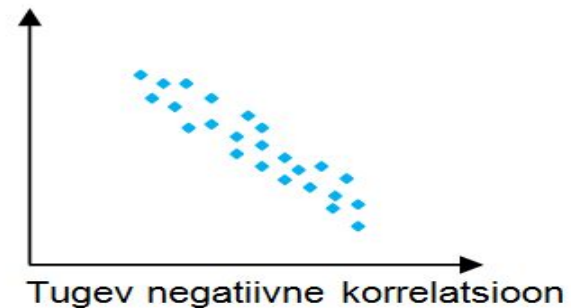
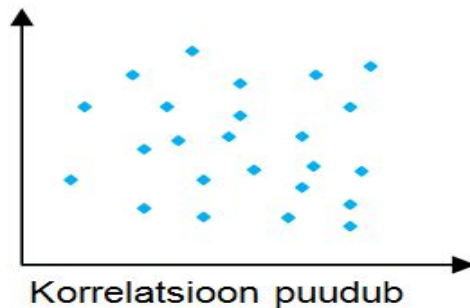
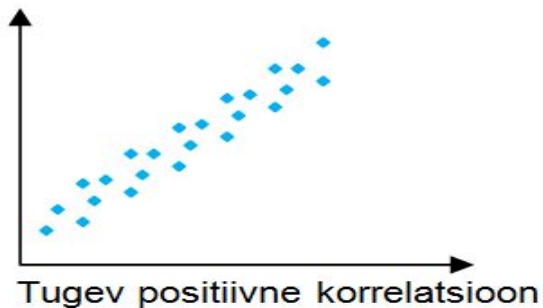




