2/6/24

7.1 Confidence Intervals

C-lvl	Z_c
80%	1.28
90%	1.645
95%	1.96
99%	2.58

1) $E = Z_c\left(\frac{\sigma}{\sqrt{n}}\right)$ maximum error

2) $\overline{X} - E < \mu < \overline{X} + E$

(the interval will be larger for higher percents confidence)

$\boxed{\uparrow w = \uparrow c}$

also $\uparrow n \uparrow c \downarrow w$

when pop $\uparrow$, the width $\downarrow$

Example: 15.6 is μ, σ is 1.8. 95% cofidence interval?

$E = 1.96\left(\frac{1.8}{\sqrt{90}}\right)$, $E = 0.37$

z for 95%

$\overline{X} - E < \mu < \overline{X} + E$

$15.6 - 0.37 < \mu < 15.6 + 0.37$

$\boxed{15.23 < \mu < 15.97}$

99% confidence interval?

$E = 2.58\left(\frac{1.8}{\sqrt{90}}\right)$, $E = 0.4895$

$15.6 - 0.49 < \mu < 15.6 + 0.49$

$15.11 < \mu < 16.09$

★ include on hw + tests

Conclusion: if we took 100 samples of n=90, we expect to catch the population mean (μ) of Julia's 2 mile jogging times 99 occasions

or many samples.... 99% occasions

Prerequsites for this to work:

Checks:

1) σ known
2) RRS
3) Independent
4) normally distributed or $n > 30$ (CLT)

★ include on hw + test

$W = 2E$

$E = W/2$

Finding Sample Size

-knowing E, finding n

$$\boxed{n = \left(\frac{Z_c \sigma}{E}\right)^2}$$

or plug back in $E = Z_c\left(\frac{\sigma}{\sqrt{n}}\right)$

T- Intervals

A24

$\boxed{\text{Tests of Significance}}$ / $\boxed{\text{T-table}}$

due to chance or something else?

__T-procedures:__ σ __unknown__

- Z formulas for significance: $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, $x \pm Z_c\left(\frac{\sigma}{\sqrt{n}}\right)$ c for confidence intervals
- if no σ, use s (sd for sample); but w/ this switch we can't use z, so we use T!

$$-t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \quad , \quad x \pm T\left(\frac{s}{\sqrt{n}}\right)$$

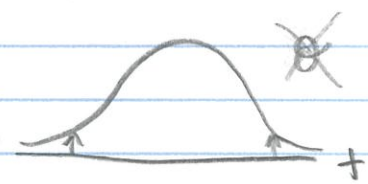
significance   confidence

__T-distribution__

- broader, tails are broder ($\uparrow$ probabiliy of getting results farther from 0, as s varies + more uncertain than σ)
- diff T-distr for every sample size
  - as sample size $\uparrow$, the sd (s) gets closer to σ, big samples more accurate
  - $\uparrow$ sample gets closer to normal curve
    - $\downarrow$ instead of saying sample size

__Degrees of Freedom__

- sample size minus 1 (n-1)
  as df $\uparrow$, t distr becomes more normal



WS Gossett
1908, Ireland
Guiness Brewery
chemist
t- distribution

$\boxed{\text{ex:}}$ confidence level notation

$\underline{287.7} < \mu < \underline{290.3}$ PCB levels

If we took 100 samples of sample size n=10, we expect to capture the popln mean (μ) of PCB levels 95 times

$\boxed{\text{Steps}}$

Go to T-table, down using degrees of freedom and the row is confidence levels (or area under tail.) That gives you Tc (which plug into equation to get confidence interval)

$$\boxed{\varepsilon = t_c\left(\frac{s}{\sqrt{n}}\right)}$$

$\boxed{\text{Checks}}$

1) RRS
2) Independent
3) normally distributed (appx) or n>30 or NQP is linear /pearsons index

7.3

based off binomials: $\boxed{\frac{r}{n} = \hat{p}}$ ← sample percent / proportion

- $\hat{p} = r/n$
- $\boxed{E = z_c\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)}$
- $\hat{p} - E < p < \hat{p} + E$
  - p/π is population % / proportion
  - $\hat{p}$ is sample % / proportion
- Checks
  1) RRS
  2) Indep
  3) np and $nq \geq 10$

  "ANAIJ"

estimating n

7.1

$n = \left(\frac{z_c \sigma}{E}\right)^2$

round up!

