Matemaatiline statistika ja modelleerimine

Tõenäosus, teoreetilised jaotused, populatsiooni parameetrite hindamine

EMÜ doktorikool
DK.0007

Tanel Kaart

Tõenäosus

Kuidas leida praktikas sündmuse toimumise tõenäosust?

• Teha väga palju (teoreetiliselt lõpmatu arv) katseid/vaatlusi – nn statistiline tõenäosus – sündmuse toimumise suhteline sagedus annab tulemuseks ligikaudse tõenäosuse.

• Tundes uuritavat objekti hästi või mõistes täielikult, mis katse käigus toimub, on võimalik tõenäosust leida mõttetöö tulemusena (st, et teatavate sündmuste toimumise tõenäosuse kohta saab teha äärmiselt usutavaid eeldusi) – nn klassikaline tõenäosus (see, mida koolis õpitakse).

• Vahel on võimalik sündmuse tõenäosust arvutada, kui teame mõne teise sündmuse toimumise tõenäosust.

Statistiline tõenäosus

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

Suhtelise sageduse kaudu leitud nn statistiline tõenäosus on teoreetilise 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

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

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.

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

Tehted tõenäosustega

1. $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
2. Tinglik tõenäosus $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B) / \mathrm{P}(B)$,
   millest $\mathrm{P}(A \cap B) = \mathrm{P}(B) \times \mathrm{P}(A \mid B) = \mathrm{P}(A) \times \mathrm{P}(B \mid A)$
3. Kui A ja B on sõltumatud, siis
   $\mathrm{P}(A \cap B) = 0$,
   $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$,
   $\mathrm{P}(A \mid B) = \mathrm{P}(A)$ ja $\mathrm{P}(B \mid A) = \mathrm{P}(B)$
4. Täistõenäosuse valem: $\mathrm{P}(A) = \sum_{i=1}^{k} \mathrm{P}(H_i) \times \mathrm{P}(A \mid H_i)$
5. Bayesi valem: $\mathrm{P}(H_i \mid A) = \dfrac{\mathrm{P}(A \mid H_i) \times \mathrm{P}(H_i)}{\sum_{j=1}^{k} \mathrm{P}(A \mid H_j) \times \mathrm{P}(H_j)}$

Teoreetilised jaotused

• Teoreetilised jaotused leitakse analüütiliselt või arvutisimulatsioonide abil, **baseeruvana uuritava tunnuse** (=juhusliku suuruse) või selle väärtuste funktsiooni (e statistiku) tekkemehhanismil e **olemusel**.

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 kirjeldatakse parameetritest sõltuvate eeskirjadega**, mille abil on võimalik leida vastava jaotusega tunnuste (statistikute) väärtuste esinemise tõenäosused.

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

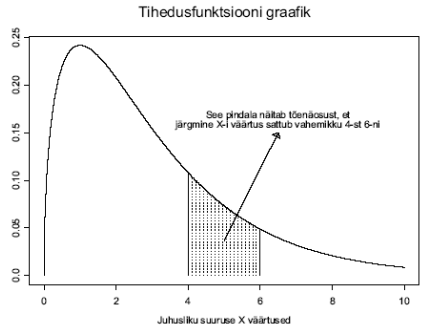
Seejuures on järeldused õiged üksnes siis, kui nad on tehtud andmetega sobivatele teoreetilistele jaotustele tuginedes (seda eriti väikeste, $n < 100$, valimite puhul)!

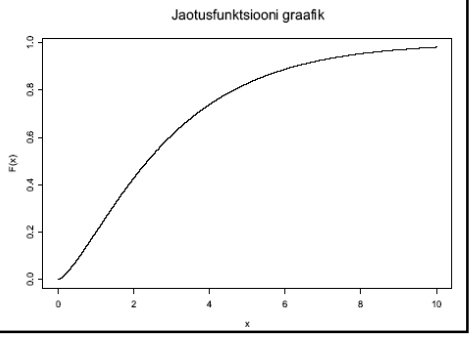
Teoreetilised jaotused

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

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

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





Tuntumaid jaotusi

Bernoulli jaotusega on kõik binaarsed tunnused, $X \sim Be(p)$, kus p on Bernoulli jaotuse parameeter (tõenäosus, et uuritav suurus omandab väärtuse 1).

