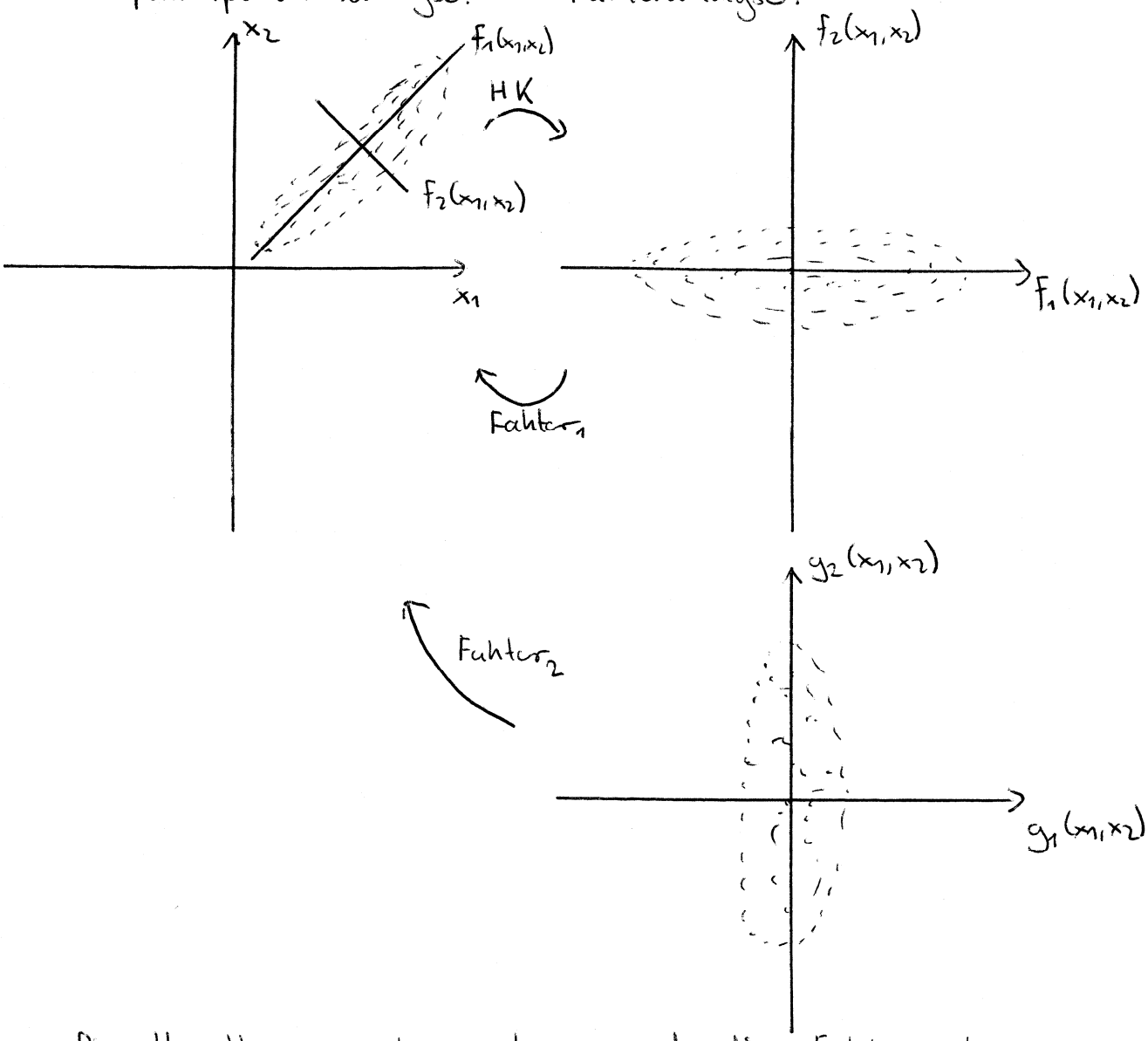
5. Übungsblatt

Hauptkomponentenanalyse vs. Faktoranalyse:



Die Hauptkomponentenanalyse und die Faktoranalyse beschreiben beides Methoden zur Dimensionsreduktion von Daten. ~~Bei der HK-Analyse bestimmen wir die HK~~ Die HK-Analyse beschreibt einen eindeutigen Weg zur Dimensionsreduktion von Daten. Dies ist nicht der Fall bei der Faktoranalyse, ~~da wir keine restr. Annahmen an die Faktoren~~ treffen - Wir ~~machen~~ keine restriktiven Annahmen an die Faktoren, so dass die Faktoren nicht eindeutig bestimmt werden können. Um eindeutige Faktoren bestimmen zu können, müssen wir zusätzliche Annahmen an die Faktoren treffen.

