Inferential Statistics and Probability
a Holistic Approach

Chapter 8
Point Estimation and
Confidence Intervals

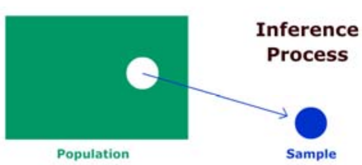
1

Inference Process

1. From a **Population**, take a sample.

2

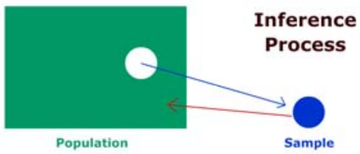
Inference Process



1. From a **Population**, take a sample.
2. Analyze the **Sample**.

3

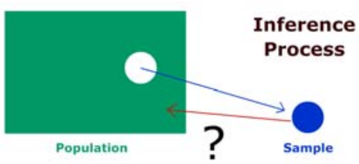
Inference Process



1. From a **Population**, take a sample.
2. Analyze the **Sample**.
3. Make an **Inference** about the Population based on the Sample.

4

Inference Process



1. From a **Population**, take a sample.
2. Analyze the **Sample**.
3. Make an **Inference** about the Population based on the Sample.
4. Measure the **Reliability** of the Inference.

5

Inferential Statistics

- Population Parameters
  - Mean = μ
  - Proportion = p
  - Standard Deviation = σ
- Sample Statistics
  - Mean = $\overline{X}$
  - Proportion = $\hat{p}$
  - Standard Deviation = s

6

Inferential Statistics

- Estimation
  - Using sample data to estimate population parameters.
  - Example: Public opinion polls
- Hypothesis Testing
  - Using sample data to make decisions or claims about population
  - Example: A drug effectively treats a disease

7

Estimation of μ

$\overline{X}$ is an unbiased point estimator of μ

Example: The number of defective items produced by a machine was recorded for five randomly selected hours during a 40-hour work week. The observed number of defectives were 12, 4, 7, 14, and 10. So the sample mean is 9.4.

Thus a point estimate for μ, the hourly mean number of defectives, is 9.4.

8

Confidence Intervals

- An Interval Estimate states the range within which a population parameter "probably" lies.
- The interval within which a population parameter is expected to occur is called a Confidence Interval.
- The distance from the center of the confidence interval to the endpoint is called the **"Margin of Error"**
- The three confidence intervals that are used extensively are the 90%, 95% and 99%.

9

Confidence Intervals

- A 95% confidence interval means that about 95% of the similarly constructed intervals will contain the parameter being estimated, or 95% of the sample means for a specified sample size will lie within 1.96 standard deviations of the hypothesized population mean.
- For the 99% confidence interval, 99% of the sample means for a specified sample size will lie within 2.58 standard deviations of the hypothesized population mean.
- For the 90% confidence interval, 90% of the sample means for a specified sample size will lie within 1.645 standard deviations of the hypothesized population mean.

10

90%, 95% and 99% Confidence Intervals for μ

- The 90%, 95% and 99% confidence intervals for μ are constructed as follows when $n \geq 30$
- 90% CI for the population mean is given by

$$\overline{X} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$

- 95% CI for the population mean is given by

$$\overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

- 99% CI for the population mean is given by

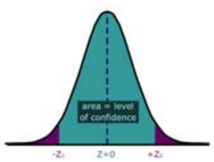
$$\overline{X} \pm 2.58 \frac{\sigma}{\sqrt{n}}$$

11

Constructing General Confidence Intervals for μ

- In general, a confidence interval for the mean is computed by:

$$\overline{X} \pm Z \frac{\sigma}{\sqrt{n}}$$



- This can also be thought of as:

Point Estimator ± Margin of Error

12

The nature of Confidence Intervals

- The Population mean μ is fixed.
- The confidence interval is centered around the sample mean which is a Random Variable.
- So the Confidence Interval (Random Variable) is like a target trying hit a fixed dart (μ).

13

EXAMPLE

- The Dean wants to estimate the mean number of hours worked per week by students. A sample of 49 students showed a mean of 24 hours with a standard deviation of 4 hours.
- The point estimate is 24 hours (sample mean).
- What is the 95% confidence interval for the average number of hours worked per week by the students?

