Оценяване на параметри

Точкови оценки

***Параметърът** е числова характеристика на популацията*

Точковата оценка е едно число (стойност), което се използва, за да се оцени неизвестната стойност на параметъра.

Зависи от извадката и има характер на случайна величина

Параметър	Точкова оценка
Средна стойност μ	*Извадково средно* $\overline{x}$
Дисперсия σ^2	*Извадкова дисперсия* s^2
Вероятност p	*Извадкова пропорция* $\hat{p}$

Неизместеност на точкова оценка

Средната стойност на статистиката= параметъра

Пример 1. Извадковото средно $\overline{X} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

Нека случайната извадка с обем n е от популация със средна стойност μ

$$E\overline{X} = \frac{EX_1 + EX_2 + \ldots EX_n}{n} = \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$

$$EX_i = \mu$$

$$E\overline{X} = \mu$$

Извадковото средно е **неизместена оценка** на параметъра "*популационна средна стойност* μ"

Пример 2. Извадкова пропорция $\hat{p}$

Извадка с обем n е направена от алтернативна популация.

$$E\hat{p} = \frac{EX}{n} = p$$

Извадковата пропорция е **неизместена оценка** на параметъра "*вероятност p*"

Доверителни интервали

Интервал, построен на основата на извадката, който съдържа неизвестния параметър с вероятност близка до 1

Пример:

Време на полет : 11 часа± 15 мин ➔ 10 ч 45 мин ~ 11 ч 15 мин.

средна стойност(математическо очакване)= точното времетраене на полета - неизвестно

Ниво на доверие

Означение: $(1-\alpha)100\%$

примери:

90% дов. инт.,

95% дов. инт.

Доверителен интервал за μ (σ известно)

Случай 1. Нормална популация

Нека е направена извадка с обем *n* от *нормално разпределение* с неизвестна средна стойност μ, и известно стандартно отклонение σ

Как да построим доверителен интервал за μ с (1-α)100% ниво на доверие ?

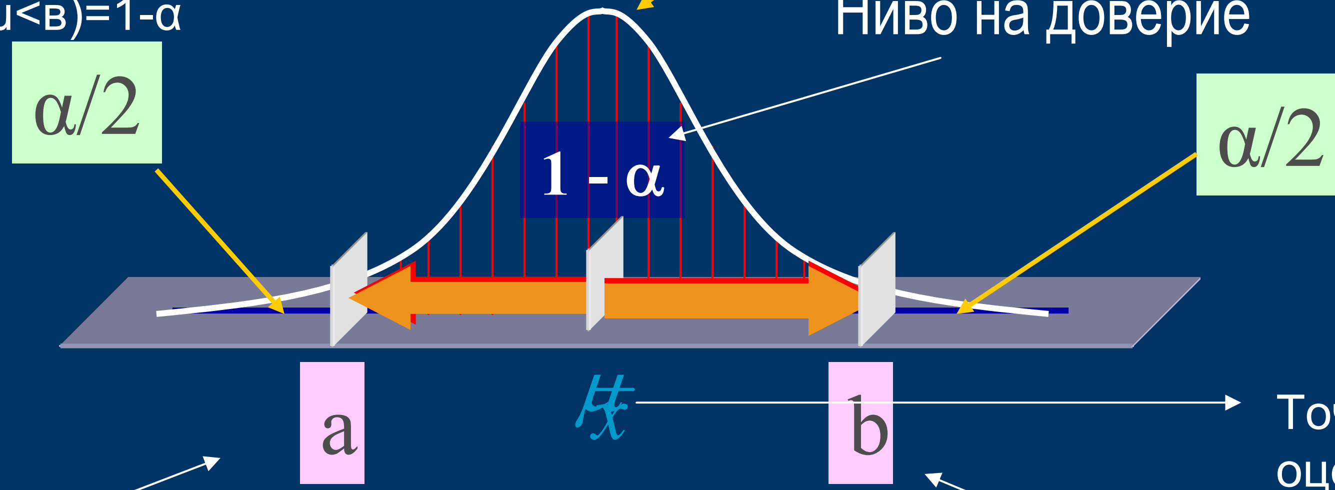
Точкова оценка на μ е извадковото средно

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \in N(\mu, \frac{\sigma^2}{n})$$

Избираме си ниво на доверие (1-α)100% и търсим доверителен интервал (а,в) така, че P(а<μ<в)=1-α

Плътност на извадковото средно

Ниво на доверие



Точкова оценка на μ е $\overline{x}$

Долна граница

Горна граница

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$z_{\frac{\alpha}{2}} = \frac{a-\overline{x}}{\frac{\sigma}{\sqrt{n}}}$$

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Пример

Нека X е количеството мляко, консумирано от българите дневно. Известно е, че това количество е нормално разпределено със стандартно отклонение 96 г. За да оценят средното количество, Асоцията на млекопроизводителите е направила случайна извадка от 576 българи и е получила средна стойност 133 грма дневно. Постройте 90% доверителен интервал.

Точкова оценка на неизвестната средна консумация е извадковото средно, $\bar{x} = 133$

1-α=0,9   α=0,1   α/2=0,05 $\Rightarrow$ $z_{0,05} = 1,645$

$$133 \pm 1,645 \frac{96}{\sqrt{576}} \Rightarrow 126,42 \leq \mu \leq 139,58$$

Интерпретация

На нивото на доверие

Пример

90% ниво на доверие

Ако построим 100 доверителни интервала, всеки на основата на различни извадки от една и съща популация, то можем да очакваме, че 90 от тях ще съдържат параметъра на популацията.

Доверителен интервал за μ (σ известно)

Случай 2. Популацията не е нормална, но обемът на извадката е достатъчно голям (по-голям от 30)

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Пример: Направено е допитване до 100 деца в предучилищна възраст като са записани броя часове седмично прекарани пред телевизора и е намерено, че средното време е 27,197. Ако е известно, че стандартното отклонение на броя часове, прекарани пред телевизора за тази група от деца е 8.0 часа, то постройте 95% доверителен интервал на средното прекарано време седмично пред телевизора.

Решение

Параметърът, който ще оценяваме е μ –средното време, прекарано пред телевизора

Следователно, трябва да построим доверителен интервал за μ при условие, че не знаем нищо за разпределнието на популацията, но обемът на извадката n=100>30

$$\boxed{\bar{x} = 27{,}197}$$

Доколкото 1 - α =0,95, α = 0,05. то $\alpha/2$ = 0,025 и $Z_{0,025}$ = 1,96

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 27{,}191 \pm z_{0{,}025} \frac{8}{\sqrt{100}}$$

$$= 27{,}191 \pm 1{,}96 \frac{8}{\sqrt{100}} = 27{,}191 \pm 1{,}57 = [25{,}621 ;\ 28{,}761]$$

Връзка между нивото на доверие и дължината на доверителния интервал

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Нека променим само нивото на доверие=> променя се само z

Случай 1:Ниво на доверие 90%=(1-α)100%

$\alpha = 0{,}1$ $\alpha/2 = 0{,}05$

$$z_{0,05} = 1{,}645$$

$$\left[\bar{x} - 1{,}645\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1{,}645\frac{\sigma}{\sqrt{n}}\right]$$

Случай 2: Ниво на доверие 95% =(1-α)100%

$\alpha = 0{,}05$ $\alpha/2 = 0{,}025$

$$z_{0,025} = 1{,}96$$

$$\left[\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right]$$

По-малко ниво на доверие = по-тесен доверителен интервал.

Ефект на обема на извадката върху ширината на доверителния интервал

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right]$$

Нека променим само обемът на извадката => променя се само **n**

Случай 1: обем n=10

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{10}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{10}} \right]$$

Случай 2: обем n=100

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{100}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{100}} \right]$$

По-голям обем на извадката => по-тесен доверителен интервал.

Доверителен интервал за μ (σ неизвестно)

Предположения

- **Стандартното отклонение на популацията е неизвестно**
- **Популацията е нормално разпределена**

Точкова оценка на σ е s $s^2 = \dfrac{\Sigma(X - \overline{X})^2}{n-1}$

Знаем, че $\overline{X} \in N(\mu, \dfrac{\sigma^2}{n})$

Но σ е неизвестно и не можем да го използваме.

Заместваме σ с точковата му оценка S.

Важно

Извадковата дисперсия **s²**, както и извадковото стандартно отклонение **s** са случайни величини.

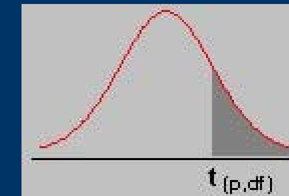
Тогава $\dfrac{\overline{X}-\mu}{\dfrac{s}{\sqrt{n}}} = \dfrac{\overline{X}-\mu}{s}\sqrt{n}$

Случайна величина

е **частно** на две случайни величини

Има специално разпределение, което се нарича =>

t-разпределение



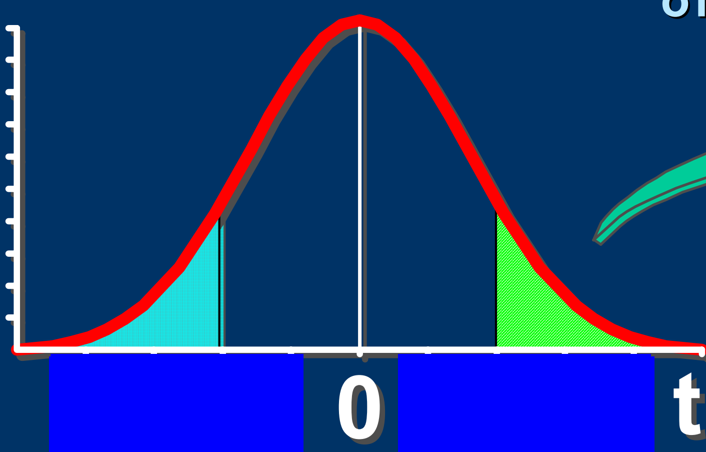
Има степени на свобода

Плътността е симетрична относно 0, т.е. 0=медиана

В t – таблица са дадени лицата под графиката на плътността

Нека ст.св. = 2

Намерете двете точки, които отсичат 20% от графиката на плътността



	Дясна опашка		
Ст. св.	.25	.10	.05
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

Означение

$$t_{n,\alpha} : \quad P(T > t_{n,\alpha}) = \alpha$$

n - степени на свобода

α лице на дясна опашка

$\boxed{t_{5;0,005} = ?}$ = 4,03214

$\boxed{t_{4;0,9} = ?}$ = - 1,533206

df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.619 2
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688

Обратно към

Доверителен интервал за μ (σ неизвестно)

Предположения

- **Стандартното отклонение на популацията е неизвестно**
- **Популацията е нормално разпределена**

$$T = \frac{\overline{X} - \mu}{s}\sqrt{n} \in t(n-1)$$ **t** разпределение с **n-1** степени на свобода

Използваме таблицата за t -разпределение

Доверителния интервал :

Станд. Откл.

$$\overline{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

Тези стойности за получени от таблиците за t- разпределението

Пример

Организаторите на компютърна зала искат да знаят колко време седмично студентите прекарват в залата. За целта се избират случайно 16 студенти и се намира, че тяхното средно седмично време е 24 часа със стандартно отклонение 4 часа. Намерете 95% доверителен интервал на неизвестното средно време на всички студенти, ако се знае, че времето прекарано в компютърната зала е нормално разпределено.

Средната стойност на популацията е неизвестна. *Точкова оценка* на тази стойност е извадковото средно 24 часа.

$\alpha = 0{,}05$

$\alpha/2 = 0{,}025$

Решение

Какво е дадено? **Популацията е нормално разпределена**

Извадковото станд. откл. s=4

Използваме t-разпределение, степени на свобода=16-1=15

от t-таблицата

$$t_{15,0.025} = 2{,}131$$

$$\overline{X} \pm 2{,}131\frac{s}{\sqrt{n}} = 24 \pm 2{,}131\frac{4}{\sqrt{16}}$$

$$= 24 \pm 2{,}131$$

Доверителният интервал е (21,869; 26,131).

Връзка между нивото на доверие и дължината на доверителния интервал

Разглеждаме *t (15)*

Ниво на доверие =95%

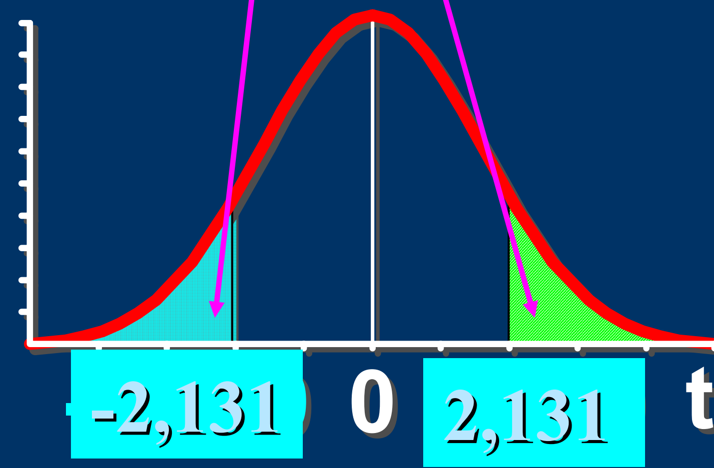
α = 0,05

α /2 = 0,025

Ниво на доверие =90%

α = 0,1

α /2 = 0,05



По=малко ниво на доверие (1-α) –
по-тесен доверителен интервал.

Ефект на обема на извадката върху ширината на доверителния интервал

$$\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

Ако се променя само обемът, то в този случай се променя и стойността на t и стойността на s => не може да се направи извод за ширината

Сравни с доверителния интервал с известна дисперсия!!!

Определяне обема на извадката

Колко голяма да е извадката, за да оценим средната стойност?

Оценяването става с доверителен интервал

Максималната грешка при оценяването = половината дължина на интервала

μ

Доверителен интервал

Случай 1. Знае се стандартното отклонение σ, $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ грешката е E

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < E$$

$$n > \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$$

Случай 2. Не се знае стандартното отклонение- предварително се правят някакви оценки.

Например, може да апроксимираме σ с s и да използваме нормалното разпределение (голямо n)

А ако се оценява процент , то

$$n > \left(\frac{z_{\frac{\alpha}{2}}\sqrt{p(1-p)}}{E}\right)^2 = \left(0{,}5\frac{z_{\frac{\alpha}{2}}}{E}\right)^2$$

Пример

Нека X е количеството мляко, консумирано от българите дневно. Известно е, че това количество е нормално разпределено със стандартно отклонение 16 г. Трябва да се направи извадка, с помощта на която да се построи 90% доверителен интервал, при които грешката да не е повече от 0,01 грама.

Колко е обемът на извадката????

$\sigma = 4$ 1-α=0,9 α/2=0,05 $z_{0,05} = 1,645$

$$n > \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1,645*4}{0,01}\right)^2 = 432\ 964$$

Поне 432 965 българи трябва да се изберат по случаен начин

Ако увеличим грешката от 0,01 на 1 грам, то $n > \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1,645*4}{1}\right)^2 = 43,2964$

Само 44, обемът се намалява

Тестване хипотези относно популационна пропорция

$\hat{p}$ p

Пропорцията е частно или процент, който показва каква част от популацията или от извадката притежава определена характеристика.

- **Предположения**
  - Извадка с обем **n** е направена от алтернативна популация: всеки елемент или притежава или не притежава дадена характеристика
  - $n\hat{p}(1-\hat{p}) \geq 10$ където $\hat{p} = \dfrac{x}{n}$

x =броя на индивидите от извадката, които притежават характеристиката

Нека $\sigma^2 = n\hat{p}(1-\hat{p}) \geq 10$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Стандартно нормално разпределение

Твърдението е по отношение на популационната пропорция

H_0: $p = p_0$

H_1: $p \neq p_0$

Двустранен тест

H_0: $p \geq p_0$ H_1: $p < p_0$

Лявостранен тест

H_0: $p \leq p_0$ H_1: $p > p_0$

Дясностранен тест

статистика

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} \in N(0,1)$$

Тестване на хипотези

Критична област

р-стойност

Пример

Критична област

В миналото, 15% от получателите на определен вид писмо с молба за помощ са отговаряли с дарение. Асоциация “Помогни на детето” е подготвила за изпращане ново писмо, което счита, че би довело до увеличение на даренията. За целта, писмото е изпратено до случайно избрани 200 човека и 45 от тях правят дарение. При ниво на значимост 0,05 може ли да се заключи, че новото писмо е по-ефективно?

$$\hat{p} = \frac{x}{n} = \frac{45}{200} = 0,225$$

$$n\hat{p}(1-\hat{p}) = 200(0,225)(0,775) = 34 \geq 10$$

Решение

1. Нулева и алтернативна хипотеза

$H_0: p \leq 0,15$ $H_1: p > 0,15$

2. Ниво на значимост

$\alpha = 0,05$

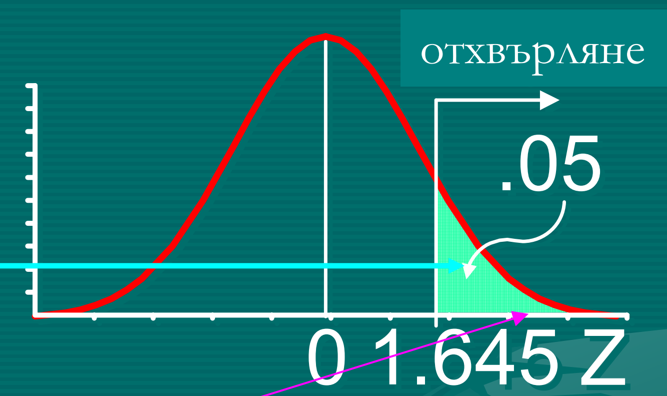
3. Статистика и извадково разпределение

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_o(1 - p_0)}{n}}} = \frac{0,225 - 0,15}{\sqrt{\frac{0,15(1 - 0,15)}{200}}} = 2,97$$

4. Критична област

Критична област (1,6645 ; ∞)

$\alpha = 0{,}05$



5. Извод

z = 2,97 е в критичната област, затова отхвърляме H_0

6. Интерпретация

Повече от 15% даряват.Новото писмо е по-ефективно.

Пример p-стойност

Списание публикува, че 90% от домакинствата купуват елха за Коледа. За да се провери твърдението се прави случайна извадка от 500 домакинства, на които се задава въпроса “Имахте ли елха миналата година?” 461 от тях отговорили “да.”

$$\hat{p} = \frac{x}{n} = \frac{461}{500} = 0{,}922$$

$$n\hat{p}(1-\hat{p}) = 500(0{,}922)(0{,}078) = 36{,}9 \geq 10$$

Решение

1. Нулева и алтернативна хипотеза

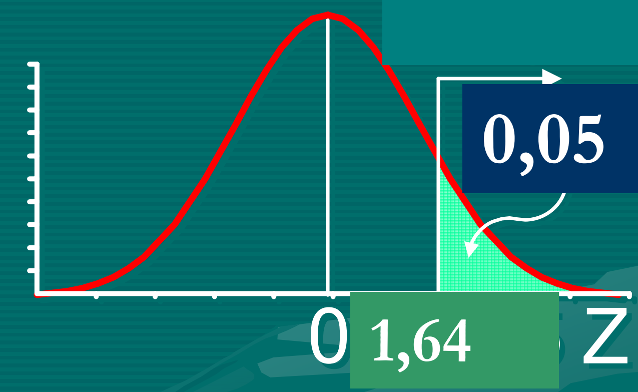
H_0: p= 0,9

H_1: p≠0,9

2. Статистика и извадково разпределение

$$z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_o(1-p_0)}{n}}}=\frac{0{,}922-0{,}9}{\sqrt{\frac{0{,}9(0{,}1)}{500}}}=1{,}64$$

4. p-стойност



5. Извод

p-стойност=2*0,05=0,1 не е достатъчно малко число, *затова не отхвърляме Но*

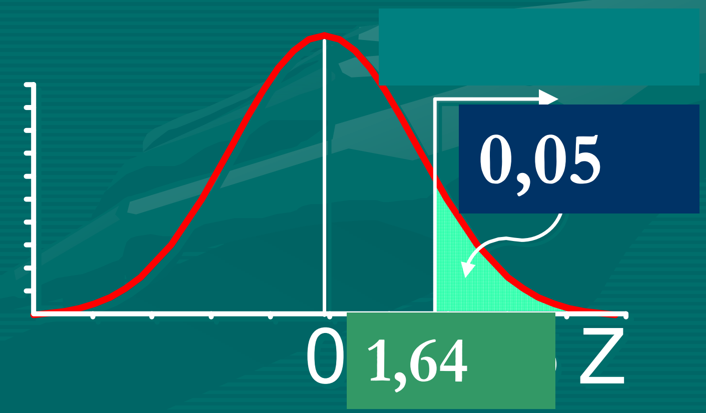
6. Интерпретиране

Няма основание да отхвърлим твърдението че точно 90% от домакинствата са купили елха.

Изменение : ако искаме да проверим дали повече от 90% са купили елха, то:

H_0: p= 0,9 **H_1: p>0,9**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_o(1-p_0)}{n}}} = \frac{0{,}922 - 0{,}9}{\sqrt{\frac{0{,}9(0{,}1)}{500}}} = 1{,}64$$

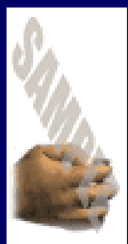


р-стойност=0,05=0,1 е достатъчно малко число, ***затова отхвърляме Но***

Повече от 90% от домакинствата са купили елха.

Теория на вероятностите и приложна статистика

(спец. Информатика)



Доц. д-р Снежана Христова каб. 342

E-mail : snehri@uni-plovdiv.bg

http://www.fmi-plovdiv.org/pmm/Informatika_10/index.htm

Вероятност ???

Статистика ???

Проверка на знанията:

Контролни работи: 2 контролни работи. Поправка на контролна работа не се допуска. При отсъствие по уважителна причина (след представяне на официален документ) контролната може да се направи допълнително по споразумение с преподавателя.   Максимум точки на всяка контролна 50

Малки контролни работи: През семестъра по време на упражненията ще бъда правени 10-минуни контролни (общо 5 на брой с максимален общ брой точки 30)

Домашни работи: През семестъра ще бъдат дадени 2 домашни задания, които трябва да се предадат в определен срок. Всяко от тях съдържа 5 задачи, като само една, избрана по случаен начин от студент ще бъде оценявана. Максимален брой точки на всяка домашна е 10 точки

Изпит: *първа част* (вероятности) Максимум 80+50 точки
и *втора част* (статистика) Максимум 70 + 50 точки.
Всяка част на изпита се състои от теоретична част   и една задача.
Максимум 250 точки

Пример 1

Нека оценките на студент са: контролна 1 -*25 точки*, контролна 2 - *28 точки*, домашна 1 – *10 точки*, домашна 2 – *10 точки*, общо точки от 10-минутните тестове (лекции и упр.) – *11 точки*, изпит ТВ – *15 точки*, изпит ПС- *30 точки*, и

$$Оценка = 2 + \frac{25+28+10+10+11+15+30}{100} = 3{,}29$$

крайна оценка Среден 3

ПРИМЕР 2

Нека оценките на студент са:
контролна 1 -*38 точки*, контролна 2 - *42 точки*, домашна 1 – *10 точки*, домашна 2 – *10 точки*, общо точки от 10-минутните тестове (лекции и упр.) – *11 точки*, **Не се явява на изпит**

$$Оценка = 2 + \frac{38+42+10+10+11}{100} = 3{,}11$$

крайна оценка Среден 3

Времетраене на лекциите

9:15-10:20

10:30- 11:40

Вероятност ???

Всекидневно
- Тото 6 от 49
- Тото 5 от 35

Технологиите

Конструирането на атомен реактор - надежност

Финанси

Цена на акции

Компютри ???

Мярка за възможността нещо да се случи

С. Христова

Основни понятия в ТВ

Опит — Всеки процес, който може да се повтори и чийто резултати са неизвестни

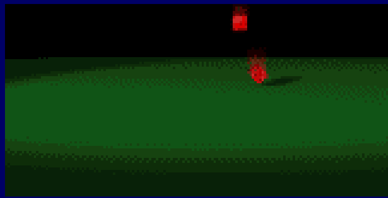
Елементарно събитие — Всеки изход на даден опит

Пространство от събития, S — Съкупност от всички елементарни изходи, свързани с даден опит

Събитие — Всяка съвкупност от елементарни събития (всяко подмножество на S).

Обикновено събитията се означават с главни латински букви: *A*, *B*,*C*,...

Един изход е благориятен за събитие A, ако е елемент на A



Примери

опит: Хвърляне на монета един път. S={Л,Г}, A={Г}, B={Л}

Опит: хвърляне на зарче един път. S={1,2,3,4,5,6},
A={нечетен брой точки върху зара}={1,3,5}
B={поне 5 точки на зара}={5,6}
C= {по-малко от 4 точки на зара}={1,2,3}

Опит: избор на листче измежду 4 листчета с написани числата от 1 до 4 върху тях. S={1,2,3,4},
A={нечетно число върху листчето}={1,3}
B={число по-голямо от 4 върху листчето}=празно

Опит: избор на семейство измежду всички с две деца.
S={BB,AA,BA,AB}, където A-момиче, B-момче
A={семейството има едно момче}={BA, AB}
B={семейството има поне едно момче}={BA, AB, BB}

Опит: стрелба по кръгова мишена.
S={всички точки от кръга}
A={попадение в десятката}={точките от кръга, които са означени с 10}

Опит: Хвърляне на два различни зара едновременно
Елементарно събитие(изход)= наредени двойки от вида (i,j)
S={11,12,13,14,15,16,21,22,23,…,66},
A={еднакви числа върху зарчетата}={11,22,33,44,55,66}
B={сумата от точките е 7}={16,61,25,52,34,43}

Опит: Хвърляне на един зар два пъти

Опит: Хвърляне на монета до поява на лице.
S={Л, ГЛ,ГГЛ,ГГГЛ,ГГГГЛ, ….. },
A={точно един герб}={ГЛ}
B={поне един герб}={ГЛ,ГГЛ,ГГГЛ,ГГГГЛ, ….. }

Видове пространства от елем. изходи S

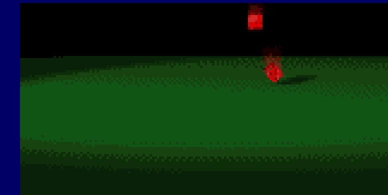
крайномерни

Изброими безкрайни

неизброими

С. Христова

Практическо приложение



Опит Всеки процес, който може да се повтори и чийто резултати са неизвестни

Цена на акции

Невъзможно е опитът да се повтори при едни и същи условия много, много пъти => теоретически считаме, че е възможно

S={всички положителни числа},
A={цената е по-висока от 10 лв}={всички числа >10}

Времето

S={дъжд, сняг, ясно}

Но може и друга интерпретация

Видове събития

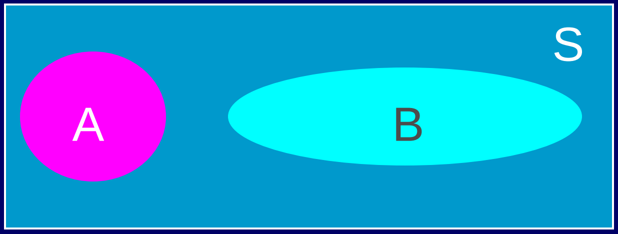
Достоверно

Състои се от всички изходи, свързани с даден опит= S

Невъзможно ∅

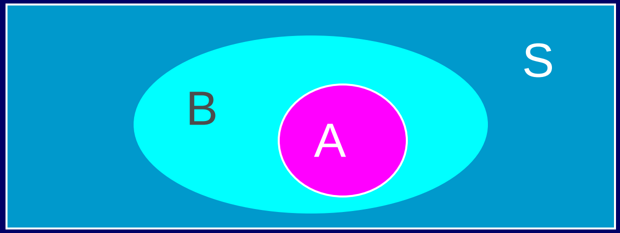
Несъвместими

Събития, които нямат общи благоприятни изходи



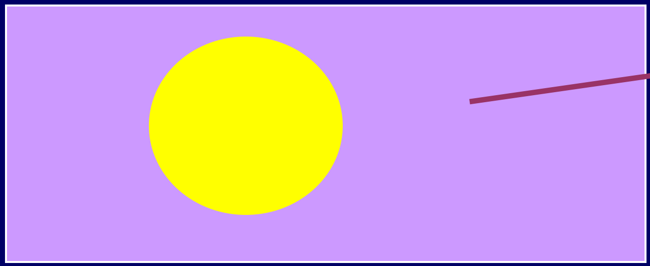
А влече В

Всеки благоприятен изход на А е благоприятен изход и за В



Допълнение

Събитието $\overline{A}$ се нарича допълнение на събитието А, ако се състои от всички изходи на пространството S, които не принадлежат на А



Допълнение на събитието А

Казваме, че настъпва събитието А, ако след изпълнение на опита се наблюдава изход от А (благоприятен за А)

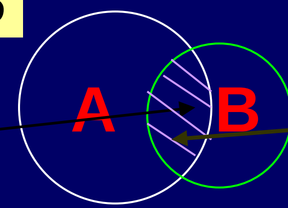
$\overline{A}$ и А са несъвместими

Действия със събития

Нека А,В са събития

$A \cap B$

Сечение ***на две събития:***

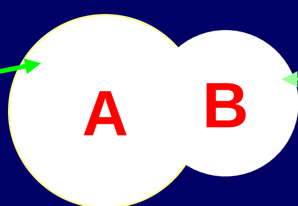


и

А и В е събитие, което се състои от всички изходи, които принадлежат както на А така и на В.

