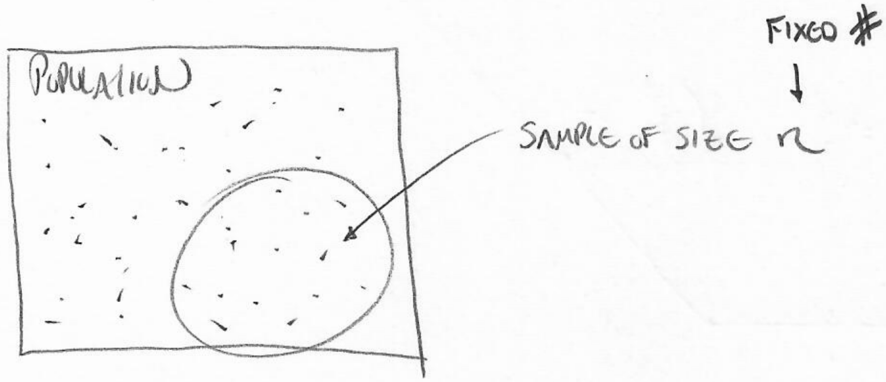
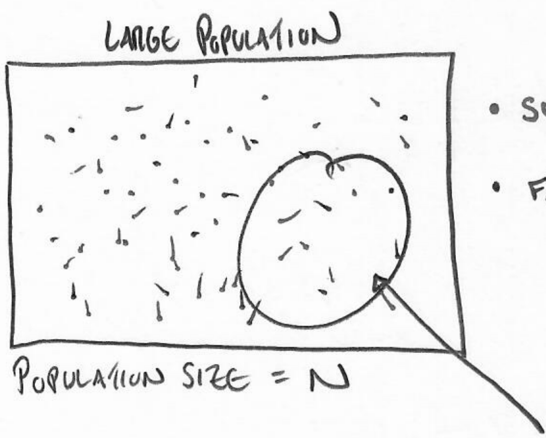


§ 7.6 THE SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION



- SUCCESS
- FAILURE

POPULATION PROPORTION
(PROPORTION OF SUCCESSES IN POPULATION)

$$p = \frac{\text{\# SUCCESSES}}{N}$$

SAMPLE PROPORTION $\hat{p} = \dfrac{\text{\# SUCCESSES IN SAMPLE}}{\text{SAMPLE SIZE}} = \dfrac{x}{n}$

NOW TAKE A RANDOM SAMPLE OF SIZE n &
COUNT THE # OF SUCCESSES IN THE SAMPLE, X.

IF $\dfrac{n}{N} \leq .05$ THEN X IS APPROXIMATELY BINOMIAL.

(OTHERWISE IT'S HYPERGEOMETRIC)

(OO OOO) PICK 2 MARBLES, X = # RED. ⟶ (2 MILL. 3 MILL.) PICK 2 MARBLES X = # RED

x	0	1	2
$p(x)$	$\dfrac{C^2_0 C^3_2}{C^5_2}$	$\dfrac{C^2_1 C^3_1}{C^5_2}$	$\dfrac{C^2_2 C^3_0}{C^5_2}$

x	0	1	2
$p(x)$	$C^2_0 p^0 q^2$	$C^2_1 p^1 q^1$	$C^2_2 p^2 q^0$

BINOMIAL R.V. # TRIALS $= n$

PROB. SUCCESS p

PROB FAILURE $q = 1 - p$

$np \geq 5$

$nq \geq 5$ ⟶ APPROX NORMAL WITH

MEAN $\mu = np$

STAND. DEV. $\sigma = \sqrt{npq}$

SO, GIVEN A LARGE POPULATION WITH PROPORTION OF SUCCESS p,

X = # OF SUCCESSES IN A SAMPLE OF SIZE n $\left(\frac{n}{N} \leq .05 \right)$

IS APPROXIMATELY NORMAL WITH MEAN $\mu = np$ ⟶ DIV. BY n

STAND. DEV. $\sigma = \sqrt{npq}$ ⟶ DIV BY n

$\Rightarrow$ SAMPLE PROPORTION $\hat{p} = \frac{X}{n}$ ← # SUCCESSES IN SAMPLE ← SIZE OF SAMPLE

IS APPROX. NORMALLY DISTRIBUTED WITH MEAN p

AND STANDARD DEV = STANDARD ERROR S.E. $\sqrt{\frac{pq}{n}}$

ex. SUPPOSE A MANUFACTURING PLANT PRODUCES 1,000 DEVICES.

