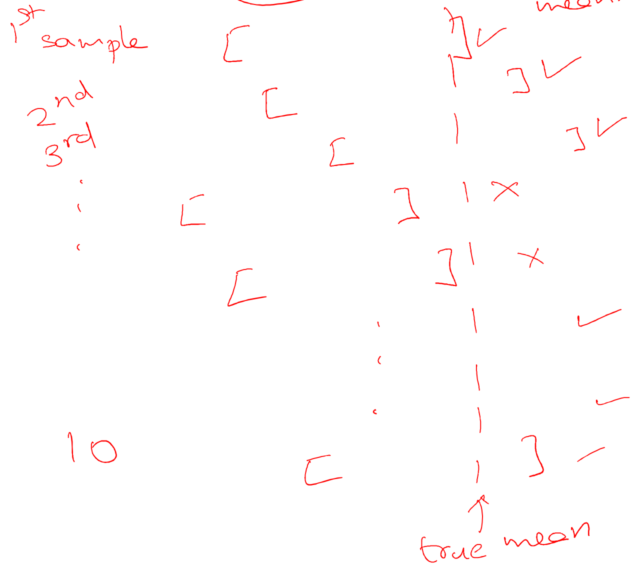
Statistical Intervals

Confidence intervals: interval estimate for a population parameter

(Not necessary that CI contains true mean!)

90% CI

9/10 times my interval will capture the mean.



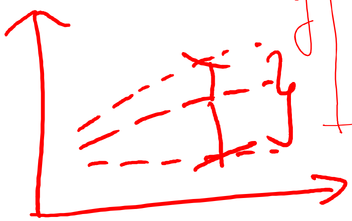
Tolerance interval: interval that bounds value within some proportion

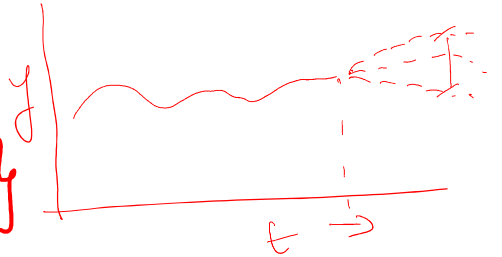
Prediction interval: bounds future observations (predictions)

Precision of estimate

Narrower the interval ⟺ Better the precision

- reduce % confidence ✗
- increase sample size ✓





90% [ ] 75% [ ]

CI for the mean of a RV $\sim N(\_\_, \_\_)$, with known variance σ^2

$l \leq \mu \leq u$, $\quad P(l \leq \mu \leq u) = \{(1-\alpha)\}$ Confidence level

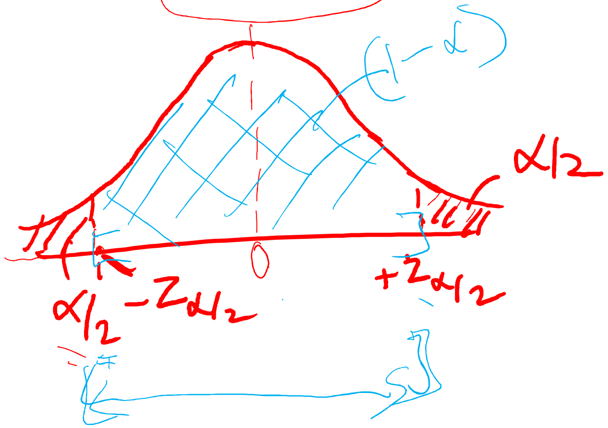
95%, $\alpha = 5\%$

α – significance level

std. normal variate

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$P(l \leq \mu \leq u) = (1-\alpha) = P\left\{-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\}$$



$(1-\alpha)$ CI: $\quad \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$

Precision of CI: $\boxed{2\, z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}}$ $\quad$ ↑ precision ↑ n - sample size

Error, $E = |\bar{X} - \mu| \leq z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$

Determine sample size required given a certain error tolerance..

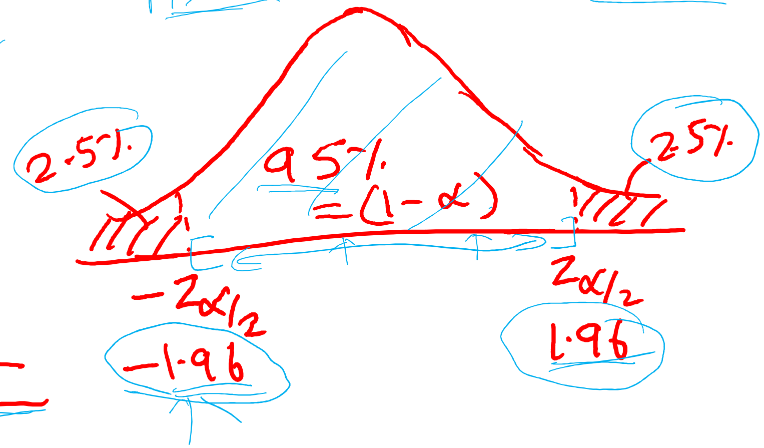
sample size, $n = \left(\frac{z_{\alpha/2}\cdot\sigma}{E}\right)^2$