14

EXAMPLE *continued*

- Using the 95% CI for the population mean, we have

$$24 \pm 1.96(4/7) = 22.88 \text{ to } 25.12$$

- The endpoints of the confidence interval are the confidence limits. The lower confidence limit is 22.88 and the upper confidence limit is 25.12

15

EXAMPLE *continued*

- Using the 99% CI for the population mean, we have

$$24 \pm 2.58(4/7) = 22.53 \ to \ 25.47$$

- Compare to the 95% confidence interval. A higher level of confidence means the confidence interval must be wider.

16

Selecting a Sample Size

- There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population. They are:
  - The degree of confidence selected.
  - The maximum allowable error.
  - The variation of the population.

17

Sample Size for the Mean

- A convenient computational formula for determining n is:

$$n = \left(\frac{Z\sigma}{E}\right)^2$$

- where E is the allowable error (margin of error), Z is the z score associated with the degree of confidence selected, and σ is the sample deviation of the pilot survey.
- σ can be estimated by past data, target sample or range of data.

18

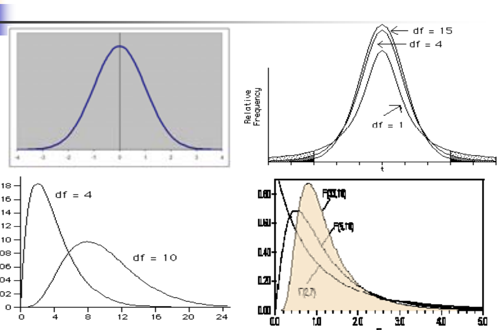
EXAMPLE

- A consumer group would like to estimate the mean monthly electric bill for a single family house in July. Based on similar studies the standard deviation is estimated to be $20.00. A 99% level of confidence is desired, with an accuracy of $5.00. How large a sample is required?

$$n = [(2.58)(20)/5]^2 = 106.5024 \approx 107$$

19

Normal Family of Distributions: Z, t, χ^2, F

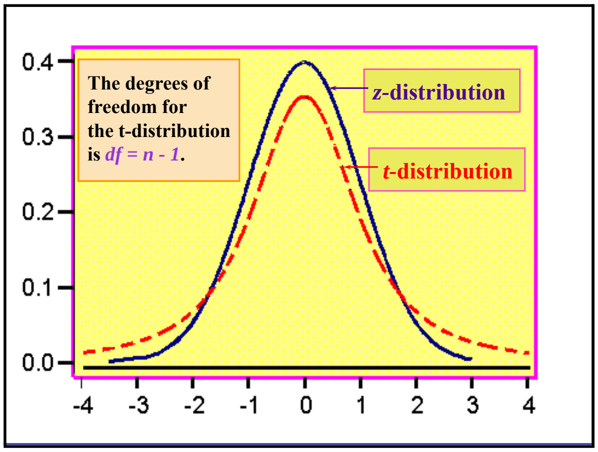


20

Characteristics of Student's *t*-Distribution

- The *t*-distribution has the following properties:
  - It is continuous, bell-shaped, and symmetrical about zero like the z-distribution.
  - There is a **family** of *t*-distributions sharing a mean of zero but having different standard deviations based on **degrees of freedom**.
  - The t-distribution is more spread out and flatter at the center than the z-distribution, but approaches the z-distribution as the sample size gets larger.

21



Confidence Interval for μ (σ unknown)

Formula to find a confidence interval using the t-distribution for the appropriate level of confidence:

$$\overline{X} \pm t\left(\frac{s}{\sqrt{n}}\right) \quad df = n-1$$

23

Example – Confidence Interval

- In a random sample of 13 American adults, the mean waste recycled per person per day was 5.3 pounds and the standard deviation was 2.0 pounds.
- Assume the variable is normally distributed and construct a 95% confidence interval for μ.

24

Example- Confidence Interval

level of confidence = 95%
df=13-1=12
t=2.18

$$5.3 \pm 2.18 \frac{2.0}{\sqrt{13}}$$

$$5.3 \pm 1.2 = (4.1, 6.5)$$

25

Confidence Intervals, Population Proportions

- Point estimate for proportion of successes in population is: $\hat{p} = \frac{X}{n}$
- X is the number of successes in a sample of size n.
- Standard deviation of $\hat{p}$ is $\sqrt{\frac{(p)(1-p)}{n}}$
- Confidence Interval for p:

$$\hat{p} \pm Z \cdot \sqrt{\frac{p(1-p)}{n}} \approx \hat{p} \pm Z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

26

Population Proportion Example

- In a May 2006 AP/ISPOS Poll, 1000 adults were asked if "Over the next six months, do you expect that increases in the price of gasoline will cause financial hardship for you or your family, or not?"
- 700 of those sampled responded yes!
- Find the **sample proportion** and **margin of error** for this poll. (This means find a 95% confidence interval.)

