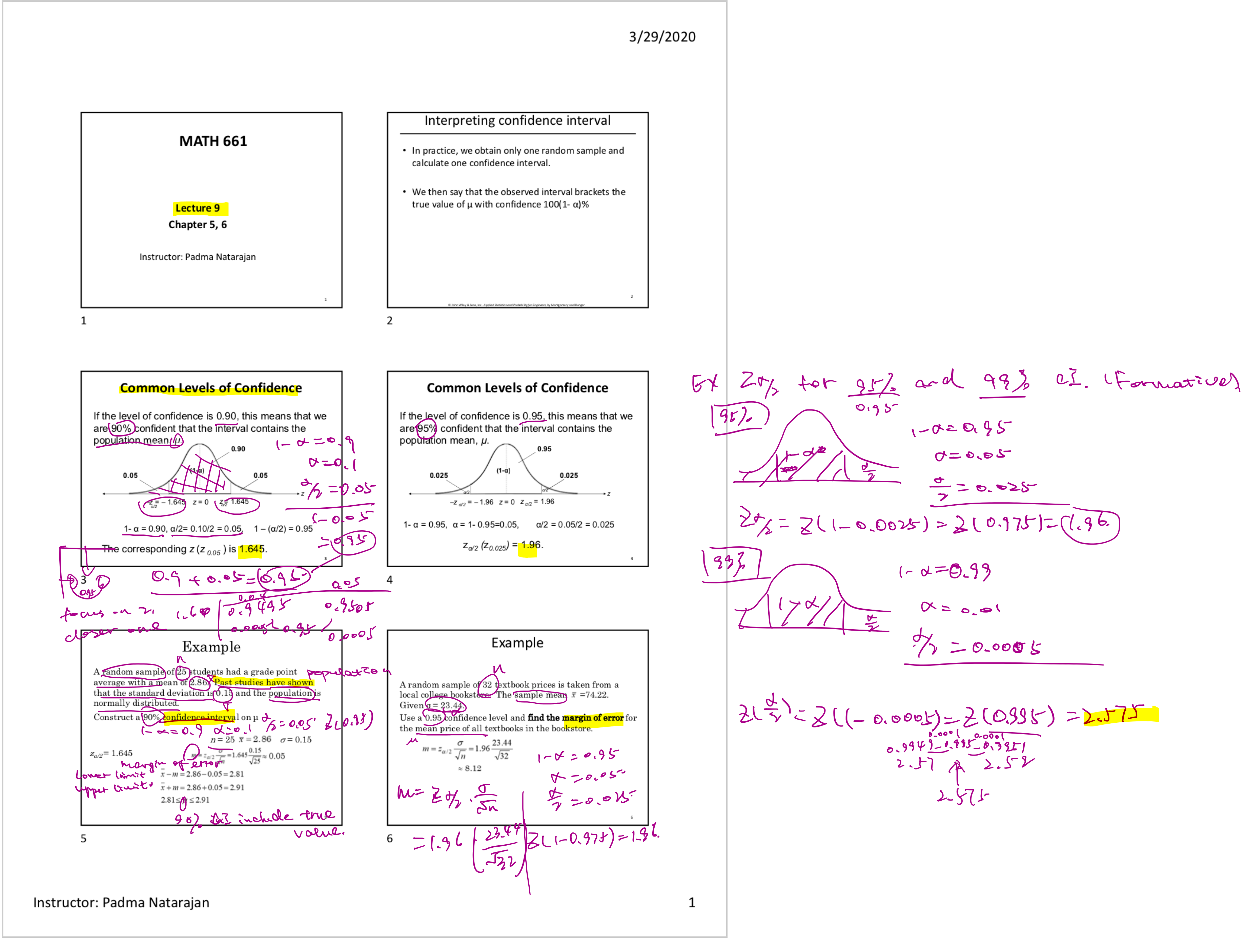
L9.slides

2020年4月2日 星期四 下午5:38

Lecture9_print



MATH 661

Lecture 9
Chapter 5, 6

Instructor: Padma Natarajan

1

Interpreting confidence interval

- In practice, we obtain only one random sample and calculate one confidence interval.
- We then say that the observed interval brackets the true value of μ with confidence 100(1- α)%

2

Common Levels of Confidence

If the level of confidence is 0.90, this means that we are 90% confident that the interval contains the population mean μ.

1- α = 0.90, α/2= 0.10/2 = 0.05, 1 – (α/2) = 0.95

The corresponding z ($z_{0.05}$) is 1.645.

1 − α = 0.9
α = 0.1
α/2 = 0.05
1 − 0.05 = 0.95

0.9 + 0.05 = 0.95

focus on z, closer one
1.64 | 0.9495 0.9505
avoid 0.95 / 0.0005

3

Common Levels of Confidence

If the level of confidence is 0.95, this means that we are 95% confident that the interval contains the population mean, μ.

1- α = 0.95, α = 1- 0.95=0.05, α/2 = 0.05/2 = 0.025

$z_{\alpha/2}$ ($z_{0.025}$) = 1.96.

4

Example

A random sample of 25 students had a grade point average with a mean of 2.86. Past studies have shown that the standard deviation is 0.15 and the population is normally distributed.

Construct a 90% confidence interval on μ

n = 25 $\bar{x}$ = 2.86 σ = 0.15

$z_{\alpha/2}$= 1.645

$m = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.645\frac{0.15}{\sqrt{25}} \approx 0.05$

$\bar{x} - m = 2.86 - 0.05 = 2.81$

$\bar{x} + m = 2.86 + 0.05 = 2.91$

$2.81 \le \mu \le 2.91$

n, population, 1 − α = 0.9, α = 0.1, α/2 = 0.05, z(0.95)
margin of error
Lower limit
upper limit
90% CI include true value.

5

Example

A random sample of 32 textbook prices is taken from a local college bookstore. The sample mean $\bar{x}$ =74.22. Given σ = 23.44.

Use a 0.95 confidence level and **find the margin of error** for the mean price of all textbooks in the bookstore.

