

Στατιστική II

Γιώργος Τσιρογιάννης

Τμήμα Διοίκησης Επιχειρήσεων Αγροτικών
Προϊόντων και Τροφίμων,
Πανεπιστήμιο Πατρών



Διάλεξη 4η

- Πληθυσμοί & δείγματα
- Τυχαία δειγματοληψία & ανάθεση
- Πείραμα και έρευνα
- Κατανομή δειγματοληψίας του μέσου



8^ο - 9^ο κεφάλαιο

Πληθυσμός

- Ολοκληρωμένο σύνολο παρατηρήσεων
- Πραγματικός πληθυσμός: όλες οι πιθανές παρατηρήσεις είναι εφικτές κατά την δειγματοληψία
 - Πχ Ηλικίες κατοίκων του νομού Αχαΐας
 - Βαθμολογία σε εθνικές εξετάσεις
 - Πλήθος παιδιών ανά οικογένεια στην Αττική
 - Βαθμολογία αποφοίτων του ΔΕΑΠΤ

Υποθετικός πληθυσμός

- Δεν είναι παρούσες όλες οι παρατηρήσεις κατά την δειγματοληψία
 - Πχ Θέλουμε να μελετήσουμε το βάρος των *Mus* (genus) κατά την γέννησή τους. Θεωρούμε υποθετικό πληθυσμό όλων των ποντικών και εκείνα που έχουμε στο εργαστήριό ως ένα τυχαίο δείγμα

Δείγμα

- Οποιοδήποτε υποσύνολο (παρατηρήσεων) ενός πληθυσμού
 - Συνήθως πολύ μικρότερο σε μέγεθος
 - Πχ Ηλικίες κατοίκων του νομού Αχαΐας : υποσύνολο 1500 κατοίκων
 - Βαθμολογία σε εθνικές εξετάσεις: υποσύνολο 300 υποψήφίων
 - Το βέλτιστο μέγεθος δείγματος εξαρτάται μεταξύ των άλλων από τα συμπεράσματα που θέλουμε να εξάγουμε και την ακρίβεια τους

Ερωτήσεις

- Είναι σχέσεις δείγματος/πληθυσμού (αριστερά προς δεξιά):
 - Φοιτητές στην τελευταία σειρά, φοιτητές στην τάξη
 - Κάτοικοι της Λάρισας, κάτοικοι Αχαΐας
 - 20 πειραματόζωα σε ένα πείραμα, παρόμοια με εκείνα που χρησιμοποιούνται, τα οποία θα μπορούσαν να υποβληθούν στο ίδιο πείραμα
 - Όλοι οι πρόεδροι των ΗΠΑ, όλοι οι εγγεγραμμένοι Δημοκρατικοί
 - Δυο ρίψεις ενός κέρματος, όλες οι πιθανές ρίψεις ενός κέρματος

Τυχαία δειγματοληψία

- Αν σε κάθε στάδιο της δειγματοληψίας η διαδικασία επιλογής εγγυάται ότι όλες οι δυνατές παρατηρήσεις στον πληθυσμό έχουν ίσες πιθανότητες να συμπεριληφθούν στο δείγμα.
- Η τυχαιότητα αποτυπώνει την επιλογή και όχι κάποιο πρότυπο
- Δεν υπάρχει εξασφάλιση ότι είναι αντιπροσωπευτική του πληθυσμού

Ερωτήσεις

- Κάνουμε τυχαία επιλογή 10 καρτών από 52 μιας τράπουλας
 - Το τ.δ. αναπαριστά με ακρίβεια τα σημαντικά χαρακτηριστικά ολόκληρης της τράπουλας
 - Κάθε κάρτα της τράπουλάς έχει ίσες πιθανότητες να επιλεγεί
 - Είναι αδύνατο τα επιλεχθούν 10 φύλλα κούπες
 - Οποιοδήποτε αποτέλεσμα, όσο απίθανο και αν είναι μπορεί να εμφανισθεί



