TECH 646 Analysis of Research in Industry and Technology

PART IV

Analysis and Presentation of Data:

Data Presentation and Description; Exploring, Displaying, and Examining Data; Hypothesis Testing; Measures of Association; Multivariate Analysis; Presenting Insights and Findings

Ch. 17 Hypothesis Testing (1 of 2)

Lecture note based on the text book and supplemental materials:

Cooper, D.R., & Schindler, P.S., *Business Research Methods* (11th edition), 2011, McGraw-Hill/Irwin

Paul I-Hai Lin, Professor

http://www.etcs.pfw.edu/~lin

A Core Course for M.S. iln Technology Program

Purdue University Fort Wayne

1

Exploring, Displaying, and Examining Data

Learning Objectives … Understand

- The nature and logic of hypothesis testing.
- A statistically significant difference.
- The six-step hypothesis testing procedure.
- The difference between parametric and nonparametric tests and when to use each.
- The factors that influence the selection of an appropriate test of statistical significance.
- How to interpret the various test statistics.

2

Inductive and Deductive Reasoning

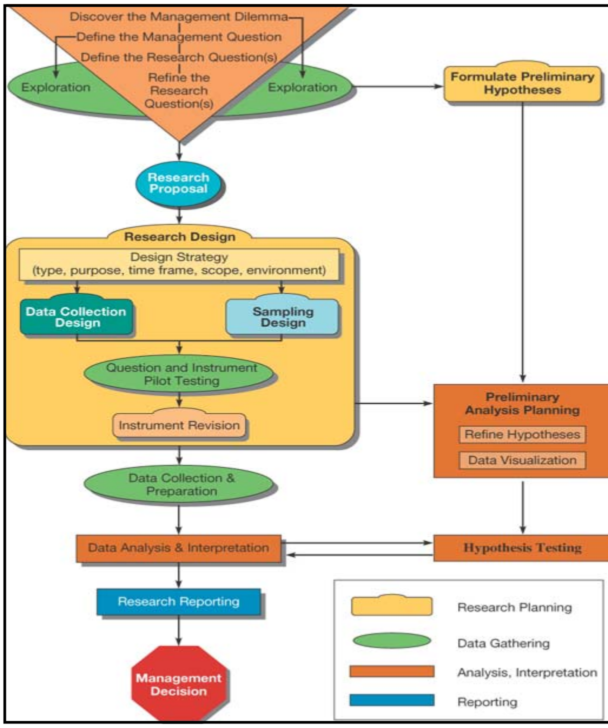
- Inductive Reasoning
  - From specific facts to general, but tentative, conclusions.
  - Using probability estimates, we can qualify our results and state the degree of confidence.
  - Statistical inference - an application of inductive reasoning
  - Allows us to reason from evidence found in the sample to conclusions we wish to make about the population.
- Deductive Reasoning
  - Conclusion must necessarily follow from the premises given.

3

Statistical Procedures

- Inferential statistics
  - Includes the estimation of population values and the testing of statistical hypotheses.
- Descriptive statistics
  - Simply describe the characteristics of the data by giving frequencies, measures of central tendency, and dispersion.

4



Hypothesis Testing and the Research Process

5

Hypothesis (review from Chapter 3)

- Hypothesis
  - A declarative statement about the relationship between two or more variables
  - A tentative and speculative/conjectural nature
- Classification of Hypothesis
  - Descriptive Hypothesis
  - Relational Hypothesis
- The Role of the Hypothesis
  - Guide the direction of the study
  - Identify “facts” about that are relevant
  - Suggest which form of “research design”
  - Provide a framework for organizing the conclusion

6

Hypothesis (review from Chapter 3)

- Descriptive Hypothesis
  - Rewrite research questions as statements that describe the existence, size, form, or distribution of some variables
  - Encourage thinking about "the likely relationship to be found", and "the implications of a supported or rejected findings"
  - Useful for testing statistical significance

7

Hypothesis (review from Chapter 3)

- Relational Hypothesis
  - Statements that describe a relationship between two variables with respect to some case
  - Correlational relationship (unspecified relationship)
    - Example: Foreign (variable) cars are perceived by American consumers (case) to be more of better quality (variable) than domestic cars.
    - Relationship between two variables: Country-Of-Origin, and Perceived-Quality is not specified
    - Make no claim – on Cause-Effect relationship
  - Predictable relationship (explanatory, or causal relationship)

