Statistics

Lecture 11

Feb 19-8:47 AM

Review of SG 19

Clear all lists.

Store 1, 3, and 5 in L1.

Use [1-Var Stats] with L1 only to find

$\mu = \bar{x} = 3$ $\sigma = \sigma_{\bar{x}} = 1.633$ $\sigma^2 = \sigma_{\bar{x}}^2 = 2.6 = \frac{8}{3}$

Take all Samples of Size 2 with replacement from 1, 3, and 5.

1,1	1,3	1,5
3,1	3,3	3,5
5,1	5,3	5,5

9 Samples of Size 2.

Find $\bar{x}$ of each Sample.

1	2	3
2	3	4
3	4	5

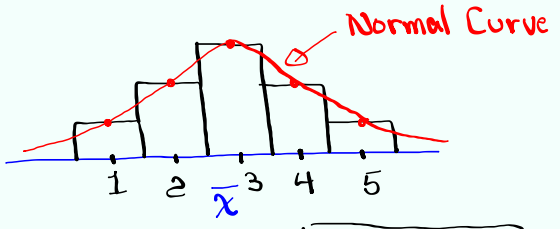
9 means

$\bar{x}$	$P(\bar{x})$
1	1/9
2	2/9
3	3/9
4	2/9
5	1/9

Nov 7-8:02 AM

$\bar{x}$	$P(\bar{x})$
1	1/9
2	2/9
3	3/9
4	2/9
5	1/9

Draw Prob. dist. Histogram



$\bar{x}$ → L2 , $P(\bar{x})$ → L3 , use 1-Var Stats with L2 & L3. Find

$\mu_{\bar{x}} = 3$ $\quad \sigma_{\bar{x}} = 1.155$ $\quad \sigma^2_{\bar{x}} = 1.\overline{3} = \frac{4}{3} = \frac{8}{3} \cdot \frac{1}{2}$

Central Limit theorem CLT

$\mu_{\bar{x}} = \mu$ $\quad \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ $\quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Nov 7-8:10 AM

Exam Scores have a normal dist. with mean of 85 and Standard dev. of 12.

$N(85, 12)$ $\quad n=9$

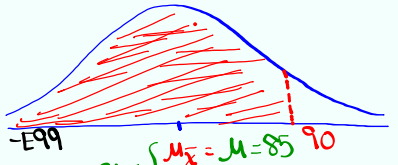
If we randomly select 9 exams

$\mu_{\bar{x}} = \mu = 85$ CLT $\quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{9}} = \frac{12}{3} = 4$ CLT

Find the prob. that their mean ($\bar{x}$) Score is below 90.

$\bar{x} < 90$

$P(\bar{x} < 90)$



CLT $\begin{cases} \mu_{\bar{x}} = \mu = 85 \\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{9}} = 4 \end{cases}$

= normalcdf (-E99, 90, 85, 4)

= .894

(-) 2nd [)]

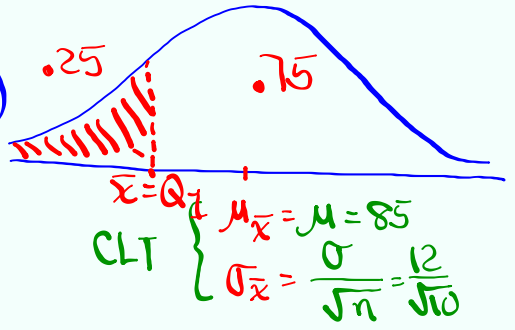
Nov 7-8:19 AM

Find $\bar{x} = Q_1$ for randomly selected group of 10 exams. Round to whole #.

$\bar{x} = Q_1$

$= \text{invNorm}(.25, 85, 12/\sqrt{10})$

$= 82.444 \approx \boxed{82}$



CLT $\begin{cases} \mu_{\bar{x}} = \mu = 85 \\ \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{12}{\sqrt{10}} \end{cases}$

SG 19 ✓

Nov 7-8:26 AM

Ages of students are normally dist. (SG 20) with mean of 32 and standard dev. of 8.

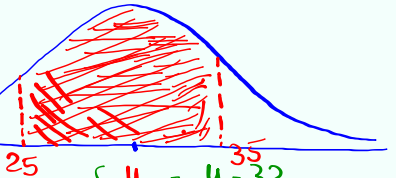
$N(32, 8)$ $n=5$

If we randomly select 5 students, find the prob. that their mean age ($\bar{x}$) is between 25 and 35 yrs.

