6.50 The Rh Factor In a certain population, 15% of the people have Rh-negative blood. A blood bank serving this population receives 92 blood donors on a particular day.

a. What is the probability that 10 or fewer are Rh-negative?

b. What is the probability that 15 to 20 (inclusive) of the donors are Rh-negative?

c. What is the probability that more than 80 of the donors are Rh-positive?

BINOMIAL: $\frac{\text{SAMPLE SIZE}}{\text{POP. SIZE}} < .05$ ✓ ASSUME TRUE.

n : 92 TRIALS
p : .15 PROB SUCCESS
q : .85 PROB FAILURE

MEAN $\mu = np = (92)(.15) = 13.8$

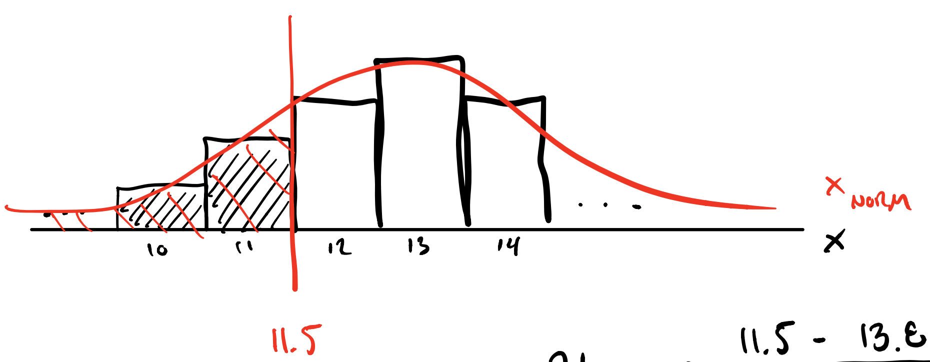
STAND DEV. $\sigma = \sqrt{npq} = \sqrt{(92)(.15)(.85)} = \sqrt{11.73}$

R.V. X = # SUCCESSES

(c) 92 PEOPLE : P(MORE THAN 80 FAILURES)

= P(LESS THAN 92 - 80 = 12 SUCCESSES)

$P(X < 12) \approx P(X_{NORM} \le 11.5)$



$= P\left(Z \le \frac{11.5 - 13.8}{\sqrt{11.73}}\right) = \ldots$ NormCDF(...)

§7.4, 7.5 Central Limit Theorem

Sampling Distribution: Distribution of a random variable

$\bar{X}$ = sample mean.

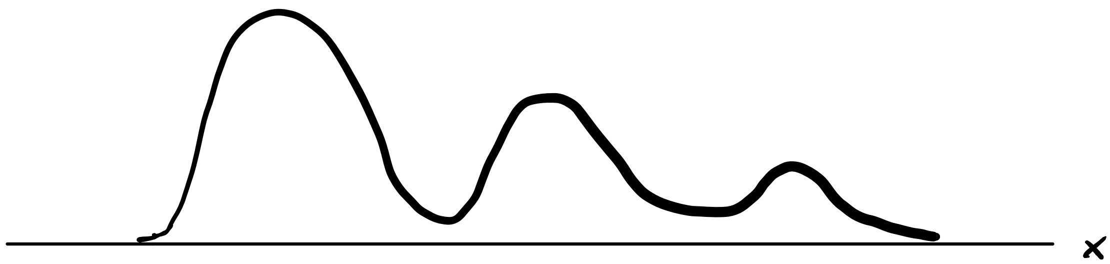
<u>Experiment</u> is to take an average (mean) of a sample

① Gathering a sample of size n

② Collect a measurement for each object in your sample

③ Record a single value for $\bar{X}$ = mean of those inividual measurements.

e.g. Let X = length of fish in Lake A.



$\bar{X}$ is the mean length of n randomly selected fish

IDEA

Averages of measurements of samples (sample means) tend to be close to the population mean.

Extreme case: sample size = population size.

DISTRIBUTION FOR X	DISTRIBUTION FOR $\bar{X}$ (SAMPLE MEAN) i.e. SAMPLING DISTRIBUTION
Distribution of individual measurements taken from a population	Distribution of averages (or sums) of n measurements taken from a random sample of n individuals.
The experiment is to record a single measurement.	The experiment is to take n measurements and average them together to produce the mean
The distribution of x can have any shape!	The distribution for sample means x-bar is approximately normal!

