclusive and does not support Alternative Hypothesis.



46

46

EXAMPLE – General Question

- A food company has a policy that the stated contents of a product match the actual results.
- A General Question might be "Does the stated net weight of a food product match the actual weight?"
- The quality control statistician decides to test the 16 ounce bottle of Soy Sauce.

47

47

EXAMPLE – Design Experiment

- A sample of n=36 bottles will be selected hourly and the contents weighed. Assume σ = 0.5
- Ho: μ=16 Ha: $\mu \neq 16$
- The Statistical Model will be the one population test of mean using the Z Test Statistic.
- This model will be appropriate since the sample size insures the sample mean will have a Normal Distribution (Central Limit Theorem)
- We will choose a significance level of α = 5%

48

48

EXAMPLE – Conduct Experiment

- Last hour a sample of 36 bottles had a mean weight of 15.88 ounces.
- From past data, assume the population standard deviation is 0.5 ounces.
- Compute the Test Statistic

$$Z = [15.88 - 16]/[.5/\sqrt{36}] = -1.44$$

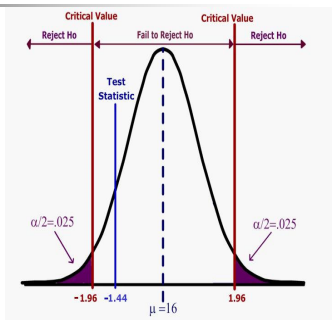
- For a two tailed test, The Critical Values are at Z = ±1.96

49

49

Decision – Critical Value Method

- This two-tailed test has two Critical Value and Two Rejection Regions
- The significance level (α) must be divided by 2 so that the sum of both purple areas is 0.05
- The Test Statistic does not fall in the Rejection Regions.
- Decision is **Fail to Reject Ho.**



50

50

Computation of the p-Value

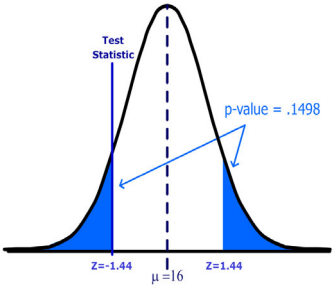
- One-Tailed Test: p-Value = P{z ≥ absolute value of the computed test statistic value}
- Two-Tailed Test: p-Value = 2P{z ≥ absolute value of the computed test statistic value}
- Example: Z= 1.44, and since it was a two-tailed test, then p-Value = 2P {z ≥ 1.44} = 0.0749) = .1498. Since .1498 > .05, do not reject H_0.

51

51

Decision – p-value Method

- The p-value for a two-tailed test must include all values (positive and negative) more extreme than the Test Statistic.
- p-value = .1498 which exceeds α = .05
- Decision is **Fail to Reject Ho.**



52

52

p-value form Minitab (shown as p)

One-Sample Z: weight

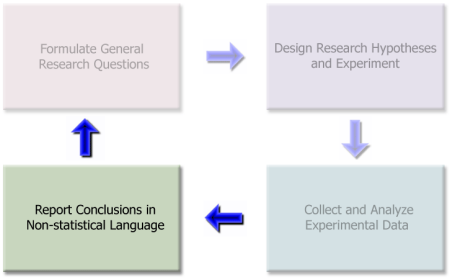
Test of μ = 16 vs ≠ 16
The assumed standard deviation = 0.5

Variable	N	Mean	StDev	SE Mean	Z	**P**
weight	36	15.8800	0.4877	0.0833	-1.44	**0.150**

53

53

Hypotheses Testing – Procedure 4



54

54

Converting Decision to Conclusion

- Conclusion if Decision is Reject Ho:
- <Ha in the Context of Problem>

- Conclusion if Decision is Fail to Reject Ho:
- "There is insufficient evidence to conclude" <Ha in the Context of Problem>

55

55

Example - Conclusion

- Decision: Fail to Reject Ho

- There is insufficient evidence to conclude that the mean amount of soy sauce being filled into bottles is not 16 ounces.
- There is insufficient evidence to conclude machine that fills 16 ounce soy sauce bottles is operating improperly.