2% OF THE DEVICES PRODUCED ARE DEFECTIVE.

IN AN ORDER OF 100 DEVICES, FIND THE PROBABILITY THAT LESS THAN 3% OF THE DEVICES ARE DEFECTIVE.

POPULATION: ALL DEVICES

$p = .02$ (DEFECTIVE)

$q = .98$ (NOT DEFECTIVE)

SAMPLE:

SIZE: $n = 100$

SAMPLE PROPORTION $\hat{p}$

"p-HAT"

$\hat{p}$ IS APPROXIMATELY NORMALLY DISTRIBUTED WITH

MEAN $p = .02$

STANDARD ERROR $S.E. = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(.02)(.98)}{100}}$

FIND $P(\hat{p} < .03) = P(z < .71) = \boxed{.7611}$

$$z = \frac{\text{RAN. VAR.} - \text{MEAN}}{\text{S.E.}}$$

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} = \frac{.03 - .02}{\sqrt{\dfrac{(.02)(.98)}{100}}}$$

7.45

POPULATION : ALL M&M'S / PROPORTION OF POP. THAT ARE BROWN

POP PROP. $p = .13$

$q = .87$

PACKAGE = RANDOM SAMPLE OF SIZE $n = 55$

X = # BROWN M&M'S IN PACKAGE

(a) $\approx$ APPROX. NORMAL RANDOM VARIABLE

$\hat{p} = \frac{X}{n}$ PROPORTION OF BROWN M&M'S IN THE SAMPLE (SAMPLE PROPORTION)

APPROX. NORMAL DISTRIBUTION WITH

MEAN $p = .13$

S.E. $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(.13)(.87)}{55}} = .0453$

(b) $P(\hat{p} < .2) = P\left(z < \frac{.2 - .13}{\sqrt{\frac{(.13)(.87)}{55}}}\right)$

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

$= P(z < 1.54) = \boxed{.9382}$

(c) $P(\hat{p} > .35) = P\left(z > \dfrac{.35 - .13}{\sqrt{\dfrac{(.13)(.87)}{55}}}\right)$

$= P(z > 4.85) = 1 - P(z < 4.85)$

$\approx 1 - 1 = \boxed{0}$ $\quad .0000000357\ldots$

$$P(z \le z_0) = \begin{cases} \approx 0 & \text{IF } z_0 < -3.49 \\ \text{TABLE} & \text{IF } -3.49 \le z_0 \le 3.49 \\ \approx 1 & \text{IF } z_0 > 3.49 \end{cases}$$

(d) BY EMPERICAL RULE

$\hat{p}$ WILL LIE $\underbrace{\text{WITHIN 2 S.E.'S OF MEAN}}$

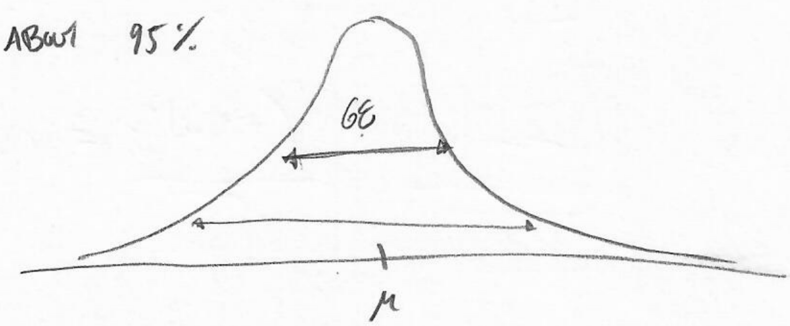
$$p - \overset{1.96}{2}\,\text{S.E} \le \hat{p} \le p + \overset{1.96}{2}\,\text{S.E.}$$

95 % OF THE TIME

$.13 - 2(.0453)$ $\qquad$ $.13 + 2(.0453)$

$.0394$ $\qquad$ $.2206$

ABOUT 95%



§ 8.3 TYPES OF ESTIMATORS

INFERENTIAL STATISTICS: USING KNOWLEDGE ABOUT A SAMPLE TO MAKE (INFER) ESTIMATES ABOUT A POPULATION.

IN REALITY, POPULATION PARAMETERS

- MEAN μ
- PROPORTION p
- STANDARD DEV. σ

USUALLY UNKNOWN.

SO WE USE SAMPLE STATISTICS

- SAMPLE MEAN $\bar{x}$
- SAMPLE PROPORTION $\hat{p}$
- SAMPLE STAND. DEV. s

TO ESTIMATE THE POPULATION PARAMETERS.