E.g. Speeds of vehicles: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3

$n = 10$

$\sim N$, $\sigma^2 = 1 \text{ kmph}^2$

Construct a 95% CI for the mean.



$$\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad \sqrt{n} = \sqrt{10}$$

20

64.46

$$63.84 \le \mu \le 65.08 \leftarrow 95\% \text{ CI}$$

64.46

Tolerance interval: 95%

$$64.46 - 1.96 \cdot 1 \le \mu \le 64.46 + 1.96 \cdot 1$$

E.g What should be the sample size to keep error within a limit of 1 kmph with 95% confidence

$$N = \left(\frac{z_{\alpha/2} \cdot \sigma}{E = \frac{1}{2}}\right)^2 = \left(\frac{1.96 \cdot 1}{0.5}\right)^2$$

$\simeq$ 16 samples

Large Sample CI (n is large ⟺ CLT is applicable)

$X_1, X_2, \dots X_N$ — random sample not necessarily from Normal dist.
with a sample mean $\bar{X}$, variance S^2

From CLT : $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \approx$ approx. Normal

$$\bar{X} - z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}$$ $(1-\alpha)100\%$ CI

CI of mean, $\boxed{RV \sim N(-,-)}$, variance is unknown or sample size is small ✓

$\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \approx$ t-dist

Student's t-distribution with $(n-1)$ d.o.f

t-dist $\Rightarrow$ normal dist for large n

$X_1, X_2, \dots \sim$ Normal distribution

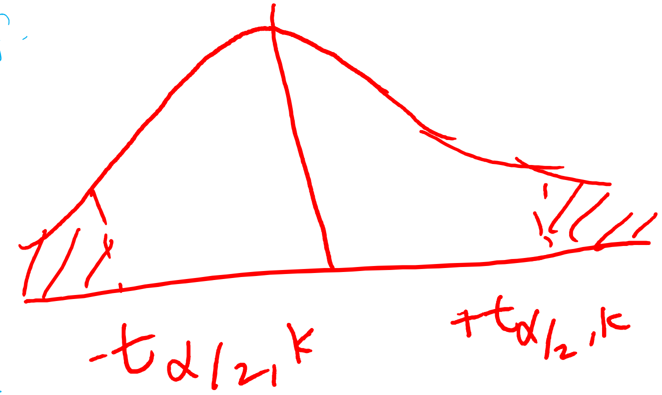
sample $\sim N(\cdot, \cdot)$

PDF of a t-dist: $f(t) = \dfrac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma[k/2]} \cdot \dfrac{1}{[(t^2/k)+1]^{\frac{k+1}{2}}}$

$$\boxed{\Gamma(r) = \int_{-\infty}^{\infty} x^{r-1} e^{-x}\,dx}$$

k - d.o.f.

(n-1) d.o.f.



CI for the mean

when variance is unknown / small sample

X_i's $\sim N(-,-)$

$$\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

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

S^2 - sample variance

CI on the variance of a normally dist. RV

χ^2 dist : $\boxed{X_1, X_2, \dots X_n \sim N(\mu, \sigma^2)}$

& let s^2 be the sample variance, then

$$\chi^2 = \boxed{\frac{(n-1)s^2}{\sigma^2}}$$ statistics has a χ^2 dist with $(n-1)$ dof

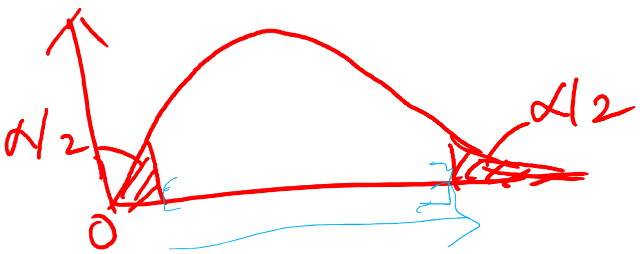
PDF of $$\chi^2 = \frac{1}{2^{k/2}\,\Gamma(k/2)}\, x^{(k/2)-1} e^{-x/2}$$

$x > 0$

k - d.o.f.

$$E[x] = k, \quad V[x] = 2k$$

CI on variance



$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \boxed{\sigma^2} \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2),n-1}}$$

est. pop variance

Large sample CI for a population proportion

Let X obs. out of N trials be 'True' / 'Yes'

$\hat{p} = \frac{X}{N}$ is a point est. of 'p'

Use CLT! Sampling dist. of $\hat{p}$ is approx. Normal (from CLT)

$E[\ ] = p$ & $V[\ ] = \frac{p(1-p)}{n}$

CI on a binomial pop.

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \le p \le \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Thumb rule: np & $n(1-p) \geq 5$ - we can wk above.

E.g. (8-8)

$N = 85$, engine crankshaft bearing

10 bearing do not meet specification for roughness. Compute 95% CI for p

$\hat{p} = \frac{10}{85} = 0.12$

$$0.12 \pm 1.96\sqrt{\frac{0.12(1-0.12)}{85}}$$

$$\boxed{5\% \le p \le 18.9\%}$$