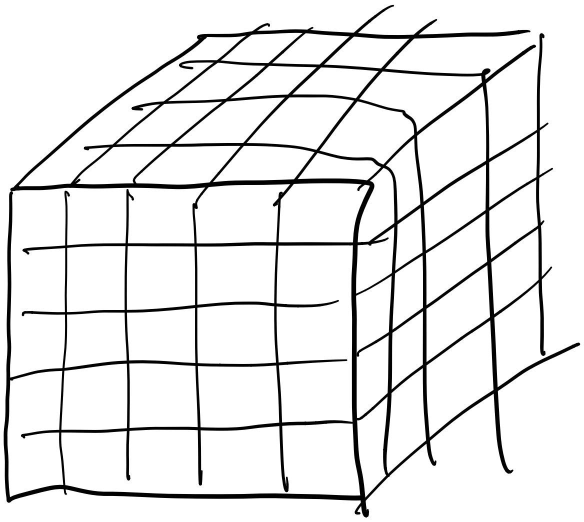
7.29 Bacteria in Water Use the Central Limit Theorem to explain why a ~~Poisson~~ random variable—say, the number of a particular type of bacteria in a cubic foot of water—has a distribution that can be approximated by a normal distribution when the mean μ is large. (HINT: One cubic foot of water contains 1728 cubic inches of water.)



\# BACTERIA IN 1 ft^3 OF H_2O

$$= \underbrace{\sum \text{\# BACTERIA IN } 1 \text{ in}^3}_{1,728 \text{ CUBIC IN}}$$

LET $X =$ # BACT. IN 1 in^3 H_2O.

POPULATION IS A LOT OF in^3 OF H_2O

→ SAMPLE OF 1,728 in^3 OF H_2O

SAMPLE SUM $\sum x_i$

≥ 30

⇒ APPROX. NORMAL DISTRIBUTION.

The number of bacteria in one cubic foot is actually the sum of thumper of bacteria in 1728 cubic inches. When adding the measurements of more than 30 individuals to make a single sample sum, the sampled sum is approximately normally distributed.

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

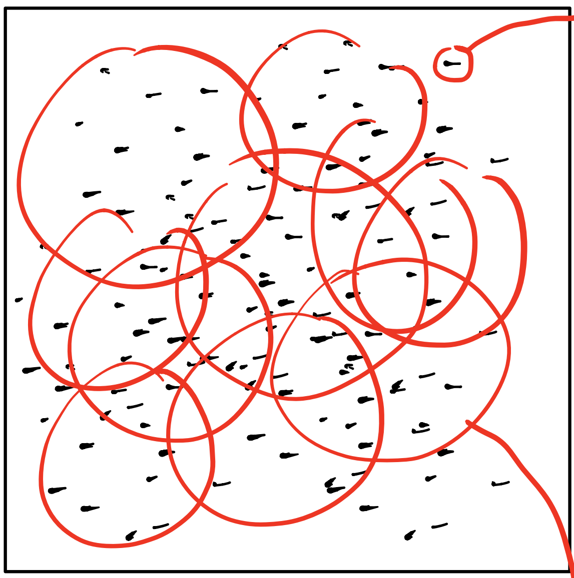
$$\frac{\sigma}{\sqrt{n}}$$

The approximation becomes more accurate as n becomes large.

POPULATION

POP SIZE $N = 1000$

SAMPLE SIZE $n = 40$



SINGLE INDIVIDUAL IN POPULATION CAN BE USED TO OBTAIN A VALUE OF A R.V.

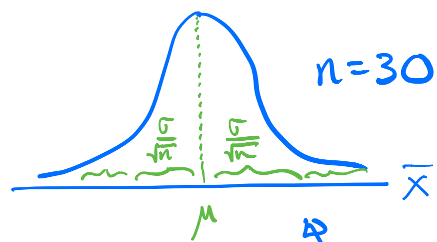
X

CONSIDERING ALL INDIVIDUALS & ALL VALUES OF X, WE CAN CALCULATE MEAN μ & STAND. DEV. σ FOR X.

