Бизнис статистика

Гранични теореми

Биномна распределба

- Нека A е даден настан со веројатност $p = P(A)$.
- Случајната променлива X – број на појавувања на настанот A во серија од n независни и еднакви експерименти, има биномна распределба со параметри n и p, т.е. $X \sim B(n,p)$.
- Да го разгледаме следниот пример. Нека $X \sim B(1000,\ 0.02)$. Тогаш множеството вредности на X е $R_X = \{0,1,\ldots,1000\}$, а соодветните веројатности се

$$P\{X = i\} = \binom{1000}{i} 0.02^i \cdot 0.98^{1000-i}, \qquad i = 0,1,\ldots,1000$$

- Ако треба да се пресмета, на пример, веројатноста $P\{250 \le X \le 520\}$, тогаш таа веројатност ќе се добие со следната сума

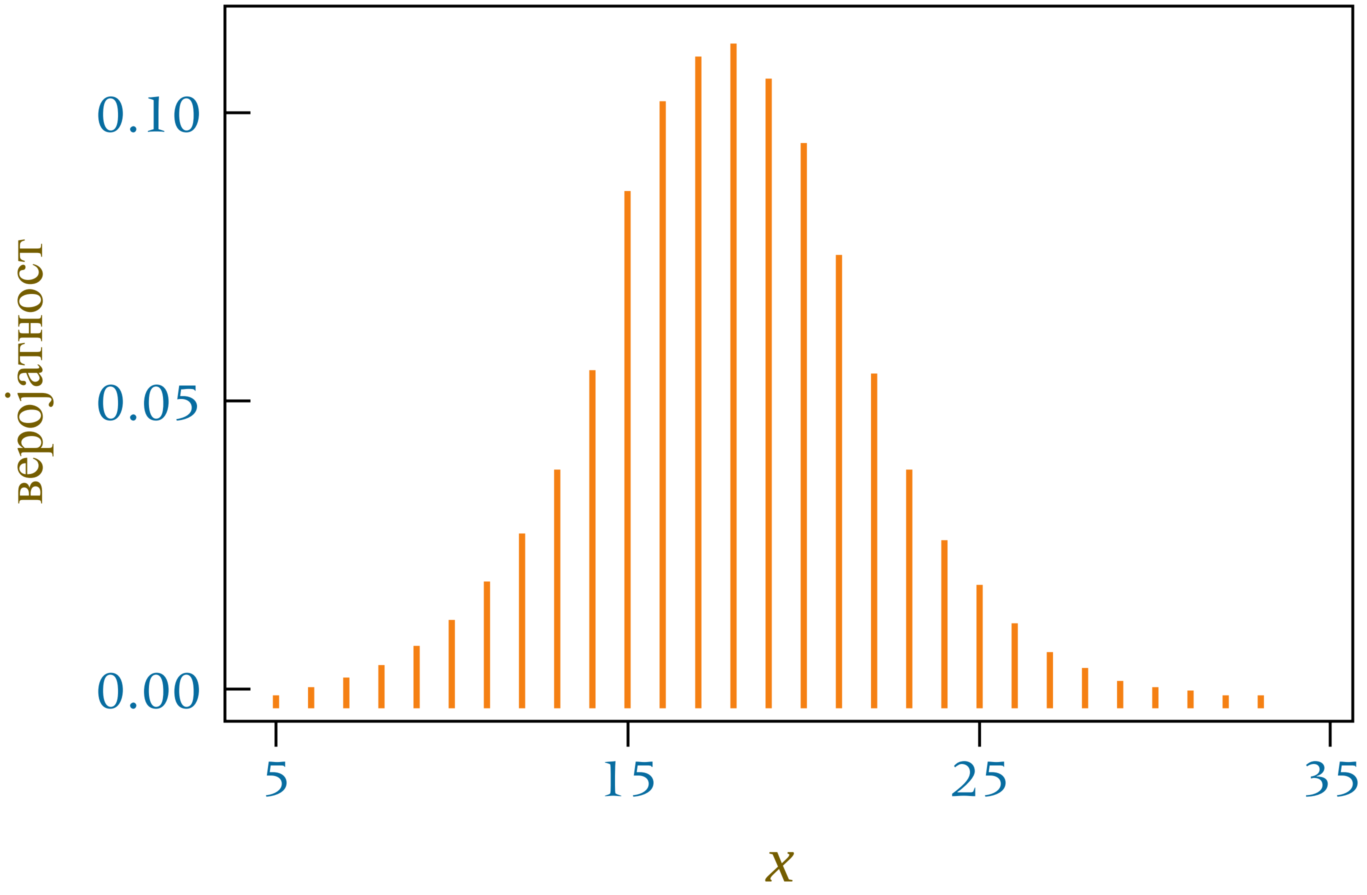
$$P\{250 \le X \le 520\} = \sum_{i=250}^{520} \binom{1000}{i} 0.02^i \cdot 0.98^{1000-i}.$$

Апроксимација на биномна распределба со нормална

$$P\{250 \le X \le 520\} = \sum_{i=250}^{520} \binom{1000}{i} 0.02^i \cdot 0.98^{1000-i}.$$

- Оваа веројатност тешко ќе може да се пресмета рачно или со калкулатор, а ако се користи некоја компјутерска програма при пресметките ќе се направат одредени грешки при заокружување.
- Се поставува прашање, ако веќе се прават грешки при заокружување дали е можно да се најдат некои формули за приближно определување на ваквите веројатности во кои бројот на пресметки ќе биде значително помал.
- Се покажува дека за голем број n (број на експерименти во серијата), обликот на биномна распределба е приближно ист со обликот на нормална распределба.
- Нормалната добро ја апроксимира биномната кога $np > 10$ и $n > 30$.

Биномна со n = 60, p = 0.3



Апроксимација на биномна распределба со нормална

- За да се примени апроксимација на биномна распределба со нормална, потребно е двата параметри n и p на биномната распределба да се конвертираат во двата параметри μ и σ^2 на нормалната распределба.
- Согласно формулите за математички очекувања и дисперзии на познати распределби, добиваме:

$$\mu = np, \qquad \sigma^2 = npq.$$

- Сега, за големи вредности на n, ако $X \sim B(n, p)$, а $Y \sim N(np, npq)$, тогаш обликот на овие две распределби ќе биде многу сличен.

$$P\{a < X < b\} \approx P\{a < Y < b\}$$

Апроксимација на биномна распределба со нормална

$$P\{a < X < b\} \approx P\{a < Y < b\}$$

- Со нормирање на Y се добива дека

$$Z = \frac{Y - np}{\sqrt{npq}} \sim N(0,1)$$

- Врз основа на претходното, ако $np > 10$ и $n > 30$, добиваме

$$\begin{aligned} P\{a < X < b\} &\approx P\{a < Y < b\} \\ &= P\left\{\frac{a-np}{\sqrt{np(1-p)}} < \frac{Y-np}{\sqrt{npq}} < \frac{b-np}{\sqrt{np(1-p)}}\right\} \\ &= P\left\{\frac{a-np}{\sqrt{np(1-p)}} < Z < \frac{b-np}{\sqrt{np(1-p)}}\right\} \\ &= \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right) \qquad (1) \end{aligned}$$

каде што вредностите $\Phi(x)$ се читаат од таблица за нормална нормирана распределба.

