



UPPSALA  
UNIVERSITET

U.U.D.M. Project Report 2016:20

# Multivariat databehandling och dataanalys

– en faktoranalys av slutbetyg i åk 9 inom Lgr11 på kommunnivå

Robin Samuelsson

Examensarbete i matematik, 15 hp

Handledare: Jesper Rydén

Examinator: Magnus Jacobsson

Juni 2016

A large, faint watermark of the Uppsala University seal is visible in the bottom right corner of the page. The seal features a sun with rays, a cross, and the Latin motto "ALIIENSIS GRATIA VERITAS".

Department of Mathematics  
Uppsala University





UPPSALA  
UNIVERSITET

UPPSALA UNIVERSITET  
MATEMATISKA INSTITUTIONEN

EXAMENSARBETE D I MATEMATIK, 15 HP

# Multivariat databehandling och dataanalys

- en faktoranalys av slutbetyg i åk 9 inom  
Lgr11 på kommunivå

*Robin Samuelsson*

under handledning av  
Jesper RYDÉN

8 juni 2016

## Sammanfattning

År 2011 antogs en ny läroplan för grundskolan där man även inkluderade ett nytt betygssystem. Sedan tidigare finns det utbildningssociologiska studier (till exempel Lidegran (2009)) som visat på hur vissa geografiska områden kan inneha ett koncentrerat utbildningskapital eller andra typer av kapital. Den här studien söker potentiella strukturer i slutbetygen för årskurs 9 inom den nya läroplanen för definierade geografiska områden med hjälp av multivariat dataanalys. Metoder som faller inom multivariat dataanalys har länge använts och utvecklats av bland annat psykometriker och används inom så vitt skilda områden som kemi och sociologi. Just faktoranalys som används i denna studie har även tidigare används för studier av betyg och tester, bland annat av statistikern och psykologen Charles Spearman under början av 1900-talet.

In 2011 a new curriculum and grading system was adopted in the Swedish educational system. Studies from sociological research (for example Lidegran (2009)) shows that certain geographical areas may concentrate certain types of symbolic capital. This study seeks potential structures among final grades for ninth grade pupils at Swedish elementary school (under the new curriculum and for defined geographical areas). This is done with multivariate data analysis. Methods of multivariate data analysis have long been used and developed by for example psychometricians and are used in such diverse areas as chemistry and sociology. Factor analysis, that is used for this study, has been used for studies involving test results and grades, for example in studies made by Charles Spearman in early 20th century.

# Innehåll

<b>1</b>	<b>Inledning</b>	<b>2</b>
1.1	Syfte och frågeställning . . . . .	2
1.2	Avgränsningar och begränsningar . . . . .	2
<b>2</b>	<b>Teori och metod</b>	<b>3</b>
2.1	Utbildningssociologi och psykometri . . . . .	3
2.2	Inledande matematik och faktormodellen . . . . .	5
2.3	Att skatta modellens parametrar . . . . .	10
2.4	Att bestämma antal faktorer . . . . .	11
2.5	Faktorrotation och faktorvärden . . . . .	12
2.6	Reliabilitet och validitet . . . . .	14
2.7	Generell arbetsgång . . . . .	15
<b>3</b>	<b>Resultat</b>	<b>18</b>
3.1	Steg 1: Förberedelse av material . . . . .	18
3.2	Steg 2: Val av av faktoranalystyp . . . . .	19
3.3	Steg 3: Design och antagandekvalitet . . . . .	19
3.4	Steg 4: Analysmetod . . . . .	20
3.5	Steg 5-6: Metod för skattning, antal faktorer, rotation och analys	21
<b>4</b>	<b>Diskussion</b>	<b>26</b>
<b>5</b>	<b>Sammanfattning och kritik</b>	<b>33</b>
<b>6</b>	<b>Vidare forskning</b>	<b>34</b>
<b>7</b>	<b>Bilagor</b>	<b>35</b>
7.1	Bilaga 1: Tabeller och data . . . . .	35
7.2	Bilaga 2: Skript . . . . .	45
<b>8</b>	<b>Referenser</b>	<b>46</b>
8.1	Litteratur . . . . .	46
8.2	Hemsidor . . . . .	47
8.3	Bilder . . . . .	47

# 1 Inledning

## 1.1 Syfte och frågeställning

Idén till arbetet uppkom efter att jag läst Ida Lidegrans avhandling *Utbildningskapital: Om hur det alstras, fördelas och förmedlas* vilken jämför ett par norrorter i Stockholm med Uppsala just med grund i norrorternas ekonomiska kapital till skillnad från Uppsalas väldigt koncentrerade utbildningskapital<sup>1</sup>. Möjligen skulle liknande strukturer återspeglas hos elever och på kommunnivå.

Eftersom arbetet är av matematisk karaktär har jag utgått från vilka statistiska verktyg jag har tillgängliga och som faller inom ramarna för arbetets nivå och material. Arbetet ligger inom området tillämpad statistik och metoderna ska vara fördjupande, alltså dels ligga utanför det jag känner igen sedan tidigare kurser och dels kunna generera ett resultat för exempelvis ett tvärvetenskapligt område. Inom utbildningssociologin och psykometrin arbetar man ofta med statistik, dock oftast en form av klusteranalys eller en typ av geometrisk dataanalys kallad korrespondensanalys eller andra metoder som går under beteckningen multivariat dataanalys.

Frågeställning: Vilka möjliga latent strukturer, på kommunnivå, finns bland ämnena för åk 7-9 gällande det nya betygssystemet och läroplanen? Vad kan de bero av?

## 1.2 Avgränsningar och begränsningar

All data har hämtats från Skolverkets databas SiRiS.

I mitt arbete har jag kommit att begränsa mig till hur variationen mellan kommuner ser ut. Att just kommuner valts ut beror dels på att kommunerna styr grundskolan rent ekonomiskt och dels för att kunna avgränsa studien rent geografiskt (det ger även ett bra urval för statistisk data tack vare det stora antalet kommuner). Variablerna är genomsnittliga slutbetygspoäng, beräknade av Skolverket, för årskurs 9 i samtliga ämnen där det finns fullständig data för alla kommuner. Poängen går från 0-20 för betygen F-A<sup>2</sup>. De ämnen som faller inom det kriteriet redovisas under resultatsektionen i detta arbete. Samtliga tillgängliga huvudmän för datan i gällande databas har valts för arbetet vilket alltså innebär att både kommunala skolor och friskolor utövar underlag för de genomsnitt Skolverket beräknat från kommunernas slutbetygsresultat för respektive ämne. Året 14/15 har valts eftersom

---

<sup>1</sup>Lidegran, Ida, 2009, *Utbildningskapital: Om hur det alstras, fördelas och förmedlas*, Acta Universitatis Upsaliensis, s. 231-232

<sup>2</sup>F=0, E=10, D=12.5, C=15, B=17.5 och A=20.

detta är den senaste årskullen (och totalt andra) av högstadieelever, som betygssatts enligt den nya betygsskalan och läroplanen (och därmed de nya kursplanerna) hela högstadieperioden, erhållit ett slutbetyg för högstadiet. Eftersom slutbetyget i årskurs 9 utgör ett samlande betyg för elevens högstadiestudier skulle ett elevunderlag som bedömts enligt två olika läroplaner kunna vara något missvisande i arbetet.

## 2 Teori och metod

Teorin är uppdelade i två block. Det ena behandlar den humanistiska eller samhällsvetenskapliga teorin som kan användas för tolkning av resultaten och det senare behandlar matematiken för studien och den matematiska metoden. Utbildningssociologi och psykometri är tätt sammankopplade med tillämpad statistik och utgör därför ett naturligt underlag i undersökningar om testresultat, personlighet, egenskaper och åsikter. I det här fallet behandlar studien ett summerat resultat med definierade värden (slutbetyg) från bedömning av en yrkeskår med olika inriktningar (ämneslärare) vilket är ett bra material för faktoranalys som metod.

### 2.1 Utbildningssociologi och psykometri

I det utbildningssociologiska vetenskapsfältet görs ofta studier över olika kapitalformer, fält och habitus, allt efter Pierre Bourdieus teoretiska ramverk från hans verk såsom *La reproduction* från 1970. Här grundlade Bourdieu ett antal begrepp och ramverk som hjälpt till att förklara hur exempelvis eliter kan utnyttja utbildning som system för att ”befästa sin egen ställning”<sup>3</sup>. Ett av dessa begrepp är just kapitalbegreppet. Ett kapital kan vara materiellt men även symboliskt, gällande tillgångar, dock måste det tillskrivas ett värde av en social grupp och därmed ha en marknad<sup>4</sup>. En av de kanske mest kända kapitalformerna är det kulturella kapitalet. Det kulturella kapitalet behandlar ofta dominansförhållanden i samhället<sup>5</sup>.

Elevers betyg grundas i lärarnas bedömning men även i de förutsättningar eleven har med sig från hemmiljön. Bourdieu talade här om ett socialt kapital där kontakter och sociala nätverk utgjorde ett kapital i sig där de ingående in-

---

<sup>3</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15, s. 3

<sup>4</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15, s. 6-7

<sup>5</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15, s. 13



dividerna kan dra nytta av varandras symboliska kapital genom att konvertera det sociala kapitalet till exempelvis kulturellt eller ekonomiskt kapital.<sup>6</sup> På så sätt kan vedertagna och särskilt starka kontakter såsom familjeband utgöra en grundsten för elevernas förutsättningar i utbildningsväsendet. Det symboliska kapitalet kan emellertid även lagras i titlar eller institutioner såsom skolor (Bourdieu använder begreppet tillstånd för att beskriva de olika nivåer som kapital kan existera i<sup>7</sup>). För att erhålla dessa tillgångar krävs dock att man har det symboliska kapital som kan konverteras till titeln eller skolan man vill erhålla. Eftersom tillstånden skapas och bibehålls genom dispositioner och relationer blir dessa ofta föremål för strider mellan olika aktörer som vill hävda och förflytta sina positioner inom, och därmed även förändra, vad Bourdieu kallar fältet ("ett system av relationer mellan positioner"<sup>8</sup>).<sup>9</sup>

Inom psykologin använde man redan under tidigt 1900-tal barn och ungdomars resultat på diverse prov för att utröna övergripande faktorer som kunde förklara samband mellan resultaten. Under den här tiden var psykometriska metoder väldigt populära för kvalitetsgranskning inom psykologin.<sup>10</sup> En av de psykologer som arbetade med psykometrin var Charles Spearman som i sitt arbete om generell intelligens lade grunderna till den så kallade faktoranalysen som metod.<sup>11</sup> Metoden utvecklades sedan av psykologen Cyril Burt och ingenjören Louis Leon Thunderstone. Den skulle förklara de mönster man kunde finna i korrelationer mellan en uppsättning variabler genom att identifiera ett antal bakomliggande faktorer.<sup>12</sup> Spearman använde sina resultat för att klassificera intelligens och gjorde så med hjälp av två faktorer; en för generell intelligens, g-faktorn, och en för specifik intelligens, s-faktorn. Spearman menade på att en individs intelligens kunde beskrivas som summan av g-faktorn och en uppställning s-faktorer. G-faktorn var en starkt bidragande orsak till positiva resultat på intelligenstester medan s-faktorer var specifika för olika ämnen, förmågor eller prov.<sup>13</sup> Forskningsresultatet var

---

<sup>6</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15,s. 15

<sup>7</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15,s. 16

<sup>8</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15,s.19

<sup>9</sup>Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15,s. 19

<sup>10</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 481

<sup>11</sup>Teigen, Karl Halvor, 2006, En psykologihistoria, Liber, s. 296-299

<sup>12</sup>Teigen, Karl Halvor, 2006, En psykologihistoria, Liber, s. 300

<sup>13</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s. 434-435

dock kontroversiellt då många psykologer ansåg att intelligenstest endast visade på helt oberoende mentala förmågor och därför inte kunde utgöra ett resultat av någon typ av mer generell intelligens.<sup>14</sup> Senare forskare har dock modifierat idén om generell intelligens och det har bland annat föreslagits en intelligensmodell med sju faktorer (Thurstones Test of Primary Mental Abilities) och andra har till och med föreslagit att intelligens kan förklaras med upp till 150 faktorer. Dock finner man fortfarande inom Thurstones modell ett beroende mellan faktorerna vilket skulle kunna förklaras med hjälp av den tidigare föreslagna generella intelligensen.<sup>15</sup> Flertalet test och skalor för intelligens togs fram under tidigt 1900-tal, bland annat Stanford-Binet-skalan vilket ligger till grund för IQ och vars metoder används än idag dock något reviderade. 1986-års Stanford-Binet delar in intelligens i fyra färdigheter: verbalt, kvantitativt och visuellt resonemang samt korttidsminne.<sup>16</sup> En annan skala som utvecklades för att bättre beskriva intelligens hos vuxna, Wechslers vuxenintelligensskala, delade istället in intelligens i en verbal del och en utförandedel<sup>17</sup>. Det utvecklades även tester för grupper där man kunde testa färdigheter hos en större grupp individer på en gång, exempelvis skriftliga prov. I USA är exempelvis SAT ett exempel på sådant test<sup>18</sup> och i Sverige har vi nationella prov och högskoleprov. När jämförelser mellan betyg och SAT gjorts i USA har korrelationerna visat sig vara väldigt låga (strax över 0.3) vilket dock påverkas av att SAT görs av de som vill påbörja collegestudier och de som haft låga betyg därmed undviker att göra SAT.<sup>19</sup> I Sverige kan emellertid vissa av de nationella proven testas mot betygen eftersom ett antal av dessa är obligatoriska för alla elever.