Dies kann zum Beispiel sein, dass jede ursprünglichen Variablen auf möglichst wenige Faktoren gewichtet (Varimax) bzw. das eine ursprüngliche Variable mit möglichst wenigen Faktoren beschrieben wird (Orthomax)

```
. loadingplot, title("Hauptkomponentenanalyse")

. graph save loading1,replace
(file loading1.gph saved)

.
. *1.3 Faktoranalyse mit Methode der Hauptkomponenten
. factor VALUEOFHOME LOWERSTATUS, pcf mineigen(0.0005)
(obs=506)

Factor analysis/correlation                       Number of obs    =        506
    Method: principal-component factors           Retained factors =          2
    Rotation: (unrotated)                         Number of params =          1

    Beware: solution is a Heywood case
            (i.e., invalid or boundary values of uniqueness)

    ------------------------------------------------------------------------------
        Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    ------------+-----------------------------------------------------------------
        Factor1 |      1.73766      1.47533            0.8688       0.8688
        Factor2 |      0.26234            .            0.1312       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(1)  =  396.33 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

    -------------------------------------------------
        Variable |  Factor1   Factor2 |   Uniqueness
    -------------+--------------------+--------------
     VALUEOFHOME |  -0.9321    0.3622 |     -0.0000
     LOWERSTATUS |   0.9321    0.3622 |     -0.0000
    -------------------------------------------------

.
. loadingplot, title("Faktoranalyse (unrotiert)")

. graph save loading2,replace
(file loading2.gph saved)

.
. predict f1 f2
(regression scoring assumed)

Scoring coefficients (method = regression)

    ----------------------------------
        Variable |  Factor1   Factor2
    -------------+--------------------
     VALUEOFHOME | -0.53642   1.38056
     LOWERSTATUS |  0.53642   1.38056
    ----------------------------------


. scatter f2 f1, title("Faktoranalyse (unrotiert)")

. graph save score2, replace
(file score2.gph saved)

.
. *1.4 Faktoranalyse mit Varimax-Rotation
. rotate, varimax

Factor analysis/correlation                       Number of obs    =        506
    Method: principal-component factors           Retained factors =          2
    Rotation: orthogonal varimax (Kaiser off)     Number of params =          1

    Beware: solution is a Heywood case
```

(i.e., invalid or boundary values of uniqueness)

Factor	Variance	Difference	Proportion	Cumulative
Factor1	1.00000	0.00000	0.5000	0.5000
Factor2	1.00000	.	0.5000	1.0000

LR test: independent vs. saturated: chi2(1) = 396.33 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
VALUEOFHOME	0.9152	-0.4030	-0.0000
LOWERSTATUS	-0.4030	0.9152	-0.0000

Factor rotation matrix

	Factor1	Factor2
Factor1	-0.7071	
Factor2	0.7071	0.7071

```
. loadingplot, title("Faktoranalyse (Varimax-Rotation)")

. graph save loadingv,replace
(file loadingv.gph saved)

. predict f1v f2v
(regression scoring assumed)
```

Scoring coefficients (method = regression; based on varimax rotated factors)

Variable	Factor1	Factor2
VALUEOFHOME	1.35551	0.59690
LOWERSTATUS	0.59690	1.35551

```
. scatter f2v f1v, title("Faktoranalyse (Varimax-Rotation)")

. graph save scorev, replace
(file scorev.gph saved)

. *1.5 Faktoranalyse mit Quartimax-Rotation
. rotate, quartimax
```

Factor analysis/correlation — Number of obs = 506
Method: principal-component factors — Retained factors = 2
Rotation: orthogonal quartimax (Kaiser off) — Number of params = 1

Beware: solution is a Heywood case
(i.e., invalid or boundary values of uniqueness)

Factor	Variance	Difference	Proportion	Cumulative
Factor1	1.73766	1.47533	0.8688	0.8688

```
    Factor2  |      0.26234            .            0.1312       1.0000
-------------------------------------------------------------------------
LR test: independent vs. saturated:  chi2(1)  =  396.33 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances

-------------------------------------------------
    Variable |   Factor1    Factor2 |   Uniqueness
-------------+--------------------+--------------
 VALUEOFHOME |   -0.9321     0.3622 |      -0.0000
 LOWERSTATUS |    0.9321     0.3622 |      -0.0000
-------------------------------------------------

Factor rotation matrix

--------------------------------
             | Factor1  Factor2
-------------+------------------
     Factor1 |  1.0000
     Factor2 |  0.0000   1.0000
--------------------------------

.
. loadingplot, title("Faktoranalyse (Quartimax-Rotation)")

. graph save loadingq,replace
(file loadingq.gph saved)

.
. predict f1q f2q
(regression scoring assumed)

Scoring coefficients (method = regression; based on quartimax rotated factors)

----------------------------------
    Variable |   Factor1   Factor2
-------------+--------------------
 VALUEOFHOME |  -0.53642   1.38056
 LOWERSTATUS |   0.53642   1.38056
----------------------------------


. scatter f2q f1q, title("Faktoranalyse (Quartimax-Rotation)")

. graph save scoreq, replace
(file scoreq.gph saved)

.
. graph combine score1.gph score2.gph scorev.gph scoreq.gph

. graph export score.png, replace
(file score.png written in PNG format)

.
. graph combine loading1.gph loading2.gph loadingv.gph loadingq.gph

. graph export loading.png, replace
(file loading.png written in PNG format)

.
. /********************/
. /* 2. Faktoranalyse */
. /********************/
.
. * 2.1 mit Hilfe der Methode der Hauptkomponenten
. factor VALUEOFHOME - LOWERSTATUS, pcf
(obs=506)
```

Vorgeschlagene Anzahl an Faktoren

Factor analysis/correlation
Method: principal-component factors
Rotation: (unrotated)