$P(25 < \bar{x} < 35)$

$= \text{normalcdf}(25, 35, 32, 8/\sqrt{5})$

$= \boxed{.774} = \boxed{77.4\%}$



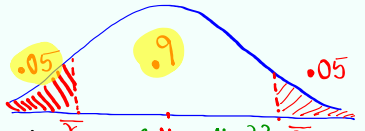
CLT $\begin{cases} \mu_{\bar{x}} = \mu = 32 \\ \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{8}{\sqrt{5}} \end{cases}$

Nov 7-8:31 AM

Find $\bar{x}_1, \bar{x}_2$ two means, rounded to 1-dec., that separate the middle 90% from the rest for randomly selected group of 4 students.

$1 - .9 = .1$

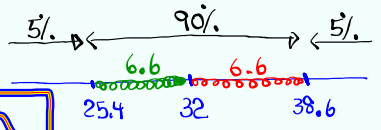
$.1 \div 2 = .05$



$\bar{x}_1 = P_5 = \text{invNorm}(.05, 32, 4) \approx 25.4$

CLT $\begin{cases} \mu_{\bar{x}} = \mu = 32 \\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = \frac{8}{2} = 4 \end{cases}$

$\bar{x}_2 = P_{95} = \text{invNorm}(.95, 32, 4) \approx 38.6$



SG 20

Nov 7-8:38 AM

SG 21

$Z_{\alpha/2}$ Critical value that has a right area of $\alpha/2$.

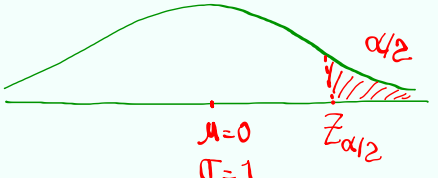
$\alpha \rightarrow$ Alpha

$0 < \alpha < 1$

Significance level.

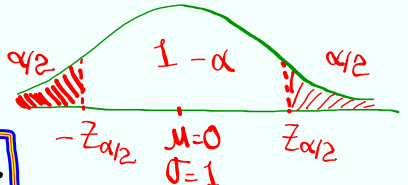
$\alpha/2 \rightarrow$ Area on the right

$1-\alpha \rightarrow$ Area in the middle



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

Confidence level



How to find $Z_{\alpha/2}$: use invNorm

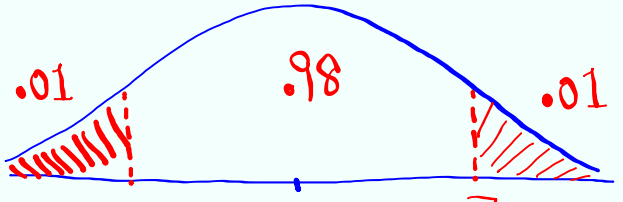
Nov 7-9:04 AM

Find $z_{.01}$ → Right tail Area

$\frac{\alpha}{2} = .01 \rightarrow \alpha = .02$ → Significance level

$1 - \alpha = .98$ Middle Area

$.98 = 98\%$ Confidence level



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

Nov 7-9:10 AM

Find $\pm z_{\alpha/2}$ for $\alpha = .04$.

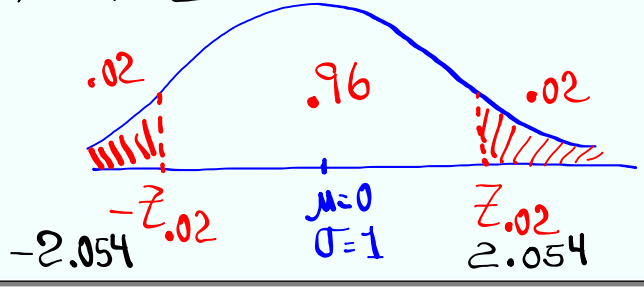
↑ Significance level

$1 - \alpha = .96$ Middle Area

$.96 = 96\%$ Confidence level

$\alpha/2 = .02$ area of each tail

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



Nov 7-9:15 AM

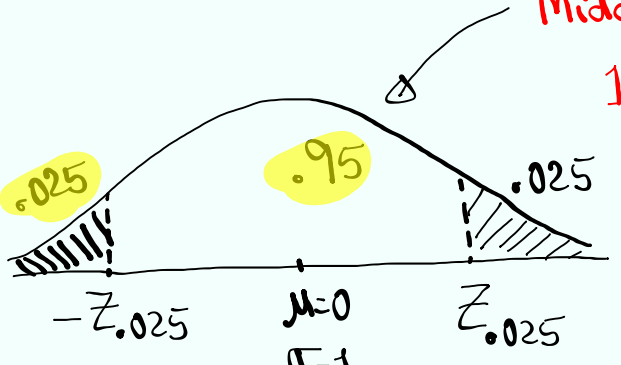
Find $Z_{\alpha/2}$ for 95% Confidence level.