## 2.2 Inledande matematik och faktormodellen

När man vill göra en studie över ett större antal variabler (manifesterade variabler) med många observationer och datas korrelationsmatris är svårtolkad kan antingen principalkomponentanalys (PCA) eller faktoranalys användas

---

<sup>14</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s. 434

<sup>15</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s. 434-435

<sup>16</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s. 431

<sup>17</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s. 432

<sup>18</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s. 433

<sup>19</sup>Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace, s.434

(PCA fungerar även som skattningsmetod för vissa fall av faktoranalysen<sup>20</sup>, mer om detta senare). Faktoranalys liknar PCA i och med att båda metoderna fördelar manifesterade variabler på ett antal, på förhand, dolda komponenter genom att studera hur dessa komponenter förklarar variationen hos variablerna. Båda metoderna fungerar även reducerande. Faktoranalysen förklarar de manifesterade variablerna med hjälp av så kallade latent variabler (faktorer). De manifesterade variablerna kan då grupperas mellan eller laddas på faktorerna (en manifesterad variabel kan emellertid tillhöra flera faktorer) beroende på hur pass stor andel av dess variation som kan förklaras med hjälp av faktorn.<sup>21</sup> <sup>22</sup> Även PCA utnyttjar variationerna hos de manifesterade variablerna men de fördelas då istället över komponenter. De båda metoderna skiljer sig främst i avseende på hur mycket av variationen som används för analysen vilket blir avgörande vid skattning av de ingående parametrarna i faktoranalysmodellen.<sup>23</sup> Detta förklaras närmare i kommande sektion. Vanligtvis måste dock ett antal antaganden göras i faktoranalysens inledande del vilket även brukar vara kritiken mot metoden<sup>24</sup>.

Faktoranalys kan ses som en metod för att skapa regressionsmodeller med  $m$  stycken variabler, här samlade i vektorn  $\mathbf{X}^T = [X_1, X_2, \dots, X_m]$ , vars korrelationer kan förklaras av de  $k$  stycken faktorerna eller latent variablerna  $\mathbf{F}^T = [F_1, F_2, \dots, F_k]$  där  $k < m$ . Väntevärdesvektorn för variablerna  $X_1, X_2, \dots, X_m$  betecknas  $\boldsymbol{\mu}$  och kovariansmatrisen för de samma med  $\boldsymbol{\Sigma}$ . Regressionsmodellerna kan då beskrivas med ekvationerna

$$\begin{aligned} X_1 - \mu_1 &= \Lambda_{11}F_1 + \Lambda_{12}F_2 + \dots + \Lambda_{1k}F_k + \epsilon_1 \\ X_2 - \mu_2 &= \Lambda_{21}F_1 + \Lambda_{22}F_2 + \dots + \Lambda_{2k}F_k + \epsilon_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ X_m - \mu_m &= \Lambda_{m1}F_1 + \Lambda_{m2}F_2 + \dots + \Lambda_{mk}F_k + \epsilon_m \end{aligned}$$

Dessa utgör själva faktormodellen vilken även kan skrivas på matrisform enligt följande

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{\Lambda}\mathbf{F} + \boldsymbol{\epsilon}$$

---

<sup>20</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 68-69

<sup>21</sup>Martin Paul och Bateson, Patrick, 2007, Measuring behaviour, Cambridge University Press, s. 116-117

<sup>22</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 65

<sup>23</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 68-69

<sup>24</sup>Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall, s. 83

där

$$\mathbf{\Lambda} = \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} & \cdots & \Lambda_{1,k} \\ \Lambda_{2,1} & \Lambda_{2,2} & \cdots & \Lambda_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{m,1} & \Lambda_{m,2} & \cdots & \Lambda_{m,k} \end{pmatrix}$$

är laddningsmatrisen. Faktorladdningen,  $\Lambda_{i,j}$ , anger här laddningen på faktor  $F_j$  för variabeln  $X_i$  och kan förstås som sambandet mellan variabeln och faktorn. Emellertid kan dessa endast antas representera korrelationen mellan variabel och faktor om ortogonal rotation genomförs vid rotationen av faktorerna (mer om detta senare)<sup>25</sup>.

Storheterna  $\epsilon_1, \epsilon_2, \dots, \epsilon_m$  är faktormodellens residualer. Dessa kallas de specifika faktorerna eftersom de är specifika för respektive variabel och variansen för  $\epsilon_i$  representerar den unika variansen för variabel  $X_i$ . De specifika faktorerna kan beskrivas med vektorn  $\boldsymbol{\epsilon}^T = [\epsilon_1, \epsilon_2, \dots, \epsilon_m]$ .<sup>26 27</sup>

Utöver detta brukar ett antal antaganden göras, bland annat att faktormodellen är ortogonal (vilket även innebär att de ingående faktorerna är oberoende varandra), samt följande egenskaper (från att faktorerna är standardiserade<sup>28</sup>), hos de specifika faktorerna  $\boldsymbol{\epsilon}$  och faktorerna  $\mathbf{F}$ , som följer av den ortogonala faktormodellen<sup>29</sup>:

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \text{ (med } n \times 1 \text{ element)}$$

och

$$\text{Cov}(\boldsymbol{\epsilon}) = E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \boldsymbol{\Psi} = \begin{pmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Psi_m \end{pmatrix}$$

där  $\boldsymbol{\Psi}$  alltså är en  $m \times m$ -diagonalmatrix och  $\psi_i$  är den unika variansen för  $X_i$ .

För  $\mathbf{F}$  görs dessa antaganden:

$$E(\mathbf{F}) = \mathbf{0} \text{ (med } k \times 1 \text{ element)}$$

och

$$\text{Cov}(\mathbf{F}) = E[\mathbf{F}\mathbf{F}^T] = \mathbf{I}$$

---

<sup>25</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 75

<sup>26</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 66

<sup>27</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 482

<sup>28</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 66

<sup>29</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 483

Dessutom gäller att  $\mathbf{F}$  och  $\boldsymbol{\epsilon}$  är oberoende och därmed att

$\text{Cov}(\boldsymbol{\epsilon}\mathbf{F}) = \text{E}(\boldsymbol{\epsilon}\mathbf{F}^T) = \mathbf{0}$  (nollmatris med  $m \times k$  element).

I övrigt kan man anta att  $\boldsymbol{\mu} = \mathbf{0}$  eftersom analysen behandlar kovariansstrukturen för  $X_1, X_2, \dots, X_m$ <sup>30</sup>. Många gånger görs även antagande om normalfördelning för  $\mathbf{F}$  och  $\boldsymbol{\epsilon}$  vilket då innebär normalfördelning för  $\mathbf{X}$ .<sup>31</sup>

Eftersom

$$\begin{aligned} (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T &= (\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\epsilon})^T \\ &= \boldsymbol{\Lambda}\mathbf{F}(\boldsymbol{\Lambda}\mathbf{F})^T + \boldsymbol{\epsilon}(\boldsymbol{\Lambda}\mathbf{F})^T + \boldsymbol{\Lambda}\mathbf{F}\boldsymbol{\epsilon}^T + \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \end{aligned}$$

så gäller följande för kovariansmatrisen (även från tidigare utskrivna egenskaper)  $\boldsymbol{\Sigma}$ :

$$\begin{aligned} \boldsymbol{\Sigma} &= \text{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \\ &= \boldsymbol{\Lambda}\text{E}(\mathbf{F}\mathbf{F}^T)\boldsymbol{\Lambda}^T + \text{E}(\boldsymbol{\epsilon}\mathbf{F}^T)\boldsymbol{\Lambda}^T + \boldsymbol{\Lambda}\text{E}(\mathbf{F}\boldsymbol{\epsilon}^T) + \text{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) \\ &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \end{aligned}$$

Sambandet mellan variablerna och faktorerna kan beskrivas med faktorladdningarna genom att

$$\text{Cov}(\mathbf{X}, \mathbf{F}) = \text{E}(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}^T = \boldsymbol{\Lambda}\text{E}(\mathbf{F}\mathbf{F}^T) + \text{E}(\boldsymbol{\epsilon}\mathbf{F}^T) = \boldsymbol{\Lambda}$$

vilket följer av

$$(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}^T = (\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\epsilon})\mathbf{F}^T = \boldsymbol{\Lambda}\mathbf{F}\mathbf{F}^T + \boldsymbol{\mu}\mathbf{F}^T$$

Variansen för en vald variabel,  $X_i$ , kan fås ur diagonalen på kovariansmatrisen  $\boldsymbol{\Sigma}$  och kan därför beräknas som

$$\sigma_i^2 = \sum_{j=1}^k \Lambda_{ij}^2 + \psi_i$$

varav den första delen av ekvationen brukar betecknas som, och ges av,

$$h_i^2 = \sum_{j=1}^k \Lambda_{ij}^2.$$

<sup>30</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 66

<sup>31</sup>Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall, s. 83

$h_i^2$  kallas komunalitet och visar på andelen delad varians genom de gemensamma faktorerna för  $X_i$  och övriga variabler som laddar på faktorerna. Den unika variansen,  $\psi_i$ , representerar andelen varians som är unik för variabeln  $X_i$ .

För att finna faktorerna och de specifika faktorerna måste emellertid  $\Psi$  och  $\Lambda$  skattas (låt estimatorerna betecknas som  $\hat{\Psi}$  och  $\hat{\Lambda}$  för respektive parameter). Detta kan göras genom att  $\Sigma$  först skattas med hjälp av datas kovariansmatris,  $\mathbf{S}$ , eller korrelationsmatris  $\mathbf{R}$ , om de manifesterade variablerna är standardiserade<sup>32</sup>. Men först är det bra om man utvärderar vald data. Detta kan bland annat göras med Kaiser-Meyer-Olkins sampling adequacy (KMO med MSA). KMO med MSA testar hur stor andel summan av de kvadrerade korrelationerna utgör av den totala summan för de kvadrerade korrelationerna tillsammans med de kvadrerade partialkorrelationerna vilket genererar ett värde mellan 0 och 1.

Det vill säga låt

$$\mathbf{R} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,n} \end{pmatrix}$$

vara korrelationsmatrisen för variablerna  $X_1, X_2, \dots, X_m$  och

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & p_{m,2} & \cdots & p_{m,n} \end{pmatrix}$$

vara en matris med partialkorrelationerna för  $X_1, X_2, \dots, X_m$ . Låt därefter diagonalelementen i  $R$  och  $P$  vara 0. Då gäller följande:

$$KMO = \frac{\sum_{i=1}^m (\sum_{j=1}^n (r_{i,j})^2)}{\sum_{i=1}^m (\sum_{j=1}^n (r_{i,j})^2) + \sum_{i=1}^m (\sum_{j=1}^n (p_{i,j})^2)}$$

samt

$$MSA_{X_i} = \frac{\sum_{i=1}^n (r_{i,j}^2)}{\sum_{i=j}^n (r_{i,j}^2) + \sum_{i=1}^n (p_{i,j})^2}$$

MSA testar alltså i grunden samma sak men med summeringar över respektive variablers kvadrerade korrelation med andra variabler istället för den totala

<sup>32</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 67-68

summan av alla korrelationer vilket alltså genererar värden för respektive variabel<sup>33</sup>. Variabler med högt MSA är användbara för faktoranalysen eftersom ju närmre 1 värdet ligger desto lägre är partialkorrelationerna. I partialkorrelationer låter man en variabel vara konstant för att se hur pass stor påverkan den har på andra variablers korrelation, denna bör då vara låg (alla värden över 0.7 är för höga) för att korrelationerna ska kunna förklaras med faktorer snarare än genom några av de enskilda variablerna.<sup>34</sup> Alltså bör KMO-värdet och MSA-värdena vara höga för att en faktoranalys ska vara lämplig för variablerna. Författarna av *Multivariate data analysis* använder 0.8 som en minimigräns för bra värden och skriver att värden under 0.5 bör indikera att variablerna är direkt olämpliga för faktoranalys.<sup>35</sup>

## 2.3 Att skatta modellens parametrar

Skattningen av parametrarna kan ske genom flera metoder varav de vanligaste är principalfaktoranalys (PFA) och maximum likelihood-faktoranalys (MLFA)<sup>36</sup>.  $\mathbf{S}$  är utbytbar mot  $\mathbf{R}$  i beräkningarna i de fall som beskrevs ovan. PFA utgår från den reducerade kovariansmatrisen  $\mathbf{S}_r = \mathbf{S} - \hat{\Psi}$  (eller reducerade korrelationsmatrisen  $\mathbf{R}_r = \mathbf{R} - \hat{\Psi}$ ) där  $R_{ii}$ , det vill säga diagonalelementen för vald matris ges av kommunaliteterna  $h_i^2 = 1 - \hat{\psi}_i$ . Detta innebär att  $\mathbf{R}_r$  kan beskrivas med skattade laddningar  $\hat{\Lambda}_{ij}$  från laddningsmatrisen  $\hat{\Lambda}_r$  genom  $\mathbf{R}_r = \hat{\Lambda}_r \hat{\Lambda}_r^T$ . Följande estimat används här:

$$\hat{\psi}_i = 1 - \sum_{j=1}^k \hat{\Lambda}_{ij}^2$$

samt

$$\hat{\Lambda}_r = [\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1, \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2, \dots, \sqrt{\hat{\lambda}_k} \hat{\mathbf{e}}_k]$$

där  $(\sqrt{\hat{\lambda}_i} \hat{\mathbf{e}}_i), i = 1, 2, \dots, k$  "are the (largest) eigenvalue-eigenvector pairs determined from  $\mathbf{R}_r$ ."<sup>37</sup> Då egenvärde-egenvektor-paren bestämts används des-

<sup>33</sup>Nakazawa, Minato, 2011, R practice: Factor analysis, <http://minato.sip21c.org/swtips/factor-in-R.pdf> (senast kontrollerad 26/05/2016)

<sup>34</sup>Snedecor, George W. and Cochran, William G., 1967, Statistical methods, Iowa state university press, s. 400-401

<sup>35</sup>Hair, Joseph F., et al., 2006, Multivariate data analysis, Pearson Prentice Hall, s. 114-115

<sup>36</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 68

<sup>37</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 495

sa för att beräkna nya kommunaliteter genom

$$h_i^2 = \sum_{j=1}^k \Lambda_{ij}^2$$

vilket itereras tills ett visst konvergenzkriterium uppnåtts (exempelvis tills summan av korrelationsmatrisens diagonal inte varierar mer<sup>38</sup>). Om kommunaliteterna, under processen, överstiger respektive manifesterade variablers varians uppstår ett så kallat *Heywood case* där minst en specifik varians är negativ vilket ger ett otolkbart resultat.<sup>39</sup>

I MLFA används MLE för att skatta parametrarna. Här används antagandet om multivariat normalitet för data samt funktionen

$$\Gamma = \ln |\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}| + \text{trace}(\mathbf{S}|\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}|^{-1}) - \ln |\mathbf{S}| - m$$

för att skatta parametrarna till faktormodellen. Då  $\mathbf{S} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}$  är funktionen  $\Gamma = 0$ . I övriga fall är  $\Gamma > 0$  vilket innebär att en minimering av funktionen  $\Gamma$  ger estimat för de sökta parametrarna. Minimeringen görs genom att maximera funktionen  $L = -\frac{1}{2}n\Gamma + \Phi$  där  $\Phi$  är en funktion av observationerna<sup>40</sup>. En längre redogörelse av metoden görs av Johnson och Wichern i *Some computational details for maximum likelihood estimation* från *Applied multivariate statistical analysis*<sup>41</sup>.

