Kvantitatiivsed mudelid käitumisteadustes - praktikum nr 6: struktuurivõrrandite mudelid

Martin Kolnes, Dmitri Rozgonjuk, Karin Täht

26 aprill 2017

**Struktuurivõrrandite mudelid** (SEM) on üsna mitmekesine meetodite rühm, mis omab sarnasusi nii faktor- kui regressioonanalüüsiga. Nagu uurivas faktoranalüüsis saame vaadeldud muutujate kaudu defineerida liitmuutujaid ehk faktoreid. Struktuurivõrrandite mudeliste kontekstis nimetatakse neid tüüpiliselt **latentseteks muutujateks**. Aga mudelisse saab lisada veel muid vaadeldud muutujaid (mis ei osale latentsete muutujate defineerimises) ja uurida muutujatevahelisi regressioonseoseid.  
Kinnitav faktoranalüüs on selline struktuurivõrrandite mudel, mis sisaldab ainult latentsete muutujate osa ja milles puuduvad regressioonseosed. Samas saab teha ka sellise mudeli (nn teeanalüüs, ingl k *path analysis*), mis sisaldab ainult vaadeldud muutujaid ja nende vahelisi regressioonseoseid ilma igasuguste latentsete muutujateta. Selles praktikumis kasutame R-i lisamoodulit *lavaan* (nimetus tuleb sõnadest *latent variable analysis*).

install.packages("lavaan")

Kasutame esialgu lavaani mooduliga kaasa tulevat klassikalist näidisandmestikku nimega *HolzingerSwineford1939*, mis sisaldab 7. ja 8. klassi õpilaste kohta üheksat vaimse võimekuse alatesti skoori. Laadime lavaani lisamooduli ja vaatame andmestiku esimesi ridu.

library(lavaan)  
head(HolzingerSwineford1939)

## id sex ageyr agemo school grade x1 x2 x3 x4 x5  
## 1 1 1 13 1 Pasteur 7 3.333333 7.75 0.375 2.333333 5.75  
## 2 2 2 13 7 Pasteur 7 5.333333 5.25 2.125 1.666667 3.00  
## 3 3 2 13 1 Pasteur 7 4.500000 5.25 1.875 1.000000 1.75  
## 4 4 1 13 2 Pasteur 7 5.333333 7.75 3.000 2.666667 4.50  
## 5 5 2 12 2 Pasteur 7 4.833333 4.75 0.875 2.666667 4.00  
## 6 6 2 14 1 Pasteur 7 5.333333 5.00 2.250 1.000000 3.00  
## x6 x7 x8 x9  
## 1 1.2857143 3.391304 5.75 6.361111  
## 2 1.2857143 3.782609 6.25 7.916667  
## 3 0.4285714 3.260870 3.90 4.416667  
## 4 2.4285714 3.000000 5.30 4.861111  
## 5 2.5714286 3.695652 6.30 5.916667  
## 6 0.8571429 4.347826 6.65 7.500000

Nagu näha kannavad meid huvitavad muutujad nimesid x1 kuni x9, mis meile kuigi palju ei ütle. Alloleva käsu abil peaks RStudio akna alumises parempoolses osas avanema andmestiku täpsem kirjeldus, milles on ära toodud alatestide sisu.

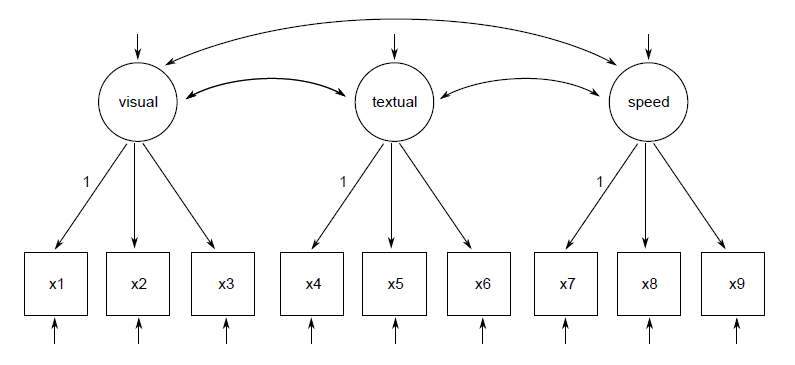
help(HolzingerSwineford1939)

Selle käsu jooksutamine peaks avama *Help* akna selle andmestiku kirjeldusega. Seal on ära toodud ka muutujate nimed:

* id - Identifier
* sex - Gender
* ageyr - Age, year part
* agemo - Age, month part
* school - School (Pasteur or Grant-White)
* grade - Grade
* x1 - Visual perception
* x2 - Cubes
* x3 - Lozenges
* x4 - Paragraph comprehension
* x5 - Sentence completion
* x6 - Word meaning
* x7 - Speeded addition
* x8 - Speeded counting of dots
* x9 - Speeded discrimination straight and curved capitals