Πίνακες τυχαίων αριθμών

00000	10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	24805	20636	10402	00822	91665	
00002	08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
00003	99019	02529	09376	70715	38311	31165	88676	74397	04436	27659
00004	12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
00005	69065	74717	34072	76850	36697	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
00007	85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
00010	98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
00012	83452	99634	06288	98033	13746	70078	18475	40610	68711	77817
00013	88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
00015	65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
00017	74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
00020	91499	14523	68479	91479	27686	46162	83554	94750	89923	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
00024	63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
00025	61196	90446	26457	47774	51924	33729	65394	59593	42582	60527
00026	15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
00027	94557	28573	67897	54387	44431	91190	42592	92927	45973	
00028	42481	16213	97344	08721	16868	48767	03071	12059	25701	46670
00029	23523	78317	73208	89837	68935	91416	26252	29663	05522	82562
00030	04493	52494	75246	33824	45862	51025	61962	79335	65337	12472
00031	00549	97654	64051	88159	96119	63896	54692	82391	23287	29529
00032	35963	15307	26898	09354	33351	35462	77974	50024	90130	39333
00033	59808	08391	45427	26842	83609	49700	13021	24892	78565	20106
00034	46058	85236	01390	92286	77281	44077	93910	83647	70617	42941
00035	32179	00597	87379	25241	05567	07007	86743	17157	85394	11838
00036	69234	61406	20117	45204	15956	60000	18743	92423	97188	96338
00037	19565	41430	01758	75379	40419	21585	66674	36806	84962	85207
00038	45155	14938	19476	07246	43667	94543	59047	90033	20826	69541
00039	94864	31994	36168	10851	34888	81553	01540	35456	05014	51176
00040	98086	24826	45240	28404	44999	08896	39094	73407	35441	31880
00041	33185	16232	41941	50949	89435	48581	88695	41944	37548	73043
00042	80951	00406	96382	70774	20151	23387	25016	25298	94624	61171
00043	79752	49140	71961	28296	69861	02591	74852	20539	00387	59579
00044	18633	32537	98145	06571	31010	24674	05455	61427	77938	91936
00045	74029	43902	77557	32270	97790	17119	52527	58021	80814	51748
00046	54178	45611	80993	37143	05335	12969	56127	19255	36040	90324
00047	11664	49883	52079	84827	59381	71539	09973	33440	88461	23356
00048	48324	77928	31249	64710	02295	36870	32307	57546	15020	09994
00049	69074	94138	87637	91976	35584	04401	10518	21615	01848	76938
00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813
00055	08358	69910	78542	42785	13661	58873	04618	97553	31223	08420
00056	28306	03264	81333	10591	40510	07893	32604	60475	94119	01840
00057	53840	86233	81594	13628	51215	90290	28466	68795	77762	20791
00058	91757	53471	61613	62669	50263	90212	55781	76514	83483	47055
00059	89415	92694	00397	58391	12607	17646	48949	72306	94541	37408