# Korrelatsioonanalüüs

# Korrelatsioon - seose hindamine



-1

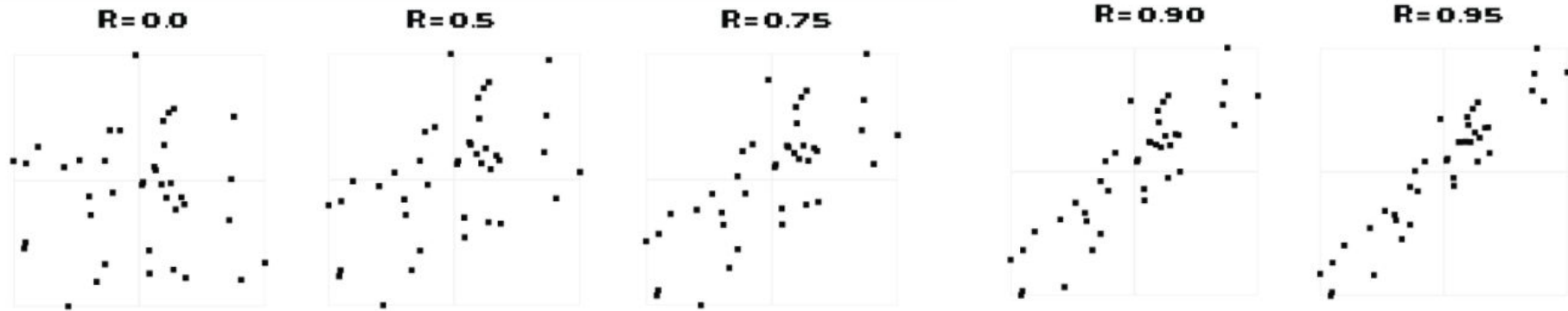
0

+1

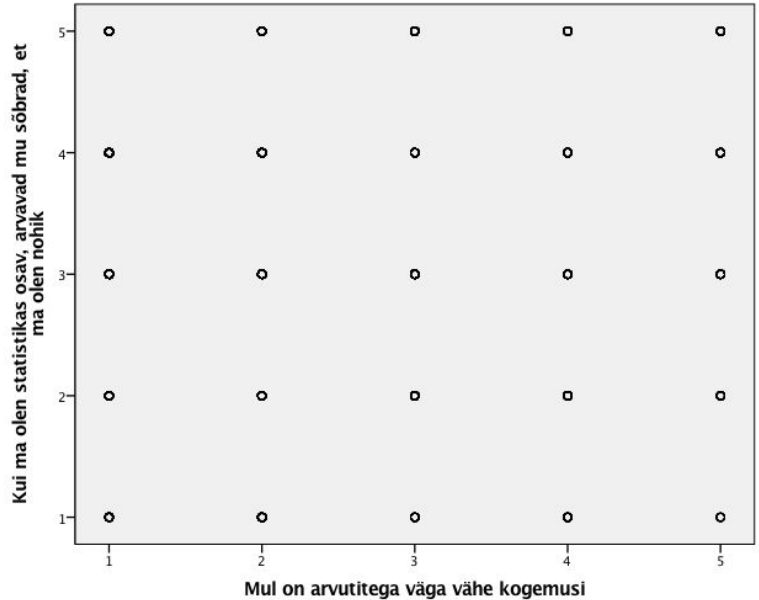
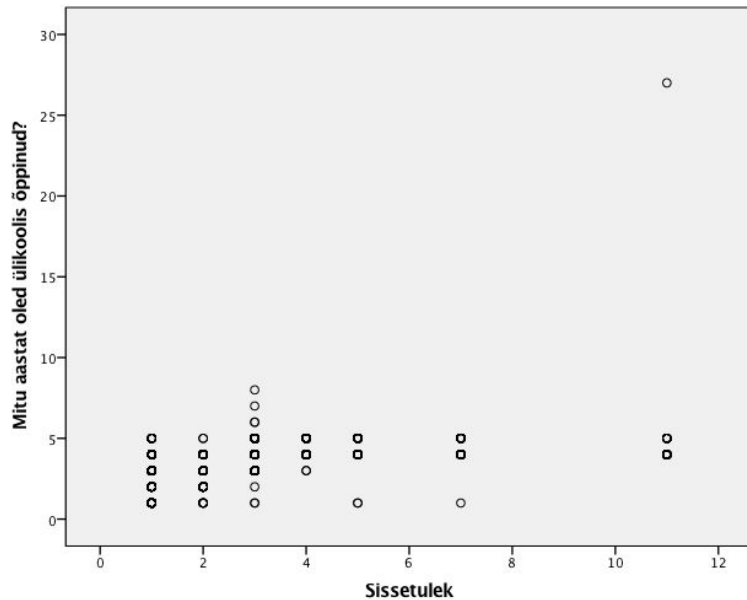
-1...-0,9	-0,9...-0,7	-0,7...-0,3	-0,3...-0,2	-0,2...-0,1	-0,1...0,1	0,1...0,2	0,2...0,3	0,3...0,7	0,7...0,9	0,9...1
Väga tugev negatiivne seos	Tugev negatiivne seos	Keskmine negatiivne seos	Nõrk negatiivne seos	Väga nõrk negatiivne seos	Seos puudub	Väga nõrk positiivne seos	Nõrk positiivne seos	Keskmine positiivne seos	Tugev positiivne seos	Väga tugev positiivne seos



