

Jaotusfunktsioonid

Allan Sims

31.01.2025

Jaotusfunktsioonid

Statistikas nimetatakse **jaotusfunktsiooniks** funktsiooni, mis kirjeldab juhusliku suuruse X tõenäosusjaotust. See määrab iga reaalarvu x korral tõenäosuse, et juhusliku suuruse X väärtus on väiksem või võrdne x -ga. Formaalselt väljendub jaotusfunktsioon kui $F(x) = P(X \leq x)$, seostes iga x -i väärtuse vastavusse tõenäosuse $P(X \leq x)$.

Jaotusfunktsiooni saab kasutada nii pidevate kui ka diskreetsete juhuslike suuruste puhul. Pideva juhusliku suuruse korral on jaotusfunktsioon integraal tema tihedusfunktsioonist. Diskreetse juhusliku suuruse korral on jaotusfunktsioon kumulatiivne tõenäosuste summa.

Jaotusfunktsioonil on järgmised omadused:

- See on *monotoonselt kasvav*, st $F(x) \leq F(y)$ kui $x \leq y$.
- See on *paremalt pidev*, st $\lim_{x \rightarrow y^+} F(x) = F(y)$ iga y korral.
- See läheneb 0-le kui x läheneb $-\infty$ -le, ja 1-le kui x läheneb $+\infty$ -le.

Jaotusfunktsiooni saab kasutada juhusliku suuruse tõenäosusjaotuse visualiseerimiseks. Selle abil saab arvutada ka tõenäosusi, et juhuslik suurus langeb teatud vahemikku. Näiteks tõenäosus, et X langeb vahemikku $[a, b]$, on $F(b) - F(a)$.

Jaotusfunktsioon on statistikas väga oluline tööriist. See aitab mõista juhuslike suuruste käitumist ja teha järeldusi andmete kohta.

Lisaks ülaltoodule on jaotusfunktsioonil veel teisi olulisi omadusi ja kasutusvõimalusi. Näiteks:

- Jaotusfunktsiooni abil saab leida juhusliku suuruse kvantiile. Kvantiil on väärtus, millest väiksemaks või võrdseks jääb juhusliku suuruse väärtus teatud tõenäosusega.
- Jaotusfunktsiooni abil saab arvutada juhusliku suuruse keskväärtust, dispersiooni ja teisi momente.
- Jaotusfunktsiooni abil saab genereerida juhuslikke arve, mis järgivad teatud jaotust.

Jaotusfunktsioon on statistikas fundamentaalne mõiste ja sellel on lai valik rakendusi.

Teoreetilisi jaotusi kasutatakse hüpoteeside testimisel statistilistes analüüsides selleks, et teha kindlaks, kas vaatlusandmete hajuvus on juhuslik või on olemas mingi muster või seos. Oluline on märkida, et enamik statistilisi teste eeldab teatud jaotuse tundmist, mille põhjal tehakse järeldusi populatsiooni kohta.

Kõige tuntum teoreetiline jaotus on normaaljaotus, mis on sümmeetriline kõver ning annab tihti paljudele looduslikele nähtustele lähedase jaotuse. Normaaljaotus on oluline paljudes statistilistes testimistes, näiteks t-testis või ANOVA analüüsis. Kui andmed järgivad normaaljaotust, saab kasutada erinevaid statistilisi meetodeid.

Lisaks normaaljaotusele on olulised ka teised jaotused, nagu näiteks binoomjaotus (kasutatakse binomiaalsete katsete puhul), eksponentsiaaljaotus (kasutatakse elupikkuste analüüsimisel) ja Poissoni jaotus (kasutatakse haruldaste sündmuste analüüsimisel).

Hüpoteeside testimisel võrreldakse empiirilisi andmeid teoreetiliste jaotustega, et teha järeldusi populatsiooni kohta. Näiteks võib uurida, kas kahe rühma keskmiste vahel on oluline erinevus või kas mingi seos on

juhuslik või mitte. Sellisel juhul kasutatakse statistilisi teste, et hinnata, kas tulemus võib olla juhuslik või on tõenäoliselt tingitud tegelikust mõjust või seosest.

Hüpotees

Hüpotees on statistikas väide või eeldus, millele toetudes teostatakse uuringuid, analüüsi või katseid. Hüpotees võib olla kas uurija poolt esitatud eeldus, mida on vaja tõestada või ümber lükata või juba olemasolev teooria, mida soovitakse kontrollida.