Πλήθος ψηφίων

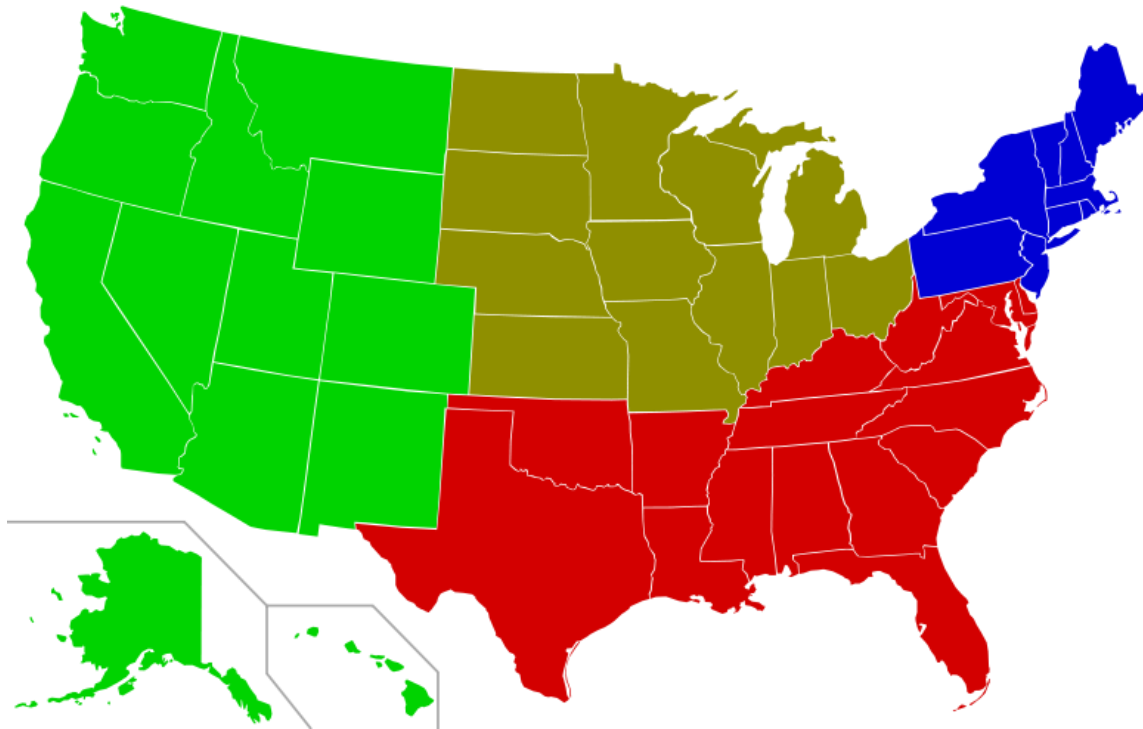
- Επιλέγουμε εκείνο το μικρότερο πλήθος ψηφίων, που υπερκαλύπτει το μέγεθος του πληθυσμού
 - Πχ για πληθυσμό 679 φοιτητών επιλέγουμε 3 ψηφία στο δεκαδικό σύστημα

10097	32533	76520	13586
37542	04805	64894	74296
08422	68953	19645	09303
99019	02529	09376	70715
12807	99970	80157	36147
66065	74717	34072	76850
31060	10805	45571	82406
85269	77602	02051	65692

ΤΔ με βάση τον πίνακα

- Επιλέγουμε μέγεθος δείγματος
- Απαριθμούμε τα σημεία (κατάλογος πληθυσμού)
- Επιλέγουμε τυχαίο σημείο έναρξης στον πίνακα
- Σταθερή κατεύθυνση κίνησης
- Μόλις συναντήσουμε αριθμό εντός περιοχής κρατάμε αντίστοιχο σημείο
- Επαναλαμβάνουμε έως ότου συμπληρώσουμε το ζητούμενο πλήθος για το δείγμα:
 - Ανάλογα με την τακτική δειγματοληψίας απορρίπτουμε επαναλήψεις

Απουσία καταλόγου πληθυσμού



Τυχαία επιλογή εντός της κάθε περιοχής



Μικρά γεωγραφικά τμήματα αντιστοιχίζονται σε μια θέση σταθμοσκόπησης



Καταρτίζονται κατάλογοι για τις «μικρές» που επελέγησαν

Τυχαία ανάθεση αντικειμένων

- Αν οι συνθήκες του πειράματος δεν μας επιτρέπουν τυχαία δειγματοληψία, μπορούμε να τα αναθέσουμε τυχαία
 - Όλες οι διαμορφώσεις είναι ισοπίθανες
 - Οι πιθανές ομάδες έχουν τυχαίες διαφορές, αλλά παρόμοιες ως προς τις ελεγχόμενες μεταβλητές
- Πιθανοί τρόποι:
 - Στρίψιμο νομίσματος όταν «έρχεται» ένα αντικείμενο προς ταξινόμηση
 - Ανισορροπία του πλήθους
 - Στρίψιμο νομίσματα ανά δύο αντικείμενο: το πρώτο στο αποτέλεσμα της ριψης του νομίσματος, το επόμενο στην άλλη ομάδα
 - Ισορροπία του πλήθους
 - Ιδέες για 3 ομάδες;