## 2.4 Att bestämma antal faktorer

En testvariabel kan beräknas för att bestämma antalet faktorer för faktormodellen:

$$U = \min(\Gamma)(n + 1 - \frac{1}{6}(2m + 5) - \frac{2}{3}k).$$

$U$  kan sedan testas mot  $\chi^2$ -fördelning med  $\nu$  frihetsgrader för att  $k$  faktorer ska gälla för faktormodellen. Antalet frihetsgrader beräknas enligt

$$\nu = \frac{1}{2}(m - k)^2 - \frac{1}{2}(m + k)$$

<sup>38</sup>Revelle, William, 12/05/2016, Package 'psych', s. 108, <https://cran.r-project.org/web/packages/psych/psych.pdf> (senast kontrollerad 27/05/2016)

<sup>39</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 69

<sup>40</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 69

<sup>41</sup>Johnson, Richard A. and Wichern, Dean W., 2007, *Applied multivariate statistical analysis*, Pearson Prentice Hall, s. 527-530



och följande hypoteser ställs mot varandra:

$H_0$  : det är tillräckligt med  $k$  gemensamma faktorer för faktormodellen

$H_1$  :  $\neg H_0$

En exakt metod för att avgöra antalet faktorer har emellertid ej utvecklats<sup>42</sup>, flera alternativ till  $\chi^2$ -testet har därför framförts<sup>43</sup>, bland annat kan man använda så kallade *scree plots* där antalet faktorer plottas mot egenvärden. Det är då brukligt att behålla alla faktorer med egenvärden  $\geq 1$  eftersom detta innebär att de då representerar lika mycket varians som en enda variabel. Metoden är dock vanligare för PCA och med ett stort antal observationer (fler än 50) kan för många faktorer extraheras.<sup>44</sup>

## 2.5 Faktorrotation och faktorvärden

För att tydliggöra sambanden mellan faktorer och manifesterade variabler, genom faktorladdningarna, kan en faktorrotation göras. Ett problem med faktormodellen som här kommer utnyttjas är det att det ej finns en unik lösning för laddningsmatrisen då det är fler än en faktor<sup>45</sup>, detta eftersom korrelationsmatrisen alltid kan återskapas av samtliga laddningar som skattas med en ortogonaltransformation<sup>46</sup>. Exempel på detta är om vi låter  $\mathbf{A}$  vara en ortogonalmatrix, av storlek  $k \times k$ , och denna läggs till i den ursprungliga ekvationen för faktormodellen på följande sätt:

$$\mathbf{X} = (\mathbf{\Lambda}\mathbf{A})(\mathbf{A}^T\mathbf{F}) + \boldsymbol{\epsilon}$$

De nya laddningarna och faktorerna kan då betecknas som  $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{A}$  samt  $\mathbf{F}^* = \mathbf{A}^T\mathbf{F}$  vilket medför att  $\boldsymbol{\Sigma}^* = \mathbf{\Lambda}^*\mathbf{\Lambda}^{*T} + \boldsymbol{\Psi}$ . Men eftersom  $\mathbf{A}\mathbf{A}^T = \mathbf{I}$  så är

$$\boldsymbol{\Sigma}^* = \mathbf{\Lambda}^*\mathbf{\Lambda}^{*T} + \boldsymbol{\Psi} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \boldsymbol{\Psi} = \boldsymbol{\Sigma}$$

Kovariansmatrisen ändras alltså inte vid införandet av en ortogonalmatrix i faktormodellens ekvation och därmed ej heller den specifika variansen eller

---

<sup>42</sup>Hair, Joseph F., et al., 2006, Multivariate data analysis, Pearson Prentice Hall, s. 119

<sup>43</sup>För andra sätt att testa antalet faktorer än de som tas upp här; Multivariate data analysis på sidorna 119-121 (se referenserna)

<sup>44</sup>Hair, Joseph F., et al., 2006, Multivariate data analysis, Pearson Prentice Hall, s. 120

<sup>45</sup>Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall, s. 85

<sup>46</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 495

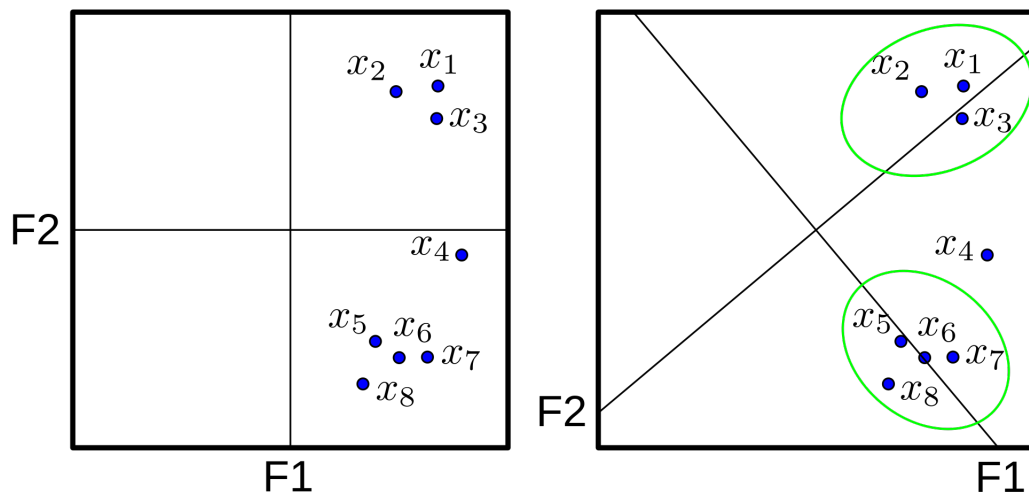
kommunaliteterna<sup>47</sup> vilket medför att det inte spelar någon roll matematiskt huruvida  $\Lambda^*$  eller  $\Lambda$  används. Det är detta som utnyttjas i faktorrotation där rotationsmatrisen är just en ortogonalmatris och ger nya (förhoppningsvis tydligare) faktorladdningar utan att förändra kommunaliteterna. Med två gemensamma faktorer kan rotationen studeras och bestämmas grafiskt. Man roterar då koordinataxlarna en viss vinkel  $\Theta$ , antingen motsols eller medsols, med hjälp av rotationsmatriserna

$$\Upsilon = \begin{pmatrix} \cos \Theta & \sin \Theta \\ -\sin \Theta & \cos \Theta \end{pmatrix}$$

eller

$$\Upsilon = \begin{pmatrix} \cos \Theta & -\sin \Theta \\ \sin \Theta & \cos \Theta \end{pmatrix}$$

där den första roterar medsols och den andra motsols. Hur rotationen går till grafiskt kan ses i figur 1. Men när antalet gemensamma faktorer är fler än



Figur 1: Illustration av ortogonal rotation. Av Fjalnes [CC0], via Wikimedia Commons

två till antalet måste en utvärdering göras av de roterade laddningarna utan grafisk representation.

Faktorrotation kan antingen göras ortogonalt eller skevt<sup>48</sup>. En ortogonal rotation antar att faktorerna är okorrelerade medan korrelation mellan fak-

<sup>47</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 504

<sup>48</sup>Oblique rotation på engelska

torer kräver skev rotation. Skev rotation medför då även att antagandet om ortogonalitet för faktormodellen förkastas.

En viktig skillnad, som medförs av detta, mellan att tolka resultatet efter skev rotation och ortogonal rotation, är att laddningar i den skeva rotationen ej kan tolkas som korrelationer mellan variablerna och faktorer. Detta kan emellertid göras vid en ortogonal rotation. För de två typerna av rotation finns flera tekniker varav de vanligaste är varimax (ortogonal) och promax (skev) men även quartimax och oblimin förekommer ofta för respektive typ av rotation<sup>49, 50</sup>.

Slutligen kan faktorvärden predikteras för varje manifesterad variabel och observation<sup>51</sup>. Det finns flera metoder för att finna faktorvärdena, för exempel på detta se Everitt (2005) eller Johnson och Wichern (2007) i referenserna. Faktorvärdena kan sedan användas för andra metoder i efterföljande studier.

## 2.6 Reliabilitet och validitet

Då studien görs i ett gränsområde mellan sociologi och tillämpad statistik krävs en diskussion kring validiteten och reliabiliteten. Reliabilitet behandlar upprepningsbarheten hos studien och baseras på hur pass mycket fel som kan finnas i studiens data jämfört med de verkliga värdena.<sup>52</sup> Dessa fel kan i denna studien antingen komma från själva databasen då myndigheten har mätt fel (felen återupprepas då vid ett nytt genomförande om någon skulle göra om studien och använder sig av samma databaser). De kan även uppstå vid varje överföringssteg av data från databasen till dokument eller program. Jag har emellertid klippt ut och klistrat in datan direkt från en enda databas för att minimera fel i detta.

Ett tredje sätt är vid standardiseringen av vissa variabler, exempelvis när ett antal görs om till andel för en viss variabel. Alla observationer är uppmätta med samma skala och samma gränser (en variabel kan anta 0 som lägst och 20 som högst) och är därmed redan standardiserade. Ett sista sätt är vid avrundningar; de eventuella avrundningar som skett har gjorts av Skolverket innan datan publicerats men värdena har då avrundats på samma sätt oavsett variabel eftersom samtliga variabler är hämtade ur samma

---

<sup>49</sup>För vidare beskrivningar av dessa; s. 75 i An R and S-Plus® Companion to multivariate analysis samt s. 507-513 i Applied multivariate statistical analysis (se referenserna)

<sup>50</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 75

<sup>51</sup>I An R and S-Plus® Companion to multivariate analysis menar Everitt på att faktorerna bör ses som slumpvariabler så det är därför fel att tala om en skattning av faktorvärden här

<sup>52</sup>Martin, Paul and Bateson, Patrick, 2007, Measuring behaviour, Cambridge University Press, s. 72-73

datamängd. Avrundningsfelen kommer dock då påverka hela variabeln men här kan det, beroende på hur avrundningen ser ut förekomma olika stora fel, i förhållande mellan enstaka observationer.

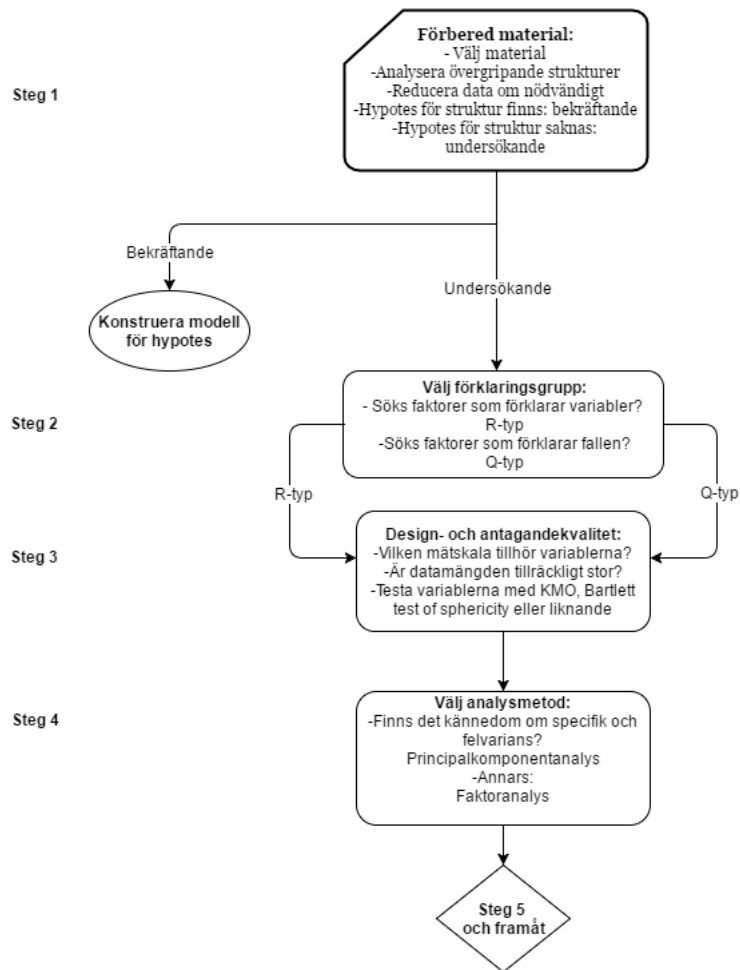
Validiteten behandlar huruvida det som ska mätas verkligen mäts eller snarare om mätningarna är relevanta för frågeställningen.<sup>53</sup> Här ligger problematiken snarare i valet av variabler och metod. Variablerna har valts för samtliga ämnen på högstadiet där data funnits för samtliga kommuner i Sverige. Se sektionen Resultat för mer om detta.