- **Nullhüpotees (H0):** See on väide, mida püütakse ümber lükata. Nullhüpotees eeldab, et uuritavas populatsioonis puudub mingi mõju, seos või erinevus. Näiteks võib nullhüpotees väita, et kahe rühma keskmised väärtused on võrdsed või et tunnus ei mõjuta teise tunnuse jaotust. Nullhüpotees on alati konkreetne väide, mille kohta on võimalik arvutada tõenäosust.
- **Sisukas ehk alternatiivne hüpotees (H1):** See on väide, mida uurija soovib tõestada. See hüpotees esitab oletuse mingi mõju, seose või erinevuse olemasolu kohta populatsioonis. Näiteks võib alternatiivne hüpotees väita, et kahe rühma keskmised väärtused on erinevad või et tunnus mõjutab teise tunnuse jaotust. Sisukaid hüpoteese võib olla ka rohkem kui üks.
- Hüpoteesipaar püstitatakse üldkogumi jaoks.
- Hüpoteeside püstitus sõltub uurimisküsimusest ja sellest, kas oodatakse kindlasuunalist kõrvalekallet või mitte.
 - **Kahepoolne hüpotees:** Kontrollitakse, kas parameeter (nt keskvärtus) erineb mingist väärtusest. Näiteks H0: $\mu = \mu_0$ ja H1: $\mu \neq \mu_0$.
 - **Ühepoolne hüpotees:** Kontrollitakse, kas parameeter on suurem või väiksem mingist väärtusest. Näiteks H0: $\mu \leq \mu_0$ ja H1: $\mu > \mu_0$ (parempoolne) või H0: $\mu \geq \mu_0$ ja H1: $\mu < \mu_0$ (vasakpoolne).
- **Olulisuse nivoo (α):** See on tõenäosus, millega ollakse nõus tegema I liiki viga ehk lükkama tagasi kehtiva nullhüpoteesi. Tavaliselt kasutatakse olulisuse nivood 0,05 (5%) või 0,01 (1%), kuid see sõltub uurimisvaldkonnast. Mida väiksem on olulisuse nivoo, seda väiksem on tõenäosus teha I liiki viga. Olulisuse nivoo on maksimaalne lubatud p-väärtus väärtus.
- **Kriitilised väärtused:** Need on teststatistiku väärtused, mis määratlevad kriitilise piirkonna. Kriitilised väärtused leitakse teststatistiku jaotusest vastavalt valitud olulisuse nivool.
 - Kahepoolse hüpoteesi korral on kaks kriitilist väärtust, mis asuvad jaotuse mõlemal pool.
 - Ühepoolse hüpoteesi korral on üks kriitiline väärtus, mis asub kas jaotuse paremal või vasakul pool.
 - Kriitilised väärtused on seotud ka usaldusvahemiku laiusega.
- **Teststatistiku väärtuse võrdlemine kriitilise väärtusega:** Võrreldakse, kas empiiriline teststatistiku väärtus langeb kriitilisse piirkonda.
 - Kui empiiriline väärtus ületab kriitilise väärtuse (langeb kriitilisse piirkonda), siis lükatakse nullhüpotees tagasi ja võetakse vastu sisukas hüpotees.
 - Kui empiiriline väärtus ei lange kriitilisse piirkonda, siis ei ole alust nullhüpoteesi tagasi lükata. Jäädakse nullhüpoteesi juurde, aga see ei tähenda, et nullhüpotees on tõestatud.
- **Olulisuse tõenäosus (p-väärtus):** See on tõenäosus saada antud või veelgi äärmuslikum valim, eeldusel, et nullhüpotees on tõene. Mida väiksem on p-väärtus, seda väiksem on tõenäosus, et nullhüpotees kehtib.
 - Kui p-väärtus on väiksem kui olulisuse nivoo ($p < \alpha$), siis lükatakse nullhüpotees tagasi.
 - Kui p-väärtus on suurem või võrdne olulisuse nivoo ($p \geq \alpha$), siis ei lükata nullhüpoteesi tagasi.
 - Olulisuse tõenäosust saab kasutada nullhüpoteesi sobivuse hindamiseks antud valimiga.
- Teststatistiku empiirilise väärtuse võrdlemine kriitilise väärtusega ja olulisuse tõenäosuse võrdlemine olulisuse nivoo on ekvivalentsed meetodid.

Tuleb arvestada, et nullhüpoteesi ei saa “tõestada”, kuna statistiliste testidega ei saa nullhüpoteesi tõestada samamoodi, nagu matemaatikas võidakse tõestada väiteid. Selle asemel saab statistiliste testide abil hinnata, kas on piisavalt tõendeid nullhüpoteesi tagasilükkamiseks. Kui testi tulemus näitab, et andmed on nullhüpoteesiga vastuolus, siis võidakse nullhüpotees tagasi lükata. Kui aga testi tulemus ei näita olulist vastuolu nullhüpoteesiga, siis ei lükata nullhüpoteesi tagasi. Siiski, see ei tähenda, et nullhüpotees oleks “tõestatud”; pigem öeldakse, et ei leitud piisavalt tõendeid nullhüpoteesi ümberlükkamiseks. Oluline on mõista, et nullhüpoteesi mitte-tagasilükkamine ei tõesta selle paikapidavust. See viitab lihtsalt sellele, et andmed ei andnud piisavalt alust nullhüpoteesi ümber lükata. Samuti on oluline mees pidada, et statistiliste testide tulemused annavad ainult tõenäosusliku hinnangu ning need tuleb alati panna laiemasse konteksti koos teiste tõendite ja teadmistega valdkonnast.

Jaotusfunktsioon

Jaotusfunktsioon on viis näidata, kui tõenäoline on, et mingi arv (mida me nimetame muutujaks) on väiksem või võrdne kindla arvuga x . See aitab meil mõista, kuidas arvud (või muutujad) ühes grupis paiknevad või jaotuvad, näiteks kui me tahame teada, kui tõenäoline on, et metsa pindala on teatud suurusega või väiksem.

Jaotusfunktsiooni olulised omadused on:

1. Vahemik: jaotusfunktsioon võtab väärtuseid 0 ja 1 vahemikus, mis tähendab, et tõenäosus on alati selles vahemikus.
2. Mittelangus: jaotusfunktsioon ei kasva kunagi tagurpidi; see liigub alati kas võrdse kõrguse juurde või kõrgemale.
3. Ühekordsus: Ükskõik millised on väärtused, jaotusfunktsiooni koguulatus on alati 1.

Jaotusfunktsiooni abil saame lahendada paljusid statistilisi ülesandeid. See aitab meil leida näiteks keskmisi väärtusi, hajuvust (ehk kui laiali arvud on jaotunud), tõenäosusi (kui tõenäoline on mingi sündmus), ja isegi määrata arvude jaotuse erinevaid tasemeid. See tähendab, et jaotusfunktsioon on tööriist, mis aitab meil mõista, kuidas andmed käituvad, ja teha selle põhjal täpseid otsuseid või järeldusi.

```
# Näide R keskkonnas
# Laadime peatükis kasutatavad paketid
library(dplyr)
library(ggplot2)
library(ggpmisc)
```

- **Andmete jaotus.** ECDF graafik annab intuiitiivse ülevaate andmete jaotusest, näidates, kui suur osa andmestikust on alla või võrdne iga võimaliku väärtusega. See on eriti kasulik andmete jaotuse mõistmiseks ilma eeldamata mingit konkreetset jaotustüüpi.