Сума ***на две събития :***

$A \cup B$



или

А или В е събитие, което се състои от всички изходи, които принадлежат или на *А*, или на *В*, или и на двете

Примери

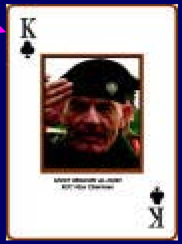
Карта е избрана по случаен начин от колода от 52 карти

A={избраната карта е черна}   B={избраната карта е пика}

C={избраната карта е поп}   D={избраната карта е спатия}

B или C= Избраната карта е пика или поп $= B \cup C$

B и C= Избраната карта е поп пика $= B \cap C$



A или B= Избраната карта е черна $= A \cup B$

A и B= Избраната карта е пика $= A \cap B$

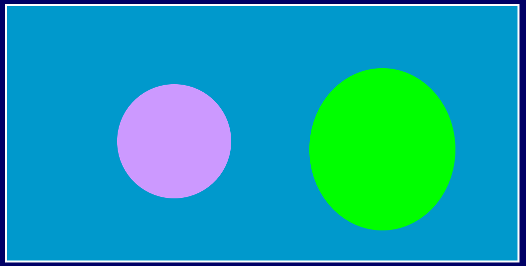
B или D= A $= B \cup D$

B и D= празно $= B \cap D$

Свойства

Нека A и B са несъвместими събития

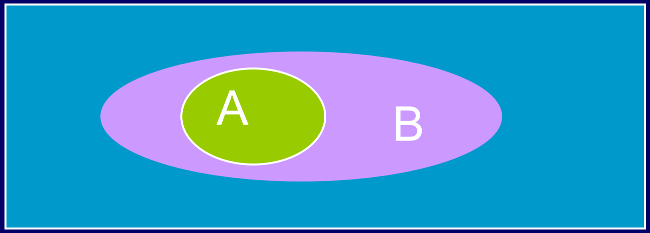
$A \cap B$ = A и B= невъзможното



$A \cap \overline{A}$ = невъзможното

$A \cup \overline{A} =$ S

Нека A влече B



$A \cap B$ = A и B= A

$A \cap S$ = A и S= A

$A \cup B$ = A или B= B

$A \cup S$ = A или S= S

Вероятност

Вероятността на едно събитие А, ще означаваме с *P*(*A*). Тя изразява възможността това събитие на настъпи.

Дефиниция

На всяко събитие А се съпоставя число P(A) за което:

1. **P(A)≥0**
2. **P(S)=1**
3. **P е (безкрайно) адитивна, т.е. ако** $A_1, A_2, \dots$ **е крайна или безкрайна редица от несъвместими събития, то**

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

Свойства на вероятността

Свойство 1

P(*невъзможното*)=0

Доказателство Нека A=S, B= ∅ => те са несъвместими и сумата им дава S => от аксиома 3

$$1 = P(S) = P(S \cup \varnothing) = P(S) + P(\varnothing) = 1 + P(\varnothing)$$

$$\xrightarrow{\text{аксиома 1}} \quad P(\varnothing) = 0$$

Свойство 2

Ако A влече B ⇒ **P(A)≤P(B)**

Свойство 3

$$P(\overline{A}) = 1 - P(A)$$

Доказателство A или $\overline{A}$ =S $\xrightarrow{\text{от аксиома 3}}$ $P(\overline{A}) + P(A) = 1$

Свойство 4

0 ≤ P(A)≤1

Доказателство **Ако A влече S** → **P(A)≤1**

Частен случай

Класическа вероятност

Само за крайномерни пространства

$$P(A) = \frac{k}{n}$$

k- брой благоприятни изходи на събитието А
n – брой на всички възможни изходи

Примери

опит: Хвърляне на монета един път. S={Л,Г},
A={Г}, B={Л}

P(A)=1/2, P(B)=1/2, P(S)=1

Опит: хвърляне на зарче един път. S={1,2,3,4,5,6},
A={нечетен брой точки върху зара}={1,3,5}
B={поне 5 точки на зара}={5,6}
C= {по-малко от 4 точки на зара}={1,2,3}

P(A)=3/6=0,5 P(B)=2/6=1/3 P(C)=3/6=0,5

Опит: избор на семейство измежду всички с две деца.
S={BB,AA,BA,AB}, където A-момиче, B-момче
A={семейството има едно момче}={BA, AB}
B={семейството има поне едно момче}={BA, AB, BB}

P(A)=2/4=0,5 P(B)=3/4=0,75

Опит: избор на листче измежду 4 листчета с написани числата от 1 до 4 върху тях. S={1,2,3,4},
A={нечетно число върху листчето}={1,3}
B={число по-голямо от 4 върху листчето}=празно

P(A)=2/4 P(B)= 0

Опит: стрелба по кръгова мишена.
S={всички точки от кръга}
A={попадение в десятката}={точките от кръга, които са означени с 10}

P(A)=?????

Класическата вероятност е неприложима; пространството е безкрайно

Пример

Два различни зара са хвърлени едновременно един път.

А/ Пространството от събития?

Има *36* възможни изхода:

(1,1), (2,1), (3,1), (4,1), (5,1), (6,1), (1,2), (2,2), . . ., (5,6), (6,6)

Б/ каква е вероятността да се паднат еднакъв брой точки и на двата зара?

Има 6 възможни изхода:(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)

Вероятността е 6/36 = 1/6

В/ Каква е вроятността сумата от точките да е 7?

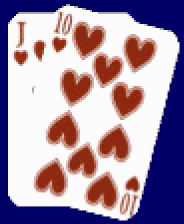
Има 6 възможни изхода

(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)

Вероятността е 6/36 = 1/6

Опит

Карта е избрана случайно от колода (тесте) карти (52 карти)



А/ Каква е вероятността избраната карта да е червена?

P(A)=26/52=1/2

Б/ Каква е вероятността избраната карта да е купа?

P(B)=13/52=1/4

В/ Каква е вероятността избраната карта да е поп?

P(C)=4/52=1/13

Г/ Каква е вероятността избраната карта да е червен поп?

P(D)=2/52=1/26

Д/Каква е вероятността избраната карта да е поп пика?

P(E)=1/52

Опит

Пет карти са избрани случайно от колода карти (52 карти)

А/ Каква е вероятността точно една от избраните карти да е червена?

Колко са всички възможни изходи- начини за избор на 5 карти измежду 52 ?

Как да броим ???

Техники за пресмятане

Принцип на умножението

- Ръководство на клуб – председател, касиер и секретар се избират по случаен начин измежду 4 човека: Ана, Борис, Васил и Георги. По колко различни начини може да се избере ръководството?

24

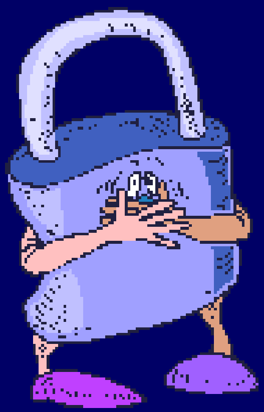
- Ръководството – председател, касиер и секретар се избират по случаен начин измежду 4 човека: Ана, Борис, Васил и Георги. Ана не иска да бъде председател . По колко различни начини може да се избере ръководството?

18

Примери

За да се отвори един катинар е необходимо да се въведе код от 4 цифри. Колко различни кода са възможни?

10 000



Ако кодът на катинар се състои от 4 различни цифри, то колко кода са възможни?

5 040

А ако кодът се състои от буква (25 латински букви), последвана от 3 различни цифри, то колко кода са възможни?

25(10)(9)(8)

Колко различни пароли, които имат повече от 3 , но не повече от 6 символа са възможни, ако се използват само цифрите 0 и 1?

с 4 символа – 2^4

с 5 символа – 2^5

с 6 символа – 2^6

Общо: 2^4+2^5+2^6=2^4(1+2+4)=112

Колко различни пароли, които имат повече от 3 , но не повече от 6 символа са възможни, ако се използват само цифрите 0, 1, 2, 3, 4, 5?

с 4 символа – 6^4

с 5 символа – 6^5

с 6 символа – 6^6

Общо:
6^4+6^5+6^6=6^4(1+6+36)=55728

Комбинации

не- наредено множество от елементи

без повторение

$$C_k^{\,n} = \binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)\dots(n-k+1)}{k!}$$

Пример

По колко различни начина могат да се изберат трима плувци от 10-членен отбор?

$$C_3^{\,10} = \frac{10(9)(8)}{3!} = \frac{720}{6} = 120$$

Пример (избор на тим)

Измежду 5 мъже и 7 жени трябва да се изберат петима, за да работят върху проект.

Колко различни 5-членни групи могат да се изберат?

$$C_{12}^{\,5} = \frac{12!}{5!(12-5)!} = \frac{12(11)(10)(9)(8)}{1(2)(3)(4)(5)} = 792$$

Колко различни 5-членни групи могат да се изберат, ако трима са мъже, а останалите две са жени?

$$C_5^{\,3} = \frac{5(4)(3)}{3!} = 10$$

$$C_7^{\,2} = \frac{7(6)}{2!} = 21$$

210

мъже

жени

Ако двама души настояват или да работят заедно или да не са в групата, то **колко различни 5-членни групи могат да се изберат**?

1 начин за двамата

$$\frac{10(9)(8)}{3!} = 120$$

$$+ \quad C_{12-2}^{\,5} = \frac{10(9)(8)(7)(6)}{5!} = 252$$

- Колко различни множества от n елемента съществуват (множествата могат да се състоят от по един, два, или повече елемента)?

От един елемент – 1 множество
От два елемент – n(n-1)/2 множества
От три елемент – n(n-1)(n-2)/(1*2*3) множества и т.н.

Общо:

$$C_n^1 + C_n^2 + C_n^3 + \ldots + C_n^n = (1+1)^n = 2^n$$

Опит

Пет карти са избрани случайно от колода карти (52 карти)

А/ Каква е вероятността точно една от избраните карти да е червена?

Колко са всички възможни изходи- начини за избор на 5 карти измежду 52 ?

$$C_5^{52} = \frac{52!}{5!(52-5)!} = \frac{52(51)(50)(49)(48)}{1(2)(3)(4)(5)} = 2598960$$

Колко са всички възможни изходи, при които една от картите е червена, а другите 4 са черни?

$$26(C_4^{26}) = 26\frac{26(25)(24)(23)}{1(2)(3)(4)} = 388700$$

$$P = \frac{388700}{2598960} \approx 0,15$$

Продължение

Б/ Каква е вероятността точно една от избраните карти да е купа?

Колко са всички възможни изходи, при които една от картите е купа, а другите 4 не са купи?

$$13(C_4^{39}) = 13\frac{39(38)(37)(36)}{1(2)(3)(4)} = 1069263$$

$$P = \frac{1069263}{2598960} \approx 0{,}41$$

В/ Каква е вероятността точно една от избраните карти да е поп?

Колко са всички възможни изходи, при които една от картите е поп, а другите 4 не са поп?

$$4(C_4^{48}) = 4\frac{48(47)(46)(45)}{1(2)(3)(4)} = 778320$$

$$P = \frac{778320}{2598960} \approx 0{,}3$$

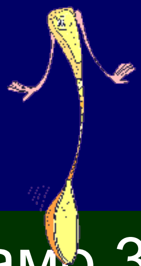
В/ Каква е вероятността точно три от избраните карти да са поп?

Колко са всички възможни изходи, при които три от картите са поп, а другите 2 не са поп?

$$\left(C_3^{4}\right)(C_2^{48}) = \frac{4(3)(2)}{1(2)(3)}\,\frac{48(47)}{1(2)} = 4512$$

$$P = \frac{4512}{2598960} \approx 0{,}0017$$

Предизвикателни задачи


задача 1

Студентка има 5 вилици, от които само 3 харесва, 4 ножа и 6 лъжици. Каква е вероятността, ако избира комплект по случаен начин, да избере вилица, която й харесва? n= 5*4*6=120 k= 3*4*6=72 P=72/120=3/5

задача 2

Асансьор със 7 пътника тръгва от партера и спира на всеки от 9-те етажа на сграда. Каква е вероятността поне двама от пътниците да слязат на един и същ етаж? Има 9 възможности за всеки пътник: $n = 9^7$

A= *поне двама слизат на един и същ етаж*

$\overline{A}$ = *всички слизат на различни етажи*

За 1-вия пътник има 9 възм. ,
за 2-рия има 8 възм. и т.н

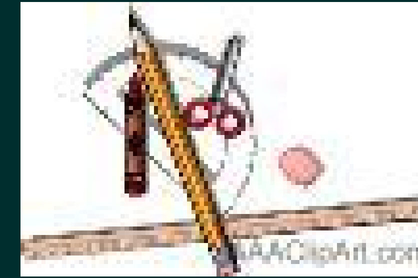
$k = 9.8.7....3$

$P(\overline{A}) = k / n = 0,038$

P(A)=1-0,038=0,962

Благодаря за вниманието!

Случайни величини



Определение

Нека S е множеството от всички елементарни събития.

Случайна величина е числова функция, дефинирана върху множеството S,

т.е. тя съпоствя на всеки елементарен изход реално число

Примери:

Опит: Хвърляне на монета един път

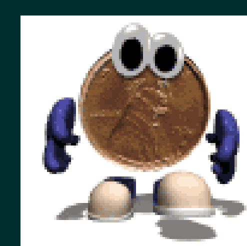
S= {Л, Г}

X={брой лица}

0	1

Y={брой гербове}

1	0

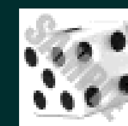


Опит: хвърляне на зарче един път

S= {1, 2, 3, 4, 5, 6}

1, 2, 3, 4, 5, 6

X={брой на точките на лицевата страна на зара}



S= {1, 2, 3, 4, 5, 6}

1 0

Y={брой паднали се лицеви страни с точно една точка върху тях}

Опит: Случайно избрано топче от кутия с 5 червени и 2 сини

X={брой сини топчета измежду избраните}

Стойности : 1, 0

Опит: Време на събуждане точно определена сутрин

Стойности :безброй много

Пример: Определете случайна величина

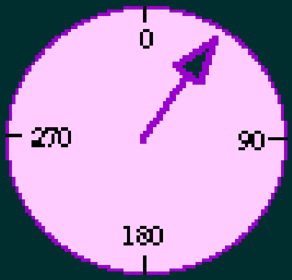
1. монета се подхвърля два пъти

X={ брой паднали се лица}

2. Монета се подхвърля, при което тя се завърта.

X={време от момента на подхвърляне до момента на покой }

3. Кръг, разделен на 4 части е завъртян по посока на часовниковата стрелка



X={квадранта, в който показва стрелката, след като кръгът спира}

4. Студент се явява на изпит.

X={оценката}

5. Студент отива на среща с приятелка.

X={времето, когато приятелката пристига}

Теорема . ***Линейна комбинация, произведение, минимум, максимум и функция на сл.в. е сл.в.***

Видове случайни величини

Дискретни

Случайна величина, която приема само краен брой или изброимо стойности

Дискретната случайна величина обикновено е случайна величина, чийто стойности са резултат от броене.

Непрекъснати

Случайна величина, чийто стойности са всички числа от даден интервал (или интервали), които могат да са крайни или безкрайни

Непрекъсната случайна величина обикновено е случайна величина, чийто стойности са резултат от измервания.

Примери

Дискретни

- брой студенти в клас.
- брой деца в семейство.
- Брой жилищни заеми, дадени от банка миналата седмица.

Непрекъснати

- разстоянието, което студентите изминават от дома им до входа на университета .
- Времето, което се пътува с автобус от Панаира до университета.

- Времетраенето на един изпит.
- Дължината на телефонен разговор с приятелка.

пример: Дефинирайте случайна величина, свързана с опита и определете нейния вид и множеството от възможни стойности

1. Зар е подхвърлен и броят на точките е записан.

X={ брой точки}

2. Зар е подхвърлен и времето докато застане в покой е засечено.

X={време }

4. Измерена е температурата сутринта в 7 часа.

X={температура}

5. В партида от 1 000 чифта пантофи, произведени в дадена фирма, са отчетени броят на дефектните.

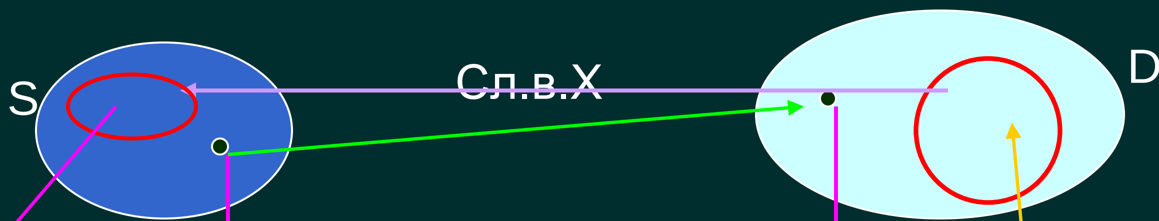
X={брой дефектни}

Непрекъснати

Дискретни

Функция на разпределение

Нека X е сл.в., дефинирана в пространството от ел.изходи S и със стойности в множеството от числа D (крайно, изброимо или неизброимо)



ел. събитие

число

Събитие {X < a}

вероятност

Нека *a* е произволно число => M={всички числа от D, които са ≤ a}

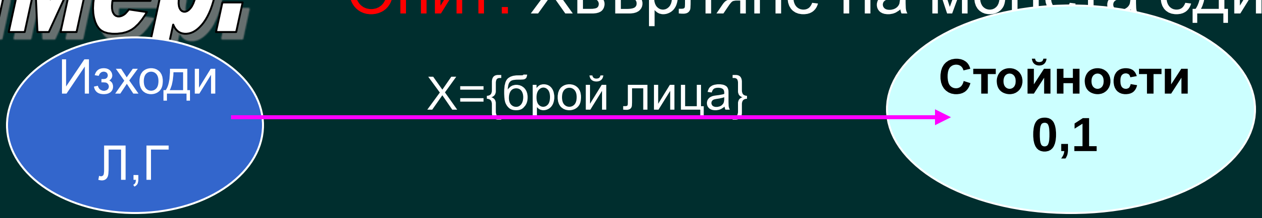
F(a)=P(X ≤ a) {
дефиниционно множество: всяко реално a
множество от стойности: [0.1]

F(a)=P(X≤a) => вероятност

Дефиниция: Ф.р. на една сл. в. X е **F(x)= P({ω : X(ω) ≤ x) = P(X ≤ x)** за всяко реално x

Пример:

Опит: Хвърляне на монета един път



F(-1)=P(X ≤ -1)= P(невъзможното)=0 F(-3)=P(X ≤ -3)= P(невъзможното)=0

Няма изход на който да се съпоставя число ≤ -1

Ако x<0 , то F(x)=P(X ≤ x)= P(невъзможното)=0

F(0,3)=P(X ≤ 0,3)= P(Г)=0,5 F(0,8)=P(X ≤ 0,8)= P(Г)=0,5

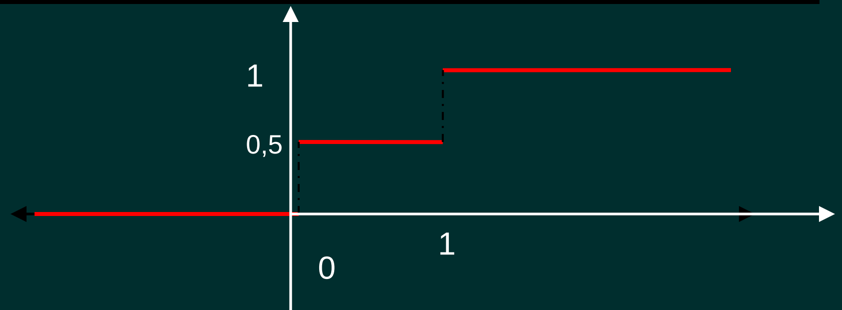
На Г се съпоставя числото 0, което е < 0,3

Ако 0<x<1 , то F(x)=P(X ≤ x)= P(Г)=0,5

F(2)=P(X ≤ 2)= P(Г,Л)=1 F(7)=P(X ≤ 7)= P(Г,Л)=1

На Г се съпоставя числото 0, на Л се съпоставя числото 1, и двете са < 2

Ако x>1 , то F(x)=P(X ≤ x)= P(Л, Г)=1



$$F(x)=\begin{cases} 0 & \text{Ако } x<0 \\ 0,5 & \text{Ако } 0 \le x < 1 \\ 1 & \text{Ако } x \ge 1 \end{cases}$$

Пример:

Опит: хвърляне на зарче един път

X={брой на точките на лицевата страна на зара}

S= {1; 2; 3; 4; 5; 6} → 1; 2; 3; 4; 5; 6

$$F(x)=\begin{cases} 0 & \text{Ако } x < 1 \\ 1/6 & \text{Ако } 1 \le x < 2 \\ 2/6 & \text{Ако } 2 \le x < 3 \\ 3/6 & \text{Ако } 3 \le x < 4 \\ 4/6 & \text{Ако } 4 \le x < 5 \\ 5/6 & \text{Ако } 5 \le x < 6 \\ 1 & \text{Ако } x \ge 6 \end{cases}$$

F(2,78)=P(X<2,78)= P(**1,2**)=2/6

На 1 се съпоставя числото 1, на 2 се съпоставя числото 2, и двете са < 2,78

Свойства на ф.р.

Дефиниция: Ф.р. на една сл. в. X е F(x)=**P(X ≤ x)** за всяко реално x

Свойство 1.

Дефиниционно множество : множеството на реалните числа

Свойство 2.

Множество от стойности : [0,1]

Свойство 3.

ф. р . е непрекъсната отдясно, т. е .

$$\lim_{\substack{s\to x\\ s>x}} F(s) = F(x+0) = F(x)$$

Свойство 4.

$$\lim_{x\to-\infty} F(x) = 0 \qquad \lim_{x\to+\infty} F(x) = 1$$

Понеже P(X < - ∞)=0 и P(X < + ∞)=1

F (- ∞) = 0

F (+ ∞) = 1

Свойство 5. $P(X > a) = 1 - F(a)$

От дефиницията $P(X>a)=1-P(X\leq a)=1-F(a)$

Свойство 6. $P(a < X \leq в) = F(b)-F(a)$

Свойство 7.

$$F(x)=F(x+)$$

Функцията на разпределение F(x) е непрекъсната **отдясно**

Свойство 8.

Функцията на разпределение F(x) е НЕНАМАЛЯВАЩА

$\Rightarrow F(a)=P(X \leq a) \geq P(X \leq b)= F(b)$

Нека a>b

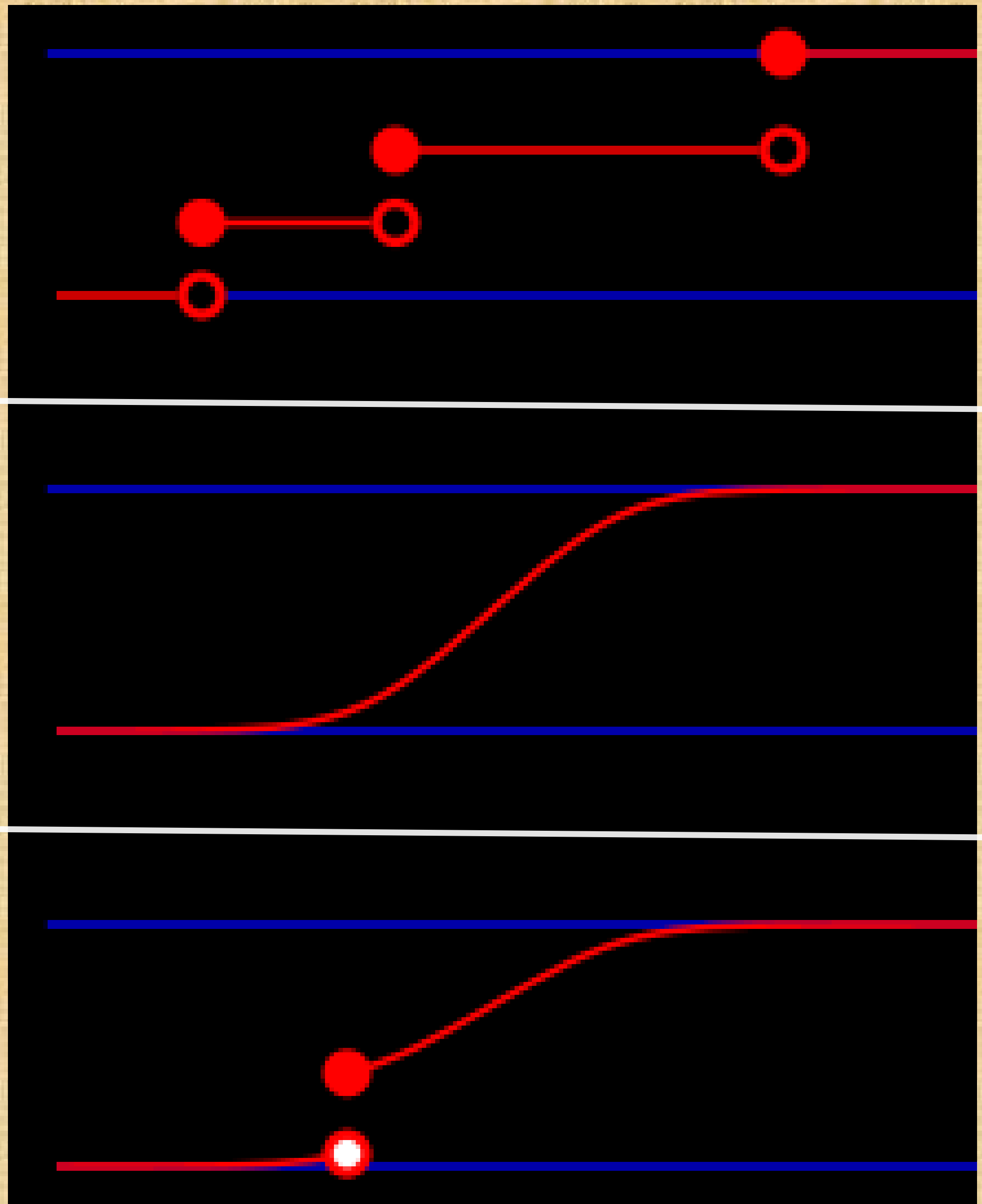
ВАЖНО!!!

Нека функцията на разпределение F(x) е константа в даден интервал (а,в)

$$P(a<X \leq b)=F(b)-F(a)=0$$

Случайната величина X не приема стойности в интервала (а,в)

Графика на функция на разпределение случайна величина

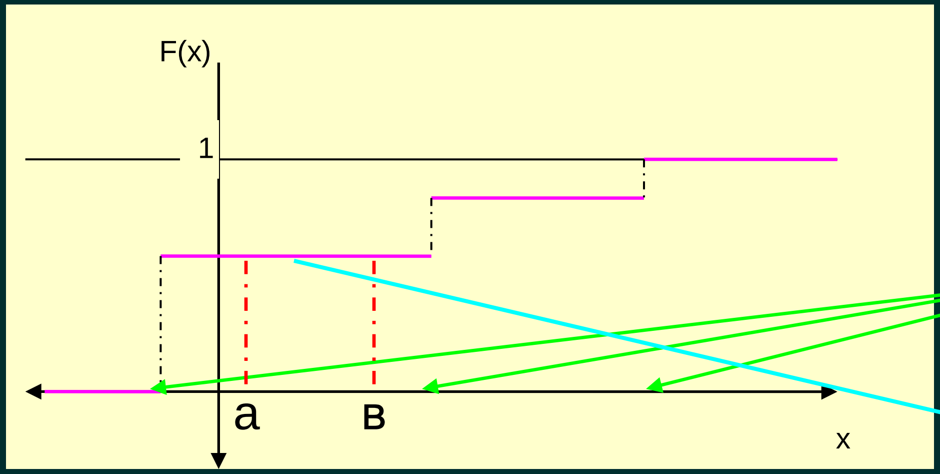


Дискретна
случайна
величина

непрекъсната
случайна
величина

случайна
величина
Смесен тип

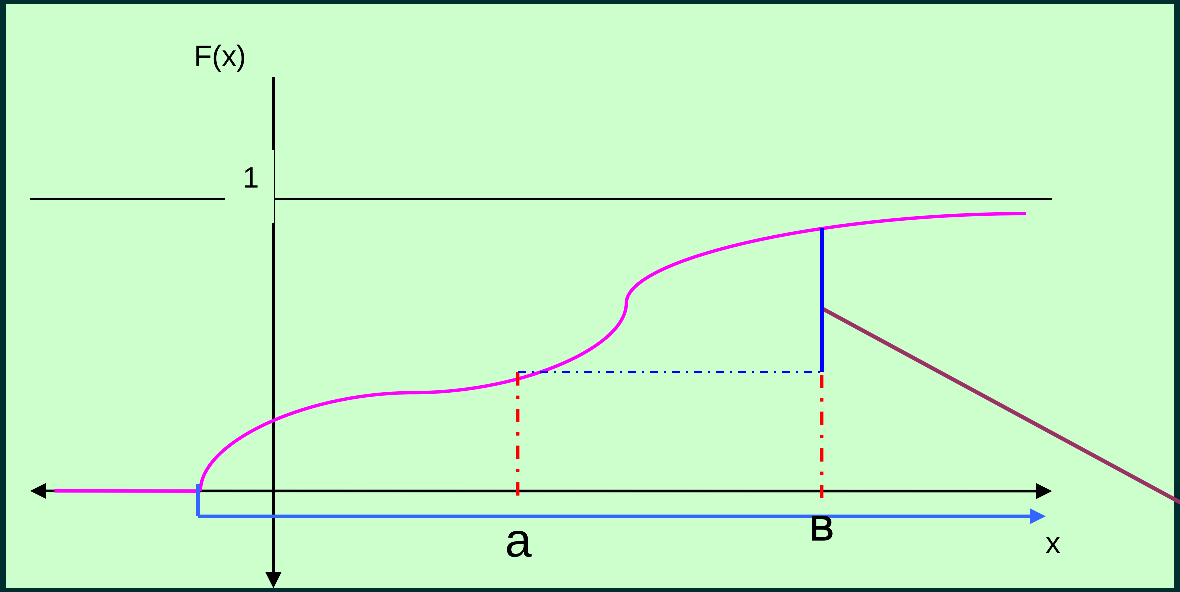
Графика на ф.р.



Дискретна сл.в.

Стойности на сл.в.

P(а < X≤ в)= F (в) - F (а) = 0



непрекъсната сл.в.

