Do I measure what I want to measure?

Questionnaire evaluation with Factor analysis and Cronbach's alpha

Melanie Hof

How do we measure

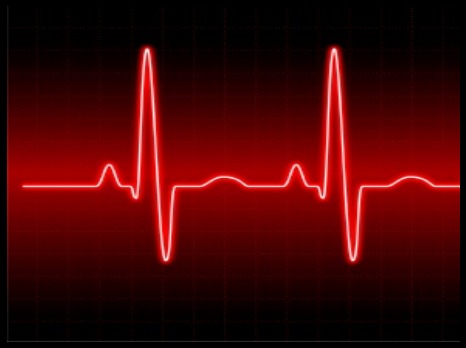
?

- Direct:
  - Writing is something I

1 2 3 4 5 6 7

- Indirect:
  - heart rate

- Direct:
  - Writing is something, I
    like 1 2 3 4 5 6 7 dislike
- Indirect:
  - heart rate



- Quasi-direct:
  - I love writing

    completely agree – agree – don't know – disagree – completely disagree

    (Likert scale)

Overview

- Description study
- Factor analysis
  - validity
- Cronbach's alpha
  - reliability

Study

- 114 seventh graders
  - half control group
  - half experimental group
- learn to write expository texts in Dutch class
  - experimental group: wrote these texts in Dutch, Science and History class
  - control group: just in Dutch class

Study

- measured effect of intervention on:
  - writing ability
  - knowledge growth
  - attitude against writing
    - confidence
    - passion

FACTOR ANALYSIS

Correlation matrix (R) (Field, 2009, p.629)

	Talk about other person	Social skills	Interest	Talk about oneself	Selfish	Liar
Talk about other person	1.000	.772	.646	.074	-.131	.068
Social skills	.772	1.000	.879	-.120	.031	.012
Interest	.646	.879	1.000	.054	-.101	.110
Talk about oneself	.074	-.120	.054	1.000	.441	.361
Selfish	-.131	.031	-.101	.441	1.000	.277
Liar	.068	.012	.110	.361	.277	1.000

Used to

- understand structure latent variables

Correlation matrix (R) (Field, 2009, p.629)

	Talk about other person	Social skills	Interest	Talk about oneself	Selfish	Liar
Talk about other person	1.000	.772	.646	.074	-.131	.068
Social skills	.772	1.000	.879	-.120	.031	.012
Interest	.646	.879	1.000	.054	-.101	.110
Talk about oneself	.074	-.120	.054	1.000	.441	.361
Selfish	-.131	.031	-.101	.441	1.000	.277
Liar	.068	.012	.110	.361	.277	1.000

Used to

- understand structure latent variables
- construct questionnaire to measure latent variables

Correlation matrix (R) (Field, 2009, p.629)

	Talk about other person	Social skills	Interest	Talk about oneself	Selfish	Liar
Talk about other person	1.000	.772	.646	.074	-.131	.068
Social skills	.772	1.000	.879	-.120	.031	.012
Interest	.646	.879	1.000	.054	-.101	.110
Talk about oneself	.074	-.120	.054	1.000	.441	.361
Selfish	-.131	.031	-.101	.441	1.000	.277
Liar	.068	.012	.110	.361	.277	1.000

Used to

- understand structure latent variables
- construct questionnaire to measure latent variables
- reduce data set
  - solving problem of multicollinearity in multiple regression

Prerequisites

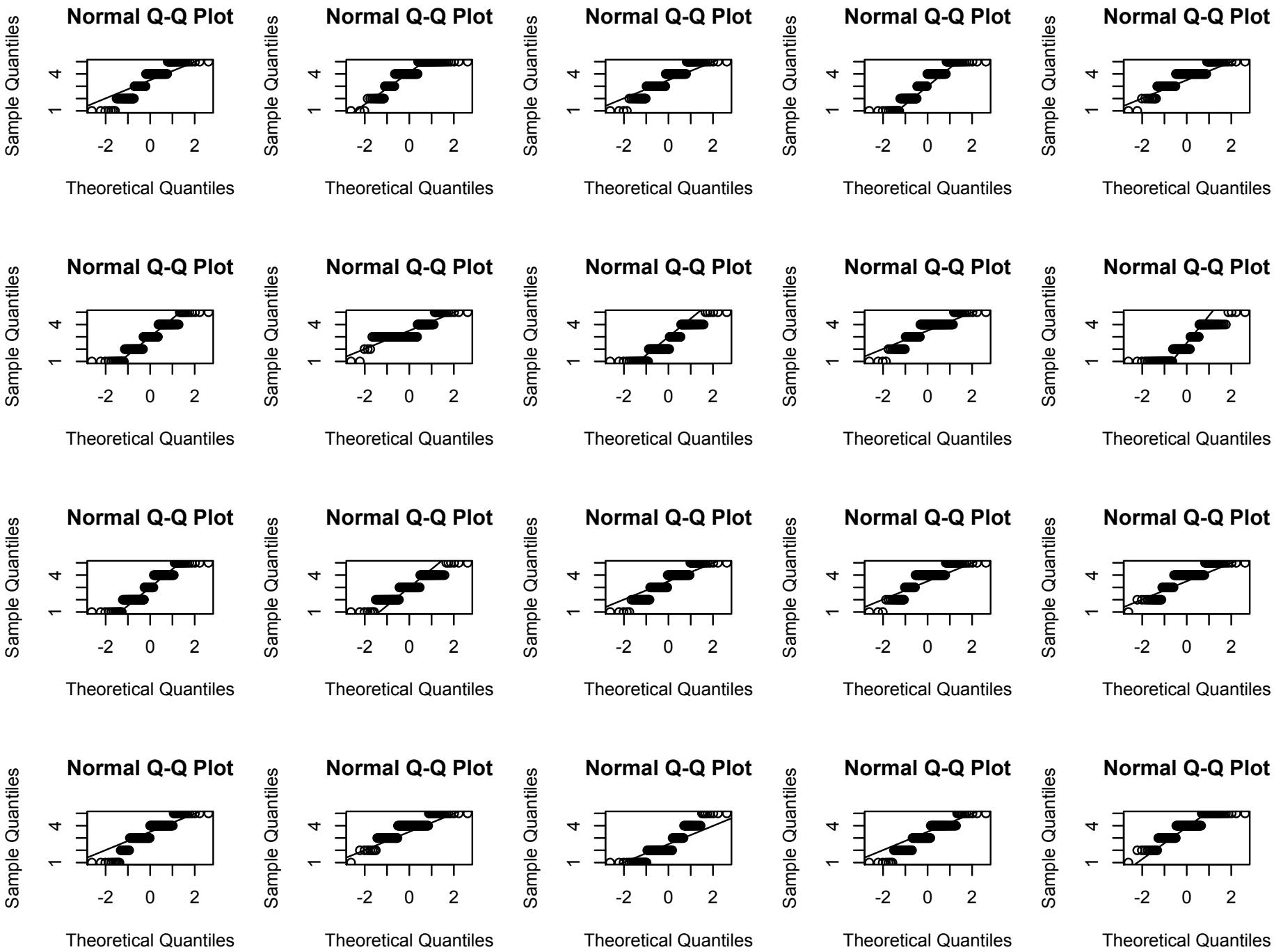
- sample size: 114
  - rule of thumb: “at least 10-15 per variable”
  - depends on communalities between variables
  - KMO: .922 > superb

