Biomeetria

Populatsioon, valim, punkt-hinnang.
Tõenäosuse mõiste.
Teoreetilised jaotused

Populatsioon *versus* valim

Üldkogum (**populatsioon**) on reaalselt olemasolev või ka abstraheeritud objektihulk, mille/kelle kohta soovitakse uurimistöö tulemusena sisulisi järeldusi teha (<u>peab olema enne uuringut fikseeritud nii ajas kui ka ruumis</u>).

Valim on teatava eeskirja järgi moodustatud hulk üldkogumisse kuuluvatest objektidest e mõõtmiseks valitud üldkogumi osa;
kui valim on piisavalt esinduslik, siis võime eeldada seal esinevate seaduspärade kehtimist enam-vähem sarnasel kujul ka üldkogumis.

<u>Lihtsaim reegel</u>: igal üldkogumi objektil on võrdne tõenäosus valimisse sattuda.

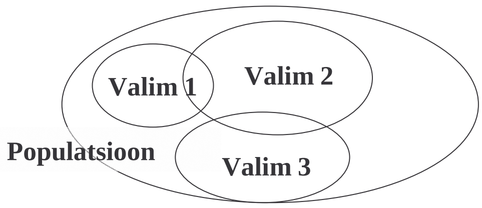
Näiteid:

	<u>Populatsioon</u>	<u>Valim</u>
1.	Eesti vetes kudevad lõhed	100 Eesti vetest kinni püütud ja ära mõõdetud lõhet
2.	Eesti maatõugu veised	Eesti maatõugu veised
3.	Eesti sigade pekipaksuse muutus ajavahemikul 1995-2005	34521 ajavahemikus 1995-2005 sooritatud sigade pekipaksuse mõõtmist

Populatsioon *versus* valim

Kõikne uuring: mõõdetakse kõiki üldkogumi objekte.
Valikuuring: mõõdetakse teatud väikest osa üldkogumist (valimit).

Statistika olemus: võtame teatud reeglite järgi osa üldkogumist (valimi), analüüsime seda ja teeme järeldusi kogu üldkogumi kohta!



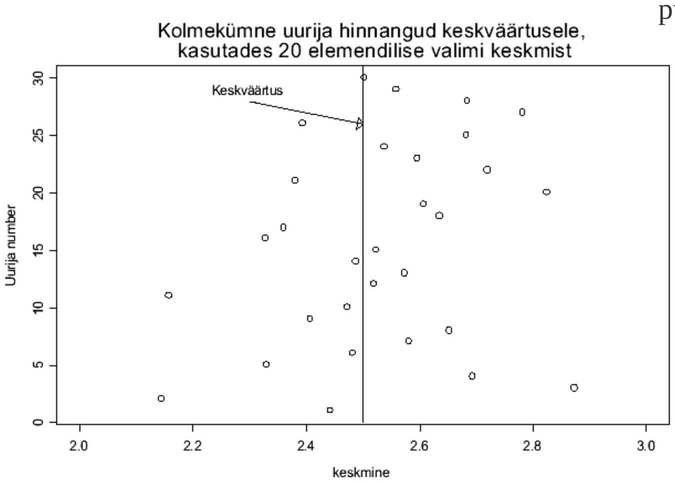
Populatsioon *versus* valim, punkthinnangud

Valimi alusel leitud arvkarakteristikud on vastavate populatsiooni parameetrite punkthinnangud.

Näiteks:

$\bar{x} = \hat{\mu},$

$s = \hat{\sigma}.$



Tõenäosus

Statistika valetab alati!
Küsimus on üksnes selles, kui palju või vähe.
Viimase leidmiseks on vaja tõenäosuse mõistet ning eeskirju selle leidmiseks.

Tõenäosus [*probability*] on sündmuse toimumise mõõt skaalal 0-st 1-ni.
Võimatu sündmuse tõenäosus on 0 ja kindla sündmuse tõenäosus on 1.