Стойности на сл.в.

P(а < X≤ в)= F (в) - F (а)

Една функция може да е ф.р. на случайна величина, ако ПРИТЕЖАВА свойства 1-8

Задача Може ли функцията

$$F(x)=\begin{cases}0 & при \quad x<0\\ e^{-2x} & при \quad 0\le x<1\\ 1 & при \quad x\ge 1\end{cases}$$

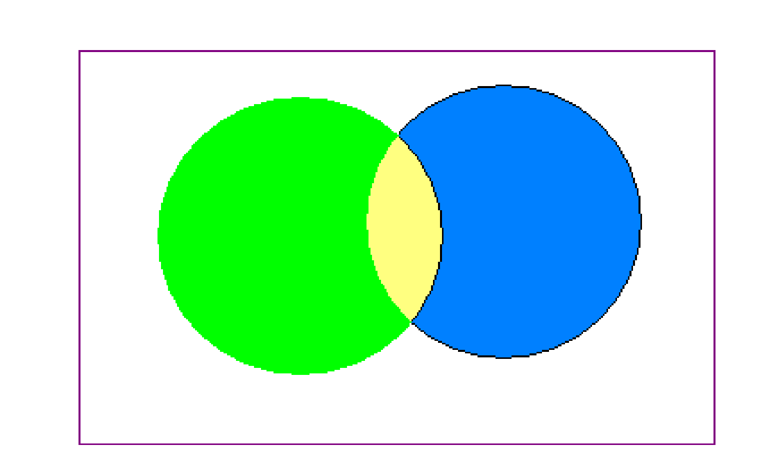
да бъде функция на разпределение на случайна величина?

НЕ може да е ф.р, не е изпълнено свойство 8, не е НЕНАМАЛЯВАЩА

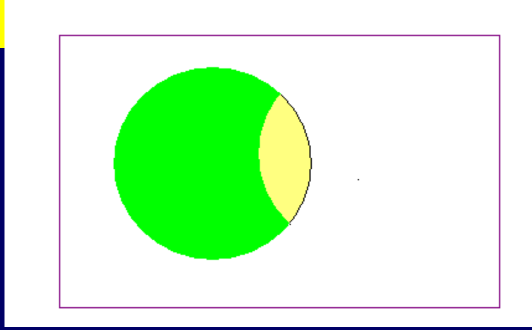
Формула за събиране на вероятности

$P(A \text{ или} B) = P(A) + P(B) - P(A \text{ и } B)$

P(*A* <u>или</u> *B*) =

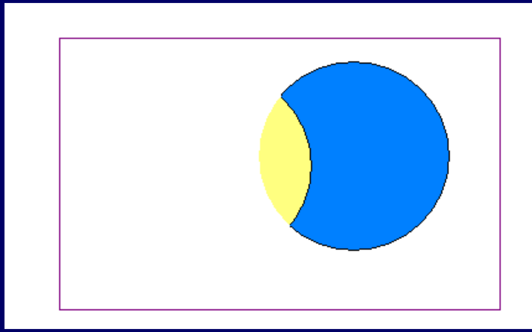


P(*A*)



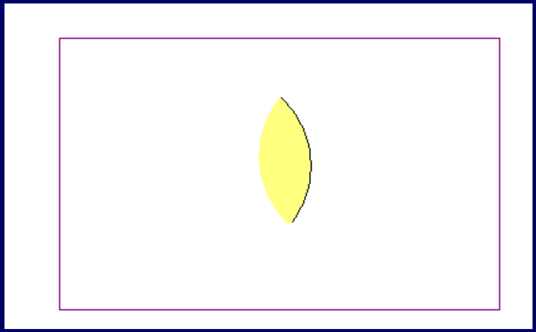
+

P(*B*)



–

P(*A* <u>и</u> *B*)



При три събития

Разглеждаме събитията A, B и C

P(A или B или C) = ????

$$P(A \cup B \cup C) = P(A) + P(B \cup C) - P(A \cap [B \cup C]).$$

$$P(B \cup C) = P(B) + P(C) - P(BC).$$

$$A \cap [B \cup C] = (A \cap B) \cup (A \cap C)$$

$$\begin{aligned} P(A \cap [B \cup C]) &= P([A \cap B] \cup [A \cap C]) \\ &= P(AB \cup AC) \\ &= P(AB) + P(AC) - P(ABC). \end{aligned}$$

$$\begin{aligned} P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ - P(AB) - P(BC) - P(AC) \\ + P(ABC) \end{aligned}$$

Обобщение

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{\kappa=1}^{n} P(A_\kappa) - \sum_{\substack{\kappa, j=1 \\ k<j}}^{n} P(A_k A_j) + \sum_{\substack{\kappa, j, i=1 \\ k<j<i}}^{n} P(A_k A_j A_i) - \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n)$$

С. Христова

Пример: събиране на вероятности

Известно е, че 25% от жителите на един град четат вестник “Новинар”, 20% четат “Дневник”, 13% четат “За вас”, 10% четат и “Новинар” и “Дневник”, 8% четат и “Новинар” и “За вас”, 5% четат и “Дневник” и “За вас”, и 4% четат трите.

Ако един жител е избран случайно, то каква е вероятността той/тя да не четат вестник изобщо?

A= чете “Новинар”, B=чете “Дневник”, C=чете “За вас”, E=не чете

$\overline{E}=AUBUC$

$P(\overline{E})=P(AUBUC)=P(A)+P(B)+P(C)$
$-P(AB)-P(AC)-P(BC)$
$+P(ABC)=$

$0,25+0,2+0,13-0,1-0,08-0,05+0,04=0,39$

$P(A)=1-0,39=0,61$

Примери:

Карта е избрана от колода от 52 карти.
A= {картата е червена} B= { картата е пика}
C= { картата е поп}

Намерете вероятността избраната карта да е червена или пика.

P(A или B)= P(A)+P(B)-P(A и B)=.5+.25-0=.75

A и B= невъзможно

Намерете вероятността избраната карта да е червена или поп.

P(A или C)=P(A)+P(C)-P(A иC)=1/2+4/52-2/52=(26+4-2)/52

A и C= {червен поп}

Геометрична вероятност

За безкрайномерни пространства

Нека да може да се установи **взаимно еднозначно съответствие** между S и геометричен обект върху права, в равнина или в пространството.

=>на всяко събитие съответства подмножество на този геометричен обект

$$P(A)=\frac{\mu(A)}{\mu(S)}$$

μ е мярка на множество(геометричен обект)

Опит: стрелба по кръгова мишена.

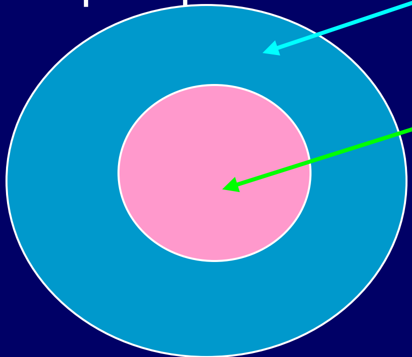
Върху права - μ е дължина на отсечка

S={всички точки от кръга}

В равнината - μ е лице на фугура

A={попадението е по-близо до центъра отколкото до контура}

В пространството - μ е обем на тяло



$$P(A)=\frac{\text{лице на малък кръг}}{\text{лице на голям кръг}}=\frac{\pi r^2}{\pi(2r)^2}=\frac{1}{4}$$

Вероятност ли е???

$$P(A) = \frac{\mu(A)}{\mu(S)}$$

Проверка:

1. $P(A) \geq 0$

ДА

2. $P(S)=1$

ДА

3. P е (безкрайно) адитивна, т.е. ако $A_1, A_2, \ldots$ е крайна или безкрайна редица от несъвместими събития, то

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

ДА

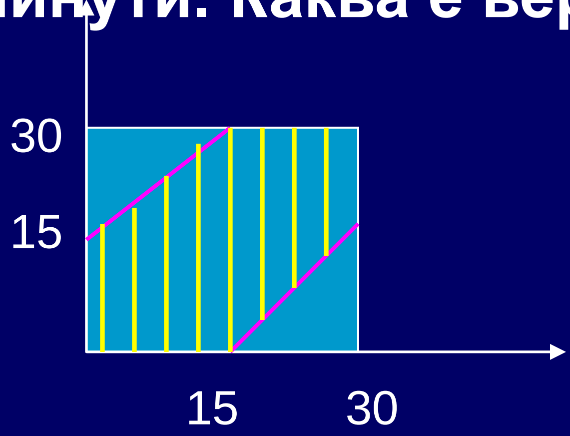
ДА, това е вероятност

Ако A и B са несъвместими, то множествата, които им съответстват нямат общи точки => от геометрията =>

мярката на обеденението на двете множества е сума от мерките на двете множества .

Задача за срещата

Иванчо и Марийка си определят среща пред Ректората между 10:00 и 10:30 като всеки чака не повече от 15 минути. Каква е вероятността да се срещнат?



A=двамата ще се срещнат

x-време на пристигане на Марийка

y- време на пристигане на Иванчо

Ел. събитие : (x,y)

S→квадрат

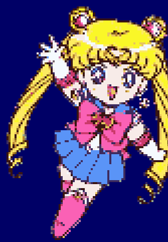
Ще се срещнат ако |x-y|≤15

A→защрихована част

Лицето на квадрата= 30(30)=900

Лицето на защрихованата част= 900-(15)15=675

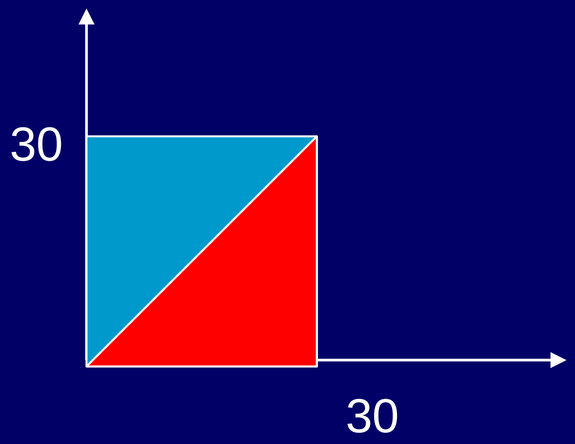
P(A)=675/900=0,75



С. Христова

Каква е вероятността Марийка да пристигне след Иванчо?

Ако $y \leq x$



x-време на пристигане на Марийка

y- време на пристигане на Иванчо

P(A)=450/900=0,5

Условна вероятност

Да разгледаме **вероятностен опит** с **пространство от елементарните изходи** *S*

Нека В е събитие, от S (различно от невъзможното)

Каква е вероятността да настъпи събитието *А*, ***ако е известно***, че събитието *В* е настъпило ?

Означение: P(A|B)

Пример

Карта е изтеглена от колода от 52 карти. Ако е известно, че картата е червена, то каква е вероятността тя да е поп?

$$P(A \cap B) = \frac{2}{52} = \frac{1}{26}$$

$$P(A|B) = \frac{2}{26}$$

$$P(B) = \frac{26}{52} = \frac{1}{2}$$

Определение

Нека *A* и *B* са две събития от едно и също пространство S, и $P(B) > 0$. ***Условна вероятност*** на *A* при условие *B се* дефинира с равенството

$$P(A|B)=\frac{P(A\ и\ \ B)}{P(B)}=\frac{P(A\cap B)}{P(B)}$$

Пример

В едно семейство има две деца. Ако поне едно от децата е момиче, то каква е вероятността и двете да са момичета?

P(| поне едно)=P( и поне едно)/P(поне едно)

= P()/ P(поне едно)=0,25/0,75= 1/3

С.Христова

Вероятност ли е???

$$P(A \mid B) = \frac{P(A \text{ и } B)}{P(B)}$$

Проверка:

ДА, това е вероятност

1. $P(A|B) \geq 0$

ДА

2. $P(S|B)=1$

ДА

ДА

3. P е (безкрайно) адитивна, т.е. ако $A_1, A_2, \ldots$ е крайна или безкрайна редица от несъвместими събития, то

$$P\big((A_1 \cup A_2 \cup \ldots) \mid B\big) = P(A_1 \mid B) + P(A_2 \mid B) + \ldots$$

$$P\big((A_1 \cup A_2 \cup \ldots) u B\big) = P(A_1\, u\, B) + P(A_2\, u\, B) + \ldots$$

$$P\big((A_1 \cup A_2 \cup \ldots) \mid B\big) = \frac{P(A_1\, u\, B) + P(A_2\, u\, B) + \ldots}{P(B)}$$

$$= \frac{P(A_1\, u\, B)}{P(B)} + \frac{P(A_2\, u\, B)}{P(B)} + \ldots$$

Пример 1:

В курса по информатика от 80 студента само 40 са изкарали над 10 точки и на двете контролни, а 60 са изкарали над 10 точки на втората контролна. Каква е вероятността, случайно избран студент от този курс, който е изкарал на втората контролна над 10 точки, да е изкарал и на първата над 10 точки?

A=студентът е изкарал над 10 точки на първата контролна

B=студентът е изкарал над 10 точки на втората контролна

P(B)=60/80=0,75

P(A и B)=40/80=0,5

P(A|B)=?

P(A|B)= 0, 5/0,75=0,66=2/3

Независими събития

Нека *A* и *B* са събития, свързани с един и същ опит.

A и B са **независими** , ако **P(A|B)=P(A)**

Пример

Опит: Карта се избира случайно от колода от 52 карти.

A={избраната карта е червена}

B={избраната карта е дама}

Независими ли са A и B?

P(A)=0,5

P(A|B)=2/4=0,5

A и B са независими

Пример

Опит: две карти са избрани една по една от колода карти.

с връщане

Каква е вероятността ***и*** двете карти да са поп?

A= {първата карта е поп} B= {втората карта е поп}

P(A и B)=P(B|A)P(A)

P(B|A)= 4/52

P(A)= 4/52

P(A иB)=(4/52)(4/52)=0,0059

A и B са независими

без връщане

P(A и B)=P(B|A)P(A)

P(B|A)= 3/51

P(A)= 4/52

P(A и B)=(3/51)(4/52)=0,0045

A и B са зависими

Умножение на вероятности

Каква е вероятността да настъпят и двете събития?

Нека A и B са събития

P(A и B)=P(A|B) P(B)

Пример

В шкаф има смесени разноцветни чорапи- 6 черни и 4 бели. Вадим един по един със затворени очи. Каква е вероятността да извадим един черен чифт?

A={черен чорап при първото вадене}
B ={черен чорап при второто вадене}

P(A)=6/10   P(B|A)=5/9

P(A и B) =P(A)P(B|A)=(6/10)(5/9)=1/3

A и B са зависими

Втори начин:

P(AиB)= $\frac{6.5}{10.9}$

Отново независимост и произведение на вероятности

Нека *A* и *B* са събития, свързани с един и същ опит.
A и B са **независими** , ако **P(A|B)=P(A)**

Знаем, че

$$P(A \mid B) = \frac{P(A \text{ и } B)}{P(B)}$$

Нека *A* и *B* са събития, свързани с един и същ опит.
A и B са **независими** , ако **P(A и B)=P(A) P(B)**

Събитията А и В са с положителни вероятности.

Нека А и В са несъвместими $\xrightarrow{?????}$ А и В са независими

<u>Доказателство:</u> А. В – невъзможното => Р(А.В)=0

За да са независими трябва Р(А.В)=Р(А) Р(В) => 0= Р(А) Р(В) => поне една от вероятностите е 0 =>противоречие

Нека А и В са независими $\xrightarrow{?????}$ А и В са несъвместими

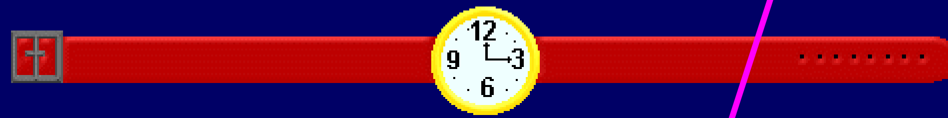
<u>Доказателство:</u> Р(А.В)=Р(А) Р(В)

За да са несъвместими трябва А.В=невъзможното и следователно 0= Р(А.В) =>Р(А)Р(В)=0 => поне една от вероятностите е 0 =>противоречие

Обобщение

$P(A \text{ и } B \text{ и } C)=P(A)P(B|A)\ P(C|(A \text{ и } B))$

пример



Вечерта три съквартирантки си оставят часовниците на масата и на сутринта, в бързината, всяка взема часовник по случаен начин.

A={първата взема своя часовник} B ={втората взема своя часовник}
C ={третата взема своя часовник}

Каква е вероятността първата да вземе своя часовник?

$P(A)= ???$

$P(A)=(1*2*1)/(3*2*1)=1/3$

Каква е вероятността втората да вземе своя часовник?

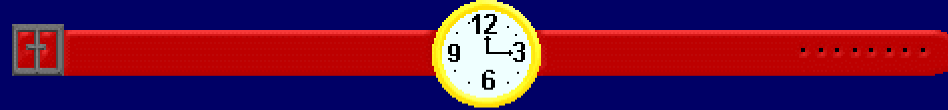
$P(B)= ???$

Каква е вероятността всяка да вземе своя часовник?

P(всяка да вземе своя)= $P(A \text{ и } B \text{ и } C)=???$

P(всяка да вземе своя)= $P(A \text{ и } B \text{ и } C)=P(A)P(B|A)\ P(C|(A \text{ и } B))=(1/3)(1/2)1=1/6$

Продължение



Каква е вероятността поне една да не вземе своя часовник?

P(поне една **не** взема)= 1- P(всяка взема своя)=???

P(поне една **не** взема)= 1- P(всяка взема своя) =1-1/6=5/6

Каква е вероятността поне една **да** вземе своя часовник?

P(поне една **да** вземе своя)=P(A или B или C)=????

P(поне една да вземе своя)=P(A или B или C)=
P(A)+P(B)+P(C) - {P(A иB)+P(AиC)+P(BиC)}+P(AиBиC)=
(1/3)+ (1/3)+ (1/3) - {(1*1*1)/(3*2*1)+ 1/6+1/6 } +1/6=2/3

Каква е вероятността нито една да не вземе своя часовник?

P(нито една не взема)=1- P(поне една да вземе своя)= ???

P(нито една не взема)=1-2/3=1/3

Обобщение на независимост

Три събития са **независими в съвкупност**, ако

$$P(AB) = P(A)P(B)$$

$$P(AC) = P(A)P(C)$$

и

$$P(BC) = P(B)P(C)$$

$$P(ABC) = P(A)P(B)P(C).$$

В общия случай, N събития са **независими в съвкупност**, ако за всяка комбинация $1 \le i \le j \le k \le \ \ldots \le N$ е изпълнено

$$\begin{gathered} P(A_i A_j) = P(A_i)P(A_j) \\ P(A_i A_j A_k) = P(A_i)P(A_j)P(A_k) \\ \vdots \\ P(A_1 A_2 \cdots A_N) = P(A_1)P(A_2)\cdots P(A_N). \end{gathered}$$

Пример

Три от стените на правилен тетраедър са боядисани съответно в бяло, в зелено и в червено, а четвъртата е с триколъра на знамето ни. Тетраедърът е подхвърлен на пода.

A= стената на която пада тетраедъра съдържа **бял** цвят

B= стената на която пада тетраедъра съдържа **червен** цвят

C= стената на която пада тетраедъра съдържа **зелен** цвят

Независими ли са A,B и C?

P(A)=P(B)=P(C)=2/4=1/2

P(A и B)=1/4= P(A)P(B)

A и B са независими

P(C и B)= 1/4= P(C)P(B)

C и B са независими

P(A и C)= 1/4= P(A)P(C)

A и C са независими

P(AиBиC)= 1/4≠1/8=P(A)P(B)P(C)

A,B,C са независими по двойки

A,B и C са зависими

Тестване на хипотези за два параметъра

Сравняване на две средни стойности

Зависими извадки

Зависими извадки - когато могат да комбинират по двойки (в някакъв смисъл)

Пример:

При измерване ефективност на нова диета, се претеглят едни и същи хора, подложени на диетата, преди и след прилагането й.

Проверка на хипотези за **две зависими** извадки

Комбинираме двете извадки в една= разликата от двете

- Използваме следната статистика

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

където $\bar{d}$ е средната стойност на разликата , s_d е стандартното отклонение на разликата, и n е броя на двойките (разликите)

Пример

- За да се изучат дневните тарифи за коли под наем на леки автомобили на компаниите Hertz и Avis в САЩ е направена случайна извадка от 8 големи града и е записана информацията за тарифата в следната таблица. При ниво на значимост 0,05 може ли да се твърди, че има разлика в тарифата на двете компании?

Пример

Град	Hertz (цена в $)	Avis (цена в $)
Атланта	42	40
Чикаго	56	52
Кливланд	45	43
Денвър	48	48
Хонолулу	37	32
Канзас	45	48
Маями	41	39
Сеатъл	46	50

Решение:

Образуваме нова извадка от разликите

Град	Hertz	Avis	*d*	*(d-средно)²*
Атланта	42	40	2	1
Чикаго	56	52	4	9
Кливланд	45	43	2	1
Денвър	48	48	0	1
Хонолулу	37	32	5	16
Канзас	45	48	-3	16
Маями	41	39	2	1
Сиатъл	46	50	-4	25
			Сума= 8	сума=70

Средно =d=1 дисп.=s^2_d=10

1. Нулева и алтернативна хипотеза

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

или $H_0 : \mu_d = 0$ $H_1 : \mu_d \neq 0$

2. Ниво на значимост

$\alpha = 0,05$

3. Статистика, извадково разпределение

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

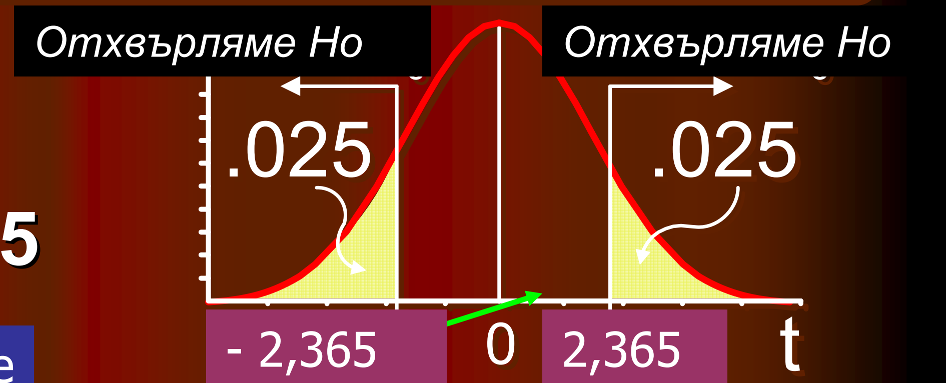
4. Критична област

Критична област

(- ∞, -2,365) и (2,365 ; ∞)

$\alpha = 0{,}05$ **$\alpha/2 = 0{,}025$**

t (7) разпределение



5. Извод

t=0,894 не е в критичната област, затова не отхвърляме H_0

6. Интерпретация на извода

Няма достатъчно основание на считаме, че има разлика в цените на Hertz и Avis.

Независими извадки (голям обем)

Предположения

1. Двете извадки са незавизими

2. Обемите на двете извадки са големи

$n_1 > 30 \qquad n_2 > 30$

Алтернативи

отхвърляме H_0 ако:

Алтернативи	отхвърляме H_0 ако:
$H_1 : \mu_1 - \mu_2 > D_0$	$z > z_\alpha$
$H_1 : \mu_1 - \mu_2 < D_0$	$z < -z_\alpha$
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\lvert z \rvert > z_{\alpha/2}$, т.е. $z > z_{\alpha/2}$ или $z < -z_{\alpha/2}$

Статистика

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Използваме s_1 и s_2 ако σ_1 и σ_2 не са известни

Пример:

Твърди се, че средната възраст на студентите от хуманитарните специалности е различна от средната възраст на студентите от техническите специалности. За да се провери твърдението, са направени две случайни извадки от по 50 студенти от хуманитарни специалности и 50 студенти от технически специалности и е записана възрастта им. От данните е получено, че средната възраст на първата група е 21 година, а средната възраст на втората е 20 години със стандартни отклонения съответно 4 и 2,5 години. При ниво на значимост 0,04, тествайте твърдението за различие в средната възраст на двете групи студенти .

Интерпретация на данните:

$$n_1 = 50 \qquad n_2 = 50$$

$$\bar{x}_1 = 21 \qquad \bar{x}_2 = 20$$

$$s_1 = 4 \qquad s_2 = 2{,}5$$

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

статистика

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}} =$$

$$\frac{21 - 20}{\sqrt{\frac{4^2}{50} + \frac{2,5^2}{50}}} = 1,499$$

Критична област : z<-2,05 или z>2,05

Z=1.499 не е в критичната област

Не отхвърляме H_0

Няма достаъчно основание да се твърди, че има съществена разлика във възрастта.

Независими извадки с малък обем

Случай1: Равни дисперсии

Предположения :

1. Двете популации са нормално разпределени
2. Обемът поне на едната от двете извадки е малък (n< 30 или m<30)
3. Двете популации имат еднакви (макар и неизвестни) дисперсии σ
4. Двете извадки, взети от тези популации са независими

Разликата на двете извадкови средни е $N(\mu_X - \mu_Y, \sigma\left(\frac{1}{n} + \frac{1}{m}\right))$

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad H_1 : \mu_1 - \mu_2 > D_0$$

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad H_1 : \mu_1 - \mu_2 < D_0$$

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad H_1 : \mu_1 - \mu_2 \neq D_0$$

Статистика: $$t = \frac{(\overline{x} - \overline{y}) - D_0}{\sqrt{s_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}}$$ където $$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{(n+m-2)}$$

- Използва се **t-разпределението с n+m-2 степени на свобода.**

Пример:

Ботаник се интересува от влиянието на определен вид тор върху растежа на стъблото на грах. Използвайки 16-дневни растения, той измерва дължината на стеблата на 11 растения, на които средното изменение на дължината е 1,03 с дисперсия 0,24. Той третира 13 растения със съответния тор в продължение на 16 дни, измерва стеблата им и намира, че средното изменение в дължините е 1,66 с дисперсия 0,35. Може ли да се твърди, че този вид тор подобрява растежа? Предполагаме една и съща популационна дисперсия.

Интерпретация на данните:

$$n = 11,\ \bar{x} = 1{,}03,\ s_X^2 = 0{,}24, \quad m = 13,\ \bar{y} = 1{,}66,\ s_Y^2 = 0{,}35$$

$H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{(n+m-2)} \rightarrow s_p^2 = \frac{(10)(0,24)+(12)(0,35)}{(11+13-2)} = 6,6$$

$$t = \frac{(\overline{x}-\overline{y}) - D_0}{\sqrt{s_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}}$$

$$t = \frac{(1,03 - 1,66) - 0}{\sqrt{0,3\left(\frac{1}{11}+\frac{1}{13}\right)}} = -0,47$$

Критична област:

Нека α=0,05

Използваме t-рапределението при степени на свобода= 22

Критична област: t<-1,7171

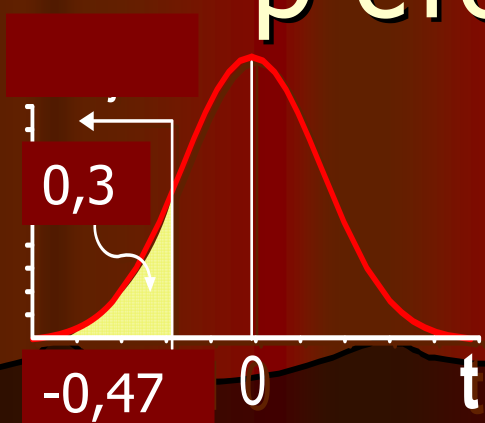
извод:

t= - 0,47 не е в критичната област, затова не отхвърляме хипотезата, т.е. Няма статистическо основание да отхвърлим твърдението, че този вид тор подобрява растежа.

р-стойност:

t= - 0,47

Лявостранен тест



р-стойността на теста =0,3

извод: 0,3>0,1, затова не отхвърляме хипотезата.

Независими извадки с малък обем

Случай 2: Различни дисперсии

Предположения :

1. Двете популации са <u>нормално разпределени</u>
2. Обемът поне на едната от двете извадки е малък (n< 30 или m<30)
3. Двете популации имат <u>различни неизвестни дисперсии</u>
4. Двете извадки, взети от тези популации са <u>независими</u>

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad H_1 : \mu_1 - \mu_2 > D_0$$

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad H_1 : \mu_1 - \mu_2 < D_0$$

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad H_1 : \mu_1 - \mu_2 \neq D_0$$

Статистика:

$$t = \frac{(\overline{x} - \overline{y}) - D_0}{\sqrt{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)}}$$

- Използва се **t-разпределението с к степени на свобода,** където к=min(n-1,m-1).

Пример:

Два града, А и Б, се разделят от една река. Местната преса публикува, че леките автомобили в град А са на повече километри от тези в град Б. За да се провери твърдението се избират по случаен начин 40 автомобила от А и се установява, че средно те са на 38 000 км със стандартно отклонение 6 000 км. Избрани са по случаен начин и 35 автомобила от Б и се намира, че средно те са на 35 000 км със стандартно отклонение 7 000 км. При ниво на значимост 0,01 може ли да се твърди, че колите в А са на повече километри, ако километрите са нормално разпределени?

Стъпка 1: Нулева и алтернативна хипотеза:

$$H_0:\ \mu_А = \mu_Б\ ; \qquad H_1:\ \mu_А > \mu_Б$$

Стъпка 2: Ниво на значимост: α=0,01

стъпка 3: Статистика:

Обемите и на двете извадки са >30, използваме *z*.

$$z = \frac{38000 - 35000}{\sqrt{\frac{(6000)^2}{40} + \frac{(7000)^2}{35}}} = 1,98$$

стъпка 4: Критична област: $z > 2,33$

Критична стойност $z_{0,01} = 2,33$

1,98 не попада в критичната област, не отхвърляме нулевата хипотеза. Няма основание да твърдим, че колите в град А са на повече километри.