ECDF graafik on väärtuslik tööriist andmete uurimisel, võimaldades kiiresti hinnata andmete jaotuse omadusi ja tuvastada andmete levikuga seotud probleeme.

```
# Loo andmestiku
set.seed(1)
df <- data.frame(
  x = c(rnorm(1000, 0, 3))
)
ggplot(df, aes(x)) +
  stat_ecdf()
```

Selgitus:

1. **set.seed(1):** See rida määrab juhuslike arvude generaatori seemne. Seemne määramine tagab, et kui käitate seda koodi uuesti, saate *samad* juhuslikud arvud ja seega *sama* graafiku. See on korduvuse jaoks ülioluline. Kui te seemet ei määra, on juhuslikud arvud (ja sellest tulenev graafik) iga kord, kui koodi käitate, erinevad.

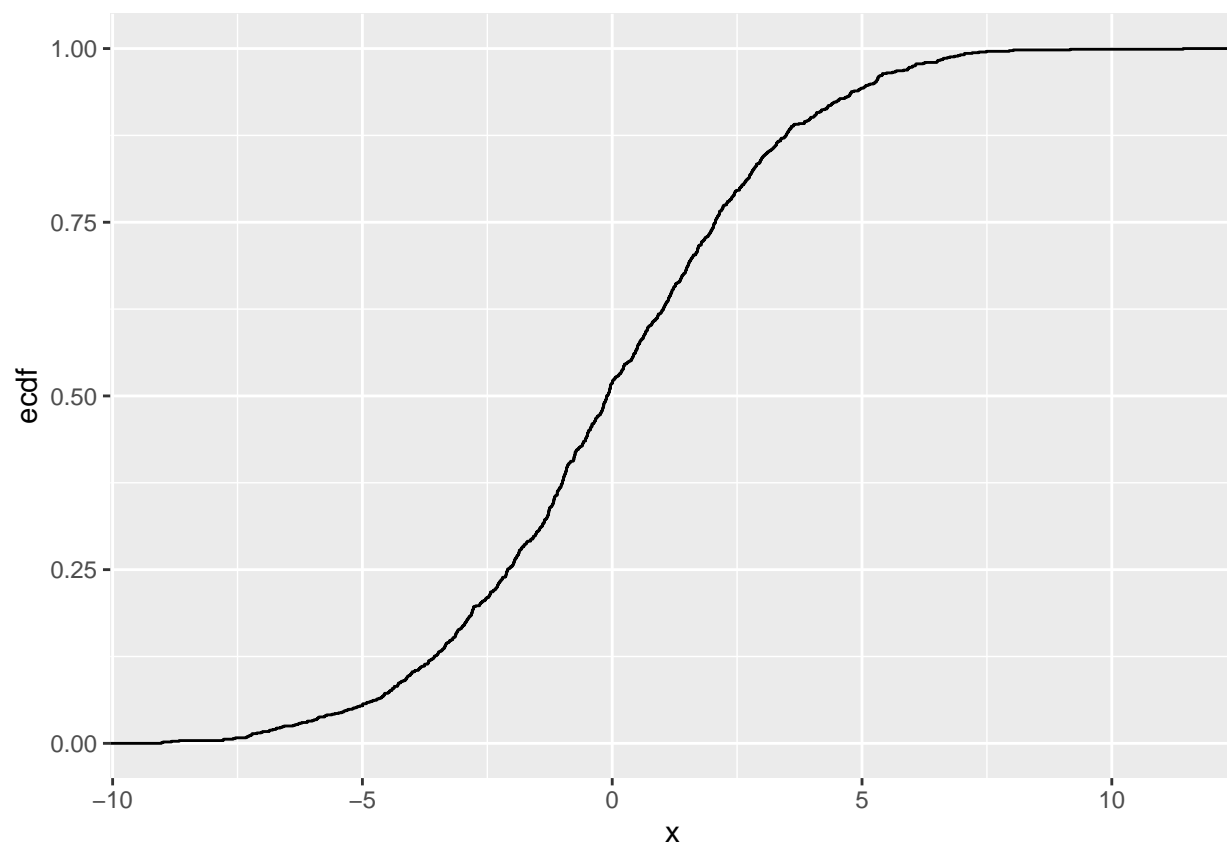


Figure 1: Empiiriline jaotusfunktsioon.

2. `df <- data.frame(x = c(rnorm(1000, 0, 3)))`: See loob andmetabeli nimega `df`. Selles andmetabelis luuakse veerg nimega `x`. Veeru `x` väärtused genereeritakse funktsiooniga `rnorm(1000, 0, 3)`.
 - `rnorm()` on funktsioon, mis genereerib juhuslikke arve normaaljaotusest.
 - 1000 määrab, et me tahame 1000 juhuslikku arvu.
 - 0 on normaaljaotuse keskvärtus.
 - 3 on normaaljaotuse standardhälve.

Seega genereerib see rida 1000 juhuslikku arvu normaaljaotusest keskvärtusega 0 ja standardhällbega 3 ning salvestab need `df` andmetabeli veergu `x`.
3. `ggplot(df, aes(x))`: See algatab graafiku loomise, kasutades `ggplot2`.
 - `ggplot(df, ...)` määrab, et graafiku andmed pärinevad `df` andmeraamist.
 - `aes(x)` seostab andmeraami veeru `x` graafiku `x`-teljega. See ütleb `ggplot2`-le, et veeru `x` väärtused määravad punktide asukoha horisontaalteljel.
4. `+ stat_ecdf()`: See lisab graafikule kihi, nimelt ECDF-i.
 - `stat_ecdf()` arvutab ECDF-i. ECDF antud väärtusel x tähistab andmepunktide osakaalu, mis on väiksemad või võrdsed x -ga.

