Ch. 7 - Normal Probability Approximations

1 Central Limit Theorem

Some terminology:

Population vs. sample, parameter vs. statistic
Population: contains the entire collection of individuals we want to study
Sample: subset of individuals selected from the population

Parameter: characteristic of interest from the population. Value of the parameter is unknown in practice
Statistic: numerical measure of the sample. We use statistics to estimate the unknown population parameter. Due to sampling variability a statistic takes on different values for different samples.

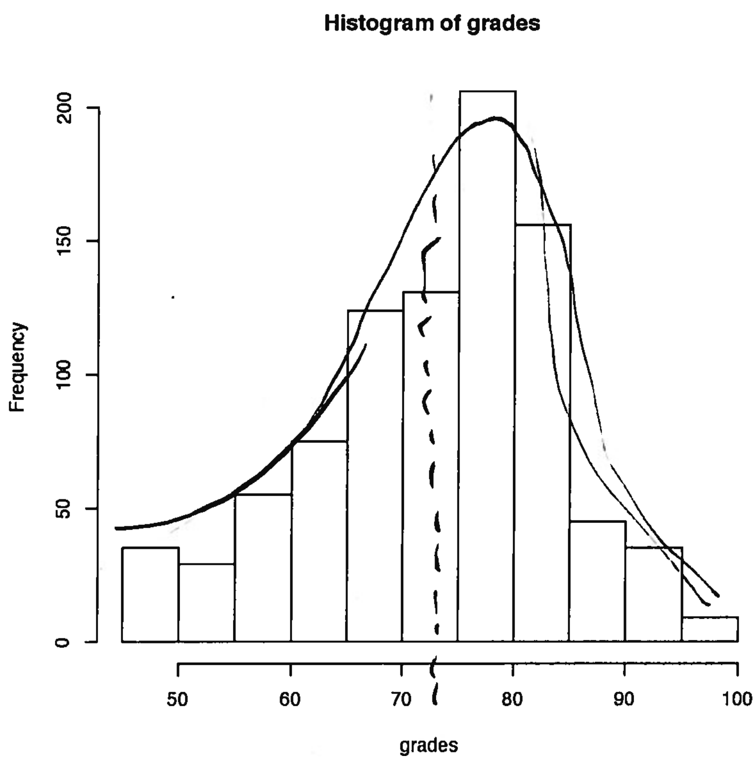
Central Limit Theorem

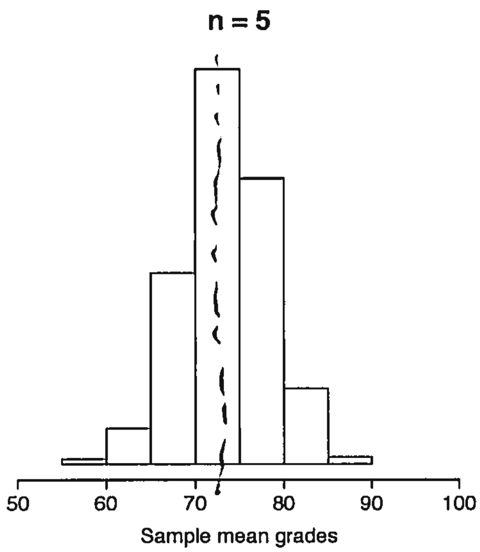
1. Let $X_1, X_2, ..., X_n$ be an independent random sample of size n taken from any distribution with mean μ and variance σ^2.

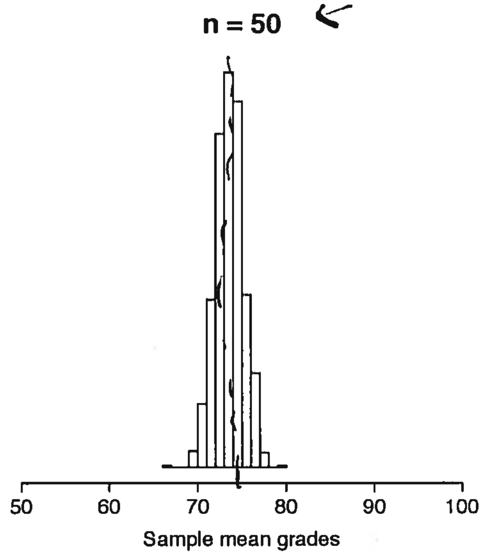
2. If n is large (book says $n \geq 20$), then