# Seose tugevuse hindamine



# Korrelatsiooniväli e hajuvusdiagramm





# Seose tugevus

## D. Rowntree

0.0 - 0.2 olematu, väga nõrk

0.2 - 0.4 nõrk

0.4 - 0.7 keskmine

0.7 - 0.9 tugev

0.9 - 1.0 väga tugev

## A. Field

0.0 - 0.1 olematu

0.1-0.2 väga nõrk

0.2 - 0.3 nõrk

0.3 - 0.7 keskmine

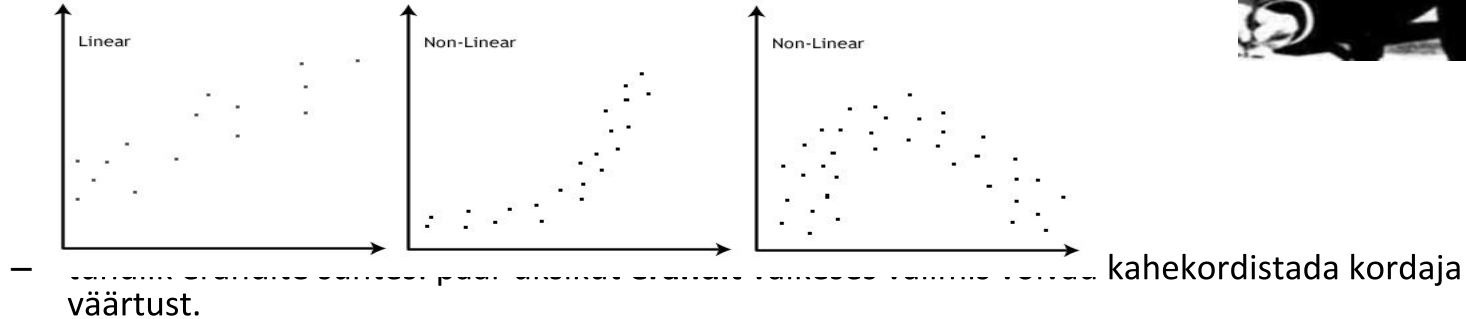
0.7 - 0.9 tugev

0.9 - 1.0 väga tugev

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

# Pearson r

- Standardiseeritud kahe tunnuse vahelise seose kordaja
- Pearsoni kordaja puudused
  - lineaarne seos: tunneb punktipilve, mis on venitatud piki sirget.



# Näide 1

Intervalltunnus: Statistika ajab mind nutma (1 täiesti nõus .... 5 üldse ei ole nõus)

Intervalltunnus: Ma ei saa statistikast aru (1 täiesti nõus .... 5 üldse ei ole nõus)

## Correlations

		Statistika ajab mind nutma	Ma ei saa statistikast aru
Statistika ajab mind nutma	Pearson Correlation	1	.402**
	Sig. (2-tailed)		.000
	N	2571	2571
Ma ei saa statistikast aru	Pearson Correlation	.402**	1
	Sig. (2-tailed)	.000	
	N	2571	2571

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Pearsoni  
korrelatsioonikordaja  
 $r=0,4$

**Keskmise tugevusega  
positiivne seos ( $r=0,4$ ).** Mida  
rohkem ei saanud vastaja  
statistikast aru, seda enam  
ajas statistika teda nutma ja  
vastupidi.

# Näide 6

Intervalltunnus: Statistika ajab mind nutma (1 täiesti nõus .... 5 üldse ei ole nõus)

Intervalltunnus: Mitu aastat oled ülikoolis õppinud?

Correlations

		Statistika ajab mind nutma	Mitu aastat oled ülikoolis õppinud?
Statistika ajab mind nutma	Pearson Correlation	1	.110**
	Sig. (2-tailed)		.000
	N	2571	2503
Mitu aastat oled ülikoolis õppinud?	Pearson Correlation	.110**	1
	Sig. (2-tailed)	.000	
	N	2503	2503

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Correlations

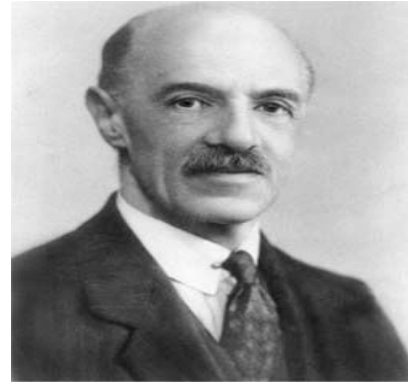
		Statistika ajab mind nutma	Mitu aastat oled ülikoolis õppinud?
Spearman's rho	Correlation Coefficient	1.000	.122**
	Sig. (2-tailed)	.	.000
	N	2571	2503
Mitu aastat oled ülikoolis õppinud?	Correlation Coefficient	.122**	1.000
	Sig. (2-tailed)	.000	.
	N	2503	2503

\*\* . Correlation is significant at the 0.01 level (2-tailed).