## 2.7 Generell arbetsgång

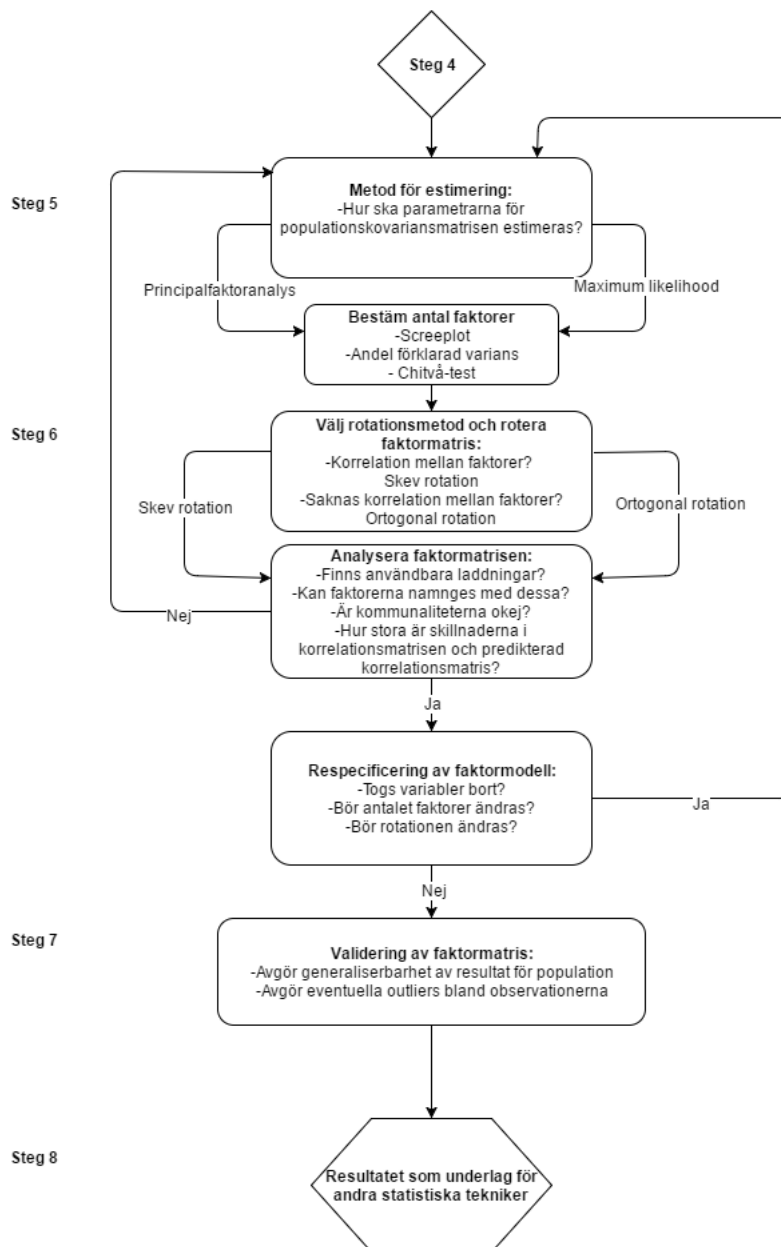
Arbetsgången för faktoranalys och PCA kan beskrivas med ett antal steg och val genom ett flödesdiagram. Jag har modifierat två redan konstruerade sådana av Hair et al. i *Multivariate data analysis*. Dessa finns i figur 2 och 3

---

<sup>53</sup>Martin, Paul and Bateson, Patrick, 2007, Measuring behaviour, Cambridge University Press, s. 73



Figur 2: Flödesschema för steg 1-4. Baserat på flödesdiagram i Multivariate data analysis av Hair et al.



Figur 3: Flödesschema för steg 5-8. Baserat på flödesdiagram i Multivariate data analysis av Hair et al.

## 3 Resultat

Genomsnittliga betygsvärden för varje kommun och ämne har hämtats från Skolverkets databas SiRiS. Genomsnittligt betyg för respektive ämne har här använts som variabler med observationer för varje kommun i Sverige där Sveriges kommuner är populationen för analysen. Då samtliga variabler har hämtats från samma datamaterial behövdes knappt någon bearbetning av datan för sammanställning. Den sökta datan extraherades ur ett enda excel-dokument och överfördes till ett nytt där det konverterades till .csv-format för inmatning i R. Allt skript som används finns summerat i bilaga 2 med referenser till källor som beskriver deras funktion.

Här nedan presenteras resultaten i enlighet med flödesschemat i figur 2 och figur 3. Då reduktion av variabler skett, eller dylik förändring av material som krävt återgång till ett tidigare steg, har detta resulterat i en ny omgång, oftast har dock flera steg gjorts samtidigt med en enda funktion (exempelvis skattning av parametrar och rotation). De presenterade resultaten kan vara avrundade; för fullständiga resultat som använts i beräkningarna, se bilaga 1. RStudio har använts för beräkningar. För eventuell information kring de skript som använts i programmet för att erhålla resultaten, se bilaga 2. Längre tabeller eller datamatriser finns i bilaga 1.

### 3.1 Steg 1: Förberedelse av material

Av ämnena som ges för grundskolans senare år (åk 7-9) valdes ämnena, där data saknades för någon kommun, bort för att data för hela den undersökta populationen (Sveriges kommuner) skulle kunna användas i analysen. Därmed gick moderna språk, svenska som andraspråk och modersmål bort från hämtad data. Följande ämnen användes i ett första steg i faktoranalysen och presenteras här i bokstavsordning: bild, biologi, engelska, fysik, geografi, hem- och konsumentkunskap (hemkunskap), historia, idrott och hälsa (idrott), kemi, matematik, musik, religionskunskap (religion), samhällskunskap, slöjd, svenska och teknik med variabelbeteckningar inom parentes. Datamängden förs in under beteckningen `rebe` (reducerade betyg) vilken återfinns i bilaga 1. Funktionen `cor(rebe)` genererar en korrelationsmatris (`correbe`) som presenteras i tabell 5 (bilaga 1).

Från korrelationsmatrisen kan man finna flera intressanta korrelationer. Språkämnen svenska och engelska har en hög korrelation ( $> 0.6$ ) vilket möjligen är väntat men även historia, geografi, matematik, religion och samhällskunskap korrelerar högt med språkämnen. För ämnet engelska gäller det även för fysik och biologi (även andra ämnen ligger nära en korrelation på 0.6). Om det hade visat sig att få variabler hade korrelationer på över 0.3

hade faktoranalys varit olämpligt för analys på datamängden<sup>54</sup>. Låga korrelationer tyder på att de specifika faktorerna kommer påverka resultatet mer än de gemensamma faktorerna<sup>55</sup>.

Eftersom det inte finns någon initial hypotes kommer analysen att vara explorativ i sin form vilket leder analysen till steg 2.

### 3.2 Steg 2: Val av faktoranalystyp

Här görs valet baserat på huruvida man vill klustra kommunerna eller finna strukturer för variablerna. Analysen behandlar strukturer mellan variablerna snarare än en klusteranalys av de olika kommunerna och därav används R-typ i studien (se figur 2). En studie med klustring av kommuner skulle emellertid kunna fungera som en potentiell uppföljningsstudie, mer om detta senare.

### 3.3 Steg 3: Design och antagandekvalitet

Gällande skalor kan en data delas in i någon (eller flera då de är delmängder av varandra) av skalorna nominalskala, ordinalskala, intervallskala och kvotskala. Datan som används i detta arbete hade kunnat anses tillhöra ordinalskalan i och med att betygen F till A kan jämföras i storleksordning med hjälp av operationerna  $<$ ,  $>$  och  $=$ . Men då betygen är omvandlade till en poängskala på 0–20 med ett genomsnitt för varje kommun kan kommunernas genomsnitt jämföras proportionerligt mot varandra med en entydig nollpunkt som här utgörs av värdet 0 vilket medför att datan då kan sägas tillhöra kvotskalan.<sup>56</sup>

Angående datamängd finns flera olika rekommendationer för detta. I *Multivariate data analysis* sätter författarna ett minimum på 50 observationer för stickprovet samt att det bör vara minst fem till tio observationer för varje variabel<sup>57</sup>. I denna undersökning utgörs materialet emellertid av hela den undersökta populationen vilken består av 290 observationer per variabel. Antalet variabler (16 stycken) är dessutom många gånger färre än antalet observationer. Generellt gäller ändå att fler observationer ger ett mer tillförlitligt resultat.

Korrelationerna räcker för att ge en överblick över möjliga strukturer men för att datan ska kunna användas med faktoranalys krävs det att man säkerställer aspekter såsom normalitet, homoskedasticitet och linjäritet. Dock

---

<sup>54</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 114

<sup>55</sup>Johnson, Richard A. and Wichern, Dean W., 2007, *Applied multivariate statistical analysis*, Pearson Prentice Hall, s. 488

<sup>56</sup>Alm, Sven E. and Britton, Tom, 2008, *Stokastik*, Liber, s. 412-413

<sup>57</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 113



är det inte alltid nödvändigt med avsaknad av multikollinearitet eftersom själva datamängden som används för faktoranalys kan delas upp i mängder under faktorerna som har en viss grad av korrelation<sup>58</sup>. För att kontrollera variablernas användbarhet för faktoranalys användes KMO med MSA (Kaiser-Meyer-Olkins sampling adequacy criteria med measure sampling adequacy). KMO-funktionen, vilken återfinns i bilaga 2, gav upphov till MSA-värden för variablerna i datamängden **rebe**. Dessa presenteras i tabell 4 (bilaga 1). Samtliga värden ligger alltså en bra bit över 0.8 och bör därför kunna anses väl lämpliga för en faktoranalys. Studien går därför vidare till steg 4.

### 3.4 Steg 4: Analysmetod

Principalkomponentanalysen skiljer sig även från faktoranalysen (även ibland kallat *common factor analysis*) genom att det främst har ett prediktionssyfte genom att summera så mycket som möjligt av den totala variansen i så få faktorer som möjligt medan faktoranalys främst används för att finna strukturer mellan variablerna<sup>59</sup>. Då principalkomponentanalys alltså använder den totala variansen hos variablerna använder faktoranalysen endast den gemensamma variansen för dem. Detta innebär att principalkomponentanalysen även kommer att generera komponenter som innehåller den unika variansen samt felvariansen medan faktoranalysen endast använder den gemensamma variansen.<sup>60</sup>

I faktoranalys kan laddningarna och faktorerna anta olika värden beroende på hur många faktorer man väljer att använda, detta gäller inte för komponentanalys där en förändring av antalet komponenter ej påverkar de ursprungliga komponenterna oavsett antal komponenter. Alltså, har man  $k$  faktorer i faktoranalysen och sedan ändrar det till  $k + 1$  stycken kommer de  $k$  första faktorerna att ändras, de första  $k$  komponenterna förändras emellertid inte av att utöka antalet till  $k + 1$  stycken komponenter.<sup>61 62</sup>

I *Multivariate data analysis* föreslår författarna att

”Component factor analysis is most appropriate when:

- *data reduction is a primary concern* (...), and

---

<sup>58</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 114

<sup>59</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 117

<sup>60</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 117-118

<sup>61</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 69

<sup>62</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 117

- (...) specific and error variance represent a *relatively small proportion* of the total variance.

Common factor analysis is most appropriate when:

- *the primary objective is to identify the latent dimensions or constructs* represented in the original variables, and
- the researcher has *little knowledge about the amount of specific and error variance (...)*.<sup>63</sup>

Den här studien söker först och främst den bakomliggande strukturerna bland de manifesterade variablerna och därför används gemensam faktoranalys<sup>64</sup>

### 3.5 Steg 5-6: Metod för skattning, antal faktorer, rotation och analys

Då funktionerna som används i följande avsnitt ger information om antal faktorer, skattar parametrar och kan faktorrottera på en och samma gång, slås steg 5 och 6 samman.

För att skatta parametrarna,  $\Lambda$  och  $\Psi$ , hos kovariansmatrisen för populationen,  $\Sigma$ , det vill säga finna  $\hat{\Lambda}$  och  $\hat{\Psi}$ , kan antingen principalfaktoranalys eller maximum likelihood användas. Maximum likelihood är en parametrisk metod och kräver därför att fördelningen är känd eftersom skattning görs med hjälp av täthets- eller fördelningsfunktionen. I det här fallet antas, som tidigare nämnts, en normalfördelning av data vilket krävs för metoden.<sup>65</sup> I figur 4 och 5 finns histogram och QQ-plot över data ur vilka man kan urskilja en normalfördelning för respektive variabel.

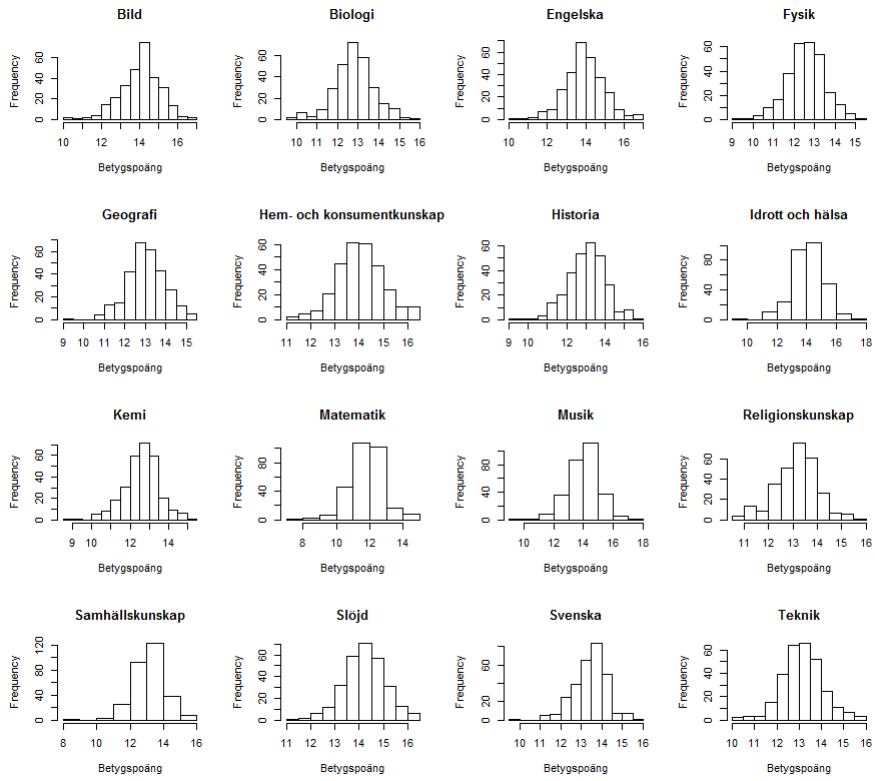
Författarna till *Applied multivariate statistical analysis* rekommenderar emellertid att man först testar principalfaktoranalysen utan och därefter med så kallad varimax-rotation varefter maximum likelihood testas och jämförs mot resultaten från principalfaktoranalysen (PFA).<sup>66</sup> Funktionen `fa()` användes för PFA och `factanal()` för maximum likelihood men först måste antalet faktorer avgöras och detta kan göras med funktionen `fa.parallel()` vilken föreslår

<sup>63</sup>Hair, Joseph F., et al., 2006, *Multivariate data analysis*, Pearson Prentice Hall, s. 118

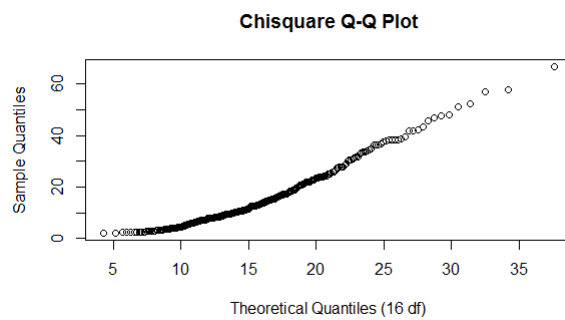
<sup>64</sup>På engelska *common factor analysis*.