POINT ESTIMATES

OR WE USE INTERVALS

- $\bar{x} \pm \varepsilon$
- $\hat{p} \pm \varepsilon$
- $s \pm \varepsilon$

TO ESTIMATE POPULATION PARAMETERS

"EPSILON"
E FOR ERROR.

INTERVAL ESTIMATES

REPORT: PRESIDENTIAL APPROVAL RATINGS 41% (± 2%)

ERROR ε

SAMPLE PROPORTION $\hat{p}$

POINT ESTIMATE

INTERVAL ESTIMATE

39% - 43%

---

§ 8.4 POINT ESTIMATES

Def: AN ESTIMATOR OF A PARAMETER IS UNBIASED IF THE MEAN OF ITS DISTRIBUTION IS THE SAME AS THE VALUE OF THE PARAMETER.

OTHERWISE, IT IS BIASED.

WE WANT TO USE UNBIASED ESTIMATORS.

WE KNOW (CENTRAL LIMIT THEOREM)

SAMPLE MEAN $\overline{X}$ HAS A (NORMAL) DISTRIBUTION WITH THE SAME MEAN AS THE POPULATION.

§ 8.3 TYPES OF ESTIMATORS

POPULATION PARAMETERS (μ, p, σ) USUALLY UNKNOWN, SO WE

USE SAMPLE STATISTICS AS <u>POINT ESTIMATES</u> $(\bar{x}, \hat{p}, s)$ OR

AS <u>INTERVAL ESTIMATES</u> $(\bar{x} \pm \varepsilon, \hat{p} \pm \varepsilon)$.

POP. PARAM. ARE APPROX. EQUAL TO THESE <u>POINTS</u>

CONFIDENT THAT POP. PARAM. LIE IN THESE <u>INTERVALS</u>

§ 8.4 POINT ESTIMATION

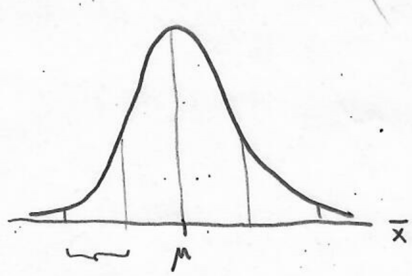
<u>DEF.</u> AN ESTIMATOR OF A PARAMETER IS <u>UNBIASED</u> IF MEAN OF ITS (SAMPLE) DISTRIBUTION IS SAME AS TRUE VALUE OF PARAMETER. OTHERWISE <u>UNBIASED</u>

WE WANT TO USE UNBIASED ESTIMATORS. PREFERABLY WITH SMALL VARIANCE. ← (CLT SAYS SAMPLE MEAN $\bar{X}$ IS DISTRIBUTION WITH SAME MEAN AS POPULATION !)

e.g. $n = 100$

$\bar{x} = 18$

$s = 5$

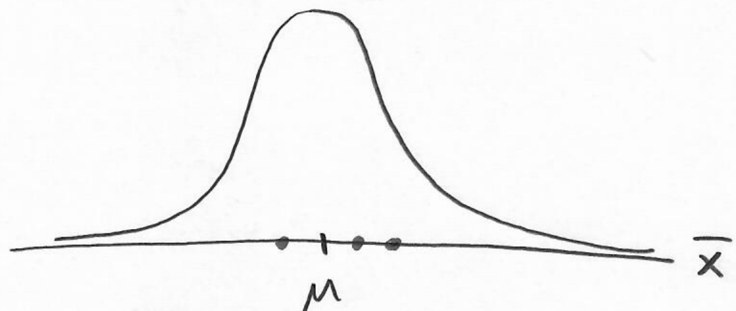


$\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = .5$

ESTIMATE $\mu \approx 18$

e.g. You want to estimate the ave. GPA of all CCNY students, by taking a random sample of 50 students.