Number of obs = 506
Retained factors = 3
Number of params = 39

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	6.54599	4.89646	0.4676	0.4676
Factor2	1.64953	0.30063	0.1178	0.5854
Factor3	1.34891	0.46237	0.0964	0.6817
Factor4	0.88654	0.03564	0.0633	0.7451
Factor5	0.85090	0.19089	0.0608	0.8058
Factor6	0.66001	0.12460	0.0471	0.8530
Factor7	0.53541	0.13233	0.0382	0.8912
Factor8	0.40308	0.12581	0.0288	0.9200
Factor9	0.27726	0.02501	0.0198	0.9398
Factor10	0.25226	0.03947	0.0180	0.9578
Factor11	0.21279	0.02980	0.0152	0.9730
Factor12	0.18299	0.04898	0.0131	0.9861
Factor13	0.13401	0.07368	0.0096	0.9957
Factor14	0.06033	.	0.0043	1.0000

$\sum_{R=4}^{14} \lambda_R^2$ klein genug

LR test: independent vs. saturated: chi2(91) = 5143.04 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Uniqueness
VALUEOFHOME	-0.6822	0.5714	0.1895	0.1722
CRIMERATE	0.6199	-0.0846	0.4589	0.3980
RESIDENTIA~E	-0.6279	-0.1901	0.4582	0.3596
NONRETAIL	0.8491	0.1632	-0.0767	0.2466
RIVER	-0.0129	0.5274	-0.1455	0.7005
NOCONCENTR~N	0.8320	0.3266	-0.0540	0.1982
NOOFROOMS	-0.5189	0.5574	0.4105	0.2516
AGE	0.7598	0.3343	-0.2332	0.2565
DISTANCE	-0.7629	-0.4613	0.1824	0.1720
ACCESSTOHI~Y	0.7763	0.0400	0.4861	0.1595
TAXRATE	0.8290	0.0114	0.3986	0.1536
PUPILTEACHER	0.5314	-0.4041	0.0005	0.5544
BLACKPEOPLE	-0.5031	0.0340	-0.4197	0.5696
LOWERSTATUS	0.7967	-0.2585	-0.1871	0.2634

```
.
. /*  Stata sagt:  3 Faktoren sind ausreichend (Kriterim: Die Summe von
> quadrierten vernachlässgten Eigenwerten ist klein genug, Skript, S.45)
>
> Wir werden 3 behalten.
>
> Interpretation:  Die Größe der geschätzten Faktorldung zeigt die wichtigkeit
> (siehe Skript, S.40)
>
> Was bei Stata in der "Uniqueness"-Spalte steht ist die Varianz des spezifichen
Faktors
> (d.h. der Hauptdiagonalelement der grosen Psi-Matrix)  */
.
. predict F1_pcf F2_pcf F3_pcf, norotated
(regression scoring assumed)
```

Scoring coefficients (method = regression)