Middle Area .95

$1-\alpha = .95$

$\alpha = .05$ ← Significance level

$\alpha/2 = .025$



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

$= \boxed{1.960}$

Nov 7-9:19 AM

$t_{\alpha/2}$ t - Dist.

Symmetric, Bell-Shape, Total Area 1

$\mu = 0$, σ unknown

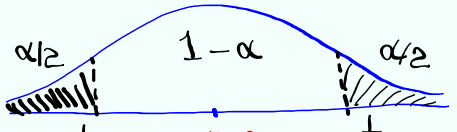
It comes with degrees of freedom

df

$\alpha \rightarrow$ Significance level

$\alpha \rightarrow 0 < \alpha < 1$

$\alpha/2 \rightarrow$ Area of each tail



$1-\alpha \rightarrow$ Middle Area

$(1-\alpha)\cdot 100\% \rightarrow$ Confidence level

How to find $t_{\alpha/2}$: use invT — Left Area & df.

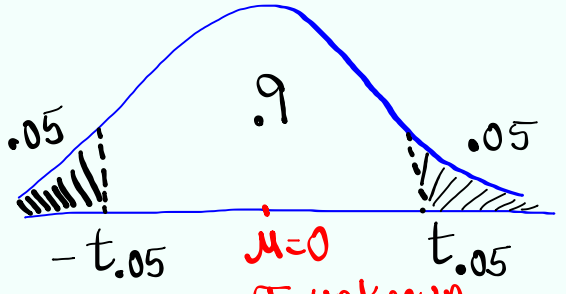
Nov 7-9:24 AM

Find $t_{.05}$ with $df = 19$.

$\alpha/2 = .05$

$\alpha = .1$

$1 - \alpha = .9$



σ unknown

$df = 19$

$t_{.05} = \text{invT}(.95, 19) = \boxed{1.729}$

left Area

df

Nov 7-9:30 AM

Find $\pm t_{\alpha/2}$ for 99% Confidence level with $df = 24$.

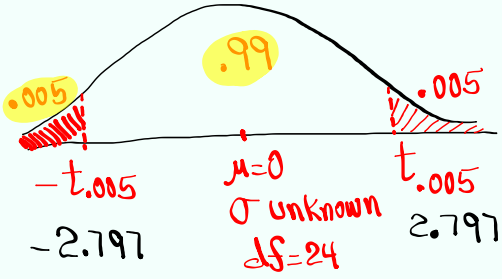
C-level

Middle Area = .99

$1 - \alpha = .99$

$\alpha = .01$ Sig. level

$\alpha/2 = .005$ area of each tail



$t_{.005} = \text{invT}(.995, 24) = \boxed{2.797}$

As df increases,

$t_{\alpha/2} \approx z_{\alpha/2}$

Nov 7-9:35 AM

What is degrees of freedom?

Consider a class of 20 students,

I bring in 20 Donuts,

Each student can choose 1 Donut.

First	student	20 choices
Second	"	19 "
Third	"	18 "
⋮		
Last	"	0 choices (1 donut)

df = 19

Nov 7-9:41 AM

You have 7 clean shirt, you only wear one clean shirt daily.

Monday	7 shirts	(7 choices)
Tuesday	6 "	(6 choices)
Wednesday	5 "	(5 ")
⋮		
Sunday	1 shirt	(0 choices)

df = 6

df will be determined by topic or will be given to us.

Nov 7-9:45 AM

Estimating Parameters SG 21

Population ⟷ Parameter

Samples ⟷ Statistic

We use statistic to estimate Parameter.