Претходниот пример

- Да се вратиме сега на претходниот пример каде $X \sim B(1000,\ 0.02)$. Се бара веројатноста $P\{250 \le X \le 520\}$. Утврдивме дека

$$P\{250 \le X \le 520\} = \sum_{i=250}^{520} \binom{1000}{i} 0.02^{i} \cdot 0.98^{1000-i}. \qquad (2)$$

- Во овој случај, $n = 1000 > 30$, а $np = 20 > 10$, па може да се примени формулата (1) за апроксимација на биномна распределба со нормална.

$$P\{250 \le X \le 520\} \approx P\left\{\frac{250-20}{\sqrt{1000 \cdot 0.02 \cdot (1-0.02)}} \le Z \le \frac{520-20}{\sqrt{1000 \cdot 0.02 \cdot (1-0.02)}}\right\}$$

$$= P\{51.95 \le Z \le 112.94\} = P\{51.95 < Z < 112.94\}$$

$$= \Phi(112.94) - \Phi(51.95) = 1 - 1$$

$$= 0$$

- Можеме да заклучиме дека иако во точната формула (2) има 271 собирок, тие се многу мали, па практично целата сума е многу блиску до 0.

Пример 1

40% од гласачите го поддржуваат кандидатот *A*. Колкава е веројатноста дека во примерок од $n = 200$ гласачи, од 76 до 80 ќе гласаат за *A*?

Решение: Нека случајната променлива *X* го определува бројот на гласачи, кои го подржуваат кандидатот *A*.

- Тогаш $X \sim B(200,\ 0.4)$. Притоа, $n = 200 > 30$, а $np = 80 > 10$, па може да се примени формулата (1) за апроксимација на биномна распределба со нормална.

$$P\{76 \le X \le 80\} \approx P\left\{\frac{76-80}{\sqrt{200(0.4)(1-0.4)}} \le Z \le \frac{80-80}{\sqrt{200(0.4)(1-0.4)}}\right\}$$

$$= P\{-0.58 \le Z \le 0\} = P\{-0.58 < Z < 0\}$$

$$= \Phi(0) - \Phi(-0.58) = 0.5000 - 0.2810$$

$$= 0.2190$$

Апроксимација со нормална распределба

- Со формулата (1) е дадена апроксимација на биномна распределба со нормална.

- Следната теорема покажува дека распределбата на просекот (аритметичката средина) на n независни и еднакво распределени случајни променливи може да се апроксимира со нормална распределба.

- Оваа теорема се нарекува ***централна гранична теорема*** и покажува зошто нормалната распределба е најраспространета распределба за опишување на реалните величини.

Математичко очекување и дисперзија на просек

- Пред да ја формулираме централната гранична теорема ќе изведеме математичко очекување и дисперзија на просек на независни и еднакво распределени случајни променливи.
- Нека X_1, X_2, …., X_n се независни и еднакво распределени случајни променливи со математичко очекување μ и стандардна девијација σ и нека

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

е просекот на тие случајни променливи.

- Просекот е случајна променлива, како функција од случајните променливи X_1, X_2, …., X_n.

Математичко очекување и дисперзија на просек

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

- Со користење на својствата на математичко очекување и дисперзија, за математичкото очекување и дисперзијата на просекот се добива:

$$E\bar{X} = E\left(\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right) = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n) = \frac{1}{n}(EX_1 + EX_2 + \ldots + EX_n) = \frac{1}{n}\cdot n\mu = \mu$$

$$D\bar{X} = D\left(\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right) = \frac{1}{n^2}D(X_1 + X_2 + \ldots + X_n) = \frac{1}{n^2}(DX_1 + DX_2 + \ldots + DX_n) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

- Стандардната девијација на просекот е $\sigma/\sqrt{n}$.

Централна гранична теорема

Централна гранична теорема. Нека $X_1, X_2, \ldots, X_n$ се независни и еднакво распределени случајни променливи со математичко очекување μ и стандардна девијација σ. Ако n доволно големо тогаш просекот $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ има приближно нормална распределба со математичко очекување μ и стандардна девијација ${}^{\sigma}/_{\sqrt{n}}$ (без оглед на распределбата на X_i).

- Значи, $\bar{X} \approx N(\mu, {}^{\sigma^2}/_{n})$. Оттука, со нормирање се добива дека

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma}\sqrt{n} \approx N(0,1)$$

Централна гранична теорема

- Оттука,

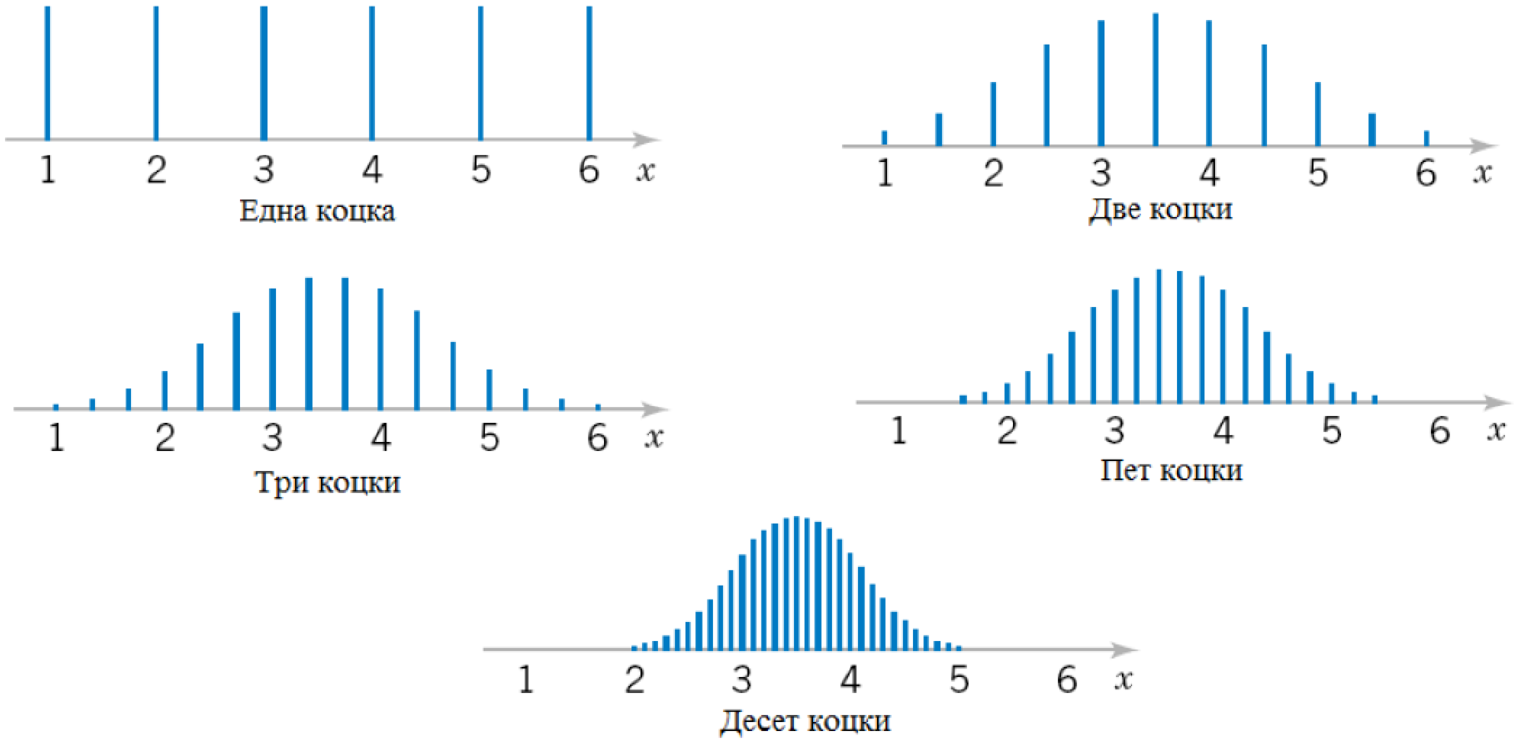
$$P\{a < \bar{X} < b\} = P\left\{\frac{a-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{b-\mu}{\sigma/\sqrt{n}}\right\}$$

