Read chapter 1.

Sample mean: $X_1, X_2, \cdots X_n$ a random sample from a popn.

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Sample variance:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

$\bar{X}$, S^2 are both sample statistics, random variables.

μ, σ^2 — are both popn parameters, numbers.

Ex: a random sample of size 6,

4, 3.2, 5, 3.9, 4.2, 3.7

$\bar{X} = 4$

$$S^2 = \frac{0.8^2 + 1^2 + 0.1^2 + 0.2^2 + 0.3^2}{5} = \_\_\_$$

$S^2,\ O^2$

chapter 8 Sampling dist

population $\xleftrightarrow{\text{inference}}$ random sample (subset of popu)

parameter (constant)

$\mu, \sigma^2, p, \beta, \lambda$

statistic (random variable)

$\bar{X}, S^2, m, R, \max,$

$\begin{cases} \text{estimation} \\ \text{hypothesis testing} \end{cases}$

what's the (sampling) dist of $\bar{X}$?

- The sampling dist of $\bar{X}$.

Ex: $X_1, X_2, X_3 \sim N(70, 10^2)$

① $P(\bar{X} > 66.67)$

$$= P\left(\frac{X_1 + X_2 + X_3}{3} > 66.67\right)$$

$$= P(\underbrace{X_1 + X_2 + X_3}_{N(210,\ 300)} > 200)$$

$$= P\left(Z > \frac{200-210}{\sqrt{300}}\right) = P\left(Z > -\frac{1}{\sqrt{3}}\right)$$

$$= P(Z > -0.58) = 0.7190$$

Thm: Let $X_1, X_2, \ldots X_n$ be a random sample of size n from $N(\mu, \sigma^2)$

then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{X_1}{n} + \frac{X_2}{n} + \cdots + \frac{X_n}{n}$$

$$E(\bar{X}) = \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = \mu.$$

$$Var(\bar{X}) = \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \cdots + \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Q: What if we take away the normal assumption in the underlying dist?

We still have $E(\bar{X}) = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$, But not the Normal part.

Ex: $X_1, X_2, \cdots X_n \sim N(80, 8^2)$.

① $P(|X_1 - 80| < 5) = P\left(\frac{|X_1 - 80|}{8} < \frac{5}{8}\right)$

$= P(|Z| < 0.625) = 1 - 2 * 0.2643$
$\qquad\qquad\qquad\qquad\qquad 5286$

$= 0.4714$

② if $n = 4$, what is $P(|\bar{X} - 80| < 5) =$

$\bar{X} \sim N\left(80, \frac{8^2}{4}\right) = N(80, 4^2)$ $= P\left(|Z| < \frac{5}{4}\right)$

$= P(|Z| < 1.25)$

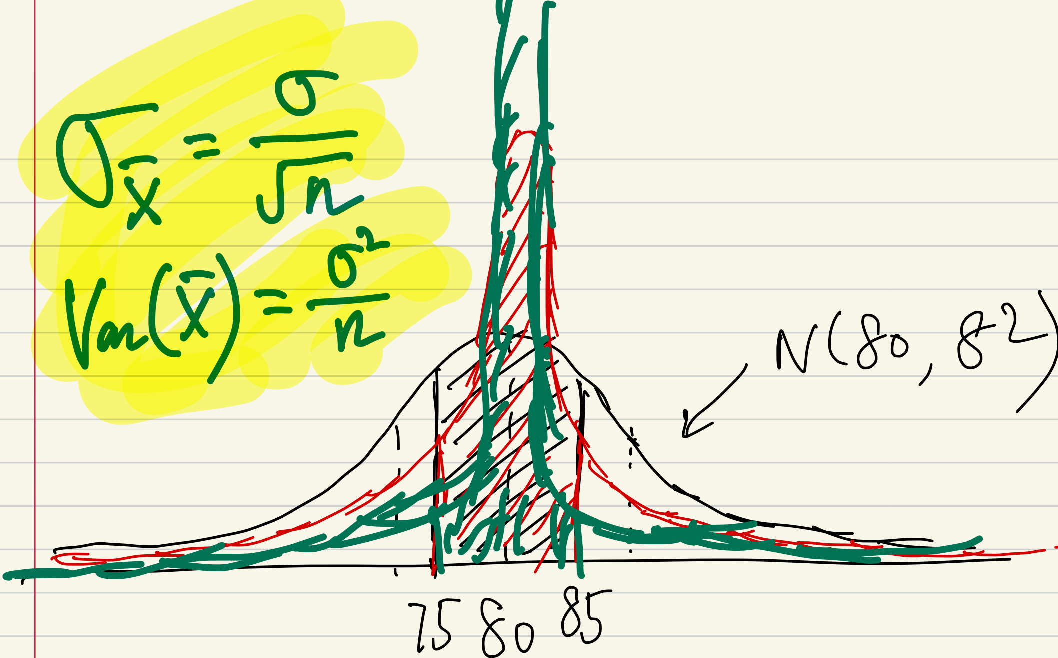
$= 0.7888$

③ if $n = 16$, what is $P(|\bar{X} - 80| < 5) = P(|Z| < 2.5)$

$\bar{X} \sim N\left(80, \frac{8^2}{16}\right) = N(80, 2^2)$ $= 0.987$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$



$\bar{X}_{n=4} \sim N(80, 4^2)$

$\bar{X}_{n=16} \sim N(80, 2^2)$

The Central Limit Theorem.

Let $X_1, X_2, \cdots X_n$ be a random sample from any dist with mean μ and variance $\sigma^2 < \infty$.
Then the limit dist of $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is Z as $n \to \infty$.

In appilication, when n is large, we say the dist of $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is approx Z.

$$\iff \bar{X} \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\iff \sum_{i=1}^{n} X_i \overset{\text{approx}}{\sim} N(n\mu, n\sigma^2)$$

$n \geq 30$ is enough.

n can be smaller here.