Tõlgendused:

tõenäosus – osakaal e protsent
(kui teatud haigus esineb 5%-l uuritavast populatsioonist, siis tõenäosus, et selle populatsiooni juhuslikult valitud esindaja on haige, on 0,05 ehk 5%);

tõenäosus – usk sündmuse või nähtuse võimalikkusesse.

Klassikaline tõenäosus

Juhuslik katse – katse, mille tulemus pole ette teada.
Juhuslik sündmus – juhusliku katse tulemus.

$$\text{Tõenäosus} = \frac{\text{Sündmuse jaoks soodsate katsetulemuste arv}}{\text{Kõigi katsetulemuste arv}}$$

Näide. Katseks on 20-tahulise täringu veeretamine, sündmuseks A on 10-ga jaguva silmade arvu saamine.

$$P(A) = 2 / 20 = 0{,}1 .$$

Šansid

Sündmuse toimumise šansid [*odds*] näitavad, mitmel juhul sündmus toimub võrreldes sellega, mitmel juhul ta ei toimu.

Näide. Kui sündmus toimub tõenäosusega 0,2 (20%) e ühel juhul viiest, siis selle sündmuse toimumise šansid on üks nelja vastu e 1:4.

Statistiline tõenäosus

Suurte arvude seadus: katseseeria lõpmatul pikenemisel läheneb sündmuse suhteline sagedus tema tõenäosusele.

Suhtelise sageduse kaudu leitud nn statistiline tõenäosus on klassikalise tõenäosuse hinnanguks (st, et ei ole konstant – muutub katseseeria pikenedes).

Näiteks veeretate 75 korda täringut ja saate 52 korda 6 silma. Antud täringuga 6 silma saamise tõenäosus on siis hinnatav suhtest

$$\hat{P}(6\ \text{silma}) = 52/75 \approx 0{,}693.$$

Näiteks, kui geneetikas räägitakse genotüübisagedusest populatsioonis, siis mõistetakse selle all tegelikult tõenäosust, et populatsioonist juhuslikult valitud isend on uuritava genotüübiga, leitud on see aga kui uuritava genotüübiga isendite suhteline sagedus valimis (katseks on siin juhusliku indiviidi valimine).

Statistiline tõenäosus

Teades mingi sündmuse tõenäosust (või ka selle tõenäosuse hinnangut), võime ennustada selle sündmuse esinemise suhtelist sagedust.

✓ Kui mingi sündmuse tõenäosus on 0,1, siis tähendab see seda, et keskmiselt kümne katse korral see sündmus esineb üks kord. Muidugi ei välista see teadmine ka võimalust, et sündmus esineb esimesel korral.

✓ Sündmus, mille tõenäosus on 0,001, on vähetõenäoline: ta esineb üks kord tuhande katse kohta. Kuigi niisuguse tõenäosusega sündmuse esinemine üksikkatsel on äärmiselt vähe tõenäone, oleks viga sellise sündmuse toimumise võimalust eirata pikkade katseseeriate korral.

Näiteks, kui mingisse haigusesse haigestumise tõenäosus on 0,001, siis on seda üsna lohutav teada igal inimesel eraldi võetuna. Ometigi haigestub 1 000 000 inimesega populatsioonis sellesse haigusesse keskmiselt 1000 inimest, mis pole sugugi väike arv.

Teoreetilised jaotused

• Teoreetilised jaotused on teatavad standardsed seaduspärad, mille kuju määrab tunnuse tekkemehhanism.

Näiteks on matemaatika mõistes oma olemuselt sarnased tunnused ‘bakterite arv 1 ml piimas’, ‘emise pesakonna suurus’, ‘edukalt talvitunud mesitarude arv mesilas’ jne, või siis ‘lõhe kasvukiirus’, ‘õhu liikumise kiirus laudas’, ‘mulla happesus’ jne.

• Teoreetilised jaotused on kirjeldatud parameetritest sõltuvate eeskirjadega, mille abil on võimalik leida vastava jaotusega tunnuse (statistikute) väärtuste esinemise tõenäosused.