8

Approaches to Hypothesis Testing

- Classical Statistics (sampling theory)
  - Accept or reject a hypothesis on the basis of sampling information alone
  - Analysis based on sample data
  - Judge whether the differences (sample/population) are statistically significant or insignificant
- Bayesian Statistics
  - Extension of classical approach
  - Analysis based on sample data
  - Also considers established subjective probability estimates

Statistical Significance

- The hybrid Toyota Prius has maintained an average of about 51 mpg for city driving, with a standard deviation of 10 mpg.
  - Researchers discover by analyzing all production vehicles that the mpg is now 51. (based on a census, no sampling involved)
  - Question: Is the difference statistically significant? If 51 significant that 50?
  - Too expensive to analyze all the manufacturer’s vehicle
- Resort to sampling: a sample of 25 cars

Statistical Significance

- Resort to sampling method
  - A sample of 25 vehicles is randomly selected
  - The average city-mpg is calculated to be 54 mpg.
  - Question:
    - Is 51 significantly different from 54 or is it only sampling error?
  - Hypothesis testing will answer it

11

Types of Hypotheses

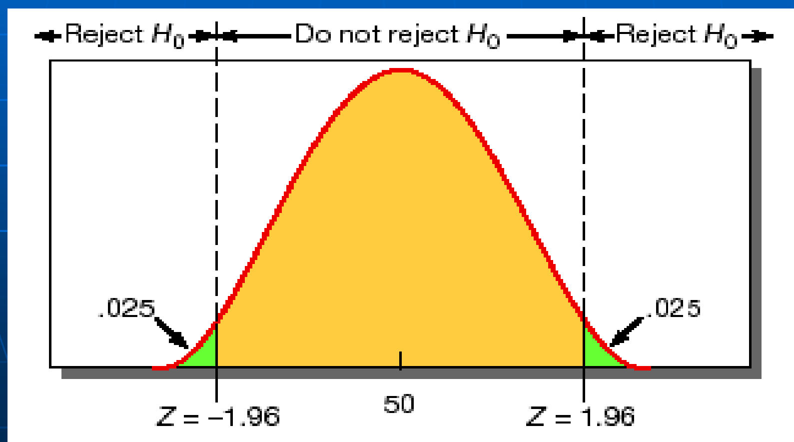
- Null Hypothesis
  - A statement - no difference exists between the parameter and the statistics being compared
  - Toyota Prius (null hypothesis: the population parameter of 50 mpg has not changed)
- Alternative Hypothesis
  - A logical opposite of the null hypothesis

- **Null**
  - H_0: $\mu =$ **50 mpg**
  - H_0: $\mu \leq$ **50 mpg**
  - H_0: $\mu \geq$ **50 mpg**
- **Alternate**
  - H_A: $\mu \neq$ **50 mpg**
  - H_A: $\mu >$ **50 mpg**
  - H_A: $\mu <$ **50 mpg**

12

Two-Tailed Test at 5% Level of Significance

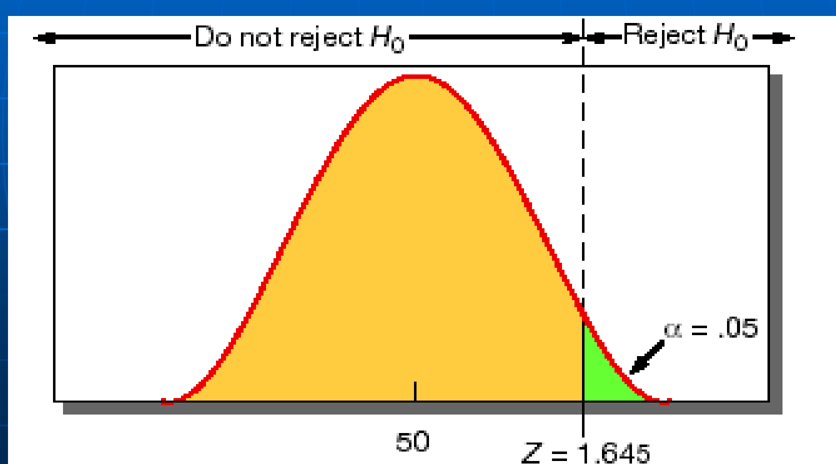
- Null H_0: $\mu = 50$ mpg
- Alternative H_A: $\mu \neq 50$ mpg (not the same case)



