Introduction to Estimation

Confidence Intervals
&
Sample Size

1

Disadvantage of Point Estimation

1. Provides Single Value
   Based on Observations from 1 Sample.
   * Sample Mean $\bar{X} = 98$ Is a Point Estimate of Unknown Population Mean.
2. Gives No Information about How Close Value Is to the Unknown Population Parameter

Which of the following statistics do you prefer?
a. 32%
b. 32% with a margin of error 3%

2

Estimation

You're interested in finding the average body temperature of healthy adults in Northeastern Ohio (the population). What would you do?

How can we estimate this average with a measure of reliability?

98 ± 1 F° | 98 ± .5 F° | 98 ± .2 F°

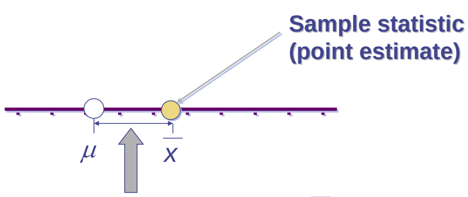
3

Interval Estimation

Margin of Error Gives Information about How Close Value Is to the Unknown Population Parameter.

4

Sampling Error

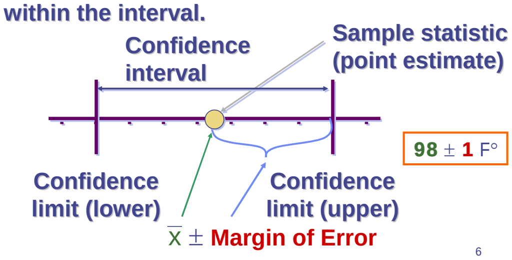


Sampling Error = $|\mu - \bar{x}|$

5

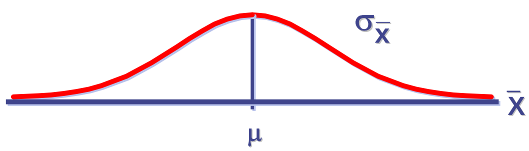
Key Elements of Interval Estimation

<u>Confidence Level</u>: A probability that the population parameter falls somewhere within the interval.



6

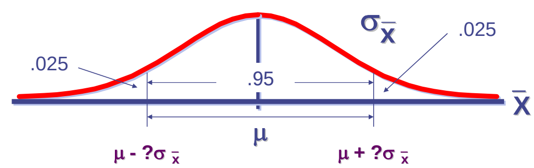
Sampling Distribution of the Mean



The sampling distribution is normal when **sampled from normally distributed population** or **having a relatively large sample**.

7

Sampling Distribution of the Mean



Within how many standard deviations of the mean will have 95% of the sampling distribution?

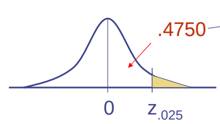
8

A Special Notation

z_α = the z score that the proportion of the standard normal distribution to the right of it is α.

$z_{.025} = 1.96$

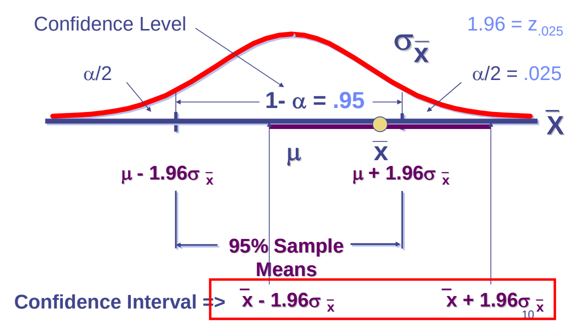
$z_{.010} = ?$



Z	.05	.06	.07
1.8	.4678	.4686	.4693
1.9	.4744	.4750	.4756
2.0	.4798	.4803	.4808
2.1	.4842	.4846	.4850

9

The Confidence Interval



10

Confidence Interval for Mean (σ Known)

- (1-α)·100% Confidence Interval Estimate for mean of a **normal** population

$$\left(\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \, , \, \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

or

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Margin of Error

"σ Known" may mean that we have very good estimate of σ.
It is not practical to assume that we know σ.