Näiteks teades, et poisslapse sünni tõenäosus on P(‘sugu’=‘poiss’) = 0,518, saame kohe leida ka tüdruku sünni tõenäosuse: P(‘sugu’=‘tüdruk’) = 1 – 0,518 = 0,482;

analoogset lahutamistehet (eeskirja) saame kasutada ka arvutamaks teatud haigusest paranemise tõenäosust – kui me teame, et keskmiselt 72% haigestunud loomadest hukkub, siis tervenemise tõenäosus avaldub kujul:
P(‘tervenemine’=‘jah’) = 1 – P(‘tervenemine’=‘ei’) = 1 – 0,72 = 0,28;

parameeriks, millest jaotuseeskiri antud näites sõltub, on ühe sündmuse toimumise tõenäosus.

Teoreetilised jaotused

• Teoreetilised jaotused on aluseks teaduslike järelduste tegemisel (statistiliste hüpoteeside kontrollimisel, sageli ka parameetrite väärtuste hindamisel ja nende hinnangute usaldusväärsuse leidmisel).

Seejuures on järeldused õiged üksnes siis, kui nad on tehtud andmetega sobivatele teoreetilistele jaotustele tuginedes (seda eriti väikeste, $n < 100$, valimite puhul)!

Näiteks pakkugu meile huvi tõenäosus, et utel sünnib 2 talle. Siis teades, et vaid ühe talle sündimise tõenäosus on 0,42, saame eelmisel lehel toodud eeskirja kasutades vale hinnangu meid huvitavale 2 talle sündimise tõenäosusele:

1 – P(‘tallede arv’=1) = 1 – 0,42 = 0,58 = P(‘tallede arv’≠1) ≠ P(‘tallede arv’=2).

Diskreetsed jaotused

Diskreetne jaotus esitatakse tõenäosusfunktsiooniga $p(k) = P(X=k)$, kus k on jaotuse võimalik väärtus.

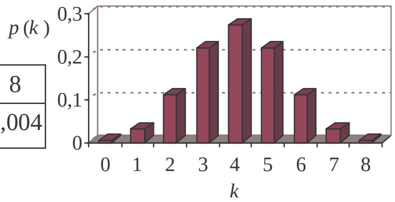
Binoomjaotus

Sündmuse toimumiste arv n-katselises katseseerias, kus igal üksikul katsel on sündmuse toimumise tõenäosus p: $X \sim B(n;p)$.

Tõenäosusfunktsioon: $p(k) = C_n^k p^k (1-p)^{n-k}$, $C_n^k = \dfrac{n!}{k!(n-k)!}$

Näide. Koeral sündis 8 kutsikat. Huvi pakkuv suurus X on isaste arv nende hulgas. Lihtsuse mõttes on eeldatud, et isase ja emase kutsika sündimise tõenäosus on võrdne (p = 0,5).

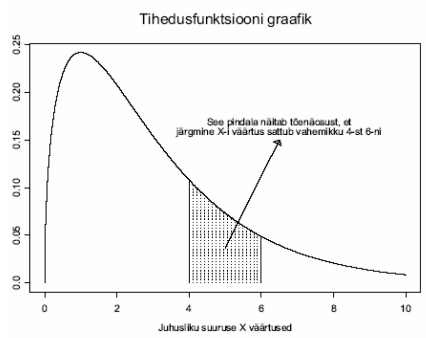
k	0	1	2	3	4	5	6	7	8
$p(k)$	0,004	0,031	0,109	0,219	0,273	0,219	0,109	0,031	0,004

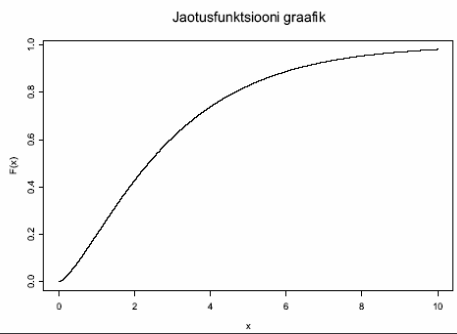


Pidevad jaotused

Pidev jaotus esitatakse tihedusfunktsiooniga $f(x) = dF(x)/dx$, kus $F(x) = P(X \leq x)$ on jaotusfunktsiooni väärtus kohal x,