56

56

Conclusions

- Conclusions need to
  - Be consistent with the results of the Hypothesis Test.
  - Use language that is clearly understood in the context of the problem.
  - Limit the inference to the population that was sampled.
  - Report sampling methods that could question the integrity of the random sample assumption.
  - Conclusions should address the potential or necessity of further research, sending the process back to the first procedure.

57

57

Conclusions need to be consistent with the results of the Hypothesis Test.

- Rejecting Ho requires a strong statement in support of Ha.
- Failing to Reject Ho does NOT support Ho, but requires a weak statement of insufficient evidence to support Ha.
- Example:
  - The researcher wants to support the claim that, on average, students send more than 1000 text messages per month
  - Ho: $\mu=1000$ Ha: $\mu>1000$
  - Conclusion if Ho is rejected: The mean number of text messages sent by students exceeds 1000.
  - Conclusion if Ho is not rejected: There is insufficient evidence to support the claim that the mean number of text messages sent by students exceeds 1000.

58

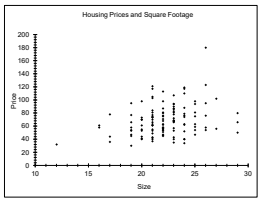
58

Conclusions need to use language that is clearly understood in the context of the problem.

- Avoid technical or statistical language.
- Refer to the language of the original general question.
- Compare these two conclusions from a test of correlation between home prices square footage and price.

Conclusion 1: By rejecting the Null Hypothesis we are inferring that the Alterative Hypothesis is supported and that there exists a significant correlation between the independent and dependent variables in the original problem comparing home prices to square footage.

Conclusion 2: Homes with more square footage generally have higher prices.



59

59

Conclusions need to limit the inference to the population that was sampled.

- If a survey was taken of a sub-group of population, then the inference applies to the subgroup.
- Example
  - Studies by pharmaceutical companies will only test adult patients, making it difficult to determine effective dosage and side effects for children.
  - "In the absence of data, doctors use their medical judgment to decide on a particular drug and dose for children. 'Some doctors stay away from drugs, which could deny needed treatment,' Blumer says. "Generally, we take our best guess based on what's been done before."
  - "The antibiotic chloramphenicol was widely used in adults to treat infections resistant to penicillin. But many newborn babies died after receiving the drug because their immature livers couldn't break down the antibiotic."

source: FDA Consumer Magazine – Jan/Feb 2003

60

60

Conclusions need to report sampling methods that could question the integrity of the random sample assumption.

- Be aware of how the sample was obtained. Here are some examples of pitfalls:
  - Telephone polling was found to under-sample young people during the 2008 presidential campaign because of the increase in cell phone only households. Since young people were more likely to favor Obama, this caused bias in the polling numbers.
  - Sampling that didn't occur over the weekend may exclude many full time workers.
  - Self-selected and unverified polls (like ratemyprofessors.com) could contain immeasurable bias.

61

61

Conclusions should address the potential or necessity of further research, sending the process back to the first procedure.

- Answers often lead to new questions.
- If changes are recommended in a researcher's conclusion, then further research is usually needed to analyze the impact and effectiveness of the implemented changes.
- There may have been limitations in the original research project (such as funding resources, sampling techniques, unavailability of data) that warrants more a comprehensive study.
  - Example: A math department modifies is curriculum based on a performance statistics for an experimental course. The department would want to do further study of student outcomes to assess the effectiveness of the new program.

62

62

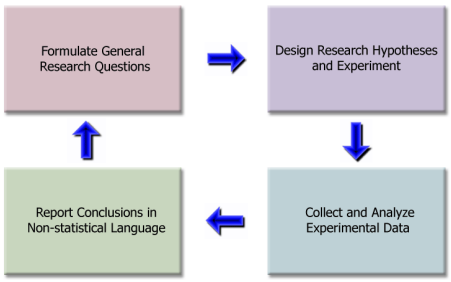
Soy Sauce Example - Conclusion

- There is insufficient evidence to conclude that the machine that fills 16 ounce soy sauce bottles is operating improperly.
- This conclusion is based on 36 measurements taken during a single hour's production run.
- We recommend continued monitoring of the machine during different employee shifts to account for the possibility of potential human error.