Пример:

Разглеждаме предишния пример, но при избрани по 15 автомобила от всеки град.

- Стъпка 1: Нулева и алтернативна хипотеза.
- H_0: $\mu_A = \mu_Б$; H_1: $\mu_A > \mu_Б$

Стъпка 2: Ниво на значимост: α=0,01

Стъпка 3: Статистика:

Тъй като обемите и на двете извадки са <30, то е необходимо да сравним статистически популационните дисперсии, дали са равни или не, т.е да тестваме хипотезата H_0: $\sigma_A = \sigma_Б$; H_1: $\sigma_A \neq \sigma_Б$

Как да тестваме хипотеза за две дисперсии

$H_0: \sigma_А = \sigma_Б$; $H_1: \sigma_А \neq \sigma_Б$

Статистика:

$$F = \frac{s_1^2}{s_2^2}$$

F-разпределение, има две степени на свобода- в числителя и в в знаменателя

Изчисляване на p-стойността може да използвате :

davidmlane.com/hyperstat/**F_table**.html

Обратно към задачата: извадковите стандартни откл. са 7000 и 6000

$$F = \frac{7000^2}{6000^2} = \frac{49}{36} = 1{,}361$$

Степени на свобода:

в числител 35-1=34

В знаменател 40-1=39

Р-стойност = 0,49697 няма основание да отхвърлим хипотезата

Тогава използваме теста за равни дисперсии, т.е.

Използваме *t-разпределение и статистиката*

$$t = \frac{(\overline{x} - \overline{y}) - D_0}{\sqrt{s_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{(n+m-2)}$$

$$s_p^2 = \frac{(14)(6000)^2 + (14)(7000)^2}{(15+15-2)} = 42\ 500\ 000$$

$S_p = 6519$

$$t = \frac{(38000 - 35000) - 0}{6519\sqrt{\left(\frac{1}{15} + \frac{1}{15}\right)}} = 1,2605$$

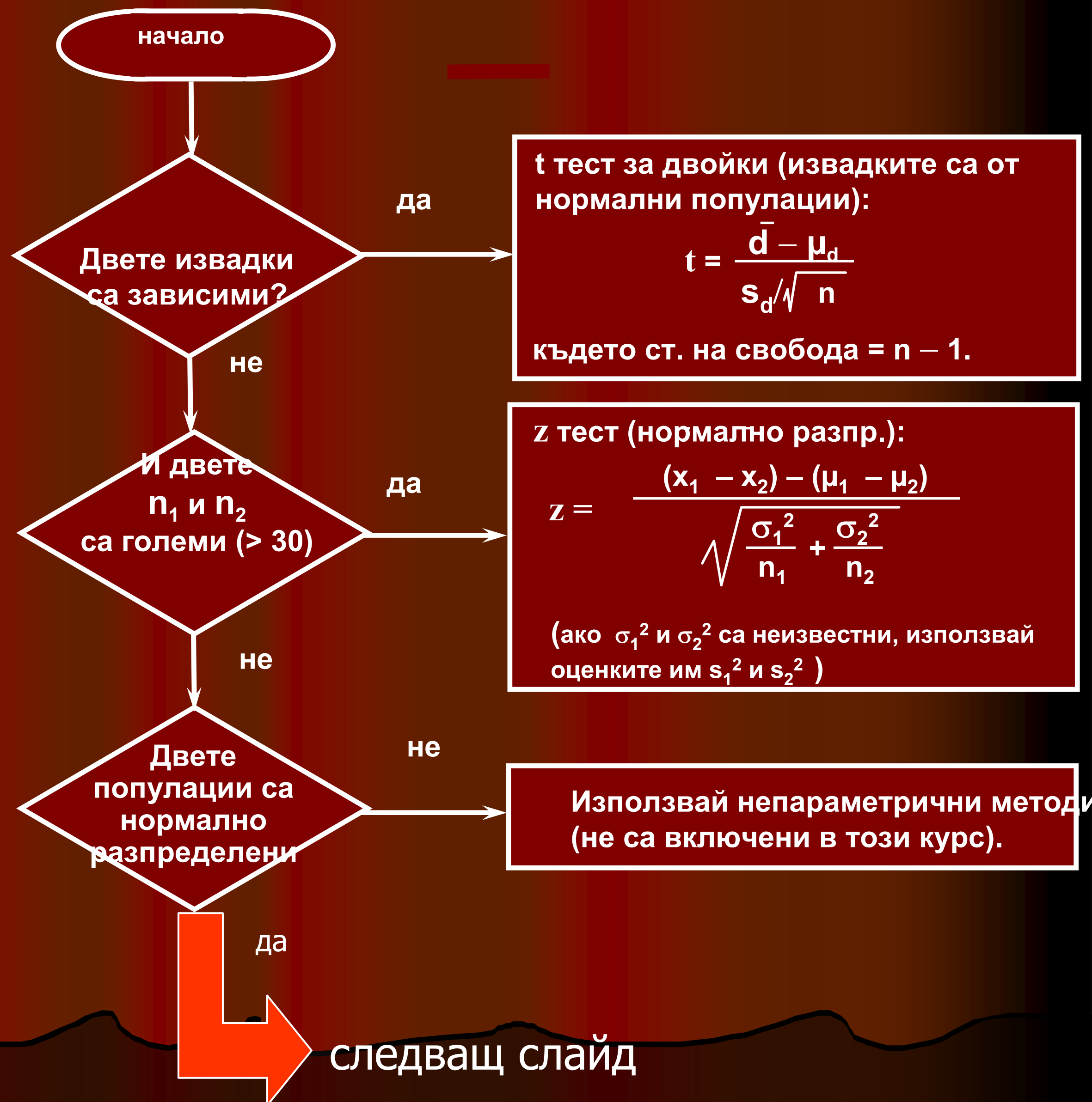
Стъпка 4: Намираме *t(28)* от таблицата и построяваме критичната област.

$$t_{0,01}(28) = 2,467$$

$$t > 2,467$$

t=1,2605 не в критичната област; не отхвърляме нулевата хипотеза. Няма основание да се счита, че автомобилите в А са на повече километри.

Тестване на хипотези за средните на две популации



продължение

Двете популации са нормално разпределени

не → **Използвай непараметрични методи (не са включени в този курс).**

да ↓

σ_1 σ_2 са известни

да → **z тест (нормално разпр.):**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

(s_1^2 и s_2^2 не могат да бъдат използвани за оценки на σ_1^2 и σ_2^2.)

не ↓

Проверяваме хипотеза за $\sigma_1^2 = \sigma_2^2$

отхвърляме $\sigma_1^2 = \sigma_2^2$ → **t тест (извадките са от нормална популация):**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

където ст. на св. = по-малкото от $n_1 - 1$ и $n_2 - 1$.

Не отхвърляме $\sigma_1^2 = \sigma_2^2$ ↓

t тест (извадките са от нормална популация):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_P^2}{n_1} + \dfrac{s_P^2}{n_2}}}$$

където $s_P^2 = \dfrac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{(n_1 - 1) + (n_2 - 1)}$

и ст. на свобода= $n_1 + n_2 - 2$

Сравняване на две пропорции

Независими извадки

Разглеждаме опити на Бернули и две биномно разпределени популации (алтернативни популации)

Първа извадка: обем n и брой успехи в нея x =>намираме $\hat{p}_1 = \frac{x}{n}$

Втора извадка: **обем m и брой успехи в нея y =>намираме** $\hat{p}_2 = \frac{y}{m}$

Предположения :

$$n\hat{p}_1 \geq 10, \quad n(1-\hat{p}_1) \geq 10$$

$$m\hat{p}_2 \geq 10, \quad m(1-\hat{p}_2) \geq 10$$

H_0: $p_1 - p_2 = D_0$ H_1: $p_1 - p_2 \neq D_0$

статистика

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n-1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{m-1}}}$$

- “Несемейните служители отсъстват по-често от работа отколкото семейните”.

За целта се избират 250 семейни служители, от които се оказва че 22 са отсъствали повече от 5 дни последната година, докато при случайно избрани 300 несемейни служители се оказало, че 35 са отсъствали повече от 5 дни. При ниво на значимост 0,05 какво може да кажете за твърдението?

Интерпретация на данните:

Първа извадка: обем n=250 и брой успехи в нея x =22>намираме

$$\hat{p}_1 = \frac{x}{n} = \frac{22}{250} = 0{,}088$$

Втора извадка: **обем m=300 и брой успехи в нея y =35>намираме**

$$\hat{p}_2 = \frac{y}{m} = \frac{35}{300} = 0{,}1167$$

$H_0: p_1 = p_2$ $H_1: p_1 < p_2$

Предположения :

$$n\hat{p}_1 = 22 \geq 10, \quad n(1-\hat{p}_1) = 228 \geq 10$$
$$m\hat{p}_2 = 35 \geq 10, \quad m(1-\hat{p}_2) = 265 \geq 10$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 - 1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2 - 1}}} = \frac{0{,}088 - 0{,}1167}{\sqrt{\frac{0{,}088(0{,}912)}{249} - \frac{0{,}1167(0{,}8833)}{299}}} = -1{,}1112$$

- α=0,05 и критичната област е

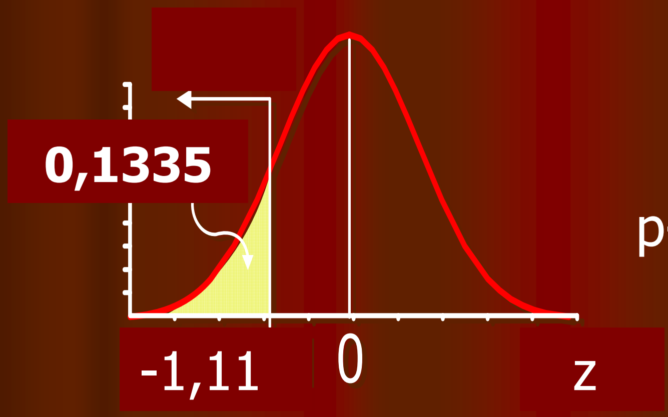
  (-∞, – 1,65).

Не отхвърляме хипотезата, т.е. няма основание да смятаме, че несемейните отсъстват по-често от работа.

p-стойност на теста

$$z=\frac{(\hat{p}_1-\hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1-1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2-1}}}=-1{,}1112$$

Лявостранен тест



p-стойност на теста=0,1335>0,1

Извод: не отхвърляме хипотезата

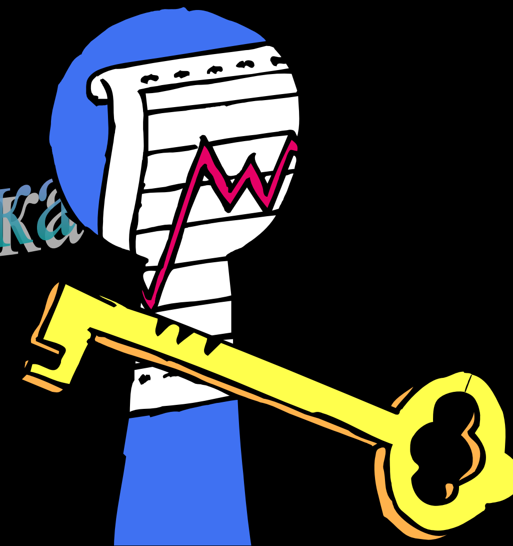
Ако отхвърлим хипотезата, то ще допуснем грешка от първи род= 0,1335

Статистиката е наука за

- събиране,
- организиране,
- обобщаване,
- анализиране, и
- интерпретиране

на данни с цел да се вземе по-ефективно решение.

Що е статистика

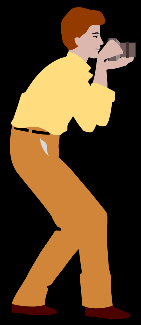


Основни елементи на статистиката

- Събиране на данни
- Обобщаване на данни
- Интерпретиране на данни
- Вземане на решения от данни

Събиране на данни

- Определяне предмета на изследване
  - Редуцира ли аспирина риска от сърдечен инфаркт?
- Наблюдения
  - Наблюдения на хора, вземащи аспирин и не вземащи аспирин



Време за някои дефиниции

Популация: Групата, която притежава изучаваната характеристика

Индивид : всеки член на популацията

Извадка **- подмножество на популацията**

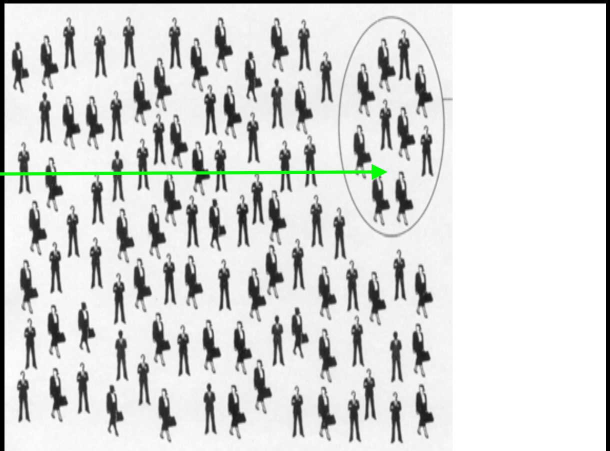
Броят n на елементите на извадката се нарича *обем* на извадката.

Пример: Изучаване въздействието на аспирина

Популация: всички хора от дадена възрастова група

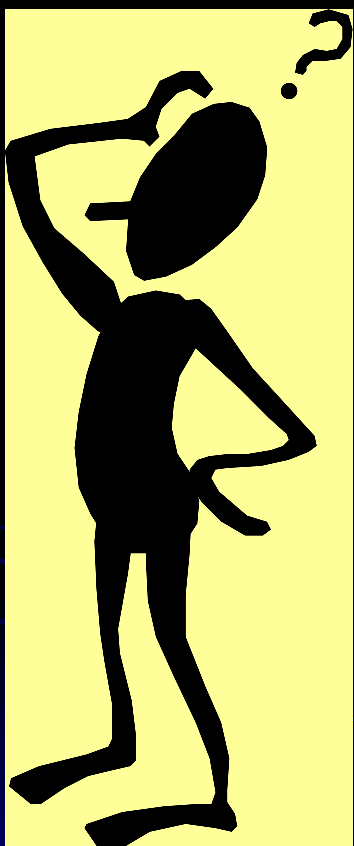
Индивид : всеки човек

Извадка : избират се само 100 човека и се набюдават



100= обем на извадката

Защо извадка?



В повечето изследвания в е трудно да се получи информация от цялата популация. Тогава на основата на извадката ние оценяваме или правим изводи за цялата популация.

Още дефиниции

променливи – изследваната характеристика(и) на индивидите в популацията

данни - Списък от стойности от наблюденията на изследваната характеристика (и)

Пример

Изследване на средния успех на студентите в ПУ

Променлива: среден успех

Данни: 3,82; 4,95; 4,82; 3,49; 5,70

Видове променливи

Качествени променливи – изследваната характеристика не се измерва числено

При наблюдението им се определят категории

Количествени променливи – измерва се числено

Дискретни променливи

Непрекъснати променливи

примери

- цвят на очи
- пол,
- Мнение относно преподавател

Сини

Кафяви

Черни и пр.

Жена

Мъж

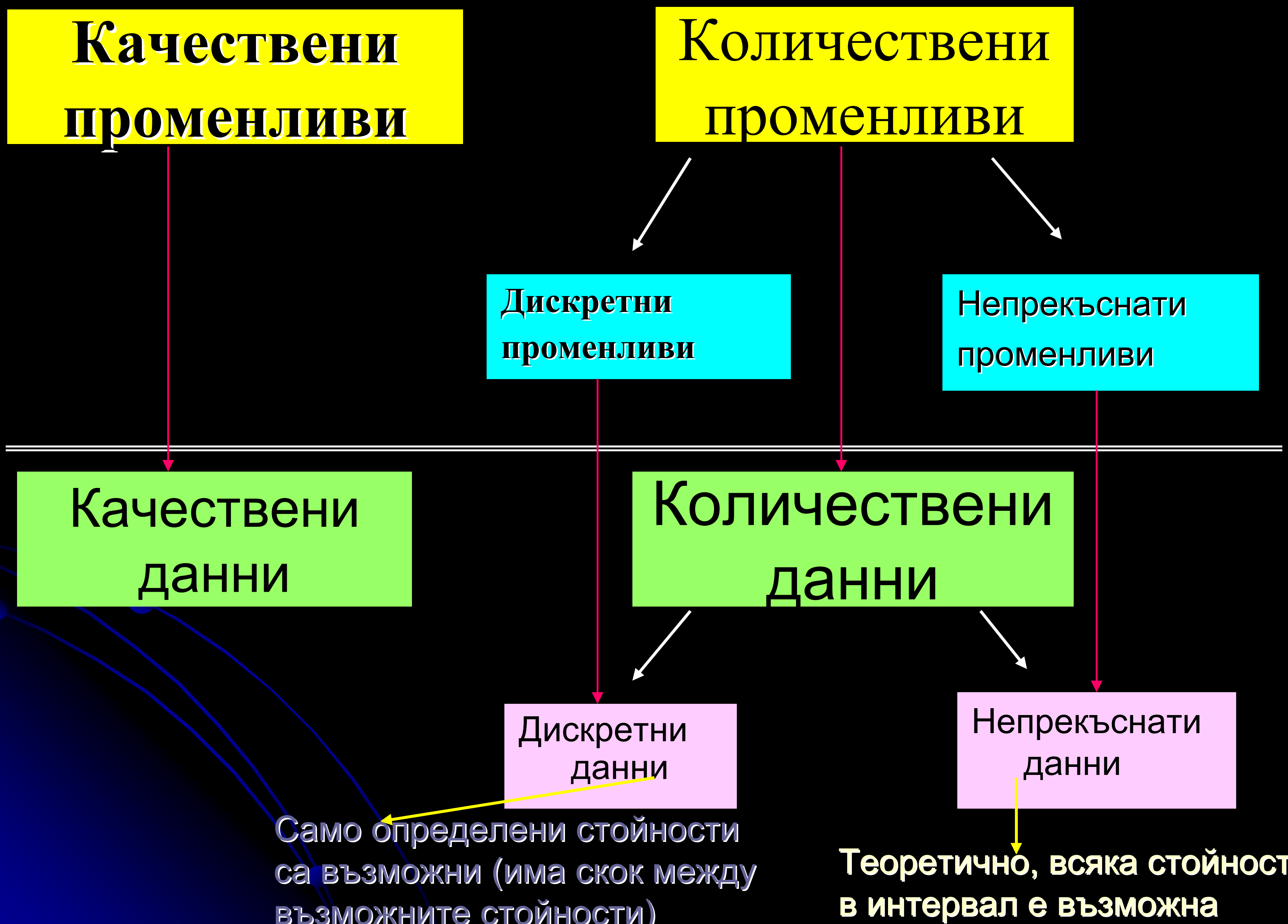
Много добро

Добро

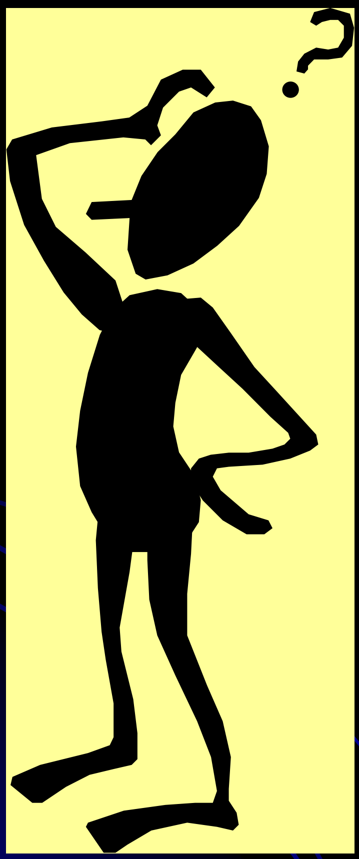
Лошо

Нямам мнение

- баланс в банкова сметка
- ръст
- успех
- възраст


Качествени променливи
Количествени променливи
Дискретни променливи
Непрекъснати променливи
Качествени данни
Количествени данни
Дискретни данни
Непрекъснати данни
Само определени стойности са възможни (има скок между възможните стойности)
Теоретично, всяка стойност в интервал е възможна

Защо е важен видът на данните?



Видът на данните
определя и
статистическия анализ,
който ще се използва

Как се събират данни?

Има различни методи, но ние ще разгледаме само **случайни извадки**.

С разглеждането на случайните извадки, ще считаме, че тя представлява адекватно популацията, което ни дава основание да считаме, че заключенията, направени от извадката са верни и за популацията. Разбира се при това има някаква несигурност—вероятност за грешка!!!

Организиране и обобщаване на данните

Графично представяне на данните

- Зависи от типа данни
- Зависи от това, какво искаме да илюстрираме
- Зависи от статистическия софтуер, с който разполагаме

Качествени данни

Избрани са по случаен начин 200 първокурсници в ПУ и е записана тяхната специалност

Пресмятат се ЧЕСТОТИ и се оформя

честотно разпределение

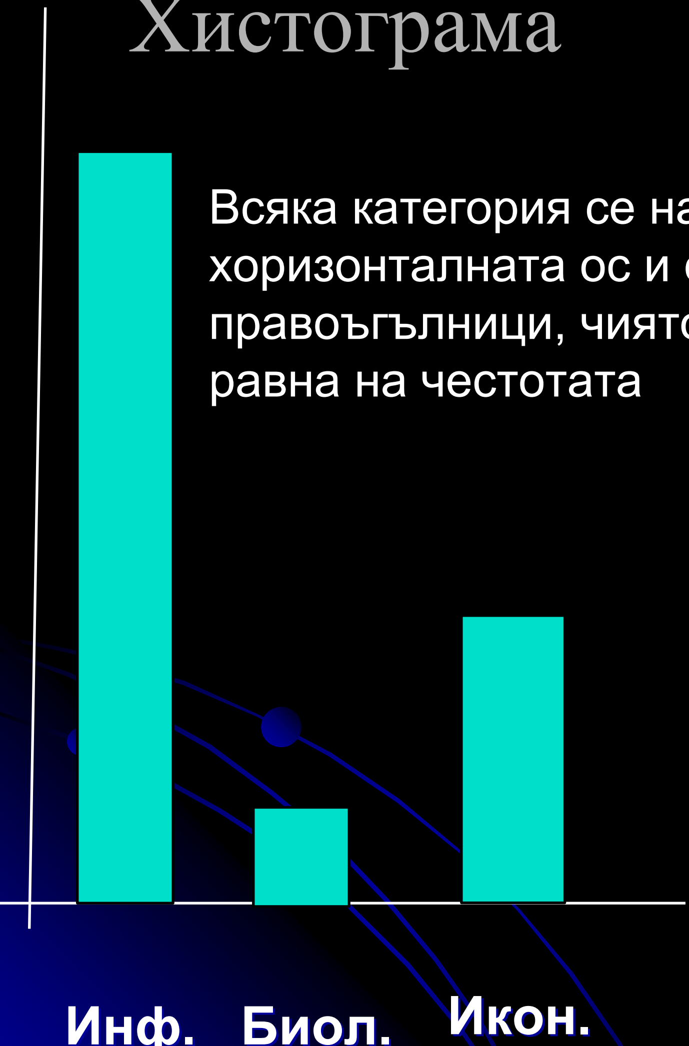
специалност	брой
информатика	130
биология	20
икономика	50
общо	200

категории

Представяне графично на събраните данни:

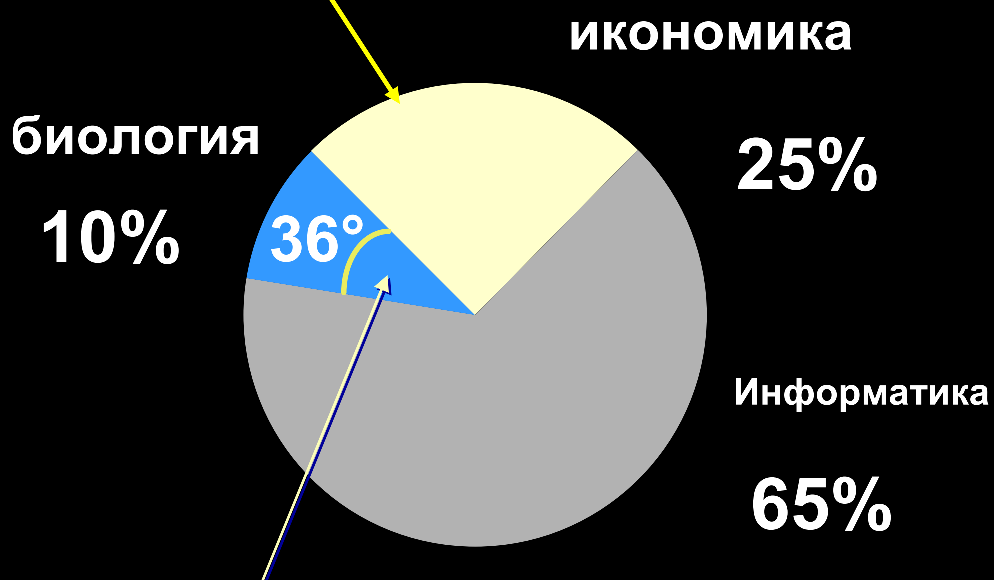
Специалности

Хистограма



Всяка категория се нанася на хоризонталната ос и се чертаят правоъгълници, чиято височина е равна на честотата

Кръг се разделя на сектори, като всеки представя различни категории.Лицето на всеки сектор е пропорционален на честотата.



(360°) (10%) = 36°

Количествени данни

Дискретни данни Графичното представяне с хистограма е подобно на качествените данни.

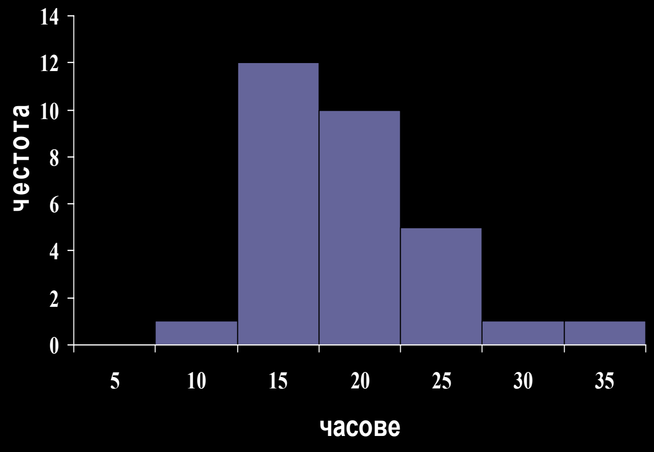
Непрекъснати данни Първо, данните се групират, като получаваме честотно разпределение

ПРИМЕР: Направена е случайна извадка от 30 студенти, които са запитани за броя часове, прекарани в учене през последната седмица:

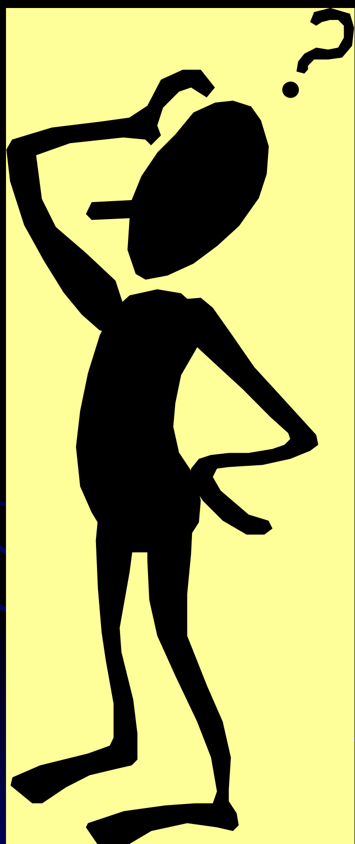
15,0	23,7	19,7	15,4	18,3	23,0	14,2	20,8	13,5	20,7	17,4
18,6	12,9	20,3	13,7	21,4	18,3	29,8	17,1	18,9	10.3	26,1
15,7	14,0	17,8	33,8	23,2	12,9	27,1	16,6			

Часове	честота	Относителна честота
7,5 до 12,5	1	1/30=0,0333
12,5 до17,5	12	12/30=0,400
17,5 до 22,5	10	10/30=0,333
22,5 до 27,5	5	5/30=0,1667
27,5 до 32,5	1	1/30=0,0333
32,5 до 37,5	1	1/30=0,0333
ОБЩО	30	30/30=1

хистограма



Защо различни методи?



Качествените и количествените данни имат съвсем различно поведение и затова се изучават по различен начин.

Описателна статистика

Какво можем да опишем?

Какво е “местоположението” или “центъра” на данните?

Как варират данните?

Мерки за “местоположението” на данните

Средна стойност

Медиана

Мода

Средна стойност

- Ако описва популацията, се нарича популационна средна стойност и се означава с μ **(параметър)**
- Ако описва извадката се нарича извадково средно и се означава с $\bar{x}$ (статистика)
- Използва се само за количествени данни.
- Съществено се влияе от всички данни.

Формула $\bar{X} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

$x_1, x_2, x_3, \ldots x_n$

Стойностите от данните

Пример: Случайно избрани студенти, взели даден тест, са получили следните точки 14, 15, 17, 16, 15

Тогава средният им брой точки е извадково средно

$$\bar{X} = \frac{\Sigma X}{n} = \frac{14 + \ldots + 15}{5} = \frac{77}{5} = 15{,}4$$

Медиана

Данните са подредени във възходящ ред.

Медианата е точка, за която 50% от данните са по-малки от нея.

- Данните, по-малки от медианата са точно толкова, колкото и данните по-големи от нея.
- Използва се само за количествени данни.

- При нечетен брой данни, медианата е =средния елемент на данните
- При четен брой данни, медианата е аритметично средното на двата средни елемента;

Примери

<u>6,72 3,46 3,60 6,44 26,70</u>

3,46 3,60 6,44 6,72 26,70 (наредени данни)

(нечетен брой данни)

Има среда

<u>Медианата</u> е 6,44

<u>6,72 3,46 3,60 6,44</u>

3,46 3,60 6,44 6,72 (наредени данни)

(четен брой данни)

$$\frac{3,60 + 6,44}{2}$$

<u>Медианата</u> е 5,02

Мода

Модата е най-често срещаната стойност в данните .

