3/13 Central Limit Theorem (cont).

If $X_1, \ldots, X_n$ iid from population with mean μ and std. deviation σ,

- $\bar{X} \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ moreso as $n \to \infty$
- $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z \sim N(0,1)$ "

$S_n = \sum_{i=1}^{n} X_i$ where $X_1, \ldots, X_n$ iid from pop with mean μ std dev σ (ie var is σ^2)

(sample sum)

$$E(S_n) = E(X_1 + \cdots + X_n)$$
$$= \sum_{i=1}^{n} E(X_i) = n \cdot \mu$$

$$Var(S_n) = Var(X_1 + \cdots + X_n)$$
$$= \sum_{i=1}^{n} Var(X_i) = n \cdot \sigma^2$$

so $SD(S_n) = \sqrt{n} \cdot \sigma$

$$\boxed{\text{CLT says} \quad S_n \sim N(n\mu, n\sigma^2) \text{ approx}}$$

$$\bar{X} = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \Leftrightarrow n\bar{X} = S_n$$

Say you have 2 populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2

take samples of sizes n_1, n_2

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$\mu = 104$ sec , $\sigma = 10$ sec.

~~P~~ sample size $n = 40$

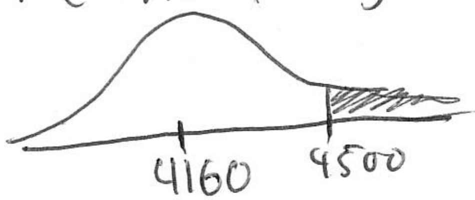
$P(S_n > 75 \text{ min} = 4500_{\text{sec}})$ $\qquad 40 \cdot 10^2$

CLT: $S_n \sim N(n \cdot \mu, n\sigma^2)$ $\quad \sigma^2_{S_n} \quad \sigma_{S_n} = \sqrt{4000}$

$\Rightarrow N(4160, 4000)$

$P(S_n > 4500) = \text{normalcdf}(4500, 100000, 4160, \sqrt{4000})$

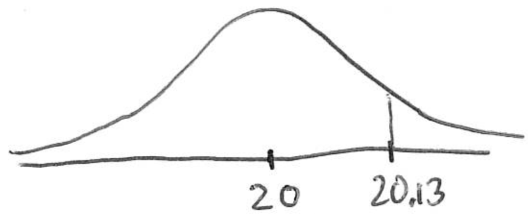
≈ 63 sec



≈ 0

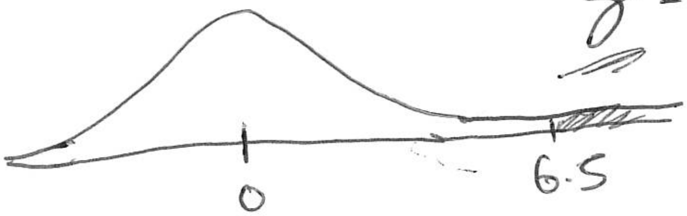
same as asking $P\left(\bar{X} > \overset{112.5}{\dfrac{4500}{40}}\right)$

if $\mu = 20$, $\bar{X} \sim N\left(20, \frac{.2^2}{100}\right)$ or $\sigma_{\bar{x}} = \frac{.2}{\sqrt{100}} = \frac{.2}{10} = .02$



~~What is~~

$$z = \frac{20.13 - 20}{.02} = 6.5$$



empirical rule
99.7% within 3σ

within

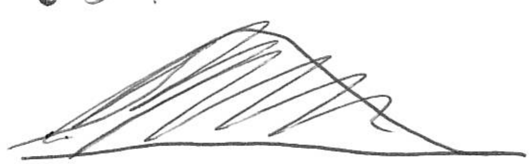
$$P(Z > 6.5) \approx 0$$

An estimator is called consistent if
T_n of θ
if $T_n \to \theta$ as $n \to \infty$
eg. $\bar{X}$ to estimate μ

Say in May avg temp is $70°$ with $\sigma = 5°$
in Dec avg temp is $30°$ with $\sigma = 7°$

May 12 it was $78°$, Dec 15 it was $40°$
Which is a more rare event?

$$z_{May} = \frac{78° - 70°}{5°} = 1.6 \qquad z_{Dec} = \frac{40 - 30}{7} = 1.429$$



$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

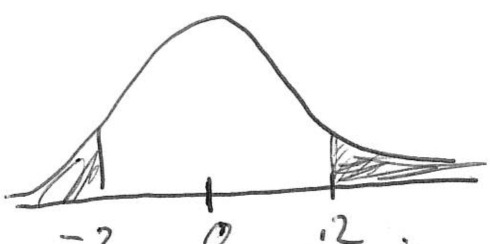
ex) Apples from orchard 1 ~~supposedly~~ have ~~mean weight of 15 oz, with~~ std. dev. of .8 oz,
From orchard 2, ~~sup~~ m~~ean supposed~~
stdev of .9 oz.
We want to see whether they apples in orchard 1 have a greater mean weight.

Sample randomly 37 from 1 get $\bar{X}_1 = 15.3$ oz
sample 45 from 2 get $\bar{X}_2 = 15.1$ oz
$\bar{x}_1 - \bar{x}_2 = .2$ oz

suppose $\mu_1 = \mu_2$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \equiv N(0, .03529)$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = .1879$$



$$P(\bar{X}_1 - \bar{X}_2 > .2) = .1436$$

inconclusive - assume equal mean weights.

normalcdf(.2, 1000, 0, .1879)