Example · toss coin 100 times. prob get ≥ 90 Hs?

· voters vote for incumbent with prob. 0.~~6~~ If we take a sample of 50 voters, what is prob ≥ 25 say they will vote ~~of~~ for incumbent?

§6.1 confidence intervals

so far: given a distribution $X \sim N(\mu, \sigma)$ we take a sample of size n, the sample mean ~~$\bar{x}$~~ ~~$\bar{x}$~~ $\bar{x}$ has distribution $N(\mu, \frac{\sigma}{\sqrt{n}})$.

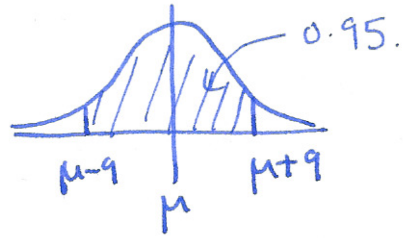
problem: in practice, we don't know μ or σ for X.

Example Lets assume we know σ, but not μ.

SRS of SAT scores of size $n = 500$. Assume $\sigma = 100$. Don't know μ.

but, $\bar{x}$ has distribution approx $N(\mu, \frac{\sigma}{\sqrt{500}})$.

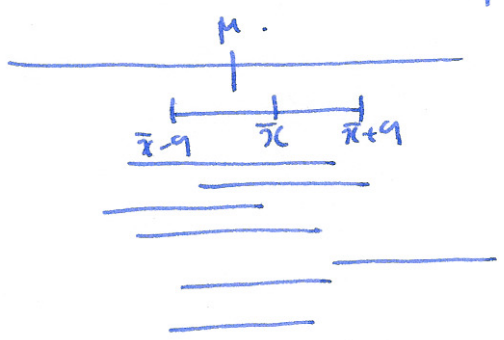
$$\frac{\sigma}{\sqrt{500}} = \frac{100}{\sqrt{500}} \approx 4.5$$



68-95-99.7 rule says that with probability 0.95 $\bar{x}$ is with 2 s.d. $= 9$ of μ. $\Leftrightarrow$ μ is within 2 s.d $= 9$ of $\bar{x}$.

conclusion: about 95% of all samples of size 500 contain μ in interval $[\bar{x}-9, \bar{x}+9]$

picture:



← approx 1 in 20 doesn't hit μ.

if $\bar{x} = 495$ then interval is $[486, 504]$.

there are two possibilities for our SRS: μ is contained in $[486, 504]$ / μ not contained in $[486, 504]$. } don't know which!

$[486, 504]$ is a 95% confidence interval for μ means:

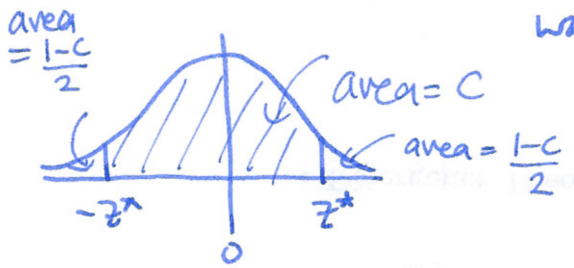
"we arrived at these numbers by a procedure which gives correct results 95% of the time".

Summary
- population dist. with mean μ (unknown).
  s.d. σ (known).

- SRS of size n with sample mean $\bar{x}$.

- the confidence interval of level C ($0 < C < 1$) is $[\bar{x}-m, \bar{x}+m]$
  where $m = z^* \frac{\sigma}{\sqrt{n}}$. and $z^* = \text{qnorm}\left(\frac{1-C}{2}\right)$.

- conclusion: $C\%$ of confidence intervals constructed in this way will contain the population mean μ.



z^* is the z value for which $N(0,1)$ has $C\%$ of the distribution lying between $-z^*$ and $+z^*$.

Q: how big does the sample need to be? $m = \frac{z^*\sigma}{\sqrt{n}} \Leftrightarrow n = \left(\frac{z^*\sigma}{m}\right)^2$.
so if you want a confidence interval of size $[\bar{x}-m, \bar{x}+m]$, choose $n \geq \left(\frac{z^*\sigma}{m}\right)^2$.

recall

C	.90	.95	.99
z^*	1.645	1.960	2.576

Warnings these procedures require:
- sample is an SRS.
- can't work with data with unknown bias.
- $\bar{x}$ sensitive to outliers.
- if n is small (≤ 15), and population not normal, $\bar{x}$ is only approximately normal ($n \geq 15$ usually OK unless very skew / many outliers).
- this assumes σ is known. Very unlikely! We'll see what to do about this later.
- this is the best case confidence interval! Assuming no errors in transcribing data, doing calculations, etc.