13

One-Tailed Test at 5% Level of Significance

- Null H_0: $\mu \leq 50$ mpg
- Alternative H_A: $\mu > 50$ mpg (greater than case)

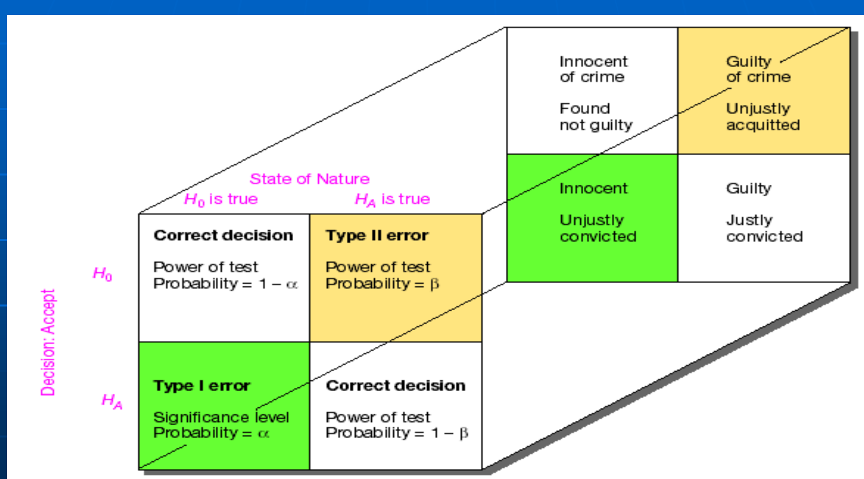


14

Decision Rule

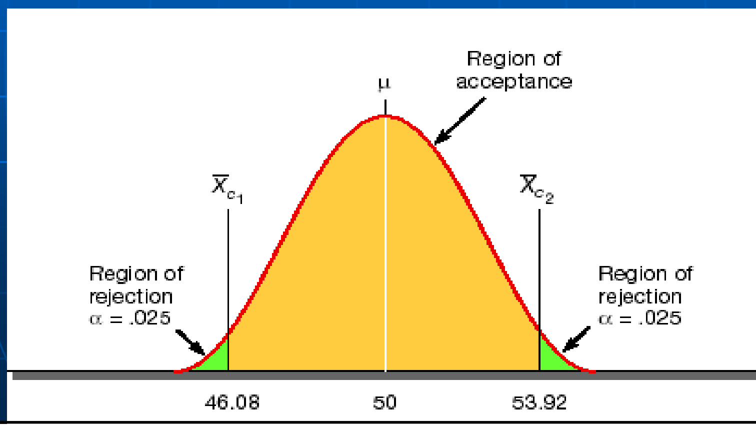
- ***Take no corrective action if the analysis shows that one "cannot reject" the null hypothesis.***

Statistical Decisions



Probability of Type I Error of 0.05

- Population mean 50 mpg. Standard deviation σ =10 mpg, Sample size n = 25 vehicles
- Standard error of the mean $\sigma_{\overline{x}} = \sigma/\sqrt{n}$ = 10/5 = 2 mpg
- Z = 1.96 => 2 x 1.96 = 3.92 mpg
- $\overline{X}_{C1}$= μ - Z, $\overline{X}_{C2}$= μ + Z



17

Critical Values

Z = 1.96 (significance level = .05)
$\overline{X}_c$ = the critical value of the sample mean
μ = the population value stated in $H_0 = 50$
$\sigma_{\overline{X}}$ = the standard error of a distribution of means of samples of 25