Prerequisites

- ✓ sample size
- interval data

Prerequisites

- ✓ sample size
- ✓ interval data
- normally distributed


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
1
4
-2
0
2

Prerequisites

- ✓ sample size
- ✓ interval data
- ✓ normally distributed
- variables should not correlate too highly

r < .8

	pass01v	pass02v	pass03v	pass04v	pass05v	pass06v
pass01v	1.0000000	0.6624426	0.6102046	-0.5623126	0.4137463	-0.3889058
pass02v	0.6624426	1.0000000	0.6220010	-0.4148258	0.3525541	-0.2959315
pass03v	0.6102046	0.6220010	1.0000000	-0.5162518	0.4744346	-0.3440221
pass04v	-0.5623126	-0.4148258	-0.5162518	1.0000000	-0.2205194	0.5098645
pass05v	0.4137463	0.3525541	0.4744346	-0.2205194	1.0000000	-0.2213123
pass06v	-0.3889058	-0.2959315	-0.3440221	0.5098645	-0.2213123	1.0000000
pass07v	0.3662825	0.2458077	0.4317847	-0.3205762	0.4838719	-0.3388872
pass08v	-0.4581692	-0.3411867	-0.3934714	0.4738564	-0.3359287	0.5555088
pass09v	0.6159883	0.5825688	0.5945829	-0.4663368	0.5447457	-0.3214989
pass10v	-0.3733701	-0.2031242	-0.3347282	0.4608984	-0.2616383	0.4390695
pass11v	-0.5248995	-0.3382241	-0.4047499	0.6089488	-0.2744794	0.4557957
pass12v	0.4161127	0.4157369	0.5139252	-0.4409926	0.2371068	-0.3090945
pass13v	0.4858500	0.4732903	0.4682252	-0.3126358	0.3290753	-0.2775079
pass14v	0.5700604	0.5339748	0.6003164	-0.3114945	0.4553965	-0.3018648
pass15v	0.4704597	0.5000866	0.6000588	-0.3505059	0.4783165	-0.3733202
pass16v	-0.4074288	-0.2880072	-0.3853428	0.6665053	-0.2136700	0.5423090
pass17v	0.5973765	0.6120596	0.6903767	-0.4340764	0.4997126	-0.3975682
pass18v	-0.4726983	-0.2956505	-0.4005931	0.5374775	-0.2333723	0.4491783
pass19v	0.5509944	0.5041410	0.5126616	-0.3588532	0.2494718	-0.4012147
pass20v	0.5158979	0.5890959	0.5696296	-0.4364437	0.4302584	-0.3998005

Prerequisites

- ✓ sample size
- ✓ interval data
- ✓ normally distributed
- ✓ variables should not correlate too highly ($r<.8$)
- variables should correlate enough