Seejuures $E(X) = p$ ja $D(X) = p \times (1 - p)$.

Binoomjaotus

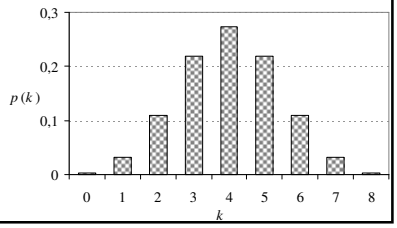
Sündmuse toimumiste arv n-katselises katseseerias, kus igal üksikul katsel on sündmuse toimumise tõenäosus p: $X \sim B(n;p)$.

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

$E(X) = np$ ja $D(X) = np(1 - p)$

Näited. 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



Tuntumaid jaotusi

Poissoni jaotus

Poissoni jaotusega on näiteks ühe päeva jooksul aset leidvate südameatakkide arv Tartu linnas, raku jagunemisel tekkivate geenimutatsioonide arv jne. Seda, et tunnus X on Poissoni jaotusega, tähistatakse $X \sim P(\lambda)$, kus λ on keskmine südameatakkide arv ühes päevas või keskmine mutatsioonide arv raku jagunemisel.

Tõenäosusfunktsioon: $p(k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0,1,...$ $\mathrm{E}(X) = \mathrm{D}(X) = \lambda$

Geomeetriline jaotus

Kui katse õnnestumise tõenäosus on p, siis katse number, millal katse esimest korda õnnestus, on geomeetrilise jaotusega juhuslik suurus.

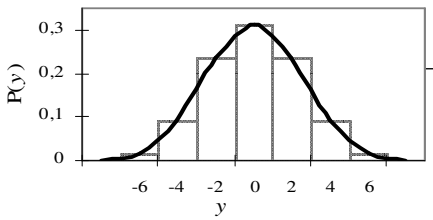
Tõenäosusfunktsioon: $p(k) = p(1-p)^{k-1},\ k = 1,2,...$ $\mathrm{E}(X) = (1-p)/p$, $\mathrm{D}(X) = (1-p)/p^2$

Normaaljaotus

Kui uuritavat tunnust mõjutavad paljud erinevad tegurid, millest ühegi mõju ei ole omaette võttes märkimisväärne, siis on uuritava tunnuse jaotus lähedane normaaljaotusele.

Näide. Alleelid *a*, *b*, *c* vähendavad ja *A*, *B*, *C* suurendavad fenotüübiväärtust *y* 1 võrra.

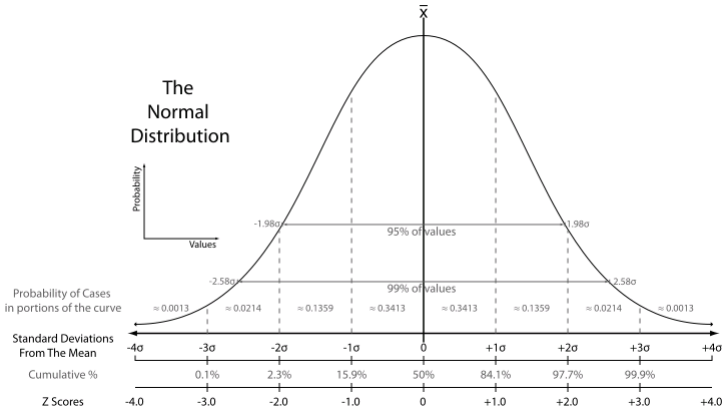
Genotüübid (*Aa* ja *aA* jne on kirjas ühe variandina)	*y*	P(*y*)
AABBCC	6	1/64
AABBCc, AABbCC, AaBBCC	4	6/64
AABBcc, AABbCc, AaBBCc, AAbbCC, AaBbCC, aaBBCC	2	15/64
AABbcc, AaBBcc, AAbbCc, AaBbCc, aaBBCc, AabbCC, aaBbCC	0	20/64
AAbbcc, AaBbcc, AabbCc, aaBBcc, aaBbCc, aabbCC	-2	15/64
Aabbcc, aaBbcc, aabbCc	-4	6/64
aabbcc	-6	1/64



Normaaljaotus

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

kus μ on uuritava tunnuse keskväärtus ja σ^2 dispersioon, $X \sim N(\mu,\sigma^2)$.