# Spearman roo



- MPAR korrelatsioonikordaja
- Astakorrelatsioonikordaja
  - intervalltunnused ei vasta normaaljaotusele (ka erandlikud väärtused)
  - Järjestustunnus(ed)
  - asendab väärtused järjekorranumbritega ning kasutab Pearsoni kordaja valemit => Spearmani kordaja < Pearsoni kordaja (üldjuhul)



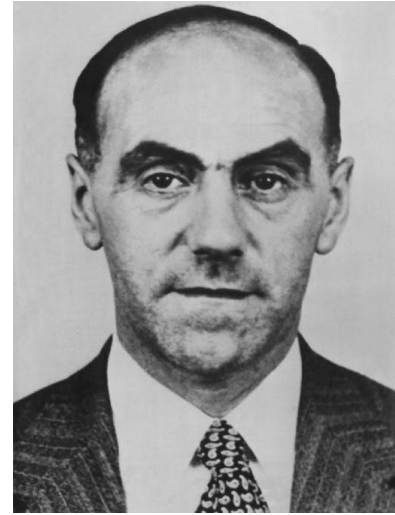
# PAR kordajad

- **Pearsoni** kordaja **puudused**
  - Lineaarne seos: tunneb punktipilve, mis on venitatud piki sirget.
  - Tundlik erandite suhtes.
- **Spearman** e. astakkorrelatsioonikordaja
  - Pidevad tunnused ei ole normaaljaotusega (ka erandlikud väärtused)
  - Järjestustunnus
  - Spearmanni kordaja > Pearsoni kordaja (tavaliselt)
- **Kendall**
  - Järjestustunnused, palju sarnaseid väärtuseid, lühikesed skaalad
  - Samasuunaliste ja vastassuunaliste paaride analüüs.

---

# Cramer V

- Nimitunnuste seose tugevuse uurimiseks.  
Kordaja ei näita seose suunda, ainult tugevust.
- Phi  
2x2 binaarsete tunnuste korral  
Ruutjuur hii-ruudu väärtuse jagatisest valimi suurusega



# Näide 11

Järjestustunnus: Sissetulek

Nimitunnus: Töötamine (2 ei töötä, 3 täistööajaga, 4 osalise koormusega)

Kas sa töötad? \* Sissetulek Crosstabulation

			Sissetulek						rohkem kui 4000\$	Total
			kuni 1000\$	kuni 1500\$	kuni 2000\$	kuni 2500\$	kuni 3000\$	kuni 4000\$		
Kas sa töötad?	ei töötä	Count	234	490	260	132	19	5	6	1146
		% within Kas sa töötad?	20.4%	42.8%	22.7%	11.5%	1.7%	0.4%	0.5%	100.0%
	töötab täistööajaga	Count	198	362	174	81	18	4	9	846
		% within Kas sa töötad?	23.4%	42.8%	20.6%	9.6%	2.1%	0.5%	1.1%	100.0%
	töötab osalise koormusega	Count	58	90	61	58	44	40	45	396
		% within Kas sa töötad?	14.6%	22.7%	15.4%	14.6%	11.1%	10.1%	11.4%	100.0%
Total		Count	490	942	495	271	81	49	60	2388
		% within Kas sa töötad?	20.5%	39.4%	20.7%	11.3%	3.4%	2.1%	2.5%	100.0%

## Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.429	.000
	Cramer's V	.303	.000
N of Valid Cases		2388	



<b>KORDAJA</b>	<b>ANDMED</b>	<b>LISATINGIMUS</b>
PEARSON	I + I	Seose kuju – lineaarne Erandlikud väärtused puuduvad (ei domineeri)
SPEARMAN	I + I, I + J, J + J	Seose kuju ei ole lineaarne ja jaotusel on erandlikud väärtused (I + I)
CRAMER	I + N, I + B, N + B, N + N, N + J, B + J	Tõlgendatakse vaid seose tugevust, mitte suunda