72 DIGITS

$$C^{1000}_{40} = 5\ldots \left(5.56 \times 10^{72}\right)$$

NEW RANDOM VARIABLE $\bar{X}$ SAMPLE MEAN



TO DETERMINE THE VALUE OF $\bar{X}$,

1. COLLECT A RANDOMLY SELECTED SAMPLE OF n INDIVIDUALS FROM POPULATION.
2. ADD UP THE VALUES X FOR THE INDIVIDUALS & DIVIDE BY n TO PRODUCE $\bar{X}$.

$\bar{X}$ IS APPROX NORMALLY DISTRIBUTED IF EITHER

(1) X IS NORMALLY DISTRIBUTED

(2) $n \geq 30$

NORMAL

MEAN μ

STAND. DEV. $\frac{\sigma}{\sqrt{n}}$

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 $n \geq 30$.

7.23 A random sample of size $n = 40$ is selected from a population with mean $\mu = 100$ and standard deviation $\sigma = 20$.

a. What will be the approximate shape of the sampling distribution of $\bar{x}$?

b. What will be the mean and standard deviation of the sampling distribution of $\bar{x}$?

(a) NORMAL

(b) MEAN FOR $\bar{X}$ = MEAN FOR X = μ = 100

STAND. DEV. FOR $\bar{X}$ = $\dfrac{\text{STAND DEV FOR } X}{\sqrt{n}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{20}{\sqrt{40}}$

7.26 Faculty Salaries Suppose that college faculty with the rank of professor at public 2-year institutions earn an average of \$71,802 per year[7] with a standard deviation of \$4000. In an attempt to verify this salary level, a random sample of 60 professors was selected from a personnel database for all 2-year institutions in the United States.

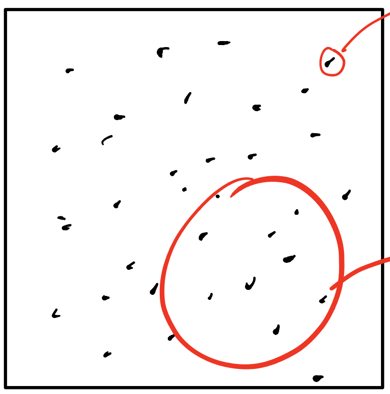
a. Describe the sampling distribution of the sample mean $\bar{x}$.

b. Within what limits would you expect the sample average to lie, with probability .95?

c. Calculate the probability that the sample mean $\bar{x}$ is greater than \$73,000? $\approx 1\%$

d. If your random sample actually produced a sample mean of \$73,000, would you consider this unusual? What conclusion might you draw?