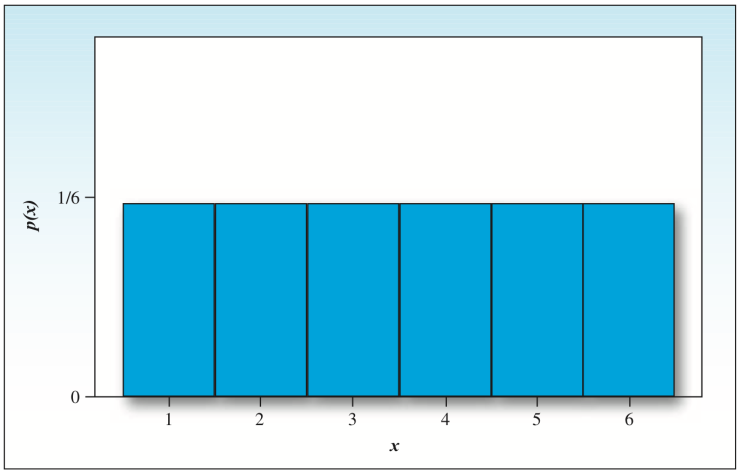
ROLL 2 DICE & RECORD $\bar{X}$ = MEAN OF 2 DICE

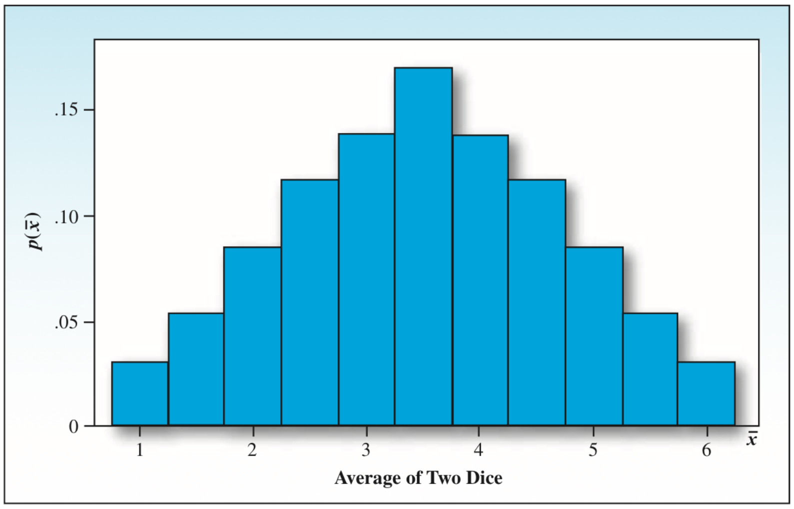
	1	2	3	4	5	6
1	1	1.5	2	2.5	3	3.5
2	1.5	2	2.5	3	3.5	4
3	2	2.5	3	3.5	4	4.5
4	2.5	3	3.5	4	4.5	5
5	3	3.5	4	4.5	5	5.5
6	3.5	4	4.5	5	5.5	6

$P(1) = 1/36$

$P(3.5) = 1/6$

$P(6) = 1/36$





ARBITRARY

Central Limit Theorem

If random samples of n observations are drawn from a nonnormal population with finite mean μ and standard deviation σ, then, when n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately normally distributed, with mean μ and standard deviation

$$S.E = \frac{\sigma}{\sqrt{n}}$$

The approximation becomes more accurate as n becomes large.

STANDARD DEVIATION FOR $\bar{X}$

S.E. (STANDARD ERROR) IS SMALLER THAN THE STANDARD DEV. FOR THE POPULATION (INDIVIDUAL MEASUREMENTS).