Tihedusfunktsioon

Tihedusfunktsioon on statistikas kasutatav funktsioon, mis kirjeldab tõenäosust, et juhuslik suurus langeb mingisse kindlasse vahemikku. Tihedusfunktsiooni saab kasutada tõenäosusjaotuse (nt normaaljaotus, Bernoulli jaotus jne) karakteriseerimiseks. See näitab, kui tõenäoline on konkreetse väärtuse saamine selle jaotuse järgi.

Tihedusfunktsioon on graphiline joon, mis esitab tõenäosuse jaotuse ühtlase tervikuna. Üksiku väärtuse tõenäosus on antud tihedusfunktsiooni abil vahemikuna, mida funktsioon katab selles punktis. Tavaliselt on tihedusfunktsioon positiivne ja integreerub üle vahemiku 1, mis tähendab, et kõikidele võimalikele väärtustele vastavate tõenäosuste summa on alati 1.

Metsanduses võib tihedusfunktsiooni kasutada näiteks puistu tiheduse jaotuse kirjeldamisel, kus tihedusfunktsioon näitab puude arvu tõenäosust kindlas pindalas. Tihedusfunktsioonide abil saab hinnata erinevaid statistilisi parameetreid nagu keskmine, mediaan, variatsioon jt, ning teha ennustusi ja otsuseid metsaressursi haldamisel ja planeerimisel.

Tihedusfunktsioon (inglise keeles probability density function, lühidalt PDF) on statistikas oluline mõiste, mida kasutatakse tõenäosusjaotuse kirjeldamiseks. Tihedusfunktsioon annab meile informatsiooni selle kohta, kuidas on tõenäosus mingi juhusliku muutuja väärtuse jaotustel. Allpool on toodud mõned olulised statistilised omadused tihedusfunktsiooni kohta:

Integreeruvus: tihedusfunktsioon peab olema integreeruv kogu kandva ala suhtes, st selle integraal alates miinimum- kuni maksimumväärtuseni peab olema võrdne ühega. Teisisõnu, kogu võimaliku vaadeldava mahu (nt kogu metsaala, kus uuritakse puude tihedust) peab olema 100%.

Positiivsus: Tihedusfunktsiooni väärtused peavad olema alati positiivsed. See tähendab, et tõenäosus väärtuse sattumiseks konkreetse intervalli ei saa olla negatiivne.

Normaliseerumine: Tihedusfunktsioon peab olema normaliseeritud, mis tähendab, et alati, kui me integreerime tihedusfunktsiooni kogu liinil või tasandil, siis tulemuseks peab olema üks. See tagab, et tihedusfunktsioon annab meile täpse tõenäosuse, et vaadeldav suurus langeb mingisse kindlasse vahemikku.

Mitte-negatiivsus: Tihedusfunktsioon ei või mitte kunagi võtta negatiivseid väärtusi. See tähendab, et tõenäosus mingi suuruse väärtuse jaoks on alati null või suurem.

Tipukoha lokaliseerimine: Tihedusfunktsiooni tippkoht näitab moodi ehk suurimat tõenäosust mingi väärtuse saamiseks. Mood on statistikas näitaja, mis näitab, milline väärtus esineb andmehulgas kõige sagedamini ehk

mis väärtusega esineb andmehulkade hulgas enim kordi. Mood on üks keskmiste näitajate tüüpe, mis aitab anda ülevaadet andmete esinemissagedusest konkreetse väärtuse juures.

Need omadused on olulised tihedusfunktsiooni mõistmisel ja kasutamisel statistilistes analüüsides, sealhulgas metsanduses, kus tihedusfunktsiooni abil saab kirjeldada erinevaid muutujaid nagu puu tihedus, metsa elujõulisus jne.

Näide R-i koodist, mis genereerib normaaljaotuse tihedusfunktsiooni graafiku:

```
# Loo me andmestiku
set.seed(1)
df <- data.frame(
  x = c(rnorm(1000, 0, 3))
)
ggplot(df, aes(x)) +
  geom_density()
```