Variable	Factor1	Factor2	Factor3
VALUEOFHOME	-0.10422	0.34642	0.14051
CRIMERATE	0.09470	-0.05129	0.34017

```
                                    Boston.log
    RESIDENTIA~E |  -0.09593   -0.11524    0.33971
       NONRETAIL |   0.12971    0.09894   -0.05690
           RIVER |  -0.00196    0.31975   -0.10789
    NOCONCENTR~N |   0.12710    0.19798   -0.04002
       NOOFROOMS |  -0.07927    0.33792    0.30429
             AGE |   0.11607    0.20267   -0.17291
        DISTANCE |  -0.11654   -0.27964    0.13524
    ACCESSTOHI~Y |   0.11859    0.02425    0.36034
         TAXRATE |   0.12665    0.00689    0.29553
    PUPILTEACHER |   0.08117   -0.24497    0.00034
     BLACKPEOPLE |  -0.07686    0.02062   -0.31115
     LOWERSTATUS |   0.12171   -0.15669   -0.13867
    ----------------------------------------------


. loadingplot, title("Methode der Hauptkomponenten")

. graph save loading_hk, replace
(file loading_hk.gph saved)

.
.
. * 2.2 mit Hilfe der Methode der Hauptfaktoren
. factor VALUEOFHOME - LOWERSTATUS, ipf
(obs=506)

Factor analysis/correlation                       Number of obs    =        506
    Method: iterated principal factors            Retained factors =         13
    Rotation: (unrotated)                         Number of params =         91

    Beware: solution is a Heywood case
            (i.e., invalid or boundary values of uniqueness)

    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      6.38752      4.97003            0.6097       0.6097
        Factor2  |      1.41750      0.27154            0.1353       0.7451
        Factor3  |      1.14596      0.58379            0.1094       0.8544
        Factor4  |      0.56217      0.26183            0.0537       0.9081
        Factor5  |      0.30033      0.10810            0.0287       0.9368
        Factor6  |      0.19224      0.02699            0.0184       0.9551
        Factor7  |      0.16525      0.04485            0.0158       0.9709
        Factor8  |      0.12040      0.02204            0.0115       0.9824
        Factor9  |      0.09836      0.04561            0.0094       0.9918
       Factor10  |      0.05275      0.02673            0.0050       0.9968
       Factor11  |      0.02602      0.01868            0.0025       0.9993
       Factor12  |      0.00734      0.00706            0.0007       1.0000
       Factor13  |      0.00028      0.00059            0.0000       1.0000
       Factor14  |     -0.00031            .           -0.0000       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(91) = 5143.04 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

-------------------------------------------------------------------------------
-------------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3   Factor4   Factor5   Factor6
Factor7   Factor8   Factor9  Factor10  Factor11  Factor12  Factor13

-------------+-----------------------------------------------------------------
-------------------------------------------------------------
     VALUEOFHOME |  -0.6801    0.6208    0.1459   -0.0837    0.0608    0.1197
-0.0873    0.0917    0.0306   -0.0582    0.0275    0.0422   -0.0041
       CRIMERATE |   0.5832   -0.0494    0.3213    0.0972   -0.2083    0.0368
-0.1314    0.0604   -0.0281    0.1389   -0.0158    0.0214    0.0016
    RESIDENTIA~E |  -0.6081   -0.1467    0.4009    0.3579    0.0924    0.0751
```

```
0.0990     0.0809   -0.0516    0.0547    0.0506   -0.0144   -0.0049
       NONRETAIL |    0.8401    0.1422   -0.0764   -0.0102    0.2415   -0.1217
0.0704     0.1550    0.0857    0.0463    0.0316    0.0154    0.0069
          RIVER |   -0.0104    0.2645   -0.0772    0.0408    0.1267    0.1070
-0.0279   -0.0352    0.1978    0.0756   -0.0479   -0.0322   -0.0043
    NOCONCENTR~N |    0.8338    0.3189   -0.1010    0.2246    0.1167   -0.1006
0.0251    -0.1840   -0.0486    0.0201    0.0578    0.0109   -0.0043
       NOOFROOMS |   -0.5024    0.5498    0.2912   -0.0714   -0.2046   -0.0611
0.1886     0.0012    0.0204    0.0014    0.0247   -0.0276    0.0048
          AGE |    0.7485    0.2968   -0.2659    0.0452   -0.1358    0.1766
0.2049    -0.0232   -0.0320    0.0234   -0.0421    0.0253   -0.0009
       DISTANCE |   -0.7636   -0.4649    0.2473    0.0694    0.0808    0.0217
0.1470    -0.1082    0.1016    0.0003   -0.0220    0.0412    0.0030
    ACCESSTOHI~Y |    0.7874    0.0754    0.5655   -0.1031    0.0193    0.1250
-0.0951   -0.1360    0.0473   -0.0333    0.0315   -0.0056    0.0048
        TAXRATE |    0.8381    0.0372    0.4617   -0.0310    0.1862   -0.0029
0.0718     0.0617   -0.0931   -0.0622   -0.0818   -0.0097   -0.0023
    PUPILTEACHER |    0.5081   -0.3394    0.0832   -0.5033   -0.0316    0.0240