↑ 2.32 STAND DEV. ABOVE MEAN



X = SALARY OF INDIVIDUAL

MEAN FOR X = μ = 71,802

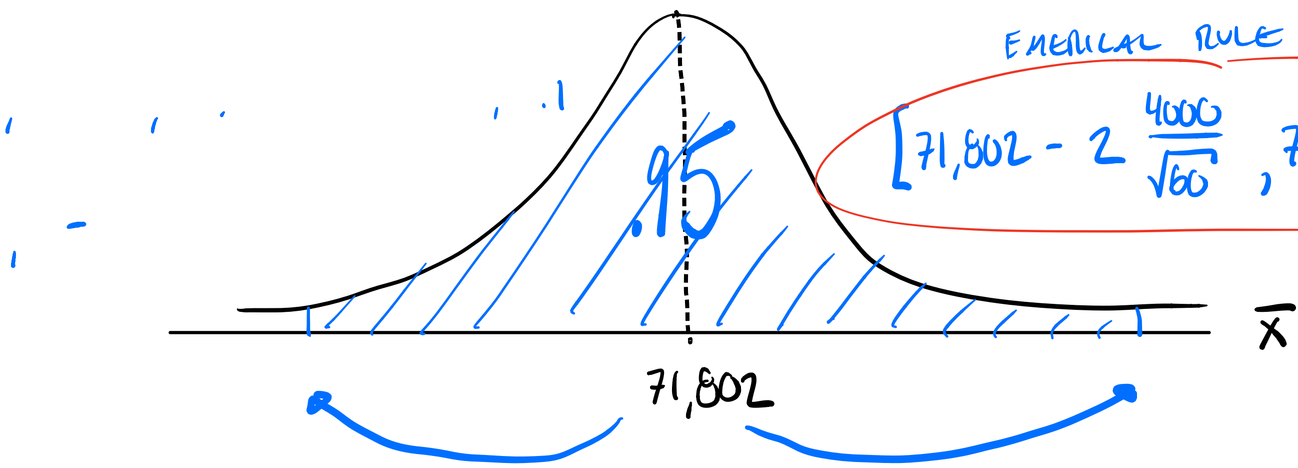
STAND DEV. FOR X = σ = 4,000

$\bar{x}$ SAMPLE MEAN $\left(\bar{x} = \dfrac{\sum_{i=1}^{60} x_i}{60}\right)$

(a) NORMAL (WHY? CENTRAL LIMIT THM, $n \geq 30$)

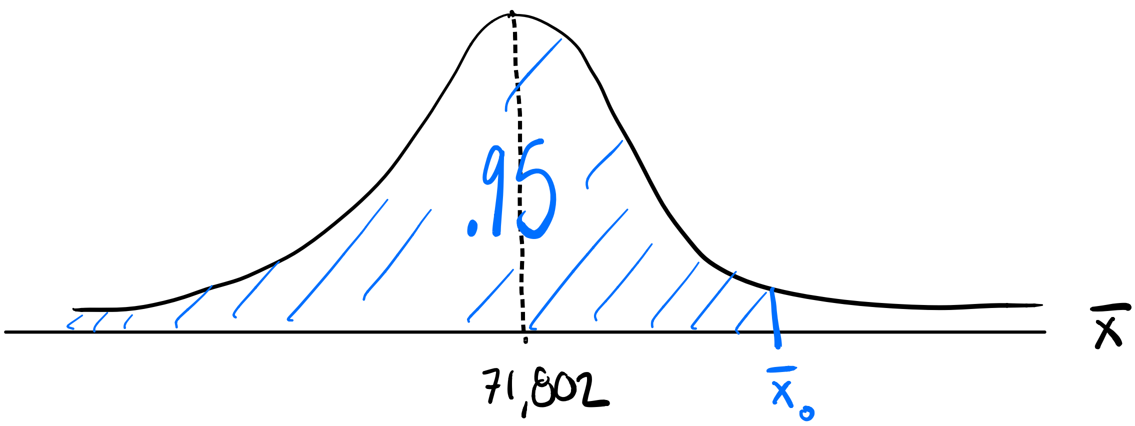
EMPIRICAL RULE

$$\left[71{,}802 - 2\frac{4000}{\sqrt{60}},\ 71{,}802 + 2\frac{4000}{\sqrt{60}}\right]$$



-2 STAND DEV.

+2 STAND.



(c) $P(\bar{x} > 73{,}000)$ $\quad \bar{x} \sim \text{NORMAL}\left(\underset{\text{MEAN}}{71{,}802}, \underset{\text{STAND DEV.}}{\dfrac{4000}{\sqrt{60}}}\right)$

STANDARDIZE (Z-SCORE)