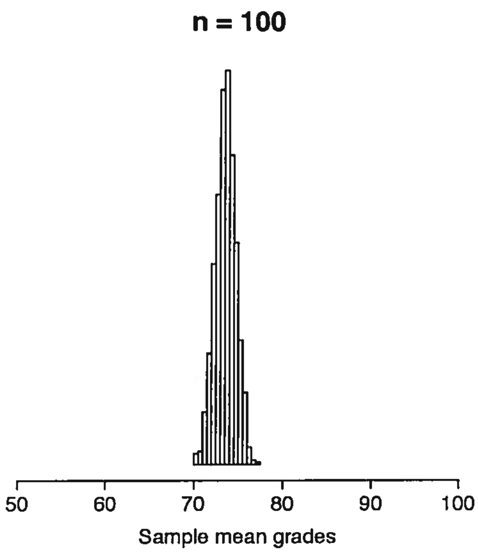
$$\bar{X} \overset{approx}{\sim} N(\mu, \frac{\sigma^2}{n})$$

In words: For a sufficiently large, independent random sample taken from a population with mean μ and standard deviation σ, the sample mean follows approximately a Normal model with mean μ and standard deviation $\sigma/\sqrt{n}$, even if the underlying distribution of the individual observations in the population is not Normal.









$\frac{\sigma^2}{n}$

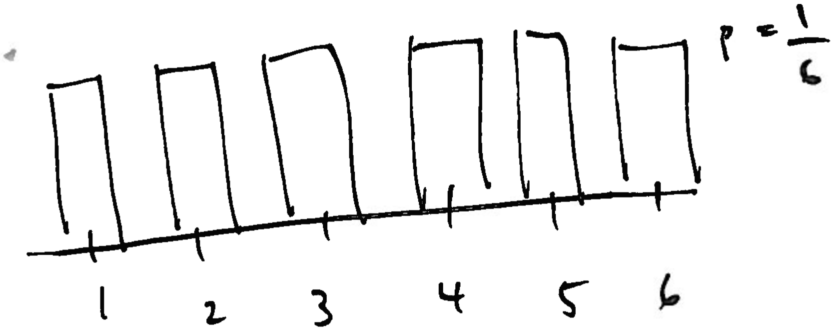
Note that if a question ask about a sum instead of an average, you can still use CLT. Let T be the sum of an independent random sample $X_1, X_2, ..., X_n$.

$$T = X_1 + X_2 + ... + X_n$$

$$T = \sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \implies T = n\bar{X}$$

$$\begin{aligned}
E(T) &= E(n(\bar{X})) \\
&= nE((\bar{X})) = n\mu \\
Var(T) &= Var(n\bar{X}) \\
&= n^2 Var(\bar{X}) \\
&= n^{\cancel{2}} \sigma^2 / \cancel{n} \\
&= n\sigma^2 \\
T &\sim N(n\mu, n\sigma^2)
\end{aligned}$$

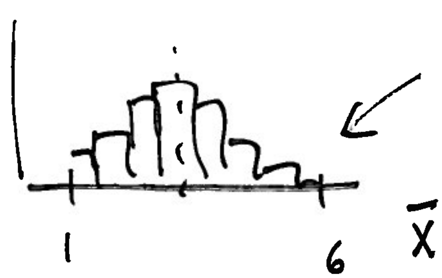
ex. Fair die



Roll 3 die : $n = 3$

Sample	Values	Mean
1	{1, 2, 3}	2
2	{2, 2, 2}	2
3	{6, 6, 6}	6
4	{1, 5, 6}	4
⋮	⋮	⋮

Roll {6, 6, 6}



$n = 100$?

Example 1. Example 7.1 from your text.

A system consists of 25 independent parts connected in such a way that the ith part automatically turns–on when the $(i-1)th$ part burns out. The expected lifetime of each part is 10 weeks and the standard deviation is equal to 4 weeks.

(a) Calculate the expected lifetime and standard deviation for the the system.

(b) Calculate the probability that the system will last more than its expected life.

(c) Calculate the probability that the system will last more than 1.1 times its expected life.

(d) What are the (approximate) median life and interquartile range for the system?

Solution:

$n = 25$

Let X_i denote the lifetime of the ith component

a) $E(X_i) = 10$ weeks $\qquad X_i$? distribution is unknown.

$SD(X_i) = 4$ weeks

$T = X_1 + \ldots + X_{25}$ be the lifetime of the entire system.

$$E(T) = E(X_1 + \ldots + X_{25}) = 25\,E(X_i)$$
$$= 25 \times 10 = \underline{\underline{250}}$$

$$Var(T) = Var(X_1 + \ldots + X_{25}) = Var(X_1) + \ldots + (X_{25})$$
(X_is independent)
$$= 400$$
$$SD(T) = 20 \text{ weeks}$$

b) $P(T > E(T)) = P(T > 250) = ?$

Note: CLT: $n = 25 > 20$

$$\bar{X} \sim N\left(\mu = \underline{\underline{10}}, \frac{\sigma^2}{n} = \frac{16}{25}\right) \text{ approx.}$$

$$T \sim N(\underline{\underline{n}}\mu = 250,\ \underline{\underline{n}}\sigma^2 = 400)$$