11

Confidence Interval of Mean (σ unKnown and n ≥ 30)

- (1-α)·100% Confidence Interval Estimate for mean of a population when sample size is relative large

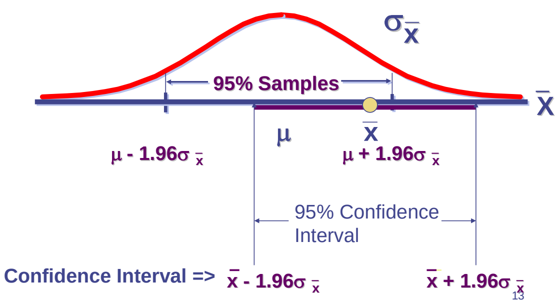
$$\left(\bar{X} - Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \, , \, \bar{X} + Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \right)$$

or

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

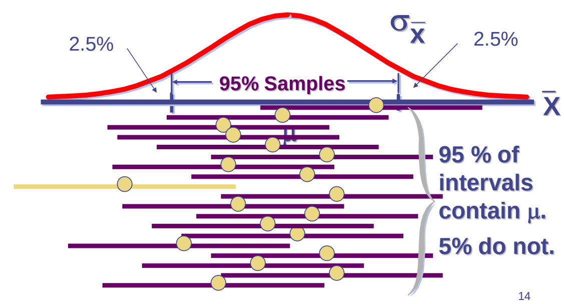
12

The Confidence Interval



13

The Confidence Interval



14

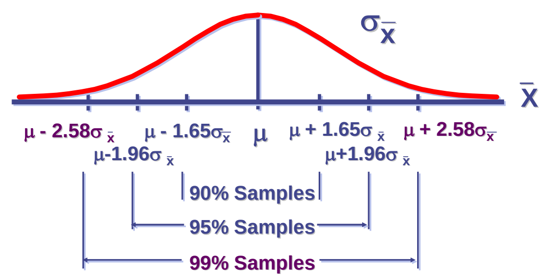
Factors Affecting Interval Width

$$\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \;,\; \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

1. Data Dispersion
   Measured by σ
2. Sample Size
   Affects standard error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
3. Level of Confidence $(1 - \alpha)$
   Affects $Z_{\alpha/2}$

15

Size of Interval



16

Thinking Challenge

- You're a Q/C inspector for Coca Cola. A random sample of **100** bottles showed **X = 1.99** liters and **s** = 0.**05** liters. What is the **90%** confidence interval estimate of the true **mean** amount in 2-liter bottles?

17

Confidence Interval Solution*

$$1-\alpha = .90,\ \alpha = 1-.90 = .1,\ \alpha/2 = .05$$

$$Z_{\alpha/2} = Z_{.05} = 1.64$$

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$1.99 \pm 1.64 \cdot \frac{.05}{\sqrt{100}} \Rightarrow 1.99 \pm 0.008$$

$$\Rightarrow (\ 1.982\,,\ 1.998\ )$$

18

Interpretation

◈We can be 95% confident that the population mean is in (1.982, 1.998).

◈We can be 95% confident that the maximum sampling error using 1.99 for estimating the population mean is within 0.008.

19

Thinking Challenge

Example: A city uses a certain noise index to monitor the noise pollution at a certain area of the city. A random sample of **100** observations from randomly selected days around noon showed an **average index** value of **$\bar{x}$ = 1.99** and **standard deviation s = 0.05**. Find the **90%** confidence interval estimate of the **average** noise index at noon.

