



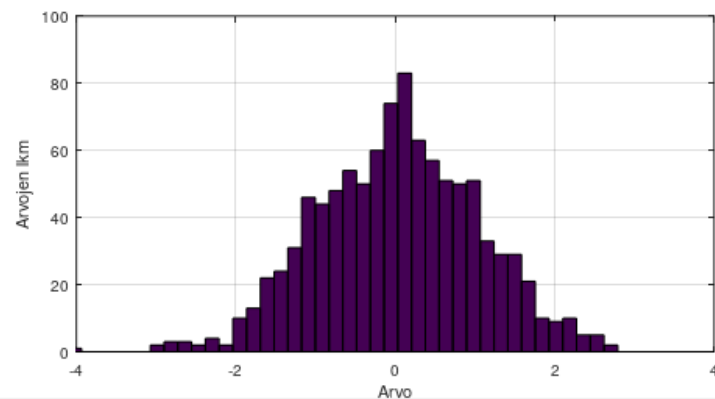
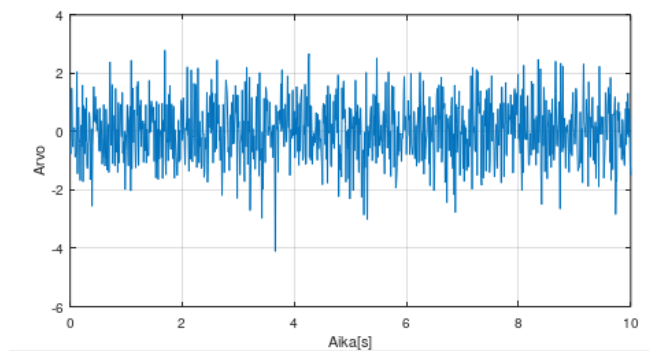
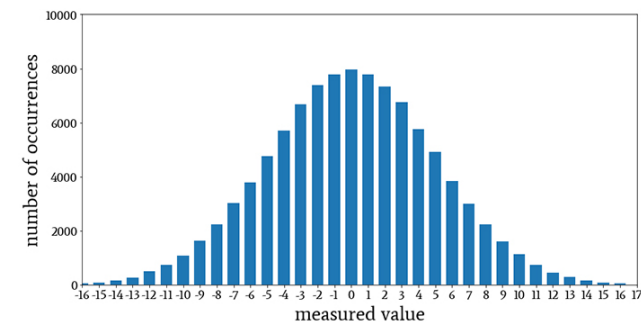
**KAMK • University
of Applied Sciences**

Digitaalisen signaalinkäsittelyn perusteet

**Signaalinkäsittelyn tekniikoita ja tilastollisia
tunnuslukuja – osa 1**

Normaalijakauma

- Normaalijakauma tunnetaan myös nimillä Gaussin jakauma tai Gaussin kellokäyrä
- Se on jatkuva todennäköisyysjakauma, joka on symmetrinen keskiarvon suhteen
- Normaalijakautuneen datan arvot ovat täten todennäköisemmin lähellä keskiarvoa kuin etäällä keskiarvosta



Keskiarvo, mediaani ja moodi

- **Aritmeettinen keskiarvo** on lukujen summa jaettuna niiden lukumäärällä
- **Mediaani** on jakauman "tyypillinen" arvo, suuruusjärjestykseen asetettujen havaintoarvojen keskimmäinen arvo
 - Jos havaintoarvoja on parillinen määrä, mediaani on kahden keskimmäisen keskiarvo tai jompikumpi keskimmäisistä arvoista
- **Moodi** eli tyyppiarvo on aineistossa useimmin esiintyvä arvo

Komennot Octavessa:

- mean, <https://octave.sourceforge.io/nan/function/mean.html>
- median, <https://octave.sourceforge.io/octave/function/median.html>
- mode, <https://octave.sourceforge.io/octave/function/mode.html>

Varianssi

- Varianssi kuvaa, kuinka paljon satunnaismuuttujan arvot keskimäärin vaihtelevat odotusarvosta.
 - Kun arvot keskittyvät odotusarvon ympärille tiiviisti, on varianssin arvo pieni
 - Kun arvot ovat hajallaan odotusarvon ympärillä, on varianssin arvo suuri

Population Variance	Sample Variance
$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$ <p> σ^2 = population variance x_i = value of i^{th} element μ = population mean N = population size </p>	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ <p> s^2 = sample variance x_i = value of i^{th} element \bar{x} = sample mean n = sample size </p>

Komento Octavessa:

- `var`

<https://octave.sourceforge.io/octave/function/var.html>

Keskihajonta (1/2)

- Keskihajonta eli satunnaismuuttujan standardipoikkeama eli kuvaa keskimääräistä poikkeamaa odotusarvosta.
- Keskihajonta on varianssin neliöjuuri
- Etuna varianssiin verrattuna on tulkinnan helppous, sillä keskihajonnan asteikko vastaa mittausten asteikkoa.