# Eeltöö

Lavaani mooduli funktsioonidega tutvumiseks teeme alustuseks suhteliselt lihtsa kinnitava **faktoranalüüsi mudeli**, milles üheksa vaadeldud muutuja abil defineerime 3 latentset muutujat (faktorit). Antud juhul on tegemist suhteliselt lihtsa mudeliga, aga tihti võivad mudelid minna üsna keeruliseks. Sellest tulenevalt on reeglina mõistlik enne andmete analüüsimise kallale asumist teha mudelisse kaasatavate muutujate vaheliste seoste kohta skeem (nt paberi peal). Praegusel juhul näeks mudeli diagramm välja järgnev.



Joonis 1.

Jooniselt on näha, et vaadeldud muutujad x1-x3 mõõdavad latentset muutujat nimega *visual*, x4-x6 latentset muutujat nimega *textual* ja x7-x9 latentset muutujat nimega *speed*. Vastavalt tavale tähistavad SEM-i diagrammidel nelinurgad vaadeldud muutujaid ja ringid või ellipsid latentseid muutujaid. Latentsete muutujate vahelised kaared tähistavad nendevahelisi korrelatsioone. Väikesed noolekesed muutujate juures tähistavad dispersioone, mis samuti kuuluvad mudeli poolt välja arvutatavate parameetrite hulka. Vaadeldud muutujate puhul on tegemist jääkdispersiooniga ehk selle osaga muutuja variatiivsusest, mis ei ole määratud latentse faktori poolt. Latentsetele muutujatele on vaja anda ka skaala ja selleks fikseeritakse iga faktori puhul ühe sellele laaduva vaadeldud muutuja laadung. Seda tähistavad number 1-d mõne noole juures.

Mudeli diagramm aitab ka hinnata, kas analüüsi kaasatavas andmestikus on piisavalt infot, et arvutada välja kõik need mudeli parameetrid, mida meil vaja läheb. Jooniselt saame kokku lugeda, et meil on vaja hinnata 6 faktorlaadungit (9 vaadeldavat muutujat miinus need 3, mille laadungid on fikseeritud), 3 korrelatsiooni ja 12 dispersiooni. See teeb kokku 21 parameetrit. Maksimaalne parameetrite arv sõltub mudelisse kaasatavate vaadeldud muutujate arvust ja seda saab arvutada valemi p(p+1)/2 abil, milles "p" on mudelisse kaasatavate vaadeldud muutujate arv. Praegusel juhul anname mudelile ette 9 muutujat ja sellest tulenevalt maksimaalne vabade parameetrite arv mudelis oleks . Järelikult pole antud juhul meil probleemi sellega, et tahaksime mudelil lasta hinnata liiga suurt arvu vabu parameetreid.

SEM eeldab ka, et analüüsi kaasatavad muutujad oleksid normaaljaotusega ja et nad ei oleks liiga tugevalt seotud (ühte meetodi multikollineaarsuse kindlakstegemiseks vaatasime eelmises praktikumis). Lisaks sellele võib SEM-i puhul probleeme tekitada see, kui analüüsi kaasatavad muutujad on väga erinevatel numbrilistel skaaladel (nt mõned muutujad ühekohalised komakohtadega arvud ja teised kolmekohalised arvud). Erinevalt uurivast faktoranalüüsist, mis teostab arvutusi standardiseeritud korrelatsioonikordajate peal, kasutab SEM kovariatsioonikordajaid, mis sisaldavad infot ka muutujate skaala kohta. Väga erinevate suurusjärkude korral tasub korrutada või jagada mõned muutujad mingi koefitsiendiga, et kõik muutujad oleksid enam-vähem sarnastes suurusjärkudes. ühe- ja kolmekohaliste arvudest koosnevate muutujate näite puhul saab ühekohalised muutujad korrutada läbi 100-ga.

# Mudeli kirjeldamine

Järgnevalt peaksime mudeli viima R-ile ja lavaanile arusaadavale kujule. Võimalik, et teil on varasemast praktikumist meeles, kuidas käis mudelite kirjeldamine lineaarsete mudelite puhul funktsioonis *lm*. Lavaani mudeli kirjeldamise süntaks on sellele üsna sarnane.