7.3

preliminary estimate → $\boxed{n = pq\left(\frac{z_c}{E}\right)^2}$

smaller n

or

$\boxed{n = .25\left(\frac{z_c}{E}\right)^2}$

starting from scratch, n bigger

7.1 (CI w/ θ)

Zc	Cl/l
1.28	80%
1.645	90%
1.96	95%
2.58	99%

$\varepsilon = Z_c\left(\frac{\theta}{\sqrt{n}}\right)$

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

$n = \left(\frac{Z_c \theta}{\varepsilon}\right)^2$ round up

Checks

1) RRS
2) Indep
   - $n \le N/10$ ~ it is rsnbl...
3) normal → ②
4) θ known → ②

note... $PI\left(\frac{3(\bar{x} - \text{med})}{s}\right)$ or NQP

7.2 (CI w/o θ)

$\varepsilon = t_c\left(\frac{s}{\sqrt{n}}\right)$ $\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon$

df = n−1 (round ↓ if not on table)

find tc by df + cl/l

label ci's

Checks

1) RRS
2) Indep
   - $n \le N/10$ ~ it is rsnbl..
3) normal → ①
4) θ unknown / s known

7.3 (cl w/ proportions)

$\varepsilon = Z_c\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$ $(\hat{p} = r/n)$

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

?? $n = pq\left(\frac{zc}{\varepsilon}\right)^2$ ← if given preliminary estimate

or $n = .25\left(\frac{zc}{\varepsilon}\right)^2$ ← if not, larger estimate

Checks

1) RRS
2) Indep
   - $n \le N/10$ ~ it is rsnbl..
3) np, nq > ≥10, thus ANAIJ

On calculator

Stat → tests → 7.1 (z int), 7.2 (T int), 7.3 (1-prop z int) ← x is r

William Gossett, 1908 Ireland, Guinness Brewery

Conclusion

If we took — samples all of size n = x, we expect to capture popln mean (μ) ______ on — occasions

8.1 (Test Overview)

Type I: H6 when its true : α

Type II: Ho✓ when false : B

4 ingredients of test: Ho, Ha, CV, + sampe stat/point estimate (+ pvalue)
turns into cv

8.2 Part 1 (σ Known)

$\alpha = 0.05$ auto-go-to when not stated

z-test α	one-tail	two-tail
0.01	± 2.33	± 2.58
0.05	± 1.645	± 1.96
0.10	± 1.28	

Conclusion

- __ the null, __ the alternate, α = __
- 3x __ statistical evidence to suggest __ ∴, __
- p-lvl = __ □ α = __

8.2 Part 2 (σ unknown)

t-test instead of z

to find tc, use df + tail area(α)

for p-value, you will get a range

$\bar{X} \to Z: \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$\bar{X} \to t: \frac{\bar{x}-\mu}{s/\sqrt{n}}$

8.2 (p-values)

- For Z: when 2 tail, x by 2!!
  small has to go (compared w/ α)
  "weun" or "seun"
- For t: use calculated t-value and see where it falls on row (use df for row, + see what α's it falls between) - get a range!

Same checks as intervals

8.3 (Z test for ps)

$\hat{p} = r/n$, binomials

$\hat{p} \to Z: \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$

Checks

1) RRS
2) indep
   - $n \leq N/10$ ✓ isnbl...
3) np, nq ≥ 10, ANAIJ

2/6/24

6.6 Normal Appx to Binomial

- Fits binomial situation (n, and set p)
- $np > 5$ and $nq > 5$, then r has a binomial distribution that is approximated by a normal distribution with

CLT was mean of a group of things, this is %.

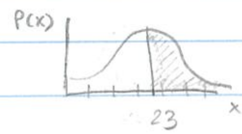
  - $\mu = np$
  - $\theta = \sqrt{\mu q}$
- more n values approaches normal distribution and can be approximated to normal
- Continuity correction
  - r if left point : subtract 0.5 to obtain the corresponding x-value
    $X = r - 0.5$
  - r if right point : add 0.5 to get the corresponding x value
    $X = r + 0.5$
    example, $P(6 \le r \le 10)$ would be $P(5.5 \le x \le 10.5)$ (as histogram includes lower + upper bound) *
    if $P(r \le 6)$ would be $P(x \le 6.5)$, if $P(r < 6)$ would be $P(x < 5.5)$

Example: $n = 40$, $p = 0.5$

a) $np = 20 > 5$, thus normal appx is justified (ANAIJ)
$nq = 20 > 5$, thus normal appx is justified (ANAIJ)

b) compute μ and θ of normal distr approximated
$\mu = np = 20$
$\theta = \sqrt{\mu q} = 3.16$

c) continuity correction for $r \ge 23$ to normal variable x
23 is lower boundary, so $P(x > 22.5)$



Sampling distribution of $\hat{p}$:

$\hat{p} = r/n$ $\bar{x}$ for binomial

- $\mu_{\hat{p}} = p$
- $\theta_{\hat{p}} = \sqrt{\frac{pq}{n}}$