Statistics

Lecture 14

Feb 19-8:47 AM

Consider a population normally distributed with the mean of 125 and standard deviation of 15.

$N(125, 15)$

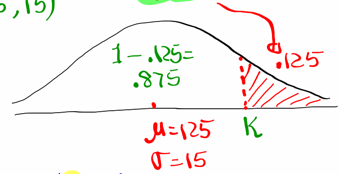
1) Find K such that $P(x > K) = .125$

Right Area

$K = \text{invNorm}(.875, 125, 15)$

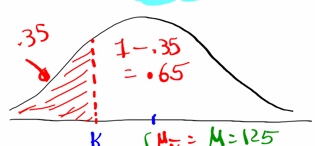
$= 142.255$

$\approx \boxed{142}$



2) Find K such that $P(\bar{x} < K) = .35$ for randomly selected group of 6.

Left area



CLT $\begin{cases} \mu_{\bar{x}} = \mu = 125 \\ \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{15}{\sqrt{6}} \end{cases}$

$K = \text{invNorm}(.35, 125, 15/\sqrt{6})$

$= 122.640 \approx \boxed{123}$

Feb 7-4:32 PM

Population ⟷ Parameter

SG 22-23

Sample ⟷ Statistic

Estimating Parameters

we can estimate

1) Population Proportion
P

2) Population Mean
μ

3) Population Standard deviation
σ

we use

Sample Proportion
$\hat{p}$ p-hat

Sample Mean
$\bar{x}$ x-bar

Sample standard deviation
S

Point-estimate →

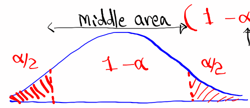
Feb 7-4:42 PM

when we estimate a parameter, we find range of values which is Confidence Interval

Every confidence interval comes with a Confidence level.

↳ is the middle area of the dist. curve and it is

Middle area $(1-\alpha)\cdot 100\%$

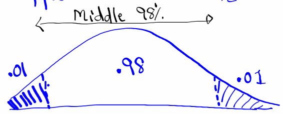


Alpha $0 < \alpha < 1$

↳ Significance level

Suppose $\alpha = .02$ $\alpha/2 = .01$, $1-\alpha = .98$

Middle 98%



α Significance level

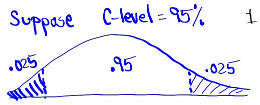
$(1-\alpha)\cdot 100\%$ C-level

98%

Suppose C-level = 95% $1-\alpha = .95$

$\alpha = .05$ Sign. level

$\alpha/2 = .025$ area of each tail



α → Significance level

$\alpha/2$ → area of each tail

$1-\alpha$ → Middle area

$(1-\alpha)\cdot 100\%$ → C-level

If α not given → use .05

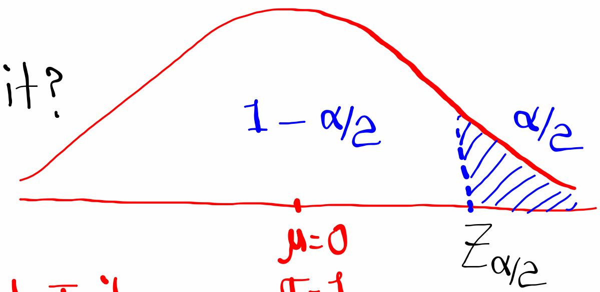
If C-level not given use 95%

Feb 7-4:49 PM

$Z_{\alpha/2}$ is a value with Standard Normal Prob. dist with right-tail area of $\alpha/2$.

How do we find it?