**1.** regressioonseose tähistus

**2.** Latentse muutuja defineerimine (kus x1, x2 ja x3 on latentsele tunnusele laaduvad vaadeldud muutujad):

**3.** Muutujate-vahelise korrelatsiooni saame mudelisse kahekordse tilde abil:

Neid kolme tüüpi seoseid läheb mudelite kirjeldamisel kõige sagedamini vaja. Antud juhul näeks meie 3-faktoriline mudel välja järgnev:

mudel1 <- "  
visual =~ x1 + x2 + x3  
textual =~ x4 + x5 + x6  
speed =~ x7 + x8 + x9"

Mudeli kirjeldus peab paiknema jutumärkide (või ülakomade) vahel ja iga mudeli valem omaette real. Pikemate mudelite korral on vahel selguse huvides kasulik mudelit liigendada, jättes valemite vahele tühje ridu või kommentaare. Nagu mujalgi R-i koodis on kommentaarid "#" sümboliga algavad read.

# Mudeli väljund

Olles mudeli ära kirjeldanud saame sobitada mudeli oma andmetele kasutades funktsiooni *cfa* (nagu *confirmatory factor analysis*), millele anname ette mudeli kirjelduse ja argumendi *data* abil tabeli nime, millest vastavaid muutujaid otsida.

fit1 <- cfa(mudel1, data=HolzingerSwineford1939)

Seejärel saame funktsiooni *summary* abil uurida mudeli väljundit. Lavaani mudelite puhul tasub enamasti funktsiooile *summary* anda veel 2 argumenti: *fit.measures=TRUE* annab meile lisaks hii-ruudule veel täiendavaid mudeli sobitusastme näitajaid, ning *standardized=TRUE* annab lisaks standardiseerimata mudeli koefitsientidele meile need ka standardiseeritud kujul. (Reeglina huvitavadki meid rohkem standardiseeritud kordajad.)

summary(fit1, fit.measures=TRUE, standardized=TRUE)

## lavaan (0.5-23.1097) converged normally after 35 iterations  
##   
## Number of observations 301  
##   
## Estimator ML  
## Minimum Function Test Statistic 85.306  
## Degrees of freedom 24  
## P-value (Chi-square) 0.000  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.931  
## Tucker-Lewis Index (TLI) 0.896  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3737.745  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 21  
## Akaike (AIC) 7517.490  
## Bayesian (BIC) 7595.339  
## Sample-size adjusted Bayesian (BIC) 7528.739  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.092  
## 90 Percent Confidence Interval 0.071 0.114  
## P-value RMSEA <= 0.05 0.001  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.065  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## visual =~   
## x1 1.000 0.900 0.772  
## x2 0.554 0.100 5.554 0.000 0.498 0.424  
## x3 0.729 0.109 6.685 0.000 0.656 0.581  
## textual =~   
## x4 1.000 0.990 0.852  
## x5 1.113 0.065 17.014 0.000 1.102 0.855  
## x6 0.926 0.055 16.703 0.000 0.917 0.838  
## speed =~   
## x7 1.000 0.619 0.570  
## x8 1.180 0.165 7.152 0.000 0.731 0.723  
## x9 1.082 0.151 7.155 0.000 0.670 0.665  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## visual ~~   
## textual 0.408 0.074 5.552 0.000 0.459 0.459  
## speed 0.262 0.056 4.660 0.000 0.471 0.471  
## textual ~~   
## speed 0.173 0.049 3.518 0.000 0.283 0.283  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## .x1 0.549 0.114 4.833 0.000 0.549 0.404  
## .x2 1.134 0.102 11.146 0.000 1.134 0.821  
## .x3 0.844 0.091 9.317 0.000 0.844 0.662  
## .x4 0.371 0.048 7.779 0.000 0.371 0.275  
## .x5 0.446 0.058 7.642 0.000 0.446 0.269  
## .x6 0.356 0.043 8.277 0.000 0.356 0.298  
## .x7 0.799 0.081 9.823 0.000 0.799 0.676  
## .x8 0.488 0.074 6.573 0.000 0.488 0.477  
## .x9 0.566 0.071 8.003 0.000 0.566 0.558  
## visual 0.809 0.145 5.564 0.000 1.000 1.000  
## textual 0.979 0.112 8.737 0.000 1.000 1.000  
## speed 0.384 0.086 4.451 0.000 1.000 1.000