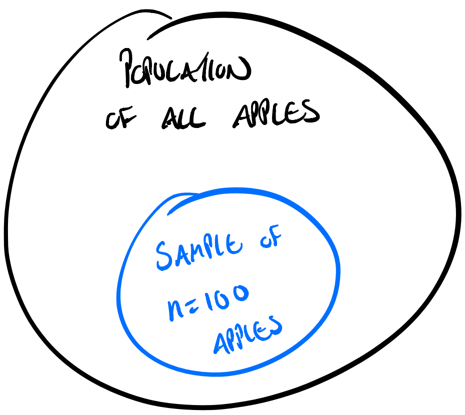
MEAN FOR SAMPLE MEANS = POPULATION MEAN μ

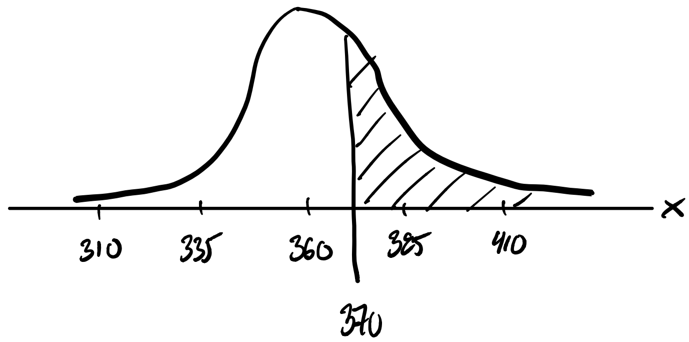
A SAMPLE SIZE n GETS LARGER, SAMPLE MEAN $\bar{X}$ HAS LESS & LESS VARIATION

STAND DEV. $\frac{\sigma}{\sqrt{n}}$ GETS SMALLER ex.

Suppose there is an apple orchard and let x be the weight a randomly selected apple. Suppose that mean weight of all apples in the orchard is 360g, and the standard deviation is 25g.

1. Assuming the distribution of weights of apples is normal, find the probability of selecting a single apple that weighs more than 370g.
2. Find the probability of selecting 100 apples with an average weight above 370g.





$$P(x \geq 370) = P\left(z \geq \frac{370-360}{25}\right)$$
$$= 1 - P(z \leq .4)$$
$$= 1 - .6554 = .3446$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5 871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879

$$z = \frac{\bar{x} - \mu}{S.E.}$$

$$P(\bar{x} \geq 370) = P\left(z \geq \frac{370-360}{2.5}\right) = P(z \geq 4) = 1 - P(z \leq 4)$$
$$\approx 1 - 1 \approx 0$$

$\bar{X}$ HAS MEAN $\mu = 360$ (SAME AS POPULATION)

STAND. DEV. OF $\bar{X}$ = S.E. $= \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{100}} = 2.5$

STANDARD ERROR

POPULATION	SAMPLE MEANS
MEAN μ	MEAN μ
STAND. DEV. σ	STAND DEV $\rightarrow$ STANDARD ERROR $S.E = \frac{\sigma}{\sqrt{n}}$

7.15 Random samples of size n were selected from populations with the means and variances given here. Find the mean and standard deviation of the sampling distribution of the sample mean in each case:

a. $n = 36, \mu = 10, \sigma^2 = 9$ $\rightarrow$ $\sigma = 3$

b. $n = 100, \mu = 5, \sigma^2 = 4$ $\rightarrow$ $\sigma = 2$

c. $n = 8, \mu = 120, \sigma^2 = 1$ $\rightarrow$ $\sigma = 1$

(a) MEAN FOR SAMPLE MEANS $\bar{X}$ IS $\mu = 10$

STANDARD DEVIATION FOR SAMPLE MEANS $\bar{X}$ IS $S.E. = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = \frac{3}{6} = .5$

(b) $\mu = 5$

$S.E = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = .2$

(c) $\mu = 120$

$S.E. = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{8}} \approx .3536$

7.25 Suppose a random sample of $n = 25$ observations is selected from a population that is normally distributed with mean equal to 106 and standard deviation equal to 12.

a. Give the mean and the standard deviation of the sampling distribution of the sample mean $\bar{x}$.

b. Find the probability that $\bar{x}$ exceeds 110.

c. Find the probability that the sample mean deviates from the population mean $\mu = 106$ by no more than 4.