20

Confidence Interval Solution*

$$1-\alpha = .90,\ \ \alpha = 1-.90 = .1,\ \ \alpha/2 = .05$$

$$Z_{\alpha/2} = Z_{.05} = 1.64$$

$$\bar{X} \pm Z_{\alpha/2}\cdot\frac{s}{\sqrt{n}}$$

$$1.99 \pm 1.64\cdot\frac{.05}{\sqrt{100}} \Rightarrow 1.99 \pm 0.008$$

$$\Rightarrow (\ 1.982\ ,\ 1.998\ )$$

21

Finding Sample Sizes for Estimating μ

C.I.: $\bar{x} \pm z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$

Margin of Error $= B = Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$

$$n = \frac{z^2_{\alpha/2}\cdot\sigma^2}{B^2}$$

***B* = Margin of Error or Bound**

22

Sample Size Example

What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

$$n = \frac{Z_{.05}{}^2\sigma^2}{B^2} = \frac{(1.645)^2(45)^2}{(5)^2} = 219.2 \cong 220$$

23

Thinking Challenge

◈You plan to survey residents in your county to find the average health insurance premium that they are paying. You want to be **95%** confident that the sample **mean** is within ± **$50**.
A pilot study showed that σ was about **$400**. What **sample size** should you use?

24

Sample Size Solution*

$$n = \frac{Z_{0.025}^{2}\sigma^2}{B^2}$$

$$= \frac{(1.96)^2(400)^2}{(50)^2}$$

$$= 245.86 \cong 246$$

25

Confidence Interval of Mean (σ unKnown and n ≥ 30)

- $(1-\alpha)\cdot 100\%$ Confidence Interval Estimate for mean of a population when sample size is relative large

$$\left(\bar{X} - Z_{\alpha/2}\cdot\frac{s}{\sqrt{n}} \;,\; \bar{X} + Z_{\alpha/2}\cdot\frac{s}{\sqrt{n}} \right)$$

or

Review

$$\bar{X} \pm Z_{\alpha/2}\cdot\frac{s}{\sqrt{n}}$$

26

Confidence Interval Mean (σ Unknown & n< 30)

1. Assumptions
   Population Standard Deviation Is Unknown
   Population Must Be **Normally Distributed**
2. Use Student's t Distribution
3. Confidence Interval Estimate

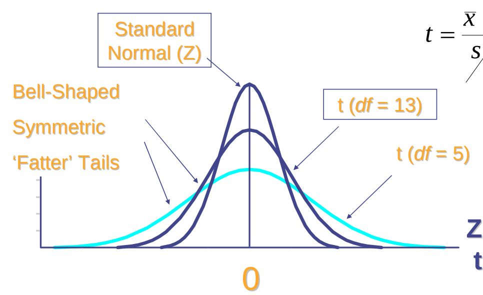
$$\left(\bar{X} - t_{\alpha/2, n-1}\cdot\frac{S}{\sqrt{n}} ,\; \bar{X} + t_{\alpha/2, n-1}\cdot\frac{S}{\sqrt{n}} \right)$$

$$\bar{X} \pm t_{\alpha/2,\, n-1}\cdot\frac{S}{\sqrt{n}}$$

27

Student's t Distribution

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

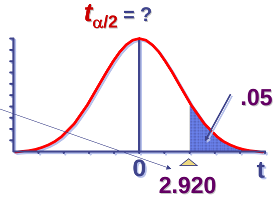


28

Student's t Table

df	Confidence interval	80%	90%	95%
	one tail	.10	.05	.025
	two tail	.20	.10	.05
1		3.078	6.314	12.706
2		1.886	2.920	4.303
3		1.638	2.353	3.182

For a 90% C.I.:
***n* = 3**
***df* = *n* - 1 = 2**
α = .10
α/2 =.05
$t_{\alpha/2}$ **= ?**



t values

29

Estimation Example Mean (σ Unknown)

A random sample of weights of 25 subjects, has a sample mean 140 and sample standard deviation 8. Set up a 95% confidence interval estimate for μ.