$$= P\left\{\frac{a-\mu}{\sigma/\sqrt{n}} < Z < \frac{b-\mu}{\sigma/\sqrt{n}}\right\}$$

$$= \Phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a-\mu}{\sigma/\sqrt{n}}\right) \qquad (2)$$

Илустрација на централна гранична теорема

- Централната гранична теорема ќе ја илустрираме на следниот пример.
- Нека $\bar{X}$ е просек од вредностите добиени при фрлање на n коцки.
- За n = 1, 2, 3, 5, 10, се добиваат следните распределби.



Илустрација на централна гранична теорема

- Од претходната слика може да се воочи дека обликот на распределбата многу се доближува до обликот на нормална распределна, дури и за $n = 3$, $n = 5$.

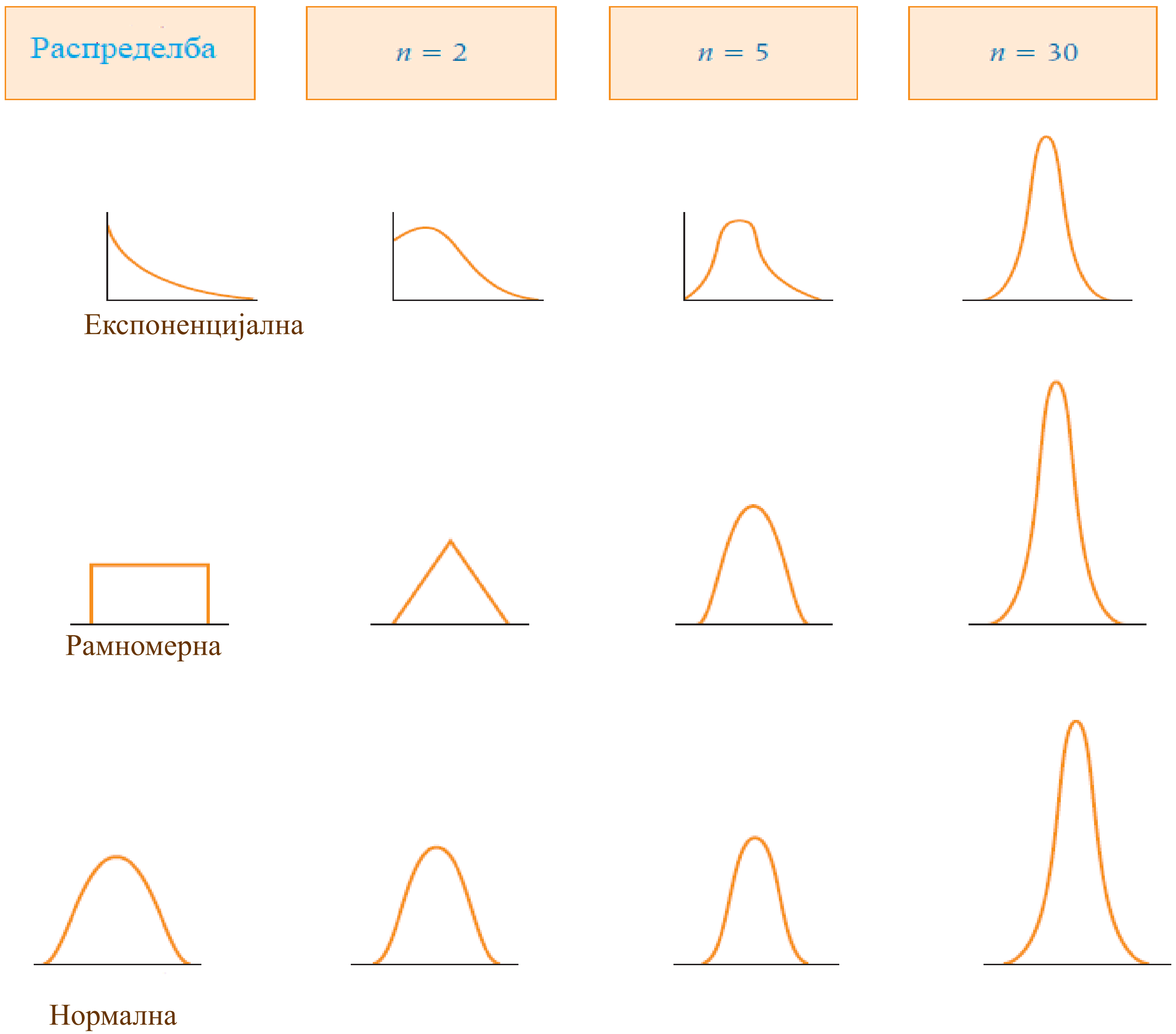
- Но, она што мора да се воочи е дека распределбата на просеците е дискретна, а нормалната распределба е од апсолутно-непрекинат тип.

- Оттука, за приближувањето да биде подобро, треба n да е поголемо.

- Од многу практични примери, за различни распределби е заклучено дека ако $n > 30$, може да се смета дека приближувањето со нормална распределба е доволно добро.

Илустрација на централна гранична теорема

- Нека $\bar{X}$ е просек на n случајни променливи кои се независни и еднакво распределени.
- На цртежот се дадени различни распределба за $X_1, \ldots, X_n$, за $n = 2$, 5 и 30.
- Може да се воочи дека за $n = 30$, распределбата на просекот тежи кон нормална нормирана распределба, без оглед која е распределбата на случајните променливи $X_1, \ldots, X_n$.



Пример 2

Нека X_1, X_2, …., X_{36} се независни и еднакво распределени случајни променливи со математичко очекување μ = 8 и стандардна девијација σ = 3. Да се определи веројатноста дека просекот на овие случајни променливи е помеѓу 7.8 и 8.2.