Calculation	Formula	Notes
Population Standard Deviation	$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$	μ = population average X = individual values in population N = count of values in population
Sample Variance	$s^2 = \frac{\sum (x_i - \bar{x})^2}{(n - 1)}$	\bar{x} = sample average x = individual values in sample n = count of individual values in sample
Sample Standard Deviation	$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n - 1)}}$	\bar{x} = sample average x = individual values in sample n = count of individual values in sample

Komento Octavessa:

- std
- <https://octave.sourceforge.io/octave/function/std.html>



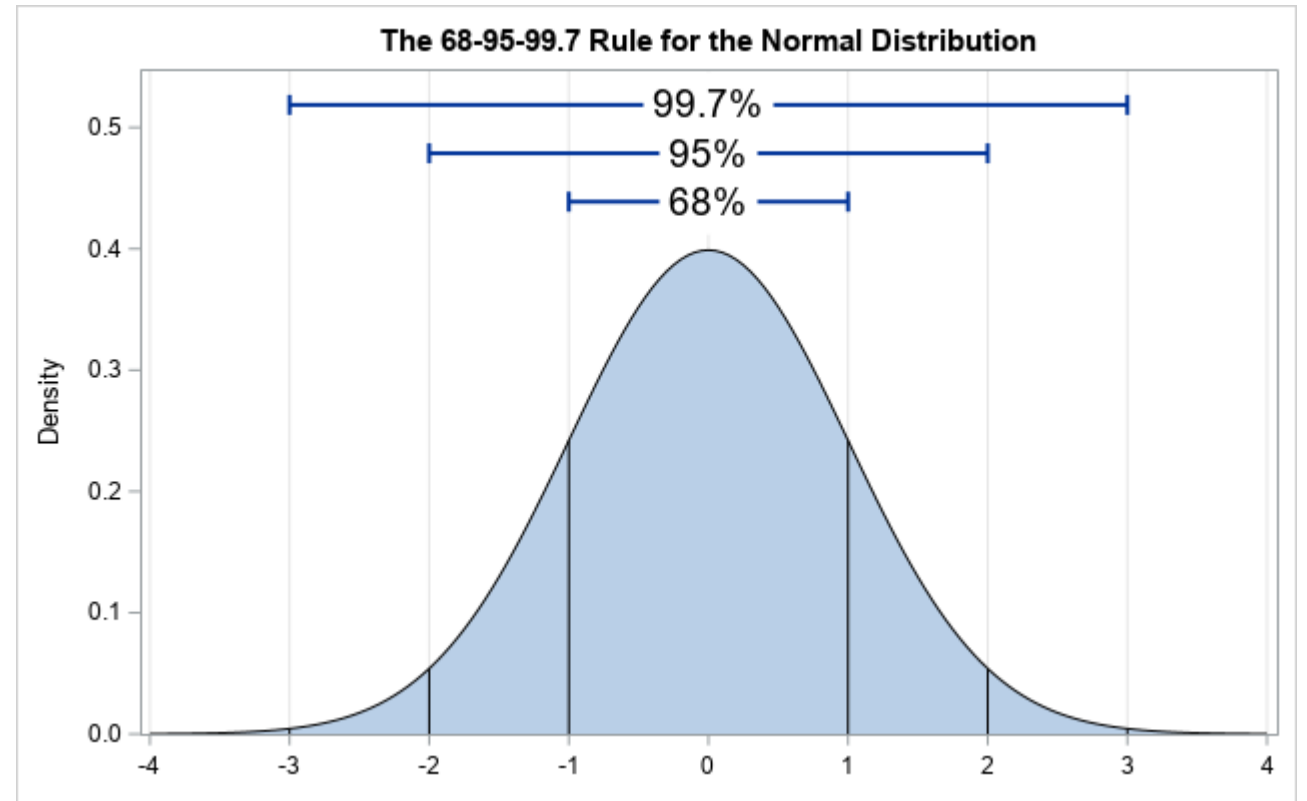
KAMK • University
of Applied Sciences

68-95-99.7 -sääntö

Normaalijakaumasta tehdyistä satunnaisista havainnoista

- ✓ **99,7 %** sijaitsee kolmen keskihajonnan sisällä keskiarvosta
- ✓ **95 %** kahden keskihajonnan sisällä keskiarvosta
- ✓ **68 %** yhden keskihajonnan sisällä keskiarvosta