r >.3

	pass01v	pass02v	pass03v	pass04v	pass05v	pass06v
pass01v	1.0000000	0.6624426	0.6102046	-0.5623126	0.4137463	-0.3889058
pass02v	0.6624426	1.0000000	0.6220010	-0.4148258	0.3525541	-0.2959315
pass03v	0.6102046	0.6220010	1.0000000	-0.5162518	0.4744346	-0.3440221
pass04v	-0.5623126	-0.4148258	-0.5162518	1.0000000	-0.2205194	0.5098645
pass05v	0.4137463	0.3525541	0.4744346	-0.2205194	1.0000000	-0.2213123
pass06v	-0.3889058	-0.2959315	-0.3440221	0.5098645	-0.2213123	1.0000000
pass07v	0.3662825	0.2458077	0.4317847	-0.3205762	0.4838719	-0.3388872
pass08v	-0.4581692	-0.3411867	-0.3934714	0.4738564	-0.3359287	0.5555088
pass09v	0.6159883	0.5825688	0.5945829	-0.4663368	0.5447457	-0.3214989
pass10v	-0.3733701	-0.2031242	-0.3347282	0.4608984	-0.2616383	0.4390695
pass11v	-0.5248995	-0.3382241	-0.4047499	0.6089488	-0.2744794	0.4557957
pass12v	0.4161127	0.4157369	0.5139252	-0.4409926	0.2371068	-0.3090945
pass13v	0.4858500	0.4732903	0.4682252	-0.3126358	0.3290753	-0.2775079
pass14v	0.5700604	0.5339748	0.6003164	-0.3114945	0.4553965	-0.3018648
pass15v	0.4704597	0.5000866	0.6000588	-0.3505059	0.4783165	-0.3733202
pass16v	-0.4074288	-0.2880072	-0.3853428	0.6665053	-0.2136700	0.5423090
pass17v	0.5973765	0.6120596	0.6903767	-0.4340764	0.4997126	-0.3975682
pass18v	-0.4726983	-0.2956505	-0.4005931	0.5374775	-0.2333723	0.4491783
pass19v	0.5509944	0.5041410	0.5126616	-0.3588532	0.2494718	-0.4012147
pass20v	0.5158979	0.5890959	0.5696296	-0.4364437	0.4302584	-0.3998005

Barlett's test

$$R_{\text{- matrix}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

> cortest.bartlett([dataset])

$\chi^2 (190) = 1263.862, p = 7.117332e\text{-}158$

Prerequisites

- ✓ sample size
- ✓ interval data
- ✓ normally distributed
- ✓ variables should not correlate too highly ($r<.8$)
- ✓ variables should correlate enough

Correlation matrix (R) (Field, 2009, p.629)

	Talk about other person	Social skills	Interest	Talk about oneself	Selfish	Liar
Talk about other person	1.000	.772	.646	.074	-.131	.068
Social skills	.772	1.000	.879	-.120	.031	.012
Interest	.646	.879	1.000	.054	-.101	.110
Talk about oneself	.074	-.120	.054	1.000	.441	.361
Selfish	-.131	.031	-.101	.441	1.000	.277
Liar	.068	.012	.110	.361	.277	1.000