<sup>65</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 69

<sup>66</sup>Johnson, Richard A. and Wichern, Dean W., 2007, *Applied multivariate statistical analysis*, Pearson Prentice Hall, s. 488, 520

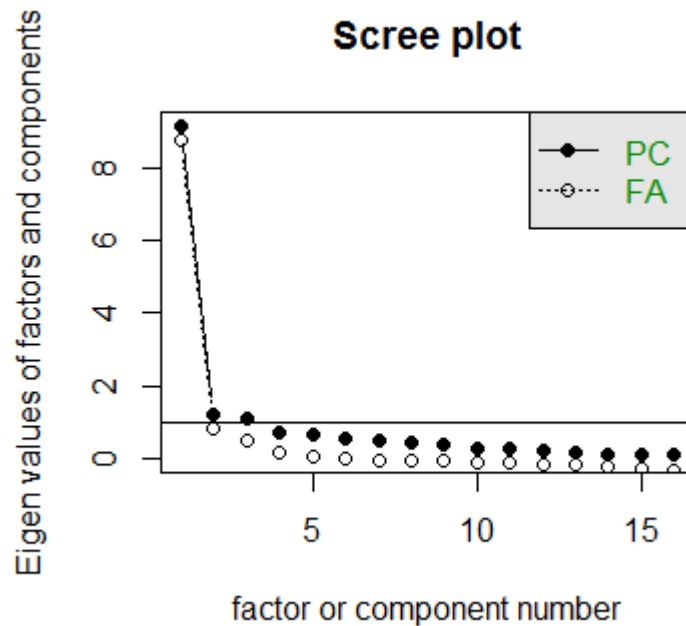


Figur 4: Histogram över datamängden ingående variabler för datamängd rebe



Figur 5: QQ-plot över ingående variabler för datamängd rebe

3 faktorer<sup>67</sup>. Ett annat sätt att avgöra antalet faktorer är genom en *scree plot*<sup>68</sup>. För studiens data finns en sådan i figur 6 .



Figur 6: *Scree plot* för korrelationsmatrisen av datamängden *rebe*. PC är principalkomponentanalys och FA är faktoranalys i figuren.

Från figur 6 skulle man troligtvis endast välja en faktor (för faktoranalys) eller möjligen tre (dock för principalkomponentanalys utifrån scree plot) vilka då har följande egenvärden avrundade till tre värdesiffror likt korrelationsmatrisen: 9.16, 1.20 samt 1.10<sup>69</sup>.

Även  $\chi^2$ -testet kontrollerades för att se hur många faktorer som krävdes för ett signifikant resultat. Dessa fås ut genom funktionen `fa()` eller `factanal()` vilka även ger laddningar, communaliteter och annan information. Everitt (2005) skriver dock att det formella testet som ges i den här typen av funk-

<sup>67</sup>Funktionen `fa.parallel()` extraherar fler faktorer (itererar) så länge egenvärdena är högre än de för en slumpmässig datamängd av samma storlek som den observerade. För mer info, se bilaga 2

<sup>68</sup>Egenvärdena plottas mot antalet faktorer eller principalkomponenter. För mer information se bilaga 2.

<sup>69</sup>Att just tre väljs beror utav att de har egenvärden som är större än 1. *Scree plots* är emellertid vanligare för PCA. Se tidigare avsnitt Att bestämma antal faktorer.

tioner<sup>70</sup> bör tas med en nypa salt och att antalet faktorer som ger ett signifikant resultat bör ses som den övre gränsen för antalet faktorer som kan ingå i modellen snarare än det exakta antalet<sup>71</sup>.

Rent praktiskt testades funktionen `fa()` med PFA först utan rotation och med iteration från en faktor upp till antal faktorer som gav signifikant resultat i  $\chi^2$ -testet. PFA visade sig dock endast kunna iterera upp till åtta faktorer. Maximum likelihood användes genom funktionen `factanal()` och gav ett signifikant resultat vid sex faktorer<sup>72</sup>. Eftersom funktionen `fa.parallel()` föreslog tre faktorer för faktormodellen och sex faktorer var signifikant i maximum likelihood enligt `factanal()`-funktionen bör modellen innehålla endera av dessa alternativ. För att reducera antalet faktorer valde jag sedan att jämföra laddningarna mellan PFA och maximum likelihood på tre faktorer samt sex faktorer för att kontrollera om de gav liknande resultat. Resultaten av `fa()` och `factanal()` med tre faktorer, utan rotation och grundat på korrelationsmatrisen för manifesterade variabler finns i tabell 6 (bilaga 1). Efter detta har en faktorrotation gjorts för att förbättra tolkningsbarheten hos resultatet. En varimax-rotation gjordes här för att behålla ortogonaliteten så att laddningarna kan tolkas som korrelationer mellan variabler och faktorer (det föreslås även av Johnson och Wichern som en standardprocedur<sup>73</sup>). Det är dock viktigt att påminna om att detta även medför att faktorerna är okorrelerade med varandra. Resultatet från varimaxrotationen finns i tabell 7 (bilaga 1). Till denna finns även, i tabell 11 (bilaga 1), värden på laddningssumman för varje faktor, andel varians som förklaras av varje faktor och så vidare. Funktionerna kördes sedan med sex faktorer utan och därefter med varimax-rotation av vilket resultatet återfinns i tabell 8 och 9 (bilaga 1). Att testa modellerna mot varandra kan göras på flera sätt. Först kan antalet relevanta faktorer kontrolleras med korrelation mellan faktorer och principalaxlar för att se när större skillnader börja uppträda i diagonalen hos korrelationsmatrisen. Det brukar emellertid föredras att man kontrollerar detta med faktorkongruens snarare än korrelation eftersom korrelation (Pearsons) baseras på standardavvikelser från medelvärden för faktorladdningarna medan faktorkongruens utgår från standardavvikelser från noll och

---

<sup>70</sup>Han använder `factanal()` som fungerar på samma sätt som `fa()`. Skillnaden ligger i hur de predikterar faktorvärden för skeva faktorrotationer. Mer om detta i bilaga 2.

<sup>71</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 83

<sup>72</sup>Ett p-värde på 0.0683 för nollhypotesen Sex faktorer är tillräckligt". Sju faktorer gav det något högre p-värde 0.346 för samma nollhypotes vilket ökar till 0.478 vid åtta faktorer och sedan faller vid fler än åtta faktorer.

<sup>73</sup>Johnson, Richard A. and Wichern, Dean W., 2007, *Applied multivariate statistical analysis*, Pearson Prentice Hall, s. 520

då kan tolkas som korrelationer mellan faktorerna i de olika modellerna snarare än korrelationerna mellan faktorladdningarna<sup>74</sup>. Dessutom rekommenderas faktorkongruens specifikt för explorativ faktoranalys, vilket är metoden för denna studie. Faktorkongruensen fås utav cosinus för vinkeln mellan de två vektorerna som jämförs (i det här fallet faktorladdningarna) och kan för faktorladdningsvektorerna  $\Lambda_i$  och  $\Lambda_j$  beräknas med funktionen  $\Omega$  genom

$$\Omega(\Lambda_i, \Lambda_j) = \frac{\sum \Lambda_{i,l} \Lambda_{j,l}}{\sqrt{\sum \Lambda_{i,l}^2 \sum \Lambda_{j,l}^2}}$$

med  $\Lambda_{i,l}$  och  $\Lambda_{j,l}$  som laddningarna på faktor  $i$  och  $j$  för variabel  $l$  där  $l = 1, \dots, 16$ .<sup>75</sup> Faktorkongruensens fördel är att den för faktorerna  $x$  och  $y$  är ”insensitive to scalar multiplication of  $x$  and  $y$ . This implies that it measures factor similarity independently of the mean absolute size of loadings: It can be high when loadings are near zero and vice versa.”<sup>76, 77</sup> Det finns flera olika gränser för vilka värden på faktorkongruenserna som är bra. Det har föreslagits allt från 0.8 till 0.98 att faktorerna ska kunna anses vara identiska eller att resultatet är väldigt bra<sup>78</sup>. Beräknas faktorkongruensen för tre- och sex-faktormodellerna med varimax-rotation och både maximum likelihood och PFA fås värdena i figur 7 (bilaga 1). De första tre faktorerna uppifrån och från vänster kommer från 3-faktormodellen med maximum likelihood-skattning, de tre följande principalaxlarna kommer från 3-faktormodellen med PFA-skattning och resterande kommer från respektive 6-faktormodell. De båda skattningsmetoderna genererar faktorer (eller principalaxlar) som har en hög (mellan 0.97 och 1.00) faktorkongruens för respektive faktorspar eller principalaxel-faktorspar i ordningen  $i = 1, 2, 3$ . Exempelvis har faktor 1 för 3-faktormodellen med maximumlikelihood-skattning och principalaxel 1 för 3-faktormodellen med PFA-skattning en faktorkongruens på 1.00. Faktorkongruenserna för samtliga par från samma ordning finns i bilaga 1 i figur 7. Av den anledningen kan en skattningsmetod bestämmas så länge samma

<sup>74</sup>Jensen, A.R., 1998, *The g factor: The science of mental ability*, CT:Praeger, s. 99

<sup>75</sup>Lorenzo-Seva, Urbano and Berge, Jos M.F. ten, 2006, Tucker’s congruence coefficient as a meaningful index of factor similarity, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), s. 57-58

<sup>76</sup>Lorenzo-Seva, Urbano and Berge, Jos M.F. ten, 2006, Tucker’s congruence coefficient as a meaningful index of factor similarity, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), s. 57

<sup>77</sup>Andra fördelar följer även av faktorkongruens vilka beskrivs av Lorenzo-Seva och Berge i Tucker’s congruence coefficient as a meaningful index of factor similarity.

<sup>78</sup>Lorenzo-Seva, Urbano and Berge, Jos M.F. ten, 2006, Tucker’s congruence coefficient as a meaningful index of factor similarity, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), s. 58

antalet faktormodeller fortsätter vara tre eller sex (eftersom en förändring av antal faktorer även förändrar faktorerna till skillnad från en PCA<sup>79</sup>). Från figur 7 (bilaga 1) kan man även dra slutsatsen att de tre första faktorerna är identiska oavsett om det är en 3-faktormodell eller 6-faktormodell däremot ser man att vid fyra faktorer är inte längre principalaxlarna, och deras respektive faktorer i den andra modellen, längre identiska (faktorkongruensen är 0.86 för den fjärde faktorn och principalaxeln och sjunker kraftigt vid den femte och sjätte faktorn). Jag har emellertid valt att testa 3-faktormodellen och 6-faktormodellen på ett ytterligare sätt vilket beskrivs nedan. Laddningarna för 6-faktormodellen (varimax-roterad) presenteras även i 9 (bilaga 1). För jämförbarhet presenteras laddningar för en varimax-roterad 1-faktormodell med maximum likelihood-skattning i tabell 10 (bilaga 1).<sup>80</sup> Ett annat sätt att utvärdera antalet faktorer på är att prediktera korrelationsmatriser med hjälp av faktormodellerna och jämföra de med de korrelationsmatriserna baserade på observationerna<sup>81</sup>. Skillnaderna för respektive modell (endast för med maximum likelihood av anledningen i föregående stycke) finns i bilaga 1 (figurerna 8, 9 samt 10). Skillnaderna har beräknats på samma tillvägagångssätt som Everitt (2005) gör i *An R and S-Plus® Companion to multivariate analysis* med druguse-exemplet<sup>82</sup>.

Steg 7 och 8 bör utföras vid fortsatta studier där man önskar använda modellen med andra metoder. Det kan då vara användbart att söka en generaliserbarhet hos modellen och prediktera faktornvärden men för denna studie har jag valt att avsluta metoden vid steg 6.

## 4 Diskussion

Om inget annat skrivs i följande avsnitt syftar respektive modell på en maximum likelihood-skattad och varimax-roterad modell. Som visas i resultatdelen med faktorkongruens verkar 3-faktormodellen och 6-faktormodellen ha närmast identiska tre första faktorer. Tittar man däremot på skillnader i predikterad korrelationsmatris och beräkning med observationer i figurerna 8 och 9 ser man ganska små skillnader vad gäller hur bra modellerna är

<sup>79</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 85

<sup>80</sup>Här med `fa()`-funktionen med maximum likelihood-skattning. `fa()` och `factanal()` fungerar på samma sätt med maximum likelihood och ortogonal rotation men `fa()` används med 1-faktormodell av tekniska skäl.

<sup>81</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s. 84

<sup>82</sup>Everitt, Brian, 2005, *An R and S-Plus® Companion to multivariate analysis*, Springer-Verlag London Limited, s.82-84

på att prediktera korrelationerna mellan variablerna. 6-faktormodellen ger en något bättre prediktion än 3-faktormodellen. Utifrån prediktionerna av korrelationsmatriserna skulle alltså 6-faktormodellen anses vara bäst för prediktion men de båda modellerna ligger inte långt ifrån varandra gällande hur väl de predikterar korrelationsmatrisen. En 3-faktormodell visar emellertid på samma strukturer med de tre första faktorerna som en 6-faktormodell.

Innan rotation gjordes verkar många variabler ladda på första faktorn eller principalaxeln för 3- och 6-faktormodellerna. Väljer man 1-faktormodellen laddar även många ämnen högt på faktorn (med varimax-rotation) men detta återkommer jag till senare.

Efter rotation blir laddningarna mycket lägre för flera ämnen på första faktorn. Detta kan bero på att det finns en bakomliggande faktor som beskriver den generella intelligens som Charles Spearman talar om i sin forskning som g-faktorn eller en liknande generell egenskap som gäller för en bredd av skolämnen. Den döljs isåfall vid rotation i och med att generella faktorer oftast försvinner vid varimax-rotation ”because the factor variance is redistributed”<sup>83</sup>. För att lättare kunna diskutera de specifika modellerna, följer tabeller för 1-faktormodellen, 3-faktormodellen och 6-faktormodellen (med maximum likelihood-skattning och varimax-rotation) i tabell 3, 1 respektive 2, där laddningar som är lägre 0.6 har tagits bort. Det här har baserats på en sammanställning av forskning som använder faktoranalys som metod där det, utifrån sammanställning, föreslås att ha 0.6 som avgränsning för laddningar<sup>84</sup>.