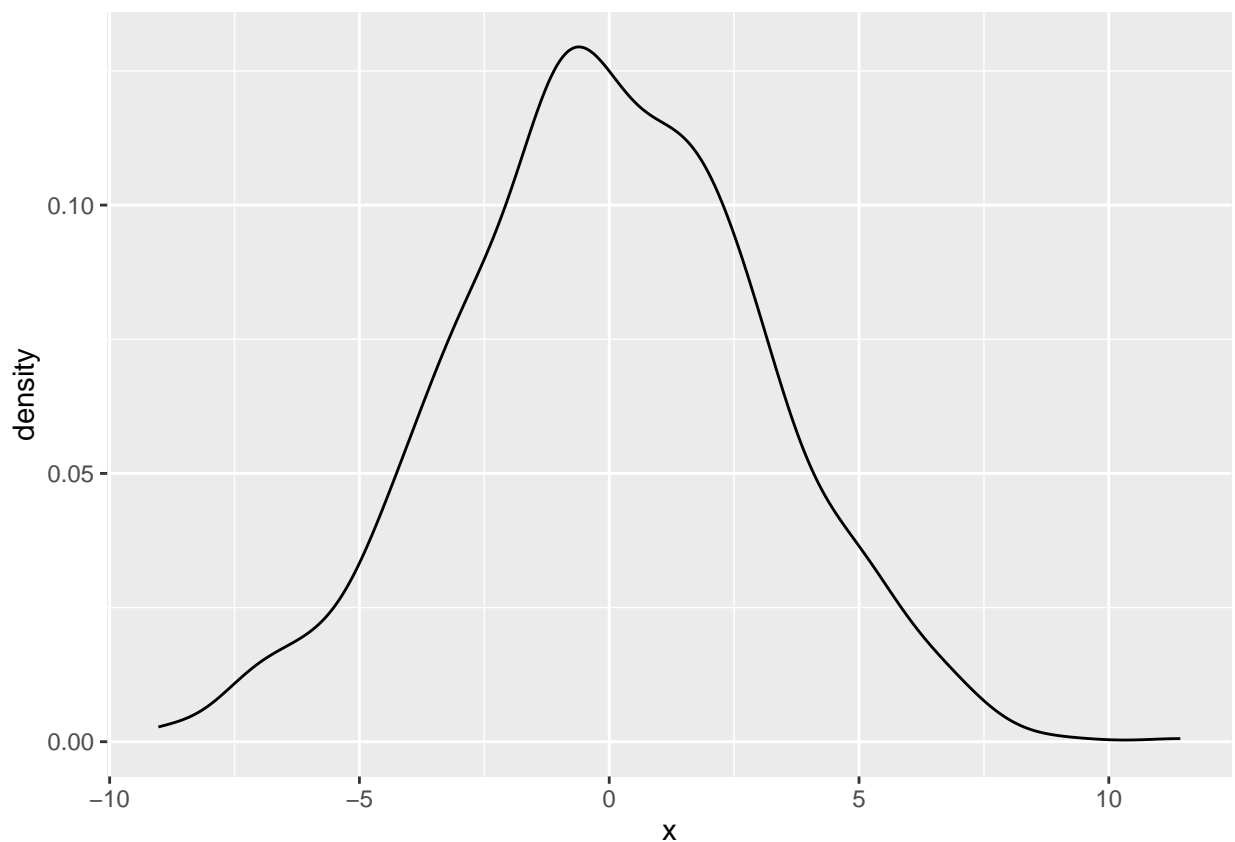


Figure 2: Empiiriline tihedusfunktsioon.

See kood genereerib normaaljaotuse tihedusfunktsiooni graafiku, kus x-teljel on andmete väärtused ja y-teljel tihedusfunktsiooni väärtused.

Normaaljaotus

Normaaljaotus ehk Gaussi jaotus on tõenäosusjaotus, mida iseloomustab sümmeetriline kellakujuline tõenäosustiheduse graafik ning parameetrid keskvärtus (mean) ja standardhälve (standard deviation).

Normaaljaotuse tõenäosustiheduse funktsioon on defineeritud järgmiselt:

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

kus: - $f(x)$ on tõenäosustiheduse funktsioon; - σ on standardhälve, mis näitab, kui hajus on andmete hulk; - μ on keskvärtus, mis näitab andmete keskmist väärtust; - e on Euleri arv, kus $e \approx 2.71828$.

Normaaljaotuse oluline omadus on see, et suure hulga juhuslike sündmuste kogum tulemusena on nende keskmine jaotunud normaalselt. Samuti on oluline märkida, et normaaljaotus on sümmeetriline ning selle standardiseeritud kuju $N(0, 1)$ on standardnormaaljaotus, mille keskvärtus on 0 ja standardhälve on 1.

Normaaljaotus on laialdaselt kasutatav statistikas erinevate nähtuste kirjeldamisel ning analüüsimisel, sh metsanduses, kus seda võib kasutada näiteks puude kõrguse või läbimõõdu jaotuste modelleerimisel ning hindamisel.

Normaaljaotust kasutatakse paljudes erinevates valdkondades, sealhulgas statistikas, majanduses ja teaduses. Metsanduses võib normaaljaotust kasutada näiteks metsa populatsiooni omaduste hindamiseks või metsanduslike nähtuste, nagu puude kõrgus või läbimõõt, jaotuse kirjeldamiseks.

Siin on mõned põhjused, miks normaaljaotust kasutatakse:

- Andmete kirjeldamine: Normaaljaotust kasutatakse tihti andmete kirjeldamiseks, kuna paljud looduslikud ja inimtegevuse tulemusena tekkinud nähtused järgivad sarnast jaotusmustrit. Normaaljaotus võimaldab kompaktselt kirjeldada, kuidas andmed on keskmise ümber hajunud.
- Statistiliste meetodite rakendamine: Paljud statistilised meetodid eeldavad, et andmed järgivad normaaljaotust. Näiteks on t-test ja ANOVA analüüsid välja töötatud eeldusel, et andmed on normaalselt jaotunud. Seega aitab normaaljaotuse eeldamine tagada statistiliste meetodite usaldusväärsuse.
- Prognoosimine ja ennustamine: Normaaljaotuse kasutamine võimaldab prognoosida tulevasi väärtusi ja sündmusi, eeldusel, et andmed järgivad seda jaotust. See võib aidata metsanduses näiteks prognoosida metsaressursi kasvu või hindamist.

Kokkuvõtvalt võib öelda, et normaaljaotuse kasutamine aitab andmeid paremini mõista, analüüsida ja tõlgendada, ning tagada, et nende tehtud järeldused oleksid usaldusväärsed ja põhineksid adekvaatsetel statistilistel mudelitel.