0.1023     0.0227    0.0267    0.0377    0.0539    0.0004   -0.0074
     BLACKPEOPLE |   -0.4627    0.0134   -0.2242   -0.1649    0.2104    0.2247
0.0020    -0.0299   -0.1297    0.0578    0.0185   -0.0116    0.0061
     LOWERSTATUS |    0.7897   -0.2958   -0.1785    0.2477   -0.0977    0.1802
0.0182     0.0817    0.0701   -0.0923    0.0449   -0.0118    0.0014

------------------------------------------------------------------------------
-------------------------------------------------------------

    ----------------------------
        Variable |   Uniqueness
    -------------+--------------
     VALUEOFHOME |      0.0829
       CRIMERATE |      0.4583
    RESIDENTIA~E |      0.2810
       NONRETAIL |      0.1552
           RIVER |      0.8446
    NOCONCENTR~N |      0.0779
       NOOFROOMS |      0.2724
             AGE |      0.1828
        DISTANCE |      0.0821
    ACCESSTOHI~Y |     -0.0040
         TAXRATE |      0.0190
    PUPILTEACHER |      0.3487
     BLACKPEOPLE |      0.5920
     LOWERSTATUS |      0.1311
    ----------------------------

.
. /*  Wir verwenden die iterative Methode der Hauptfaktoren.  Anzahl an Faktoren
ist
> vorgegeben.  Wir nehmen 3 aus der vorigen Analyse.  Da k=3, in der
Output-Tabelle die
> Eigenwerte ab 4. inklusive mahen kein Sinn.  */
.
.
. predict F1_ipf F2_ipf F3_ipf, norotated
(regression scoring assumed)

Scoring coefficients (method = regression)


------------------------------------------------------------------------------
-------------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3   Factor4   Factor5   Factor6
Factor7   Factor8   Factor9  Factor10  Factor11  Factor12  Factor13

-------------+----------------------------------------------------------------
-------------------------------------------------------------
     VALUEOFHOME | -0.18742   0.59236   0.08452   0.00660   0.38221   0.50857
```

```
-0.14414   0.40476   0.18201  -0.35371   0.08661   0.26361  -0.03108
       CRIMERATE |  0.01591  -0.01539   0.07078   0.13220  -0.22115  -0.05863
-0.11075   0.18628  -0.07390   0.29064  -0.06947   0.08681  -0.00314
    RESIDENTIA~E | -0.06381  -0.02152   0.19980   0.41675   0.02642   0.14730
 0.06714   0.16225  -0.13278   0.24895   0.25630  -0.08413  -0.01209
       NONRETAIL |  0.10499   0.06047  -0.08598  -0.11101   0.42035  -0.21595
 0.07754   0.36456   0.70090   0.31199   0.30735   0.09625   0.05176
           RIVER | -0.00030   0.03393  -0.02562   0.01943   0.09240   0.06744
 0.00093  -0.00841   0.15817   0.08055  -0.09885  -0.05864  -0.00684
    NOCONCENTR~N |  0.11145   0.32714  -0.20239   0.61255   0.40015  -0.43250
 0.14963  -0.89445  -0.14755   0.08422   0.32204   0.12049  -0.03655
       NOOFROOMS | -0.04083   0.21599   0.16719  -0.00254  -0.41347  -0.23474
 0.47780   0.03043   0.07367   0.00060   0.07653  -0.13010   0.01947
             AGE |  0.08576   0.16965  -0.15563  -0.03543  -0.27469   0.53917
 0.61689  -0.11937  -0.02186   0.14092  -0.23586   0.13437   0.00210
        DISTANCE | -0.18499  -0.34117   0.25758   0.22827   0.41297   0.11141
 0.69368  -0.51838   0.50590  -0.04231  -0.07151   0.31778   0.01182
    ACCESSTOHI~Y |  0.20799   0.03721   0.74196  -0.38740  -0.52608   0.77555
-0.96676  -1.21650   0.94316   0.18526   0.72018  -0.02947   0.07459
         TAXRATE |  0.13698  -0.03735   0.39006   0.08722   0.86396  -0.40836
 0.89487   1.12786  -1.19538  -0.59304  -1.09539  -0.05364  -0.06699
    PUPILTEACHER |  0.03245  -0.10056  -0.07063  -0.43999  -0.00661  -0.00325
 0.22826   0.05616  -0.01251   0.08536   0.20221   0.03761  -0.03245
     BLACKPEOPLE | -0.01558  -0.01214  -0.06921  -0.11303   0.22095   0.27747
-0.00051  -0.09152  -0.17508   0.09070   0.04816  -0.04242   0.01430
     LOWERSTATUS |  0.12591  -0.19252  -0.14511   0.46889  -0.20203   0.56494
 0.02137   0.42968   0.29554  -0.48885   0.20455  -0.01293  -0.00115

-------------------------------------------------------------------------------
-------------------------------------------------------------

.
. loadingplot, title("Methode der Hauptfaktoren")

. graph save loading_hf, replace
(file loading_hf.gph saved)

.
. corr F1_pcf F1_ipf
(obs=506)

             |   F1_pcf   F1_ipf
-------------+------------------
      F1_pcf |   1.0000
      F1_ipf |   0.9954   1.0000

. corr F2_pcf F2_ipf
(obs=506)

             |   F2_pcf   F2_ipf
-------------+------------------
      F2_pcf |   1.0000
      F2_ipf |   0.9458   1.0000

. corr F3_pcf F3_ipf
(obs=506)

             |   F3_pcf   F3_ipf
-------------+------------------
      F3_pcf |   1.0000
      F3_ipf |   0.9167   1.0000

.
. /*  Ergebnis:  Hochkorreliert!  => So gut wie egal welche Schätzstrategie zu
```