Έρευνα και πείραμα


- Έρευνα: μελέτη και στατιστική συμπερασματολογία από δείγμα που έχει ληφθεί από ένα πραγματικό πληθυσμό
- Πείραμα: δεδομένα/δείγμα που έχουν ενταχθεί σε ομάδες
- Και στα δύο είναι σημαντικό να εξασφαλίζουμε ισοπίθανη δειγματοληψία & ανάθεση

Πληθυσμός

- Ολοκληρωμένο σύνολο παρατηρήσεων
- Πραγματικός πληθυσμός: όλες οι πιθανές παρατηρήσεις είναι εφικτές κατά την δειγματοληψία
 - Πχ Ηλικίες κάτοικων του νομού Αχαΐας
 - Βαθμολογία σε εθνικές εξετάσεις
 - Πλήθος παιδιών ανά οικογένεια στην Αττική
 - Βαθμολογία αποφοίτων του ΔΕΑΠΤ

Υποθετικός πληθυσμός

- Δεν είναι παρούσες όλες οι παρατηρήσεις κατά την δειγματοληψία
 - Πχ Θέλουμε να μελετήσουμε το βάρος των Mus (genus) κατά την γέννησή τους. Θεωρούμε υποθετικό πληθυσμό όλων των ποντικών και εκείνα που έχουμε στο εργαστήριό ως ένα τυχαίο δείγμα



Κατανομή δειγματοληψίας του μέσου



Παράδειγμα



- Έστω ότι οι φοιτητές που λαμβάνουν μέρος στις εθνικές εξετάσεις έχουν μέσο 500 και τυπική απόκλιση 110 σε εθνικό επίπεδο.
- Τι συμβαίνει σε τοπικό επίπεδο;
- Σε τυχαίο δείγμα (τ.δ.) 100 τοπικών φοιτητών έχουν μέσο 533.

Τι είναι η κατανομή δειγματοληψίας του μέσου

- Είναι η κατανομή που αναπαριστά την μεταβλητότητα των δειγματικών μέσων
- Πλαίσιο αναφοράς για την γενίκευση των συμπερασμάτων για τον πληθυσμό
- **Αναφέρεται στην κατανομή των πιθανοτήτων των μέσων για όλα τα πιθανά τυχαία δείγματα δεδομένου μεγέθους από κάποιον πληθυσμό**

Χρησιμότητα

- Μας επιτρέπει να διαπιστώσουμε αν δεδομένης της μεταβλητότητας μεταξύ όλων των πιθανών δειγματικών μέσων, ο παρατηρούμενος δειγματικός μέσος μπορεί να θεωρηθεί κοινό ή σπάνιο αποτέλεσμα.
- Δηλαδή πως συγκρίνεται ο υπολογισμένος μέσος, με τους μέσους αν κάναμε δειγματοληψία του προς εξέταση πληθυσμού

Κατανομή δειγματοληψίας του μέσου

- Βασίζεται σε όλα τα πιθανά δείγματα συγκεκριμένου μεγέθους
- Πχ για 1000 φοιτητές, ένα δείγμα μεγέθους 100 ~ 10^{300}
- Δεν χρειάζεται να τα υπολογίσουμε όλα

Υπολογισμός από την αρχή

- Πολύ μικρός πληθυσμός $\{2,3,4,5\}$
- Ζητάμε κατανομή μέσου για δείγματα μεγέθους 2