Решение: Бидејќи n > 30, може да се примени централна гранична теорема. Според оваа теорема $\bar{X}$ има приближно нормална распределба со математичко очекување μ = 8 и стандардна девијација $\sigma/\sqrt{n} = 3/\sqrt{36} = 0.5$.

- Сега,

$$P\{7.8 < \bar{X} < 8.2\} = P\left\{\frac{7.8-8}{3/\sqrt{36}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}}\right\}$$

$$= P\{-0.4 < Z < 0.4\} = \Phi(0.4) - \Phi(-0.4)$$

$$= 0.65542 - 0.34458 = 0.31084$$

Бизнис статистика

Распределба на статистики на примерок

Да се потсетиме: популација и обележје

- Во статистиката, множеството еднородни објекти или резултати од некоја операција, кои имаат една или повеќе заеднички карактеристики се нарекува ***популација***. Заедничката карактеристика се нарекува ***обележје***.

- На пример, за популацијата "деца во Скопје", обележја се: висината, тежината, успехот, полот, националноста, бојата на очите и сл., а за популацијата "компјутери на ФИНКИ", обележја се: брзина на процесорот, големина на RAM меморијата, капацитет на дискот, итн.

- Може да се воочи дека вредноста што ја добива обележјето е променлива величина и таа не може со сигурност да се предвиди за конкретна единка од популацијата. Затоа, ќе сметаме дека секое обележје е случајна променлива.

Примерок

- Како што е познато од претходно, при вршење на статистички испитувања, скоро никогаш не е возможно е да се испитаат сите елементи од популацијата. Се случува испитувањата да траат долго, да бидат поврзани со трошење на големи финансиски средства или уште повеќе, при испитување на одредени објекти да дојде до нивно уништување.
- Затоа, се избира дел од популацијата на кој се вршат сите испитувања и притоа треба да се донесат заклучоци кои ќе важат за целата популација со одредена веројатност.
- Делот од популацијата на кој се вршат потребните испитувања се вика ***примерок***.
- Бројот на елементи во примерокот се вика ***обем на примерокот***.

Случаен примерок

- Нека за дадена популација се набљудува обележјето X и нека X_1 е вредноста на X набљудувана при првото изведување на експериментот, X_2 при второто, итн., X_n вредноста набљудувана во n-тото изведување на експериментот.
- Да воочиме дека вака добиените случајни променливи X_1, X_2, …. , X_n имаат иста распределба како и обележјето X, т.е. се еднакво распределени и се независни.

Векторот $(X_1, X_2, \ldots, X_n)$ е ***случаен примерок*** за обележјето X, ако X_1, X_2, …. , X_n се независни случајни променливи и сите X_i имаат иста распределба како и обележјето X.

Случаен примерок. Статистика

- Кога случајните променливи $X_1, X_2, \ldots, X_n$ ќе добијат конкретни вредности од популацијата, тогаш се добива една *реализација на случајниот примерок*.
- На пример, нека $(X_1, X_2, X_3, X_4, X_5)$ е случаен примерок со обем 5 за обележјето X – висина на студентите на ФИНКИ. Ако случајно се изберат 5 такви студенти и се измери нивната висина, тогаш ќе се добие една реализација на примерокот. На пример, (192, 184, 170, 185, 190) е една реализација на примерокот, (187, 190, 167, 178, 194) е друга конкретна реализација на примерокот врз основа на 5 други случајни избори и сл.

- Статистика на примерок е секоја функција од примерокот.
- Статистиката е случајна променлива бидејќи е функција од случајните променливи X_1, X_2, …, X_n.

Распределби на статистики

- Во продолжение ќе определиме распределби на некои статистики од примерок.

Распределби на статистики

- Распределба на просекот на примерок
- Распределба на пропорции во примерок
- Распределба на дисперзија на примерок

Распределби на просек на примерок

- Распределби на статистики
  - Распределба на просекот на примерок
  - Распределба на пропорции во примерок
  - Распределба на дисперзија на примерок

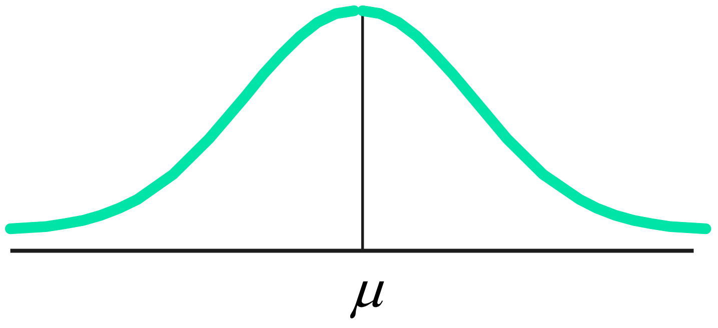
Распределба на просекот, ако обележјето X има нормална распределба

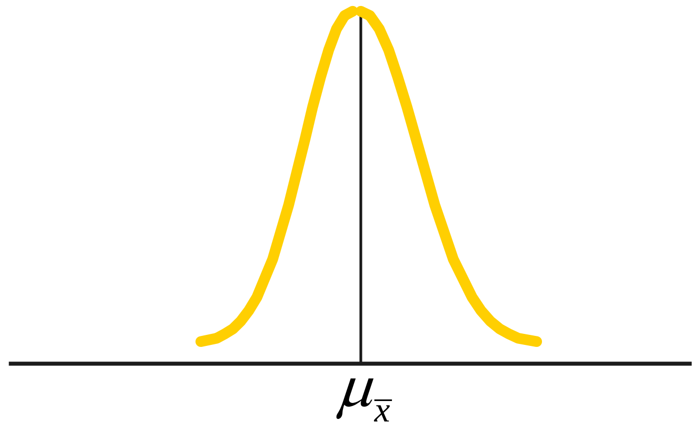
- Ако обележјето X има нормална распределба со математичко очекување μ и стандардна девијација σ, распределбата на просекот $\bar{X}$ на примерокот е исто така нормална распределба со математичко очекување $\mu_{\bar{X}} = \mu$ и стандардна девијација $\sigma_{\bar{X}} = {}^{\sigma}/_{\sqrt{n}}$.

- Оттука, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ има нормална нормирана распределба.