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a).$$



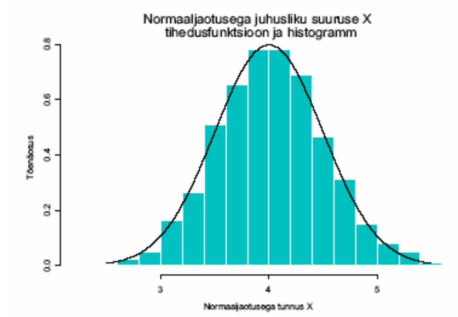


Normaaljaotus

Tihedusfunktsioon:

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

Tähistus: $X \sim N(\mu;\sigma^2)$



Normaaljaotuse tähtsus statistikas:

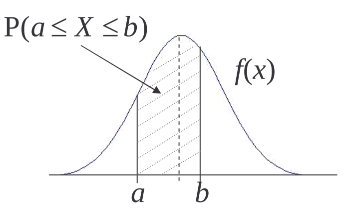
- Paljud mõõdetud tunnused on ligikaudu normaaljaotusega (näiteks juhul, kui uuritavat tunnust mõjutab hulk üksikult võttes väikese mõjuga tegureid).
- Paljud matemaatilise statistika meetodid eeldavad tunnuse jaotumist vastavalt normaaljaotuse seaduspäradele.
- Suurte valimite korral on paljud normaaljaotusega tunnuste tarvis loodud meetoditest rakendatavad sõltumata jaotusest.

Normaaljaotus

Kui uuritav tunnus on normaaljaotusega, siis ligikaudu
68,3% väärtustest jäävad vahemikku $\mu \pm \sigma$;
95,5% vahemikku $\mu \pm 2\sigma$ ja
99,7% vahemikku $\mu \pm 3\sigma$.

Standardne normaaljaotus: $X \sim N(0;1)$

Kui $X \sim N(\mu;\sigma^2)$, siis $\frac{X-\mu}{\sigma} \sim N(0,1)$



Tähistus: $\Phi(x) = P(X \le x)$
(jaotusfunktsiooni väärtus kohal x)

Omadus: $\Phi(-x) = 1 - \Phi(x)$

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

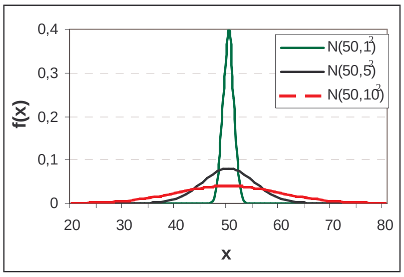
Normaaljaotus

Normaaljaotuse jaotusfunktsioon $\Phi(x)$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Normaaljaotus

Näide. Vere kogus indiviidi 50 ml vereproovis $X \sim N(50;\sigma^2)$, kus σ iseloomustab proovivõtmise täpsust.



Kui suur on tõenäosus, et proovi maht erineb 50 ml-st enam kui 5 ml võrra?

$$\mathrm{P}(X<45 \cup X>55)=1-\mathrm{P}(45\le X\le 55)=?$$

$\sigma=1$:

$$\mathrm{P}(45\le X\le 55)=\Phi\left(\frac{55-50}{1}\right)-\Phi\left(\frac{45-50}{1}\right)$$

$$=\Phi(5)-\Phi(-5)=\Phi(5)-[1-\Phi(5)]=2\Phi(5)-1=2*0,9999997-1=0,9999994$$

$$\mathrm{P}(X<45 \cup X>55)=1-0,9999994=0,00000057=5,76\mathrm{E}-07$$

$\sigma=5$:

$$\mathrm{P}(45\le X\le 55)=\Phi\left(\frac{55-50}{5}\right)-\Phi\left(\frac{45-50}{5}\right)=2\Phi(1)-1=2*0,8413-1=0,6827$$

$$\mathrm{P}(X<45 \cup X>55)=1-0,6827=0,3173$$

Normaaljaotus

Normaaljaotuse eelduse kontrollimine graafiliselt

Histogramm Tõenäosuspaber [*Normal prababilty plot*]