Mudeli väljundi ülaosas on näha valimi suurus ning hii-ruut-statistik (*Minimum Function Test Statistic*) koos vabadusastmete arvu ja p-väärtusega. Hii-ruut testib antud juhul nn täpse sobitumise hüpoteesi, mille kohaselt ei erine mudeli põhjal arvutatud kovaritasioonimaatriks andmestiku põhjal arvutatud kovariatsioonimaatriksist. **Me tahame, et hii-ruut oleks võimalikult väike ja p-väärtus võimalikult suur.** Kui me tüüpiliselt tahame, et p-väärtus oleks alla 0.05-e, siis praegu tahame, et see oleks 0.05-st suurem. Hii-ruudu puhul tuuakse sageli välja, et see on tundlik valimi suuruse suhtes: suurte valimite korral kipub p-väärtus isegi heade mudelite korral jääma < 0.05. Arvamused lähevad lahku, kui suurt valimit lugeda antud juhul suureks: mõnede arvates esineb seda probleemi alles tuhandetesse küündivate valimite puhul, teiste arvates võib see hakkata rolli mängima kohe kui valimi suurus ületab paarisaja inimese piiri. üsna kindel on aga see, et paarisaja osalejaga valimite puhul see probleemiks olla ei tohiks. Praegusel juhul näitab hii-ruut, et mudel andmetele kuigi hästi ei sobi.

Järgnevalt on ära toodud täiendavad sobitusastme näitajad, mida me funktsioonilt *summary* tellisime argumendi *fit.measures abil*. Sobitusastme näitajad võib laias laastus jagada kahte rühma: lisanduva sobitusastme indeksid (*incrementalt indices*) ja absoluutse sobitusastme indeksid (*absolute fit indices*). Esimesed mõõdavad sobitusastme paranemist võrreldes sellise mudeliga, milles vaadeldud muutujate vahelised korrelatsioonid puuduvad. Teised mõõdavad, kui hästi suudab meie poolt koostatud mudel taastada tegelike andmete põhjal arvutatud kovariatsioonimaatriksi. Lisanduva sobitusastme indeksistest on väljundis toodud TLI (*Tucker-Lewis index*) ja CFI (*comparative fit index*). Nende puhul näitavad mudelis head sobitusastet väärtused 0.95 ja üles. Absoluutse sobitusastme indeksitest on ära toodud RMSEA (*root mean square error of approximation*) ja SRMR (*standardized root mean square residual*), mille puhul head sobitusastet näitavad väärtused 0.05 ja alla. RMSEA puhul on ära toodud ka 90% usalduspiirid. Hea mudeli korral oleks alumine piir alla 0.05 ja ülemine mitte üle 0.10. Ka kõigi nende näitajate põhjal võime öelda, et sellisel kujul antud mudel andmetega ei sobi.

Sobitusastme indeksite järel tuleb viimasena tabel mudeli parameetrite hinnangutega. Tabelis on kõigepealt toodud fakorlaadungid (alajaotus *Latent variables*), seejärel faktorite vahelised seosed (alajaotus *Covariance*) ja dispersioonid (*Variances*). Standardiseerimata parameetrid on tulbas nimega *Estimates* ja standardiseeritud parameetrid tulbas nimega *Std.all*. Lisaks sellele on tabelis toodud ka standardvead (*Std.err*), z-väärtused (mis on standardiseerimata parameeter jagatuna st.veaga) ning p-väärtused. (Tulbas *Std.lv* on parameetrite väärtused, siis kui standardiseerida ainult latentsed muutujad).

Nagu näha on muutujate x1, x4 ja x7 standardiseerimata laadung 1.000 ja mõned näitajad on neist ridadest puudu. See tuleb sellest, struktuurivõrrandite mudelite puhul tuleb latentsele muutjale anda skaala. Kõige sagedamini kasutatav võimalus selleks on anda mõne laaduva vaadeldud muutuja skaala fikseerides selle sandardiseerimata laadungi 1-ga. Lavaan on seda automaatselt teinud, fikseerides iga latentse muutuja puhul esimese indikaatori standardiseerimata laadungi. Kuidas parameetreid tõlgendada? Standardiseerimata laadungid on tõlgendatavad regressioonikoefitsientidena. Kui muutuja x2 standardiseerimata laadung faktorile visual on 0.554 siis võib faktori väärtuse 1-punktilise suurenemise korral oodata, et x2 suureneb 0.554 punkti võrra.