Распределба на просекот, ако обележјето X има нормална распределба

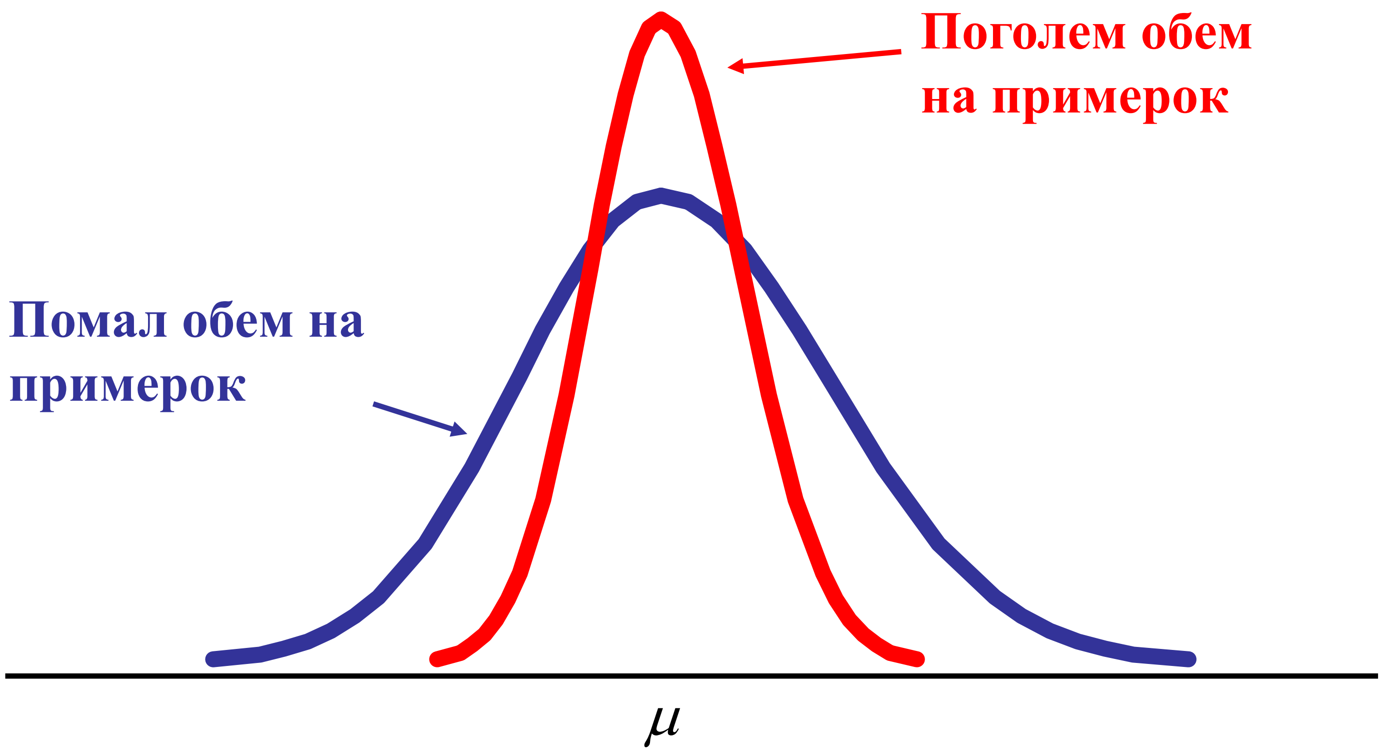
- Врз основа на претходното, математичкото очекување на просекот на примерокот е еднакво на математичкото очекување на обележјето.
- Дисперзијата на просекот е n пати помала од дисперзијата на обележјето.
- На првата слика е дадена распределбата на обележјето, а на втората распределбата на просекот на примерокот.





Распределба на просекот, ако обележјето X има нормална распределба

- Бидејќи дисперзијата на просекот на примерокот е $\sigma_{\bar{X}}^2 = {\sigma^2}/{n}$, јасно е дека дисперзијата се намалува со зголемување на обемот n на примерокот.



t - распределба

- Ако обележјето има нормална $N(\mu, \sigma^2)$ распределба, тогаш статистиката

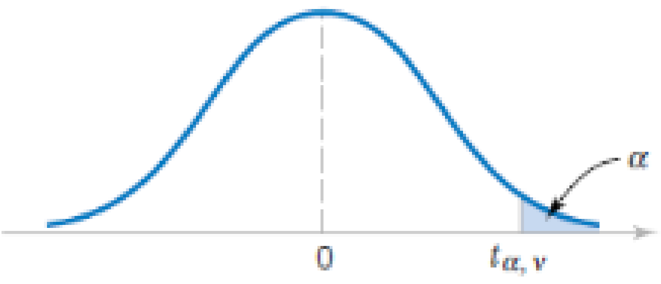
$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{S}\sqrt{n}$$

има Студентова t-распределба со $n - 1$ степени на слобода. Означуваме $T \sim t_{n-1}$.

- t-распределбата зависи од еден параметар и тој се нарекува ***број на степени на слобода***.
- Да воочиме дека оваа статистика се разликува од претходната Z статистика во тоа што стандардната девијација на обележјето σ е заменета со стандардната девијација S на примерокот.
- Оваа статистика ќе ја користиме подоцна, кога стандардната девијација на обележјето нема да биде позната.

Таблица за *t*-распределба

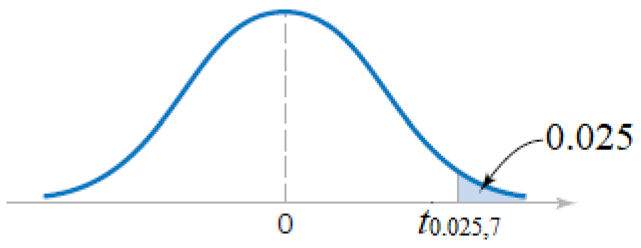
ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646



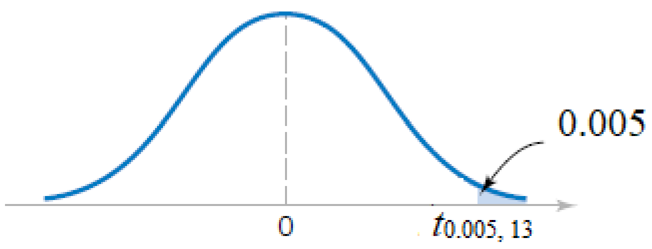
Читање од таблица за *t*- распределба

- Од таблицата за даден број на степени на слобода ν и дадено α, се чита вредноста $t_{\alpha,\nu}$ така што $P\{T > t_{\alpha,\nu}\} = \alpha$.

$t_{0.025,\ 7} = 2.365$



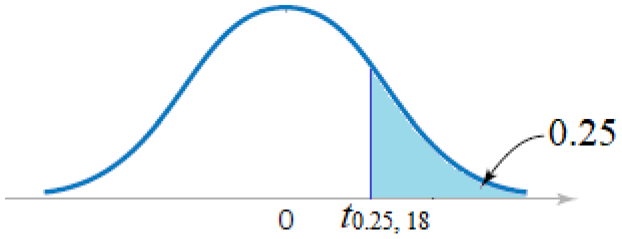
$t_{0.005,\ 13} = 3.012$



Читање од таблица за t- распределба

- Ако е дадено дека $T \sim t_{18}$, се поставува прашање, како да се определи вредноста x така што $P\{T > x\} = 0.25$.
- Согласно, табелираните вредности во табелата за t-распределба, $x = t_{0.25,\,18} = 0.688$