Keskihajonta (2/2)



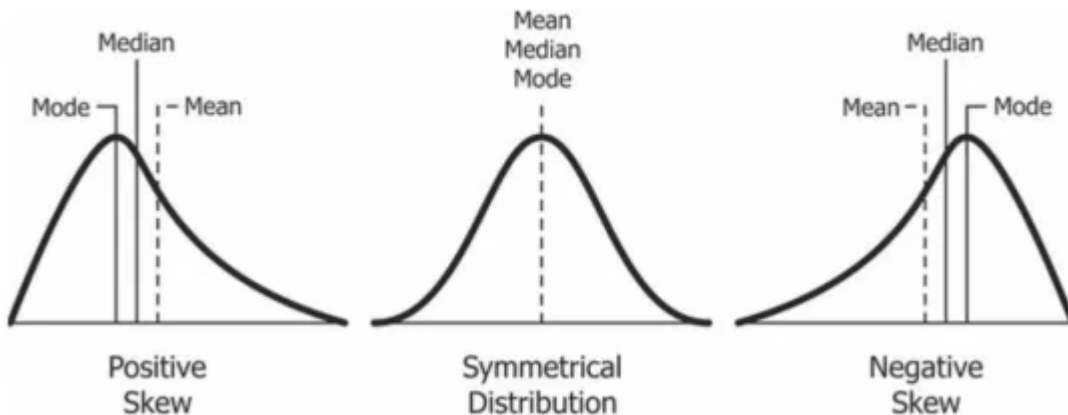
<https://blogs.sas.com/content/iml/2019/07/22/extreme-value-normal-data.html>

Vinous (skewness)

- Vinous (skewness) mittaa normaalijakauman asymmetriaa
- Vinous kertoo, kuinka paljon tarkastelun kohteena oleva todennäköisyysjakauma poikkeaa normaalijakaumasta, jonka vinous on 0.

Vinouden tyypit

- Symmetrinen: vinouskerroin on lähellä nollaa, sekä mediaani ja keskiarvo ovat likipitään samansuuruiset
- Negatiivinen vinous: Histogrammin vasen “häntä” on pidempi, ja valtaosa havainnoista keskittyy oikealle puolelle.
- Positiivinen vinous: Histogrammin oikea “häntä” on pidempi, ja valtaosa havainnoista keskittyy vasemmalle puolelle.



The skewness can be calculated from the following formula:

$$skewness = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(N-1)s^3}$$

where:

- σ is the standard deviation
- \bar{x} is the mean of the distribution
- N is the number of observations of the sample

Komento Octavessa:

- skewness
- <https://octave.sourceforge.io/octave/function/skewness.html>

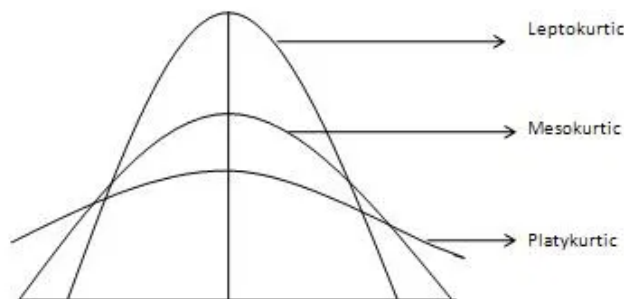
LÄHDE: <https://www.r-bloggers.com/2020/11/skewness-and-kurtosis-in-statistics/>

Kurtoosi tai huipukkuus (kurtosis)

- Kurtoosi kuvaa jakauman muotoa ja on sen huipukkuuden mitta

Huipukkuuden tyypit

- Mesokurtic = normaalijakauma, kurtoosin arvo ~3
- Leptokurtic: Jakaumalla on normaalijakaumaa terävämpi huippu ja paksummat häntäosat. Kurtoosin arvo on tällöin suurempi kuin 3
- Platykurtic: Jakaumalla on normaalijakaumaa matalampi ja laajempi huippu sekä ohuimmat häntäosat. Kurtoosin arvo on tällöin pienempi kuin 3



The kurtosis can be derived from the following formula:

$$kurtosis = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)s^4}$$

where:

- σ is the standard deviation
- \bar{x} is the mean of the distribution
- N is the number of observations of the sample

Komento Octavessa:

- kurtosis
- <https://octave.sourceforge.io/octave/function/kurtosis.html>

LÄHDE: <https://www.r-bloggers.com/2020/11/skewness-and-kurtosis-in-statistics/>

Neliöllinen keskiarvo (Root mean square, RMS)