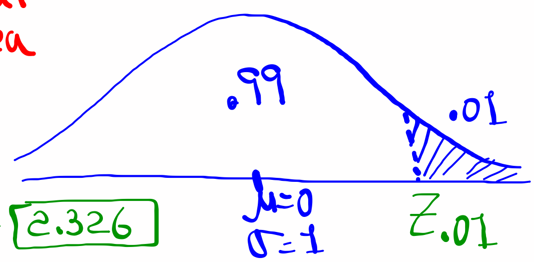
invNorm



Find $Z_{.01}$ — Right-Tail Area

$\alpha/2 = .01$

$\alpha = .02$

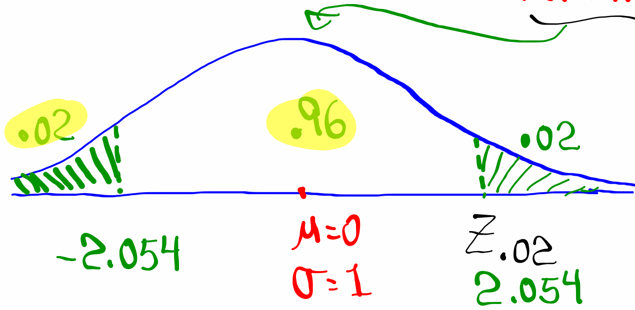


$Z_{.01} = \text{invNorm}(.99, 0, 1) = \boxed{2.326}$

Feb 7-4:59 PM

Find $Z_{\alpha/2}$ for $\underbrace{96\% \text{ C-level}}$.

Middle Area .96



$1 - \alpha = .96$

$\alpha = .04$

$\alpha/2 = .02$ Right-Tail area

$Z_{.02} = \text{invNorm}(.98, 0, 1) = \boxed{2.054}$

Feb 7-5:04 PM

find $Z_{\alpha/2}$ when α not given.

C-level

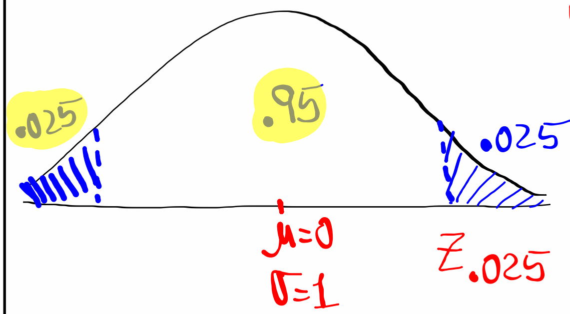
$(1-\alpha)\cdot 100\%$

95%

use $\alpha = .05$

$\alpha/2 = .025$

Area of each tail



$Z_{.025} = \text{invNorm}(.975, 0, 1) = \boxed{1.960}$

≈ 2

Usual Range

$\mu \pm 2\sigma$

Feb 7-5:08 PM

Estimating Population Proportion p:

$\cdots < p < \cdots$

$\hat{p} - E < p < \hat{p} + E$

Sample Proportion

Point-estimate

Margin of error

$\hat{p} = \frac{x}{n}$ # of favorable responses

Sample Size

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

$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Critical value for $(1-\alpha)\cdot 100\%$

C-level

Feb 7-5:13 PM

I Surveyed 100 Students, and 80 Said they Plan to vote on March 5.

$n = 100 \qquad x = 80$

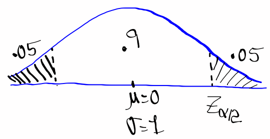
$\hat{p} = \frac{x}{n} = \boxed{.8} \qquad \hat{q} = 1 - \hat{p} = \boxed{.2}$

$\hat{p} - E < p < \hat{p} + E$

$.8 - E < p < .8 + E$

Suppose we wish to have 90% C-level

$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$



$= 1.645 \cdot \sqrt{\frac{(.8)(.2)}{100}}$

$= .0658 \approx \boxed{.07}$

$.8 - .07 < p < .8 + .07$

$Z_{\alpha/2} = \text{invNorm}(.95, 0, 1) = \boxed{1.645}$

$\boxed{.73 < p < .87}$

---

using TI:

$\boxed{\text{STAT}}$ $\boxed{\text{TESTS}}$ ↓ $\boxed{\text{1-PropZInt}}$

x: 80
n = 100
C-level: .9
$\boxed{\text{Calculate}}$

$\boxed{.73 < p < .87}$

$E = \frac{.87 - .73}{2} = \boxed{.07}$

$\hat{p} = \frac{.87 + .73}{2} = \boxed{.8}$

Feb 7-5:19 PM

I Surveyed 240 Students and 65% of them had iphone.