•Данните могат да имат повече от една мода.
•Подходяща е за всеки вид данни, но най-често се използва при качествени данни или дискретни данни с малък брой възможни стойности

а. 5 5 5 3 1 5 1 4 3 5

Модата е 5

б. 1 2 2 2 3 4 5 6 6 6 7 9

Бимодална :две моди 2 и 6

в. 1 2 3 6 7 8 9 10

Няма мода

Мерки за “разсейването” на данните

Тези мерки са подходящи само за количествени данни.

Дисперсия

- Ако описва популацията, се нарича популационна дисперсия и се означава с σ^2 **(параметър)**
- Ако описва извадката се нарича извадкова дисперсия и се означава с s^2 **(статистика)**
- Използва се само за количествени данни.
- Съществено се влияе от всички данни.

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

Стандартно отклонение

- Ако описва популацията, се нарича популационна дисперсия и се означава с **σ (параметър)**
- Ако описва извадката се нарича извадкова дисперсия и се означава с s (статистика)
- Мерните единици са същите както и мерните единици на данните.
- Измерва отклонението на данните от тяхната средна стойност.
- Съществено зависи от всички данни.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Пример

Фирма понякога наема носачи за почасова работа, като им плаща в зависимост от предлагането. Случайно са избрани 5 човека, работили почасово в тази фирма, на които се е оказало, че фирмата е плащала по

7 лв, 5 лв, 11 лв, 8лв, 6 лв на час.

Намерете стандартното отклонение .

$$\overline{X} = \frac{\Sigma X}{n} = \frac{37}{5} = 7,40$$

$$s^2 = \frac{\Sigma\left(X - \overline{X}\right)^2}{n-1} = \frac{(7-7,4)^2 + \ldots + (6-7,4)^2}{5-1}$$

$$= \frac{21,2}{5-1} = 5,30$$

$$s = \sqrt{s^2} = \sqrt{5,30} = 2,30$$

Извадково стандартно отклонение

Важно!!!

Извадковите характеристики

- Извадково средно
- Извадкова дисперсия (изв. станд. откл.)

зависят от извадката =>

те имат поведение на случайни величини.

Като случайни величини те си имат разпределение.

Разпределението на извадковото средно

Нека $\overline{X} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

Разглеждаме частен случай:

Нека случайната извадка с обем n е от нормално разпределена популация със средна стойност μ и дисперия σ², т.е. N(μ,σ²)

$$EX_i = \mu$$

$$\bar{X} = \frac{X_1 + X_2 + \dots X_n}{n}$$

$$Дисп.X_i = \sigma^2$$

$$E\bar{X} = \frac{EX_1 + EX_2 + \dots EX_n}{n} = \frac{\mu + \mu + \dots + \mu}{n} = \mu$$

Средна стойност на извадковото средно

$$Дисп\bar{X} = \frac{Дисп.X_1 + ДиспX_2 + \dots ДиспX_n}{n^2} = \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Дисперсия на извадковото средно

Извадковото средно е нормално разпределено, т.е. $\bar{X} \in N(\mu, \frac{\sigma^2}{n})$

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

Ако популацията не е нормално разпределена, то съгласно ЦГТ

Извадка с обем n е направена от алтернативна популация- всеки елемент притежава или не притежава дадена характеристика.

статистика

Нека X= брой индивиди от извадката, които притежават характеристиката

$$\hat{p} = \frac{x}{n}$$

Разглеждаме опити на Бернули – n опита и p=P(Успех) =P(отделния елелемент да притежава характеристиката)

X={брой *Успехи* в тези n опита}

биномно разпределение

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

При голямо n

Стандартно нормално

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{p(1-p)}}\sqrt{n}$$

Статистиката pˆ има средна стойност p и дисперсия

$$Дисп.\hat{p} = \frac{p(1-p)}{n}$$

Дискретни разпределения

Дискретна случайна величина = която приема **краен брой** или **изброимо много** стойности

Ред на разпределение на дискретна сл.в.= съвкупност от стойности+вероятности ; може да бъде във формата на

- таблица

стойност (x)	X_1	X_2	....	X_n
вероятност (p)	$P(X=x_1)$	$P(X=x_2)$	......	$P(X=x_n)$

- математическа формула

$$P(S_n = k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}$$

Свойства на реда на разпределение

стойност (x)	x_1	x_2	…….	x_n
вероятност (p)	$P(X=x_1)$	$P(X=x_2)$	……	$P(X=x_n)$

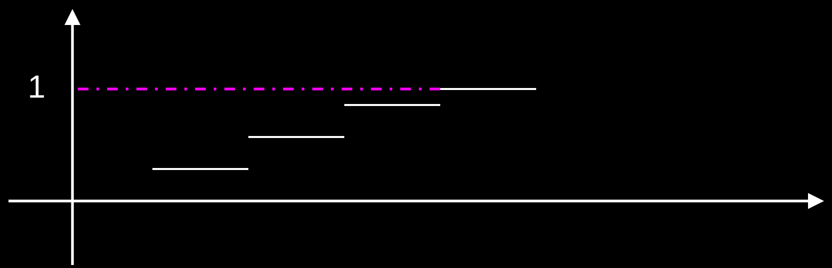
Свойство 1. $x_i \neq x_k$

Свойство 2. $\sum_i p_i = 1$ $\quad 0 \le p_i = P(X = x_i) \le 1$

Функция на разпределение

$$F(x) = \begin{cases} 0 & при \quad x < x_1 \\ p_1 & при \quad x_1 < x \le x_2 \\ p_1 + p_2 & при \quad x_2 < x \le x_3 \\ p_1 + p_2 + p_3 & при \quad x_3 < x \le x_4 \\ \dots\dots\dots\dots\dots\dots & \\ 1 & при \quad x > x_n \end{cases}$$

$$F(x) = P(X < x) = \sum_{j;\, x_j < x} p_j$$



Примери

Опит: Хвърляне на монета един път

X={брой лица}

Y={брой гербове}

стойност (x)	0	1
вероятност (p)	0.5	0.5

$$F(x)=\begin{cases}0 & при \quad x\le 0\\ 0{,}5 & при \quad 0<x\le 1\\ 1 & при \quad x>1\end{cases}$$

Опит: Хвърляне на зар един път

X={брой точки върху зара}

x	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Y={брой лицеви страни с точно една точка}

x	0	1
P	5/6	1/6

$$F(x)=\begin{cases}0 & при \quad x\le 0\\ 5/6 & при \quad 0<x\le 1\\ 1 & при \quad x>1\end{cases}$$

Опит: Случайно се избира топче от кутия с 5 червени и 2 сини топчета

X={брой червени топчета измежду избраните}

x	0	1
p	2/7	5/7

Може ли следната таблица да е ред на разпределение на сл.в.?

x	0	1	2	3
P	0.4	0.6	0.3	0.1

не

x	0	1	2	3
P	0.4	0.35	0.1	- 0.15

не

x	0	1	2	3
P	0.4	0.35	0.1	0.15

да

x	0	1	2	3
P	.4	.6	3.0	.1

не

Пример

Опит: Хвърляне на зар два пъти

X={максималните точки, които се появяват на зара в двете хвърляния}

Стойности на X=> 1,2,3,4,5,или 6

(1,1)

(2,1) (1,2), (2,2)

(3,1) (1,3), (3,2) (2,3), (3,3)

x	1	2	3	4	5	6
p	1/36	3/36	5/36	7/36	9/36	11/36

$$F(x)=\begin{cases} 0 & при \quad x \le 1 \\ 1/36 & при \quad 1 < x \le 2 \\ 4/36 & при \quad 2 < x \le 3 \\ 9/36 & при \quad 3 < x \le 4 \\ 16/36 & при \quad 4 < x \le 5 \\ 25/36 & при \quad 5 < x \le 6 \\ 1 & при \quad x > 6 \end{cases}$$

У={минималните точки, които се появяват на зара в двете хвърляния}

Стойности на У=> 1,2,3,4,5,или 6

x	1	2	3	4	5	6
p	11/36	9/36	7/36	5/36	3/36	1/36

Задача:

Сл. в. X има ф.р.

$$F(x)=\begin{cases}0 & при \quad x\le -1\\ 0{,}02 & при \quad -1<x\le 0\\ 0{,}08 & при \quad 0<x\le 3\\ 0{,}1 & при \quad 3<x\le 7\\ 0{,}3 & при \quad 7<x\le 10\\ 0{,}6 & при \quad 10<x\le 16\\ 1 & при \quad x>16\end{cases}$$

Какъв тип е сл.в.?

Дискретен

Стойности: -1; 0; 3; 7; 10; 16

Ред на разпределение

x	-1	0	3	7	10	16
p	0,02	0,06	0,02	0,2	0,3	0,4

Средна стойност (математическо очакване)

$EX = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$

Средната стойност дава информация за средата на стойностите на случайната величина.

Пример

Опит: Хвърляне на монета един път

стойност	0	1
вероятност	0.5	0.5

X={брой лица}

EX=0(0.5)+(1)(0.5)=0.5

Свойства

$E(c) = c$

E(cX)=c(EX)

$E(X+Y) = EX + EY$

$E(X.Y) = EX.EY \ \ ako \ \ X \ и \ \ Y \ \ са \, независими$

Моделиране на хазартни игри

Том и Ники играят игра: Том хвърля зар един път. Ако се паднат 5 точки, Том плаща 1 лев на Ники, в противен случай Ники плаща 1 лев на Том.

Колко е очакваната печалба на Том ?

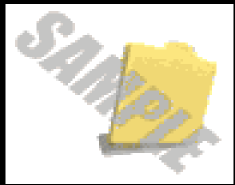
Нека X={печалба на Том}

Разпределение

стойност(x) (лв)	**- 1**	**1**
вероятност (p)	**1/6**	**5/6**

EX= (- 1) (1/6)+(1) (5/6)=4/6=.6666

Интерпретация: Ако двете момчета играят тази игра много пъти, то в някои от тях Том ще плати 1 лв, в някои ще получи 1 лв, но в крайна сметка средната му печалба ще бъде 67 ст.



Пример

Полица”Живот” осигурява плащането на определена сума при смърт на притежателя на полица. Нека например, застраховка “Живот” за 49 годишен мъж е 35 лв за година, като в случай на злополука се изплащат 25 000 лв. Ако е известно, че смъртността при 49-годишните мъже в съответния регион е 135 на 100 000, пресметнете очакваната печалба на застрахователната компания.

Нека X={печалба на компанията}

Стойности на X: 35 и (35-25 000)

стойност(x) (лв)	**35**	**-24965**
вероятност (p)	**0,99865**	**0,00135**

Разпределение

EX=35(0,99865)-24965(0,00135)=1,25 лв. печалба от всеки застрахован

Пример

Ако вероятността да има наводнение следващата година е p, то колко трябва да се плаща за застроховка, за да може застрахователната компания да има печалба поне 10% от евенталната сума, която се изплаща в случай на наводнение .

Сумата, която евентуално ще се изплати : A

Стойността на застраховката: B

Нека X={печалба на компанията}

Стойности на X: B и (B-A)

стойност(x) (лв)	**B**	**B-A**
вероятност (p)	**1-p**	**p**

$$EX=B(1-p)+(B-A)p=(B-Ap)$$ лв. печалба от всеки застрахован

$B-Ap= 0,1A$

$B\geq (0,1+p)A$

Очакване на функция от д.сл.в.

Нека X е дискретна случайна величина в ред на разпределение

x	x_1	x_2	....	x_n
p	$P(X=x_1)$	$P(X=x_2)$	......	$P(X=x_n)$

g(x) е реална функция → g(X) е случайна величина

x	$g(x_1)$	$g(x_2)$	....	$g(x_n)$
p	$P(X=x_1)$	$P(X=x_2)$	......	$P(X=x_n)$

$$Eg(X) = g(x_1)p_1 + g(x_2)p_2 + \ldots + g(x_n)p_n$$

Опит: Хвърляне на зар един път

Y={брой лицеви страни с точно една точка}

x	0	1
P	5/6	1/6

$$E(X^2) = 0^2\frac{5}{6} + 1^2\frac{1}{6} = \frac{1}{6}$$

Специален пример

X

x	0
p	1

EX=0

Y

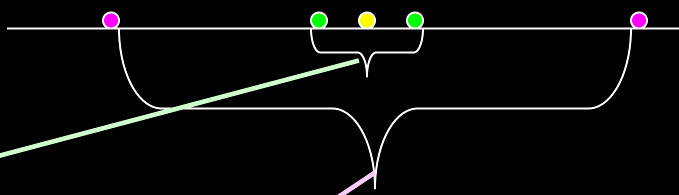
X	- 1	1
p	0,5	0,5

EY=0

Z

X	- 100	100
p	0,5	0,5

EZ=0



Харктеристика, която ги разграничава

Дисперсия

$$\sigma^2 = E(X - EX)^2$$

$$\sigma^2 = \sum_i (x_i - EX)^2 p_i$$

Дисперсията измерва степента на разсейване на стойностите на разпределението.

Свойства

$$\sigma^2(c) = 0$$

$$\sigma^2(cX) = c^2 \quad \sigma^2(X)$$

$$\sigma^2(X+У) = \sigma^2(X) + \sigma^2(У) \quad ако \quad X \, и \, У \, са \, независими$$

Стандартно отклонение = квадратен корен от σ^2.

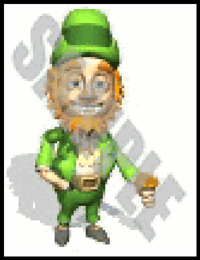
$$\sigma^2 = E(X^2) - (EX)^2$$

Опит: Хвърляне на монета един път

X={брой лица}

стойност (x)	0	1
Вероятност (p)	0.5	0.5

$$\sigma^2 = (0-0.5)^2 0.5 + (1-0.5)^2 0.5$$

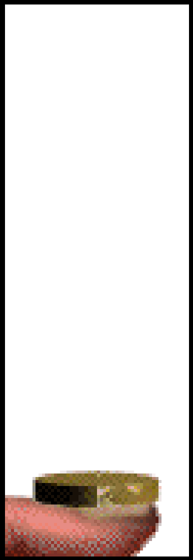


Опит: Хвърляне на монета 3 пъти

X={брой лица}

S={ЛЛЛ, ЛЛГ, ЛГЛ, ЛГГ, ГЛЛ, ГЛГ, ГГЛ, ГГГ}

3 2 2 1 2 1 1 0



Брой лица X	вероятност $P(x)$
0	1/8
1	3/8
2	3/8
3	1/8

Очакване

EX= 0(1/8)+1(3/8)+2(3/8)+3(1/8)=12/8=1.5

Видове дискретни разпределения

Равномерно дискретно

стойност (x)	X_1	X_2	…….	X_n
вероятност (p)	1/n	1/n	……	1/n

математическо очакване

$$EX = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Средно аритметично

Пример: Хвърляне на зарче един път.

X={брой паднали се точки}

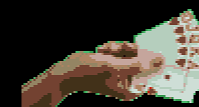
Бернулиево разпределение

Опит: Два възможни изхода : У (успех) и Н(неуспех)

Р(У)=р   Р(Н)=1-р

Х=брой успехи

случайна величина



стойност	0	1
вероятност	1-р	р

ЕХ=р

Дисперсия = р(1-р)

Пример: Избор на карта от колода от 52 карти.

стойност	0	1
вероятност	48/52=12/13	4/52=1/13

Брой дами измежду избраните

ЕХ=4/52

Дисперсия = 12/169

Задача.

Нека Х приема само стойности а и в с вероятности р и 1-р съответно.

а/ Докажете, че У=(Х-в)/(а-в) е Бернулиево разпределена

Х =>

стойност	в	а
вероятност	1-р	р

Стойности на У

0, 1

стойност	0	1
вероятност	1-р	р

а/ Намерете дисперсията на У

Дисперсия на У= р(1-р)

Биномно разпределение

$Bi(n,p)$

Разглеждаме *n* опити на Бернули:

1. Опитите са независими.

2. **Всеки опит има само два възможни изходи, У и Н.**

3. Вероятността за успех във всеки отделен опит е постоянна: ***P(У)=p***

X=брой успехи при тези опити

x	0	1	2	3	...	*n*
p	p_0	p_1	p_2	p_3	.....	p_n

$$p_k = P(S_n = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}, \quad k = 0,1,\dots,n$$

X_k=брой успехи при к-тия опит

$$X = X_1 + X_2 + \dots + X_n$$

$EX=EX_1+EX_2+\dots+EX_n=p+p+\dots+p=np$

Дисперсия= $p(1-p)+ p(1-p)+\dots+ p(1-p)= np(1-p)$

Бернулиево разпределение

Пример: Зарче се подхвърля 5 пъти.

X=брой паднали се “2 точки” при тези опити

Биномно разпределение

x	0	1	2	3	4	**5**
p	p_0	p_1	p_2	p_3	p_4	p_5

$$p_0 = P(S_5 = 0) = \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^5 = \frac{3125}{7776}$$

$$p_1 = P(S_5 = 1) = \frac{5}{1}\left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^4 = \frac{3125}{7776}$$

$$p_2 = P(S_5 = 2) = \frac{5(4)}{2!}\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = \frac{1250}{7776}$$

$$p_3 = P(S_5 = 3) = \frac{5(4)(3)}{3!}\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = \frac{250}{7776}$$

$$p_4 = P(S_5 = 4) = \frac{5(4)(3)(2)}{4!}\left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^1 = \frac{25}{7776}$$

$$p_5 = P(S_5 = 5) = \frac{5(4)(3)(2)(1)}{5!}\left(\frac{1}{6}\right)^5 \left(\frac{5}{6}\right)^0 = \frac{1}{7776}$$

EX=np=5(1/6)=5/6

Дисперсия= np(1-p)=

5(1/6(5/6)=25/36

Станд. Откл.= 5/6

Поасоново разпределение

$Po(\lambda)$

Случайната величина X е Поасоново разпределена, ако

x	0	1	2	...	*n*	...
p	p_0	p_1	p_2	...	p_n	...

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$$

λ е параметър

Примери:

- брой печатни грешки на страница
- брой клиенти, влизащи в даден офис на определен ден
- брой дефектни изделия, измежду произведените определен ден във фирма
- смъртност за даден период в даден регион
- брой земетресения в даден регион през определен период

EX=λ

Дисперсия на X=λ

Пример

Известно е, че средно 3,5 урагана преминават през даден регион. Каква е вероятността следващата година да има поне два урагана в този регион?

X=брой урагани през следващата година

Поасоново разпределена

P(X≥2)=?

P(X≥2)= P(X=2)+ P(X=3)+ P(X=4)+......

P(X≥2)=1-P(X<2)=1-P(X=0)-P(X=1)

EX=3,5= λ

$$p_k = e^{-3,5}\frac{3,5^k}{k!}$$

$$P(X \geq 2) = 1 - e^{-3,5}\frac{3,5^0}{0!} - e^{-3,5}\frac{3,5^1}{1!} = 1 - e^{-3,5}(1+3,5) = 0,86$$

Връзка между Биномно и Поасоново разпределение

При n опити по схемата на Бернули

X={ *брой успехи* } е биномно разпределена Bi(n,p)

При n голямо и p достатъчно малко: $\lambda = np$ и $X \sim Po(\lambda)$

Пример: X=брой новородени в дадена област, които са по-дълги от 58 см.

X е биномно разпределена с n=брой новородени в областта

Долколкото n е голямо, а p е малко, то $X \sim Po(\lambda)$

Пример: X=брой печеливши билети в една лотария

Пример: X=брой печатни грешки в един документ

Пример: X=брой жители на даден регион , по-възрастни от 90 години

Пример

Атомите на радиоактивните елементи се разпадат случайно. Ако всеки грам от даден елемент разпада 3,9 алфа частици за секунда, то каква е вероятността през следващата секунда не повече от 1 алфа- частица да се разпадне от един грам (emitt)

Приятен обед

Всеки грам съдържа голям брой атоми.

Успех=разпадането на алфа-частицата през следващата секунда

X= брой разпаднали се алфа-частици през следващата секунда

Bi(n, p) → EX=3,9 → np=3,9

Доколкото n е голямо => p е малко

Можем да използваме Exp(λ=3,9)

$$P(X \le 1) = \big(P(X = 0) + P(X = 1)\big) =$$

$$\frac{(3{,}9)^0 e^{-3{,}9}}{0!} + \frac{(3{,}9)^1 e^{-3{,}9}}{1!} = e^{-3{,}9}(1 + 3{,}9) = 0{,}099$$

Нормално Разпределение

Нормално разпределена случайна величина има плътност

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

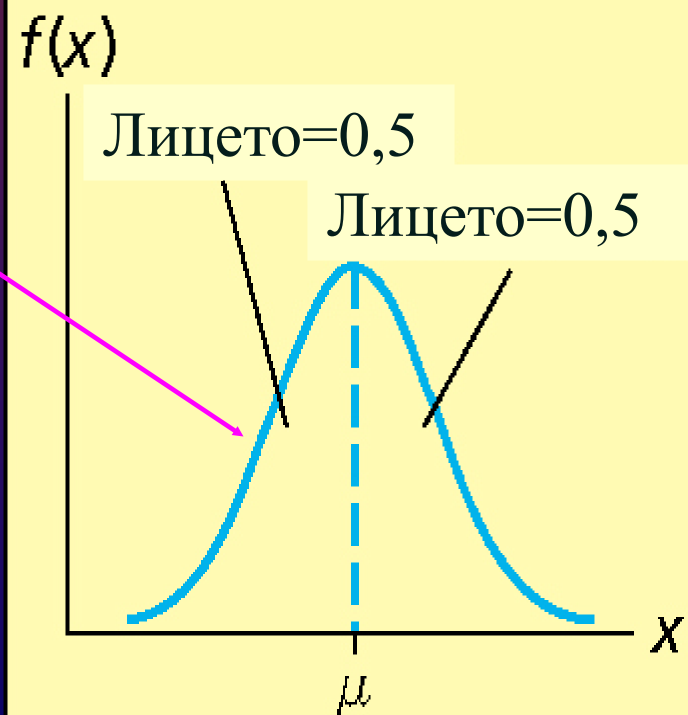
за всяко реално x

$N(\mu, \sigma^2)$

μ и σ са параметри

Нормалната крива е симетрична относно μ и общата й площ под кривата над Ox е 1

Графиката на плътността е камбанка.



Функция на разпределение

$$F(x) = \int_{-\infty}^{x} f(s)ds = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds$$

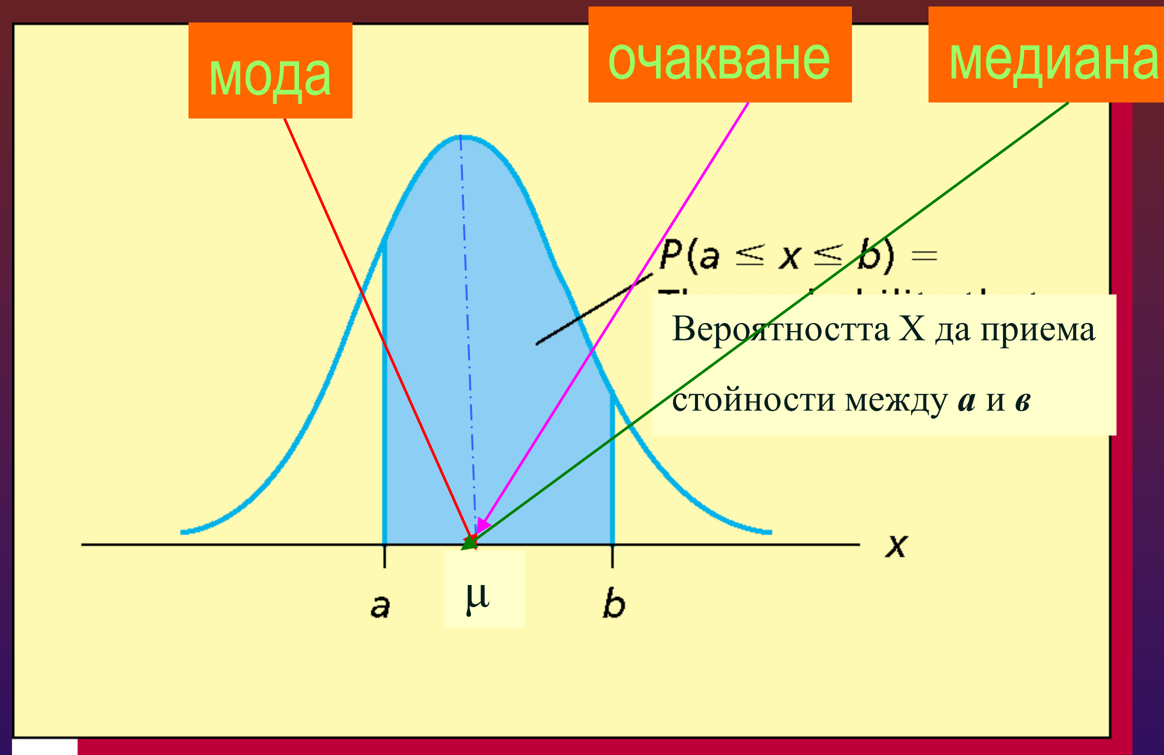
Характеристики на нормалното разпределение

Стойности- цялата реална ос

•Модата, медианата, средната стойност са равни и са точно в средата на основата на камбанката

Височината на камбанката – зависи от стандартното отклонение σ

$$h = \frac{1}{\sigma\sqrt{2\pi}}$$



-по-голямо σ, по-ниска камбанка, но по-широка основа

- по-малко σ, по-висока камбанка, но по-тясна основа

Пример

Нека $X \in N(-1;25)$

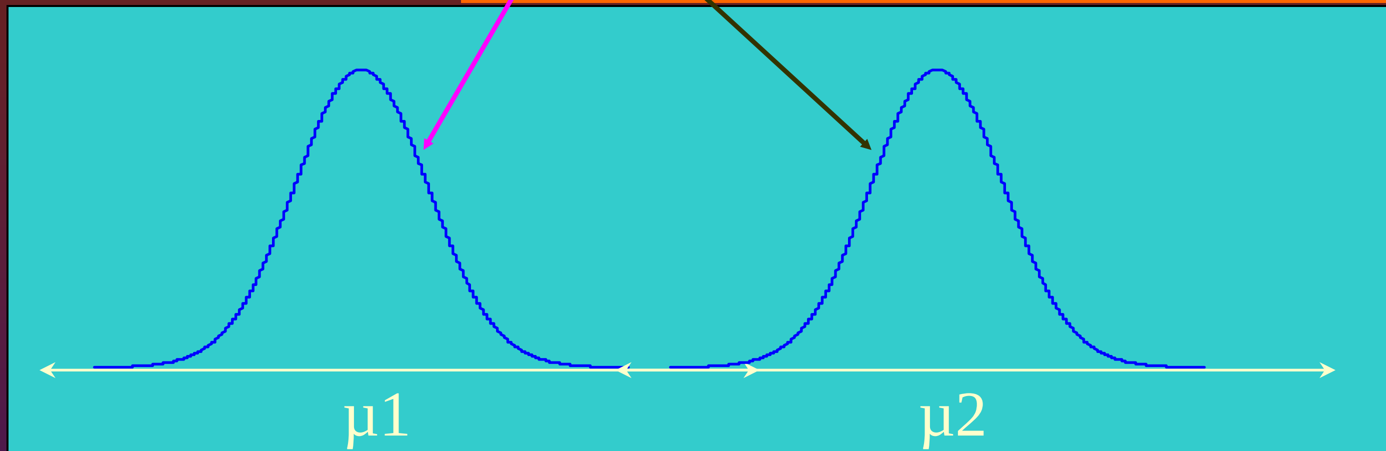
а/ Намерете плътността $f(x) = \frac{1}{5\sqrt{2\pi}} \exp(-\frac{(x+1)^2}{50})$

б/ Намерете средната стойност и стандартното отклонение

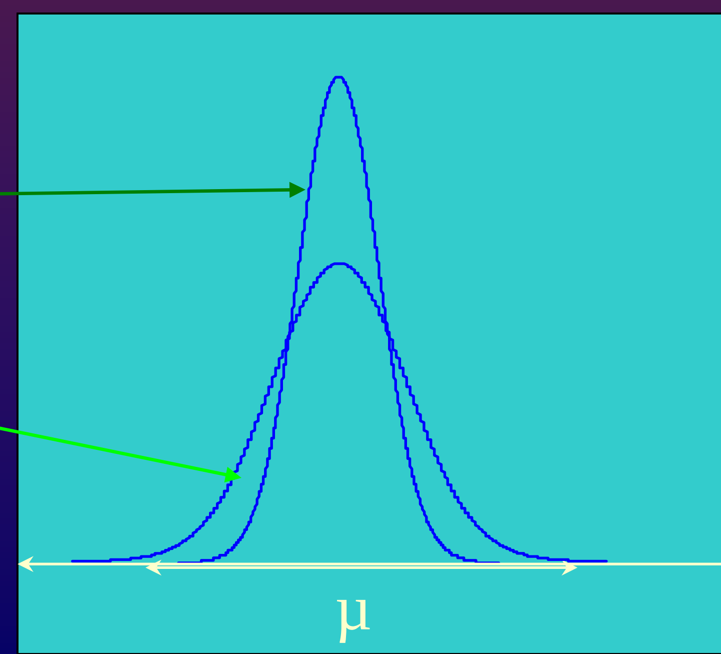
EX= -1 $\sigma^2 = 25$ $\sigma = 5$

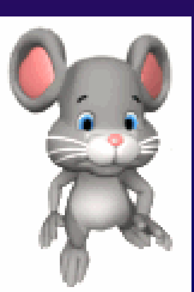
Разположение на камбанката

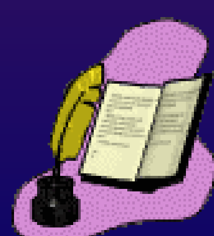
Разглеждаме две нормални разпределения:
$\mu1 < \mu2$ но имат равни станд. отклонения



Разглеждаме две нормални разпределния:
имат едни и същи средни стойности,
но $\sigma1 > \sigma2$







Важно !!!