$m = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\frac{23.44}{\sqrt{32}}$
≈ 8.12

1 − α = 0.95
α = 0.05
α/2 = 0.025

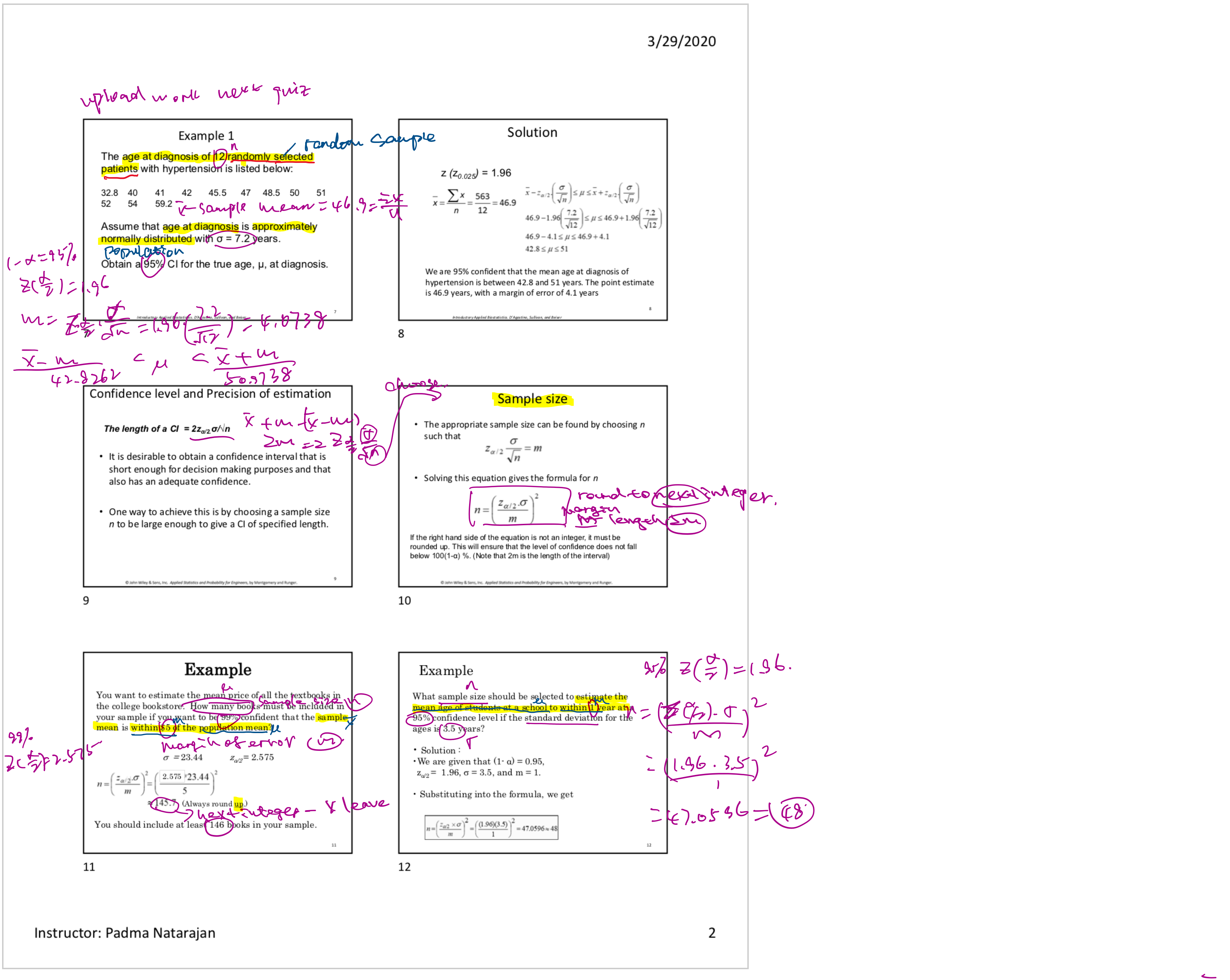
$m = z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$
$= 1.96\left(\frac{23.44}{\sqrt{32}}\right)$ | $z(1-0.975) = 1.96$

6

EX $z_{\alpha/2}$ for 95% and 99% CI. (Formative)

95% (0.95)
1 − α = 0.95
α = 0.05
α/2 = 0.025
$z_{\alpha/2} = z(1-0.025) = z(0.975) = 1.96$

99%
1 − α = 0.99
α = 0.01
α/2 = 0.0005
$z(\frac{\alpha}{2}) = z(1-0.0005) = z(0.995) = 2.575$
0.9949 0.995 0.9951
2.57 2.58
2.575



upload work next quiz

Example 1

The age at diagnosis of 12 randomly selected patients with hypertension is listed below:

32.8 40 41 42 45.5 47 48.5 50 51
52 54 59.2

Assume that age at diagnosis is approximately normally distributed with σ = 7.2 years.

Obtain a 95% CI for the true age, μ, at diagnosis.

random sample
$\bar{x}$ sample mean = 46.9 = $\frac{\sum x}{n}$
population
1 − α = 95%
$z(\frac{\alpha}{2}) = 1.96$
$m = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 1.96\left(\frac{7.2}{\sqrt{12}}\right) = 4.0738$
$\bar{x} - m < \mu < \bar{x} + m$
42.8262 50.9738

7

Solution

z ($z_{0.025}$) = 1.96

$\bar{x} = \frac{\sum x}{n} = \frac{563}{12} = 46.9$

$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$46.9 - 1.96\left(\frac{7.2}{\sqrt{12}}\right) \le \mu \le 46.9 + 1.96\left(\frac{7.2}{\sqrt{12}}\right)$

$46.9 - 4.1 \le \mu \le 46.9 + 4.1$

$42.8 \le \mu \le 51$

We are 95% confident that the mean age at diagnosis of hypertension is between 42.8 and 51 years. The point estimate is 46.9 years, with a margin of error of 4.1 years

8

Confidence level and Precision of estimation

The length of a CI **= $2z_{\alpha/2}\sigma/\sqrt{n}$**

- It is desirable to obtain a confidence interval that is short enough for decision making purposes and that also has an adequate confidence.
- One way to achieve this is by choosing a sample size n to be large enough to give a CI of specified length.