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

$$-1.96 = \frac{\overline{X}_c - 50}{2}$$

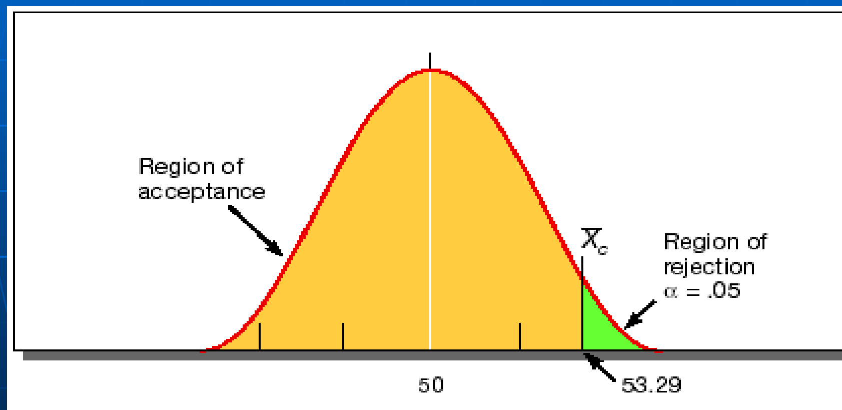
$$\overline{X}_c = 46.08$$

$$1.96 = \frac{\overline{X}_c - 50}{2}$$

$$\overline{X}_c = 53.92$$

Exhibit 17-4 Probability of Making A Type I Error

- Accept a α risk of 5%
- Z = 1.645 = ($\overline{X}_C$- μ)/2   => $\overline{X}_C$= 50 + 3.29

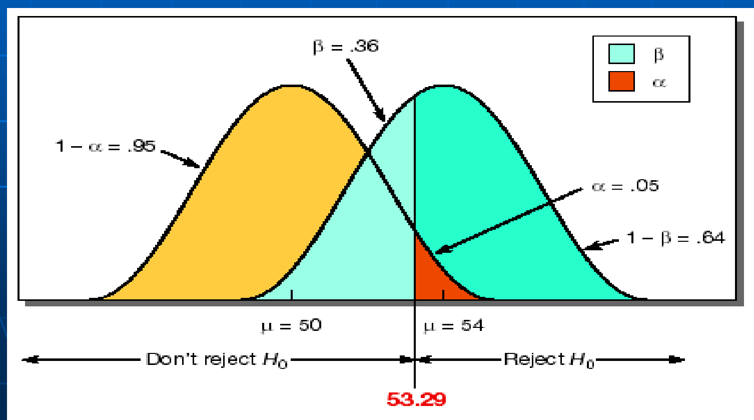


Factors Affecting Probability of Committing a Type II Error (β)

1. True value of parameter
2. Alpha level selected
3. One or two-tailed test used
4. Sample standard deviation
5. Sample size

Probability of Making A Type II Error (β)

- Would commit a Type II error (β) by accepting the null hypothesis (μ = 50) when in truth it had changed
- If μ moved from 50 to 54, what is probability of making a Type II error if the critical value is set at 53.29 mpg?



21

Probability of Making A Type II Error (β) cont.

- Would commit a Type II error (β) by accepting the null hypothesis (μ = 50) when in truth it had changed
- If μ is actually moved from 50 to 54, what is probability of making a Type II error if the critical value is set at 53.29 mpg?
- $P(A_2)S_1 = \propto = 0.05$ (assume a one-tailed alternative hypothesis)
- $P(A_1)S_2 = \beta = ?$

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$ = 10/5 = 2 mpg

$Z = \frac{\bar{X}-\mu}{\sigma_{\bar{x}}}$= (53.29-54)/2 = -0.355

Use Exhibit D-1 (Areas of the Standard Normal Distribution)

Interpolate between 0.35 and 0.36 Z scores to find 0.355Z score.

The area between the mean and Z is 0.1387.

β is the tail area, or the area below the Z.

β = 0.5 – 0.1387 = 0.36 (There is a 36% probability of a Type II error if the μ is 54)

22

Probability of Making A Type II Error (β) cont.