Нека $X \in N(\mu, \sigma^2)$ Разглеждаме $Z = \dfrac{X-\mu}{\sigma} = \dfrac{1}{\sigma}X - \dfrac{\mu}{\sigma}$

константа

Сл. в-на, която е нормално разпределена

$EZ = ???$

$$EZ = E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}EX - E\frac{\mu}{\sigma} = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0$$

$Дисп.Z = ???$

$$Дисп.Z = Дисп.\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}Дисп.X - Дисп.\frac{\mu}{\sigma} = \frac{1}{\sigma^2}\sigma^2 - 0 = 1$$

Стандартно нормално разпределение

$$Z = \frac{X-\mu}{\sigma} \in N(0,1)$$

Стандартно нормално разпределение

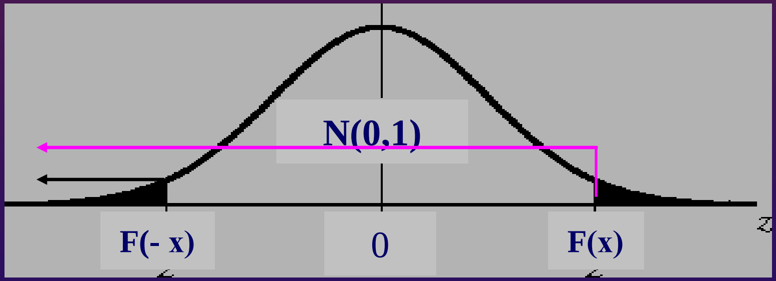
Функция на разпределение $$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{s^2}{2}}ds$$

Свойства на ф.р.

$$F(\infty)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{s^2}{2}}ds=1$$

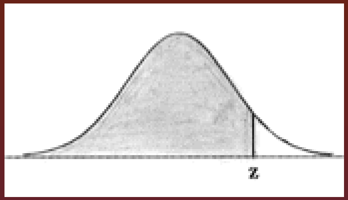
$$F(0)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-\frac{s^2}{2}}ds=0,5$$

F(-x)=1-F(x)



В таблица са дадени само стойностите на лицето под кривата наляво от положителни x

Лице под нормалната крива наляво от (стойности на ф.р. на станд. норм. разпр.)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

P(Z<1,23)=0,8907

P(Z>1,23)=

1-0,8907=0,1093

P(Z< - 1,23)=

1-P(Z<1,23)=

P(Z> - 1,23)=

P(Z< 1,23)=

0,8907

Пример: Нека X е нормално разпределение със средна стойност $\mu=50$ и стандартно отклонение $\sigma= 4$

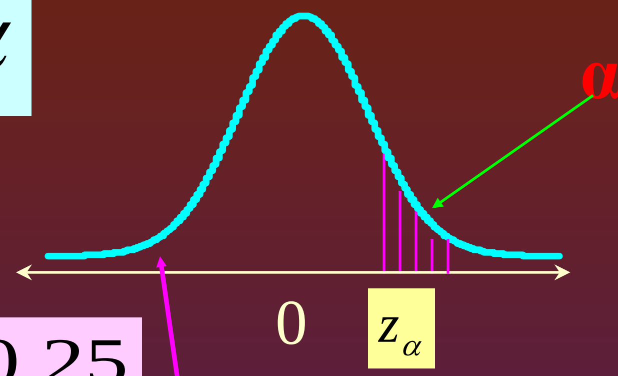
P(X<45)= $P(Z=\frac{X-\mu}{\sigma}<\frac{45-50}{4})$ =P(Z<-1,25)= 1-P(Z<1,25)=1-0,8944= 0,1056

P(X>47)= $P(Z=\frac{X-\mu}{\sigma}>\frac{47-50}{4})$ =P(Z>-0,75)=P(Z<0,75)= 0,7734

P(45<X<51)= $P(\frac{45-50}{4}<Z=\frac{X-\mu}{\sigma}<\frac{51-50}{4})$

= P(-1,25<Z<0,25) = P(Z<0,25)- P(Z<-1,25)= P(Z<0,25)- 1+P(Z<1,25)= 0,5987-1+0,8944=0,4931

Означение z_α $P(Z > z_\alpha) = \alpha$



Пример

$z_{0,25} = ?$

$P(Z > z_{0,25}) = 0,25$

$$P(Z \le z_{0.25}) = 1 - P(Z > z_{0.25}) = 1 - 0,25 = 0,75$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549

$z_{0,25} = 0,67$

От таблицата

$-z_\alpha = z_{1-\alpha}$

$z_{0.9} = ???$

$z_{0.9} = -1.28$

В таблицата търсим 0,9

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

Пример

Нека X е нормално разпределение със средна стойност $\mu=50$ и стандартно отклонение $\sigma= 4$

Намерете точката, на дясно от която се намират 80% от стойностите на X

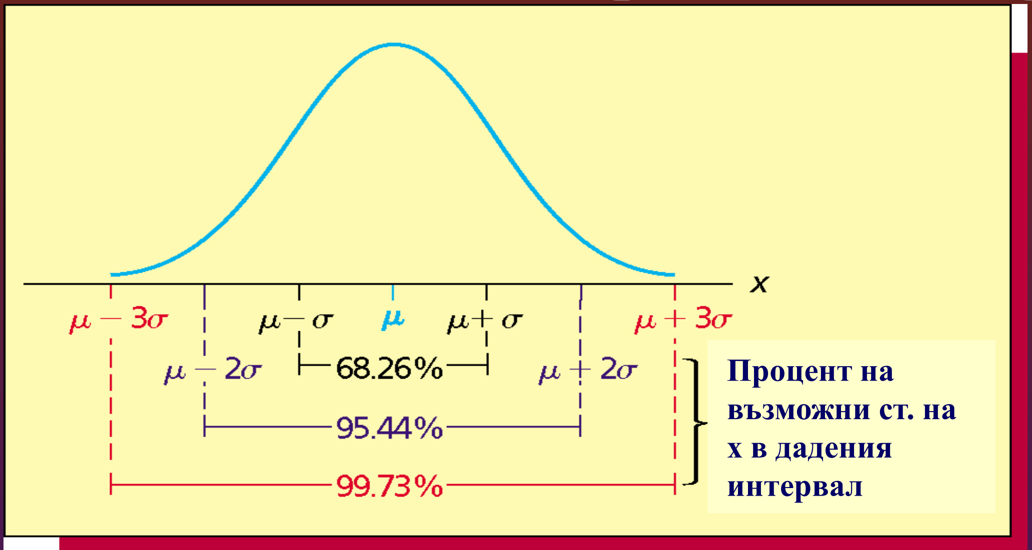
$P(X>a)=0,80 \longrightarrow a<50$ $\quad P(X<a)=1-0,8=0,2$

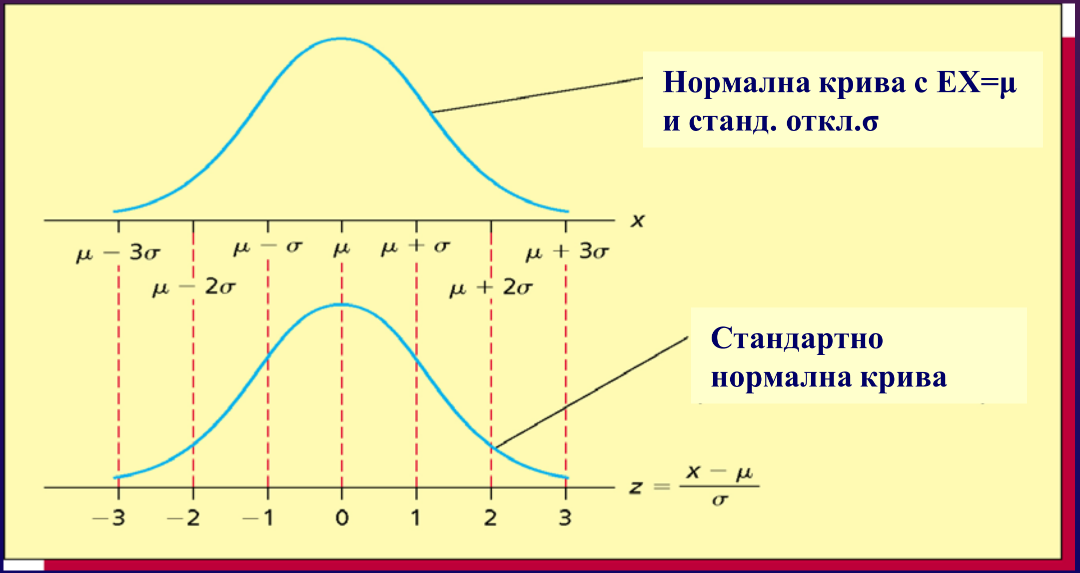
$$P(Z=\frac{X-50}{4}<\frac{a-50}{4})=0,2$$

В z-таблицата няма вероятност <0,5 => търсим точка с лице 0,8 =>0,84

$(a-50)/4= - 0,84$ или $a=(-0,84)4+50=46,64$

Три важни лица под нормалната крива





Централна гранична теорема

$X_1, X_2, \ldots, X_n$ са независими еднакво разпределени сл.в. със средни стойности **μ** и станд. откл. **σ**

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \to \infty]{} N(0,1)$$

Доказателство: (за норм. сл.в.)

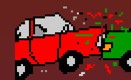
$\dfrac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$ е норм. сл.в. като сума от норм. сл. в.

$$E\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{EX_1 + EX_2 + \ldots + EX_n - n\mu}{\sigma\sqrt{n}} = \frac{\mu + \mu + \ldots + \mu - n\mu}{\sigma\sqrt{n}} = 0$$

$$Дисп\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{ДиспX_1 + ДиспX_2 + \ldots + ДиспX_n - Дисп(n\mu)}{\sigma^2 n} = \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2 - 0}{\sigma^2 n} = 1$$

$\dfrac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$ е $N(0,1)$

Приложение



Застрахователна компания има 10 000 застраховани, като очаква средно на полица да бъдат изплатени през следващата година по 25 лв. със стандартно отклонение 80 лв. Намерете вероятността следващата година сумата за изплатени злополуки да надвишава 270 000 лв.

X_1= сумата, която ще се изплати за злополука на първия клиент

Случайна величина

Дефинираме 10 000 сл.в., които са независими и еднакво разпределени.

Край

Общата сума, която ще изплати, е сумата от всички тези 10 000 сл. в-ни

$$P\left(\sum_{i=1}^{10000} X_i > 270\,000\right) = ?$$

$$Z = \frac{\sum_{i=1}^{10000} X_i - 10000(25)}{80\sqrt{10000}} \in N(0,1)$$

Съгласно ЦГТ, понеже 10 000 е достатъчно голямо число, можем да използваме приблизително N(0,1)

$$P\left(\sum_{i=1}^{10000} X_i > 270\,000\right) \approx P\left(Z > \frac{270000 - 10000(25)}{80\sqrt{10000}}\right) = P(Z > 2{,}5) = 1 - P(Z < 2{,}5) = 1 - 0{,}9938 = 0{,}0062$$

Непрекъснати разпределения

Дефиниция: Казваме, че сл.в. X е непрекъсната, ако съществува интегруема функция f, дефинирана в R такава, че за всяко реално x

Плътност на непрекъсната случайна величина

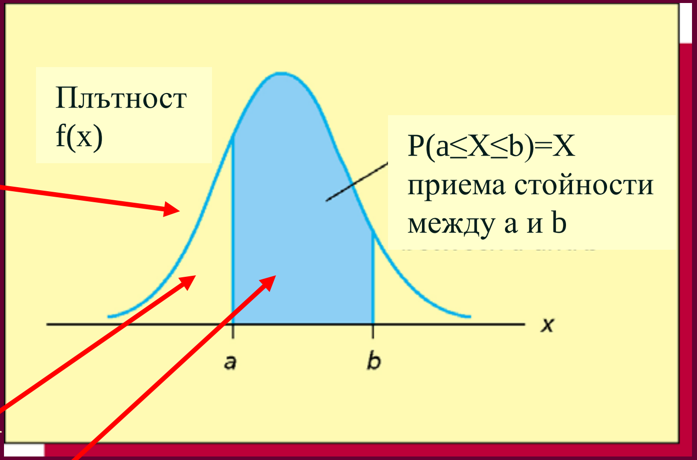
$$F(x) = \int_{-\infty}^{x} f(s)ds$$

Функция на разпределение

$$f(x) = F'(x)$$

Свойства на плътността

Дефинирана в R



$F\ е\ растяща \Rightarrow F'(x) \geq 0$

f(x) е неотрицателна функция – графиката ѝ е над и по абсцисната ос

$$F(\infty) = \int_{-\infty}^{\infty} f(s)ds = 1$$

Лицето между графиката на f(x) и абсцисната ос е 1

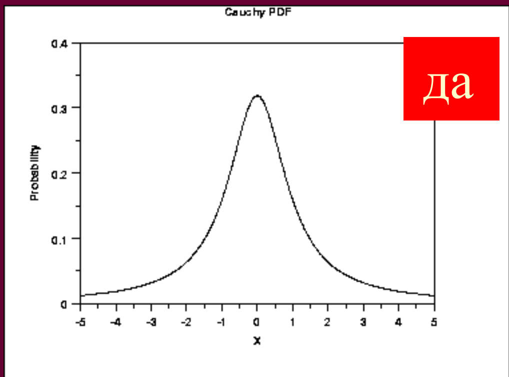
$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(s)ds$$

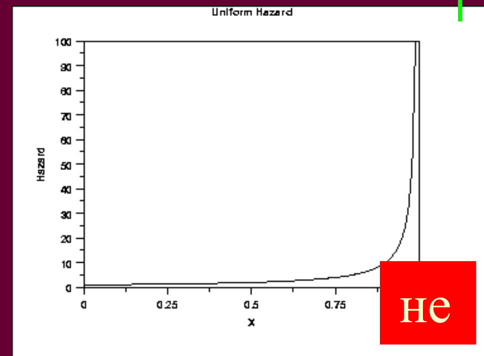
P(X=a)=0

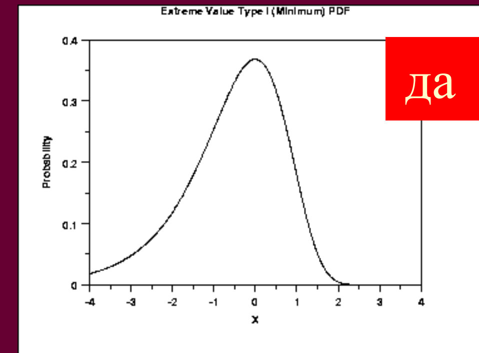
Лицето между графиката и абсцисната ос над интервала [a,b] е равен на вероятността P(a<X≤b)

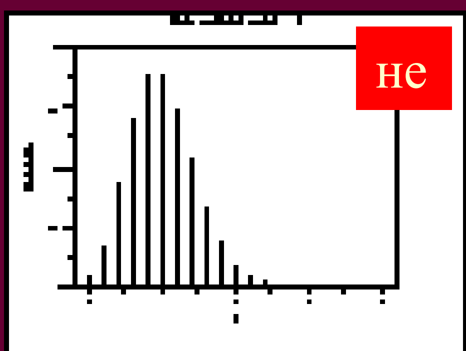
Може ли следната графика да е графика на плътност на непрекъснатата с.в. ?

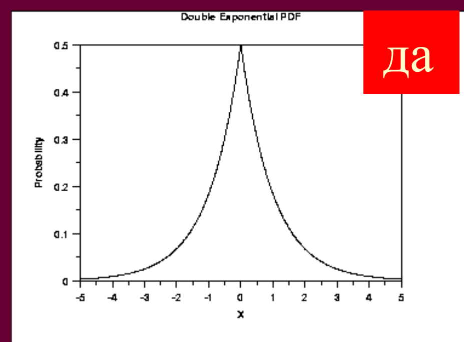
Ако да, къде приема стойности сл.в. ?

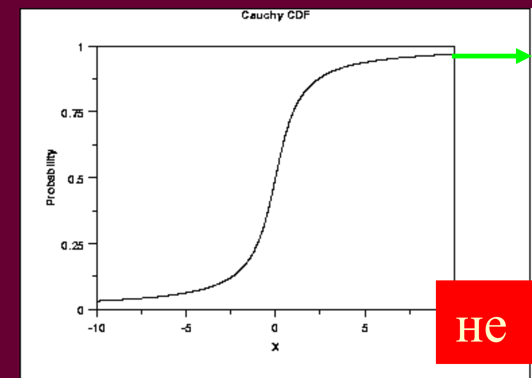


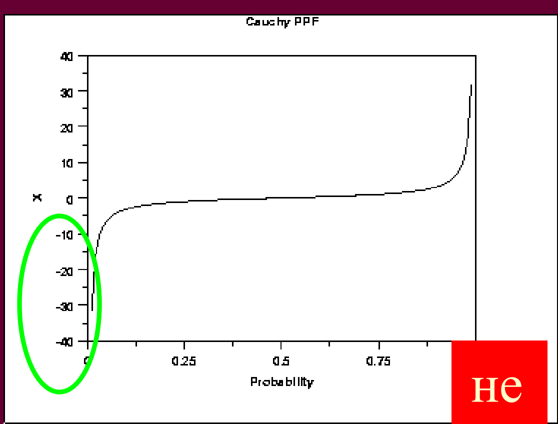


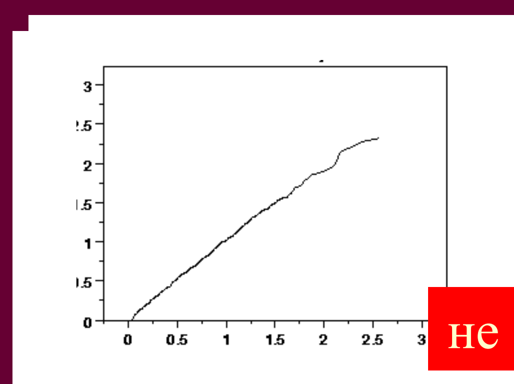


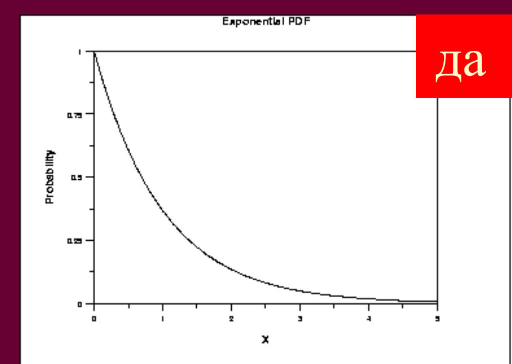












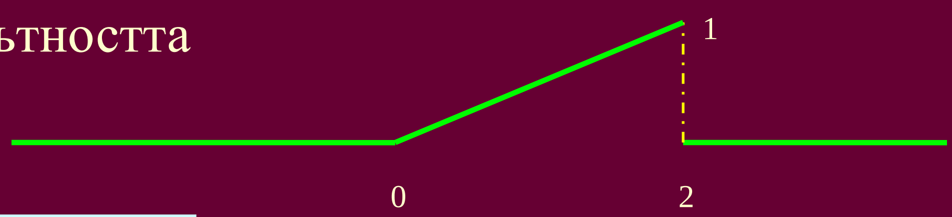
Пример 1:

Плътността на непрекъснатата сл. в. X е

$$f(x)=\begin{cases} \boxed{0{,}5x} & \text{при } 0<x<2 \\ \boxed{0} & \text{в противен случай} \end{cases}$$

Намерете стойността на константата *c* $\displaystyle\int_{-\infty}^{\infty} f(s)ds = 1$ $\displaystyle\int_{0}^{2} cx\,dx = 1$ ***2c=1*** ***c=0,5***

Начертайте графиката на плътността



Немерете ф.р.

При x≤0 $\displaystyle F(x) = \int_{-\infty}^{x} f(s)ds = \int_{-\infty}^{x} 0ds = 0$

При 0<x≤2 $\displaystyle F(x) = \int_{-\infty}^{x} f(s)ds = \int_{-\infty}^{0} 0ds + \int_{0}^{x} 0{,}5sds = 0{,}25x^2$

При x>2 $\displaystyle F(x) = \int_{-\infty}^{x} f(s)ds = \int_{-\infty}^{0} 0ds + \int_{0}^{2} 0{,}5sds + \int_{2}^{x} 0ds = 1$

Каква е вероятността сл. в. да приема стойности по-малки от 1?

$\displaystyle \mathrm{P}(\mathrm{X}<1) = \int_{-\infty}^{1} f(s)ds = \int_{0}^{1} 0{,}5sds = 0{,}25$

или P(X<1)=P(X≤1)=F(1)=0,25

Някои изводи

Непрекъсната случайна величина: стойностите ѝ са всички числа от даден интервал (или интервали), крайни или безкрайни.

Функцията на разпределение на непрекъсната случайна величина, като интеграл, е непрекъсната функция.

Може ли да се твърди, че сл. в-ни са или дискретни или непрекъснати?

X има ф.р.

$$F(x)=\begin{cases}0 & при \quad x<0\\ \dfrac{x^2}{2} & при \quad 0\le x<1\\ \dfrac{1}{4}(4x-x^2) & при \quad 1\le x<2\\ 1 & при \quad x\ge 2\end{cases}$$

Прекъсната в т. 1

X е нито дискретна нито непрекъсната

Числови характеристики

Мода=най-вероятна стойност- точка на локален максимум на плътността

Медиана=точка, за която 50% от стойностите на случайната величина са по-големи и 50% са по-малки от нея – точка, която разполовява лицето под графиката на плътността

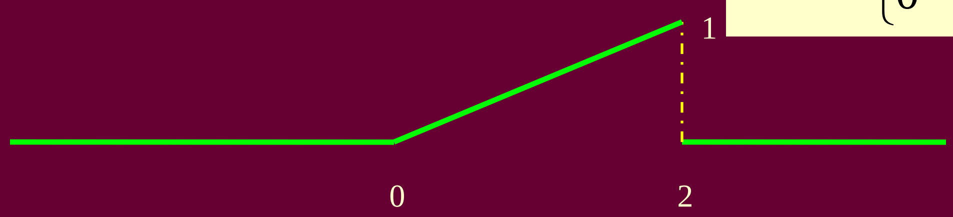
$$M: \quad 0{,}5 = \int_{-\infty}^{M} f(x)dx$$

Пример:

Сл.в. X има плътност

$$f(x) = \begin{cases} 0 & при \quad x < 0 \\ 0{,}5x & при \quad 0 \le x < 2 \\ 0 & при \quad x \ge 2 \end{cases}$$

Мода=2



Медиана

$$M: \quad 0{,}5 = \int_{-\infty}^{M} f(x)dx = \int_{0}^{M} 0{,}5x dx$$

$0{,}25M^2=0{,}5$ $\quad M=1{,}41$

Математическо очакване

$$E(c) = c$$

$$EX = \int_{-\infty}^{\infty} x f(x) dx$$

$$E(cX) = cX$$

$$E(X+Y) = EX + EY$$

$$E(X.Y) = EX.EY \quad \textit{ако } X \textit{ и } У \textit{ са независими}$$

Дисперсия

$$\sigma^2(c) = 0$$

$$\sigma^2(X) = E(X - EX)^2 = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx$$

$$\sigma^2(cX) = c^2 \quad \sigma^2(X)$$

$$\sigma^2(X+У) = \sigma^2(X) + \sigma^2(У) \quad \textit{ако } X \textit{ и } У \textit{ са независими}$$

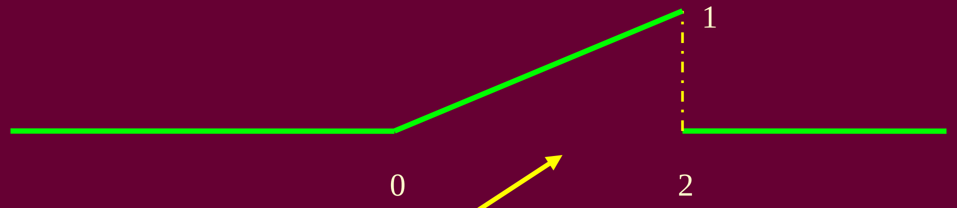
Стандартно отклонение

$$\sigma = \sqrt{\sigma^2}$$

Пример 1:

Плътността на една сл. в. е

$$f(x)=\begin{cases} 0{,}5x & \text{при } 0<x<2 \\ 0 & \text{в противен случай} \end{cases}$$



Намерете **средната стойност** на случайната величина

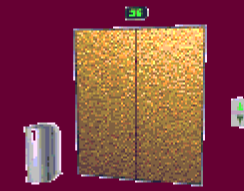
$$EX=\int_{-\infty}^{\infty} xf(x)dx=\int_{0}^{2} x(0{,}5x)dx=4/3=1.33$$

Намерете **дисперсията** на случайната величина

$$\sigma^2(X)=E(X-EX)^2=\int_{-\infty}^{\infty}(x-EX)^2 f(x)dx=\int_0^2 (x-4/3)^2(0{,}5x)dx=2/9$$

Намерете **стандартното отклонение** на случайната величина

$$\sigma(X)=\sqrt{\sigma^2}=\sqrt{\frac{2}{9}}=\frac{\sqrt{2}}{3}=0{,}4714$$



Равномерно разпределение

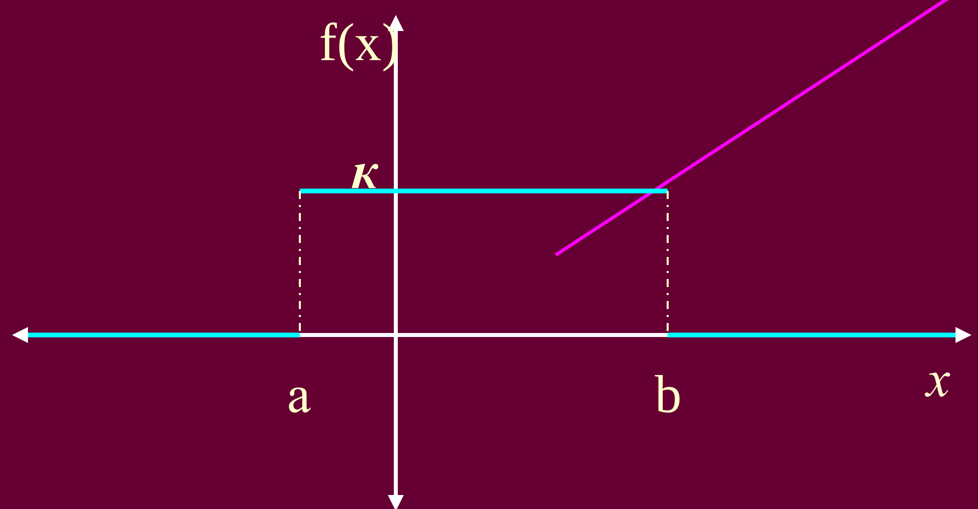
Определение

Случайна величина , чиято плътност е константа (различна от 0) върху даден интервал и нула извън този интервал, се нарича ***равномерно разпределена*** върху този интервал.

X е равномерно разпределена върху (а, в) ако има плътност

$$f(x)=\begin{cases} \kappa & \text{при } a\le x\le b \\ 0 & \text{в прот. сл.} \end{cases}$$

Графика на плътността на равномерно разпределение върху (а,в)



Лицето=1

$$k=\frac{1}{b-a}$$

Основни характеристики

Функция на разпределение

$$F(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a}, & x\in[a,b]\\ 1 & x>b\end{cases}$$

Средна стойност

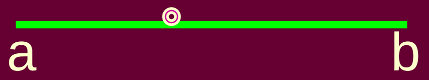
$$EX=\int_{-\infty}^{\infty}xf(x)dx=\int_a^b x\frac{1}{b-a}dx=\frac{a+b}{2}$$

$$\textit{Дисперсия}=\frac{(b-a)^2}{12}$$

мода ⇒ НЯМА

медиана ⇒ Среда на интервала

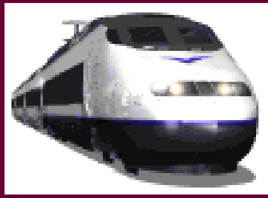
Примери



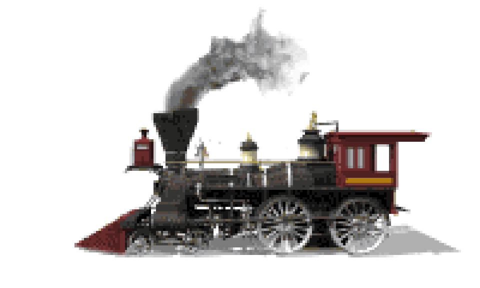
Опит: Избор на случайно число от интервала (a,b).
Нека X={избраното число}

Опит: Клиенти на банка пристигат по случаен начин на гишето.
X={времето на пристигане на клиента}

Опит: Влакчетата на метрото се движат на интервал през 10 минути. Всеки пътник пристига на спирката в случайно време.
X={времето, което пътник чака до идването на следващото влакче}



ПРИМЕР: Влаковете за град А на определена гара пристигат през 45-минутни интервали, започвайки сутрин в 7 часа.