Ainult ühele faktorile laaduvate vaadeldud muutujate puhul võib standardiseeritud laadungeid tõlgendada faktori ja muutuja vahelise korrelatsioonikordajana. Nagu korrelatsioonikordaja puhul ikka saame seda ruutu võttes kindlaks teha kui suure osa tunnuse variatiivsusest seos ära määrab. Tunnuse x2 puhul on korrelatsioon faktoriga visual 0.424 ja faktor määrab ära 0.42422=0.179 ehk umbes 18 protsenti tunnuse x2 variatiivsusest. Muutujate puhul, mis laaduvad rohkem kui ühele faktorile (praeguses mudelis meil selliseid pole) ei saa standardiseeritud laadungeid tõlgendada korrelatsioonikordajatena vaid standardiseeritud regressioonikordajatena, mille puhul hoitakse korreleeritud faktoreid kontrolli all. Neid ruutu võttes ei saa kindlaks teha seletatava variatiivsuse osakaalu. Alajaotuses *Covariances* olevad standardiseerimata parameetrid kujutavad endas faktorite-vahelisi kovariatsioonikordjaid ja standardiseeritud parameetrid korrelatsioonikordajaid. Antud mudeli puhul võib näiteks näha et faktori visual korrelatsioon faktoriga textual on 0.459 ja faktoriga speed 0.471. Vaadeldud muutujate puhul näitavad dispersioonid (*Variances*) standardiseeritud kujul seda osa tunnuse variatiivsusest, mida faktor ei seleta. Nt tunnuse x2 puhul ei suuda faktor visual seletada tervelt 82.1 protsenti variatiivsusest. Parameetrite tabeli puhul tasub ka kontrollida, et ei esineks negatiivseid jääkdispersioone ja standardiseeritud parameetreid absoluutväärtusega > 1. Selliseid väärtusi ei ole võimalik tõlgendada ja nende esinemine annab märku probleemidest mudeli või andmetega (nt liiga väike valim või liiga tugevalt korreleeritud muutujad). Mudeli väljundit saab tellida ka osade kaupa. Seda võib vaja minna näiteks siis kui soovime teada saada mõnda sellist sobitusastme näitjat, mida funktsioon summary välja ei trükkinud. Nt üks suhteliselt sageli raporteeritav sobitusastme näitaja on GFI (*goodness of fit index*, hea väärtus 0.95 ja üles). Selle ja suure hulga muid sobitusindekseid saame kätte funktsiooni fitMeasures abil:

fitMeasures(fit1)

## npar fmin chisq   
## 21.000 0.142 85.306   
## df pvalue baseline.chisq   
## 24.000 0.000 918.852   
## baseline.df baseline.pvalue cfi   
## 36.000 0.000 0.931   
## tli nnfi rfi   
## 0.896 0.896 0.861   
## nfi pnfi ifi   
## 0.907 0.605 0.931   
## rni logl unrestricted.logl   
## 0.931 -3737.745 -3695.092   
## aic bic ntotal   
## 7517.490 7595.339 301.000   
## bic2 rmsea rmsea.ci.lower   
## 7528.739 0.092 0.071   
## rmsea.ci.upper rmsea.pvalue rmr   
## 0.114 0.001 0.082   
## rmr\_nomean srmr srmr\_bentler   
## 0.082 0.065 0.065   
## srmr\_bentler\_nomean srmr\_bollen srmr\_bollen\_nomean   
## 0.065 0.065 0.065   
## srmr\_mplus srmr\_mplus\_nomean cn\_05   
## 0.065 0.065 129.490   
## cn\_01 gfi agfi   
## 152.654 0.943 0.894   
## pgfi mfi ecvi   
## 0.503 0.903 0.423

Aga eraldi saab tellida ka standardiseerimata parameetrite hinnanguid...

parameterEstimates(fit1)

## lhs op rhs est se z pvalue ci.lower ci.upper  
## 1 visual =~ x1 1.000 0.000 NA NA 1.000 1.000  
## 2 visual =~ x2 0.554 0.100 5.554 0 0.358 0.749  
## 3 visual =~ x3 0.729 0.109 6.685 0 0.516 0.943  
## 4 textual =~ x4 1.000 0.000 NA NA 1.000 1.000  
## 5 textual =~ x5 1.113 0.065 17.014 0 0.985 1.241  
## 6 textual =~ x6 0.926 0.055 16.703 0 0.817 1.035  
## 7 speed =~ x7 1.000 0.000 NA NA 1.000 1.000  
## 8 speed =~ x8 1.180 0.165 7.152 0 0.857 1.503  
## 9 speed =~ x9 1.082 0.151 7.155 0 0.785 1.378  
## 10 x1 ~~ x1 0.549 0.114 4.833 0 0.326 0.772  
## 11 x2 ~~ x2 1.134 0.102 11.146 0 0.934 1.333  
## 12 x3 ~~ x3 0.844 0.091 9.317 0 0.667 1.022  
## 13 x4 ~~ x4 0.371 0.048 7.779 0 0.278 0.465  
## 14 x5 ~~ x5 0.446 0.058 7.642 0 0.332 0.561  
## 15 x6 ~~ x6 0.356 0.043 8.277 0 0.272 0.441  
## 16 x7 ~~ x7 0.799 0.081 9.823 0 0.640 0.959  
## 17 x8 ~~ x8 0.488 0.074 6.573 0 0.342 0.633  
## 18 x9 ~~ x9 0.566 0.071 8.003 0 0.427 0.705  
## 19 visual ~~ visual 0.809 0.145 5.564 0 0.524 1.094  
## 20 textual ~~ textual 0.979 0.112 8.737 0 0.760 1.199  
## 21 speed ~~ speed 0.384 0.086 4.451 0 0.215 0.553  
## 22 visual ~~ textual 0.408 0.074 5.552 0 0.264 0.552  
## 23 visual ~~ speed 0.262 0.056 4.660 0 0.152 0.373  
## 24 textual ~~ speed 0.173 0.049 3.518 0 0.077 0.270