$$1-\alpha = .95, \quad \alpha = 1-.95 = .05, \quad \alpha/2 = .025,$$

$$t_{\alpha/2,\, df=24} = t_{0.025} = 2.064$$

$$\bar{X} \pm t_{\alpha/2, n-1}\cdot\frac{S}{\sqrt{n}}$$

$$140 \pm 2.064\cdot\frac{8}{\sqrt{25}} \Rightarrow 140 \pm 3.31$$

$$\Rightarrow (136.69 \,,\, 143.31\;)$$

30

Thinking Challenge

The numbers of community hospital beds per 1000 population that are available in each different regions of the country is normally distributed. A random sample 6 regions were selected and the rates of beds per 1000 were recorded and they are

3.6, 4.2, 4.0, 3.5, 3.8, 3.1.

Find the **90%** confidence interval estimate of the **mean** bed-rate in the country.

31

Confidence Interval Solution*

$$\bar{X} \pm t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}$$

$\bar{X} = 3.7$

$s = 0.38987$ $\quad n = 6, \quad \text{df} = n - 1 = 6 - 1 = 5$

$1-\alpha = .90, \quad \alpha = 1 - .90 = .1, \quad \alpha/2 = .05,$

$t_{\alpha/2, df=5} = t_{0.05} = 2.015$

$$3.7 \pm 2.015 \cdot \frac{.38987}{\sqrt{6}} \Rightarrow 3.7 \pm .321$$

$(3.379, 4.021)$

32

Confidence interval with *z*-score:

The $(1-\alpha)\%$ confidence interval estimate for population mean:

- Assumption: If sampled from normal population with known variance, σ,

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Assumption: If large sample and if unknown variance, s replaces σ,

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

33

Confidence interval with *t*-score:

The $(1-\alpha)\%$ confidence interval estimate for population mean:

Assumption: If sampled from normal population with unknown variance, σ,

$$\bar{x} \pm t_{\alpha/2, df=n-1} \cdot \frac{s}{\sqrt{n}}$$

(If sample size is large the normality assumption is insignificant.)

$t \to z$ *as sample becomes large*

34

Average Weight for Female Ten Year Children In US

Info. from a random sample: $n = 10$, $\bar{x} = 80$ lb, $s = 18.05$ lb, assume weight is **normally distributed**, find the 95% confidence interval estimate for average weight.

Data:

73.80	50.00	101.40	67.20	102.20
97.80	81.00	93.40	63.20	70.00

How do we know whether normality assumption is OK?

35

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ght (pounds) of particip	.171	10	.200*	.930	10	.452

. This is a lower bound of the true significance.

. Lilliefors Significance Correction

Both are greater than 0.05, normality assumption is acceptable.

36

Average Weight for Female Ten Year Children In US

Info. from a random sample: n = 10, $\bar{x}$ = 80 lb, s = 18.05 lb, assume weight is normally distributed, find the 95% confidence interval estimate for average weight.

$$t_{\alpha/2} = t_{.05/2} = t_{0.025}, \quad \text{d.f.} = 10 - 1 = 9, \quad t_{0.025,\,9} = 2.262$$

$$\bar{x} \pm t_{\alpha/2, df=9} \cdot \frac{s}{\sqrt{n}} \Rightarrow 80 \pm 2.262 \cdot \frac{18.05}{\sqrt{10}}$$

$$80 \pm 12.91 \quad \Rightarrow (67.09,\ 92.91)$$

37

Weight for Ten Year Old

Descriptives 80 ± 12.91

	What is your sex?			Statistic	Std. Error
nt (pounds) rticipant	female	Mean		80.0000	5.70840
		95% Confidence Interval for Mean	Lower Bound	67.0867	
			Upper Bound	92.9133	
		5% Trimmed Mean		80.4333	
		Median		77.4000	
		Variance		325.858	
		Std. Deviation		18.05153	
		Minimum		50.00	
		Maximum		102.20	
		Range		52.20	
		Interquartile Range		32.5000	
		Skewness		-.148	.687
		Kurtosis		-1.229	1.334
	male	Mean		86.8600	3.96048
		95% Confidence Interval for Mean	Lower Bound	77.9008	
			Upper Bound	95.8192	