Пътник пристига в случайно време на гарата

Времето, което пътникът чака следващия влак

Случайна величина

Равномерно разпределена в интервала (0,45)

плътност

$$f(x) = \begin{cases} 0 & \textit{при} \quad x < 0 \;\textit{или}\; x > 45 \\ \dfrac{1}{45} & \textit{при} \quad 0 \le x \le 45 \end{cases}$$

Функция на разпределение

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{45}, & x \in [0,45] \\ 1 & x > 45 \end{cases}$$

намерете вероятността пътникът да чака по-малко от 5 минути следващия влак

P(X<5)=5/45

намерете вероятността пътникът да чака повече от 20 минути

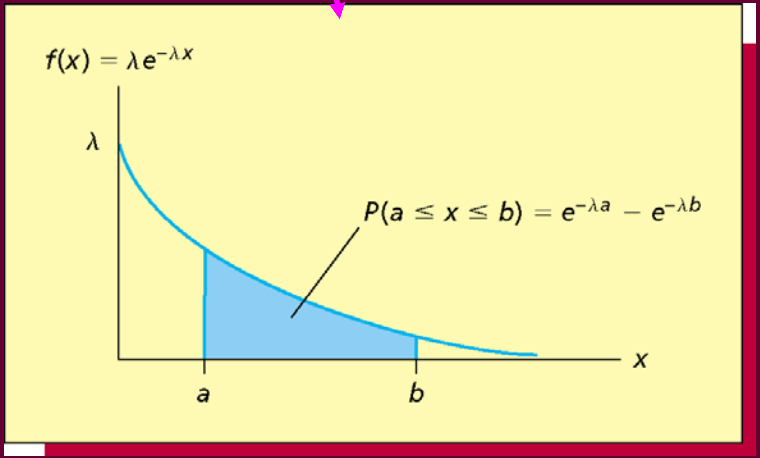
P(X>20)=25/45

Експоненциално разпределение

Случайна величина , чиято плътност $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$
се нарича ***експоненциално разпределена.***

Свойства

Стойности на случайната величина= $(0,\infty)$



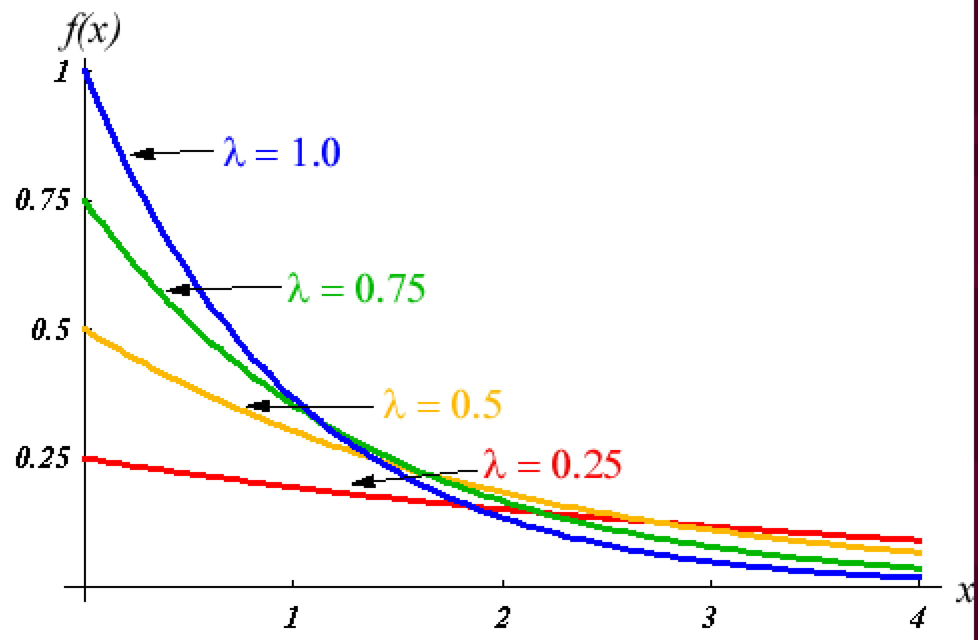
Вероятността намалява, когато интервала се премества надясно от 0, т.е. ако по абсцисната ос се нанася времето, то с течение на времето вероятността намалява

Приложение

Време на безотказна работа на апаратура

Време до настъпването на определено събитие

Графики на плътност на
експоненциално разпределение



Функция на разпределение

$$F(x)=\int_{-\infty}^{x} f(s)ds=\int_{0}^{x}\lambda e^{-\lambda s}ds=-e^{-\lambda s}\,|^{x}{}_{0}=1-e^{-\lambda x}$$

Функция на разпределение

$$F(x)=\begin{cases}1-e^{-\lambda x} & x\geq 0\\ 0 & x<0\end{cases}$$

Средна стойност

$$EX=\int_{-\infty}^{\infty} sf(s)ds=\int_{0}^{\infty} s\lambda e^{-\lambda s}ds=-xe^{-\lambda x}\Big|_{0}^{\infty}+\int_{0}^{\infty} e^{-\lambda s}ds=0-\frac{e^{-\lambda x}}{\lambda}\Big|_{0}^{\infty}=\frac{1}{\lambda}$$

Дисперсия

$$\sigma^2(X)=\frac{1}{\lambda^2}$$

Стандартно отклонение

$$\sigma(X)=\frac{1}{\lambda}$$

Пример

Средно електрическите крушки от даден вид издържат по 10 часа непрекъснато светене (преди да изгорят)

Времето на безотказна работа на крушката (до нейното изгаряне)

Случайна величина

Експоненциално разпределение с EX=10

λ=0.1

$$F(x)=\begin{cases}1-e^{-0,1x} & , \quad x\geq 0\\ 0 & x<0\end{cases}$$

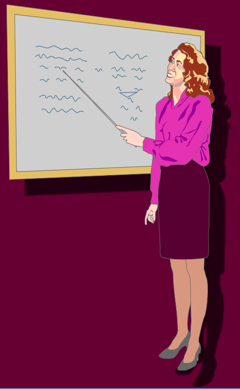
Нека студент влиза в стаята си , сменя изгоралята крушка и започва да чете за изпит 5 часа. Каква е вероятността, той да чете без да му се налага да сменя крушката?

P(X≥5)=1-P(X<5)=1-F(5)= $e^{-\frac{5}{10}}$ =.607

Пример

Нека експоненциално разпределение има плътност

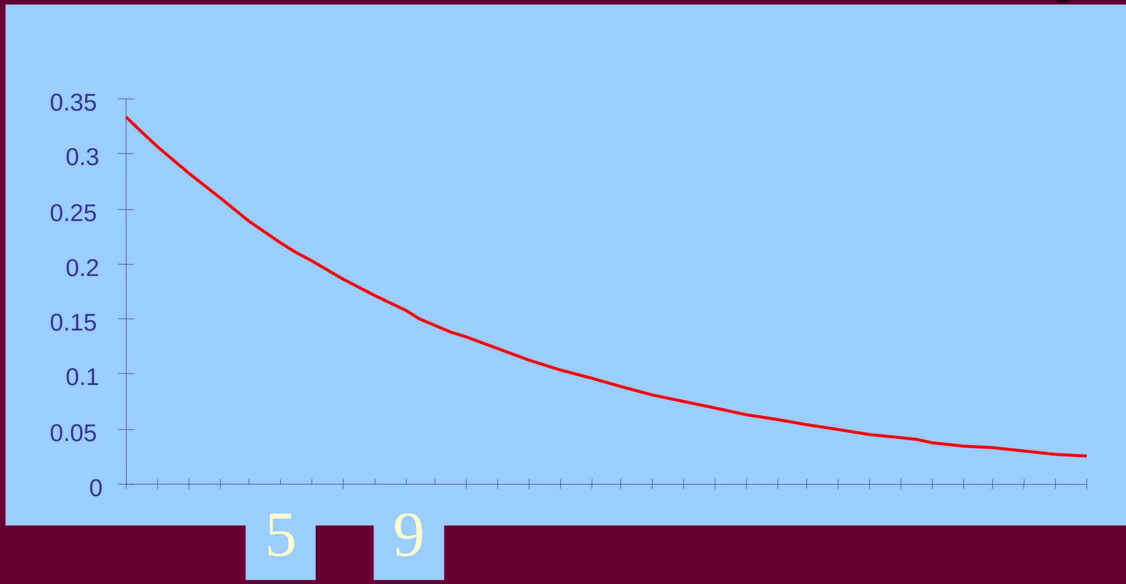
$$f(x) = \begin{cases} 0{,}34e^{-0{,}34x} & при \;\; x \geq 0 \\ 0 & при \;\; x < 0 \end{cases}$$



Начертайте графиката на плътността

Намерете средната стойност

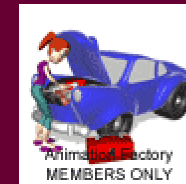
EX=1/0,34=2,9



Намерете функцията на разпределение

$$F(x) = \begin{cases} 1 - e^{-0{,}34x} & при \;\; x \geq 0 \\ 0 & при \;\; x < 0 \end{cases}$$

Пример



Допускаме, че пробегът на една кола(в км), докато акумулаторът й се изтощи е експоненциално разпределен със средна стойност 20 000 км. Ако човек решава на направи обиколка на Европа за около 7 000 км, то каква е вероятността да не му се налага да сменя акумулатор по пътя. Какво можем да кажем ако разпределнието не е експоненциално

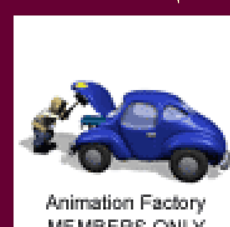
λ = 1/20 =0,05 (в хил. км)

$$f(x) = \begin{cases} 0{,}05e^{-0{,}05x} & при \ \ x \geq 0 \\ 0 & при \ \ x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-0{,}05x} & при \ \ x \geq 0 \\ 0 & при \ \ x < 0 \end{cases}$$

P(времето >7)=1-F(7)= $e^{-0{,}05(7)} = e^{-0{,}35} = 0{,}7$

Ако разпределението не е експоненциално, то ни е необходима допълнителна информация



Общи задачи

Случайна величина X има плътност

$$f(x)=\begin{cases}0 & при \quad x\le 0 \quad или \quad x\ge 2\\ cx^2 & при \quad 0<x<2\end{cases}$$

Намерете стойността на с. C=0,375

Намерете средната стойност на X.

$$EX=\int_{-\infty}^{\infty} xf(x)dx=\int_0^2 x(0{,}375x^2)dx=1.5$$

Намерете дисперсията и стандартното отклонение на X.

$$\sigma^2(X)=\int_0^2 (x-1{,}5)^2(0{,}375x^2)dx=0{,}15$$

0,3873

Намерете P(X<0,5).

$$\int_0^{0{,}5}(0{,}375x^2)dx=0{,}016$$

Намерете медианата M на X.

$$\int_0^{M}(0{,}375x^2)dx=0{,}5 \qquad \frac{0{,}375M^3}{3}=0{,}5$$

M=1,59

Намерете модата на X

Мода=2 => виж графиката на плътността

Знае се, че средната продължителност на безотказна работа на електрическата система на автомобил е експоненциално разпределена със средна стойност 1000 часа.

Каква е вероятността, ако сменим тази компонента, то в следващите 600 часа на работа на системата да не се налага тази част отново да се сменя?

EX=1/λ=1000=> λ=0,001

$$F(x)=\begin{cases}1-e^{-0,001x} & при \;\; x\geq 0\\ 0 & при \;\; x<0\end{cases}$$

P(X>600)=1-P(X<600)=1-F(600)= $e^{-0,001(600)}=e^{-0,6}=0,55$

Пример 2:

Бензиностанция се зарежда един път седмично. Обемът на седмичните продажби (в хил. тонове) е случайна величина в плътност

$$f(x)=\begin{cases} 5(1-x)^4 & \text{при } 0<x<1 \\ 0 & \text{в противен случай} \end{cases}$$

Намерете стойността на константата *c* $\Rightarrow \int_{-\infty}^{\infty} f(s)\,ds = 1 \Rightarrow \int_0^1 c(1-x)^4\,dx = 1 \Rightarrow \frac{c}{5} = 1$

Каква е вероятността седмично да се продават повече от 500 тона?

500=0,5 (хил. тона) P(X>0,5)= $\int_{0,5}^{1} 5(1-x)^4\,dx = 0{,}5^5 = 0{,}03125$

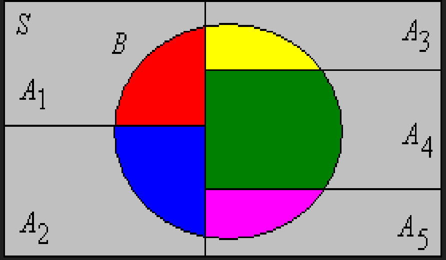
Колко трябва да е капацитетът на резорвоарите, така че вероятността да бъде изчерпвана газта седмично да е 0,01?

a=? че P(X≥a)=0,01 $\int_{a}^{1} 5(1-x)^4\,dx = (1-a)^5 = 0{,}01$ ⇒a=0,602 или 602 тона

Нека е даден опит с пространство S и събития A1, A2, ...An такива, че

- *те са несъвместими*
- *сумата им дава цялото пространство*

Пълна група



Всяко събитие и неговото допълнение са пълна група

Формула на пълната вероятност

Нека е даден опит с пространство S и **пълна група** от събития $A_1, A_2, \ldots A_n$. Нека B е друго събитие в S

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$

Доказателство

$$S = \bigcup_{k=1}^{n} A_k \Rightarrow B = B \cap S = \bigcup_{k=1}^{n} B \cap A_k \Rightarrow P(B) = \sum_{k=1}^{n} P(B \cap A_k)$$

$$P(B \cap A_k) = P(A_k)P(B \mid A_k) \Rightarrow P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$

пример

На масата има 20 непрозрачни плика, от които единият съдържа 20 лв. Двама студенти един след друг избират по един плик. Кой има по-голям шанс да избере печелившия плик – първия или втория избиращ студент?

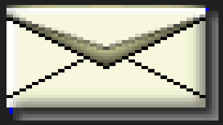
A={първия избира печелившия плик}
B ={втория избира печелившия плик}

P(A)= 1/20

P(B)=???

Разглеждаме събитията A и $\bar{A}$ – те образуват пълна група

$$P(B)= P(B|A)P(A)+ P(B|\bar{A})P(\bar{A})=$$
$$(0)*(1/20) + (1/19)*(1-1/20)= 1/20$$



Шансът на първия да избере печелившия плик = шанса на втория

Формула на Бейс

Нека е даден опит с пространство S и **пълна група** от събития $A_1, A_2, ...A_n$. Нека B е друго събитие в S

$$P(A_j \mid B) = \frac{P(A_j)P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

Доказателство

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{P(B)}$$

$P(A_1|B)=???$

В една група 60% са момичета. При това 30% от момичетата са от Пловдив, а 50% от момчетата са от Пловдив. Избран е един студент, който се оказва от Пловдив. Каква е вероятността студентът да е момиче?

A_1= момиче $P(A_1)=0,6$

A_2= момче $P(A_2)=0,4$

B=от Пловдив

$P(B|A_1)=0,3$ $P(B|A_2)=0,5$

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$P(A_1 \mid B) = \frac{0,6(0,3)}{0,6(0,3) + 0,4(0,5)} = 0,47$$

Пример:

Нека след провеждане на избори в Б-я се знае, че 50% от гласуващите в област А са гласували за партията “НБ-напред в бъдещето”, 60% от гласуващите област Б са гласували за “НБ” и 35 % от област С са гласували за “НБ”. От друга страна се знае, че само 40% от жителите на А са гласували, само 25% от жителите на Б са гласували, и само 35% жителите на С са гласували.

Ако случайно избран жител е гласувал за “НБ”, то каква е вероятността той да е жител на областта Б?

B=Избирателят е гласувал за “НБ”

A1=избирателят е от област А

A2=избирателят е от област Б

A3=избирателят е от област С

P(B|A1)=0,5 P(A1)=0,4

P(B|A2)=0,6 P(A2)=0,25

P(B|A3)=0,35 P(A3)=0,35

P(A2|B)=???

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3)}$$

$$P(A_2 \mid B) = \frac{(0,60)*(0,25)}{(0,50)*(0,40) + (0,60)*(0,25) + (0,35)*(0,35)} = 0,3175$$

Опити на Бернули

1. **Съвкупност от краен брой *n* опити**

2. Опитите са независими.

3. **Всеки опит трябва да има само два възможни изходи, успех У и неуспех Н.**

4. Вероятността за успех във всеки отделен опит е постоянна:

P(У)=p

Всеки изход от *n* **опити на Бернули е наредена** *n* –торка от У и Н.

Колко е вероятността да има точно **к** У (успеха)

Тъй като опитите са независими, то вероятността на всеки отделен опит е

$$p^k(1-p)^{n-k}$$

Броят на всички изходи с точно **к** У е

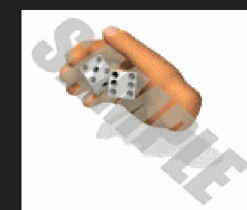
$$C_n^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$$

Вероятността да се наблюдават **k** успехи в **n** опита на Бернули:

$$P(S_n = k) = C_n^k p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

Кои от следните са опити на Бернули?

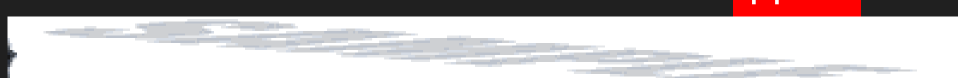
Играч хвърля двойка зарчета 10 пъти последователно. Да



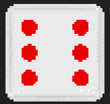
Избор на 5 студента по случаен начин измежду група от 20 студента.

изборът е без връщане не

Изборът е с връщане да



Пример



Зарче се подхвърля 7 пъти на масата. Каква е вероятността точно 5 пъти да се падне “шестица”?

7 Бернулиеви опита; Успех=6; Р(У)=1/6 ; точно 5 успеха

n=7 **k=5**

$$P(S_7 = 5) = C_7^{\,5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^{7-5} = 0{,}00188$$

Каква е вероятността поне 5 пъти да се падне “шестица”

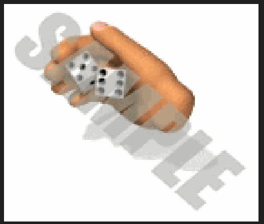
n=7 **k≥5**

$$P(S_7 \geq 5) = P(S_7 = 5) + P(S_7 = 6) + P(S_7 = 7) =$$

$$C_7^{\,5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^2 + C_7^{\,6}\left(\frac{1}{6}\right)^6\left(\frac{5}{6}\right)^1 + C_7^{\,7}\left(\frac{1}{6}\right)^7\left(\frac{5}{6}\right)^0 =$$

$$0{,}00188 + 0{,}000125 + 0{,}0000036 = 0{,}002$$

Двойка зарчета (бяло и червено) се подхвърлят 7 пъти на масата. Каква е вероятността точно 5 пъти да се паднат еднакъв брой точки?



7 Бернулиеви опита; Успех=да се паднат еднакъв брой точки ; точно 5 успеха

p=P(У)=6/36=1/6

$$P(S_7 = 5) = C_7^{\,5} (\frac{1}{6})^5 (\frac{5}{6})^{7-5} = 0,00188$$

Каква е вероятността точно 5 пъти сумата от точките на двата зара да е 8?

7 Бернулиеви опита; Успех=сумата от точките е 8 ; точно 5 успеха

p=P(У)=(5)/36

$$P(S_7 = 5) = C_7^{\,5} (\frac{5}{36})^5 (\frac{31}{36})^{7-5} = 0,0008$$

край

Тестване не хипотези

Хипотезата е недоказано твърдение за популационння параметър

Параметър е числова характеристика на популацията. Може да бъде:

- Популационна средна стойност,
- Популационна пропорция,
- Популационнна дисперсия

•*Пример:* Верно ли е, че средната седмична сума, която студентите дават за алкохол е 10 лв?

Обща идея за тестване на хипотези

Първо, формулираме две хипотези

Нулева хипотеза

- “Убеждението” което подлежи на тестване
- Винаги съдържа знаците : $=, \leq,$ или $\geq$
- Означение: H_0

Алтернативна хипотеза

- Противоположна на нулевата хипотеза
- Винаги съдържа знаците: $\neq, <,$ or $>$
- Означение H_1

Пример:

H_0: $\mu \geq 3$ H1: μ<3

Как да разграничим двете хипотези

- **Нулевата хипотеза** винаги предполага “няма смяна в сегашното положение”
- **Алтернативната хипотеза** е обикновено твърдението, което статистика се опитва да докаже.

Важно!!!

Заради теоретичните статистически методи, в нулевата хипотеза вместо знаците : $=$, $\leq$, или $\geq$ винаги се пише $=$

Тестване на хипотези

В статистиката винаги се предполага, че **нулевата хипотеза е верна**.

Тогава, решението се основава на съществуването на достатъчно основание

- Ако има достатъчно основание, то **отхвърляме нулевата хипотеза**
- Ако няма достатъчно основание, **не отхвърляме нулевата хипотеза.**

Ако нулевата хипотеза се отхвърли, то понякога на практика се казва, че резултатите потвърждават алтернативната хипотеза. Но това от статистическа гледна точка не е верно.

ВАЖНО

- В статистиката никога не можем да **докажем** едно или друго твърдение.
- Можем само да заключим, че има **достатъчно основание** да отхвърлим едно или друго твърдение.
- В статистиката няма значение какво решение вземаме, винаги има **шанс от грешки!**

Грешки при тестването на хипотези

	верно	
извод	**Нулева хипотеза**	**алтернатива**
Не отхвърляме нулевата хипотеза	OK	Грешка от втори род
Отхвърляме нулевата	Грешка от първи род	OK

Определение: Видове грешки

- **Грешка от I род**: Нулевата хипотеза се отхвърля, когато е верна.
- **Грешка от II род**: Нулевата хипотеза не се отхвърля, когато тя не е верна.
- Винаги има шанс от допускане на една от тези грешки. => как да минимизираме шанса от тяхното допускане!

Основни идеи за тестване

Първи начин:

Критична област ⇒ **(Област на отхвърляне):**

Идея: да намалим грешката от I род:

- **Грешка от I род**: Нулевата хипотеза се отхвърля, когато е верна.

Ниво на значимост

- **Означение** α
- **Избираме α** = P(грешка от I род) = P(H_0 да се отхвърли, ако H_0 е верна) => **Типични стойности 0,01 0,05 0,1**
- **Построяваме област на отхвъляне (критична област), така че лицето над нея под кривата на плътността =**α

Използване на p-стойност за проверка на хипотези

- **p-стойността** представя колко възможно е да се наблюдават такива екстремни извадки, ако нулевата хипотеза беше вярна.
- p-стойността е вероятност, т.е. това е число между 0 и 1.
- Близко до 0 означава “невъзможно”
- И така, ако p-стойността е “малка” (типично, по-малко от 0,05), тогава отхвърляме нулевата хипотеза.

Основни идеи

H_0: $\mu = 12$

H_1: $\mu < 12$

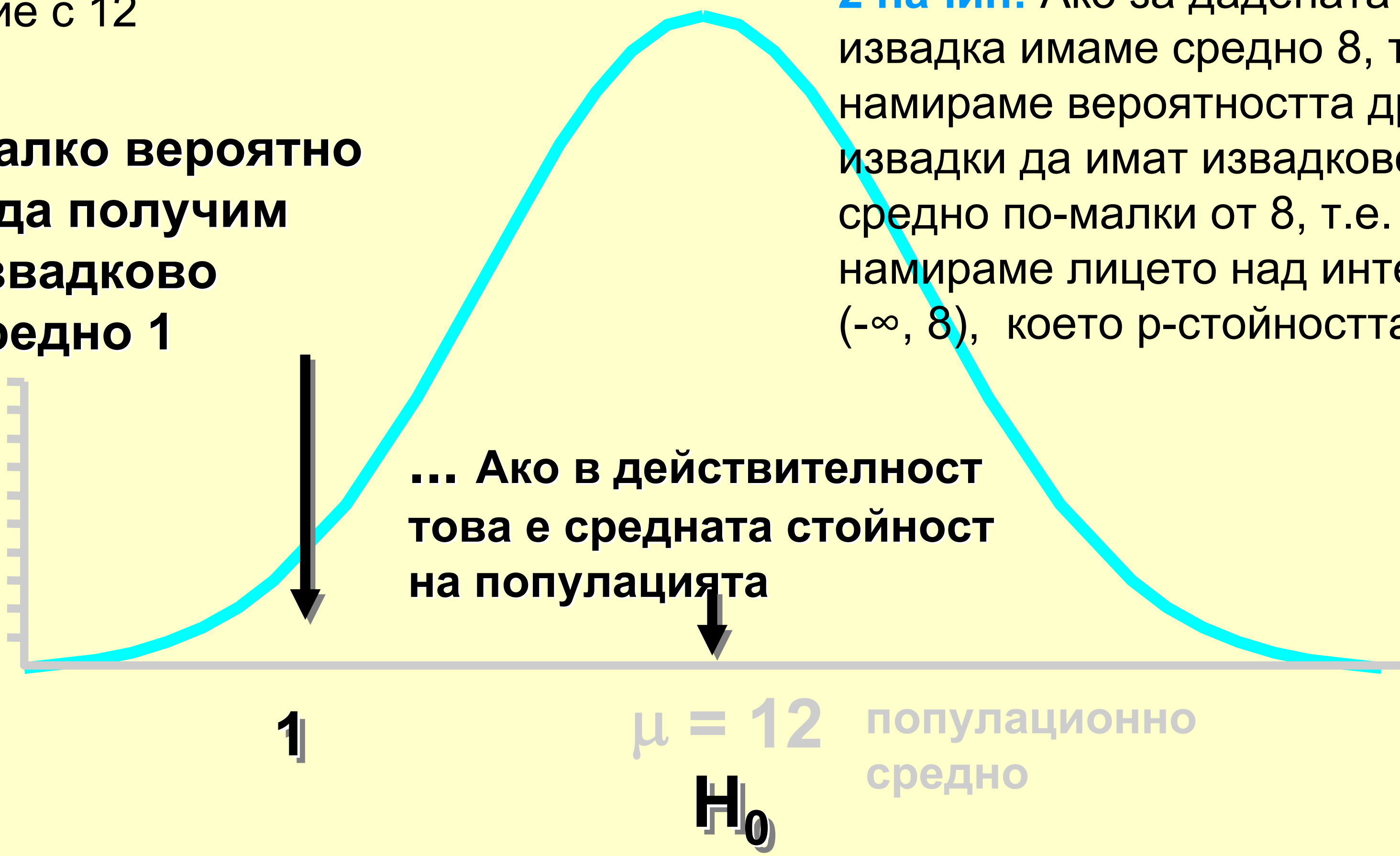
Ще отхвърлим нулевата, ако извадковото средно е много малко в сравнение с 12

1 начин: Построяваме независимо от извадката област от вида (-∞, a), която е област на отхвърляне, т.е. лицето над този интервал е избраното от нас число α

2 начин: Ако за дадената извадка имаме средно 8, то намираме вероятността други извадки да имат извадково средно по-малки от 8, т.е. намираме лицето над интервала (-∞, 8), което p-стойността

Малко вероятно е да получим извадково средно 1

… Ако в действителност това е средната стойност на популацията



Тестване на хипотези относно популационната средна стойност μ

Видове тестове

$$H_0: \mu = константа$$

Зависи от вида на алтернативната хипотеза

Едностранен тест

$$H_1: \mu < константа$$

Лявостранен тест

или

$$H_1: \mu > константа$$

дясностранен тест

Двустранен тест

$$H_1: \mu \neq константа$$

еквивалентно на

H_1: μ > константа

μ < константа

Два начина

Критична област

р-стойност

Стъпки при тестване на хипотези

критична област

1. Напишете нулевата H_0 и алтернативната H_1 хипотеза

2. Определете нивото на значимост α

3. Определете статистиката и извадковото разпр.

4. Получете критичната област

5. Направете извод

6. Интерпретация на извода

Тестване не хипотези за неизвестното популационно средното μ (σ е известно)

Предположения

– Популацията е нормално разпределена

– Ако не е нормална,то при голям обем на извадката (ЦГТ) можем да апроксимираме с нормална ($n \geq 30$)

- Популационната дисперсия (станд.откл.) σ^2 е известно

Да припомним, че $\bar{X} \in N(\mu, \frac{\sigma^2}{n})$

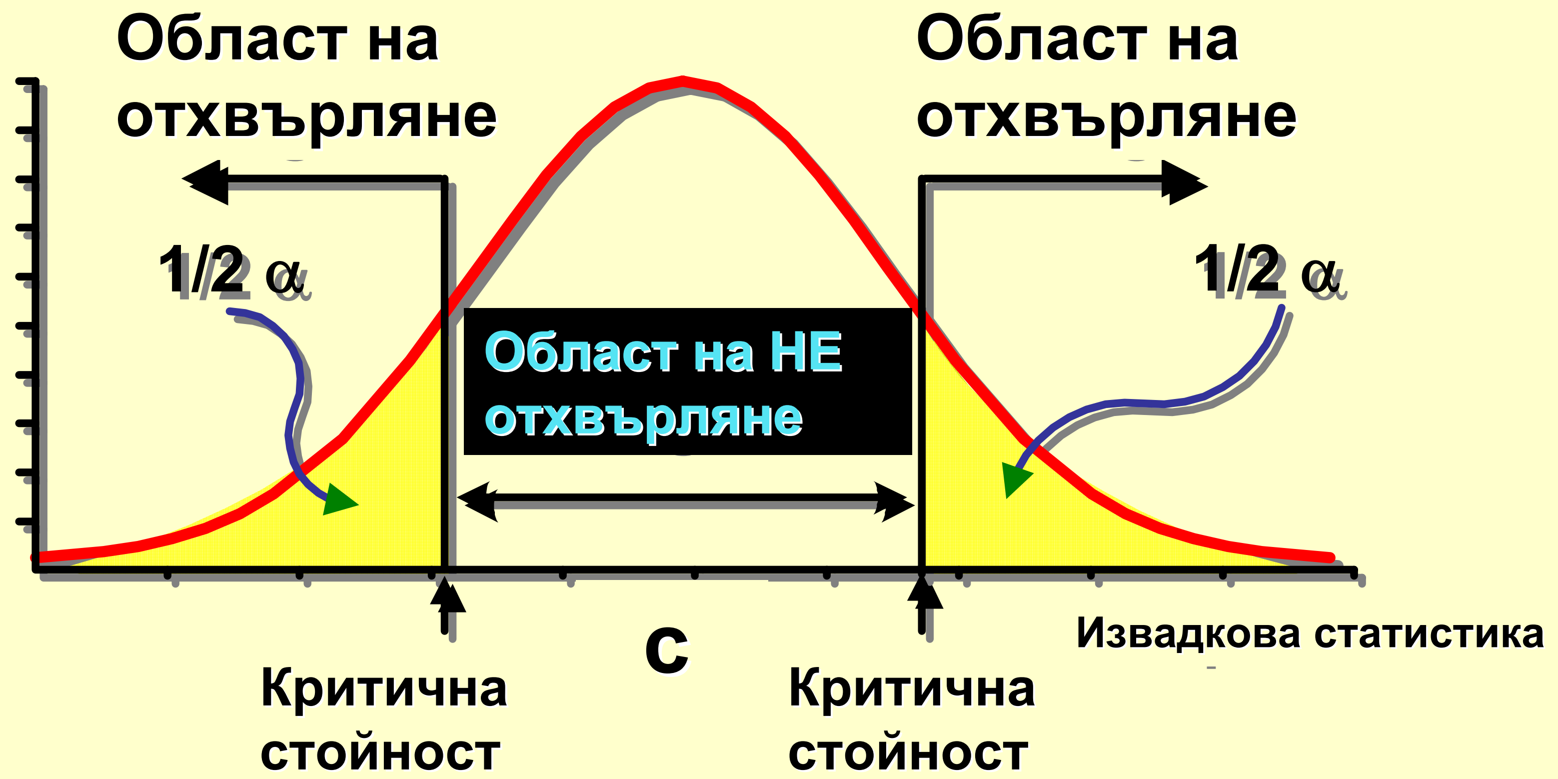
Използваме Z-статистиката

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Двустранен Z –тест за средното μ
(σ Известно)

H_0: $\mu = c$ $\qquad$ H_1: $\mu \neq c$ (означава < или >)

Избираме ниво на значимост α



Критична област при двустранен Z – тест за средното μ (σ Известно)

Използваме избраното от нас ниво на значимост α, за да намерим

$$z_\alpha : \qquad P(Z > z_\alpha) = \alpha$$

Критичната област се състои от два интервала (два клона)

(- ∞, - z_α) и (z_α , +∞).