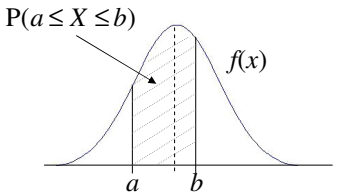
Normaaljaotus

Normaaljaotusega juhuslike suuruste lineaarkombinatsioon on samuti normaaljaotusega (muutuvad vaid parameetrite väärtused).

Sagedasemaks lineaarteisenduseks on standardiseerimine

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

kus $N(0,1)$ on standardne normaaljaotus, mille jaotusfunktsiooni $\Phi(z) = \mathrm{P}(Z \le z)$ väärtused on tabuleeritud. Seejuures kehtivad seosed



$$F(x)=\Phi\left(\tfrac{x-\mu}{\sigma}\right) \text{ ja } \Phi(-z)=1-\Phi(z)$$

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

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

Tabel. Standardse normaaljaotuse enamkasutatavad jaotusfunktsiooni väärtused

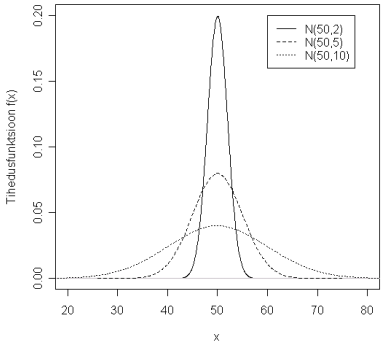
$\Phi(z) = \alpha$	0,005	0,025	0,05	0,5	0,95	0,975	0,995
$z_\alpha\ (q_\alpha)$	-2,58	-1,96	-1,64	0	1,64	1,96	2,58

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?

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



$\sigma = 1$:

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

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

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

$\sigma = 5$:

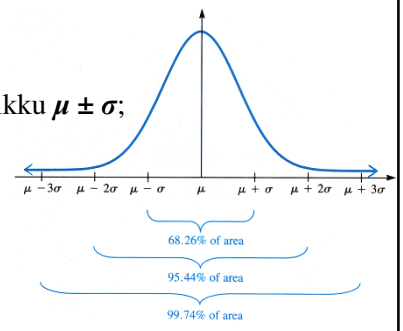
$$P(45 \le X \le 55) = \Phi\left(\frac{55-50}{5}\right) - \Phi\left(\frac{45-50}{5}\right) = 2\Phi(1) - 1 = 2 * 0,8413 - 1 = 0,6827$$

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

Normaaljaotus

Kui uuritav tunnus on normaaljaotusega, siis

- ligikaudu **68,3%** väärtustest jäävad vahemikku $\boldsymbol{\mu \pm \sigma}$;
- **95,4%** vahemikku $\boldsymbol{\mu \pm 2\sigma}$ ja
- **99,7%** vahemikku $\boldsymbol{\mu \pm 3\sigma}$.



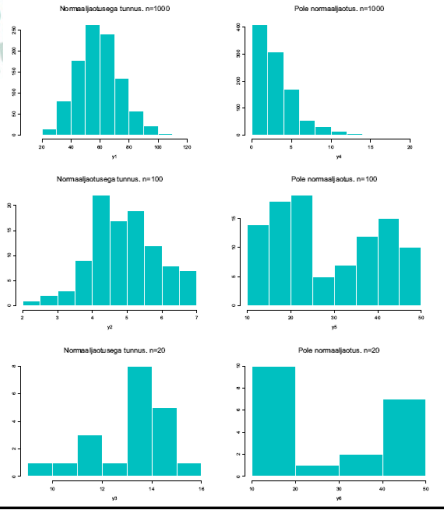
Normaaljaotuse tähtsus statistikas:

- Paljud mõõdetud tunnused on ligikaudu normaaljaotusega.
- Paljud matemaatilise statistika meetodid eeldavad tunnuse jaotumist vastavalt normaaljaotuse seaduspäradele.
- Suurte valimite korral on paljud normaaljaotusega tunnuste tarvis loodud meetoditest rakendatavad sõltumata jaotusest – näiteks läheneb mistahes jaotusega tunnuse keskmise (ja ka summa) jaotus valimi mahu kasvades normaaljaotusele.