Hohe Korrelation

```
> benutzen.  Hauptfaktoren sind u.U. genauer wegen des iterativen Schätzung */
.
. twoway (scatter F1_pcf F1_ipf) (scatter F2_pcf F2_ipf) (scatter F3_pcf F3_ipf)

. graph export sc_pcf_ipf.png, replace
(file sc_pcf_ipf.png written in PNG format)

.
. * 2.3 mit Hilfe einer Maximum-Likelihood-Schätzung
. factor VALUEOFHOME - LOWERSTATUS, ml factor(3)
(obs=506)
Iteration 0:   log likelihood =  -275.8847
Iteration 1:   log likelihood = -220.64216
Iteration 2:   log likelihood = -219.51298
Iteration 3:   log likelihood = -219.38435
Iteration 4:   log likelihood = -219.35723
Iteration 5:   log likelihood = -219.34915
Iteration 6:   log likelihood = -219.34629
Iteration 7:   log likelihood = -219.34521
Iteration 8:   log likelihood = -219.34479
Iteration 9:   log likelihood = -219.34463
Iteration 10:  log likelihood = -219.34457
Iteration 11:  log likelihood = -219.34454
Iteration 12:  log likelihood = -219.34453
Iteration 13:  log likelihood = -219.34453
Iteration 14:  log likelihood = -219.34453
Iteration 15:  log likelihood = -219.34453
Iteration 16:  log likelihood = -219.34453

Factor analysis/correlation                       Number of obs    =        506
    Method: maximum likelihood                    Retained factors =          3
    Rotation: (unrotated)                         Number of params =         39
                                                  Schwarz's BIC    =    681.524
    Log likelihood = -219.3445                    (Akaike's) AIC   =    516.689

    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |        6.10719      4.85260            0.7096       0.7096
        Factor2  |        1.25460      0.00995            0.1458       0.8554
        Factor3  |        1.24465            .            0.1446       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(91) = 5143.04 Prob>chi2 = 0.0000
    LR test:  (3) factors vs. saturated:  chi2(52) =  432.19 Prob>chi2 = 0.0000
              H0: k=3
Factor loadings (pattern matrix) and unique variances

    -----------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3 |  Uniqueness
    -------------+------------------------------+--------------
     VALUEOFHOME |  -0.6595    0.6927   -0.0659 |       0.0809
       CRIMERATE |   0.6148    0.0481    0.1445 |       0.5988
    RESIDENTIA~E |  -0.5123    0.0692    0.4921 |       0.4905
       NONRETAIL |   0.8136    0.0461   -0.2451 |       0.2758
           RIVER |  -0.0301    0.1958   -0.1750 |       0.9301
    NOCONCENTR~N |   0.7943    0.1175   -0.3495 |       0.2332
       NOOFROOMS |  -0.4603    0.5795    0.0379 |       0.4509
             AGE |   0.6744    0.0512   -0.4821 |       0.3102
        DISTANCE |  -0.6878   -0.2353    0.5796 |       0.1357
    ACCESSTOHI~Y |   0.8542    0.2856    0.2824 |       0.1091
         TAXRATE |   0.9131    0.2165    0.2446 |       0.0595
    PUPILTEACHER |   0.5148   -0.2105    0.1655 |       0.6633
     BLACKPEOPLE |  -0.4790   -0.0012   -0.0736 |       0.7651
     LOWERSTATUS |   0.7352   -0.3833   -0.1491 |       0.2903
    -----------------------------------------------------------

.
. /*  Wir fangen bei 3 Faktoren an, denn - laut dem Ergebnis bei Meth. von HK -
```

```
es ist klar, daß
> weniger als Faktoren zu unvollständige Darstellung bieten
>
> Wir sehen aber, daß der LRT (Skript, S.48) lehtn die k=3 ab:  chi2(42) =
418.68 Prob>chi2 = 0.0000.
>
> Wir müssen also noch ein Faktor hinzufügen:  */
.
. factor VALUEOFHOME - LOWERSTATUS, ml factor(4)
(obs=506)
Iteration 0:   log likelihood = -195.74265
Iteration 1:   log likelihood = -130.69468
Iteration 2:   log likelihood = -129.60009
Iteration 3:   log likelihood = -129.51423
Iteration 4:   log likelihood = -129.50662
Iteration 5:   log likelihood = -129.50593
Iteration 6:   log likelihood = -129.50586
Iteration 7:   log likelihood = -129.50585
Iteration 8:   log likelihood = -129.50585

Factor analysis/correlation                       Number of obs    =        506
    Method: maximum likelihood                    Retained factors =          4
    Rotation: (unrotated)                         Number of params =         50
                                                  Schwarz's BIC    =    570.339
    Log likelihood = -129.5059                    (Akaike's) AIC   =    359.012

    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      6.15344      4.86534            0.6673       0.6673
        Factor2  |      1.28810      0.05297            0.1397       0.8070
        Factor3  |      1.23513      0.69108            0.1340       0.9410
        Factor4  |      0.54405            .            0.0590       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(91) = 5143.04 Prob>chi2 = 0.0000