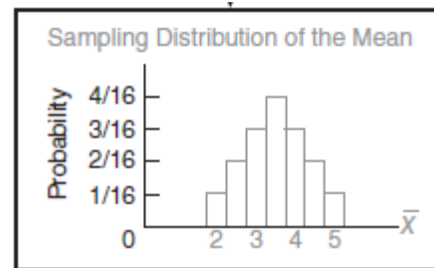
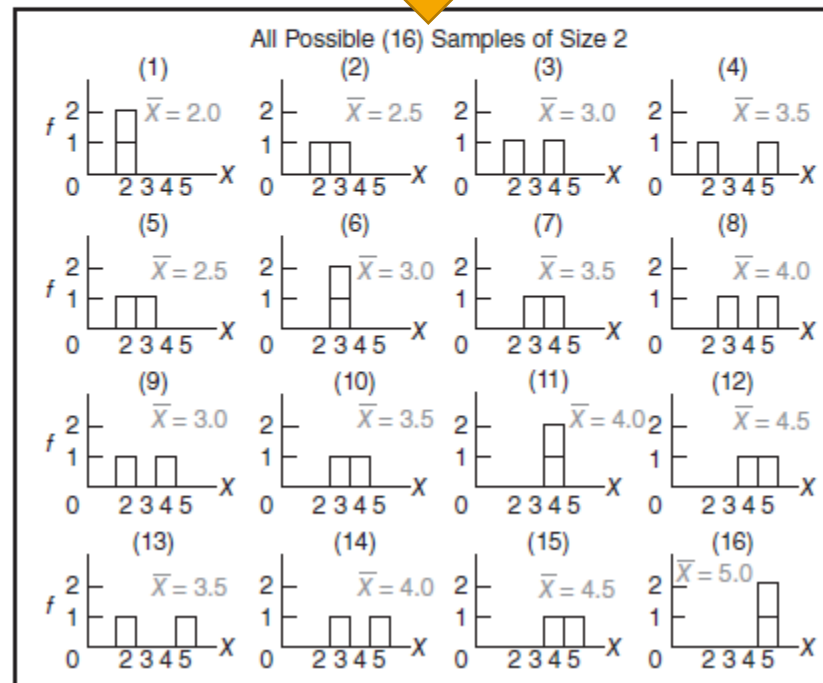
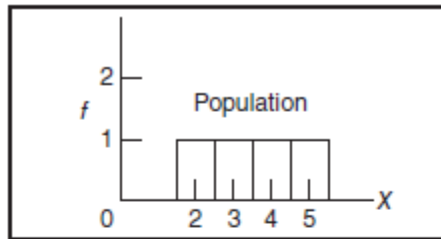
Υπολογισμός από την αρχή

Πίνακας όλων των ζευγών

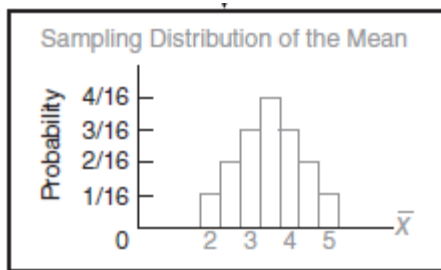
	ALL POSSIBLE SAMPLES	MEAN (\bar{X})	PROBABILITY
(1)	2,2	2.0	$\frac{1}{16}$
(2)	2,3	2.5	$\frac{1}{16}$
(3)	2,4	3.0	$\frac{1}{16}$
(4)	2,5	3.5	$\frac{1}{16}$
(5)	3,2	2.5	$\frac{1}{16}$
(6)	3,3	3.0	$\frac{1}{16}$
(7)	3,4	3.5	$\frac{1}{16}$
(8)	3,5	4.0	$\frac{1}{16}$
(9)	4,2	3.0	$\frac{1}{16}$
(10)	4,3	3.5	$\frac{1}{16}$
(11)	4,4	4.0	$\frac{1}{16}$
(12)	4,5	4.5	$\frac{1}{16}$
(13)	5,2	3.5	$\frac{1}{16}$
(14)	5,3	4.0	$\frac{1}{16}$
(15)	5,4	4.5	$\frac{1}{16}$
(16)	5,5	5.0	$\frac{1}{16}$