... ja standardiseeritud hinnanguid.

standardizedSolution(fit1)

## lhs op rhs est.std se z pvalue  
## 1 visual =~ x1 0.772 0.055 14.041 0  
## 2 visual =~ x2 0.424 0.060 7.105 0  
## 3 visual =~ x3 0.581 0.055 10.539 0  
## 4 textual =~ x4 0.852 0.023 37.776 0  
## 5 textual =~ x5 0.855 0.022 38.273 0  
## 6 textual =~ x6 0.838 0.023 35.881 0  
## 7 speed =~ x7 0.570 0.053 10.714 0  
## 8 speed =~ x8 0.723 0.051 14.309 0  
## 9 speed =~ x9 0.665 0.051 13.015 0  
## 10 x1 ~~ x1 0.404 0.085 4.763 0  
## 11 x2 ~~ x2 0.821 0.051 16.246 0  
## 12 x3 ~~ x3 0.662 0.064 10.334 0  
## 13 x4 ~~ x4 0.275 0.038 7.157 0  
## 14 x5 ~~ x5 0.269 0.038 7.037 0  
## 15 x6 ~~ x6 0.298 0.039 7.606 0  
## 16 x7 ~~ x7 0.676 0.061 11.160 0  
## 17 x8 ~~ x8 0.477 0.073 6.531 0  
## 18 x9 ~~ x9 0.558 0.068 8.208 0  
## 19 visual ~~ visual 1.000 0.000 NA NA  
## 20 textual ~~ textual 1.000 0.000 NA NA  
## 21 speed ~~ speed 1.000 0.000 NA NA  
## 22 visual ~~ textual 0.459 0.064 7.189 0  
## 23 visual ~~ speed 0.471 0.073 6.461 0  
## 24 textual ~~ speed 0.283 0.069 4.117 0

Kui mudel andmetega ei sobi, siis üks võimalus probleeme diagnoosida, on vaadata mudeli jääkide maatriksit. Jäägid kujutavad endast erinevusi tegelike andmete põhjal arvutatud ja mudeli põhjal taastatud kovariatsioonikordajate vahel. Jääke, mida struktuurivõrrandite mudelite puhul uurida saab, on erinevat tüüpi. Vaatame selliseid, mida nimetatakse **korrelatsioonijääkideks** (ehk siis kovariatsioonikordajad arvutatakse ümber korrelatsioonideks, mida on lihtsam tõlgendada). Korrelatsioonijäägid saame funktsiooni *residuals* abil, määrates argumendi *type* väärtuseks "cor". Tabelist peaksime üles otsima jääkkorrelatsioonid absoluutväärtusega > 0.10. Need nii-öelda suured jäägid näitavad meile, milliste muutujate vaheliste seoste seletamisega mudel väga hästi hakkama ei saa.

residuals(fit1, type="cor")

## $type  
## [1] "cor.bollen"  
##   
## $cor  
## x1 x2 x3 x4 x5 x6 x7 x8 x9   
## x1 0.000   
## x2 -0.030 0.000   
## x3 -0.008 0.094 0.000   
## x4 0.071 -0.012 -0.068 0.000   
## x5 -0.009 -0.027 -0.151 0.005 0.000   
## x6 0.060 0.030 -0.026 -0.009 0.003 0.000   
## x7 -0.140 -0.189 -0.084 0.037 -0.036 -0.014 0.000   
## x8 -0.039 -0.052 -0.012 -0.067 -0.036 -0.022 0.075 0.000   
## x9 0.149 0.073 0.147 0.048 0.067 0.056 -0.038 -0.032 0.000  
##   
## $mean  
## x1 x2 x3 x4 x5 x6 x7 x8 x9   
## 0 0 0 0 0 0 0 0 0