    LR test:   4 factors vs. saturated:  chi2(41) =  254.83 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

    ------------------------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3   Factor4 |   Uniqueness
    -------------+--------------------------------------+--------------
     VALUEOFHOME |  -0.6419    0.6954   -0.0813   -0.0471 |       0.0956
       CRIMERATE |   0.6123    0.0434    0.1681    0.0572 |       0.5917
    RESIDENTIA~E |  -0.5327    0.0967    0.4854    0.3715 |       0.3333
       NONRETAIL |   0.8180    0.0238   -0.2293    0.0167 |       0.2775
           RIVER |  -0.0209    0.1988   -0.1776    0.0481 |       0.9262
    NOCONCENTR~N |   0.8077    0.1155   -0.3514    0.2084 |       0.1674
       NOOFROOMS |  -0.4495    0.5949    0.0263   -0.0278 |       0.4426
             AGE |   0.6872    0.0359   -0.4605    0.0140 |       0.3143
        DISTANCE |  -0.7067   -0.2168    0.5566    0.0840 |       0.1367
    ACCESSTOHI~Y |   0.8567    0.2720    0.3097   -0.0411 |       0.0944
         TAXRATE |   0.9071    0.1969    0.2616    0.0312 |       0.0690
    PUPILTEACHER |   0.5197   -0.2588    0.2175   -0.5553 |       0.3072
     BLACKPEOPLE |  -0.4764   -0.0020   -0.0829   -0.1319 |       0.7487
     LOWERSTATUS |   0.7303   -0.3974   -0.1264    0.1346 |       0.2746
    ------------------------------------------------------------------------

. /*  4 Faktoren - Wieder Ablehnung:  chi2(32) =  242.80 Prob>chi2 = 0.0000  */
.
. factor VALUEOFHOME - LOWERSTATUS, ml factor(5)
(obs=506)
Iteration 0:   log likelihood = -166.01585
Iteration 1:   log likelihood = -72.640624
Iteration 2:   log likelihood = -69.621227
Iteration 3:   log likelihood = -68.989567
Iteration 4:   log likelihood = -68.843021
Iteration 5:   log likelihood = -68.809437
```

```
Iteration 6:   log likelihood =  -68.80121
Iteration 7:   log likelihood = -68.799008
Iteration 8:   log likelihood = -68.798374
Iteration 9:   log likelihood = -68.798182
Iteration 10:  log likelihood = -68.798122
Iteration 11:  log likelihood = -68.798103
Iteration 12:  log likelihood = -68.798097
Iteration 13:  log likelihood = -68.798096

Factor analysis/correlation                       Number of obs    =        506
    Method: maximum likelihood                    Retained factors =          5
    Rotation: (unrotated)                         Number of params =         60
                                                  Schwarz's BIC    =    511.188
    Log likelihood =  -68.7981                    (Akaike's) AIC   =    257.596

    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      5.81568      4.25275            0.6132       0.6132
        Factor2  |      1.56293      0.22648            0.1648       0.7780
        Factor3  |      1.33645      0.81298            0.1409       0.9189
        Factor4  |      0.52347      0.27780            0.0552       0.9741
        Factor5  |      0.24567            .            0.0259       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(91) = 5143.04 Prob>chi2 = 0.0000
    LR test:   5 factors vs. saturated:  chi2(31) =  135.19 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

-------------------------------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3   Factor4   Factor5 |
Uniqueness
-------------+-----------------------------------------------------+----------
     VALUEOFHOME |  -0.5635    0.5294    0.5379   -0.0300    0.0401 |
0.1103
       CRIMERATE |   0.6319    0.0687   -0.0994    0.0342   -0.2142 |
0.5391
    RESIDENTIA~E |  -0.4743    0.4582   -0.2754    0.4095    0.0243 |
0.3210
       NONRETAIL |   0.7832   -0.3114    0.2015    0.0795    0.2755 |
0.1669
           RIVER |  -0.0129    0.0256    0.2665    0.0406   -0.0078 |
0.9264
    NOCONCENTR~N |   0.7711   -0.2854    0.3321    0.1918   -0.0549 |
0.1739
       NOOFROOMS |  -0.3772    0.5130    0.3997   -0.0224   -0.0247 |
0.4338
             AGE |   0.6342   -0.3919    0.3507   -0.0099   -0.1183 |
0.3071
        DISTANCE |  -0.6638    0.3312   -0.5454    0.0981    0.0539 |
0.1396
    ACCESSTOHI~Y |   0.9264    0.3090   -0.0433   -0.0461   -0.0758 |
0.0366
         TAXRATE |   0.9412    0.1624   -0.0464    0.0833    0.1336 |
0.0609
    PUPILTEACHER |   0.5115   -0.1085   -0.3366   -0.5136    0.1869 |
0.3146
     BLACKPEOPLE |  -0.4788    0.0125    0.0590   -0.1187    0.1228 |
0.7379
     LOWERSTATUS |   0.6678   -0.4783   -0.1961    0.1088   -0.1660 |
0.2476

-------------------------------------------------------------------------------

. /*  5 Faktoren - Immer noch Ablehnung:  chi2(23) =  122.84 Prob>chi2 = 0.0000
*/
```