$\bar{x} \pm m$ $(\bar{x}+m) - (\bar{x}-m)$
$2m = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

9

choose

Sample size

- The appropriate sample size can be found by choosing n such that

$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = m$

- Solving this equation gives the formula for n

$n = \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2$

round to next integer.
margin
length 2m

If the right hand side of the equation is not an integer, it must be rounded up. This will ensure that the level of confidence does not fall below 100(1-α) %. (Note that 2m is the length of the interval)

10

Example

You want to estimate the mean price of all the textbooks in the college bookstore. How many books must be included in your sample if you want to be 99% confident that the sample mean is within $5 of the population mean?

σ =23.44 $z_{\alpha/2}$= 2.575

$n = \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2 = \left(\frac{2.575 \cdot 23.44}{5}\right)^2$

= 145.7 (Always round up.)

You should include at least 146 books in your sample.

sample size (n)
margin of error (m)
99% $z(\frac{\alpha}{2}) = 2.575$
next integer − & leave

11

Example

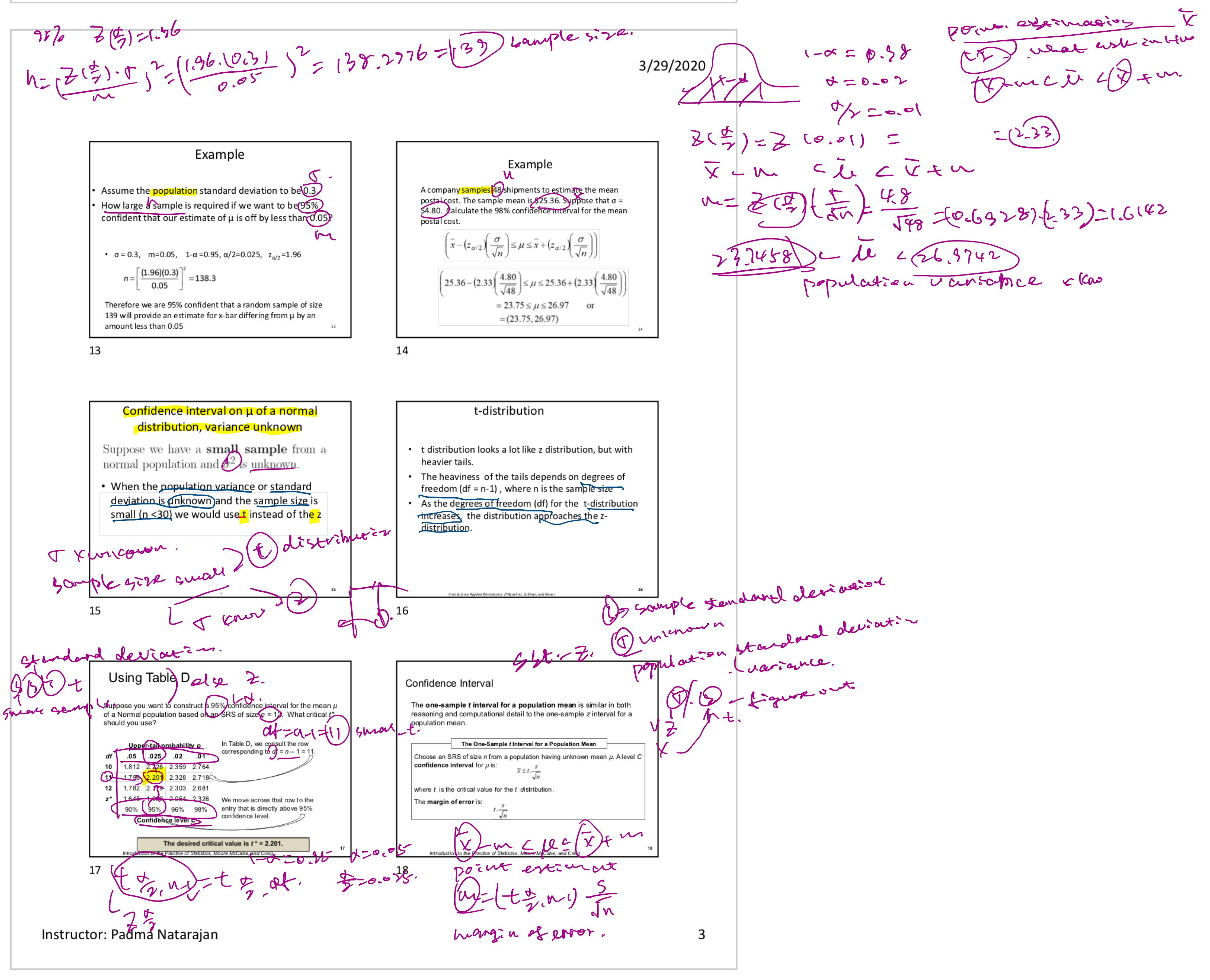
What sample size should be selected to estimate the mean age of students at a school to within 1 year at 95% confidence level if the standard deviation for the ages is 3.5 years?

- Solution :
- We are given that (1- α) = 0.95, $z_{\alpha/2}$= 1.96, σ = 3.5, and m = 1.
- Substituting into the formula, we get

$n = \left(\frac{z_{\alpha/2} \times \sigma}{m}\right)^2 = \left(\frac{(1.96)(3.5)}{1}\right)^2 = 47.0596 \approx 48$

95% $z(\frac{\alpha}{2}) = 1.96$.
$n = \left(\frac{z(\frac{\alpha}{2})\cdot\sigma}{m}\right)^2$
$= \left(\frac{1.96\cdot 3.5}{1}\right)^2$
$= 47.0596 = 48$

12



95% $z(\frac{\alpha}{2}) = 1.96$
$n = \left(\frac{z(\frac{\alpha}{2})\cdot\sigma}{m}\right)^2 = \left(\frac{1.96\cdot(0.3)}{0.05}\right)^2 = 138.2976 = 139$ sample size.