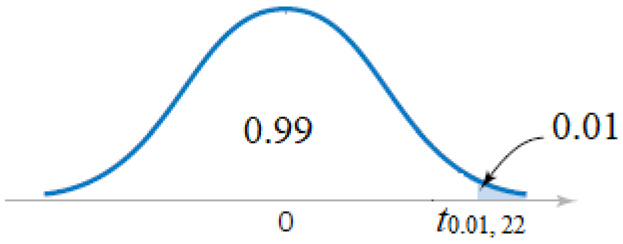
$$x = t_{0.25,\,18} = 0.688$$



Читање од таблица за *t*- распределба

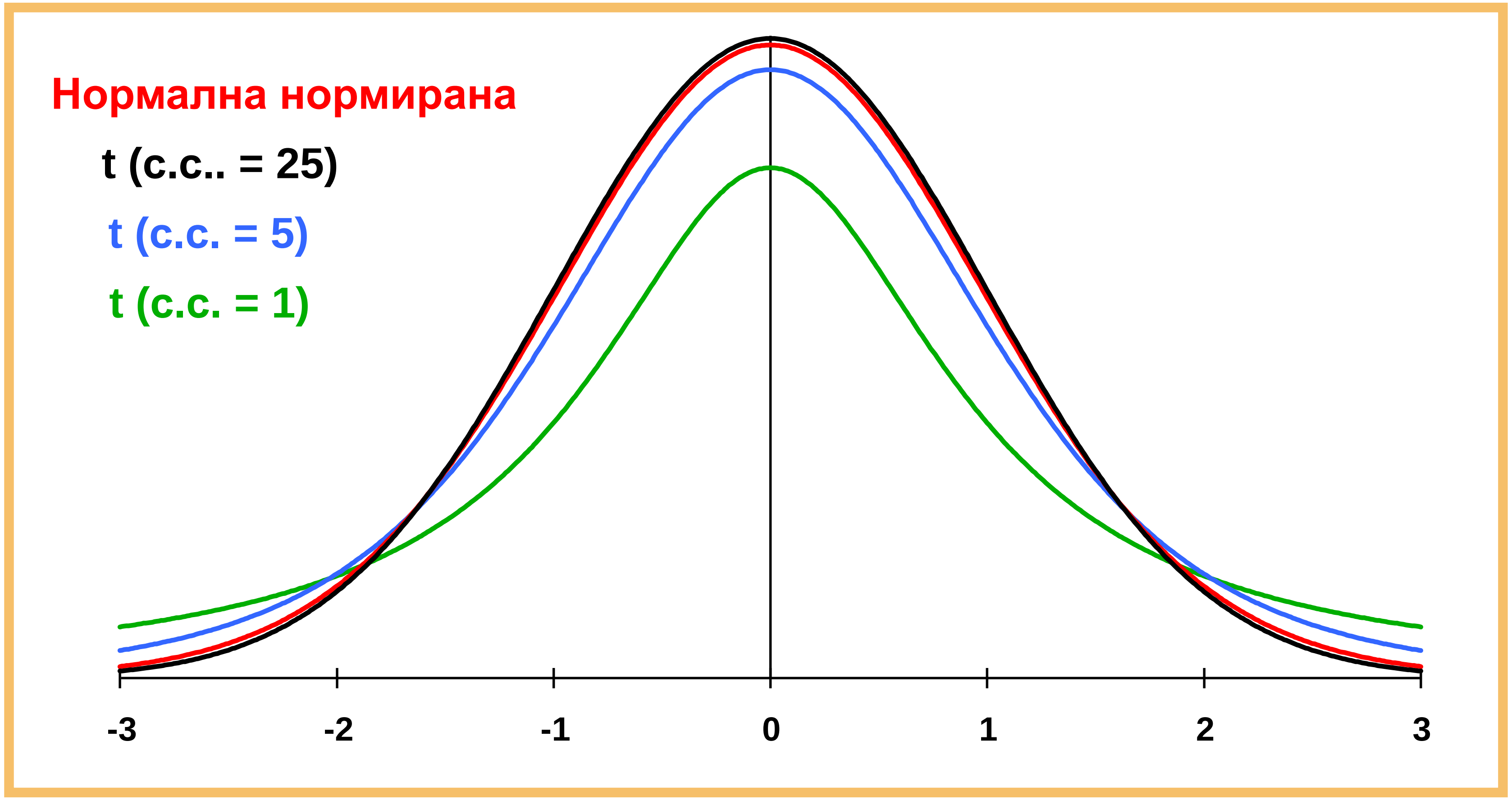
- Ако е дадено дека $T \sim t_{22}$, се поставува прашање, како да се определи вредноста x така што $P\{T < x\} = 0.99$?
- Врз основа на претходното, $P\{T > x\} = 0.01$, па бараната вредност $x = t_{0.01,\,22} = 2.508$.

$t_{0.01,\,22} = 2.508$



Споредба на *t*-распределба со нормална нормирана распределба

- Може да се воочи дека како се зголемува бројот на степени на слобода, графиците на густината на *t*-распределба се повеќе се доближуваат до графикот на нормална нормирана распределба.
- Апроксимацијата се смета за добра, ако $n > 30$.



Апроксимација на *t*-распределба со нормална нормирана распределба

- Врз основа на претходното, може да заклучите дека за $n \leq 30$, статистиката

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{S}\sqrt{n} \quad \sim \quad t_{n-1}$$

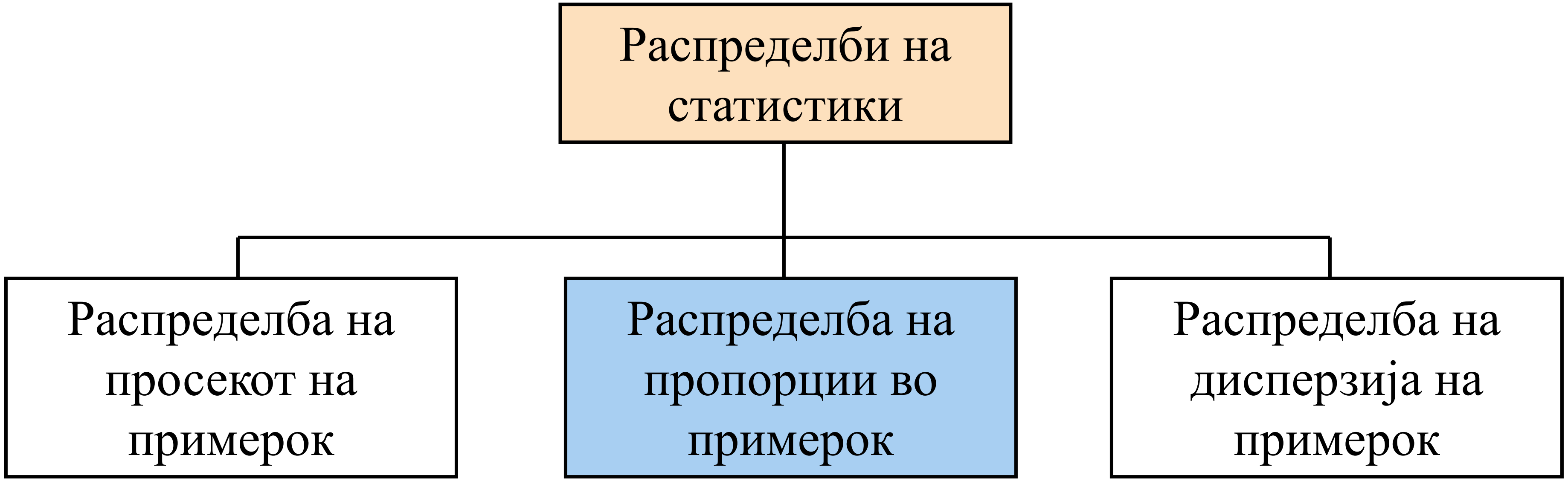
- Но, ако $n > 30$, тогаш распределбата на оваа статистика може да се апроксимира со нормална нормирана распределба. Во овој случај, ако е потребно да се прочитаат одредени вредности од таблица (како во претходните примери), тие се читаат од таблица на нормална нормирана распределба.