Total variance for a variable

=

common variance + unique variance

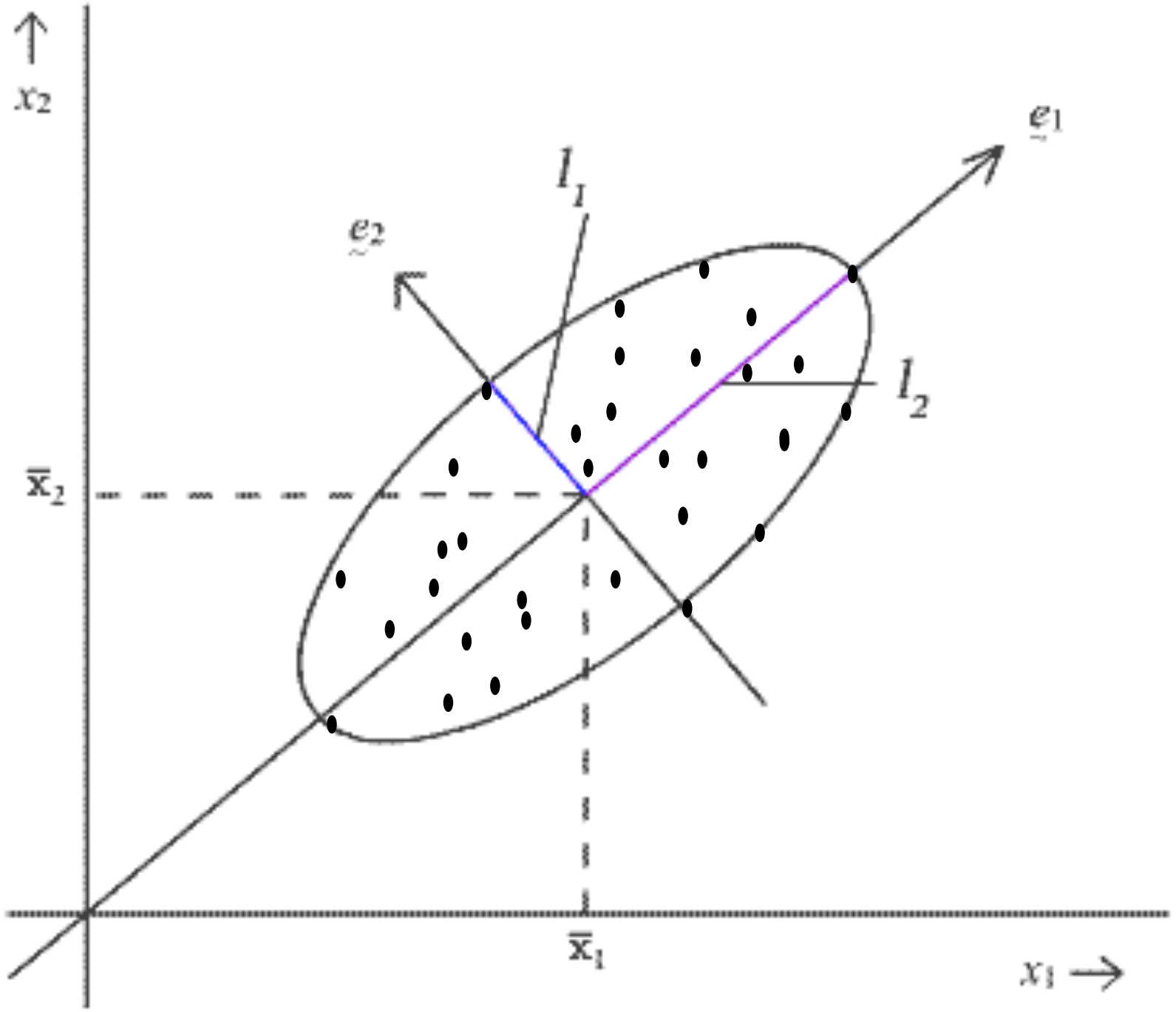
+ random variance

Total variance for a variable

=

communality + unique variance

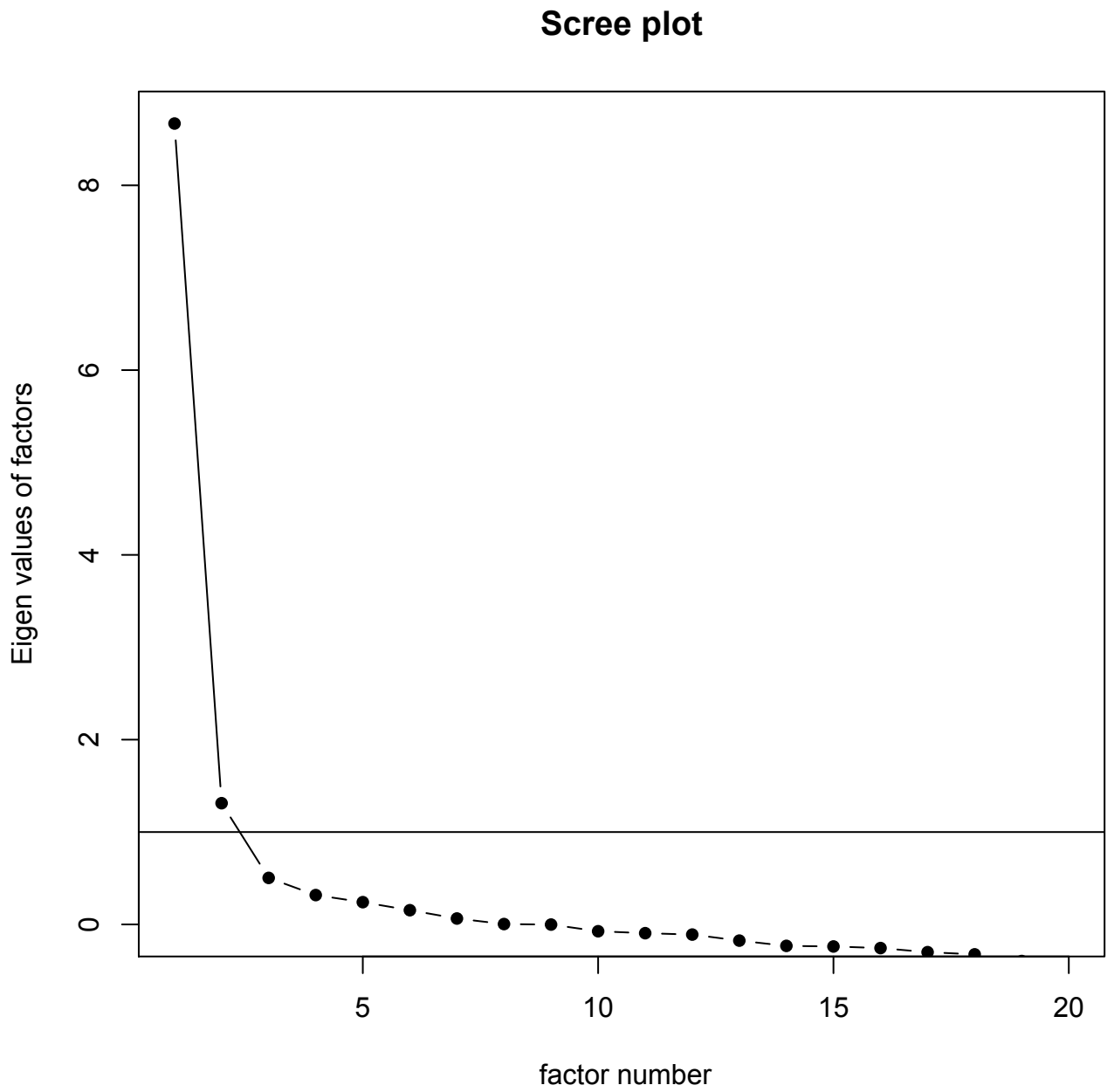
+ random variance


x_2
$\underset{\sim}{e}_1$
l_1
$\underset{\sim}{e}_2$
l_2
$\bar{x}_2$
$\bar{x}_1$
x_1

How many factors to retain?

- screeplot

```
> scree(cor(passion, use="complete.obs"), factors=T, pc=F)
```



How many factors to retain?

- screeplot: 2
- Guttman-Kaiser rule: factors with an eigenvalue larger than 1

Factor analysis

```
> m31 = factanal(na.omit(passion), factors=3, rotation='oblimin')
> print(m31)

                Factor1 Factor2 Factor3
SS loadings       5.963   3.539   0.585
Proportion Var    0.298   0.177   0.029
Cumulative Var    0.298   0.475   0.504
```

How many factors to retain?