Как се взема решение

Ако статистиката $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$ за конкретната извадка

- има стойност в критичната област – то отхвърляме H_0
- ако няма стойност в КО – няма основание да отхвърлим H_0

•Дали средно всяка кутия съдържа 368 грама?

•Направена е случайна извадка от 25 кутии, които са претеглени и е получено средното им тегло : 372,5 грама.

•Компанията знае от предварителни наблюдения, че теглото е нормално разпределено със стандартно отклонение 15 грама. Тествайте с ниво на значимост 0,05.

1. Напишете нулевата и алтернативната хипотеза

$H_0: \mu = 368$ $H_1: \mu \neq 368$ **Двустранен тест**

2. Ниво на значимост $\alpha = 0,05$

3. Статистика и извадково разпределение

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372,5 - 368}{\frac{15}{\sqrt{25}}} = +1,50$$

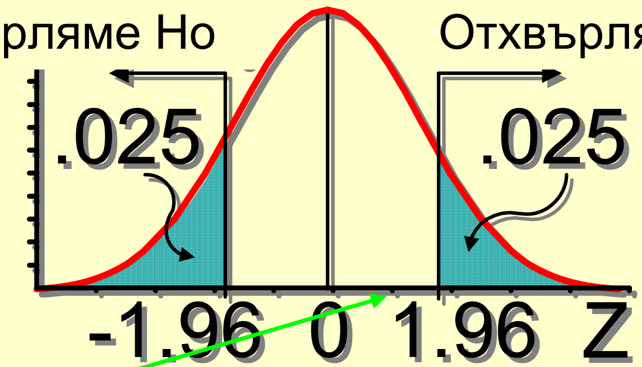
теглото е нормално разпределено със стандартно отклонение 15, т.е. σ=15

4. Критична област

Критичната област е

(- ∞, - 1,96) и (1,96, +∞).

$\alpha = \mathbf{0{,}05}$ $\alpha/2 = \mathbf{0{,}025}$



5. Извод

$z = 1{,}50$ не е в критичната област, следователно не отхвърляме H_0

6. Интерпретация

Няма достатъчно основание да отхвърлим твърдението на компанията

че средното тегло е 368

Едностранен тест за средната стойност μ (σ известно)

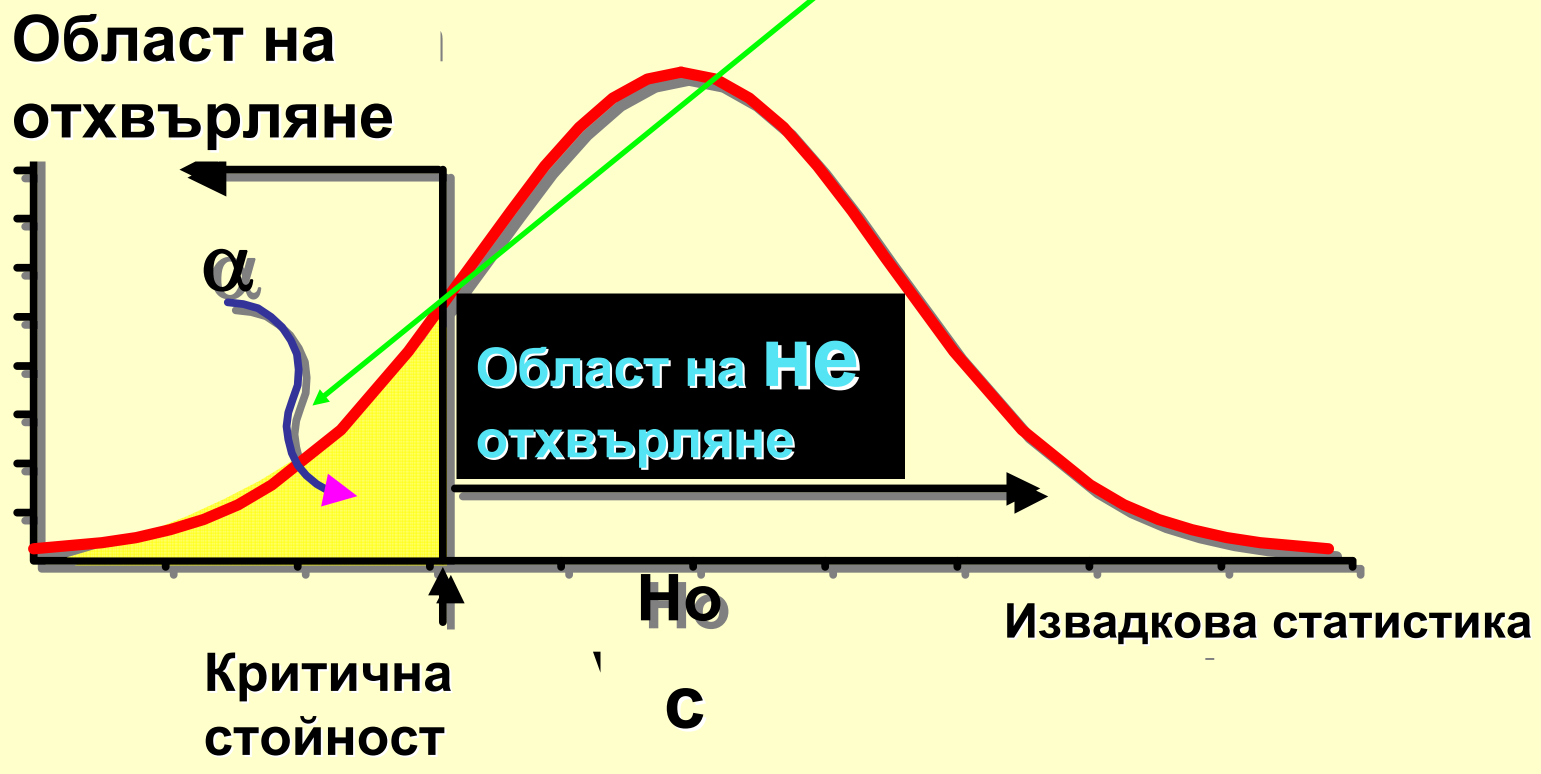
- **Предположения**
  - **Популацията е нормално разпределена**
  - **Ако не е нормална, можем да я апроксимираме с нормална (*n* ≥ 30)**
- **Алтернативната хипотеза има знаците $<$ или $>$**
- **Z-статистика**

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Лявостранен тест

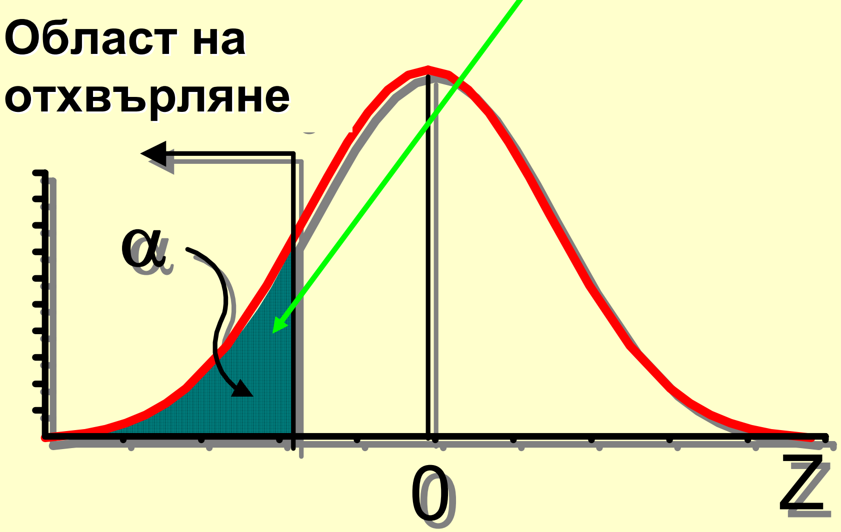
$$H_0: \mu \geq c$$

$$H_1: \mu < c$$

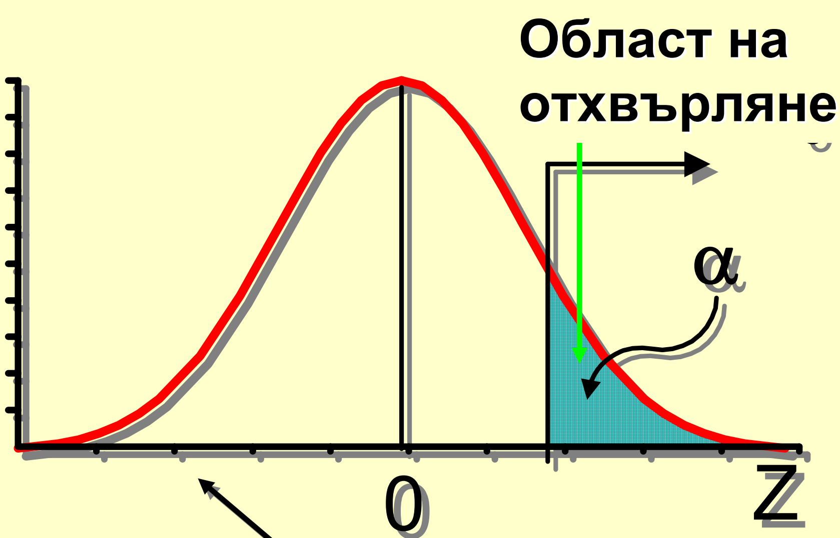


Едностранен Z тест за средна стойност

H_0: $\mu \geq 0$ H_1: $\mu < 0$



H_0: $\mu \leq 0$ H_1: $\mu > 0$



Малки стойности не противоречат на H_0 - не се отхвърля!

Дали средно всяка кутия съдържа повече от 368 грама? Случайна извадка от 25 кутии има тегло 372,5 гр. Компанията знае, че стандарното отклонение на теглото в кутиите е 15 грама. Тествайте твърдението при ниво на значимост 0,05 като предположите, че популацията е нормална.

1. Нулева и алтернативна хипотеза

H_0: $\mu = 368$ **H_1: $\mu > 368$** **Дясностранен тест**

2. Ниво на значимост

$$\alpha = 0{,}05$$

3. Статистика и извадково разпределение

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372{,}5 - 368}{\frac{15}{\sqrt{25}}} = +1{,}50$$

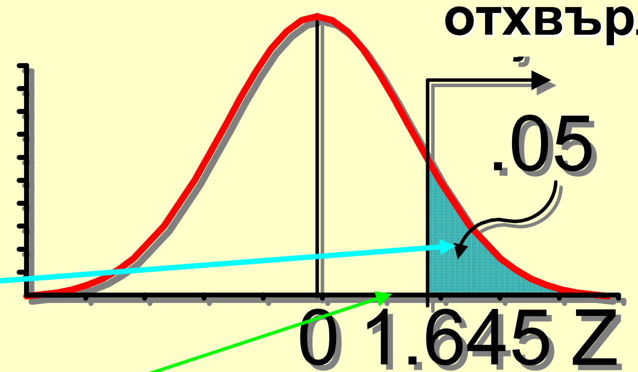
Нормална популация, σ= 15

4. Критична област

Критичната област е

(1,645, +∞).

α= **0,05**



5. Извод

z = 1,50 не е в критичната област, затова не отхвърляме H_0

6. Интерпретация

Няма достатъчно основание да отхвърлим твърдението

че средното е повече от 368

Проверка на хипотези за средната стойност μ, когато популационната дисперсия е неизвестна (σ неизвестно)

Предположения

– Популацията е нормално разпределена
– Малък обем (n≤30)

■Използваме t -разпределение и следната статистика

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

t-тест за неизвестното популационно средно

За да се провери дали машината за пълнене е точно настроена и пълни кутиите по 368 грама, се прави случайна извадка от 16 кутии. За тези кутии е намерено, че средното им тегло е 372,5 грама и стандартното отклонение е 12 грама. Знае се, че теглото на кутиите е нормално разпределено. Тествайте твърдението при ниво на съгласие 0,05 .

1. Нулева и алтернативна хипотеза

$H_0: \mu = 368$ $H_1: \mu \neq 368$ **Двустранен тест**

2. Ниво на значимост

$\alpha = 0{,}05$

3. Статистика, извадково разпределение

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{372{,}5 - 368}{\frac{12}{\sqrt{16}}} = +1.5$$

Използваме t-разпределение

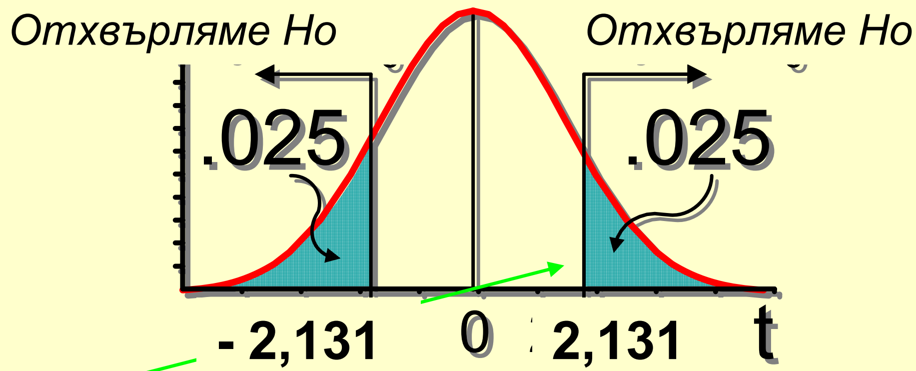
нормално разпределено тегло, σ е неизвестно

4. Критична област

Критичната област е

(- ∞, - 2,131) и (2,131, +∞).

Използваме t (15)



5. Извод

t=1,5 не попада в критичната област, затова не отхвърляме хипотезата

6. Интерпретация на извода

Няма достатъчно основание за да отхвърлим твърдението, че средното тегло е 368

Дали средният капацитет на батерии е поне 140 ампер-часа?

Случайна извадка от 20 батерии има средно 138,47 и стандартно отклонение 2,66. Допускаме, че капацитетът е нормално разпределен.

Тествайте твърдението с ниво на съгласие 0,05.

1. Нулева и алтернативна хипотеза

$H_0: \mu = 140$ **$H_1: \mu < 140$** **Лявостранен тест**

2. Ниво на значимост

$$\alpha = 0{,}05$$

3. Статистика и извадково разпределение

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{138{,}47 - 140}{\frac{2{,}66}{\sqrt{20}}} = -2{,}57$$

Нормална популация,

σ е неизвестно, n=20

Използваме t- разпределение

Степени на свобода = 20 - 1 = 19

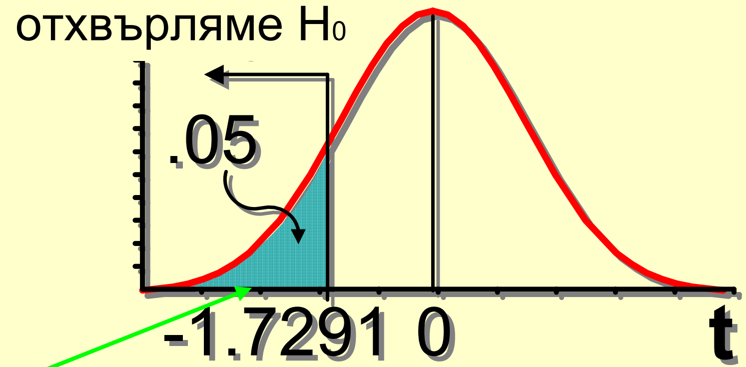
4. Критическа област

Критичната област е

(- ∞, - 1,7291)

$\alpha = \mathbf{0{,}05}$

Използваме t- таблицата



5. Извод

t= - 2,57 попада в критичната област, затова отхвърляме H_0

6. Интерпретация на извода

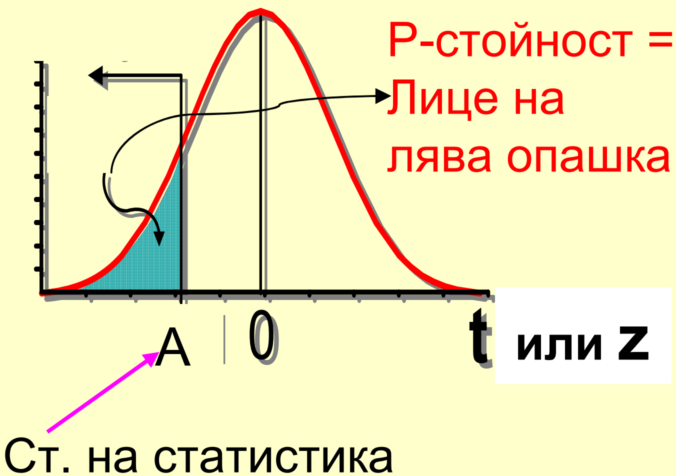
Имаме достатъчно основание да считаме, че капацитетът е по-малък от 140

Използване на p-стойност за проверка на хипотези

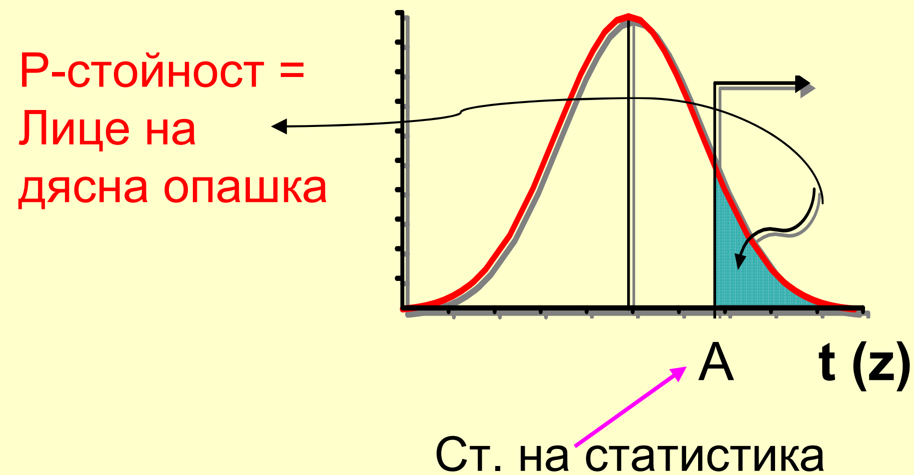
- За дадена извадка, **p-стойността** дава вероятността да се наблюдават и други екстремни извадки, ако нулевата хипотеза беше вярна.
- p-стойността е вероятност, т.е. това е число между 0 и 1.
- Близко до 0 означава “невъзможно”
- И така, ако p-стойността е “малка” (типично, по-малко от 0,05), тогава отхвърляме нулевата хипотеза.

Определяне за **р-стойност= Лице**

Едностранен, ляв тест

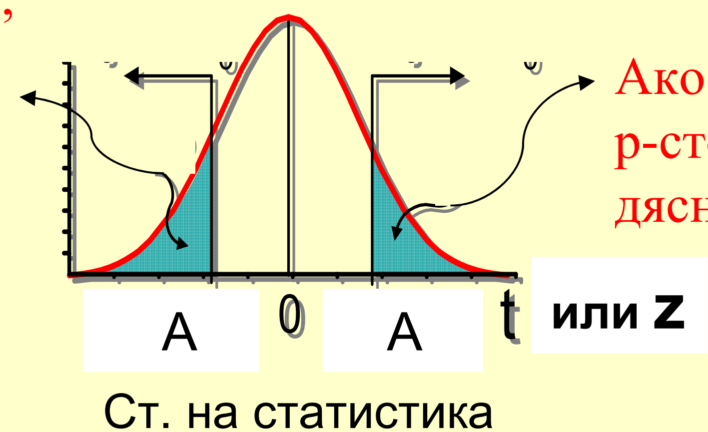


Едностранен, десен тест



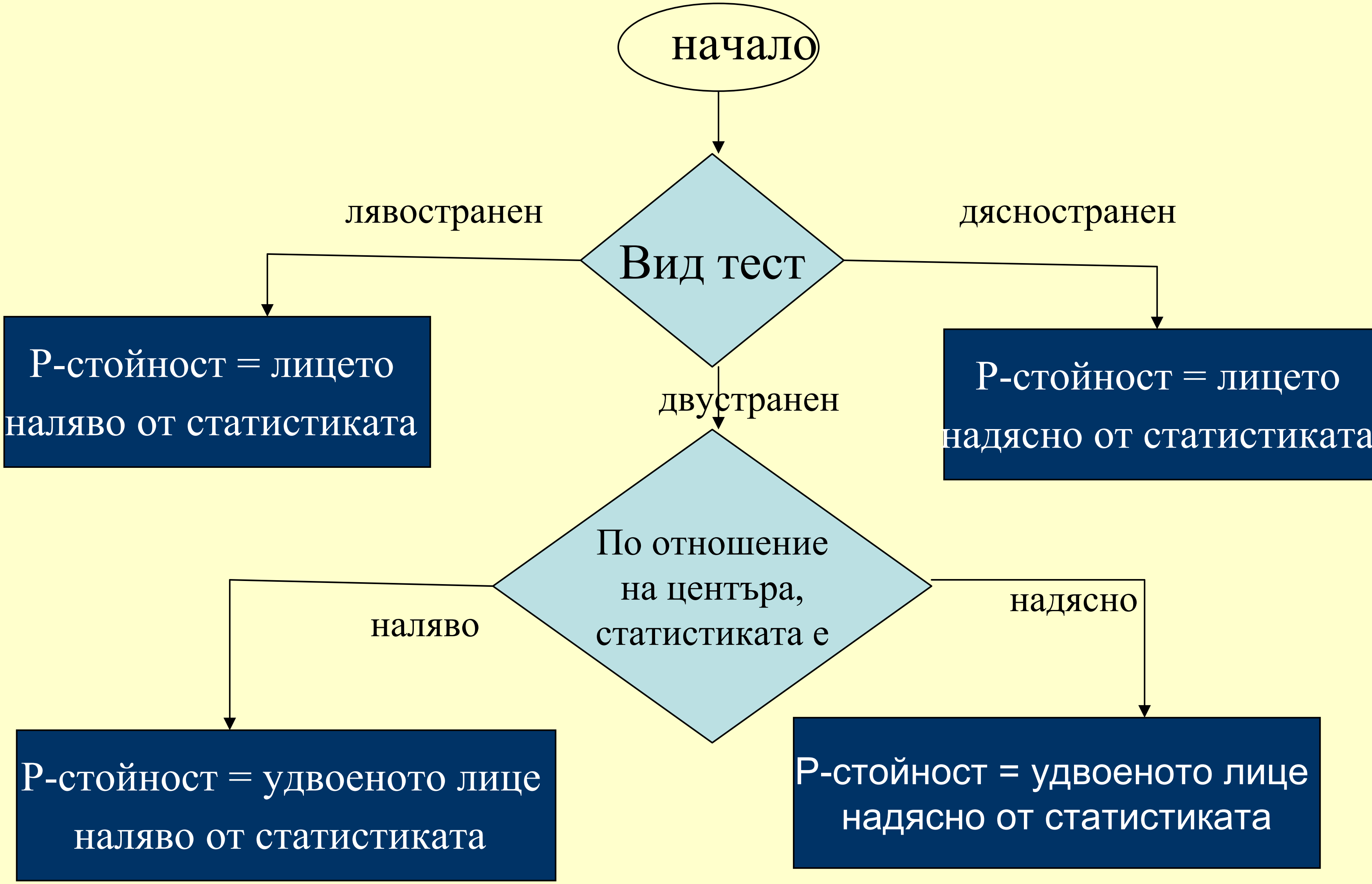
Двустранен тест

Ако *A* е отрицателно, р-стойност = 2* лицето на лявата опашка



Ако *A* е положително, р-стойност= 2* лицето на дясната опашка

Намиране на p-стойности



Отхвърляме H_0 ако стойността е малка

P-стойност	Интерпретация
По-малко от 0,01	**Голяма статистическа значимост. Много строги докателства срещу нулевата хипотеза**
От 0,01 до 0,1	**Статистически достатъчни доказателства срещу нулевата хипотеза.**
По-голямо от 0,1	**Недостатъчно основание за отхвърляне на нулевата хипотеза**

Стъпки при тестване на хипотези

(p-стойност)

1. Напишете нулевата и алтернативната хипотеза

2. Определете статистиката и извадковото разпределение

3. Пресметнете P-стойността за статистиката на теста

4. Направете извод

5. Дайте интерпретация на вашия извод

Дали средно всяка кутия съдържа 368 грама?

Направена е случайна извадка от 25 кутии, които са претеглени и е получено средното им тегло : 372,5 грама.

Компанията знае от предварителни наблюдения, че теглото е нормално разпределено със стандартно отклонение 15 грама.

1. Напишете нулевата и алтернативната хипотеза

H_0: $\mu = 368$ **H_1: $\mu \neq 368$** **Двустранен тест**

2. Статистика и извадково разпределение

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372,5 - 368}{\frac{15}{\sqrt{25}}} = +1,50$$

теглото е нормално разпределено със стандартно отклонение 15, т.е. σ=15

3. Пресметнете р-стойността на статистиката на теста

Доколкото z=1,5>0, лицето надясно от z от таблицата е 0,0668. Тъй като теста е двустранен (т.е. състои се от две опашки), ние удвояваме лицето и р=0,1336

4.Извод

Голямо р, не отхвърляме H_0

5. Интерпретиране на извода

Няма достатъчно основание, за да отхвърлим твърдението, че средно кутиите съдържат по 368 грама.

Пример

Дали средно всяка кутия съдържа повече от 368 грама? Случайна извадка от 25 кутии има тегло $\bar{X}$ = 372,5. Компанията знае, че σ е 15 грама. Тествайте твърдението при ниво на значимост 0,05 като предположите, че популацията е нормална.

Решение

1. Нулева и алтернативна хипотеза

H_0: $\mu = 368$ **H_1: $\mu > 368$** **Дясностранен тест**

2. Статистика и извадково разпределение

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372,5 - 368}{\frac{15}{\sqrt{25}}} = +1,50$$

Нормална популация, $\sigma = 15$

3. Пресметнете p-стойността на статистиката на теста

Доколкото z=1,5>0, лицето надясно от z от таблицата е 0,0668. Тъй като теста е дясностранен (т.е. състои се от дясна опашка), то p= 0,0668

4.Извод

Тъй като p= 0,0668<0,1, то има статистически достатъчни доказателства срещу нулевата хипотеза.

5. Интерпретиране на извода

Имаме основание да считаме, че средно кутиите са са по-тежки от 368 грама

Има ли връзка между ниво на значимост и p-стойност ?

Пример. Нека

$H_0: \mu = 140$ $\qquad$ $H_1: \mu < 140$

а статистиката е : $t = -1{,}7291$ с 20 степени на свобода

Тогава p=0,05 , тъй като лицето наляво от t = - 1,7291 <0 (от t-таблицата) е 0,05.

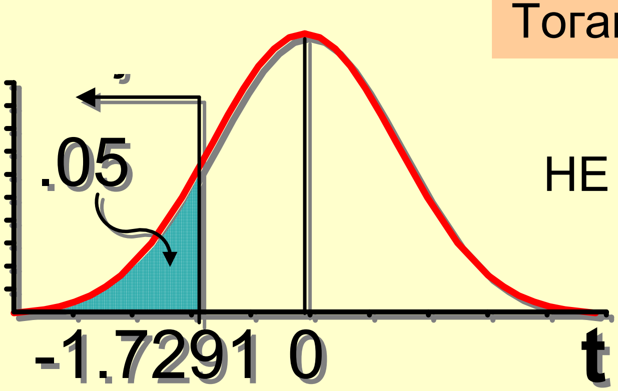
Да изберем ниво на значимост α = 0,1>0,05=p-стойността.

Тогава критичната област е (-∞, -1,325)

$t = -1{,}7291$ попада в КО-отхвърляме нулевата хипотеза, т.е. μ е по-малко от 140

Да изберем ниво на значимост α = 0,025<0,05.

Тогава критичната област е (-∞, -2,086)



$t = -1{,}7291$

НЕ попада в КО- НЕ отхвърляме нулевата хипотеза

Извод: H₀ се отхвърля с ниво на значимост α ≥ p-стойността