Statistics

Winter 2022

Lecture 13

Live QZ 4

Consider a binomial Prob. dist. with

$n = 275$ and $P = .6$

Find (Round to whole #)

1) $\mu = np$

$= 275(.6)$

$= \boxed{165}$ ✓

2) $\sigma^2 = npq$

$= 275(.6)(.4)$

$= \boxed{66}$ ✓✓

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

$= \sqrt{66}$

$\boxed{\sigma \approx 8}$ ✓✓✓

I took a Survey of 280 students, 32% of them were STEM majors. $n=280$ $\hat{p}=.32$

1) How many were stem majors?

$x = n\hat{p} = 280(.32) = 89.6$ $\boxed{x=90}$

if decimal ⇒ Round-up

2) Find 90% Conf. interval for the prop. of all students that are stem Majors. $\boxed{.276<p<.367}$

↳ C-level: .9

1-PropZInt(90, 280, .9)

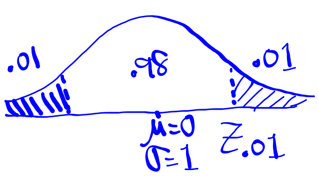
3) Find the margin of error. $E = \dfrac{.367-.276}{2} = \boxed{.0455}$

$\approx 5\%$

4) Find minimum Sample Size needed if we wish to build 98% Conf. interval for the prop. of all students that are STEM majors and error not to exceed 4%.

$E = z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n}} \Rightarrow n = \hat{p}\hat{q}\left(\dfrac{z_{\alpha/2}}{E}\right)^2$

Always round-up



$n = (.32)(.68)\left(\dfrac{2.326}{.04}\right)^2$

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

$n = 735.7975\ldots$

$\boxed{n=736}$

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

$n = (.5)(.5)\left(\dfrac{2.326}{.04}\right)^2 = 845.35\ldots$ $\boxed{n=846}$

I randomly selected 35 police officers, and their mean age was $\boxed{46.5}$ yrs.

$n = 35$
$\bar{x} = 46.5$

It is known that standard deviation of ages of all police officers is 7.9 yrs.

$\sigma = 7.9$

1) Find $\boxed{\text{conf. interval}}$ for the $\boxed{\text{mean}}$ age of $\boxed{\text{all}}$ police officers.

No C-level
use .95

σ Known
$\Rightarrow$ Z Interval
inpt: $\boxed{\text{stats}}$

$\boxed{43.9 < \mu < 49.1}$

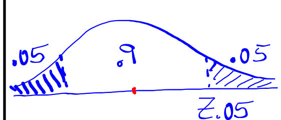
$\sigma = 7.9$
$\bar{x} = 46.5$
$n = 35$
C-level: .95

2) Find the margin of error.

$$E = \frac{49.1 - 43.9}{2} = \boxed{2.6}$$

3) Find minimum sample size needed if we wish to construct 90% conf. interval for the mean age of all police officers and margin of error not to exceed 5 yrs.

$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow$ Solve for n $\Rightarrow n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2$



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

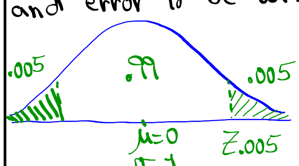
Always Round-up

$n = \left(\frac{1.645 \cdot 7.9}{5}\right)^2$

$n = 6.755\ldots$

$\boxed{n = 7}$

Redo with 99% C-level, and error to be within 2.5 yrs.



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

$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{2.576 \cdot 7.9}{2.5}\right)^2$

$n = 66.262\ldots$

$\boxed{n = 67}$

I randomly Selected 15 Students. Data below shows how many units they have completed at college.

25	32	40	45
18	10	28	35
55	50	30	48
20	30	40	

① Find $\bar{x}$ and S. Round to a whole #.

$\bar{x} \approx 34$, $S \approx 13$

② Find S^2 in reduced fraction.

$$\boxed{S^2 = \frac{17002}{105}}$$

3) Find Confidence interval for mean of completed units for all students.

σ Known ⇒ Z Interval

NO C-level ⇒ use .95

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

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

$\bar{x} = 34$

$S = 13$

$n = 15$

C-level: .95

$\boxed{27 < \mu < 41}$

4) Margin of error

$E = \dfrac{41 - 27}{2} = \boxed{7}$

5) Find min. Sample Size needed if wish to construct 97% Conf. interval for the mean, and error not to exceed 4 units.