- screeplot: 2
- Guttman-Kaiser rule: factors with an eigenvalue larger than 1: 2
- depends on proportion of variance explained

Factor analysis

```
Loadings:
        Factor1 Factor2                 Factor1 Factor2
pass01v  0.547  -0.289         pass11v          0.680
pass02v  0.747                 pass12v  0.412  -0.243
pass03v  0.727                 pass13v  0.719   0.122
pass04v          0.802         pass14v  0.739
pass05v  0.625                 pass15v  0.758
pass06v          0.641         pass16v          0.771
pass07v  0.463                 pass17v  0.917
pass08v -0.133   0.558         pass18v          0.739
pass09v  0.702  -0.115         pass19v  0.623
pass10v          0.597         pass20v  0.771

                Factor1 Factor2
SS loadings       6.141   3.534
Proportion Var    0.307   0.177
Cumulative Var    0.307   0.484
```

Factor analysis

```
> m21 = factanal(na.omit(passion), factors=2, rotation='oblimin')
> print(m21)
```

Gives maximum-likelihood factor analysis:

$\chi^2(151)=197.76$, $p=0.00636$

Factor analysis

```
> m31 = factanal(na.omit(passion), factors=3, rotation='oblimin')
> print(m31)
```

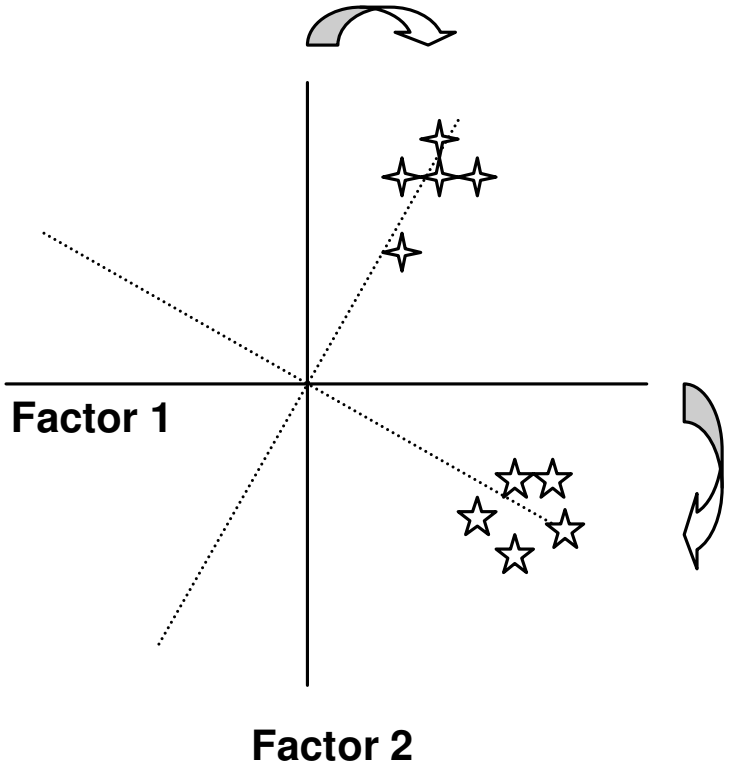
	Factor1	Factor2	Factor3
SS loadings	5.963	3.539	0.585
Proportion Var	0.298	0.177	0.029
Cumulative Var	0.298	0.475	0.504

$\chi^2(133)=153.68$, $p=0.106$

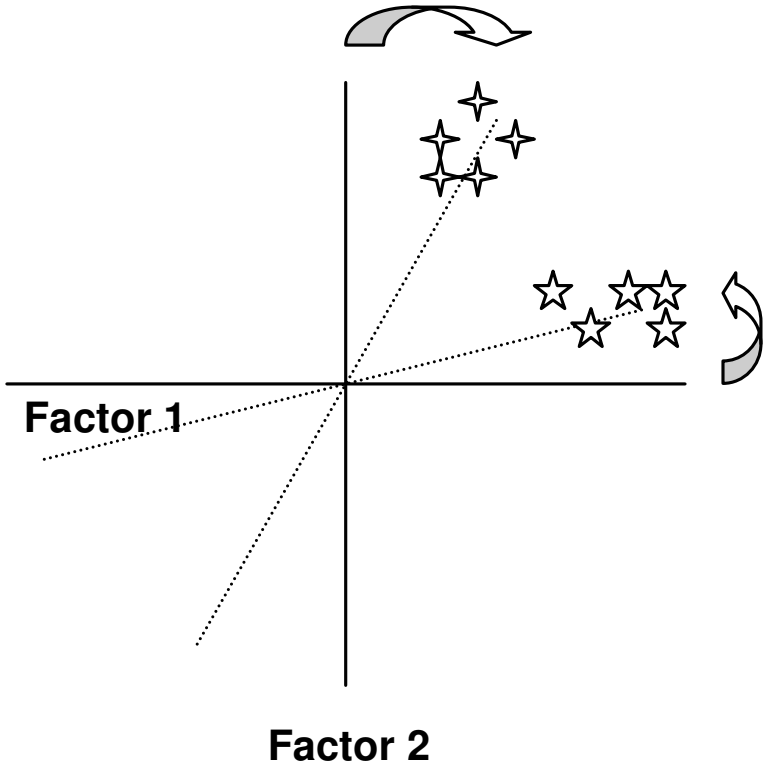
How many factors to retain?