R-is on järgnevad normaaljaotuse funktsioonid:

- `pnorm()` arvutab tõenäosuse, et juhuslik muutuja on väiksem või võrdne kui antud väärtus.
- `dnorm()` arvutab tõenäosustihedusfunktsiooni väärtuse antud punktis.
- `qnorm()` arvutab väärtuse, millest väiksem või võrdne on antud protsent andmetest.

```
pnorm(1.96, 0, 1)
```

```
## [1] 0.9750021
```

```
dnorm(1.96, 0, 1)
```

```
## [1] 0.05844094
```

```
qnorm(0.975, 0, 1)
```

```
## [1] 1.959964
```

MS Exceli keskkonnas normaaljaotusega seotud arvutuste tegemiseks saab kasutada järgmisi funktsioone:

- Näide: `NORM.DIST(1,96; 0; 1; TRUE)` annab tõenäosuse, et normaaljaotuse väärtus on väiksem või võrdne 1.96.
- Näide: `NORM.S.DIST(1,96; TRUE)` annab tõenäosuse, et standardnormaaljaotuse väärtus on väiksem või võrdne 1.96.
- Näide: `NORM.INV(0,975; 0; 1)` annab normaaljaotuse kvantiili, mis vastab 0.975 tõenäosusele.

- Näide: $NORM.S.INV(0,975)$ annab standardnormaaljaotuse kvantiili, mis vastab 0.975 tõenäosusele.

T-jaotus

T-jaotus ehk Studenti t-jaotus on statistiline jaotus, mis on nimetatud selle looja William Sealy Gosseti (Studenti) järgi. T-jaotust kasutatakse hinnangute tegemiseks populatsiooni keskmise või vahe tegeliku väärtuse kohta, kui meil on vaid valimi keskmine ja standardhälve ning väike valimi suurus.

T-jaotuse tihedusfunktsioon sõltub vabadusastmetest, mida tähistatakse tähisega $n-1$, kus n on valimi suurus. Mida suurem on vabadusastmete arv, seda lähemale liigub t-jaotuse kumerus standardnormaaljaotusele. T-jaotust kasutatakse sageli siis, kui populatsiooni standardhälve on teadmata ja valimi suurus on alla 30.

T-jaotus on oluline tööriist hüpoteeside testimisel ja usaldusintervallide arvutamisel. Selle abil saame teha järeldusi populatsiooni keskmise kohta isegi siis, kui populatsiooni kohta on vähe teavet. T-jaotust kasutatakse laialdaselt erinevates valdkondades, sealhulgas meditsiinis, majanduses ja loodusteadustes, sealhulgas metsanduses.

t-jaotus on statistikas oluline jaotus, mida kasutatakse tõenäosusteoorias ja hüpoteeside testimisel. T-jaotus sarnaneb normaaljaotusega, kuid erinevalt normaaljaotusest on t-jaotus paindlikum ning sellel on erinev arv vabadusastmeid. T-jaotust kasutatakse peamiselt siis, kui uuritav valim on väike ning populatsiooni standardhälve on teadmata.

Peamised t-jaotuse kasutamise eesmärgid on:

1. Võrdlused: T-jaotust kasutatakse kahe või enama rühma vaheliste erinevuste või keskmiste võrdlemiseks. Näiteks võib t-jaotust kasutada selleks, et teha kindlaks, kas kahe erineva metsa keskmine juurdekasv on statistiliselt oluliselt erinev.
2. Hüpoteeside testimine: T-jaotust kasutatakse hüpoteeside testimisel, et teha kindlaks, kas uuritava andmestiku tulemused on juhuslikud või kas seal on tõepoolest mingisugune seos või erinevus. Näiteks võib t-jaotust kasutada selleks, et testida, kas metsa uuendamine teatud viisil mõjutab metsa biomassi kasvu.
3. Usaldusintervallid: T-jaotust kasutatakse usaldusintervallide arvutamiseks. Usaldusintervall aitab hinnata, kui usaldusväärne on uuritava keskmise või erinevuse hinnang ning kui suured võivad olla võimalikud vead hinnangus.

Kokkuvõttes on t-jaotus oluline tööriist statistilistes analüüsides, eriti kui tegemist on väikese valimi suuruse ja/või teadmata standardhälbe või suurema vabadusastmega. Metsandustudengite jaoks on t-jaotuse mõistmine ja oskuslik kasutamine oluline, kuna see aitab neil teha täpsemaid järeldusi ja otsuseid metsandusega seotud probleemide lahendamisel.

- **Näide:** $pt(2.26, 10)$ arvutab tõenäosuse, et t-jaotuse väärtus on väiksem või võrdne 2.26, kui vabadusastmete arv
- **Näide:** $dt(2.26, 10)$ arvutab t-jaotuse tõenäosustihedust väärtuses 2.26, kui vabadusastmete arv on 10.
- **Näide:** $qt(0.975, 10)$ arvutab t-jaotuse ülemise kvantiili 0.975 väärtuse, kui vabadusastmete arv on 10.

Exceli keskkonnas t-jaotusega seotud arvutuste tegemiseks on olemas järgmised funktsioonid:

1. T.DIST:

- **Kirjeldus:** T.DIST funktsioon arvutab Studenti t-jaotuse kumulatiivse tõenäosuse.
- **Kasutamine:** $T.DIST(x; \text{vabadusastmed}; \text{kumulatiiv})$ kus x on t-jaotuse väärtus, mille kumulatiivset tõenäosust soovite arvutada, vabadusastmed tähistavad t-jaotuse vabadusastmete arvu, ja kumulatiiv on loogiline väärtus, mis määrab, kas funktsioon tagastab tõenäosuse (TRUE) või tiheduse (FALSE).
- **Näide:** $T.DIST(2.26; 10; TRUE)$ annab tõenäosuse, et t-jaotuse väärtus on väiksem või võrdne 2.26, kui vabadusastmete arv on 10.