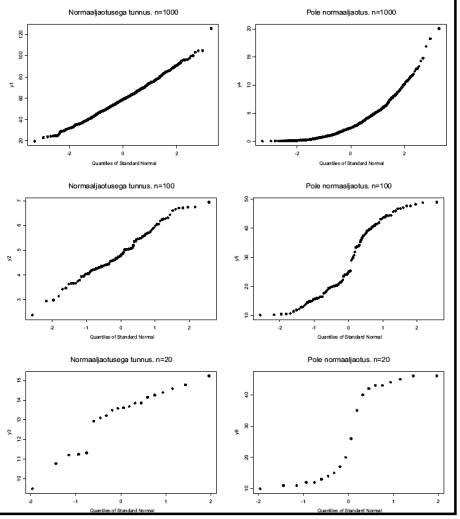
Normaaljaotus

Normaaljaotuse eelduse kontrollimine graafiliselt

Histogramm



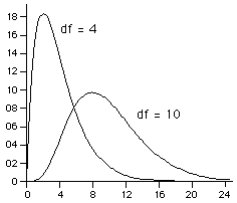
Tõenäosuspaber [*Normal prabability plot*]



Tuntumaid jaotusi

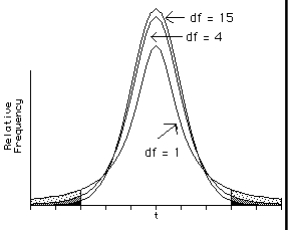
Sõltumatute standardse normaaljaotusega juhuslike suuruste $X_1,\ldots,X_n$ ruutude summa on **χ^2-jaotusega** vabadusastmete arvuga n:

$$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$



Kui juhuslik suurus $X \sim N(0,1)$ ja juhuslik suurus $Y \sim \chi^2(n)$, kusjuures X ja Y on sõltumatud, siis $Z = \dfrac{X}{\sqrt{Y/n}} \sim t(n)$,

st, et Z on ***t*-jaotusega** vabadusastmete arvuga n.

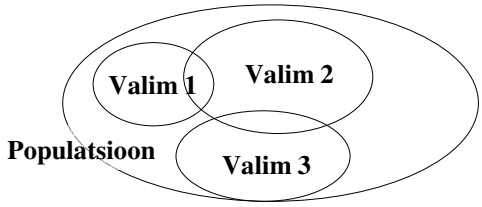


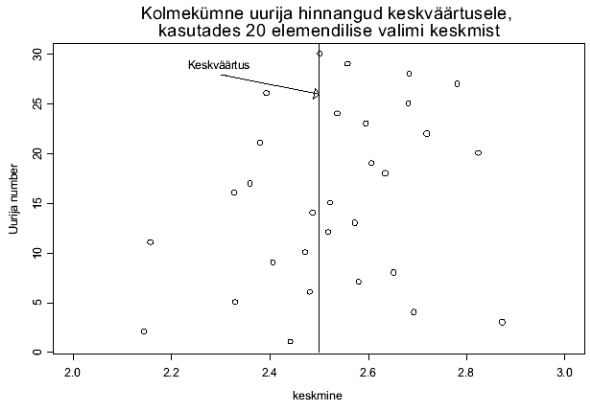
Kui $X_i \sim N(\mu,\sigma^2)$, $i=1,\ldots,n$, siis

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1),$$

kus $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ja $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$.

Populatsiooni parameetrite hindamine





Punkthinnangud

Kui huvipakkuva väärtuse hinnanguks on üks konkreetne arv, siis räägitakse, et tegemist on **punkthinnanguga**.
Märkimaks, et tegu on hinnanguga, kirjutatakse sageli arvkarakteristikut iseloomustava sümboli kohale katuseke, laineke vms; näiteks parameetri θ hinnang on $\hat{\theta}$.