1 − α = 0.98
α = 0.02
α/2 = 0.01
$z(\frac{\alpha}{2}) = z(0.01) = (2.33)$
$\bar{x} - m < \mu < \bar{x} + m$
$m = z(\frac{\alpha}{2})\left(\frac{\sigma}{\sqrt{n}}\right) = \frac{4.8}{\sqrt{48}} = (0.6928)(2.33) = 1.6142$
23.7458 < μ < 26.9742
population variance known

point estimation $\bar{x}$
(μ) what ask in the
$\bar{x} - m < \mu < \bar{x} + m$

Example

- Assume the population standard deviation to be 0.3
- How large a sample is required if we want to be 95% confident that our estimate of μ is off by less than 0.05?

- σ = 0.3, m=0.05, 1-α =0.95, α/2=0.025, $z_{\alpha/2}$=1.96

$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3$

Therefore we are 95% confident that a random sample of size 139 will provide an estimate for x-bar differing from μ by an amount less than 0.05

13

Example

A company samples 48 shipments to estimate the mean postal cost. The sample mean is $25.36. Suppose that σ = $4.80. Calculate the 98% confidence interval for the mean postal cost.

$\left(\bar{x} - (z_{\alpha/2})\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + (z_{\alpha/2})\left(\frac{\sigma}{\sqrt{n}}\right)\right)$

$\left(25.36 - (2.33)\left(\frac{4.80}{\sqrt{48}}\right) \le \mu \le 25.36 + (2.33)\left(\frac{4.80}{\sqrt{48}}\right)\right)$

$= 23.75 \le \mu \le 26.97$ or
$= (23.75, 26.97)$

14

Confidence interval on μ of a normal distribution, variance unknown

Suppose we have a **small sample** from a normal population and σ² is unknown.

- When the population variance or standard deviation is unknown and the sample size is small (n <30) we would use t instead of the z

σ unknown. sample size small → (t) distribution
σ know → (z)

15

t-distribution

- t distribution looks a lot like z distribution, but with heavier tails.
- The heaviness of the tails depends on degrees of freedom (df = n-1) , where n is the sample size
- As the degrees of freedom (df) for the t-distribution increases, the distribution approaches the z-distribution.

16

(s) sample standard deviation
(σ) unknown population standard deviation (variance).

standard deviation (s) t
small sample

Using Table D

else z.

Suppose you want to construct a 95% confidence interval for the mean μ of a Normal population based on an SRS of size n = 12. What critical t* should you use?

df	Upper-tail probability p .05	.025	.02	.01
10	1.812	2.228	2.359	2.764
11	1.796	2.201	2.328	2.718
12	1.782	2.179	2.303	2.681
z*	1.645	1.960	2.054	2.326
	90%	95%	96%	98%
	Confidence level C			

In Table D, we consult the row corresponding to df = n – 1 = 11.

We move across that row to the entry that is directly above 95% confidence level.

The desired critical value is t * = 2.201.

Introduction to the Practice of Statistics, Moore McCabe, and Craig

df = n − 1 = 11 small t.
1 − α = 0.95 α = 0.05 α/2 = 0.025
$t_{\frac{\alpha}{2}, n-1} = t_{\frac{\alpha}{2}, df}$
$z_{\frac{\alpha}{2}}$

17

s, t ← z.

Confidence Interval

The **one-sample t interval for a population mean** is similar in both reasoning and computational detail to the one-sample z interval for a population mean.

The One-Sample t Interval for a Population Mean

Choose an SRS of size n from a population having unknown mean μ. A level C **confidence interval** for μ is:

$\bar{x} \pm t^*\frac{s}{\sqrt{n}}$

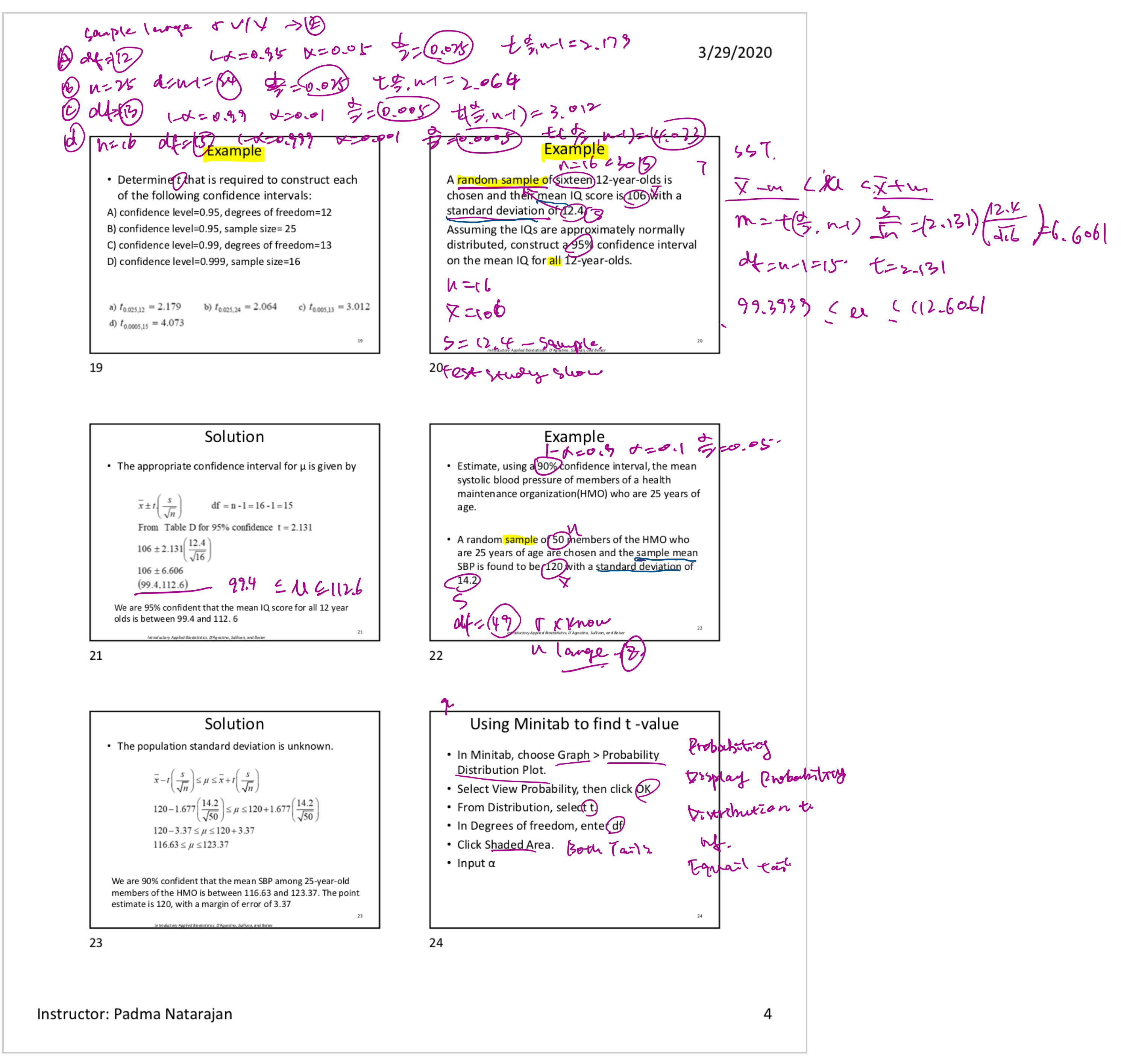
where t* is the critical value for the t distribution.