- Power of the test = (1- β)
  - 1 – probability of commit a Type II error = (1- 0.36) = 64%
- We will correctly reject the false "null hypothesis" with a 64 % probability
- A power of 64% is less than the 80 percent min percentage recommended by statisticians

23

How to Reduce Type II Error (Increase power of test)

- Shift the Critical Value Closer to the Original μ

  $Z = (\overline{X}_C - \mu)/\sigma_{\overline{X}}$
- Increase Sample Size (from 25 to 100)

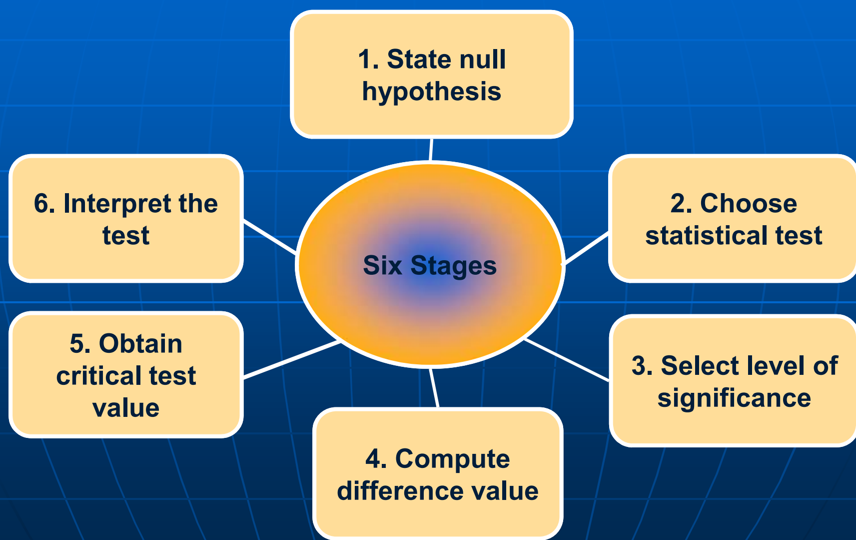
  $\sigma_{\overline{x}} = \sigma/\sqrt{n}$ = 10/10 = 1

  $Z = (\overline{X}_C - \mu)/\sigma_{\overline{X}}$ = (53.9 -54)/1 = -0.71

  β = 0.5 – 0.2612 = 0.24 (reduce type II error to 24%, power of test 76% < 80%)
- Improve both ∝ and β errors simultaneously

24

Statistical Testing Procedures



Probability Value (p values)

- Interpretation the Test
  - Method 1:
    - The conclusion is stated in terms of "Rejecting or Not Rejecting" the null hypothesis based on a reject region selected before the test is conducted.
  - Method 2:
    - Reports the extent to which the test statistics disagree with the null hypothesis.
    - Most statistical programs report the results of statistical tests as "probability value" or p-values.
    - Given that the null hypothesis is true, the p-value is the probability of observing a sample value as extreme as, or more extreme than the value actually observed.

26

Probability Value (p values)

- Interpretation the Test
  - Method 2:
    - This area represents the probability of a Type I error that must be assumed if the null hypothesis is rejected.
    - The p-value is compared to the significance level (α), and on this basis the null hypothesis is rejected or rejected.
    - If p-value < α "Reject the Null"
    - If p-value > α "Don't Reject the Null"

27

Probability Value (p values) – Exhibit 17-4

- Based on σ =10, n = 25 , 5% of risk, we compute the critical value $\bar{X}$= 53.29

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$ = 10/5 = 2 mpg

$Z = 1.96 = \frac{\bar{X}-\mu}{\sigma_{\bar{x}}}$= ($\overline{X_c}$-50)/2

$\overline{X_c}$ = 53.92

- Suppose the sample mean equals 55. Is there enough evidence to reject the null hypothesis?
  - If p-value < 0.05, the Null Hypothesis will be rejected

  $Z = 2.5 <= \frac{\bar{X}-\mu}{\sigma_{\bar{x}}}$= (55-50)/2

- The p-value is computed using the standard normal table:
  - The area between the mean and a Z value of 2.5 is 0.4938.
  - One-sided test: the p value is the area above the Z value.
  - The probability of observing a Z value at least as large as 2.5 is only 0.0062 or (0.5 – 0.4938) if the null hypothesis is true.

28

Probability Value (p values) – Exhibit 17-4

- This small p value represents the risk of rejecting a true null hypothesis.
- It is the probability of a Type I error if the null hypothesis is rejected.
- Since p-value (0.0062) < α = 0.05, the null hypothesis is rejected.
- The manufacturer can conclude that the average mpg has increased. The probability that this conclusion is wrong is 0.0062.

Summary