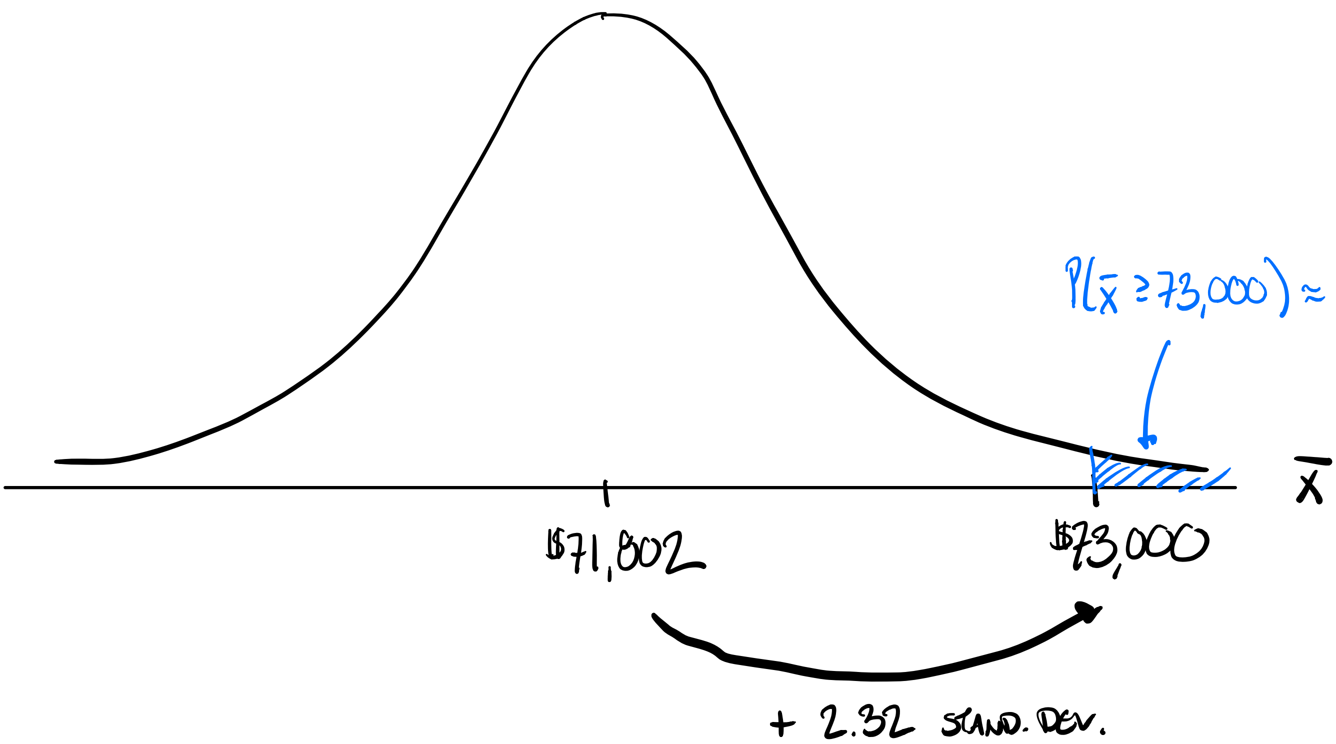
$$z = \frac{\text{R.V.} - \text{MEAN}}{\text{STAND. DEV.}}$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$P\left(z \geq \frac{73000 - 71802}{4000/\sqrt{60}}\right) = P(z \geq 2.32)$$

$$= 1 - P(z \leq 2.32) = 1 - .9898 = .0102$$

	.00	.01	.02							
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936



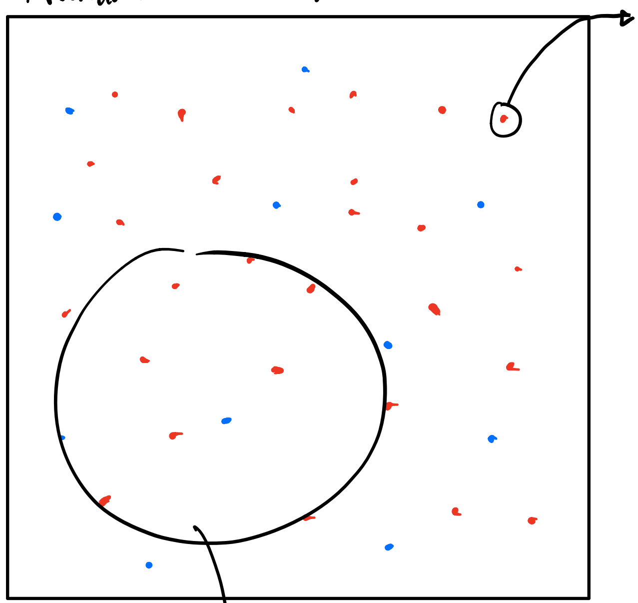
Assuming the population mean is 71802 and population standard deviation is 4000, it is unlikely to obtain a sample mean of 73000 (n=60). Since this happened on the first try, perhaps the assumptions are incorrect. Perhaps the population mean is actually higher than 71802.... or perhaps the standard deviation is higher than 4000.... or perhaps the sample was not random.