The **margin of error** is: $t^*\frac{s}{\sqrt{n}}$

Introduction to the Practice of Statistics, Moore McCabe, and Craig

σ, s − figure out n, t.
$\bar{x} - m < \mu < \bar{x} + m$
point estimate
$m = (t_{\frac{\alpha}{2}, n-1})\frac{s}{\sqrt{n}}$
margin of error.

18



sample large σ U/X → (z)
(a) df = (12) 1 − α = 0.95 α = 0.05 α/2 = (0.025) $t_{\frac{\alpha}{2}, n-1} = 2.179$
(b) n = 25 df = n − 1 = (24) α/2 = (0.025) $t_{\frac{\alpha}{2}, n-1} = 2.064$
(c) df = (13) 1 − α = 0.99 α = 0.01 α/2 = (0.005) $t_{\frac{\alpha}{2}, n-1} = 3.012$
(d) n = 16 df = (15) 1 − α = 0.999 α = 0.001 α/2 = (0.0005) $t_{\frac{\alpha}{2}, n-1} = (4.073)$

Example

- Determine t that is required to construct each of the following confidence intervals:

A) confidence level=0.95, degrees of freedom=12
B) confidence level=0.95, sample size= 25
C) confidence level=0.99, degrees of freedom=13
D) confidence level=0.999, sample size=16

a) $t_{0.025,12} = 2.179$ b) $t_{0.025,24} = 2.064$ c) $t_{0.005,13} = 3.012$
d) $t_{0.0005,15} = 4.073$

19

Example

n = 16 < 30 (t)

A random sample of sixteen 12-year-olds is chosen and their mean IQ score is 106 with a standard deviation of 12.4.

Assuming the IQs are approximately normally distributed, construct a 95% confidence interval on the mean IQ for all 12-year-olds.

n = 16
$\bar{x}$ = 106
s = 12.4 − sample

20 test study show

SST.
$\bar{x} - m < \mu < \bar{x} + m$
$m = t(\frac{\alpha}{2}, n-1)\frac{s}{\sqrt{n}} = (2.131)\left(\frac{12.4}{\sqrt{16}}\right) = 6.6061$
df = n − 1 = 15. t = 2.131
99.3939 < μ < 112.6061

Solution

- The appropriate confidence interval for μ is given by

$\bar{x} \pm t\left(\frac{s}{\sqrt{n}}\right)$ df = n -1 = 16 -1 = 15

From Table D for 95% confidence t = 2.131

$106 \pm 2.131\left(\frac{12.4}{\sqrt{16}}\right)$

106 ± 6.606

(99.4, 112.6) 99.4 ≤ μ ≤ 112.6

We are 95% confident that the mean IQ score for all 12 year olds is between 99.4 and 112. 6

21

Example

1 − α = 0.9 α = 0.1 α/2 = 0.05.

- Estimate, using a 90% confidence interval, the mean systolic blood pressure of members of a health maintenance organization(HMO) who are 25 years of age.
- A random sample of 50 members of the HMO who are 25 years of age are chosen and the sample mean SBP is found to be 120 with a standard deviation of 14.2

n, $\bar{x}$, s
df = (49) σ X know
n large (z)

22

Solution

- The population standard deviation is unknown.

$\bar{x} - t\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t\left(\frac{s}{\sqrt{n}}\right)$

$120 - 1.677\left(\frac{14.2}{\sqrt{50}}\right) \le \mu \le 120 + 1.677\left(\frac{14.2}{\sqrt{50}}\right)$

$120 - 3.37 \le \mu \le 120 + 3.37$

$116.63 \le \mu \le 123.37$

We are 90% confident that the mean SBP among 25-year-old members of the HMO is between 116.63 and 123.37. The point estimate is 120, with a margin of error of 3.37