Распределба на просекот, ако обележјето X нема нормална распределба и $n > 30$

- Ако обележјето нема нормална распределба, но примерокот е доволно голем, $n > 30$, тогаш врз основа на централна гранична теорема може да се заклучи дека просекот на примерокот има приближно нормална распределба без оглед на распределбата на обележјето.

- Како што веќе покажавме математичкото очекување на просекот на примерокот е μ, а стандардна девијација е ${}^{\sigma}/_{\sqrt{n}}$.

Распределба на пропорции во примерок



Распределба на пропорции во примерок

- Нека p е веројатноста од популацијата да земеме единка која има одредена карактеристика, т.е. пропорција на единките со одредена карактеристика во целата популација.

- Нека случајната променлива X е дефинирана со: $X = 1$ ако единката ја има карактеристиката и $X = 0$ ако ја нема карактеристиката. Тогаш

$$P\{X = 1\} = p \text{ и } P\{X = 0\} = 1-p,$$

т.е.

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

- Сега,

$$EX = 0 \cdot (1-p) + 1 \cdot p = p$$

$$EX^2 = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$DX = p - p^2 = p(1-p)$$

Распределба на пропорции во примерок

- Ако $(X_1, X_2, \ldots, X_n)$ е примерок со големина n, тогаш пропорцијата $\hat{P}$ на единките во примерокот со разгледуваната карактеристика дава оценка на пропорцијата (веројатноста p) во популацијата. Имено,

$$\hat{P} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} = \frac{\text{број на единки во примерокот кои ја имаат карактеристиката}}{\text{големина на примерокот}}$$

- Притоа, $0 \leq \hat{P} \leq 1$.

Распределба на пропорции во примерок

$$\hat{P} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X} = \frac{\text{број на единки во примерокот кои ја имаат карактеристиката}}{\text{големина на примерокот}}$$

- Значи, $\hat{P}$ е претставена како просек на случајни променливи кои се независни и еднакво распределени со Бернулиева распределба.
- Ако обемот на примерокот е доволно голем ($n > 30$), тогаш согласно централна гранична теорема, распределбата на просекот ќе тежи кон нормална распределба со математичко очекување p и стандардна девијација $\sqrt{p(1-p)/n}$.
- Со нормирање на $\hat{P}$ се добива статистика чија распределба ќе тежи кон нормална нормирана распределба, т.е.

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Пример 3

- Ако вистинската пропорција на гласачи кои подржуваат некој закон A е $p = 0.4$, колку е веројатноста дека во случаен примерок од 200 гласачи пропорцијата на подржувачите на законот е меѓу 0.40 и 0.45?

***Решение*:** Дадено е дека $n = 200$ и $p = 0.4$. Се бара веројатноста

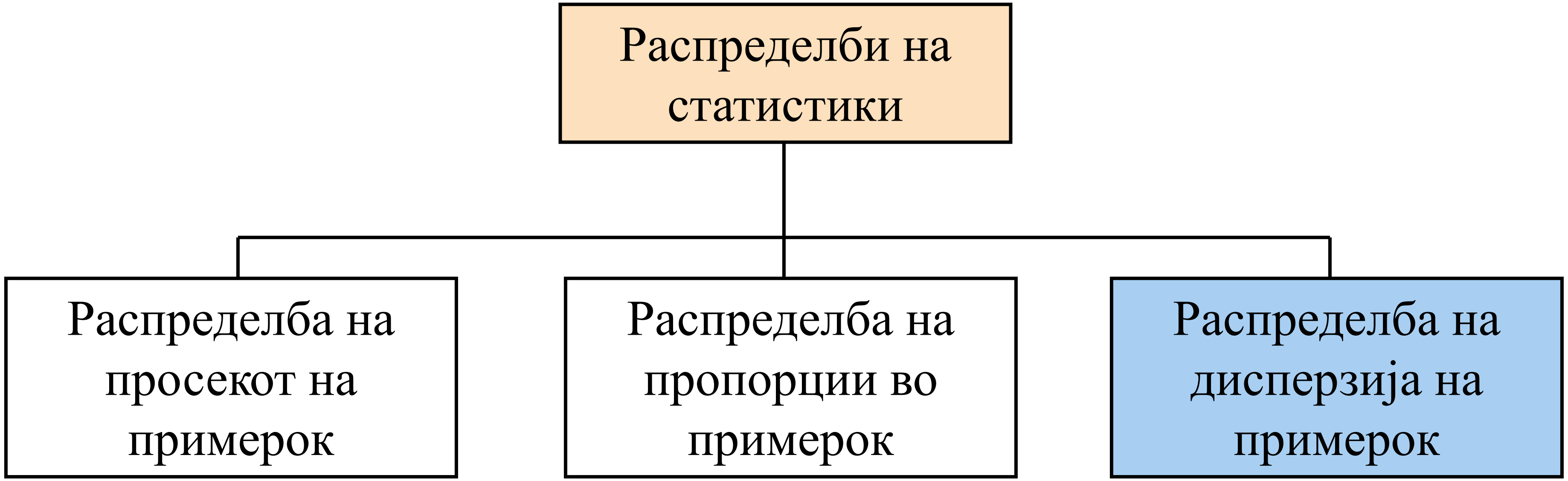
$$P\{0.40 \le \hat{P} \le 0.45\}.$$

- Добиваме:

$$P\{0.40 \le \hat{P} \le 0.45\} = P\left\{\frac{0.40-0.40}{\sqrt{0.4\cdot 0.6/200}} \le \frac{\hat{P}-p}{\sqrt{p(1-p)/n}} \le \frac{0.45-0.40}{\sqrt{0.4\cdot 0.6/200}}\right\}$$

$$= P\{0 \le Z \le 1.44\} = \Phi(1.44) - \Phi(0) = 0.92707 - 0.5 = 0.4207$$

Распределба на дисперзија на примерок



Распределба на дисперзија на примерок

- Ако (X_1, X_2, . . . , X_n) е случаен примерок од дадено обележје, тогаш

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

е дисперзијата на примерокот.

- Се покажува дека математичкото очекување на дисперзијата на примерокот S^2 е σ^2, т.е.

$$ES^2 = \sigma^2$$

- Ако обележјето има нормална $N(\mu, \sigma^2)$ распределба, тогаш статистиката

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

има χ^2-распределба (Хи-квадрат распределба) со n – 1 степени на слобода. Означуваме $\chi^2 \sim \chi^2_{n-1}$.