$n = 240$

$\hat{p} = .65 \quad \rightarrow x = n\hat{p} = 240(.65) = 156$

if decimal → Round-up

Find $\boxed{\text{Conf. interval}}$ for the prop. of all Students that have iPhone.

→ NO C-level

Use .95

$\boxed{.59 < p < .71}$

$\boxed{\text{1-PropZInt}}$

x = 156
n = 240
C-level: .95

I am 95% confident that between 59% and 71% of all Students have iphone.

$E = \frac{.71 - .59}{2} = \boxed{.06}$

$\hat{p} = \frac{.71 + .59}{2} = \boxed{.65}$

Feb 7-5:33 PM

I surveyed 175 students and 8% of them were smokers.

$n = 175$
$\hat{p} = .08 \rightarrow x = n\hat{p} = 175(.08) = 14$

Find $\boxed{99\%\ \text{Conf. interval}}$ for the prop. of all students that are smokers.

C-level

$E = \frac{.13 - .03}{2} = \boxed{.05}$

1 - PropZInt

$\boxed{.03 < P < .13}$

we are 99% confident that between 3% and 13% of all students are smokers.

$\hat{p} = \frac{.13 + .03}{2} = \boxed{.08}$

Feb 7-5:41 PM

Estimating Population Mean μ:

$\cdots < \mu < \cdots$

$\bar{x} - E < \mu < \bar{x} + E$

Sample Mean
Point-estimate

Margin of error

Case I : σ Known	
$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ TI: Z Interval inpt: $\boxed{\text{STATS}}$	

$E = \frac{\quad - \quad}{2}$, $\bar{x} = \frac{\quad + \quad}{2}$

Feb 7-6:01 PM

I randomly selected 40 textbooks, the mean price was \$125.

$n=40$ $\bar{x}=125$

It is known that standard deviation for price of all textbooks is \$20.

$\sigma=20$

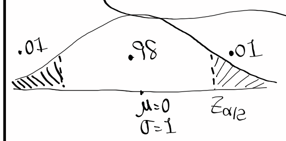
C-level .98

Find 98% Conf. interval for the mean price of all textbooks.

σ Known

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = 2.326 \cdot \frac{20}{\sqrt{40}} = \boxed{7.36} \approx 7$$



$Z_{\alpha/2} = \text{invNorm}(.99, 0, 1) = \boxed{2.326}$

$$E = \frac{132-118}{2} = \boxed{7}$$

$$\bar{x} = \frac{132+118}{2} = \boxed{125}$$

$\bar{x} - E < \mu < \bar{x} + E$

$125 - 7 < \mu < 125 + 7$

$\boxed{118 < \mu < 132}$

TI:
STAT TESTS
Z Interval
inp: Stats
$\sigma = 20$
$\bar{x} = 125$
$n = 40$
C-level: .98
Calculate

$\boxed{118 < \mu < 132}$

Feb 7-6:05 PM

I surveyed 32 students, their mean age was 28.5 yrs.

$n=32$, $\bar{x}=28.5$

It is known that standard deviation of ages of all students is 7.2 yrs.

$\sigma=7.2$

C-level: .9

Find 90% Conf. interval for the mean age of all students.

$\boxed{26.4 < \mu < 30.6}$

σ Known → Z Interval

inpt: Stats
$\sigma = 7.2$
$\bar{x} = 28.5$ ← 1-decimal
$n = 32$
C-level: .9
Calculate

$$E = \frac{30.6-26.4}{2} = \boxed{2.1}$$

$$\bar{x} = \frac{30.6+26.4}{2} = \boxed{28.5}$$

Feb 7-6:17 PM

I randomly selected 25 exams, here are the scores

80	85	60	100	70
75	65	90	92	98
78	88	55	100	94
79	89	69	59	98
100	70	80	60	90

1) Find $\bar{x}$, Round to whole #.

$\boxed{\bar{x} \approx 81}$

2) Assume $\sigma = 12$,

Find Conf. interval for the mean of all exams.