Här verkar faktor 1 utgöra någon typ av SO-faktor, faktor 2 en NO-faktor och faktor 3 en hantverksfaktor. Följer man däremot rådet, i Nathan Zhaos sammanställning, om att exkludera faktorer med färre än tre laddande variabler (efter exkludering av laddningar lägre än 0.6) så bör den tredje faktorn exkluderas. Då finns det två faktorer kvar att analysera; NO- och SO-faktorn. Detsamma gäller vid 6-faktormodellen där skillnaden är att alla tillkomna faktorer (faktor 4-6) bör exkluderas då de saknar variabler som laddar på dem samt att hemkunskap ej uppfyller kravet om laddning på 0.6 för faktor 3 längre (se tabell 2 här nedan).

NO- och SO-faktorn är alltså framträdande oavsett val av tillgängliga flerfaktormodeller. Orsakerna till att dessa två ämnesgrupper laddar så pass starkt på enskilda faktorer skulle kunna ha flera förklaringar. En förklaring utgår från eleverna: Elever som aktivt exempelvis väljer att pendla till andra

---

<sup>83</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 75

<sup>84</sup>Zhao, Nathan, 23/03/2009, The minimum sample size in factor analysis, <https://www.encyclopedia.com/entry/the-minimum-sample-size-in-factor-analysis> (senast kontrollerad 27/05/2016)



	Factor1	Factor2	Factor3
Bild			
Biologi		0.789	
Engelska			
Fysik		0.768	
Geografi	0.827		
Hemkunskap			0.655
Historia	0.815		
Idrott			
Kemi		0.841	
Matematik			
Musik			
Religion	0.816		
Samhällskunskap	0.827		
Slöjd			0.611
Svenska			
Teknik			

Tabell 1: Laddningar på respektive faktor för 3-faktormodell med maximum likelihood-skattning och varimax-rotation. Laddningar lägre än 0.6 har exkluderats.

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Bild						
Biologi		0.767				
Engelska						
Fysik		0.862				
Geografi	0.812					
Hemkunskap						
Historia	0.805					
Idrott						
Kemi		0.774				
Matematik						
Musik						
Religion	0.836					
Samhällskunskap	0.832					
Slöjd			0.713			
Svenska						
Teknik						

Tabell 2: Laddningar på respektive faktor för 6-faktormodell med maximum likelihood-skattning och varimax-rotation. Laddningar lägre än 0.6 har exkluderats.

kommuner kan dels göra det för att de bor på gränsen till en annan kommun och dels för att de gör ett aktivt val av skola eller kommun. Elever med högt kulturellt kapital har förutsättningarna för att göra detta aktiva val och kan då stärka de genomsnittliga resultaten i kommunerna de väljer och försvaga resultatet i kommunen de väljer bort, om de antas ha ett högt kulturellt kapital. Här skulle då NO-ämnena och SO-ämnena kunna antas fungera som fält där ett visst kulturellt kapital, som värderas högt inom respektive fält, koncentreras eller späds ut inom olika kommuner vilket skapar korrelationerna för just NO- eller SO-ämnena mot faktor 1 och faktor 2 i modellen. Kapital som är högt värderat inom SO-fältet är då inte lika betydelsefullt inom NO-fältet vilket skapar uppdelningen mellan faktorerna och områdena. Koncentreras kapitalet hos vissa kommuner skulle det även vara möjligt att individer utan det eftertraktade kapitalet gynnas av den stora gruppen som innehar det genom de potentiella relationer som skapas mellan grupperna. Även om du som individ saknar ett starkt nedärvt kulturellt kapital är sannolikheten stor att du känner någon, med akademikerföräldrar eller släktingar, i en kommun med hög koncentration av det eftertraktade kapitalet för fältet (möjligen kan individen då även sägas bli en del av själva fältet genom sina relationer). Implikationen här blir då att eleverna segregeras beroende på förutsättningar genom individernas kulturella kapital. Homogena grupper skapas i skolorna där några grupper utgör en elit och några en mer utsatt grupp, uppdelningar som sedermera riskeras återspeglas i samhället.

En annan förklaring skulle kunna utgå från lärarna som sätter betygen. Emil Bertilsson fann, i sin avhandling *Lärarna på skolans kungsväg*, att naturvetenskapliga lärare i Uppsala (då i matematik, fysik, kemi och biologi) oftast har mer kontakt med universitetet och många gånger utövar traditionell typ av undervisning<sup>85</sup>. Möjligen skulle detta kunna visa på homogeniteten hos just NO-lärare vilket avspeglas i elevernas prestationer och betyg. Ett sätt att tolka NO-faktorn skulle även kunna vara som akademisk faktor i och med naturvetenskapens traditionella och mer akademiska undervisningssätt. Potentiellt skulle SO-faktorn då snarare kunna ses som en allmänbildningsfaktor. En implikation av detta är att många elever bedöms inom olika system beroende på ämne och utgående från en viss typ av selektiv tradition inom ämnet. Professor Leif Östman visade redan på 90-talet hur NO-ämnena skilde sig från övriga ämnen gällande selektiva traditioner, emfaser och konceptioner<sup>86</sup>. En

---

<sup>85</sup>Bertilsson, Emil, 2007, *Lärarna på skolans kungsväg: Om det naturvetenskapliga programmet på några gymnasier i Uppsala*, SEC EDU Uppsala University, s. 54-55

<sup>86</sup>Östman, Leif, *Socialisation och mening: No-utbildning som politiskt och miljömoraliskt problem = [Socialization and meaning] : [science education as a political and environmental-ethical problem]*, Univ., Diss. Uppsala : Univ.,Uppsala, 1995 s. 114-119,151-152,154,174

ämnesgrupp eller ett ämne kan hålla på en viss diskurs och motivera samt välja ut innehåll utifrån denna diskurs (medvetet eller omedvetet). Frågan är hur rättssäkert skolsystemet är om olika ämnen har olika syften med undervisningen och gör olika val av innehållet baserat på dessa syften. Både elev- och lärarperspektivet kan skapa homogena grupper som exkluderar vissa elever (eller lärare för den delen) som inte har tillräckligt med symboliskt kapital för att kunna konkurrera inom ämnesfältet eller som har en annan inställning till ämnet än vad som utmärks genom ämnestraditionen.

Den sista modellen, 1-faktormodellen, skulle dock kunna ge en annan bild av strukturerna bland slutbetygen och ämnena. Tittar man på laddningar för denna och exkluderar laddningar som är lägre än 0.6 så erhålls tabell 3 här nedan. Här ser man att samtliga teoretiska ämnen laddar starkt på modellens enda faktor medan praktiska och estetiska ämnen exkluderats. Möjligen skulle detta kunna vara den generella faktorn som tidigare nämns och som beskriver någon typ av generell kunskap som krävs för teoretiska ämnen. En 1-faktorsmodell kan motiveras genom den *scree plot* som nämndes under resultatavsnittet. Den fungerar emellertid väldigt dåligt för prediktion vilket kan ses i skillnaderna mellan observerad och predikterad korrelationsmatris i figur 10 (bilaga 1).

	ML1
Bild	
Biologi	0.76
Engelska	0.79
Fysik	0.71
Geografi	0.92
Hemkunskap	
Historia	0.91
Idrott	
Kemi	0.69
Matematik	0.74
Musik	
Religion	0.90
Samhällskunskap	0.92
Slöjd	
Svenska	0.73
Teknik	0.60

Tabell 3: Laddningar på respektive faktor för 1-faktormodell med maximum likelihood-skattning och varimax-rotation. Laddningar lägre än 0.6 har exkluderats.

Alla tre modeller kan alltså på olika sätt motiveras för analys av strukturer, även om mycket pekar på att 6-faktormodellen eller möjligen 1-faktormodellen är bäst för fortsatt användning i uppföljande studier. Dock kan 3-faktormodellen användas för studier där samband mellan variabler genom latent struktur efterfrågas, likt denna, eftersom de tre första faktorerna då kan anses likvärdiga mot de i 6-faktormodellen och ämnena laddar likartat på de första faktorerna i de båda modellerna. Till strukturerna sett kan det dock vara bra att observera hur pass högt vissa ämnen laddade på faktorer inom modellerna i bilagorna men att de hamnade något under 0.6 och därför uteslöts ur tabellerna ovan. Exempelvis laddar matematik (0.498 och 0.573 för 6- och 3-faktormodellerna) och teknik (0.543 och 0.555 för samma modeller) högt på NO-faktorn. Ett annat intressant iakttagande är laddningarna på faktor 4 i 6-faktormodellen där engelska, matematik och svenska laddar högst (relativt laddningarna på den faktorn). Möjligen skulle denna faktor kunna klassificeras som en språkfaktor. Denna faktor ligger dock utanför resultatet av identiska faktorer med PFA-skattade modellerna som presenterades i resultatdelen.

## 5 Sammanfattning och kritik

För att sammanfatta det hela: genom den här studien har jag främst hittat strukturer i två större ämnesgrupper; SO och NO. Till detta kan det troligtvis även finnas en hantverksfaktor och en språkfaktor om modellen inkluderar fler faktorer. En generell faktor som döljs vid rotation skulle även kunna finnas som latent struktur bland variablerna och skulle även kunna vara den faktor som de teoretiska ämnena laddar på vid 1-faktormodellen. Det är emellertid viktigt att påpeka att tolkningsbarheten hos resultatet även oftast är kritiken mot faktoranalysen som metod. Jag har använt ett klassiskt och väl definierat material där matematiska antaganden även kunnat motiveras. Vid användning av material som är mer öppet för tolkning, exempelvis åsiktsgrundat material, där användaren gör matematiska antaganden utan att ha god grund för det (vilket ofta händer då antagandena många gånger ej är realistiska<sup>87</sup>) är det väldigt lätt för användaren av faktoranalysen att tolka resultatet utefter efterfrågat resultat. Till detta förekommer flertalet andra punkter med kritik, bland annat presenterat av Chatfield och Collins

---

<sup>87</sup>Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall, s. 88

(1980) i *Introduction to multivariate analysis*<sup>8889</sup> som är viktiga att tänka på när man valt sitt material. Chatfield och Collins argumenterar främst för att snarare använda PCA istället för faktoranalys och avslutar redogörelsen med ”In view of the disadvantages listed above, we recommend that FA should not be used in most practical situations, although we recognize that this view will be controversial.”<sup>90</sup>. Däremot har flera senare författare, vilka även refererats till i detta arbete, tillägnat kapitel åt metoden utan att förkasta metodens användbarhet. Everitt (2005) visar snarare på faktoranalysens likheter med PCA<sup>91</sup> och Johnson och Wichern (2007) gör en väldigt grundlig redogörelse för metoden och framför dess betydelse för beteendevetenskapen och samhällsvetenskapen men tillägger ändå en rekommendation av försiktighet i slutet av kapitlet på grund av metodens tolkningsbarhet<sup>92</sup>. Så länge en redogörelse görs över antaganden och steg som görs samt ett övervägande av lämpligheten av metoden för materialet bör ändå metoden kunna anses tillförlitlig.

## 6 Vidare forskning

Det finns ett antal sätt man kan gå vidare från resultatet av denna studie. Bland annat skulle man, likt det påpekats tidigare, kunna klustra kommuner utefter variablerna eller faktorerna. Alltså utföra en R-typ av faktoranalys alternativt gruppera kommuner med hjälp av predikterade faktorvärden från faktormodellen. Man skulle även kunna reducera antalet variabler och utanför en liknande explorativ faktoranalys utan en viss grupp av ämnen, exempelvis exkludera estetiska ämnen eller traditionellt praktiska.

Ett arbete med utgångspunkt i utbildningssociologin hade även varit intressant att se. Detta arbete har förvisso sociologiska begrepp och lånar delar av ett utbildningssociologiskt ramverk för slutanalysen men det görs inte samma djupa analys som troligtvis hade gjorts av sociologer (men däremot redogörs mycket av den matematiska bakgrunden för).

---

<sup>88</sup>Introduction to multivariate analysis

<sup>89</sup>Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall, s. 88-89

<sup>90</sup>Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall, s. 89

<sup>91</sup>Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited, s. 85,88

<sup>92</sup>Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall, s. 525-526

## 7 Bilagor

### 7.1 Bilaga 1: Tabeller och data

Tabell 4: MSA-värden för variabler i datamängden *rebe*

Variabel	MSA-värde	Variabel	MSA-värde
Bild	0.96	Kemi	0.92
Biologi	0.94	Matematik	0.95
Engelska	0.96	Musik	0.98
Fysik	0.93	Religion	0.94
Geografi	0.94	Samhällskunskap	0.94
Hemkunskap	0.94	Slöjd	0.93
Historia	0.94	Svenska	0.96
Idrott	0.95	Teknik	0.96



Tabell 5: Korrelationsmatris för datamängd rebe.