63

63

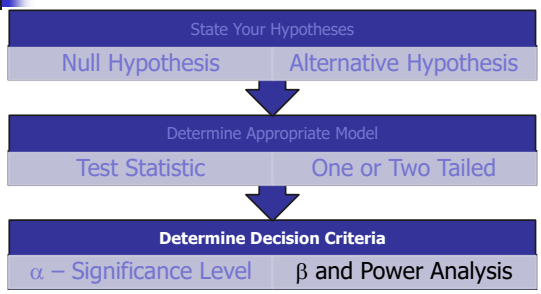
Procedures of Hypotheses Testing



64

64

Hypothesis Testing Design



65

65

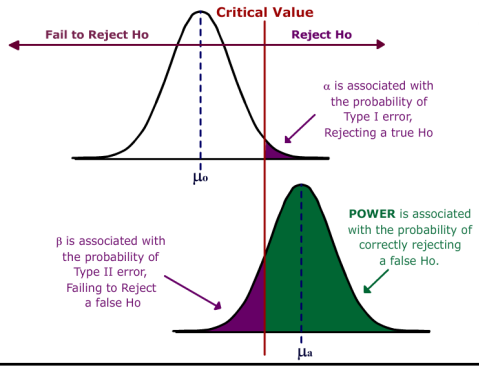
Statistical Power and Type II error

	Fail to Reject Ho	Reject Ho
Ho is true	$1-\alpha$	α Type I error
Ho is False	β Type II error	$1-\beta$ Power

66

66

Graph of "Four Outcomes"



67

67

Statistical Power (continued)

- Power is the probability of rejecting a false Ho, when $\mu = \mu_a$
- Power depends on:
  - Effect size $|\mu_o - \mu_a|$
  - Choice of α
  - Sample size
  - Standard deviation
  - Choice of statistical test

68

68

Statistical Power Example

- Bus brake pads are claimed to last on average at least 60,000 miles and the company wants to test this claim.
- The bus company considers a "practical" value for purposes of bus safety to be that the pads at least 58,000 miles.
- If the standard deviation is 5,000 and the sample size is 50, find the Power of the test when the mean is really 58,000 miles. Assume α = .05

69

69

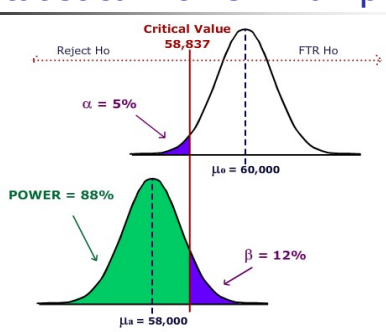
Statistical Power Example

- Set up the test
  - Ho: μ >= 60,000 miles
  - Ha: μ < 60,000 miles
  - α = 5%
- Determine the Critical Value
  - Reject Ho if $\bar{X} > 58{,}837$
- Calculate β and Power
  - β = 12%
  - Power = 1 – β = 88%

70

70

Statistical Power Example



71

71

New Models, Similar Procedures

- The **procedures** outlined for the One Sample Z test of Mean (with known population standard deviation) will apply to other models as well.
- Examples of some other one population models:
  - One Sample t-test of mean, population standard deviation unknown.
  - One sample Z-test of proportion.
  - One sample Chi-square test of variance (or standard deviation)

72

72

Testing for the Population Mean: Population Standard Deviation Unknown

- Model: One sample t-test of mean
- $$t = \frac{\bar{X} - \mu_o}{s/\sqrt{n}}$$
- The degrees of freedom for the test is n-1.
- Assumptions: $\bar{X}$ has a Normal Distribution