Populatsiooni parameeter	Hinnang valimi põhjal
keskväärtus [*expectation*] $\mathrm{E}X$, μ	keskmine $\bar{x}$
populatsiooni dispersioon $\mathrm{D}X$, $\mathrm{var}(X)$, σ^2	valimi dispersioon $\hat{\mathrm{D}}X$, $\hat{\mathrm{var}}(X)$, s^2
populatsiooni standardhälve σ	valimi standardhälve s
populatsiooni mediaan $\mathrm{med}(X)$	valimi mediaan $\hat{\mathrm{med}}(X)$
populatsiooni α-kvantiil q_α	valimi α-kvantiil $\hat{q}_\alpha$

Hinnangute leidmise meetodid

Suurima tõepära meetod (*maximum likelihood method*, ML-meetod)
• Suurima tõepära meetod baseerub teoreetilisel jaotusel, mille parameetriks (argumendiks) on hinnatav parameeter.
• Hinnanguks valitakse see parameetri väärtus, mis realiseerunud juhul (st uuritavate andmete korral) kõige paremini sobib ehk teisisõnu on antud valimi jaoks tõepäraseim väärtus.

Vähimruutude meetod (*least square method*, LS-meetod)
• Vähimruutude meetod leiab parameetri hinnangu, minimiseerides realiseerunud väärtuste (andmete) ja parameetri hinnangule vastavate väärtuste vahelise ruuterinevuse.
• Vähimruutude meetod ei eelda tihedus- või tõenäosusfunktsiooni kasutamist, mistõttu on selle abil saadavad hinnangud sageli lihtsamal kujul, võrreldes teiste hindamismeetoditega.

Momentide meetod (*method of moments*)

Nihkega ja nihketa hinnangud

Parameetri θ hinnangut nimetatakse **nihketa hinnanguks** [*unbiased*], kui $\mathrm{E}(\hat{\theta}) = \theta$; ehk hinnang on „keskmiselt õige“, puudub süstemaatiline viga.

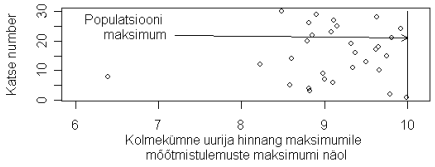
Näiteks valimi (juhusliku suuruse realiseerunud väärtuste) keskmine

$$\bar{x} = \tfrac{1}{n}\sum x_i$$

on nihketa hinnang populatsiooni (juhusliku suuruse) keskväärtusele $\mathrm{E}(X)$:

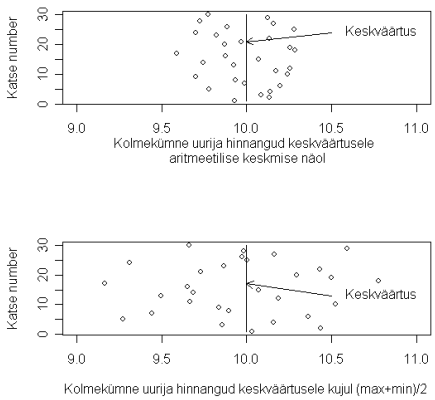
$$\mathrm{E}(\bar{x}) = \mathrm{E}\left(\tfrac{1}{n}\sum_{i=1}^{n} x_i\right) = \tfrac{1}{n}\sum_{i=1}^{n}\underbrace{\mathrm{E}(x_i)}_{\mathrm{E}(X)}$$

$$= \tfrac{1}{n}\, n\,\mathrm{E}(X) = \mathrm{E}(X)\,.$$



Efektiivsed hinnangud

Parameetri θ hinnangut $\hat{\theta}$ nimetatakse **efektiivseks hinnanguks**, kui $\text{var}(\hat{\theta})$ on vähim kõigi parameetri θ nihketa hinnangute dispersioonide hulgas; ehk, efektiivne hinnang on täpseim hinnang.



Hinnangu standardviga

Et andmete alusel leitud parameetri θ hinnang on juhuslik suurus, siis eksisteerib tal ka dispersioon $\text{var}(\hat{\theta})$. Viimane on aga jällegi tundmatu üldkogumi parameeter.
Seega, et saada tegelikkuses aimu oma andmete alusel leitud parameetri hinnangu täpsusest, tuleb andmetest hinnata ka hinnangu dispersioon, millest reeglina parema mõistetavuse huvides võetakse veel ruutjuur (et saada varieeruvuse hinnangut samal skaalal parameetri endaga).