EXAM #2 WED 11/18

- §5.2 BINOMIAL PROB. DISTRIBUTION.
- §6.1-6.3 NORMAL DISTRIBUTION
- §6.4 NORMAL APPROX. TO BINOMIAL DISTRIBUTION
- §7.4-7.5 CENTRAL LIMIT THEOREM
  SAMPLE MEANS $\bar{x}$, SAMPLE SUMS Σx
- §7.6 CENTRAL LIMIT THM FOR
  SAMPLE PROPORTIONS $\hat{p}$.

§7.6 Sampling Distribution for Sample Proportion

Population size N



- Success
- Failure

} one or the other

Population proportion $p = P(\text{success})$

$$= \frac{\#\text{ successes}}{\#\text{ successes} + \#\text{ failures}}$$

$$q = 1 - p$$

$P(\text{failure})$

Sample size n

"p-hat"

Sample proportion $\hat{p} = \dfrac{\#\text{ successes}}{n} = \dfrac{X}{n}$

div. by n

$\hat{p}$ is a random variable.

$\hat{p} = \dfrac{X}{n}$ (scale by $\frac{1}{n}$)

Approx normally distributed when

$np > 5$, $nq > 5$,

with mean for $\hat{p} = \dfrac{np}{n} = p$

stand. dev. for $\hat{p} = \dfrac{\sqrt{npq}}{n} = \sqrt{\dfrac{pq}{n}}$

§6.4

Note: # successes X is approximately normal when $np > 5$ & $nq > 5$

div. by n

Mean for X is np

div. by n

Stand. dev. for X is $\sqrt{npq}$

7.45 M&M'S An advertiser claims that the average percentage of brown M&M'S candies in a package of milk chocolate M&M'S is 13%. Suppose you randomly select a package of milk chocolate M&M'S that contains 55 candies and determine the proportion of brown candies in the package.

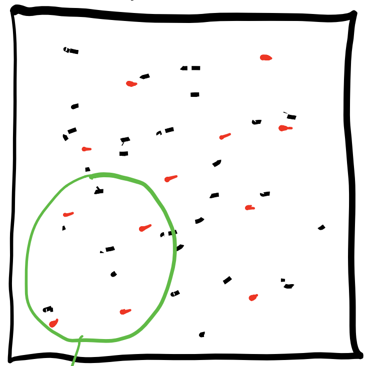
a. What is the approximate distribution of the sample proportion of brown candies in a package that contains 55 candies?

b. What is the probability that the sample percentage of brown candies is less than 20%?

c. What is the probability that the sample percentage exceeds 35%?

d. Within what range would you expect the sample proportion to lie about 95% of the time?

POPULATION OF ALL M&M'S



$p = .13$

PROP. OF POP. THAT IS BROWN.

SAMPLE OF SIZE $n = 55$

R.V. $\hat{p} = \dfrac{\#\text{ BROWN}}{n}$

SAMPLE PROPORTION

(a) NORMAL. WHY? $np = (55)(.13) > 5$

$nq = (55)(.87) > 5$

(b) $P(\hat{p} < .20)$

APPROX NORMAL WITH MEAN. $p = .13$

STAND. DEV. FOR $\hat{p}$ = S.E. $= \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(.13)(.87)}{55}}$

STANDARDIZE: $z = \dfrac{\text{R.V.} - \text{MEAN}}{\text{S.E.}} = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$

$$P\left(z \le \frac{.20 - .13}{\sqrt{\frac{(.13)(.87)}{55}}}\right) = P(z \le 1.54)$$

$$= .9382$$

1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767