73

73

Decision Rules

- Like the normal distribution, the logic for one and two tail testing is the same.
- For a two-tail test using the *t*-distribution, you will reject the null hypothesis when the value of the test statistic is greater than $t_{df,\alpha/2}$ or if it is less than - $t_{df,\alpha/2}$
- For a left-tail test using the *t*-distribution, you will reject the null hypothesis when the value of the test statistic is less than $-t_{df,\alpha}$
- For a right-tail test using the *t*-distribution, you will reject the null hypothesis when the value of the test statistic is greater than $t_{df,\alpha}$

74

74

Example – one population test of mean, σ unknown

- Humerus bones from the same species have approximately the same length-to-width ratios. When fossils of humerus bones are discovered, archaeologists can determine the species by examining this ratio. It is known that Species A has a mean ratio of 9.6. A similar Species B has a mean ratio of 9.1 and is often confused with Species A.
- 21 humerus bones were unearthed in an area that was originally thought to be inhabited Species A. (Assume all unearthed bones are from the same species.)
- Design a hypothesis test where the alternative claim would be the humerus bones were not from Species A.
- Determine the power of this test if the bones actually came from Species B (assume a standard deviation of 0.7)
- Conduct the test using at a 5% significance level and state overall conclusions.

75

75

Example – Designing Test

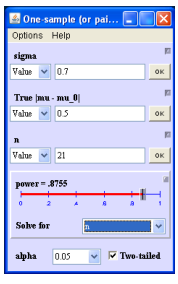
- Research Hypotheses
  - Ho: The humerus bones are from Species A
  - Ha: The humerus bones are not from Species A
- In terms of the population mean
  - Ho: $\mu = 9.6$
  - Ha: $\mu \neq 9.6$
- Significance level
  - $\alpha = .05$
- Test Statistic (Model)
  - One sample t-test of mean

76

76

Example - Power Analysis

- Information needed for Power Calculation
  - $\mu_o = 9.6$ (Species A)
  - $\mu_a = 9.1$ (Species B)
  - Effect Size $= |\mu_o - \mu_a| = 0.5$
  - $\sigma = 0.7$ (given)
  - $\alpha = .05$
  - n = 21 (sample size)
  - Two tailed test
- Results using online Power Calculator*
  - Power =.8755
  - β = 1 - Power = .1245
  - If humerus bones are from Species B, test has an 87.55% chance of correctly rejecting Ho and a maximum Type II error of 12.45%

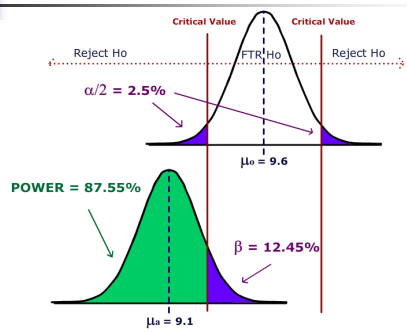


*source: Russ Lenth, University of Iowa – http://www.stat.uiowa.edu/~rlenth/Power/

77

77

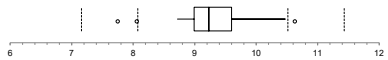
Example – Power Analysis



78

78

Example – Output of Data Analysis



Hypothesis Test: Mean vs. Hypothesized Value

9.60000	hypothesized value
9.26190	mean Data
0.66700	std. dev.
0.14555	std. error
21	n
20	df
-2.32	t
.0308	p-value (two-tailed)

P-value = .0308
α =.05
Since p-value < α
Ho is rejected and we support Ha.

79

79

Example - Conclusions

- Results:
  - The evidence supports the claim (pvalue<.05) that the humerus bones are not from Species A.
- Sampling Methodology:
  - We are assuming since the bones were unearthed in the same location, they came from the same species.
- Limitations:
  - A small sample size limited the power of the test, which prevented us from making a more definitive conclusion.
- Further Research
  - Test if the bone are from Species B or another unknown species.
  - Test to see if bones are the same age to support the sampling methodology.