χ^2-распределба

- χ^2-распределбата зависи од еден параметар и тој се нарекува број на степени на слобода.
- На сликата се дадени графиците на густината на χ^2-распределба со 3, 5 и 10 степени на слобода.

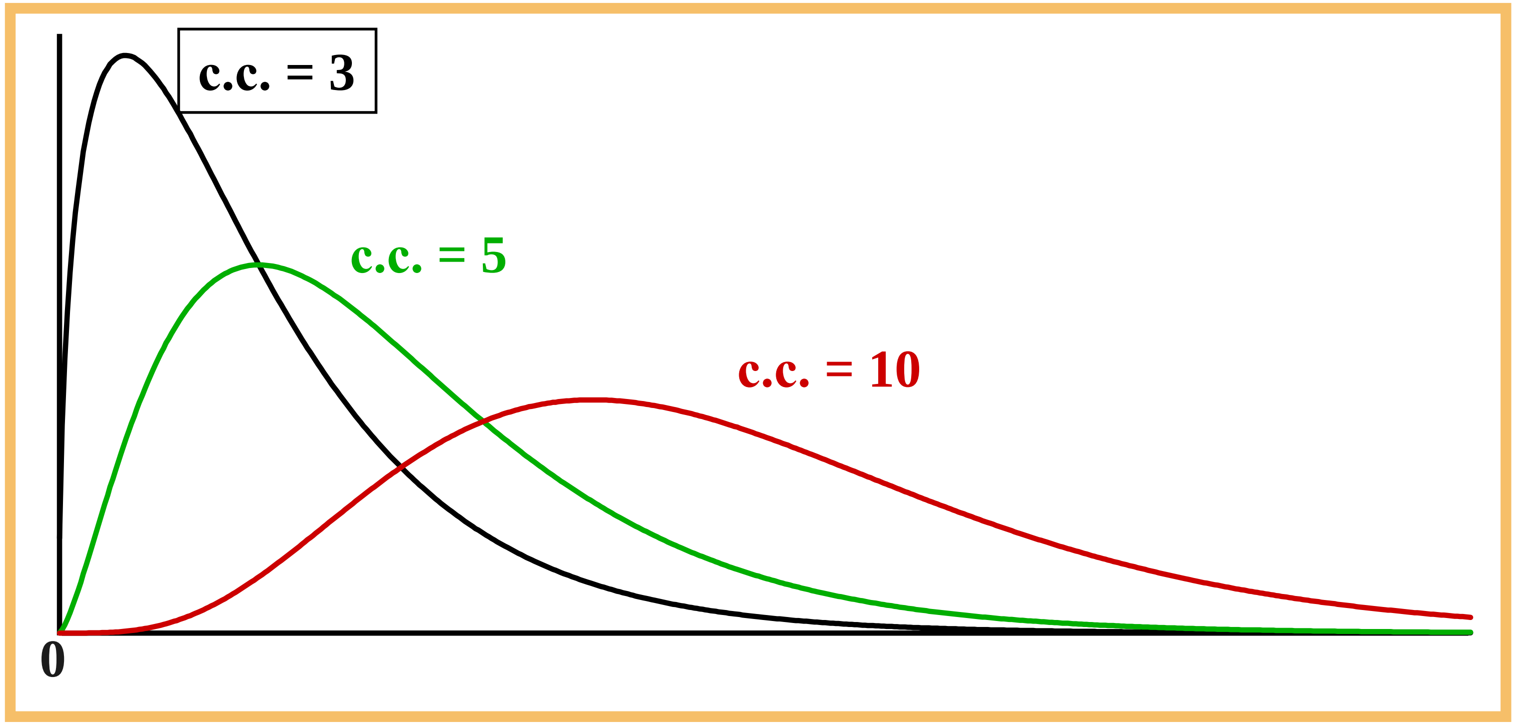
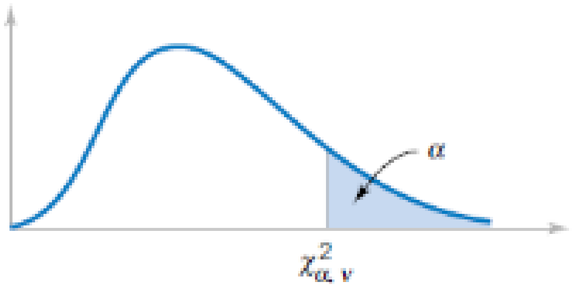


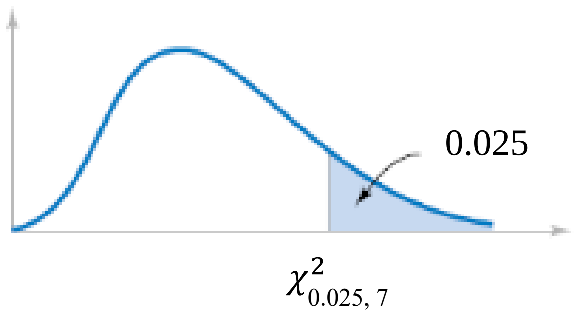
Таблица за χ^2-распределба



ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80

Читање од таблица за χ^2-распределба

- Од таблицата за даден број на степени на слобода ν и дадено α, се чита вредноста $\chi^2_{\alpha,\nu}$ така што $P\{\chi^2 > \chi^2_{\alpha,\nu}\} = \alpha$.



$$\chi^2_{0.025,\,7} = 16.01$$

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80

Пример

Комерцијален замрзнувач мора да ја одржува избраната температура со многу мало варирање. Спецификацијата бара стандардната девијација да не е поголема од 4° C. Треба да се тестира примерок од 14 замрзнувачи. Ако обележјето има нормална распределба со стандардната девијација 4, колку е вредноста K така што веројатноста дисперзијата на примерокот да ја надмине вредноста K е 0.05?

Решение:

- Треба да ја определиме вредноста K, така што $P\{S^2 > K\} = 0.05$.
- Знаеме дека за примерок со обем 14 и стандардна девијација на обележјето $\sigma = 4$, статистиката

$$\chi^2 = \frac{(14-1)S^2}{16}$$

има χ^2- распределба со 13 степени на слобода.

Пример

- Од таблицата за χ^2-распределба ја читаме вредноста $\chi^2_{0.05,13}$ така што

$$P\{\chi^2 > \chi^2_{0.05,13}\} = 0.05.$$

- Се добива дека $\chi^2_{0.05,13} = 22.36$.
- Сега, треба да ја определиме вредноста K, така што $P\{S^2 > K\} = 0.05$. Добиваме

$$P\{S^2 > K\} = P\left\{\frac{(14-1)S^2}{16} > \frac{(14-1)K}{16}\right\} = P\left\{\chi^2 > \frac{(14-1)K}{16}\right\} = 0.05$$

- Од претходните два изрази, јасно е дека $\frac{(14-1)K}{16} = \chi^2_{0.05,13} = 22.36$.
- Со решавање по K, се добива

$$\frac{(14-1)K}{16} = 22.36$$

$$K = \frac{(22.36)\cdot(16)}{(14-1)} = 27.52$$