$T \sim N(250, 400)$

$$P(T > 250) = \frac{1}{2}$$

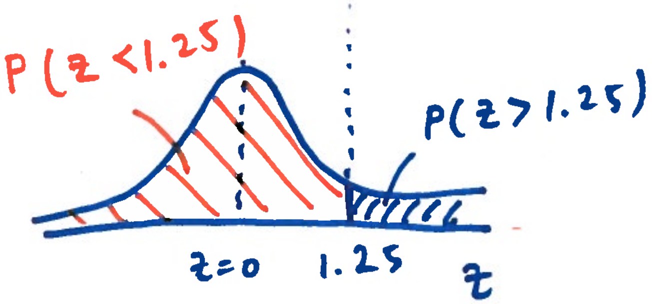
c) $P(T > 1.1E(T)) = ?$ $\qquad T \sim N(250, 400)$ approx.

$$P(T > 275) = P\left(Z > \frac{275 - 250}{\sqrt{400}}\right)$$

$$= P(Z > 1.25)$$

$$= 1 - P(Z < 1.25)$$

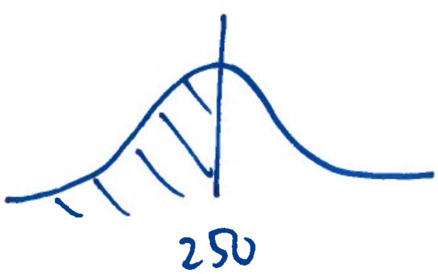
$$= 1 - 0.8944$$



$$= 0.1056$$

d) median and IQR ?