Antud juhul näeme, et muutuja x3 omab suurt jääkkorrelatsiooni muutujaga x5, muutuja x7 muutujatega x1 ja x2 ning muutuja x9 muutujatega x1 ja x3. Kui muutuja omab suuri jääk korrelatsioone mõnele teisele faktorile laaduvate muutujatega, on võimalik, et muutuja mõõdab rohkem kui ühte konstrukti ja tal tuleks lubada laaduda rohkem kui ühele faktorile. Teine võimalus on, et muutujate vahelisi seoseid seoseid ei põhjusta mitte ainult mudeli faktor(id), vaid ka mingi muu tegur (me ei pruugi isegi teada mis täpselt). Sellise võimaluse arvesse võtmiseks saab mudelis määrata mõnede muutujate jääkdispersioonid omavahel korreleeruma. Teeme proovi mõttes uue mudeli, milles määrame ka muutujatevahelised korrelatsioonid, mida faktorid ei seleta.

mudel2 <- "  
# latentsed muutujad  
visual =~ x1 + x2 + x3  
textual =~ x4 + x5 + x6  
speed =~ x7 + x8 + x9  
# jääkdispersioonide vahelised korrelatsioonid  
x7 ~~ x1 + x2  
x9 ~~ x1 + x3  
x3 ~~ x5"

Vaadake selle mudeli väljundit. Kas sobitusastme indeksid läksid paremaks või mitte?

fit2 <- cfa(mudel2, data=HolzingerSwineford1939)  
summary(fit2, fit.measures=TRUE, standardized=TRUE)

## lavaan (0.5-23.1097) converged normally after 36 iterations  
##   
## Number of observations 301  
##   
## Estimator ML  
## Minimum Function Test Statistic 40.161  
## Degrees of freedom 19  
## P-value (Chi-square) 0.003  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.976  
## Tucker-Lewis Index (TLI) 0.955  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3715.173  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 26  
## Akaike (AIC) 7482.346  
## Bayesian (BIC) 7578.730  
## Sample-size adjusted Bayesian (BIC) 7496.273  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.061  
## 90 Percent Confidence Interval 0.034 0.087  
## P-value RMSEA <= 0.05 0.227  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.043  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## visual =~   
## x1 1.000 0.859 0.747  
## x2 0.562 0.108 5.217 0.000 0.483 0.412  
## x3 0.710 0.118 6.039 0.000 0.610 0.546  
## textual =~   
## x4 1.000 0.986 0.849  
## x5 1.105 0.065 16.916 0.000 1.090 0.852  
## x6 0.931 0.056 16.763 0.000 0.918 0.840  
## speed =~   
## x7 1.000 0.680 0.623  
## x8 1.094 0.150 7.295 0.000 0.744 0.736  
## x9 0.913 0.126 7.265 0.000 0.621 0.616  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## .x1 ~~   
## .x7 -0.148 0.058 -2.546 0.011 -0.148 -0.226  
## .x2 ~~   
## .x7 -0.241 0.063 -3.800 0.000 -0.241 -0.265  
## .x1 ~~   
## .x9 0.173 0.059 2.941 0.003 0.173 0.285  
## .x3 ~~   
## .x9 0.189 0.055 3.455 0.001 0.189 0.254  
## .x5 -0.135 0.047 -2.911 0.004 -0.135 -0.216  
## visual ~~   
## textual 0.398 0.071 5.638 0.000 0.470 0.470  
## speed 0.257 0.059 4.376 0.000 0.440 0.440  
## textual ~~   
## speed 0.187 0.053 3.526 0.000 0.278 0.278  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## .x1 0.584 0.117 4.980 0.000 0.584 0.442  
## .x2 1.139 0.102 11.115 0.000 1.139 0.830  
## .x3 0.874 0.091 9.622 0.000 0.874 0.702  
## .x4 0.378 0.048 7.938 0.000 0.378 0.280  
## .x5 0.449 0.058 7.715 0.000 0.449 0.274  
## .x6 0.353 0.043 8.273 0.000 0.353 0.295  
## .x7 0.728 0.083 8.740 0.000 0.728 0.612  
## .x8 0.469 0.075 6.219 0.000 0.469 0.459  
## .x9 0.631 0.069 9.162 0.000 0.631 0.621  
## visual 0.738 0.144 5.125 0.000 1.000 1.000  
## textual 0.973 0.112 8.707 0.000 1.000 1.000  
## speed 0.462 0.096 4.830 0.000 1.000 1.000

# Parameetrite hindamise (arvutamise) meetoditest

Funktsiooni *cfa* vaikemeetodiks on ML(*maximum likelyhood*). ML arvutusmeetodi eeldusteks on: tunnuste mitmemõõtmeline normaaljaotus (iga komponentidelineaarkombinatsioon on normaaljaotusega), tunnused vähemalt intervall skaalal.

ML-i analoog on GLS (*generalized least squares*), mis on arvutuslikult lihtsam, headuse näitajad tulevad samad, eriti kui n on suur. ML ja GLS sobivad ainult mitmemõõtmelise normaaljaotuse korral, või juhtudel, kui kõrvalekalded on suhteliselt väikesed.