$\mu \approx \bar{X}$

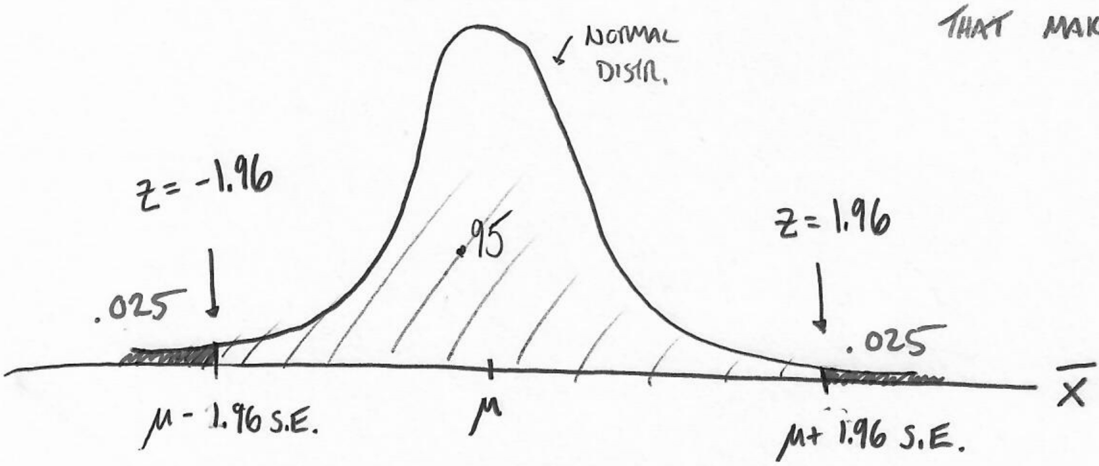


$\bar{X}$ is unbiased because it is equally likely to be overestimate / underestimate and the mean (expected value) of $\bar{X}$ is μ!

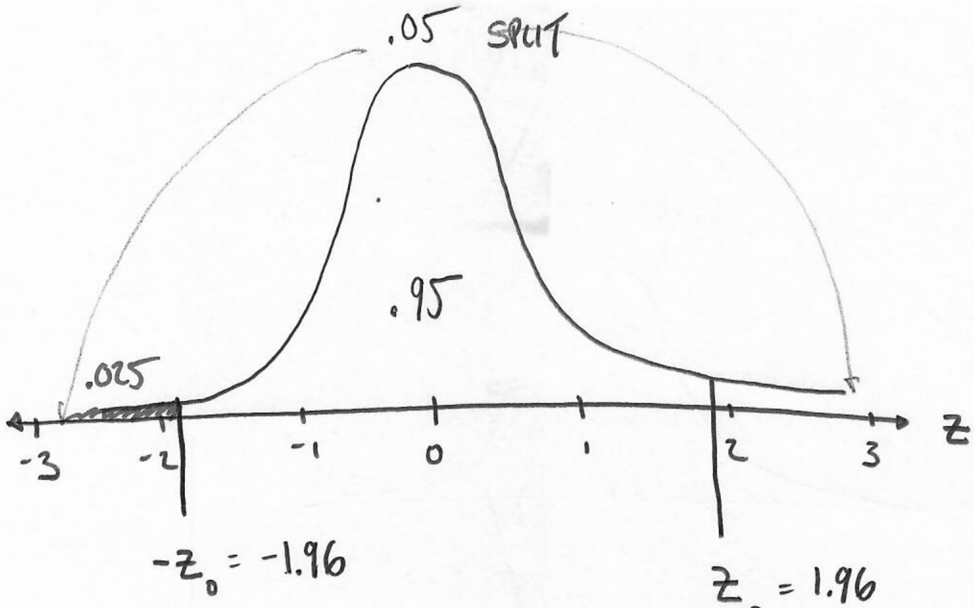
Def: The distance between an estimator & a parameter is called the error of the estimation.

$$\mu \approx \bar{X}, \qquad \mu = \bar{X} + \varepsilon$$

Error - missing piece that makes equality.



CLT: Sample means $\bar{X}$ are approx. normal with mean $= \mu$, and S.E. $= \dfrac{\sigma}{\sqrt{n}}$



$$P(-z_0 \le z \le z_0) = .95$$

$$P(-1.96 \le z \le 1.96) = .95$$

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

$$P(\mu - 1.96\, S.E. \le \bar{x} \le \mu + 1.96\, S.E.) = .95$$

POINT ESTIMATE

POPULATION MEAN $\mu \approx \bar{X}$ (SAMPLE MEAN)

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

INTERVAL ESTIMATE: $\mu \approx \bar{X} \pm \underbrace{1.96\ S.E.}_{\text{MARGIN OF ERROR}}$

$P(\bar{x}$ LIES WITHIN 1.96 S.E. OF $\mu) = .95$

$P(\mu$ LIES WITHIN 1.96 S.E. OF $\bar{x}) = .95$

HW THRU §7.6