2. T.DIST.RT:

- **Kirjeldus:** T.DIST.RT funktsioon arvutab ühepoolse tõenäosuse paremast sabast Studenti t-jaotuses.
- **Kasutamine:** T.DIST.RT(x; vabadusastmed) kus x on t-jaotuse väärtus, mille ühepoolset tõenäosust paremast sabast soovite arvutada, ja vabadusastmed on t-jaotuse vabadusastmete arv.
- **Näide:** T.DIST.RT(2,26; 10) annab tõenäosuse, et t-jaotuse väärtus on suurem kui 2,26, kui vabadusastmete arv on 10.

3. T.DIST.2T:

- **Kirjeldus:** T.DIST.2T funktsioon arvutab kahepoolse tõenäosuse Studenti t-jaotuses.
- **Kasutamine:** T.DIST.2T(x; vabadusastmed) kus x on t-jaotuse väärtus, mille kahepoolset tõenäosust soovite arvutada, ja vabadusastmed on t-jaotuse vabadusastmete arv.
- **Näide:** T.DIST.2T(2,26; 10) annab tõenäosuse, et t-jaotuse absoluutne väärtus on suurem kui 2,26, kui vabadusastmete arv on 10.

4. T.INV:

- **Kirjeldus:** T.INV funktsioon arvutab Studenti t-jaotuse kvantiili antud kumulatiivse tõenäosuse jaoks.
- **Kasutamine:** T.INV(tõenäosus; vabadusastmed) kus tõenäosus on soovitud kumulatiivne tõenäosus, ja vabadusastmed tähistavad t-jaotuse vabadusastmete arvu.
- **Näide:** T.INV(0,975; 10) annab t-jaotuse kvantiili, mis vastab 0,975 kumulatiivsele tõenäosusele, kui vabadusastmete arv on 10.

5. T.INV.2T:

- **Kirjeldus:** T.INV.2T funktsioon arvutab kahepoolse testi kriitilise väärtuse Studenti t-jaotuses.
- **Kasutamine:** T.INV.2T(tõenäosus; vabadusastmed) kus tõenäosus on soovitud kahepoolse testi tõenäosus, ja vabadusastmed on t-jaotuse vabadusastmete arv.
- **Näide:** T.INV.2T(0,05; 10) annab kahepoolse testi

F-jaotus

F-jaotus on statistiline jaotus, mis on tuletatud kahe sõltumatult ja normaalselt jaotunud muutuja jagamise tulemusena. F-jaotuse puhul hinnatakse kahe variatsiooni suhet, olles seega kasulik tööriist hüpoteeside testimisel statistilistes uuringutes.

F-jaotust kasutatakse eelkõige anova (Analüüs Variatsioon) testides, et hinnata, kas erinevate rühmade keskmised on statistiliselt oluliselt erinevad. Näiteks metsanduse valdkonnas võib F-jaotust kasutada näiteks erinevate metsamajandamise meetodite või puuliikide kasvukiiruse või mahu võrdlemisel.

F-jaotuse puhul on oluline teada kahte vabadusastet: nimetaja vabadusastet (DFn) ja jääkvabadusastet (DFd). DFn tähistab rühmade arvu miinus 1 ja DFd tähendab rühmade sees olevate üksikute vaatluste arvu miinus kõigi rühmade arv.

F-jaotuse kriitiline väärtus määratakse kindla usaldusväarsuse taseme juures ning kui arvutatud F-väärtus on suurem kui kriitiline väärtus, siis võib järeldada, et rühmade keskmised on tõepoolest erinevad.

Kokkuvõttes on F-jaotus oluline statistiline tööriist, mis võimaldab hinnata muutujate vahelisi suhteid ning teha järeldusi hüpoteeside kohta erinevate gruppide vahel.

F-jaotus on statistilise jaotus, mis on tuletatud kahe iseseisva kiirguse jaotuse ruutkorvamise jaotusena ning seda kasutatakse tihti hüpoteeside testimisel. F-jaotusel on kaks parameetrit - vabadusastmed (degrees of freedom) ja nende arv, ning seda kirjeldatakse $F(p, q)$ jaotusena, kus p ja q tähistavad vastavalt kahe erineva muutuja vabadusastmeid.

F-jaotuse eeldusteks on järgmised: 1. Andmed peavad olema normaalselt jaotunud. Kuigi F-jaotus ei eelda iseenesest normaaljaotust, on oluline, et sõltumatuid andmeid esindaksid normaalsed kõrvalekalde jaotused. 2. Andmete alusel koostatud rühmad peaksid olema sõltumatud üksteisest. 3. Andmete varieeruvused peavad olema sarnased ehk andmehulgad peaksid olema homoskedastised.

Need eeldused on olulised, et tagada F-testi usaldusväarsus ning õige tulemuse saamine hüpoteeside testimisel.

Andmete analüüsimisel kasutatakse sageli erinevaid statistilisi funktsioone, nagu df (degrees of freedom), pf (probability function), qf (quantile function) ja rf (random number generation function). Kõik need funktsioonid on saadaval programmeerimiskeeles R, mis on laialdaselt kasutatav statistilise analüüsi ja andmetöötluse tarkvara.

1. **pf:**

- **Kirjeldus:** pf funktsioon arvutab F-jaotuse kumulatiivse tõenäosuse.
- **Kasutamine:** pf(x, df1, df2) kus x on punkt, mille tõenäosust soovite arvutada, df1 on esimese andmestiku vabadusastmete arv ja df2 on teise andmestiku vabadusastmete arv.
- **Näide:** pf(3.84, 1, 30) annab tõenäosuse, et F-jaotuse väärtus on väiksem või võrdne 3.84, kui esimese andmestiku vabadusastmete arv on 1 ja teise andmestiku vabadusastmete arv on 30.

