6	Lecture - 02 08/31/2020
	telect you work at the uphone tactory the
C.F.	X, ~ Bern (+) = Bern (XN)
ودلي	TUDE QUELLE STEDENTIFICAÇÃO SUL BILLETE
.7=	K January and calcado uct
	Let's draw a second sample from the population
tody	assuming anx, = 1 ugu sus act studed took
	15 D
	$P(X_2=1 \mid X_1=1)$
0 01	Colorles Semple (n=2) show = X-love Xdl=0.
(9mg)	Reputation post Sop stab o N-1 20 Ng
	- (6) mad bil
	=> X, X,=1 ~ Bern (2(-1)
941	1) Giller-M-1 intende consulation sampled
radia d	Population soo Sample (n) on sol parat some
	Th=X,+=+Xn~Hyper(n,X,N)
	Hypergeometric distribution
	(x)(N-x)
moond	P(To=t) = (t) (n-t) parties
- lacadu	Something of the from the delegation
- V	closed Inditamites torage
reclading	Dealing with the hypergeometric is complicated
(6) mad	(but doable). What can be assume to make
	this go away?
	Let X, N -> x but $\theta = X/N$ Simplifying assumption
	3 (make the ratio constant) = x p s
(x, \dots, x)	X, x, X, Red Bern (1)
	lim P(X=1) X1=1) = lim X-1 = 0
	X3 X= 6 in set aN-) acutoculos) p el f
6	
nel verdes	"teur no "melamiles l'obitettate" - ballos
	The statistic (stellatical estimate, estimate

Prelend you work at the iphone factory, they sample new iphones to ensure they work ensure the manufacturing is working properly. check the first one X,=1, X=1, X,=1, X,00=1. What population are you sampling from? What When you estimate A, you're estimating A in a "process", le a "data generating process" (DGP), iid Bern (A). DGPs and infinite population sampling is the same thing. We no longer care about whether the population is "real", we just assume on its DGP from now on-Returning to our main goal interence i.e. knowing something about & from the data. First subgoal. point estimation, Recall, (X,+__, +xn). X,,__, Xn are random realizations from X,,__, Xn & Bern (0) e.q x=[11101] => 0=0.8 I is a realization from the rv colled a "statistical estimator" or just "estimator" The statistic (statistical estimate, estimate) is

a realization from the estimator. The distribution a is called the "Sampling distribution". This sampling distribution and properties, are vergo important mbecause it itells us a lot about our estimates One property is the estimator's expectation, the mean lover all samples of size n. [[0] = [[/n(x,+__.+xn)] = /n \ E[x,] = /n.n E[x,]

In our ind Bern(+) selling. overall $= \theta =$ $\hat{\theta}$ is unbrased" X,,--,,Xn 0 = [Bras[A] = E[A] - Di PI Bras[A] = 0 => à is unbrased & from Andres bases we define a distance function AKA "loss function error function") $(0, \infty)$ 1=0 only if 0=0 loss functions, e.g absolute error bas (L. loss) squard error loss (La loss , p>o Lp loss In (f(x, 0)) f(x; 0) dx Kyllblad - Leibler (K) loss for continuous rus

mudul.	How far away on average are we?
2466	$R(A, \theta) := E[l(A, \theta)]$ Risk of an estimator over X_1, \dots, X_n
eu all y	Risk of an estimator over X,, Xn
	estantes no tudo tol p
adl ,	Of we use squard error loss, R(A, A) = MSE [A] = E[(A-A)^2] "mean squard error (MSE)"
l x	Df the estimator is unbassed, does its MSE simplify? MSE = Voriance MSIE[A] = E[(A-B)2] = E[(A-E[A])2] = Vor'[A) Tif A is unbassed, E[A] = A
haraday	if f is unbassed, E[A]=0
1 7	For a brased estimator (ie the general case),
2004 220	$MSE[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2]$
(A ± (= E[\(\hat{\theta}^2\)] - 2 \(\theta \) E[\(\hat{\theta}\)] + \(\theta^2\) Recall Var[\(\hat{\theta}\)] = E[\(\hat{\theta}^2\)] - E[\(\hat{\theta}\)] ²
	= Vor[A] + E[A] - 20 E[A] + 02.
(200)	1) and many up to do a 10-5 (- (A))
	= Var[ê] + (E[ê] - 0)
	= Var[A] + Bias [A] Bias - variance decomposition
(M) ald so	SE[A] = Twor[A] "standard error of the estimation"