Confidence Interval Estimate of Proportion

39

Proportion Estimation

Parameter: Population Proportion p (or π)

(Percentage of people has no health insurance)

Statistic: Sample Proportion $\hat{p} = \frac{x}{n}$

x is number of successes

n is sample size

Remark: If data is coded as 1 or 0, sample mean is the same as sample proportion of 1's.

Data: 1, 0, 0, 1, 0 $\Rightarrow \bar{x} = \frac{1+0+0+1+0}{5} = \frac{2}{5} = .2 = p$

40

Confidence Interval Proportion

1. Assumptions
   - Two Categorical Outcomes (Binomial Exp.)
   - Normal Approximation Can Be Used If
     - $n\hat{p} \pm 3\sqrt{n\hat{p}(1-\hat{p})}$ Does Not Include 0 or 1
     - or expect to see at least 5 counts in each category.
2. Confidence Interval Estimate (For large sample)

$$\left(\ \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\ \right)$$

$$1 - \hat{p} = \hat{q}$$

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$$

41

Estimation Example Proportion

A random sample of 400 from a large community showed that 32 have diabetes. Set up a 95% confidence interval estimate for p, the percentage of people that have diabetes.

$$\hat{p} = \frac{32}{400} = .08, \quad n = 400, \quad z_{\alpha/2} = z_{.025} = 1.96$$

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} \qquad .08 \pm 1.96 \cdot \sqrt{\frac{.08\cdot(1-.08)}{400}}$$

$$.08 \pm .027 \ \Rightarrow 8\% \pm 2.7\%$$

$$\Rightarrow (5.3\%, 10.7\%)$$

42

Thinking Challenge

A member of a health department wish to see what percentage of people in a community will support an environmental policy. Of **200** survey forms sent and received, **35** responded that they support the policy and the rest of them do not support the policy.

Find a **90%** confidence interval estimate of the percentage of the population in this community that support t**he policy**?

43

Confidence Interval Solution*

$$\hat{p} = \frac{35}{200} = .175,\ n = 200,\ z_{\alpha/2} = 1.645$$

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

$$.175 \pm 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}}$$

$$.175 \pm .0442 = 17.5\% \pm 4.42\%$$

$$= (13.08\%,\ 21.92\%)$$

44

Example:

Researchers wish to estimate the percentage of hospital employees infected by SARS in a certain country. Out of 500 randomly chosen hospital employees, 14 were infected. Find the 95% confidence interval estimate for percentage of hospital employees infected by SARS in this country.

45

Sample Size

C.I. : $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$

Margin of Error $= B = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$

$n = \frac{z_{\alpha/2}^2}{B^2} \cdot \hat{p} \cdot (1-\hat{p})$ if pilot study is done.

or

$n = \frac{z_{\alpha/2}^2}{B^2} \cdot 0.25$ to get the largest sample to achieve the goal.

46

Sample Size (No prior information on *p*)

Sample Size Example: If one wishes to do a survey to estimate the population proportion with 95% confidence and a margin of error of 3%, how large a sample is needed?

$Z_{\alpha/2} = 1.96$; $B = .03$

$n = (1.96^2/.03^2) \times .25 = 1067.11$

A sample of size 1068 is needed.

47

Sample Size (With prior information on *p*)

Sample Size Example: If one wishes to to estimate the percentage of people infected with West Nile in a population with 95% confidence and a margin of error of 3%, how large a sample is needed? (A pilot study has been done, and the sample proportion was 6%.)

$Z_{\alpha/2} = 1.96$; $B = .03$

$n = (1.96^2/.03^2) \times .06 \times (1 - .06) = 240.7$

A sample of size 241 is needed.

48