80

80

Tests Concerning Proportion

- Proportion: A fraction or percentage that indicates the part of the population or sample having a particular trait of interest.
- The population proportion is denoted by p.
- The sample proportion is denoted by $\hat{p}$ where

$$\hat{p} = \frac{\text{number of successes in the sample}}{\text{number sampled}}$$

81

81

Test Statistic for Testing a Single Population Proportion

- If sample size is sufficiently large, $\hat{p}$ has an approximately normal distribution. This approximation is reasonable if $np_o \geq 10$ and $n(1-p_o) \geq 10$

$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$p_o = \textit{population proportion under Ho}$$

$$\hat{p} = \textit{sample proportion}$$

82

82

Example

- In the past, 15% of the mail order solicitations for a certain charity resulted in a financial contribution.
- A new solicitation letter has been drafted and will be sent to a random sample of potential donors.
- A hypothesis test will be run to determine if the new letter is more effective.
- Determine the sample size so that:
  - The test can be run at the 5% significance level.
  - If the letter has an 18% success rate, (an effect size of 3%), the power of the test will be 95%
- After determining the sample size, conduct the test.

83

83

Example – Designing Test

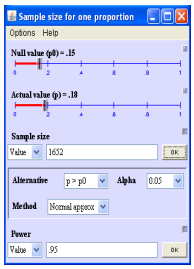
- Research Hypotheses
  - Ho: The new letter is not more effective.
  - Ha: The new letter is more effective.
- In terms of the population proportion
  - Ho: p = 0.15
  - Ha: p > 0.15
- Significance level
  - α =.05
- Test Statistic (Model)
  - One sample Z-test of proportion

84

84

Example - Power Analysis

- Information needed for Sample Size Calculation
  - p_o = 0.15 (current letter)
  - p_a = 0.18 (potential new letter)
  - Effect Size =| p_o - p_a | = 0.03
  - Desired Power = 0.95
  - α = .05
  - One tailed test
- Results using online Power Calculator*
  - Sample size = 1652
  - The charity should send out 1652 new solicitation letters to potential donors and run the test.

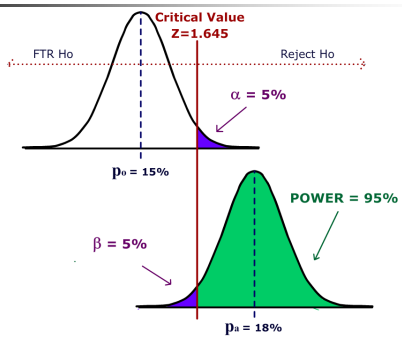


*source: Russ Lenth, University of Iowa – http://www.stat.uiowa.edu/~rlenth/Power/

85

85

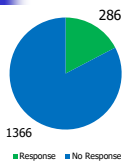
Example – Power Analysis



86

86

Example – Output of Data Analysis



Hypothesis test for proportion vs hypothesized value

Observed	*Hypothesized*	
0.1731	0.15	p (as decimal)
286/1652	248/1652	p (as fraction)
286.	247.8	X
1652	1652	n
	0.0088	std. error
	2.63	z
	.0042	p-value (one-tailed, upper)

- P-value = .0042
- α =0.05
- Since p-value < α, Ho is rejected and we support Ha.

87

87

EXAMPLE
Critical Value Alternative Method

- Critical Value =1.645 (95th percentile of the Normal Distribution.)
- H_0 is rejected if Z > 1.645
- Test Statistic: $$Z = \frac{\left(\frac{286}{1652} - .15\right)}{\sqrt{\frac{(.15)(.85)}{1652}}} = 2.63$$
- Since Z = 2.63 > 1.645, H_0 is rejected. The new letter is more effective.