Hinnangu standardhälbe hinnangut nimetatakse **hinnangu standardveaks**:

$$se(\hat{\theta}) = \sqrt{\hat{\text{var}}(\hat{\theta})}.$$

Näiteks keskväärtuse $\text{E}(X) = \mu$ hinnangu $\hat{\mu} = \bar{x}$ (valimi keskmise) dispersiooni hinnang on $\hat{\text{var}}(\hat{\mu}) = s^2/n$ ja standardviga on

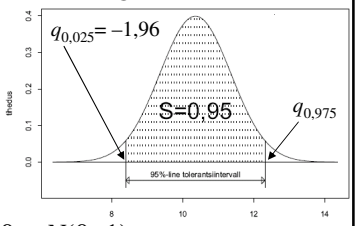
$$se(\hat{\mu}) = \frac{s}{\sqrt{n}},$$

kus s^2 on valimi dispersioon.

Tolerantsiintervall

Kogenud kalastajana teate, et teie poolt siiani püütud haugid on keskmiselt kaalunud 785 g standardhälbega 340 g. Perele õhtusöögiks kala püüdes oleks ju huvitav teada, kui palju kaalub järgmine konksu otsa jääv haug.

Täpset vastust sellele küsimusele statistika ei anna, küll aga saab leida vahemiku, millesse järgmise haugi kaal satub suure (näiteks 95%) tõenäosusega.



Väärtuste vahemik, kuhu kuuluvad 95% uuritava tunnuse väärtustest, on 95%-**tolerantsiintervall**.

Teame, et kui $X \sim N(785; 340)$, siis $(X-785)/340 \sim N(0; 1)$.

Standardse normaaljaotuse kohta teame, et 95% väärtustest jääb vahemikku $(q_{0,025}, q_{0,975}) = (-1,96; 1,96)$.

Seega

$$\mathrm{P}\left(-1,96 < \frac{X-785}{340} < 1,96\right) = 0,95$$

ja

$$\mathrm{P}(785 - 1,96 \times 340 < X < 785 + 1,96 \times 340) = \mathrm{P}(118,6 < X < 1451,4) = 0,95$$

Usaldusintervall

Vahemikhinnang (usaldusintervall, *confidence interval*, *CI*) tähendab valimi abil teatava piirkonna määramist leitud punkthinnangu ümber nii, et see piirkond kataks õige parameetri väärtuse etteantud küllalt suure tõenäosusega:

$$\mathrm{P}(\underline{\theta} < \theta < \overline{\theta}) = 1 - \alpha,$$

- kus $1 - \alpha$ on **usaldusnivoo** [*confidence level*] (ühe lähedane, ent alati ühest väiksem);
- α, mida nimetatakse **olulisuse nivooks** [*significance level*], on väike positiivne arv (tavaliselt 0,01 või 0,05);
- θ on õige, ent mitteteadaolev jaotusparameetri väärtus;
- $\underline{\theta}$ ja $\overline{\theta}$ on parameetri θ **(1-α)-usalduspiirid** (näiteks kui $\alpha = 0,05$, siis on tegu 95%-liste usalduspiiridega).

Täpsuse huvides räägitakse vahel ka alumisest ja ülemisest usalduspiirist [*lower/upper confidence limit*].

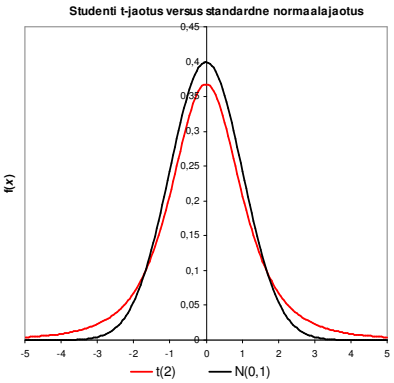
Usaldusintervall

Usaldusintervall (*confidence interval*) **keskmisele**

✓ $X \sim N(\mu,\sigma^2) \Rightarrow \bar{x} \sim N\left(\mu,\frac{\sigma^2}{n}\right)$ ehk $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Kui σ ei ole teada, siis $\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

Toodud seosed kehtivad <u>suure valimi</u> korral sõltumata uuritava tunnuse jaotusest!