(a) MEAN FOR $\bar{x}$ = MEAN FOR X = μ = 106

STANDARD DEV FOR $\bar{x}$ = STANDARD ERROR S.E. $= \dfrac{\sigma}{\sqrt{n}} = \dfrac{12}{\sqrt{25}} = 2.4$

(b) FIND $P(\bar{x} \geq 110)$ $\quad z = \dfrac{\bar{x} - \mu}{S.E.} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{x} - 106}{2.4}$

$$= P\left(z \geq \frac{110 - 106}{2.4}\right) = P(z \geq 1.67) = 1 - P(z \leq 1.67)$$

$$= 1 - .9525 = .0475$$

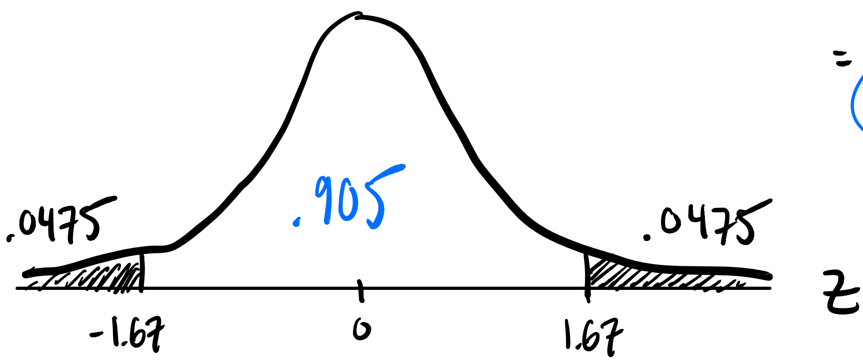
(c) FIND $P(110 - 4 \leq \bar{x} \leq 110 + 4)$

$P(106 \leq \bar{x} \leq 114)$

$P\left(\dfrac{106 - 110}{2.4} \leq z \leq \dfrac{114 - 110}{2.4}\right)$

$P(-1.67 \leq z \leq 1.67) = 1 - 2(.0475)$

$= .905$



The Central Limit Theorem can be restated to apply to the **sum of the sample measurements Σx_i,** which, as n becomes large, also has an approximately normal distribution with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$.

R.V. $\quad \Sigma x_i = n\bar{x} \qquad \left(\bar{x} = \frac{1}{n}\Sigma x_i\right)$

MEAN & S.E. FOR $\sum x_i$ ARE $n \times$ MEAN & S.E. FOR $\overline{X}$

POPULATION	SAMPLE MEAN $\overline{X}$	$\xrightarrow{\times n}$ SUM OF MEASUREMENTS $\sum x_i$
MEAN μ	MEAN μ	$\xrightarrow{\times n}$ MEAN $n\mu$
S.D σ	S.E. $\frac{\sigma}{\sqrt{n}}$	$\xrightarrow{\times n}$ S.E. $n\frac{\sigma}{\sqrt{n}} = \sqrt{n}\,\sigma$

Suppose there is an apple orchard and let x be the weight a randomly selected apple. Suppose that mean weight of all apples in the orchard is 360g, and the standard deviation is 25g. Find the probability that a sample of 100 apples has a total weight less than 35600g?

$\sum x_i$

FIND $P(\sum x_i \le 35{,}600)$

CLT $\Rightarrow$ $\sum x_i$ IS APPROXIMATELY NORMALLY DISTR.

WITH MEAN $= n\mu = 100 \times 360 = 36000$ g

S.E. $= \sqrt{n}\,\sigma = \sqrt{100} \times 25 = 250$ g

$$z = \frac{\text{R.V.} - \text{MEAN}}{\text{S.D.}}$$

$$z = \frac{\sum x_i - n\mu}{\sqrt{n}\,\sigma}$$

$$P\left(z \le \frac{35{,}600 - 36{,}000}{250}\right) = P(z \le -1.6)$$

$= .0548$

NOTE: IN ORDER TO APPLY THE CENTRAL LIMIT THEOREM ONE OF TWO CONDITIONS MUST BE MET:

(1) POPULATION IS NORMALLY DISTRIBUTED, OR

(2) SAMPLE SIZE IS LARGE ENOUGH, i.e. $n \geq 30$

THE SAMPLING DISTRIBUTION OF THE SAMPLE MEAN, $\bar{x}$

- If a random sample of n measurements is selected from a population with mean μ and standard deviation σ, the sampling distribution of the sample mean $\bar{x}$ will have mean μ and standard deviation*

$$\frac{\sigma}{\sqrt{n}}$$

- If the population has a *normal* distribution, the sampling distribution of $\bar{x}$ will be *exactly* normally distributed, *regardless of the sample size, n.*
- If the population distribution is *nonnormal,* the sampling distribution of $\bar{x}$ will be *approximately* normally distributed for large samples (by the Central Limit Theorem). Conservatively, we require $n \geq 30$.