

16.15 - 16.45 K7 Dataanalyseekspressen. Hvordan ser du ut i det multivariate rom

Egil Nodland, Universitetet i Bergen

www.uib.no

- Datasettet innsamlet i morges brukes som eksempel på

1. generell datautforskning (PCA),
2. klassifisering (SIMCA) og
3. kvantitativ modellering (PLS).

Orange
Orange
Orange
Orange
Orange

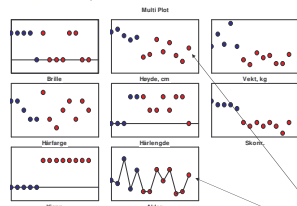
www.uib.no

Datasettet

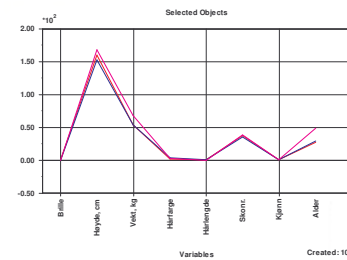
	Britle	Hayde, cm	Vekt, kg	Hårfarge	Hårlengde	Skinnr.	Kjennr.	Alder
Egil	1	186	80	4	0	45	0	43
Ingrid	0	161	53	2	0.5	38	1	30
Terje	1	182	83	3	0	44	0	67
Raymond	1	177	114	2	1	44	0	33
Ellen	0	168	68	3	0	39	1	49
Fatima	1	158	60	5	1	39	1	30
Håvard	1	189	98	4	0	44	0	39
Anne	0	178	70	1	0.5	39	1	55
Elisabeth	0	164	63	3	1	38	1	42
Marit	1	174	66	4	0.5	39	1	55
Silje	1	160	54	2	1	38	1	28
Kjersti	0	153	54	4	1	36	1	30
Dag Olav	0	180	80	2	1	43	0	54

www.uib.no

Univariate plott ?

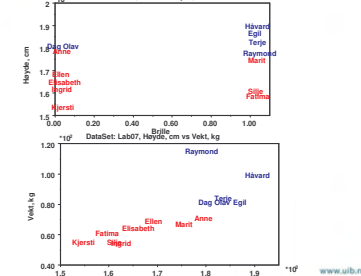


www.uib.no



www.uib.no

Bivariate plott?



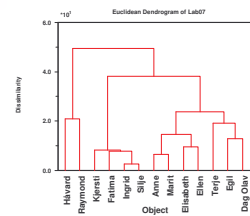
www.uib.no

Korrelasjoner mellom alle personene?

	Egil	Håvard	Terje	Anne	Elisabeth	Marit	Silje	Kjersti	Ellen
Ellen	0.997	0.989	0.996	1	0.999	0.999	0.991	0.994	1
Anne	0.995	0.985	0.997	1	0.998	1	0.989	0.992	1
Marit	0.994	0.982	0.996	1	0.998	1	0.989	0.992	0.999
Elisabeth	0.998	0.989	0.992	0.998	1	0.998	0.996	0.998	0.997
Egil	1	0.996	0.99	0.995	0.998	0.994	0.995	0.997	0.996
Terje	0.99	0.983	1	0.997	0.992	0.996	0.977	0.981	0.994
Kjersti	0.997	0.986	0.981	0.992	0.998	0.992	1	1	0.991
Silje	0.995	0.984	0.977	0.989	0.996	0.989	1	1	0.994
Håvard	0.996	1	0.983	0.985	0.989	0.982	0.984	0.986	0.989

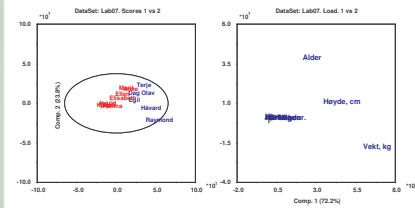
www.uib.no

Dendrogram ?



www.uib.no

Skåreplott?