2. **df:**

- **Kirjeldus:** df funktsioon arvutab F-jaotuse tõenäosustiheduse funktsiooni.
- **Kasutamine:** df(x, df1, df2) kus x on punkt, mille tõenäosustihedust soovite arvutada, df1 on esimese andmestiku vabadusastmete arv ja df2 on teise andmestiku vabadusastmete arv.
- **Näide:** df(3.84, 1, 30) annab F-jaotuse tõenäosustihedust väärtuses 3.84, kui esimese andmestiku vabadusastmete arv on 1 ja teise andmestiku vabadusastmete arv on 30.

3. **qf:**

- **Kirjeldus:** qf funktsioon arvutab F-jaotuse kvantileid.
- **Kasutamine:** qf(p, df1, df2) kus p on soovitud kvantiil, df1 on esimese andmestiku vabadusastmete arv ja df2 on teise andmestiku vabadusastmete arv.
- **Näide:** qf(0.95, 1, 30) annab F-jaotuse ülemise kvantiili 0.95 väärtuse, kui esimese andmestiku vabadusastmete arv on 1 ja teise andmestiku vabadusastmete arv on 30.

Excelis F-jaotusega seotud arvutuste tegemiseks Euroopa piirkondades, kus kümnendkoha eraldajaks on koma ja argumentide eraldajaks on semikoolon, saate kasutada järgmisi funktsioone:

1. **F.DIST:**

- **Kirjeldus:** F.DIST funktsioon arvutab F-jaotuse kumulatiivse tõenäosuse.
- **Kasutamine:** F.DIST(x; vabadusastmed1; vabadusastmed2; kumulatiiv) kus x on F-jaotuse väärtus, mille kumulatiivset tõenäosust soovite arvutada, vabadusastmed1 on esimese valimi vabadusastmete arv, vabadusastmed2 on teise valimi vabadusastmete arv, ja kumulatiiv on loogiline väärtus, mis määrab, kas funktsioon tagastab tõenäosuse (TRUE) või tiheduse (FALSE).
- **Näide:** F.DIST(3;10;5;TRUE) annab tõenäosuse, et F-jaotuse väärtus on väiksem või võrdne 3, kui esimese valimi vabadusastmete arv on 10 ja teise valimi oma on 5.

2. **F.DIST.RT:**

- **Kirjeldus:** F.DIST.RT funktsioon arvutab F-jaotuse ühepoolse tõenäosuse paremast sabast.
- **Kasutamine:** F.DIST.RT(x; vabadusastmed1; vabadusastmed2) kus x on F-jaotuse väärtus, mille ühepoolset tõenäosust paremast sabast soovite arvutada, vabadusastmed1 on esimese valimi vabadusastmete arv, ja vabadusastmed2 on teise valimi vabadusastmete arv.
- **Näide:** F.DIST.RT(3;10;5) annab tõenäosuse, et F-jaotuse väärtus on suurem kui 3, kui esimese valimi vabadusastmete arv on 10 ja teise valimi oma on 5.

3. **F.INV:**

- **Kirjeldus:** F.INV funktsioon arvutab F-jaotuse kvantiili antud kumulatiivse tõenäosuse jaoks.
- **Kasutamine:** F.INV(tõenäosus; vabadusastmed1; vabadusastmed2) kus tõenäosus on soovitud kumulatiivne tõenäosus, vabadusastmed1 on esimese valimi vabadusastmete arv, ja vabadusastmed2 on teise valimi vabadusastmete arv.
- **Näide:** F.INV(0,975;10;5) annab F-jaotuse kvantiili, mis vastab 0,975 kumulatiivsele tõenäosusele, kui esimese valimi vabadusastmete arv on 10 ja teise valimi oma on 5.

4. **F.INV.RT:**

- **Kirjeldus:** F.INV.RT funktsioon arvutab F-jaotuse kvantiili, arvestades ühepoolset tõenäosust paremast sabast.
- **Kasutamine:** F.INV.RT(tõenäosus; vabadusastmed1; vabadusastmed2) kus tõenäosus on ühepoolse testi tõenäosus paremast sabast, vabadusastmed1 on esimese valimi vabadusastmete arv, ja vabadusastmed2 on teise valimi vabadusastmete arv.
- **Näide:** F.INV.RT(0,025;10;5) annab F-jaotuse kvantiili, mis vastab 0,025 tõenäosusele paremast sabast

Hii-ruut-jaotus

Hii-ruut-jaotus (χ^2 -jaotus) on pidev tõenäosusjaotus, mida kasutatakse statistikas laialdaselt. Juhuslik suurus X järgib hii-ruut-jaotust vabadusastmete arvuga k , kui seda saab esitada k sõltumatu standard normaaljaotusega (keskväärtus 0 ja standardhälve 1) juhusliku suuruse ruutude summana.

Omadused:

- Vabadusastmete arv (k): see on jaotuse oluline parameeter, mis määrab selle kuju. Mida suurem on vabadusastmete arv, seda lamedam ja sümmeetrilisem on jaotus.
- Keskväätus: hii-ruut-jaotuse keskväätuseks on vabadusastmete arv (k).
- Dispersioon: hii-ruut-jaotuse dispersiooniks on $2k$.
- Kuju: hii-ruut-jaotus on asümmeetriline ja kaldub paremale. Väikeste vabadusastmete arvu korral on see tugevalt kaldu, aga suuremate vabadusastmete arvu korral muutub see järk-järgult sümmeetrilisemaks.

Hii-ruut-jaotust kasutatakse paljudes statistilistes rakendustes, näiteks:

- Hii-ruut-test: seda testi kasutatakse kahe kategoorilise muutuja vahelise seose tugevuse hindamiseks.
- Sobivustestid: nende testidega hinnatakse, kas vaadeldav andmestik vastab eeldatavale jaotusele.
- Usaldusintervallide konstrueerimine: hii-ruut-jaotust saab kasutada populatsiooni dispersiooni usaldusintervallide leidmiseks.