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

If σ unknown ⇒ use S

$$n = \left(\frac{Z_{\alpha/2} \cdot S}{E}\right)^2$$

$\boxed{n = 50}$

$$= \left(\frac{2.170 \cdot 13}{4}\right)^2$$

$n = 49.737\ldots$

$Z_{.015} = \text{invNorm}(.985, 0, 1) = 2.170$

Geometric Prob. dist:

It is just like binomial prob. dist except there is no $\underbrace{\text{fixed number of trials}}_{n}$

$p \to$ Prob. of Success

$q \to$ Prob. of failure

$q = 1 - p$

$x \to$ # of trial when first success happens.

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

$$\mu = \frac{1}{p} \quad \sigma^2 = \frac{q}{p^2}, \text{ and } \sigma = \sqrt{\sigma^2}$$

Consider a geometric prob. dist with $p = .2$

$q = 1 - p = 1 - .2 = \boxed{.8}$

$\mu = \frac{1}{p} = \frac{1}{.2} = \boxed{5}$

$\sigma^2 = \frac{q}{p^2} = \frac{.8}{.2^2} = \boxed{20}$

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

$= \sqrt{20} = \boxed{4.472}$

P(First Success happens on 3rd trial)

$= P(x = 3) = \text{geometpdf}(.2, 3) = \boxed{.128}$

P(first Success happens $\underbrace{\text{before the fifth}}_{x<5}$ trial)

$= P(x < 5) = P(x \leq 4) = \text{geometcdf}(.2, 4) = .5904$

$\approx \boxed{.590}$

P(first Success happens $\underbrace{\text{after the 3rd}}_{x>3}$ trial)

$= P(x > 3) = P(x \geq 4) = 1 - P(x \leq 3)$

$= 1 - \text{geometcdf}(.2, 3) = \boxed{.512}$

Poisson Prob. dist:

The mean number must be given is Some fixed interval $\Rightarrow \mu$

$x \to$ # of Successes in that interval, $x \geq 0$

$$P(x) = \frac{\mu^x}{x!} \cdot e^{-\mu} \qquad e \approx 2.718 \qquad \sigma^2 = \mu$$

Chris gets in average 25 Calls per day to do computer repair.

$\mu = 25$ — fixed interval

$\sigma^2 = \mu = 25$ 68% Range $\Rightarrow \mu \pm \sigma$

$\sigma = \sqrt{\sigma^2} = 5$ $= 25 \pm 5$

$\Rightarrow 20$ to 30

$\mu = \lambda$ ← lambda   Usual Range $\Rightarrow \mu \pm 2\sigma$ $\Rightarrow$ $\boxed{15 \text{ to } 35}$

P(He fixes exactly 30 Computers in one day)

$P(x=30) = \text{Poissonpdf}(25, 30) = \boxed{.045}$

P(He fixes at most 35 computers)

$P(x \leq 35) = \text{poissoncdf}(25, 35) = \boxed{.978}$ ✓

Testing claims:

A claim could be made for any Parameters such as mean μ, Proportion p, standard deviation σ.

Our task is to examine the claim for its validity.

If claim is false $\Rightarrow$ we reject it

If claim is true $\Rightarrow$ we fail-to-Reject it

Implies Support

With every testing, there is a possibility of making errors.

If claim is valid but we reject it.

If claim is invalid but we support it.

With every testing, there is a significance level α. $0 < \alpha < 1$

when α not given $\Rightarrow$ we use .05.

Methods of Testing:

1) Traditional Method

2) P-Value Method

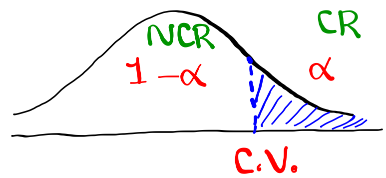
3) Confidence Interval Method

Regardless of the method, final conclusion should be the same.

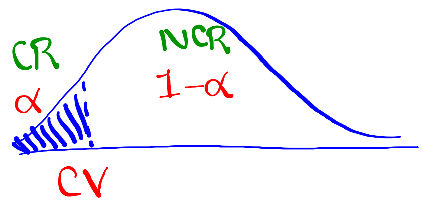
we support a valid claim OR we reject an invalid claim.