To estimate Population we use Sample

Proportion p	Point-estimate	Proportion $\hat{p}$
Mean μ		Mean $\bar{x}$
Standard Dev. σ		Standard Dev. s

Estimation will be a range of values that is called Confidence Interval

Every estimations Comes with Confidence level. If none is given, use 95% C-level.
C-level

Nov 7-10:02 AM

Confidence Interval for population Proportion:

Answer $\quad < p <$

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

$\hat{p}$: Sample Proportion, Point-estimate

E: Margin of error

$\hat{p} = \frac{x}{n}$ ← # of favorable responces, ← Sample Size

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

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

$Z_{\alpha/2}$: Critical Value for $(1-\alpha)\cdot 100\%$ C-level.

Nov 7-10:11 AM

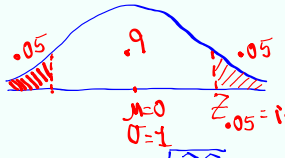
I Surveyed 100 Students, and 80 of them had iPhone.

$n = 100$, $x = 80$

$\hat{p} = \frac{x}{n} = \frac{80}{100} = .8$

$\hat{q} = 1 - \hat{p} = .2$

I want to find 90% Confidence interval for the prop. of all Students that have iPhone



$z_{.05} = \text{invNorm}(.95, 0, 1) = 1.645$

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

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

$\approx .07$

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

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

$.73 < p < .87$

I am 90% Confident that between 73% and 87% of all Students have iPhone

Using TI: $.73 < p < .87$

STAT TESTS 1-PropZInt

$x = 80$, $n = 100$, C-level: .9 Calculate

Nov 7-10:16 AM

I Surveyed 150 Students, 30 of them were Smokers. $n = 150$, $x = 30$

C-level: .99

find 99% Confidence interval for the prop. of all Students that are Smokers.

STAT TESTS 1-PropZInt

$.12 < p < .28$

$E = \frac{.28 - .12}{2} = .08$

$\hat{p} = \frac{.28 + .12}{2} = .2$

I am 99% Confident that between 12% and 28% of all Students are Smokers.

Nov 7-10:29 AM

I surveyed 250 students and 30% of them drove to campus alone.

$n = 250$

$\hat{p} = .3$

$\Rightarrow x = n\hat{p} = 250(.3) = 75$

if decimal, round-up

→ Use 95% C-level

No C-level

Find Confidence Interval for the prop. of all students that drive alone to campus.

1-PropZInt

$x = 75$

$n = 250$

C-level: .95

Calculate

$.24 < p < .36$

$E = \frac{.36 - .24}{2} = .06$

$\hat{p} = \frac{.36 + .24}{2} = .3$

Nov 7-10:36 AM

How to determine the Sample Size:

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

with some algebra, isolate n.

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

if decimal → Always round-up.

If $\hat{p}$ & $\hat{q}$ unknown, use .5 for each

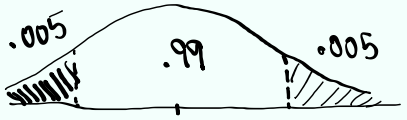
$$n = .25\left(\frac{z_{\alpha/2}}{E}\right)^2$$

Nov 7-10:43 AM

Find minimum Sample Size needed to Construct 99% Conf. interval fo pop. prop. with $\hat{p} = .2$ and $E = .05$.

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2 = (.2)(.8)\left(\frac{2.576}{.05}\right)^2$$

$$\approx 424.689 \approx \boxed{425}$$



$\mu = 0$, $\sigma = 1$

$z_{.005} = \text{invNorm}(.995, 0, 1) = \boxed{2.576}$

Suppose $\hat{p}$ was unknown

$$n = .25\left(\frac{z_{\alpha/2}}{E}\right)^2 = .25\left(\frac{2.576}{.05}\right)^2 \approx \boxed{664}$$

Nov 7-10:47 AM

Find min. Sample Size needed to Construct 90% Conf. interval for pop. Prop. and error not to exceed 4%.

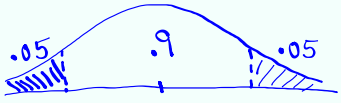
C-level: .9

$E = .04$

$\hat{p}$ unknown

$$n = .25\left(\frac{z_{\alpha/2}}{E}\right)^2 = .25\left(\frac{1.645}{.04}\right)^2$$

$$\approx \boxed{423}$$



$\mu = 0$, $\sigma = 1$

$z_{.05} = \text{invNorm}(.95, 0, 1) = \boxed{1.645}$

Redo with $E = .08$

$$n = .25\left(\frac{1.645}{.08}\right)^2 \approx \boxed{106}$$

Nov 7-10:56 AM