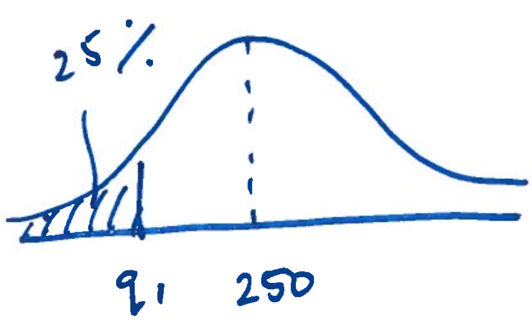
Since $T \sim N(250, 400)$ approx



median = 250 weeks

$IQR = q_3 - q_1$

Let q_1 be lower quartile

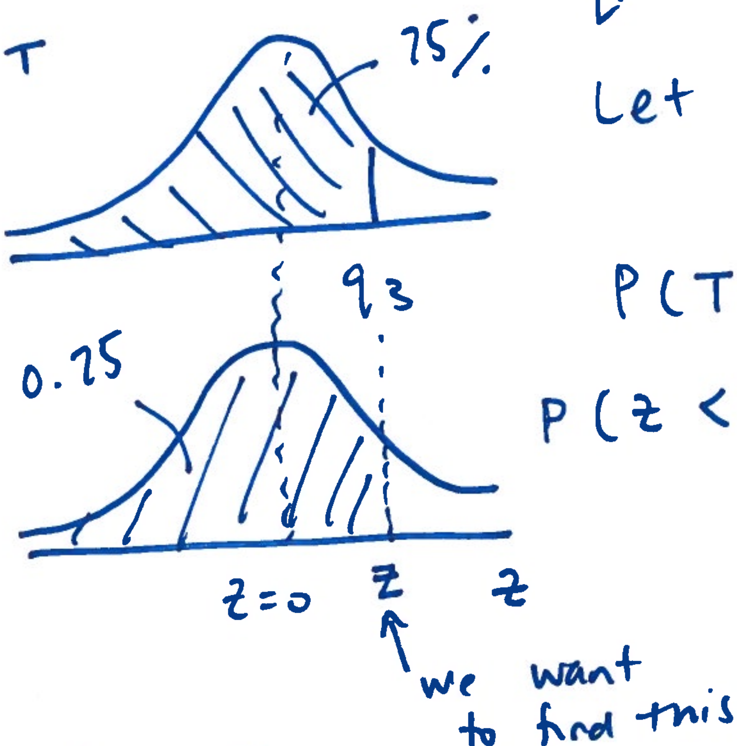


$P(T < q_1) = 0.25$

$$P\left(Z < \frac{q_1 - 250}{20}\right) = P(Z < -0.675)$$

$$\frac{q_1 - 250}{20} = -0.675$$

$$q_1 = 236.5$$



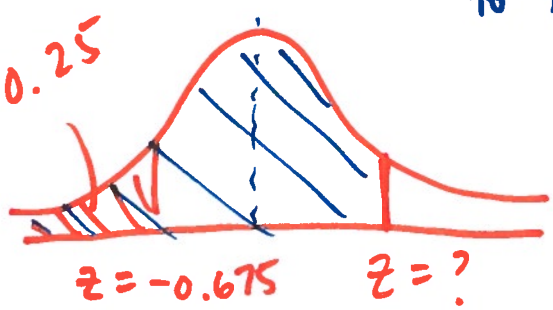
Let q_3 be upper quartile

$$P(T < q_3) = 0.75$$

$$P\left(Z < \frac{q_3 - 250}{20}\right) = P(Z < 0.675)$$

$$\frac{q_3 - 250}{20} = 0.675$$

$$q_3 = 263.50$$



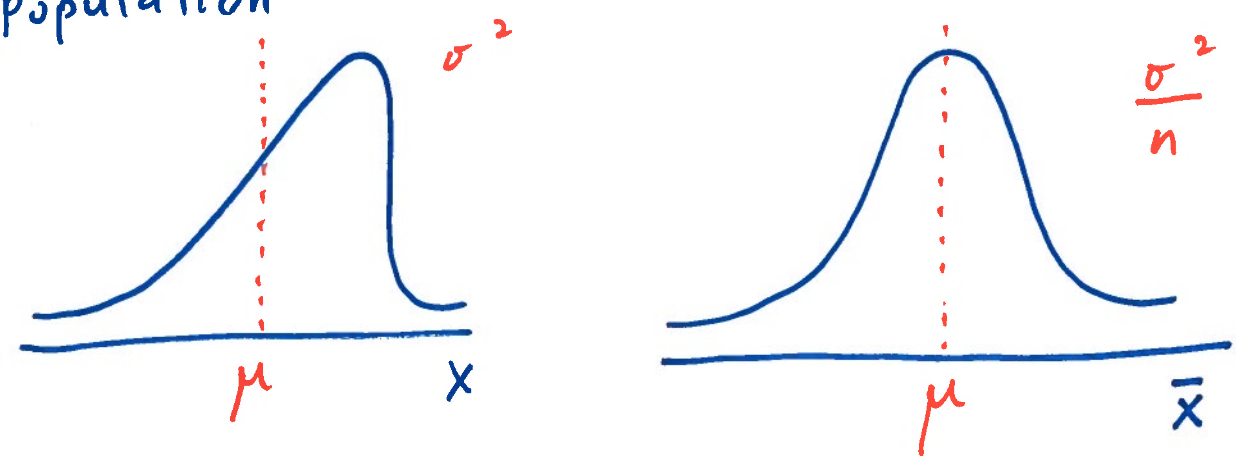
$$IQR = 263.50 - 236.5$$
$$= 27$$

Last lecture we learned about the Central Limit Theorem (CLT):

$X_1, \ldots, X_n$ is independent random sample of size $n \geq 20$ taken from <u>any</u> distribution

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

Population



Example 2. A computer generates a random variable X whose probability distribution is given in the following table:

x	0	2	4	6
$P(X = x)$	0.1	0.2	0.3	0.4

(a) Show that $Var(X) = 4$.

(b) Find $E(X^4)$ and $Var(X^2)$

(c) The sum of 100 independent observations is denoted by S. Describe fully the approximate distribution of S.

<u>Solution:</u>

a) $E(X) = (0 \times 0.1) + (2 \times 0.2) + (4 \times 0.3) + (6 \times 0.4) = 4$

$Var(X) = E(X^2) - [E(X)]^2$

$E(X^2) = (0^2 \times 0.1) + (2^2 \times 0.2) + (4^2 \times 0.3) + (6^2 \times 0.4) = 20$

$Var(X) = 20 - 4^2 = 4$

b) $E(X^4) = (0^4 \times 0.1) + (2^4 \times 0.2) + \ldots + (6^4 \times 0.4) = 598.4$

$Var(X^2) = E(X^4) - [E(X^2)]^2$

$= 598.4 - 20^2 = 198.4$

pop. mean

pop Variance of σ^2

approx.

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ for sufficiently large n.

Note: population is not normal but we know mean is 4 and variance is 4.

(part a)

$S = X_1 + X_2 + \dots + X_{100}$

$\bar{X} = \frac{\Sigma X_i}{n} = \frac{S}{n}$

by CLT,

$\frac{\sigma^2}{n}$

$\bar{X} \sim N\left(4, \frac{4}{100}\right)$ approx.

$\bar{X} \sim N(4, 0.04)$

$n\bar{X} \sim N(400, 400)$

$E(n\bar{X})$
$= nE(\bar{X})$
$= 100 \times 4$
$= 400$

$Var(n\bar{X}) = n^2 Var(\bar{X})$
$= 100^2 \times 0.04$
$= 400$

Hence $S \sim N(400, 400)$ approx.

up to test.