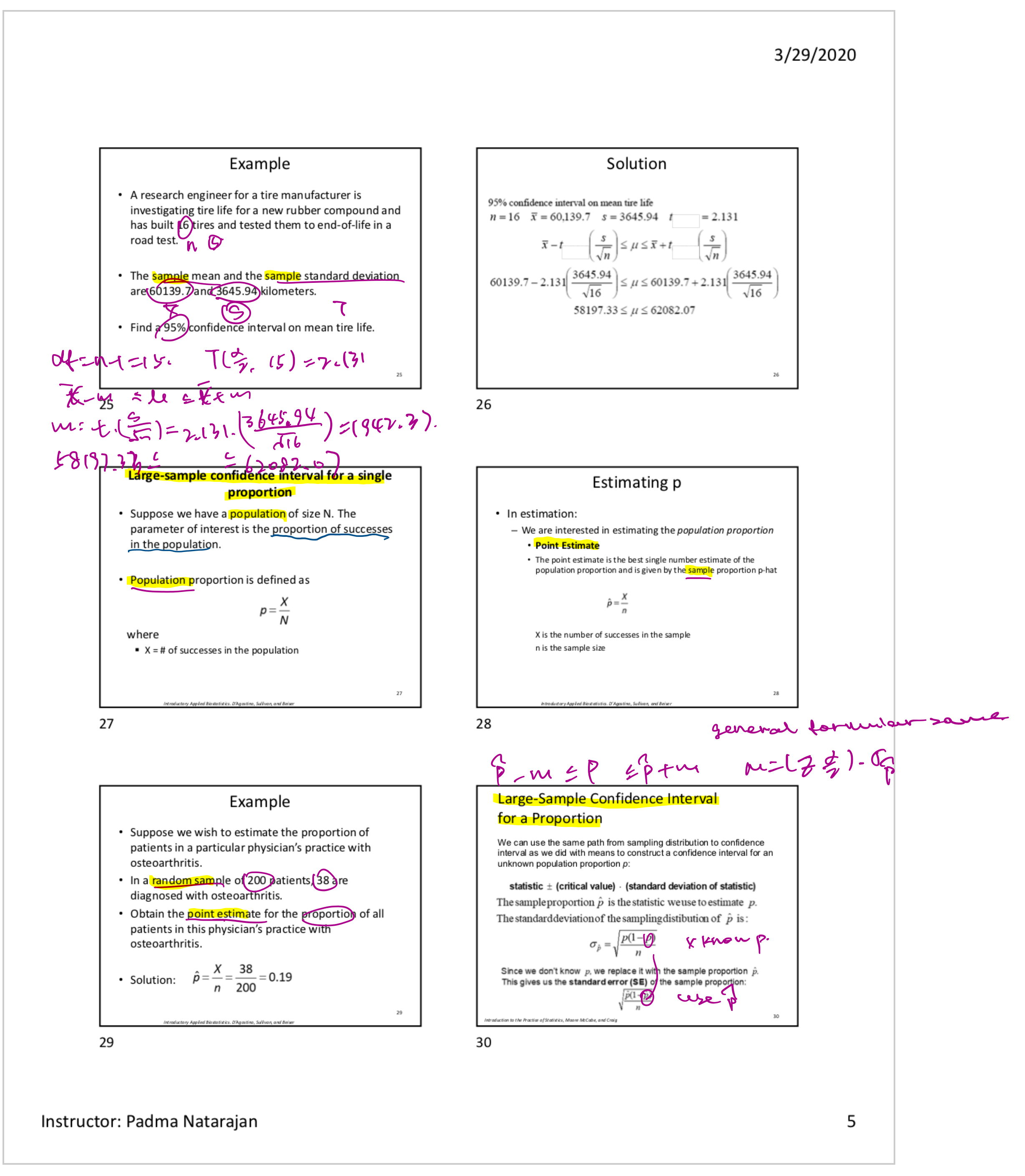
23

Using Minitab to find t -value

- In Minitab, choose Graph > Probability Distribution Plot.
- Select View Probability, then click OK
- From Distribution, select t
- In Degrees of freedom, enter df
- Click Shaded Area. Both Tails
- Input α

Probability
Display Probability
Distribution t
df.
Equal tail

24



Example

- A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test.
- The sample mean and the sample standard deviation are 60139.7 and 3645.94 kilometers.
- Find a 95% confidence interval on mean tire life.

n (16) $\bar{x}$ (s) T

df = n − 1 = 15. $T(\frac{\alpha}{2}, 15) = 2.131$
$\bar{x} - m \le \mu \le \bar{x} + m$
$m = t\left(\frac{s}{\sqrt{n}}\right) = 2.131\cdot\left(\frac{3645.94}{\sqrt{16}}\right) = (1942.3)$
58197.37 ≤ (62082.0)

25

Solution

95% confidence interval on mean tire life
n = 16 $\bar{x}$ = 60,139.7 s = 3645.94 t = 2.131

$\bar{x} - t\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t\left(\frac{s}{\sqrt{n}}\right)$

$60139.7 - 2.131\left(\frac{3645.94}{\sqrt{16}}\right) \le \mu \le 60139.7 + 2.131\left(\frac{3645.94}{\sqrt{16}}\right)$

$58197.33 \le \mu \le 62082.07$

26

Large-sample confidence interval for a single proportion

- Suppose we have a population of size N. The parameter of interest is the proportion of successes in the population.
- Population proportion is defined as

$p = \frac{X}{N}$

where
- X = # of successes in the population

27

Estimating p

- In estimation:
  - We are interested in estimating the *population proportion*
    - **Point Estimate**
    - The point estimate is the best single number estimate of the population proportion and is given by the sample proportion p-hat

$\hat{p} = \frac{X}{n}$

X is the number of successes in the sample
n is the sample size

28

Example

- Suppose we wish to estimate the proportion of patients in a particular physician's practice with osteoarthritis.
- In a random sample of 200 patients, 38 are diagnosed with osteoarthritis.
- Obtain the point estimate for the proportion of all patients in this physician's practice with osteoarthritis.
- Solution: $\hat{p} = \frac{X}{n} = \frac{38}{200} = 0.19$

29

general formula same
$\hat{p} - m \le p \le \hat{p} + m$ $m = (z_{\frac{\alpha}{2}})\cdot\sigma_{\hat{p}}$

Large-Sample Confidence Interval for a Proportion

We can use the same path from sampling distribution to confidence interval as we did with means to construct a confidence interval for an unknown population proportion p:

statistic ± (critical value) · (standard deviation of statistic)

The sample proportion $\hat{p}$ is the statistic we use to estimate p.