	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$	$V_7$	$V_8$
$V_1$ Bild	1	0.447	0.462	0.364	0.535	0.370	0.542	0.401
$V_2$ Biologi	0.447	1	0.683	0.768	0.627	0.383	0.603	0.470
$V_3$ Engelska	0.462	0.683	1	0.611	0.696	0.485	0.647	0.576
$V_4$ Fysik	0.364	0.768	0.611	1	0.591	0.333	0.586	0.424
$V_5$ Geografi	0.535	0.627	0.696	0.591	1	0.414	0.875	0.508
$V_6$ Hemkunskap	0.370	0.383	0.485	0.333	0.414	1	0.447	0.455
$V_7$ Historia	0.542	0.603	0.647	0.586	0.875	0.447	1	0.523
$V_8$ Idrott	0.401	0.470	0.576	0.424	0.508	0.455	0.523	1
$V_9$ Kemi	0.405	0.826	0.592	0.794	0.568	0.315	0.547	0.443
$V_{10}$ Matematik	0.369	0.728	0.730	0.676	0.629	0.445	0.618	0.558
$V_{11}$ Musik	0.361	0.479	0.485	0.443	0.446	0.414	0.455	0.358
$V_{12}$ Religion	0.524	0.651	0.662	0.616	0.871	0.378	0.876	0.464
$V_{13}$ Samhällskunskap	0.526	0.624	0.676	0.587	0.902	0.406	0.877	0.536
$V_{14}$ Slöjd	0.473	0.419	0.520	0.430	0.460	0.530	0.469	0.404
$V_{15}$ Svenska	0.427	0.558	0.668	0.467	0.636	0.422	0.669	0.439
$V_{16}$ Teknik	0.386	0.627	0.544	0.647	0.505	0.364	0.492	0.352

	$V_9$	$V_{10}$	$V_{11}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$	$V_{16}$
$V_1$ Bild	0.405	0.369	0.361	0.524	0.526	0.473	0.427	0.386
$V_2$ Biologi	0.826	0.728	0.479	0.651	0.624	0.419	0.558	0.627
$V_3$ Engelska	0.592	0.730	0.485	0.662	0.676	0.520	0.668	0.544
$V_4$ Fysik	0.794	0.676	0.443	0.616	0.587	0.430	0.467	0.647
$V_5$ Geografi	0.568	0.629	0.446	0.871	0.902	0.460	0.636	0.505
$V_6$ Hemkunskap	0.315	0.445	0.414	0.378	0.406	0.530	0.422	0.364
$V_7$ Historia	0.547	0.618	0.455	0.876	0.877	0.469	0.669	0.492
$V_8$ Idrott	0.443	0.558	0.358	0.464	0.536	0.404	0.439	0.352
$V_9$ Kemi	1	0.662	0.441	0.565	0.570	0.376	0.496	0.591
$V_{10}$ Matematik	0.662	1	0.478	0.603	0.601	0.474	0.632	0.545
$V_{11}$ Musik	0.441	0.478	1	0.430	0.459	0.408	0.432	0.448
$V_{12}$ Religion	0.565	0.603	0.430	1	0.860	0.423	0.646	0.472
$V_{13}$ Samhällskunskap	0.570	0.601	0.459	0.860	1	0.472	0.638	0.477
$V_{14}$ Slöjd	0.376	0.474	0.408	0.423	0.472	1	0.424	0.495
$V_{15}$ Svenska	0.496	0.632	0.432	0.646	0.638	0.424	1	0.435
$V_{16}$ Teknik	0.591	0.545	0.448	0.472	0.477	0.495	0.435	1

Tabell 6: Laddningar på respektive faktor eller principalaxel. Den övre tabellen är PFA-skattningen och den nedre maximum likelihood-skattningen. Funktionerna `factanal()` och `fa()` har körts med tre faktorer utan rotation (med korrelationsmatrisen för datamängd `rebe`).

	PA1	PA2	PA3
Bild	0.59	-0.12	0.14
Biologi	0.83	0.34	-0.16
Engelska	0.82	0.04	0.12
Fysik	0.78	0.36	-0.19
Geografi	0.87	-0.32	-0.15
Hemkunskap	0.55	-0.02	0.47
Historia	0.86	-0.34	-0.11
Idrott	0.62	-0.02	0.18
Kemi	0.77	0.41	-0.22
Matematik	0.80	0.19	0.04
Musik	0.58	0.09	0.17
Religion	0.85	-0.30	-0.22
Samhällskunskap	0.87	-0.33	-0.14
Slöjd	0.60	0.02	0.40
Svenska	0.73	-0.10	0.06
Teknik	0.66	0.26	0.05

	Factor1	Factor2	Factor3
Bild	0.581		0.153
Biologi	0.798	0.430	
Engelska	0.792	0.133	0.249
Fysik	0.753	0.414	-0.127
Geografi	0.914	-0.237	
Hemkunskap	0.494		0.490
Historia	0.901	-0.249	
Idrott	0.594		0.280
Kemi	0.741	0.498	-0.170
Matematik	0.762	0.284	0.173
Musik	0.544	0.145	0.221
Religion	0.898	-0.208	-0.112
Samhällskunskap	0.909	-0.242	
Slöjd	0.548		0.412
Svenska	0.720		0.189
Teknik	0.625	0.321	

Tabell 7: Laddningar på respektive faktor eller principalaxel. Den övre tabellen är PFA-skattning och den nedre maximum likelihood-skattning. Funktionerna `factanal()` och `fa()` har körts med tre faktorer och varimax-rotation (med korrelationsmatrisen för datamängd `rebe`).

	PA1	PA2	PA3
Bild	0.41	0.21	0.41
Biologi	0.35	0.79	0.28
Engelska	0.45	0.47	0.51
Fysik	0.31	0.79	0.22
Geografi	0.83	0.33	0.30
Hemkunskap	0.19	0.15	0.68
Historia	0.82	0.30	0.34
Idrott	0.34	0.28	0.47
Kemi	0.29	0.83	0.20
Matematik	0.36	0.59	0.44
Musik	0.25	0.34	0.44
Religion	0.83	0.36	0.23
Samhällskunskap	0.83	0.32	0.31
Slöjd	0.22	0.24	0.65
Svenska	0.51	0.33	0.41
Teknik	0.22	0.57	0.38

	Factor1	Factor2	Factor3
Bild	0.413	0.222	0.379
Biologi	0.335	0.789	0.307
Engelska	0.424	0.448	0.571
Fysik	0.326	0.768	0.240
Geografi	0.827	0.323	0.325
Hemkunskap	0.192	0.143	0.655
Historia	0.815	0.294	0.349
Idrott	0.316	0.275	0.509
Kemi	0.277	0.841	0.204
Matematik	0.330	0.573	0.504
Musik	0.245	0.333	0.443
Religion	0.816	0.359	0.260
Samhällskunskap	0.827	0.317	0.320
Slöjd	0.234	0.222	0.611
Svenska	0.484	0.311	0.474
Teknik	0.230	0.555	0.376

Tabell 8: Laddningar på respektive faktor eller principalaxel. Den övre tabellen är PFA-skattningen och den nedre maximum likelihood-skattningen. Funktionerna `factanal()` och `fa()` har körts med sex faktorer utan rotation (med korrelationsmatrisen för datamängd `rebe`).

	PA1	PA2	PA3	PA4	PA5	PA6
Bild	0.60	-0.15	0.15	0.21	0.16	0.24
Biologi	0.83	0.34	-0.15	-0.02	0.01	0.14
Engelska	0.82	0.03	0.12	-0.15	-0.12	0.02
Fysik	0.78	0.39	-0.20	0.09	0.06	-0.17
Geografi	0.87	-0.31	-0.18	0.03	0.01	-0.06
Hemkunskap	0.55	-0.04	0.42	0.02	-0.02	-0.04
Historia	0.86	-0.34	-0.13	0.04	0.01	-0.05
Idrott	0.64	-0.04	0.27	-0.35	0.32	-0.03
Kemi	0.77	0.41	-0.21	0.01	0.11	0.11
Matematik	0.81	0.20	0.05	-0.24	-0.15	-0.06
Musik	0.58	0.08	0.15	0.07	-0.07	0.01
Religion	0.85	-0.29	-0.25	0.05	-0.02	0.00
Samhällskunskap	0.86	-0.33	-0.16	0.02	0.06	-0.06
Slöjd	0.61	0.01	0.43	0.26	-0.02	-0.08
Svenska	0.73	-0.12	0.05	-0.13	-0.27	0.13
Teknik	0.67	0.27	0.05	0.21	-0.04	-0.07

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Bild	0.423	0.403	0.133		0.248	0.156
Biologi	0.813	0.184	-0.182	0.442		
Engelska	0.674	0.390	0.214	0.243		-0.113
Fysik	0.989	-0.129				
Geografi	0.679	0.644				
Hemkunskap	0.379	0.294	0.432	0.198	0.137	
Historia	0.672	0.633		-0.121		
Idrott	0.476	0.334	0.263	0.171	-0.144	
Kemi	0.818		-0.126	0.285		
Matematik	0.724	0.249	0.178	0.313	-0.170	-0.211
Musik	0.482	0.231	0.202	0.181		
Religion	0.699	0.607	-0.123	-0.118	0.126	-0.132
Samhällskunskap	0.677	0.661		-0.120	-0.102	0.128
Slöjd	0.471	0.262	0.477	0.138	0.224	0.168
Svenska	0.538	0.476	0.147	0.166		-0.193
Teknik	0.670		0.141	0.188	0.129	0.105

Tabell 9: Laddningar på respektive faktor eller principalaxel. Den övre tabellen är PFA-skattning och den nedre maximum likelihood-skattning. Funktionerna `factanal()` och `fa()` har körts med sex faktorer och varimax-rotation (med korrelationsmatrisen för datamängd `rebe`).

	PA1	PA2	PA3	PA4	PA5	PA6
Bild	0.40	0.19	0.41	0.13	0.02	0.37
Biologi	0.33	0.76	0.20	0.17	0.24	0.16
Engelska	0.42	0.42	0.38	0.29	0.39	0.01
Fysik	0.33	0.81	0.23	0.12	0.02	-0.09
Geografi	0.82	0.32	0.25	0.17	0.14	0.04
Hemkunskap	0.19	0.14	0.57	0.25	0.18	0.02
Historia	0.81	0.28	0.29	0.17	0.15	0.05
Idrott	0.28	0.23	0.30	0.68	0.12	0.06
Kemi	0.27	0.81	0.14	0.17	0.12	0.16
Matematik	0.32	0.55	0.30	0.30	0.41	-0.10
Musik	0.24	0.33	0.40	0.11	0.18	0.05
Religion	0.82	0.34	0.19	0.11	0.16	0.08
Samhällskunskap	0.82	0.31	0.25	0.21	0.11	
Slöjd	0.22	0.23	0.72	0.11	0.06	0.04
Svenska	0.48	0.27	0.30	0.14	0.49	0.07
Teknik	0.23	0.58	0.43	0.02	0.07	0.01

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Bild	0.410	0.199	0.475		0.139	
Biologi	0.333	0.767	0.233	0.294	0.291	
Engelska	0.415	0.387	0.401	0.497		
Fysik	0.308	0.862	0.217	0.136	-0.301	
Geografi	0.812	0.289	0.281	0.245		
Hemkunskap	0.185	0.128	0.598	0.258		
Historia	0.805	0.266	0.315	0.228		
Idrott	0.306	0.228	0.371	0.391		0.144
Kemi	0.283	0.774	0.195	0.221	0.101	
Matematik	0.307	0.498	0.304	0.596		
Musik	0.245	0.302	0.407	0.216		
Religion	0.836	0.332	0.216	0.193		-0.164
Samhällskunskap	0.832	0.287	0.281	0.209		0.194
Slöjd	0.211	0.207	0.713	0.128		
Svenska	0.479	0.232	0.321	0.456		
Teknik	0.219	0.543	0.422	0.133		

Tabell 10: Laddningar på respektive faktor. Funktionen  $\mathbf{fa}()$  har körts med en faktor, varimax-rotation och maximum likelihood-skattning (med korrelationsmatrisen för datamängd **rebe**).

	ML1
Bild	0.59
Biologi	0.76
Engelska	0.79
Fysik	0.71
Geografi	0.92
Hemkunskap	0.50
Historia	0.91
Idrott	0.59
Kemi	0.69
Matematik	0.74
Musik	0.54
Religion	0.90
Samhällskunskap	0.92
Slöjd	0.55
Svenska	0.73
Teknik	0.60

Tabell 11: Andel varians förklarad av 3-faktormodellerna. Maximum likelihood-baserade modellens värden har avrundats från tre till två decimaler.

	<b>PA1</b>	<b>PA2</b>	<b>PA3</b>	<b>Factor1</b>	<b>Factor2</b>	<b>Factor3</b>
<b>SS loadings</b>	4.10	3.68	2.75	3.99	3.58	2.94
<b>Proportion Var</b>	0.26	0.23	0.17	0.25	0.22	0.18
<b>Cumulative Var</b>	0.26	0.49	0.66	0.25	0.47	0.66

Tabell 12: Andel varians förklarad av 6-faktormodellerna med maximum likelihood. Maximum likelihood-baserade modellens värden har avrundats från tre till två decimaler.

	<b>Factor1</b>	<b>Factor2</b>	<b>Factor3</b>	<b>Factor4</b>	<b>Factor5</b>	<b>Factor6</b>
<b>SS loadings</b>	3.929	3.266	2.372	1.458	0.213	0.106
<b>Proportion Var</b>	0.246	0.204	0.148	0.091	0.013	0.007
<b>Cumulative Var</b>	0.246	0.450	0.598	0.689	0.702	0.709

Tabell 13: Andel varians förklarad av 6-faktormodellerna med PFA.

	<b>PA1</b>	<b>PA2</b>	<b>PA3</b>	<b>PA4</b>	<b>PA5</b>	<b>PA6</b>
<b>SS loadings</b>	3.91	3.45	2.13	0.95	0.81	0.24
<b>Proportion Var</b>	0.24	0.22	0.13	0.06	0.05	0.01
<b>Cumulative Var</b>	0.24	0.46	0.59	0.65	0.70	0.72

Tabell 14: Andel varians förklarad av 1-faktormodellerna med maximum likelihood.