88

88

Example - Conclusions

- Results:
  - The evidence supports the claim (pvalue<.01) that the new letter is more effective.
- Sampling Methodology:
  - The 1652 test letters were selected as a random sample from the charity's mailing list. All letters were sent at the same time period.
- Limitations:
  - The letters needed to be sent in a specific time period, so we were not able to control for seasonal or economic factors.
- Further Research
  - Test both solicitation methods over the entire year to eliminate seasonal effects.
  - Send the old letter to another random sample to create a control group.

89

89

Test for Variance or Standard Deviation vs. Hypothesized Value

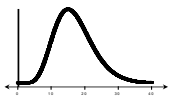
- We often want to make a claim about the variability, volatility or consistency of a population random variable.
- Hypothesized values for population variance σ^2 or standard deviation σ are tested with the χ^2 distribution.
- Examples of Hypotheses:
  - Ho: $\sigma = 10$ Ha: $\sigma \neq 10$
  - Ho: $\sigma^2 = 100$ Ha: $\sigma^2 > 100$
- The sample variance s^2 is used in calculating the Test Statistic.

90

90

Test Statistic uses χ^2 distribution

s^2 is the test statistic for the population variance. Its sampling distribution is a χ^2 distribution with n-1 d.f.



$$\chi^2 = \frac{(n-1)s^2}{\sigma_o^2}$$

91

91

Example

- A state school administrator claims that the standard deviation of test scores for 8th grade students who took a life-science assessment test is less than 30, meaning the results for the class show consistency.
- An auditor wants to support that claim by analyzing 41 students recent test scores, shown here:

57	75	86	92	101	108	110	120	155
63	77	88	96	102	108	111	122	
66	78	88	96	107	109	115	135	
68	81	92	98	107	109	115	137	
72	82	92	99	107	110	118	139	

- The test will be run at 1% significance level.

92

92

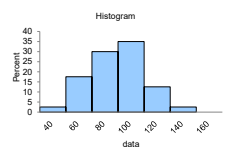
Example – Designing Test

- Research Hypotheses
  - Ho: Standard deviation for test scores equals 30.
  - Ha: Standard deviation for test scores is less than 30.
- In terms of the population variance
  - Ho: $\sigma^2 = 900$
  - Ha: $\sigma^2 < 900$
- Significance level
  - $\alpha = .01$
- Test Statistic (Model)
  - One sample Chi-square test of variance

93

93

Example – Output of Data Analysis



Chi-square Variance Test

900.000	hypothesized variance
469.426	observed variance of Data
41	n
40	df
20.86	chi-square
.0054	p-value (one-tailed, lower)

- p-value = .0054
- $\alpha = 0.01$
- Since p-value < α, Ho is rejected and we support Ha.

94

94

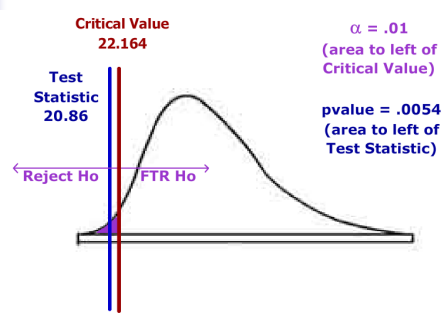
EXAMPLE
Critical Value Alternative Method

- Critical Value =22.164 (1st percentile of the Chi-square Distribution.)
- H_0 is rejected if $\chi^2 < 22.164$
- Test Statistic: $\chi^2 = \frac{(40)(469.426)}{900} = 20.86$
- Since Z = 20.86< 22.164, H_0 is rejected. The claim that the standard deviation is under 30 is supported.

95

95

Example – Decision Graph



96

96

Example - Conclusions

- Results:
  - The evidence supports the claim (pvalue<.01) that the standard deviation for 8th grade test scores is less than 30.
- Sampling Methodology:
  - The 41 test scores were the results of the recently administered exam to the 8th grade students.
- Limitations:
  - Since the exams were for the current class only, there is no assurance that future classes will achieve similar results.
- Further Research
  - Compare results to other schools that administered the same exam.
  - Continue to analyze future class exams to see if the claim is holding true.

97

97