Testing Types:

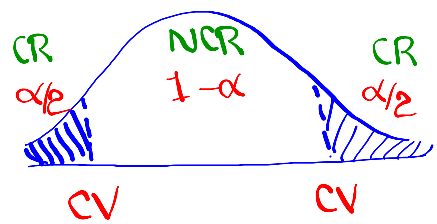
1) Right-Tail Test



2) Left-Tail Test



3) Two-Tail Test



Hypothesis Testing

1) Null Hypothesis H_0

H_0 must contain $=$ Condition

$=, \geq, \leq$

2) Alternative Hypothesis H_1

H_1 Cannot have $=$ Condition

$\neq, >, <$

Key words:

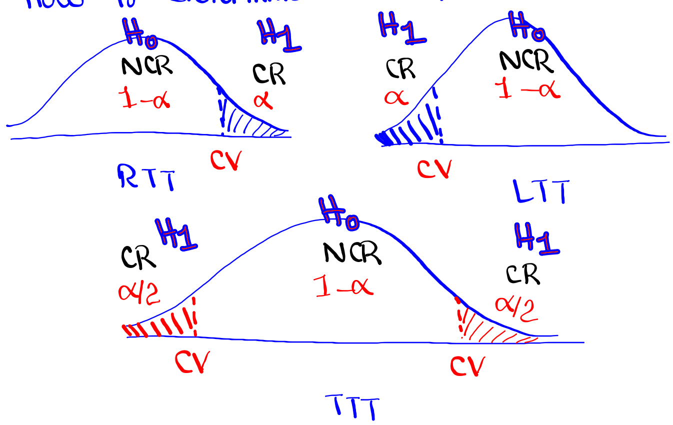
H_0: is, get, equal, at least, at most, ...

H_1: is not, not equal, different, more than, less than, exceed, above, below, ...

$H_0: =$	$H_0: \geq$	$H_0: \leq$
$H_1: \neq$	$H_1: <$	$H_1: >$
TTT	LTT	RTT
Two-Tail Test	Left-Tail Test	Right-Tail Test

H_1 tells us what type of testing.

How to determine validity of H_0 & H_1!



Possible outcomes for Ho

	Ho Valid	Ho invalid
Support Ho	Correct Decision	Type II error
Reject Ho	Type I error	Correct Decision

$P(H_0 \text{ Valid}) = 1 - \alpha = P(H_1 \text{ invalid})$

$P(H_0 \text{ invalid}) = \alpha = P(H_1 \text{ Valid})$

I claim that 45% of all Students are fully vaccinated.

H_0: $p = .45$ claim

H_1: $p \neq .45$ TTT

I claim the mean of all math exams at Mt. SAC is at least 80. ≥ 80

H_0: $\mu \geq 80$ claim

H_1: $\mu < 80$ LTT

I claim that standard deviation of ages of all police officers is more than 10 Years: >10

H_0: $\sigma \leq 10$

H_1: $\sigma > 10$ claim, RTT

CNN claims that less than 25% of all voters voted by Party line. $< .25$

H_0: $p \geq .25$

H_1: $p < .25$ claim LTT

Mt. SAC claims the mean age of all students during winter classes is not 30 yrs. $\neq 30$

$H_0: \mu = 30$

$H_1: \mu \neq 30$ claim, TTT

LA Times claims that standard deviation of salaries of all nurses is \$400/mo. $= 400$

$H_0: \sigma = 400$ claim

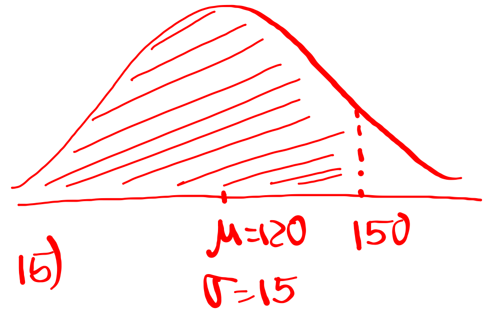
$H_1: \sigma \neq 400$ TTT

Let's go to the website.

Live QZ 5:

Given $N(120, 15)$

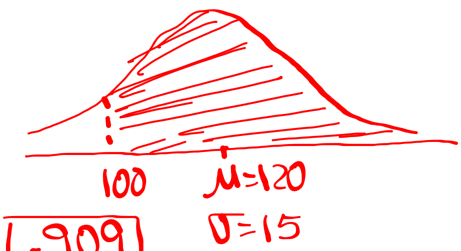
1) find $P(x < 150)$



$= \text{normalcdf}(-E99, 150, 120, 15)$

$= \boxed{.977}$

2) find $P(x > 100)$



$= \text{normalcdf}(100, E99, 120, 15) = \boxed{.909}$