	<b>ML1</b>
<b>SS loadings</b>	8.50
<b>Proportion Var</b>	0.53

	Factor1	Factor2	Factor3	PA1	PA2	PA3	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	PA1	PA2	PA3	PA4	PA5	PA6
Factor1	1.00	0.74	0.78	1.00	0.75	0.77	1.00	0.72	0.75	0.77	0.12	0.09	1.00	0.73	0.73	0.69	0.71	0.46
Factor2	0.74	1.00	0.77	0.75	1.00	0.75	0.74	1.00	0.73	0.80	0.15	0.06	0.74	1.00	0.72	0.68	0.73	0.39
Factor3	0.78	0.77	1.00	0.79	0.78	1.00	0.78	0.74	0.98	0.89	0.12	0.10	0.78	0.76	0.97	0.83	0.83	0.42
PA1	1.00	0.75	0.79	1.00	0.75	0.77	1.00	0.72	0.75	0.78	0.13	0.09	1.00	0.73	0.73	0.70	0.73	0.46
PA2	0.75	1.00	0.78	0.75	1.00	0.76	0.74	1.00	0.74	0.81	0.14	0.06	0.74	1.00	0.73	0.68	0.74	0.38
PA3	0.77	0.75	1.00	0.77	0.76	1.00	0.76	0.72	0.99	0.86	0.12	0.11	0.76	0.74	0.98	0.81	0.79	0.44
Factor1	1.00	0.74	0.78	1.00	0.74	0.76	1.00	0.71	0.74	0.76	0.13	0.09	1.00	0.72	0.72	0.68	0.71	0.47
Factor2	0.72	1.00	0.74	0.72	1.00	0.72	0.71	1.00	0.71	0.76	0.10	0.05	0.71	1.00	0.69	0.64	0.68	0.36
Factor3	0.75	0.73	0.98	0.75	0.74	0.99	0.74	0.71	1.00	0.78	0.12	0.11	0.75	0.72	1.00	0.75	0.71	0.49
Factor4	0.77	0.80	0.89	0.78	0.81	0.86	0.76	0.76	0.78	1.00	0.17	0.09	0.76	0.78	0.77	0.86	0.96	0.25
Factor5	0.12	0.15	0.12	0.13	0.14	0.12	0.13	0.10	0.12	0.17	1.00	-0.02	0.11	0.12	0.08	0.11	0.26	0.66
Factor6	0.09	0.06	0.10	0.09	0.06	0.11	0.09	0.05	0.11	0.09	-0.02	1.00	0.08	0.06	0.08	0.36	-0.09	0.03
PA1	1.00	0.74	0.78	1.00	0.74	0.76	1.00	0.71	0.75	0.76	0.11	0.08	1.00	0.73	0.73	0.68	0.71	0.45
PA2	0.73	1.00	0.76	0.73	1.00	0.74	0.72	1.00	0.72	0.78	0.12	0.06	0.73	1.00	0.71	0.65	0.70	0.37
PA3	0.73	0.72	0.97	0.73	0.73	0.98	0.72	0.69	1.00	0.77	0.08	0.08	0.73	0.71	1.00	0.72	0.70	0.45
PA4	0.69	0.68	0.83	0.70	0.68	0.81	0.68	0.64	0.75	0.86	0.11	0.36	0.68	0.65	0.72	1.00	0.70	0.35
PA5	0.71	0.73	0.83	0.73	0.74	0.79	0.71	0.68	0.71	0.96	0.26	-0.09	0.71	0.70	0.70	0.70	1.00	0.26
PA6	0.46	0.39	0.42	0.46	0.38	0.44	0.47	0.36	0.49	0.25	0.66	0.03	0.45	0.37	0.45	0.35	0.26	1.00

Figur 7: Faktorkongruenser för faktorer och principalaxlar från 3- och 6-faktormodellerna med varimax-rotation.

	Bild	Biologi	Engelska	Fysik	Geografi	Hemkunskap	Historia	Idrott	Kemi
Bild	-3.144828e-06	1.730028e-02	-2.897697e-02	-3.200684e-02	-1.192672e-03	1.043406e-02	8.166807e-03	1.656395e-02	2.682888e-02
Biologi	1.730028e-02	3.691068e-07	1.206236e-02	-2.118235e-02	-4.058727e-03	4.947146e-03	-9.093039e-03	-9.177738e-03	7.136045e-03
Engelska	-2.897697e-02	1.206236e-02	1.296339e-07	-9.205070e-03	1.445995e-02	-3.424904e-02	-3.058984e-02	2.812649e-02	-1.889574e-02
Fysik	-3.200684e-02	-2.118235e-02	-9.205070e-03	-5.041864e-07	-5.262325e-03	3.185555e-03	9.721057e-03	-1.257263e-02	8.402509e-03
Geografi	-1.192672e-03	-4.058727e-03	1.445995e-02	-5.262325e-03	2.556370e-07	-3.889814e-03	-8.311691e-03	-8.074931e-03	1.228776e-03
Hemkunskap	1.043406e-02	4.947146e-03	-3.424904e-02	3.185555e-03	-3.889814e-03	-2.220324e-06	1.915694e-02	2.174460e-02	8.050273e-03
Historia	8.166807e-03	-9.093039e-03	-3.058984e-02	9.721057e-03	-8.311691e-03	1.915694e-02	-4.219773e-07	6.255700e-03	1.959679e-03
Idrott	1.656395e-02	-9.177738e-03	2.812649e-02	-1.257263e-02	-8.074931e-03	2.174460e-02	6.255700e-03	3.011407e-06	2.022422e-02
Kemi	2.682888e-02	7.136045e-03	-1.889574e-02	8.402509e-03	1.228776e-03	8.050273e-03	1.959679e-03	2.022422e-02	-1.407734e-07
Matematik	-8.589011e-02	1.053984e-02	4.553551e-02	6.912036e-03	6.604265e-03	-3.029023e-02	4.356610e-03	3.928332e-02	-1.477860e-02
Musik	1.813990e-02	-8.422827e-04	-2.055975e-02	7.735340e-04	-7.005563e-03	2.950827e-02	3.324540e-03	-3.630791e-02	2.743576e-03
Religion	9.266085e-03	1.445976e-02	6.366768e-03	1.148138e-02	-4.172180e-03	-8.368476e-04	1.434141e-02	-2.531180e-02	-1.635835e-02
Samhällskunskap	-6.722520e-03	-1.286024e-03	3.213833e-04	-3.182546e-03	1.201120e-02	-7.784781e-03	-2.687853e-03	2.449937e-02	8.751477e-03
Slöjd	9.535393e-02	-2.165978e-02	-2.695992e-02	3.575796e-02	-3.291097e-03	5.324750e-02	-7.072228e-04	-4.168327e-02	-7.089909e-04
Svenska	-2.162092e-02	5.220664e-03	5.345671e-02	-4.377009e-02	-1.819162e-02	-2.506785e-02	1.797144e-02	-3.992345e-02	3.336377e-03
Teknik	2.549716e-02	-3.325938e-03	-1.720626e-02	5.512657e-02	1.303723e-02	-5.886406e-03	9.586400e-03	-6.497601e-02	-1.599958e-02
Matematik	-8.589011e-02	1.053984e-02	4.553551e-02	6.912036e-03	-7.735340e-04	-7.005563e-03	2.950827e-02	3.324540e-03	-3.630791e-02
Biologi	1.053984e-02	-8.422827e-04	1.445976e-02	-1.286024e-03	-2.165978e-02	5.220664e-03	-3.325938e-03	-1.720626e-02	5.512657e-02
Engelska	4.553551e-02	-2.055975e-02	6.366768e-03	3.213833e-04	-2.695992e-02	5.345671e-02	-1.720626e-02	-1.720626e-02	5.512657e-02
Fysik	6.912036e-03	7.735340e-04	1.148138e-02	-3.182546e-03	3.575796e-02	-4.377009e-02	5.512657e-02	5.512657e-02	5.512657e-02
Geografi	6.604265e-03	-7.005563e-03	-4.172180e-03	1.201120e-02	-3.291097e-03	-1.819162e-02	1.303723e-02	1.303723e-02	1.303723e-02
Hemkunskap	-3.029023e-02	2.950827e-02	-8.368476e-04	-7.784781e-03	5.324750e-02	-2.506785e-02	-5.886406e-03	-5.886406e-03	-5.886406e-03
Historia	4.356610e-03	3.324540e-03	1.434141e-02	-2.687853e-03	-7.072228e-04	1.797144e-02	9.586400e-03	9.586400e-03	9.586400e-03
Idrott	3.928332e-02	-3.630791e-02	-2.531180e-02	2.449937e-02	-4.168327e-02	-3.992345e-02	-6.497601e-02	-6.497601e-02	-6.497601e-02
Kemi	-1.477860e-02	2.743576e-03	-1.635835e-02	8.751477e-03	-7.089909e-04	3.336377e-03	-1.599958e-02	-1.599958e-02	-1.599958e-02
Matematik	-1.190609e-06	-1.660972e-02	-2.911969e-03	-1.471798e-02	-3.862239e-02	5.522392e-02	-3.908569e-02	-3.908569e-02	-3.908569e-02
Musik	-1.660972e-02	-6.783735e-07	-3.929492e-03	1.025052e-02	6.875384e-03	9.660313e-04	4.046241e-02	4.046241e-02	4.046241e-02
Religion	-2.911969e-03	-3.929492e-03	-3.753577e-08	-1.133956e-02	-6.532269e-03	1.678420e-02	-1.236304e-02	-1.236304e-02	-1.236304e-02
Samhällskunskap	-1.471798e-02	1.025052e-02	-1.133956e-02	-5.394266e-07	1.304047e-02	-1.201826e-02	-9.340797e-03	-9.340797e-03	-9.340797e-03
Slöjd	-3.862239e-02	6.875384e-03	-6.532269e-03	1.304047e-02	1.505782e-06	-4.784009e-02	8.793158e-02	8.793158e-02	8.793158e-02
Svenska	5.522392e-02	9.660313e-04	1.678420e-02	-1.201826e-02	-4.784009e-02	5.310281e-06	-2.696416e-02	-2.696416e-02	-2.696416e-02
Teknik	-3.908569e-02	4.046241e-02	-1.236304e-02	-9.340797e-03	8.793158e-02	-2.696416e-02	7.644291e-07	7.644291e-07	7.644291e-07

Figur 8: Skillnader mellan observerad och predikterad korrelationsmatris för 3-faktormodellen med maximum likelihood-skattning.





## 7.2 Bilaga 2: Skript

De flesta funktionerna kommer från paketet `psych` som är en uppsättning skript främst för multivariat dataanalys och psykometri. Funktionerna `factanal()` och `cor()` är emellertid inkluderat i R. `chisqplot()` är tagen från en annan källa (se nedan).

`cor()`

Genererar korrelationsmatrisen. Funktion inkluderad i R.

`fa()`

Faktoranalys som kan köras med både `fm="pa"` och `fm="ml"` (PFA- och maximum likelihood-skattning). Funktion hämtad från `psych`-paketet. För mer information: <http://www.inside-r.org/packages/cran/psych/docs/fa> (senast kontrollerad 26/05/2016)

`factanal()`

Funktion inkluderad i R. Fungerar som `fa()` men endast med maximum likelihood.

`fa.parallell()`

Funktionen utvärderar antalet faktorer/komponenter. Funktion hämtad från `psych`-paketet. För mer information: <http://www.inside-r.org/packages/cran/psych/docs/fa.parallell> (senast kontrollerad 26/05/2016)

`KMO()`

KMO med MSA. Funktion hämtad från `psych`-paketet. För mer information: <http://www.inside-r.org/packages/cran/psych/docs/KMO> (senast kontrollerad 26/05/2016). Även `kmo()` kan användas för liknande resultat. Denna finns beskriven på <http://minato.sip21c.org/swtips/factor-in-R.pdf> (senast kontrollerad 26/05/2016).

`chisqplot()`

Chitvå-qplot för normalitetstest. Skript hämtat från: <http://www.stat.ncsu.edu/people/bloomfield/courses/ST731/chisqplot.r> (senast kontrollerad 26/05/2016)

`factor.congruence(list())`

Funktionen jämför olika faktormodeller mot varandra (jämför uppsättningar av faktorladdningar). Se mer på `factor.congruence` i <https://cran.r-project.org/web/packages/psych/psych.pdf> (senast kontrollerad 26/05/2016) el-

ler i använd form längst ned på <http://www.inside-r.org/packages/cran/psych/docs/fa> (senast kontrollerad 26/05/2016)

### **scree()**

Egenvärden för faktorerna eller principalkomponenter plottas mot antalet faktorer eller principalkomponenter för att ge en visuell indikation på hur många faktorer som bör inkluderas i modellen. För mer information: <http://www.inside-r.org/packages/cran/psych/docs/VSS.scree> (senast kontrollerad 26/05/2016)

## **8 Referenser**

### **8.1 Litteratur**

Alm, Sven E. and Britton, Tom, 2008, Stokastik, Liber

Atkinson, Rita L. et al., 2000, Hilgard's Introduction to Psychology, Harcourt Brace

Bertilsson, Emil, 2007, Lärarna på skolans kungsväg: Om det naturvetenskapliga programmet på några gymnasier i Uppsala, SEC EDU Uppsala University

Broady, Donald, Kapitalbegreppet som utbildningssociologiskt verktyg, 1998, Skeptronhäften Nr 15

Chatfield, Christopher, and Collins, Alexander J., 1980, Introduction to multivariate analysis, Chapman and Hall

Everitt, Brian, 2005, An R and S-Plus® Companion to multivariate analysis, Springer-Verlag London Limited

Hair, Joseph F., et al., 2006, Multivariate data analysis, Pearson Prentice Hall

Jensen, A.R., 1998, The g factor: The science of mental ability, CT:Praeger

Johnson, Richard A. and Wichern, Dean W., 2007, Applied multivariate statistical analysis, Pearson Prentice Hall

Lidegran, Ida, 2009, Utbildningskapital: Om hur det alstras, fördelas och fömedlas, Acta Universitatis Upsaliensis

Lorenzo-Seva, Urbano and Berge, Jos M.F. ten, 2006, Tucker's congruence coefficient as a meaningful index of factor similarity, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2)

Martin Paul och Bateson, Patrick, 2007, *Measuring behaviour*, Cambridge University Press

Snedecor, George W. and Cochran, William G., 1967, *Statistical methods*, Iowa state university press

Teigen, Karl Halvor, 2006, *En psykologihistoria*, Liber

Östman, Leif, *Socialisation och mening: No-utbildning som politiskt och miljömoraliskt problem = [Socialization and meaning] : [science education as a political and environmental-ethical problem]*, Univ., Diss. Uppsala : Univ., Uppsala, 1995

## 8.2 Hemsidor

Nakazawa, Minato, 27/06/2011, R practice: Factor analysis, <http://minato.sip21c.org/swtips/factor-in-R.pdf> (senast kontrollerad 26/05/2016)

Revelle, William, 12/05/2016, Package 'psych', <https://cran.r-project.org/web/packages/psych/psych.pdf> (senast kontrollerad 27/05/2016)

Skolverket - SiRis, Urval - Statistikexport Skol- eller kommunnivå, grundskolan, <http://sir.is.skolverket.se>(senast kontrollerad 27/05/2016)

## 8.3 Bilder

Illustration av ortogonal rotation. Av Fjalnes [CC0], via Wikimedia Commons, [https://upload.wikimedia.org/wikipedia/commons/3/38/Orthogonale\\_faktorrotation.svg](https://upload.wikimedia.org/wikipedia/commons/3/38/Orthogonale_faktorrotation.svg)