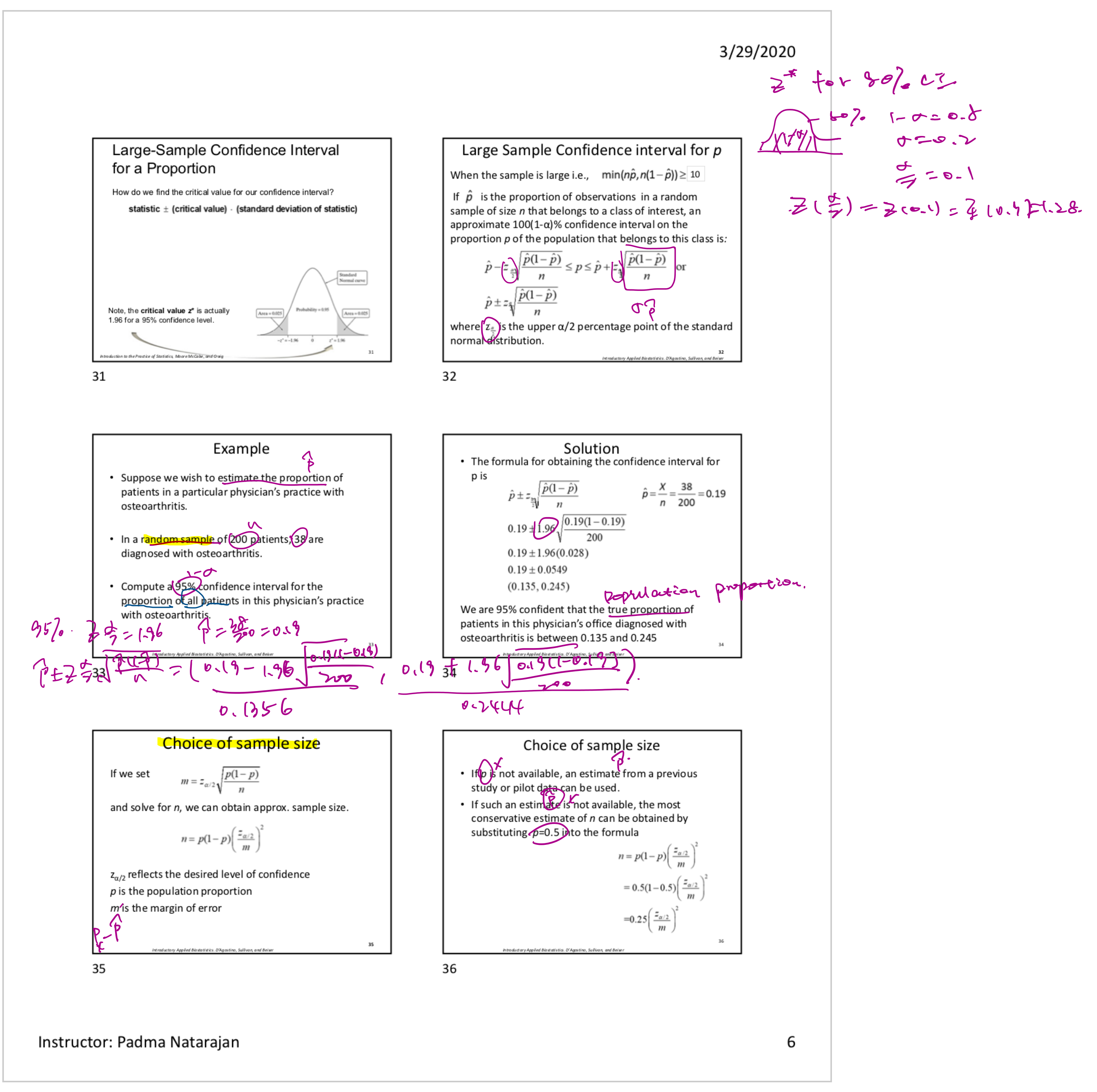
The standard deviation of the sampling distibution of $\hat{p}$ is:

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ X know p.

Since we don't know p, we replace it with the sample proportion $\hat{p}$. This gives us the **standard error (SE)** of the sample proportion:

$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ use $\hat{p}$

30



z* for 80% CI
80% 1 − α = 0.8
α = 0.2
α/2 = 0.1
$z(\frac{\alpha}{2}) = z(0.1) = z(0.9) = 1.28$

Large-Sample Confidence Interval for a Proportion

How do we find the critical value for our confidence interval?

statistic ± (critical value) · (standard deviation of statistic)

Note, the **critical value z*** is actually 1.96 for a 95% confidence level.

31

Large Sample Confidence interval for p

When the sample is large i.e., $\min(n\hat{p}, n(1-\hat{p})) \ge 10$

If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate 100(1-α)% confidence interval on the proportion p of the population that belongs to this class is:

$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ or

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

where $z_{\frac{\alpha}{2}}$ is the upper α/2 percentage point of the standard normal distribution.

32

Example

- Suppose we wish to estimate the proportion of patients in a particular physician's practice with osteoarthritis.
- In a random sample of 200 patients, 38 are diagnosed with osteoarthritis.
- Compute a 95% confidence interval for the proportion of all patients in this physician's practice with osteoarthritis.

95%. $z_{\frac{\alpha}{2}} = 1.96$ $\hat{p} = \frac{38}{200} = 0.19$

$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \left(0.19 - 1.96\sqrt{\frac{0.19(1-0.19)}{200}},\ 0.19 + 1.96\sqrt{\frac{0.19(1-0.19)}{200}}\right)$
0.1356 0.2444

33

Solution

- The formula for obtaining the confidence interval for p is

$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\hat{p} = \frac{X}{n} = \frac{38}{200} = 0.19$

$0.19 \pm 1.96\sqrt{\frac{0.19(1-0.19)}{200}}$

0.19 ± 1.96(0.028)

0.19 ± 0.0549

(0.135, 0.245)

We are 95% confident that the true proportion of patients in this physician's office diagnosed with osteoarthritis is between 0.135 and 0.245

population proportion.

34

Choice of sample size

If we set $m = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

and solve for n, we can obtain approx. sample size.