NO C-level
Use .95

$\boxed{76 < \mu < 86}$

Since σ is known
Use Z Interval
inpt: Stats

$E = \frac{86-76}{2} = \boxed{5}$

$\bar{x} = \frac{86+76}{2} = \boxed{81}$

Feb 7-6:25 PM

Estimating Population Mean μ:

$\cdots < \mu < \cdots$

$\bar{x} - E < \mu < \bar{x} + E$

Sample Mean
Point-estimate

Margin of error

Case I: σ Known	Case II: σ unknown
$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ → $df = n-1$
TI: Z Interval inpt: STATS	TI: T Interval inpt: Stats

$E = \frac{-}{2}$, $\bar{x} = \frac{+}{2}$

Feb 7-6:01 PM

Given: $n = 20$, $\bar{x} = 72$, $s = 10$,

Find 90% Conf. interval for μ.

σ known → Z Interval

$\boxed{\sigma \text{ unknown} \to \text{T Interval}}$

$\boxed{68 < \mu < 76}$

inpt: $\boxed{\text{Stats}}$

$\bar{x} = 72$ ← whole

$s = 10$

$n = 20$

C-level: .9

$E = \frac{76-68}{2} = \boxed{4}$

$\bar{x} = \frac{76+68}{2} = \boxed{72}$

Feb 7-6:37 PM

I randomly selected 10 gas stations. Here are the gas prices/gal.

4.25	4.55	3.85
4.19	3.99	4.75
4.29	4.49	3.99
	4.85	

Find

$\bar{x} = 4.32$

$s = .33$

in two decimals.

Find $\boxed{\text{Conf. interval}}$ for the mean gas price for all gas stations.

→ No C-level → use .95

σ Known → Z Interval

$\boxed{\sigma \text{ Unknown} \to \text{T Interval}}$

inpt: $\boxed{\text{Stats}}$

$\bar{x} = 4.32$ $\boxed{4.08 < \mu < 4.56}$

$s = .33$ 2-decimals

$n = 10$

C-level: .95

$\boxed{\text{Calculate}}$

Feb 7-6:41 PM

Geometric Prob. dist. SG 17

It is very Similar to binomial prob. dist but there is no n.

$p \rightarrow$ Prob. of Success

$q \rightarrow$ Prob. of failure $p + q = 1$

$$\boxed{P(x) = p \cdot q^{x-1}}$$

where x is the first attempt when Success happens. $x = 1, 2, 3, 4, \ldots$

$\mu = \frac{1}{p}$ $\sigma^2 = \frac{q}{p^2}$

$\sigma = \sqrt{\sigma^2}$

Feb 7-6:53 PM

Consider a geometric Prob. dist with $p = .4$.

$q = 1 - p = .6$

$\mu = \frac{1}{p} = \frac{1}{.4} = 2.5$

$\sigma^2 = \frac{q}{p^2} = \frac{.6}{.4^2} = 3.75$

$\rightarrow \sigma = \sqrt{\sigma^2} \approx 1.936$

$P(x) = p \cdot q^{x-1}$

$P(x = 3) = .4(.6)^{3-1} = \boxed{.144}$

First Success happens on 3rd attempt.

$\boxed{2nd}$ $\boxed{VARS}$ $\boxed{geometpdf}$ $(.4, 3) = \boxed{.144}$

$P(x \leq 3) = \text{geometcdf}(.4, 3) = \boxed{.784}$

First success could happen on 1st, 2nd, or 3rd attempt

Feb 7-6:57 PM

Suppose a basketball player makes 80% of FT.

$p = .8$

$q = .2$

$\mu = \frac{1}{p} = \frac{1}{.8} = 1.25$

$\sigma^2 = \frac{q}{p^2} = \frac{.2}{.8^2} = .3125$

$\sigma = \sqrt{\sigma^2}$

$= \sqrt{.3125}$

$\approx .559$

P(the player makes FT on his/her 4th attempt)

$P(x=4) = \text{geometpdf}(.8, 4) = \boxed{.006}$

P(The player makes FT after 2nd attempt)

$P(x>2) = P(x \geq 3) = 1 - P(x \leq 2)$

$= 1 - \text{geometcdf}(.8, 2)$

$= \boxed{.04}$

Feb 7-7:05 PM