✓ α-kvantiil q_α (t-jaotuse puhul $t_{\alpha,n-1}$, standardse normaaljaotuse puhul z_α)

$$P(X < q_\alpha) = \alpha$$

Sisuliselt sama, mis protsentiil;
näiteks 0,5-kvantiil on mediaan, sest $P(X < med) = 0{,}5$.

Usaldusintervall

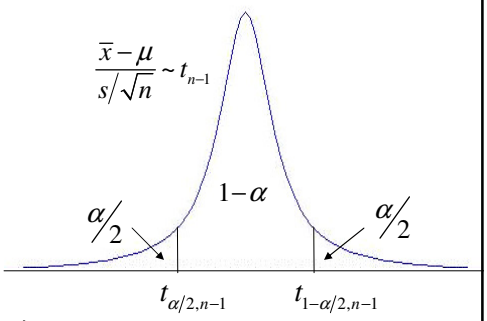
Usaldusintervall keskmisele

$$P\left(\frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{\alpha/2,n-1}\right) = \alpha/2 \quad \text{ja}$$

$$P\left(\frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{1-\alpha/2,n-1}\right) = 1-\alpha/2$$

$$\Downarrow$$

$$P\left(\bar{x} - t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

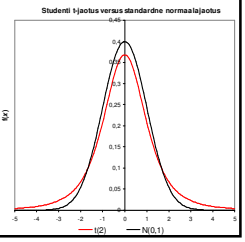


(1−α)-usalduspiirid:

$$(\underline{\mu};\overline{\mu}) = \left(\bar{x} - t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

Kui valim on suur ($n > 100$), siis võib kasutada ka normaaljaotust:

$$(\underline{\mu};\overline{\mu}) = \left(\bar{x} - q_{1-\alpha/2}\frac{s}{\sqrt{n}}; \bar{x} + q_{1-\alpha/2}\frac{s}{\sqrt{n}}\right)$$



Usaldusintervall

Näide. Kanauurijat Hans Hane huvitab, mitu muna munevad keskmiselt Eestis peetavad sassexi tõugu kanad ühe nädala jooksul. Härra Hani luges ühe nädala jooksul kokku kümne kana munad: 3 5 4 6 2 6 5 6 5 3.
95%-line usaldusintervall = ?

$\bar{x} = 4,5;\ s \approx 1,43 \quad t_{1-\alpha/2;(n-1)} = t_{0,975;9} = 2,26$

$$(\underline{\mu};\overline{\mu}) = \left(\bar{x} - t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}};\ \bar{x} + t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = \left(4,5 - 2,26\frac{1,43}{\sqrt{10}};\ 4,5 + 2,26\frac{1,43}{\sqrt{10}}\right)$$
$$= (4,5 - 2,26\times 0,45;\ 4,5 + 2,26\times 0,45) = (3,47;\ 5,53)$$

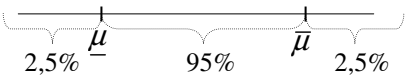
95%-lise tõenäosusega võib väita, et keskmine nädalas munetud munade arv on kuskil vahemikus 3,47-st 5,53-ni.

Suurendamaks hinnangu täpsust, tuleks uurida rohkem kanu, sest mida suurem on *n*, seda kitsamaks muutub usaldusintervall.

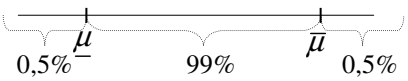
Usaldusintervall

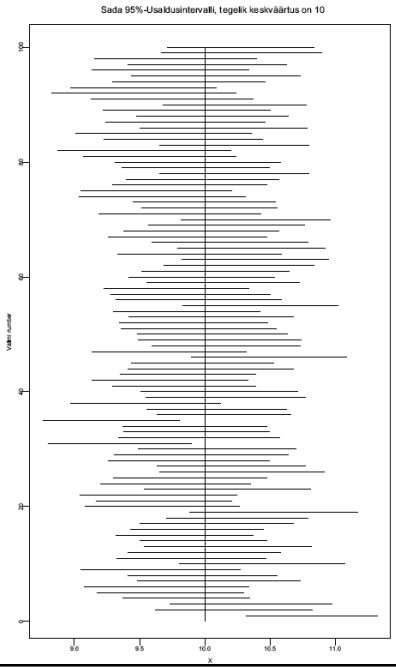
α=0,1 → 90%-usaldusintervall

α=0,05 → 95%-usaldusintervall



α=0,01 → 99%-usaldusintervall





Usaldusintervall

Table I. Values of rainbow trout blood chemistry values ($n = 45$)