- RMS-arvo on eräs lukujoukkoa tai jakaumaa kuvaavista matemaattisista keskiluvuista
- Vaihtojännitteen tehollisarvo lasketaan sähkötekniikassa neliöllisen keskiarvon avulla
- RMS-arvo käytetään myös esim. värähtelyn voimakkuuden ilmaisemiseen

Diskreeteille jakaumille RMS-arvon lauseke on muotoa

$$R(x) = \sqrt{\frac{\sum_{k=0}^N x_k^2}{N}}$$

Komento Octavessa:

- rms
- <https://octave.sourceforge.io/nan/function/rms.html>

Keskiarvon poistaminen ja keskihajonnan normalisoiminen (1/2)

- Datan esikäsittelyssä on usein mielekästä poistaa signaalista keskiarvo
- Esimerkiksi kiihtyvyyssanturin signaalissa saattaa olla mukana bias/offset, joka hankaloittaa tulkintoja
- Keskihajonnan normalisoimisella taas saadaan taas parannettua erilaisten datasettien vertailukelpoisuutta

$$x_{scaled} = \frac{x - mean}{sd}$$

Periaate

- Lasketaan havaintojoukon keskiarvo
- Lasketaan havaintojoukon keskihajonta
- Vähennetään ka kustakin havainnosta ja jaetaan tulos keskihajonnalla

Komennot Octavessa:

- mean
- std

Lisätietoa:

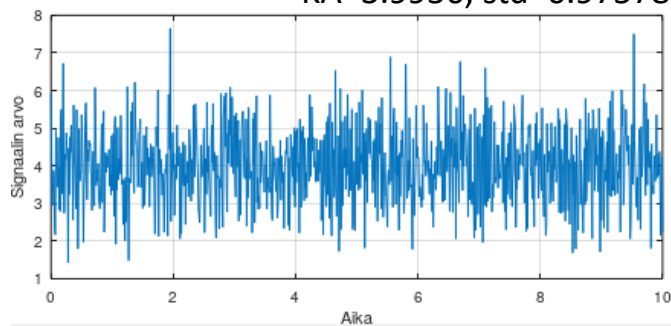
<https://stackoverflow.com/questions/61795527/removing-mean-from-data>



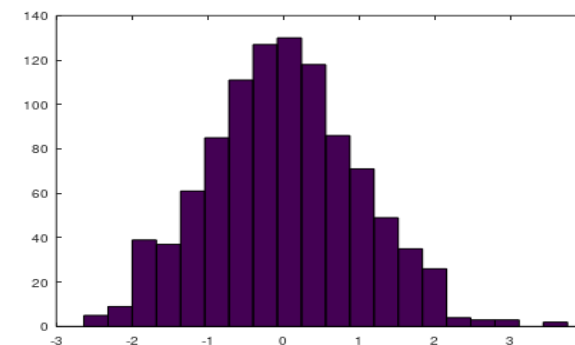
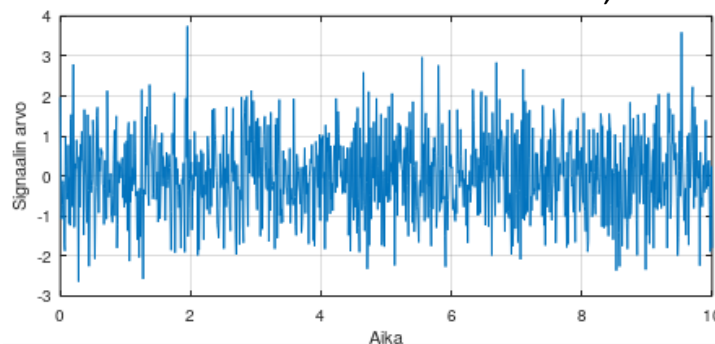
KAMK • University
of Applied Sciences

Keskiarvon poistaminen ja keskihajonnan normalisoiminen (2/2)