- screeplot: 2
- Guttman-Kaiser rule: factors with an eigenvalue larger than 1: 2
- depends on proportion of variance explained
  - maximum-likelihood factor analysis: 2

Rotation (Field, 2000, p. 439)



Orthogonal rotation (unrelated factors)



Oblique rotation (related factors)

Oblique rotation

```
> m21 = factanal(na.omit(passion), factors=2, rotation='oblimin')
> print(m21)


With factor correlations of
      MR1   MR2
MR1  1.00 -0.64
MR2 -0.64  1.00
```

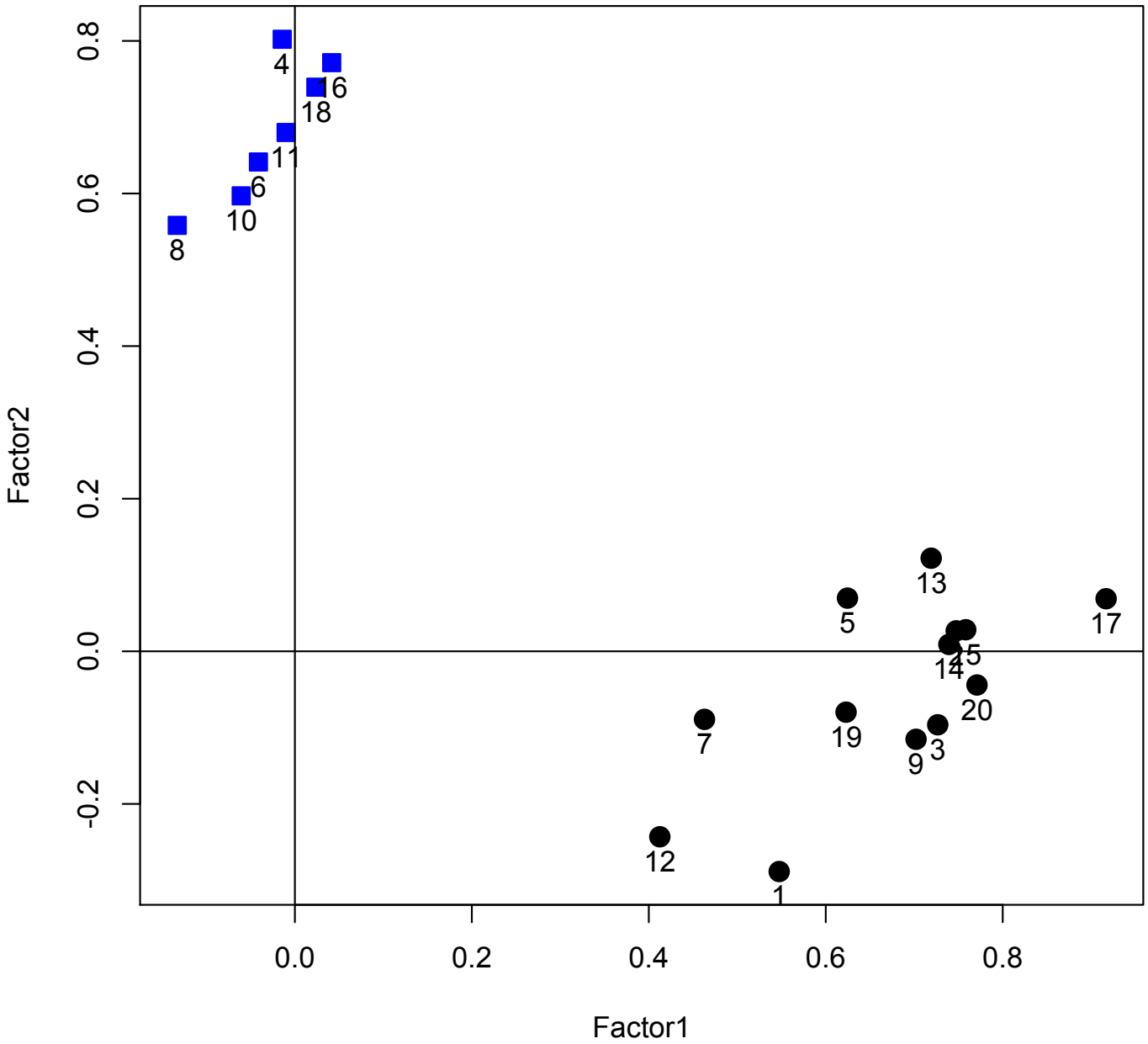
Factor analysis

```
Loadings:
        Factor1 Factor2                 Factor1 Factor2
pass01v  0.547  -0.289          pass11v          0.680
pass02v  0.747                 pass12v  0.412  -0.243
pass03v  0.727                 pass13v  0.719   0.122
pass04v          0.802         pass14v  0.739
pass05v  0.625                 pass15v  0.758
pass06v          0.641         pass16v          0.771
pass07v  0.463                 pass17v  0.917
pass08v -0.133   0.558         pass18v          0.739
pass09v  0.702  -0.115         pass19v  0.623
pass10v          0.597         pass20v  0.771
```

Factor 1 : like writing

Factor 2 : dislike writing

```
> factor.plot(m21,cluster = NULL, cut = 0, labels=NULL,title="Factor Analysis",
jiggle=F)
```



Conclusions

- difference in questionnaire = difference in factors
- total explained variance moderate: 48,4%

So:

I measure what I want to measure and many useless information.

