Lisähuomioita Standardoinnista.

Jos standardoitu arvo on itseisarvoisesti suurempi kuin 1.5 puhutaan poikkeavista havainnoista.

Jos standardoitu arvo on itseisarvoisesti suurempi kuin 3 puhutaan merkittävästi poikkeavista havainnoista.

Vinous ja huipukkuuskertoimen laskeminen (esimerkki)

Laske seuraavalle otokselle huipukkuus- ja vinouskertoimet:

11	**12**	**15**	**20**	**22**	**26**	**28**	**32**	**35**	**40**	**45**

Nyt valmiiksi laskettu keskiarvo: 26 ja otoskeskihajonta: 11.3 (n=11)

x	**11**	**12**	**15**	**20**	**22**	**26**	**28**	**32**	**35**	**40**	**45**	286
(x- $\bar{x}$)	-15	-14	-11	-6	-4	0	2	6	9	14	19	0
(x- $\bar{x})^3$	-3375	-2744	-1331	-216	-64	0	8	216	729	2744	6859	2826
(x- $\bar{x})^4$	50625	38416	14641	1296	256	0	16	1296	6561	38416	130321	281844

Vinouskerroin : $g_1 = \frac{m_3}{s^3} = \frac{\frac{1}{11} * 2826}{11.3^3} = \frac{256.91}{1442.9} = 0.1781$

➔ Lievää positiivista vinoutta eli ”häntä” on pidempi jakauman oikealla puolella.

Huipukkuuskerroin: $g_2 = \frac{m_4}{s^4} - 3 = \frac{\frac{1}{11} * 281844}{11.3^4} - 3$

$$= \frac{25622.18}{16304.74} - 3 = -1.43$$

➔ Litteyttä havaittavissa (jakauma on käytännössä tasaisesti jakautunut!)

Harjoittelua: Laske samat aineistolla 2, 4 ,6 , 8 ,10 (ka= 6 ja kh= 2.9)

Vastaus:

x	2	4	6	8	10	30
(x- $\bar{x}$)	-4	-2	0	2	4	0
(x- $\bar{x})^3$	-64	-8	0	8	64	0
(x- $\bar{x})^4$	256	16	0	16	256	544

Vinouskerroin : $g_1 = \frac{m_3}{s^3} = \frac{\frac{1}{5}*0}{2.9^3} = \frac{0}{24.4} = 0$

➔ Ei Vinoutta

Huipukkuuskerroin: $g_2 = \frac{m_4}{s^4} - 3 = \frac{\frac{1}{5}*544}{2.9^4} = \frac{108.8}{70.73} - 3 = -1.46$

➔ Litteyttä havaittavissa (Tosin näin pienellä aineistolla tulos on odotettavissa).

Kontingenssikerroin esimerkki:

Kolmessa kaupungissa kysyttiin jokaisessa vastaantulijoilta lempijäätelömakua. Saatiin seuraavat tulokset. Onko kaupungin ja lempimaun välillä yhteyttä?

$\boldsymbol{f_{ij}}$

	Suklaa	Vanilja	Mansikka	yht.
Pori	20	10	20	50
Turku	30	50	20	100
Helsinki	50	30	40	120
yht.	100	90	80	270

$\boldsymbol{e_{ij}}$

	Suklaa	Vanilja	Mansikka	yht.
Pori	50*100/270 = 18.5	50*90/270 = 17.7	50*80/270 = 14.8	50
Turku	100*100/270 = 37.0	90*100/270 = 33.3	80*100/270 = 29.6	100
Helsinki	120*100/270 = 44.4	120*90/270 = 40.0	120*80/270 = 35.6	120
yht.	100	90	80	270

$$\chi^2 = \frac{(20-18.5)^2}{18.5} + \frac{(10-17.7)^2}{17.7} + \frac{(20-14.8)^2}{14.8} +$$

$$\frac{(30-37.0)^2}{37.0} + \frac{(50-33.3)^2}{33.3} + \frac{(40-29.6)^2}{29.6} +$$

$$\frac{(50-44.4)^2}{44.4} + \frac{(30-40.0)^2}{40.0} + \frac{(40-35.6)^2}{35.6}$$

Käytetään apuna taulukointia!

f_i	e_i	$(f_i\text{-}e_i)^2$	$(f_i\text{-}e_i)^2/\,e_i$
20	**18.5**	2.25	0.12
10	**17.7**	59.29	3.35
20	**14.8**	27.04	1.83
30	**37**	49	1.32
50	**33.3**	278.89	8.38
20	**29.6**	92.16	3.11
50	**44.4**	31.36	0.71
30	**40.0**	100	2.50
40	**35.6**	19.36	0.54
			21.82

Kontingenssikerroin C: $\sqrt{\frac{\chi^2}{\chi^2+n}} = \sqrt{\frac{21.82}{21.82+270}} = \mathbf{0.274}$

Teoreettinen maksimi C_{max} = $\sqrt{\frac{q-1}{q}}$ = $\sqrt{\frac{3-1}{3}}$ = 0.82

Vertailukelpainen kontingenssikerroin: $\frac{C}{C_{max}} = \frac{0.274}{0.82}$ **= 0,33**

- ➔ **Muuttujien välillä on yhteyttä. Rajajakaumat paljastavat kaupunkien profiilit**

	Suklaa	**Vanilja**	**Mansikka**	**yht.**
Pori	**40 %**	**20 %**	**40 %**	**100 %**
Turku	**30 %**	**50 %**	**20 %**	**100 %**
Helsinki	**42 %**	**25 %**	**33 %**	**100 %**