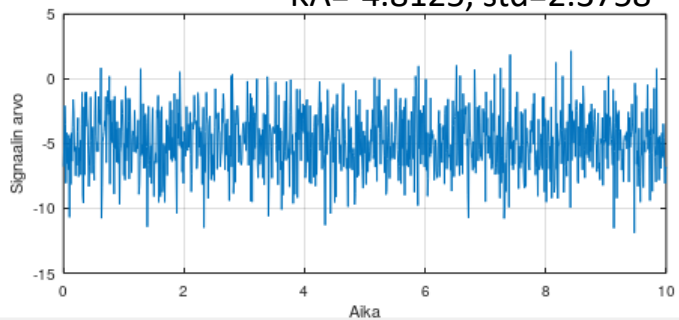
KA=3.9956, std=0.97578



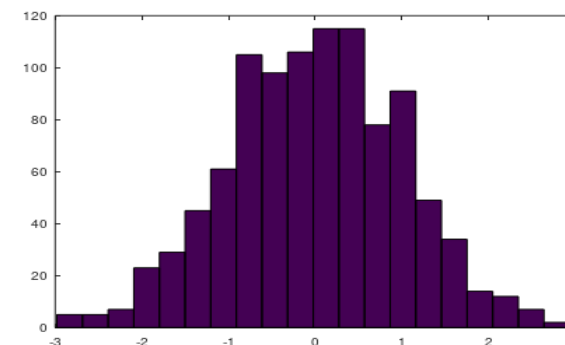
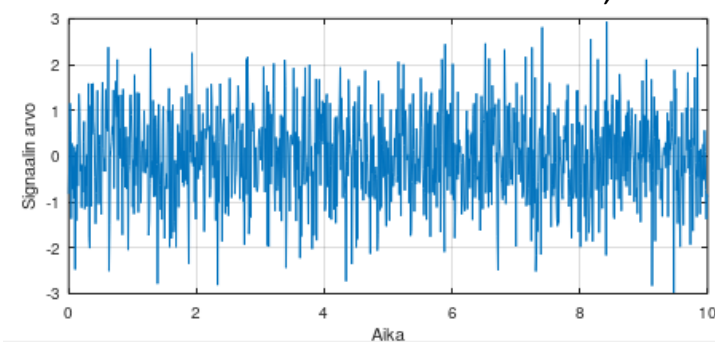
KA=0, std=1



KA=-4.8125, std=2.3758



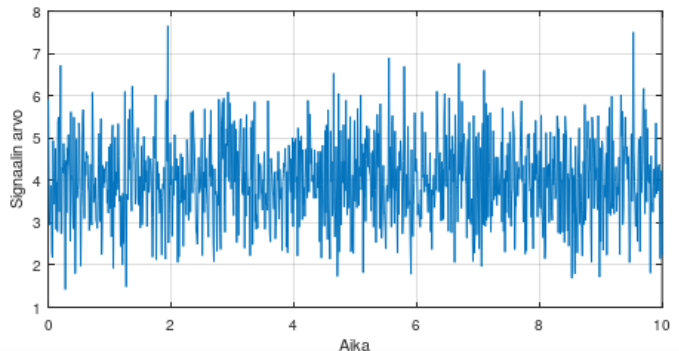
KA=0, std=1



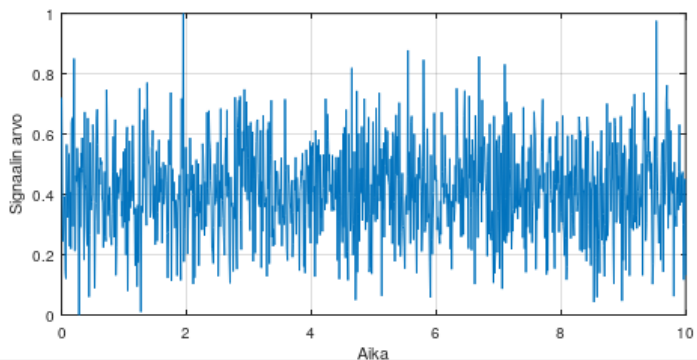
Data-aineiston skaalaaminen tietylle välille

- Tietyissä käyttötarkoituksissa data halutaan skaalata esim. välille 0...1
- Menetelmää kutsutaan nimillä feature scaling tai Min-Max feature scaling

Alkuperäinen signaali



Skaalattu signaali



Skaalaaminen välille 0...1

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

z_i = skaalattua arvo

x_i = skaalattava arvo

$\min(x)$ = havaintojoukon minimiarvo

$\max(x)$ = havaintojoukon maksimiarvo

Skaalaaminen välille a...b

$$X' = a + \frac{(X - X_{\min})(b - a)}{X_{\max} - X_{\min}}$$

Lisää aiheesta:

- [https://en.wikipedia.org/wiki/Normalization_\(statistics\)](https://en.wikipedia.org/wiki/Normalization_(statistics))
- <https://stats.stackexchange.com/questions/70801/how-to-normalize-data-to-0-1-range>
- https://docs.tibco.com/pub/spotfire/7.0.0/doc/html/norm/norm_scale_between_0_and_1.htm
- <https://www.researchgate.net/post/How-do-i-normalize-data-from-0-to-1-range>

Linkkejä ja vinkkejä

- Tilastotieteen käsitteitä ja määritelmiä: <https://www.investopedia.com/financial-term-dictionary-4769738>