www.uib.no

Maksimum-varians

	Breite	Hoeyde, cm	Vekt, kg	Hairfarge	Hairlengde	Skone	Kjenn	Alder
Egil	1	186	80	4	0	45	0	41
Åge	1	184	73	4	0	44	0	49
Anders	1	189	82	2	0	44	0	24
Pla	1	168	58	3	1	37	1	23
k1	1	169	68	4	1	38	1	46
k2	1	174	73	4	1	41	1	35
k3	1	170	85	3	1	40	1	27
hilde	0	159	57	4	1	38	1	22
amh	1	163	57	3	1	37	1	24
kristine	0	163	55	4	1	38	1	23
Sum	9	1735	688	35	7	422	7	314
Mean	1	173	69	4	1	40	1	31
Variance	0	110	132	1	0	10	0	110
Std. deviation	0	10	11	1	0	3	0	10

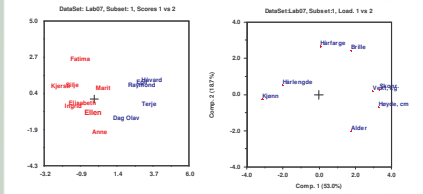
www.uib.no

Lik varians etter standardisering

	Breite	Hoeyde, cm	Vekt, kg	Hairfarge	Hairlengde	Skone	Kjenn	Alder
Egil	2	18	7	6	0	14	0	4
Åge	2	18	6	6	0	14	0	5
Anders	2	18	7	3	0	14	0	2
Pla	2	16	5	4	2	12	2	2
k1	2	16	6	6	2	12	2	4
k2	2	17	6	6	2	13	2	3
k3	2	16	7	4	2	13	2	3
hilde	0	15	5	6	2	12	2	2
amh	2	16	5	4	2	12	2	2
kristine	0	16	5	6	2	12	2	2
Sum	19	164	60	49	14	129	14	30
Mean	2	16	6	5	1	13	1	3
Variance	1	1	1	1	1	1	1	1
Std. deviation	1	1	1	1	1	1	1	1

www.uib.no

PCA etter standardisering



www.uib.no

SIMCA, klassifisering

1. "u"kontrollert, PCA avdekker mulige klasser
2. Kontrollert, vi vet på forhånd at det finnes klasser
3. Tilordning av prøver med ukjent klasseslithørighet

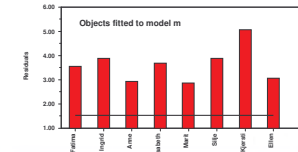
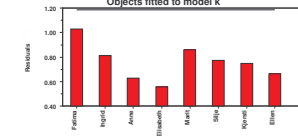
www.uib.no

2. Kontrollert klassifisering

1. Lag separate modeller for hver klasse (k/m)
2. Fjern utliggere
3. Test modellen

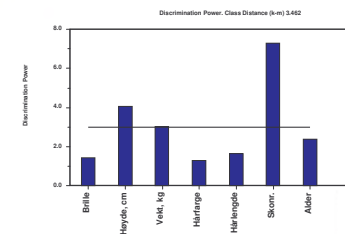
www.uib.no

Objects fitted to model k



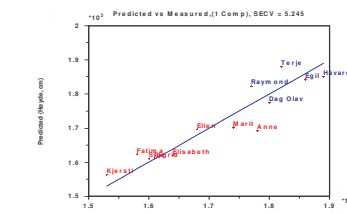
www.uib.no

Hvilke variable skiller kvinner og menn?



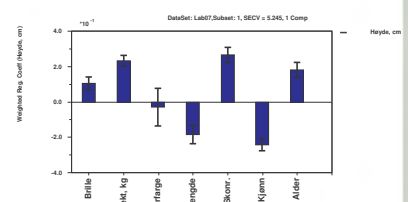
www.uib.no

Kvantitativ modellering



www.uib.no

Hvilke variabler korrelerer hvorledes med kroppshøyde



www.uib.no