	Mean ± s.e.	Median	Skewness (mean ± s.e.)	Kurtosis (mean ± s.d.)	CV (%)	95% CI (lower–upper)
Glucose (mg dl^{-1})	108·11 ± 9·98	98·00	0·62 ± 0·35	−0·21 ± 0·70	61·91	88·00–128·22
Urea* (mg dl^{-1})	4·36 ± 0·24	4·00	1·21 ± 0·35	2·24 ± 0·70	36·35	3·88–4·83
Creatinine* (mg dl^{-1})	0·29 ± 0·01	0·29	1·24 ± 0·35	3·09 ± 0·70	23·71	0·27–0·31
Total bilirubin (mg dl^{-1})	0·04 ± 0·00	0·05	0·11 ± 0·35	−0·55 ± 0·70	58·02	0·04–0·05
Aspartate aminotransferase (U l^{-1})	461·20 ± 27·62	447·00	0·65 ± 0·35	0·38 ± 0·70	40·18	405·53–516·87
Alanine aminotransferase* (U l^{-1})	12·87 ± 1·16	11·00	1·71 ± 0·35	3·63 ± 0·70	60·22	10·54–15·19
Alkaline phosphatase* (U l^{-1})	179·22 ± 19·26	131·00	1·92 ± 0·35	3·84 ± 0·70	72·10	140·40–218·04
Creatine phosphokinase* (U l^{-1})	1265·11 ± 161·70	894·00	1·21 ± 0·35	0·73 ± 0·70	85·74	939·22–1591·00
Lactate dehydrogenase* (U l^{-1})	2628·18 ± 164·75	2399·00	1·53 ± 0·35	2·62 ± 0·70	42·05	2296·15–2960·21
Gamma-glutamyl transferase (U l^{-1})	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Total protein (g dl^{-1})	3·59 ± 0·13	3·74	−0·59 ± 0·35	0·46 ± 0·70	24·64	3·32–3·85
Albumin (g dl^{-1})	1·38 ± 0·05	1·40	−0·39 ± 0·35	0·05 ± 0·70	25·17	1·27–1·48
Triglycerides (mg dl^{-1})	347·51 ± 23·56	327·00	0·46 ± 0·35	−0·73 ± 0·70	45·47	300·04–394·99
Cholesterol (mg dl^{-1})	247·38 ± 10·32	241·00	0·09 ± 0·35	0·10 ± 0·70	27·98	226·59–268·17
Ca (mg dl^{-1})	12·52 ± 0·20	12·20	0·55 ± 0·35	−0·08 ± 0·70	10·81	12·11–12·93
P (mg dl^{-1})	22·66 ± 1·19	22·50	0·45 ± 0·35	−0·67 ± 0·70	35·26	20·26–25·06
Mg (mg dl^{-1})	3·85 ± 0·11	3·82	0·07 ± 0·35	−0·84 ± 0·70	15·40	3·63–4·07
Na (mEq l^{-1})	154·07 ± 0·85	155·00	−0·53 ± 0·35	1·14 ± 0·70	3·69	152·36–155·78
K (mEq l^{-1})	3·45 ± 0·29	3·25	0·45 ± 0·35	−0·81 ± 0·70	52·55	2·87–4·03
Cl* (mEq l^{-1})	128·09 ± 1·13	130·00	−2·90 ± 0·35	11·84 ± 0·70	5·90	125·82–130·36

n.a., not assessable. *Null hypothesis (Kolmogorov–Smirnov test) was rejected.

Usaldusintervall

Näide. Kümne aastail 1996-2004 EPK-lehmade esimesel seemendusel enam kasutatud pullide spermaga seemendatud lehmade keskmine tiinestuvus, selle 95%-lised usalduspiirid ja seemendatud lehmade arv.