31

135,19

```
.
. factor VALUEOFHOME - LOWERSTATUS, ml factor(6)
(obs=506)
Iteration 0:   log likelihood = -163.88203
Iteration 1:   log likelihood = -58.911447
Iteration 2:   log likelihood = -46.574122
Iteration 3:   log likelihood = -42.473177
Iteration 4:   log likelihood = -41.466082
Iteration 5:   log likelihood = -39.124983
Iteration 6:   log likelihood = -38.879923
Iteration 7:   log likelihood = -38.857393
Iteration 8:   log likelihood = -38.855352
Iteration 9:   log likelihood =  -38.85516
Iteration 10:  log likelihood =  -38.85514
Iteration 11:  log likelihood = -38.855137
Iteration 12:  log likelihood = -38.855137

Factor analysis/correlation                       Number of obs    =        506
    Method: maximum likelihood                    Retained factors =          6
    Rotation: (unrotated)                         Number of params =         69
                                                  Schwarz's BIC    =    507.341
    Log likelihood = -38.85514                    (Akaike's) AIC   =     215.71

    Beware: solution is a Heywood case
            (i.e., invalid or boundary values of uniqueness)

    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      4.60512      3.95717            0.4683       0.4683
        Factor2  |      0.64795     -1.74139            0.0659       0.5342
        Factor3  |      2.38934      0.96524            0.2430       0.7772
        Factor4  |      1.42410      0.94529            0.1448       0.9220
        Factor5  |      0.47881      0.19086            0.0487       0.9707
        Factor6  |      0.28795            .            0.0293       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(91) = 5143.04 Prob>chi2 = 0.0000
    LR test:   6 factors vs. saturated:  chi2(22) =   76.25 Prob>chi2 = 0.0000
    (tests formally not valid because a Heywood case was encountered)

Factor loadings (pattern matrix) and unique variances

-------------------------------------------------------------------------------
---------
        Variable |  Factor1   Factor2   Factor3   Factor4   Factor5   Factor6 |
 Uniqueness

-------------+-----------------------------------------------------------+-----
---------
     VALUEOFHOME |  -0.9208    0.3901    0.0000    0.0000   -0.0000    0.0000 |
    0.0000
       CRIMERATE |   0.4553    0.0792    0.4511    0.0981    0.0000   -0.1285 |
    0.5568
    RESIDENTIA~E |  -0.4169   -0.0600   -0.1946    0.5548    0.4490   -0.0025 |
    0.2754
       NONRETAIL |   0.5881    0.1481    0.4432   -0.4507    0.1377    0.2745 |
    0.1382
           RIVER |  -0.1178    0.1711    0.0482   -0.1558    0.0402   -0.0355 |
    0.9274
    NOCONCENTR~N |   0.5527    0.2092    0.4599   -0.4868    0.1939   -0.1624 |
    0.1382
       NOOFROOMS |  -0.6985    0.1337    0.1135    0.0265    0.0322   -0.1474 |
    0.4578
             AGE |   0.5345    0.2952    0.2690   -0.4836   -0.0522   -0.1222 |
    0.3033
        DISTANCE |  -0.4108   -0.3289   -0.3985    0.6209    0.1093    0.0815 |
    0.1603
```

```
    ACCESSTOHI~Y |   0.4711    0.1337    0.8465    0.1371   -0.0367   -0.0355 |
    0.0223
         TAXRATE |   0.5460    0.0878    0.7666    0.0091    0.1225    0.1243 |
    0.0760
    PUPILTEACHER |   0.4672   -0.1988    0.3058    0.0572   -0.4208    0.3095 |
    0.3725
     BLACKPEOPLE |  -0.3764   -0.0337   -0.3098   -0.0079   -0.0998    0.1124 |
    0.7385
     LOWERSTATUS |   0.9426    0.3339   -0.0000    0.0000   -0.0000    0.0000 |
    0.0000

-------------------------------------------------------------------------------
---------

. /*  6 Faktoren:  Stata sagt - "Beware: solution is a Heywood case (i.e.,
invalid or boundary values of uniqueness)"
> Dies bedeutet daß ML Schätzer einer von Parameter liegt am rande des
Parameterraums.  Wenn wir auf "Uniqueness"-Spalte
> schauen, dan das sehen wir, das die Varainz von AGE gleich 0 ist.  Das ist
genau der Randfall!
> Wenn der wahre Parameter auf dem Rand liegt die Asylptotische Verteilung der
LR Test statistik ist nicht mehr Chi^2, sondern was kompliziertes.
> Deswegen auf dieser Stelle sollen wir mit unserer Modellselektion aufhöhren,
da wir keine genaie Testverteilung unter H0 kennen.
>
> Das Ergebnis der MLE und Modellselektion:  Mindestens 5 Faktoren! (eignetlich
mehr).  Ob wir alle 5 oder nur die erste 3 [oder gar
> nur den ersten Faktor] verwenden werden - das ist letztendlich unsere
Entscheidung.  ML, in der Regel, übetreibt
>
> factor VALUEOFHOME - LOWERSTATUS, ml factor(5)
> predict F1_mle F2_mle F3_mle F4_mle F5_mle, norotated
>
> loadingplot, title("Maximum-Likelihood")
> graph save loading_ml, replace
>
> corr F1_ipf - F3_ipf F1_mle - F5_mle
> /* a) F1_ipf und F1_mle;  b) F2_ipf und F3_mle;  c) F3_ipf und F2_mle;  -
Hochkorreliert!  Zusätzliche Faktoren bei MLE bilden zusätzliche
> Gemeinamkeit ab.  Logisch.  */
>
>
> twoway (scatter F1_ipf F1_mle) (scatter F2_ipf F3_mle) (scatter F3_ipf F2_mle)
> graph export sc_ipf_mle.png
>
>
> graph combine loading_hk.gph loading_hf.gph loading_ml.gph
> graph export loading_fk.png,replace
>
> /*  Allgemeines Ergebnis:  3 Faktoren für diesen Datensatz ist ein guter
Kompromiss.
> Alle 3 Methoden sind zugelassen um die zu berechnen. */
>
> log close

end of do-file
```

Wichtig! Wenn wir in der Übung nicht mehr bewusst

⇒ Deswegen nur 5 Faktoren

```
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--
      log:  U:\Übung5\Boston.log
  log type:  text
 opened on:  12 Jul 2011, 08:18:33

. 
. /*******************************************************************************/
. /*                                                                             */
. /*    Ü B U N G  4  -  H A U P T K O M P O N E N T E N A N A L Y S E    */
. /*                                                                             */
. /*******************************************************************************/
. 
. 
. /******************/
. /* 0. Preliminary */
. /******************/
. 
. * Daten einlesen
. use "U:\Übung5\Boston.dta", clear

. 
. 
. /************************************************/
. /* 1. Hauptkomponentenanalyse vs. Faktoranalyse */
. /************************************************/
. 
. *1.1 Scatterplot
. scatter VALUEOFHOME LOWERSTATUS, title("Daten")

. graph save scatter1,replace
(file scatter1.gph saved)

. 
. *1.2 Hauptkomponentenanalyse mit Scoreplot
. pca VALUEOFHOME LOWERSTATUS, cov

Principal components/covariance                 Number of obs    =        506
                                                Number of comp.  =          2
                                                Trace            =  135.5815
    Rotation: (unrotated = principal)           Rho              =     1.0000

    ------------------------------------------------------------------------------
       Component |   Eigenvalue   Difference         Proportion   Cumulative
    -------------+----------------------------------------------------------------
           Comp1 |      119.067      102.553             0.8782       0.8782
           Comp2 |      16.5143            .             0.1218       1.0000
    ------------------------------------------------------------------------------

Principal components (eigenvectors)

    ----------------------------------------------------
        Variable |    Comp1     Comp2 | Unexplained
    -------------+--------------------+-------------
     VALUEOFHOME |   0.8147    0.5798 |           0
     LOWERSTATUS |  -0.5798    0.8147 |           0
    ----------------------------------------------------

. 
. scoreplot, title("Hauptkomponentenanalyse")

. graph save score1, replace
(file score1.gph saved)

. 
```