$n = p(1-p)\left(\frac{z_{\alpha/2}}{m}\right)^2$

$z_{\alpha/2}$ reflects the desired level of confidence
p is the population proportion
m is the margin of error

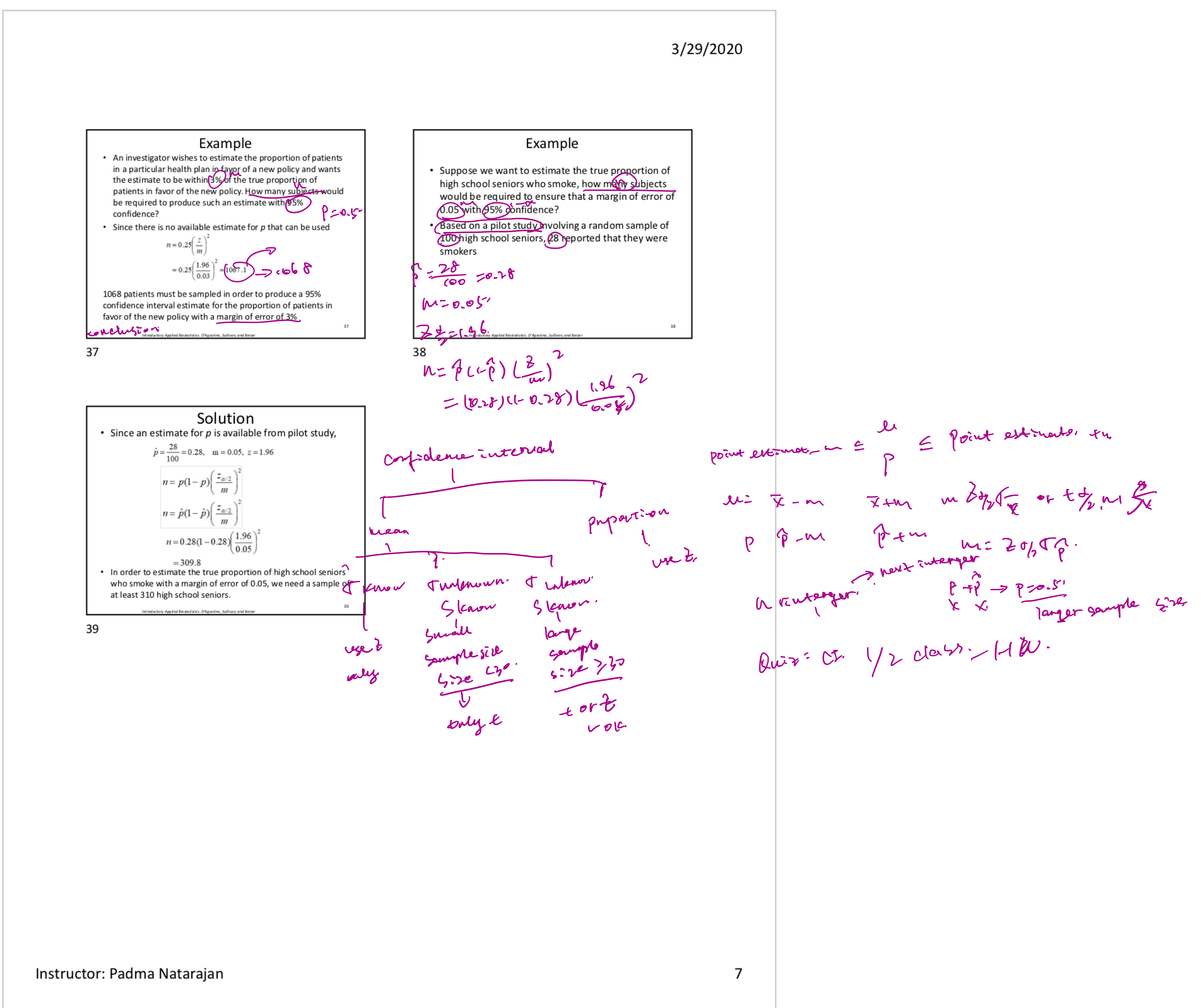
35

Choice of sample size

- If p is not available, an estimate from a previous study or pilot data can be used.
- If such an estimate is not available, the most conservative estimate of n can be obtained by substituting p=0.5 into the formula

$n = p(1-p)\left(\frac{z_{\alpha/2}}{m}\right)^2$
$= 0.5(1-0.5)\left(\frac{z_{\alpha/2}}{m}\right)^2$
$= 0.25\left(\frac{z_{\alpha/2}}{m}\right)^2$

36



Example

- An investigator wishes to estimate the proportion of patients in a particular health plan in favor of a new policy and wants the estimate to be within 3% of the true proportion of patients in favor of the new policy. How many subjects would be required to produce such an estimate with 95% confidence?
- Since there is no available estimate for p that can be used

$n = 0.25\left(\frac{z}{m}\right)^2$
$= 0.25\left(\frac{1.96}{0.03}\right)^2 = 1067.1$ → 1068

p = 0.5

1068 patients must be sampled in order to produce a 95% confidence interval estimate for the proportion of patients in favor of the new policy with a margin of error of 3%

conclusion

37

Example

- Suppose we want to estimate the true proportion of high school seniors who smoke, how many subjects would be required to ensure that a margin of error of 0.05 with 95% confidence?
- Based on a pilot study involving a random sample of 100 high school seniors, 28 reported that they were smokers

$\hat{p} = \frac{28}{100} = 0.28$
m = 0.05
$z_{\frac{\alpha}{2}} = 1.96$

38

$n = \hat{p}(1-\hat{p})\left(\frac{z}{m}\right)^2$
$= (0.28)(1-0.28)\left(\frac{1.96}{0.05}\right)^2$

Solution

- Since an estimate for p is available from pilot study,

$\hat{p} = \frac{28}{100} = 0.28$, m = 0.05, z = 1.96

$n = p(1-p)\left(\frac{z_{\alpha/2}}{m}\right)^2$

$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{m}\right)^2$

$n = 0.28(1-0.28)\left(\frac{1.96}{0.05}\right)^2$

$= 309.8$

- In order to estimate the true proportion of high school seniors who smoke with a margin of error of 0.05, we need a sample of at least 310 high school seniors.

39

confidence interval
mean
proportion
σ know
use z only
σ unknown. s know
small sample size
size < 30.
only t
σ unknown. s know.
large sample
size > 30
t or z ok
use z.

point estimate − m ≤ μ / p ≤ point estimate + m
μ: $\bar{x} - m$ $\bar{x} + m$ $m = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $t_{\frac{\alpha}{2}, n-1}\frac{s}{\sqrt{n}}$
p: $\hat{p} - m$ $\hat{p} + m$ $m = z_{\alpha/2}\sigma_{\hat{p}}$.
n integer → next integer
p → $\hat{p}$ → p = 0.5
X X larger sample size

Quiz: CI. 1/2 class. HW.