Όλοι οι συνδυασμοί δειγμάτων
Από αυτός 1 μόνο επιλέγεται

Υπολογισμός από την αρχή



Υπολογισμός από την αρχή



Μπορούμε να υπολογίσουμε την πιθανότητα κάθε δειγματικού μέσου

Μπορούμε να εξάγουμε συμπεράσματα

Τι είναι η κατανομή δειγματοληψίας του μέσου

- Είναι η κατανομή που αναπαριστά την μεταβλητότητα των δειγματικών μέσων
- Πλαίσιο αναφοράς για την γενίκευση των συμπερασμάτων για τον πληθυσμό
- Αναφέρεται στην κατανομή των πιθανοτήτων των μέσων για όλα τα πιθανά τυχαία δείγματα δεδομένου μεγέθους από κάποιον πληθυσμό

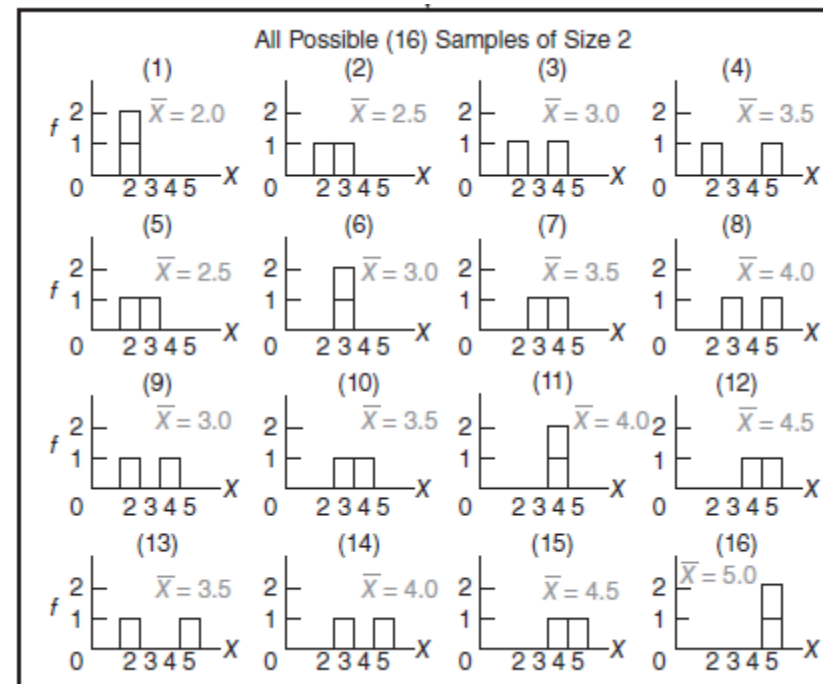
Συμβολισμοί

	TYPE OF DISTRIBUTION	MEAN	STANDARD DEVIATION	
Δείγμα	Sample	\bar{X}	s	
Πληθυσμός	Population	μ	σ	
Κατανομή δειγματοληψίας του μέσου	Sampling distribution of the mean	$\mu_{\bar{X}}$	$\sigma_{\bar{X}}$ (standard error of the mean)	(τυπικό σφάλμα του μέσου)

Ιδιότητες

$\mu_{\bar{X}} = \mu$ Ο μέσος της κατανομής δειγματοληψίας του μέσου, ισούται με το μέσο το πληθυσμού

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ Το τυπικό σφάλμα του μέσου ισούται με την τυπική απόκλιση του πληθυσμού δια της τετραγωνικής ρίζας του μεγέθους του δείγματος



Online (σημειώσεις)

- [https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_\(Shafer_and_Zhang\)/06%3A_Sampling_Distributions/6.02%3A_The_Sampling_Distribution_of_the_Sample_Mean](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/06%3A_Sampling_Distributions/6.02%3A_The_Sampling_Distribution_of_the_Sample_Mean)
- <https://online.stat.psu.edu/stat500/lesson/4/4.1>

Video

- <https://www.youtube.com/watch?v=q50GpTdFYyI>

The Sampling Distribution of \bar{X}