27

Population Proportion Example

- Sample proportion

$$\hat{p} = \frac{700}{1000} = .70 = 70\%$$

- Margin of Error

$$MOE = 1.96\sqrt{\frac{.70(1-.70)}{1000}} = .028 = 2.8\%$$

28

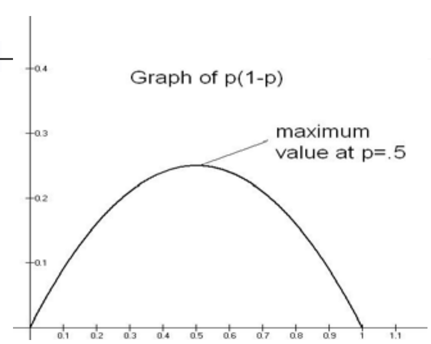
Sample Size for the Proportion

- A convenient computational formula for determining n is:

$$n = \left(p(1-p)\right)\left(\frac{Z}{E}\right)^2$$

- where E is the allowable margin of error, Z is the z-score associated with the degree of confidence selected, and p is the population proportion.
- If p is completely unknown, p can be set equal to ½ which maximizes the value of (p)(1-p) and guarantees the confidence interval will fall within the margin of error.

29



30

Example

- In polling, determine the minimum sample size needed to have a margin of error of 3% when p is unknown.

$$n = (.5)(1-.5)\left(\frac{1.96}{.03}\right)^2 = 1068$$

31

Example

- In polling, determine the minimum sample size needed to have a margin of error of 3% when p is known to be close to 1/4.

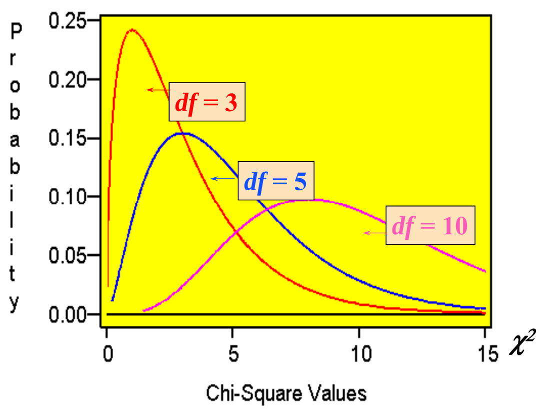
$$n = (.25)(1-.25)\left(\frac{1.96}{.03}\right)^2 = 801$$

32

Characteristics of the Chi-Square Distribution

- The major characteristics of the chi-square distribution are:
  - It is positively skewed
  - It is non-negative
  - It is based on degrees of freedom
  - When the degrees of freedom change, a new distribution is created

33



Inference about Population Variance and Standard Deviation

- s^2 is an unbiased point estimator for σ^2
- s is a point estimator for σ
- Interval estimates and hypothesis testing for both σ^2 and σ require a new distribution – the χ^2 (Chi-square)

35

Distribution of s^2

- $\dfrac{(n-1)s^2}{\sigma^2}$ has a chi-square distribution
- n-1 is degrees of freedom
- s^2 is sample variance
- σ^2 is population variance

36

Confidence interval for σ^2

- Confidence is **NO**T symmetric since chi-square distribution is not symmetric. You must find separate left and right bounds.
- We can construct a confidence interval for σ^2

$$\left(\frac{(n-1)s^2}{\chi_L^2}, \frac{(n-1)s^2}{\chi_R^2}\right)$$

- Take square root of both endpoints to get confidence interval for σ, the population standard deviation.

37

Example

- In performance measurement of investments, standard deviation is a measure of volatility or risk.
- Twenty monthly returns from a mutual fund show an average monthly return of 1% and a sample standard deviation of 5%
- Find a 95% confidence interval for the monthly standard deviation of the mutual fund.

38

Example (cont)

- df = n-1 =19
- 95% CI for σ

$$\left(\sqrt{\frac{(19)5^2}{32.8523}}, \sqrt{\frac{(19)5^2}{8.90655}}\right) = (3.8, 7.3)$$

39