CRONBACH'S ALPHA

Used to

- measure test-retest reliability
  - exactly the same person gets the same questionnaire score

Split-half reliability

4	3
4	4
4	4
4	4
3	4
4	3
3	3
4	4
4	3
3	4

Cronbach's alpha

$$\frac{N^2\overline{Cov}}{\Sigma s^2_{item} + \Sigma Cov_{item}}$$

$$Cov_{item} = (x_i - \overline{x})(y_i - \overline{y})$$

Cronbach's alpha

$$\frac{N^2\overline{Cov}}{\Sigma s^2_{item} + \Sigma Cov_{item}}$$

α=.7 to .8 > reliable measure

Preliminary steps

- reverse scores of reverse-phrased items

Cronbach's alpha

$$\frac{N^2\overline{Cov}}{\Sigma s^2_{item} + \Sigma Cov_{item}}$$

Preliminary steps

- reverse scores of reverse-phrased items
- factor analysis

Cronbach's alpha (passion)

```
> alpha(passion)

  raw_alpha average_r mean   sd
      0.93     0.42   3.3   0.36

Reliability if an item is dropped:
         raw_alpha average_r                raw_alpha    average_r
pass01v       0.93      0.41      pass11v-      0.93       0.42
pass02v       0.93      0.42      pass12v       0.93       0.42
pass03v       0.93      0.41      pass13v       0.93       0.42
pass04v-      0.93      0.42      pass14v       0.93       0.42
pass05v       0.93      0.43      pass15v       0.93       0.42
pass06v-      0.93      0.42      pass16v-      0.93       0.42
pass07v       0.93      0.43      pass17v       0.93       0.41
pass08v-      0.93      0.42      pass18v-      0.93       0.42
pass09v       0.93      0.41      pass19v       0.93       0.42
pass10v-      0.93      0.43      pass20v       0.93       0.41
```

Cronbach's alpha (dislike)

```
> alpha(passion[-c(1,2,3,5,7,9,12,13,14,15,17,19,20)])

  raw_alpha average_r mean   sd
      0.86      0.47   3    0.73

 Reliability if an item is dropped:
         raw_alpha average_r
pass04v       0.83      0.45
pass06v       0.85      0.48
pass08v       0.86      0.50
pass10v       0.85      0.48
pass11v       0.85      0.48
pass16v       0.84      0.46
pass18v       0.84      0.46
```

Cronbach's alpha (like)

```
> alpha(passion[-c(4,6,8,10,11,16,18)])

  raw_alpha average_r mean   sd
      0.93      0.49  3.6    0.72

 Reliability if an item is dropped:
        raw_alpha average_r
pass01v      0.92      0.49
pass02v      0.92      0.49
pass03v      0.92      0.48
pass05v      0.92      0.51
pass07v      0.93      0.51
pass09v      0.92      0.48
pass12v      0.92      0.51
pass13v      0.92      0.50
pass14v      0.92      0.49
pass15v      0.92      0.49
pass17v      0.92      0.47
pass19v      0.92      0.49
pass20v      0.92      0.48
```

Caution

- cannot claim unidimensionality because of high alpha
  - factors correlate highly > high α

- reliability depends on number of items

$$\frac{N^2\overline{Cov}}{\Sigma s^2_{item} + \Sigma Cov_{item}}$$

Questionnaire evaluation

- reliable:
  - dislike: α=.86
  - like: α=.93

- valid:

  I measure what I want to measure, but also many useless information.

Questionnaire evaluation

Useful

But just for this population: seventh graders Dutch secondary education

Sources

- Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Costello, A.B., & Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research and Evaluation, 10,* 1-9.
- Field, A. (2009) *Discovering Statistics using SPSS for Windows.* Sage: London.
- Field, A. (2009) *Discovering Statistics using SPSS.* Sage: London.
- O'Keefe, D.J. (2002). *Persuasion. Theory & Research.* Sage: Thousand Oaks.
- Rattray, J.C., & Jones, M.C. (2007), Essential elements of questionnaire design and development. *Journal of Clinical Nursing, 16,* 234-243.
- Tabachnik, B.G., & Fidell, L.S. (2001) *Using Multivariate Statistics* (4th ed.). Pearson: Needham Heights, MA.

APPENDIX

Correlation matrix

	pass01v	pass02v	pass03v	pass04v	pass05v	pass06v
pass01v	1.0000000	0.6624426	0.6102046	-0.5623126	0.4137463	-0.3889058
pass02v	0.6624426	1.0000000	0.6220010	-0.4148258	0.3525541	-0.2959315
pass03v	0.6102046	0.6220010	1.0000000	-0.5162518	0.4744346	-0.3440221
pass04v	-0.5623126	-0.4148258	-0.5162518	1.0000000	-0.2205194	0.5098645
pass05v	0.4137463	0.3525541	0.4744346	-0.2205194	1.0000000	-0.2213123
pass06v	-0.3889058	-0.2959315	-0.3440221	0.5098645	-0.2213123	1.0000000
pass07v	0.3662825	0.2458077	0.4317847	-0.3205762	0.4838719	-0.3388872
pass08v	-0.4581692	-0.3411867	-0.3934714	0.4738564	-0.3359287	0.5555088
pass09v	0.6159883	0.5825688	0.5945829	-0.4663368	0.5447457	-0.3214989
pass10v	-0.3733701	-0.2031242	-0.3347282	0.4608984	-0.2616383	0.4390695
pass11v	-0.5248995	-0.3382241	-0.4047499	0.6089488	-0.2744794	0.4557957
pass12v	0.4161127	0.4157369	0.5139252	-0.4409926	0.2371068	-0.3090945
pass13v	0.4858500	0.4732903	0.4682252	-0.3126358	0.3290753	-0.2775079
pass14v	0.5700604	0.5339748	0.6003164	-0.3114945	0.4553965	-0.3018648
pass15v	0.4704597	0.5000866	0.6000588	-0.3505059	0.4783165	-0.3733202
pass16v	-0.4074288	-0.2880072	-0.3853428	0.6665053	-0.2136700	0.5423090
pass17v	0.5973765	0.6120596	0.6903767	-0.4340764	0.4997126	-0.3975682
pass18v	-0.4726983	-0.2956505	-0.4005931	0.5374775	-0.2333723	0.4491783
pass19v	0.5509944	0.5041410	0.5126616	-0.3588532	0.2494718	-0.4012147
pass20v	0.5158979	0.5890959	0.5696296	-0.4364437	0.4302584	-0.3998005

Correlation matrix

	pass07v	pass08v	pass09v	pass10v	pass11v	pass12v
pass01v	0.3662825	-0.4581692	0.6159883	-0.3733701	-0.5248995	0.4161127
pass02v	0.2458077	-0.3411867	0.5825688	-0.2031242	-0.3382241	0.4157369
pass03v	0.4317847	-0.3934714	0.5945829	-0.3347282	-0.4047499	0.5139252
pass04v	-0.3205762	0.4738564	-0.4663368	0.4608984	0.6089488	-0.4409926
pass05v	0.4838719	-0.3359287	0.5447457	-0.2616383	-0.2744794	0.2371068
pass06v	-0.3388872	0.5555088	-0.3214989	0.4390695	0.4557957	-0.3090945
pass07v	1.0000000	-0.3831228	0.3535537	-0.2482759	-0.2428494	0.2683525
pass08v	-0.3831228	1.0000000	-0.4162266	0.4567412	0.4343610	-0.2557551
pass09v	0.3535537	-0.4162266	1.0000000	-0.4483098	-0.4060731	0.5518302
pass10v	-0.2482759	0.4567412	-0.4483098	1.0000000	0.4044139	-0.2629465
pass11v	-0.2428494	0.4343610	-0.4060731	0.4044139	1.0000000	-0.3808262
pass12v	0.2683525	-0.2557551	0.5518302	-0.2629465	-0.3808262	1.0000000
pass13v	0.2227190	-0.2196179	0.4614516	-0.2216292	-0.2495326	0.3788450
pass14v	0.3427935	-0.4102525	0.6127256	-0.3936322	-0.3463079	0.4865356
pass15v	0.4730720	-0.3777288	0.5421594	-0.3254854	-0.2435033	0.4041681
pass16v	-0.3098536	0.4552281	-0.3380273	0.4211897	0.4663716	-0.3793710
pass17v	0.5208122	-0.3914103	0.6368713	-0.3812316	-0.3335427	0.4430510
pass18v	-0.1968540	0.4656310	-0.4469567	0.6281839	0.4525809	-0.4152924
pass19v	0.2878531	-0.3888242	0.5373451	-0.2934744	-0.2524475	0.4500402
pass20v	0.4126059	-0.3862857	0.6243242	-0.4302263	-0.3925381	0.4101410

Correlation matrix

	pass13v	pass14v	pass15v	pass16v	pass17v	pass18v
pass01v	0.4858500	0.5700604	0.4704597	-0.4074288	0.5973765	-0.4726983
pass02v	0.4732903	0.5339748	0.5000866	-0.2880072	0.6120596	-0.2956505
pass03v	0.4682252	0.6003164	0.6000588	-0.3853428	0.6903767	-0.4005931
pass04v	-0.3126358	-0.3114945	-0.3505059	0.6665053	-0.4340764	0.5374775
pass05v	0.3290753	0.4553965	0.4783165	-0.2136700	0.4997126	-0.2333723
pass06v	-0.2775079	-0.3018648	-0.3733202	0.5423090	-0.3975682	0.4491783
pass07v	0.2227190	0.3427935	0.4730720	-0.3098536	0.5208122	-0.1968540
pass08v	-0.2196179	-0.4102525	-0.3777288	0.4552281	-0.3914103	0.4656310
pass09v	0.4614516	0.6127256	0.5421594	-0.3380273	0.6368713	-0.4469567
pass10v	-0.2216292	-0.3936322	-0.3254854	0.4211897	-0.3812316	0.6281839
pass11v	-0.2495326	-0.3463079	-0.2435033	0.4663716	-0.3335427	0.4525809
pass12v	0.3788450	0.4865356	0.4041681	-0.3793710	0.4430510	-0.4152924
pass13v	1.0000000	0.4967773	0.4450363	-0.2271384	0.5402914	-0.1845573
pass14v	0.4967773	1.0000000	0.5087498	-0.2822052	0.6123403	-0.3751856
pass15v	0.4450363	0.5087498	1.0000000	-0.4368582	0.6961890	-0.2817147
pass16v	-0.2271384	-0.2822052	-0.4368582	1.0000000	-0.3903826	0.5299966
pass17v	0.5402914	0.6123403	0.6961890	-0.3903826	1.0000000	-0.3461471
pass18v	-0.1845573	-0.3751856	-0.2817147	0.5299966	-0.3461471	1.0000000
pass19v	0.5058519	0.4511643	0.4848206	-0.3752300	0.5932185	-0.3517307
pass20v	0.5852295	0.5629514	0.5891352	-0.3954579	0.7465117	-0.3949938

Correlation matrix

	pass19v	pass20v
pass01v	0.5509944	0.5158979
pass02v	0.5041410	0.5890959
pass03v	0.5126616	0.5696296
pass04v	-0.3588532	-0.4364437
pass05v	0.2494718	0.4302584
pass06v	-0.4012147	-0.3998005
pass07v	0.2878531	0.4126059
pass08v	-0.3888242	-0.3862857
pass09v	0.5373451	0.6243242
pass10v	-0.2934744	-0.4302263
pass11v	-0.2524475	-0.3925381
pass12v	0.4500402	0.4101410
pass13v	0.5058519	0.5852295
pass14v	0.4511643	0.5629514
pass15v	0.4848206	0.5891352
pass16v	-0.3752300	-0.3954579
pass17v	0.5932185	0.7465117
pass18v	-0.3517307	-0.3949938
pass19v	1.0000000	0.5631731
pass20v	0.5631731	1.0000000