Mittenormaaljaotuslike andmete korral kaks enim kasutatavat parameetrite arvutamise meetodit: RML(*Robust ML*) ja WLS (*weighted least squares*) ei ole soovitatav väga suurte valimite korral. Eelistada pigem just väikeste valimite korral(vt. leongu konspekti).

Me saame muuta meetodi valikut argumendiga *estimator*:

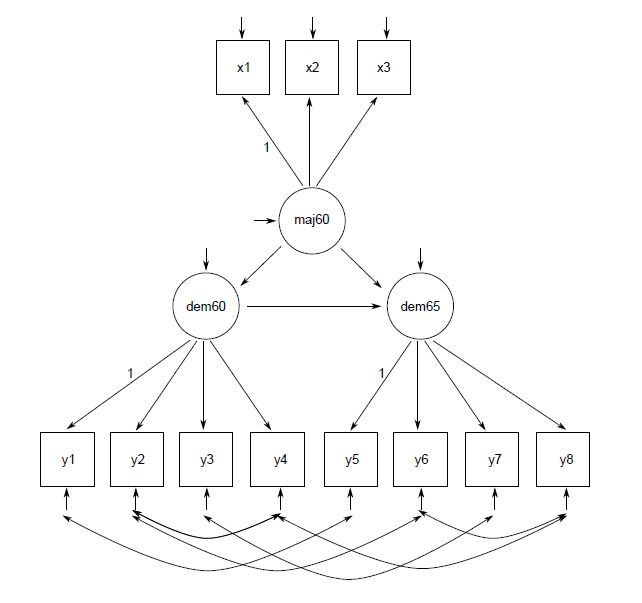
fit2 <- cfa(mudel1, data=HolzingerSwineford1939, estimator = "GLS")  
fit3 <- cfa(mudel1, data=HolzingerSwineford1939, estimator = "WLS")

# Näide 2: regressiooniseostega mudeli rakendamine

Proovime teha ka sellist mudelit, mis sisaldab lisaks latentsetele muutujatele ka regressioonseoseid. Selleks vaatame teist lavaaniga kaasa tulevat näidisandmestikku nimega *PoliticalDemocracy*. See tabel sisaldab 75 riigi kohta 11 poliitilist ja majanduslikku arengut puudutavat näitajat. Muutujad y1 kuni y4 on vastavalt ajakirjandusvabaduse tase, vaba politilise opositsiooni olemasolu, vabade valimiste toimumine ja valitud parlamendi efektiivsus mõõdetuna aastal 1960. Muutujad y5 kuni y8 on samad näitajad mõõdetuna aastal 1965. Muutujad x1 kuni x3 on vastavalt riigi sisemajanduse koguprodukt, energiatarbimine elaniku kohta ja tööstussektoris töötava tööjõu osakaal mõõdetuna aastal 1960. Andmestiku tutvustus avaneb alljärgneva käsuga:

help(PoliticalDemocracy)

Üritame andmetele sobitada alljärgneval joonisel kujutatud mudelit. Defineerime kolm latentset muutujat. Tunnused x1 kuni x3 mõõdavad latentset muutujat *maj60* (st majanduslik arengutase aastal 1960), tunnused *y1* kuni *y4* mõõdavad latentset muutujat *dem60* (st demokraatia aastal 1960) ning tunnused *y5* kuni *y8* mõõdavat latentset muutujat *dem65* (st demokraatia aastal 1965). Latentsete muutujate vahel on regressioonseosed. Demokraatiat aastal 1960 mõjutab majanduse arengutase aastal 1960. Demokraatiat aastal 1965 mõjutavad nii majandus kui demokraatia aastal 1960. Latentsed muutujad ei suuda antud juhul seletada kõiki demokraatia indikaatorite vahelisi seoseid. Seetõttu on mudelisse lisatud ka nende jääkdispersioonide vahelised korrelatsioonid. Muutuja *y1* korreleerub *y5*-ga, *y2* korreleerub *y4*-ga ja *y6*-ga, *y3* korreleerub *y7*-ga, *y8* korreleerub *y4*-ga ja *y6*-ga. Mudelisse lisatud 11 vaadeldud muutujat võimaldavad hinnata vaba parameetrit. Meil on vaja hinnata 31 parameetrit: 8 faktorlaadungit, 14 dispersiooni, 3 regressiooni- ja 6 korrelatsioonikordajat. (Nagu esimese näite puhul fikseeritakse ka praegu iga latentse muutuja puhul esimese indikaatori standardiseerimata laadung, selleks et omistada latentsele muutujale skaala. Seetõttu ei olegi meil vaja hinnata 11 vaid ainult 8 faktorlaadungit). Järelikult infot on piisavalt. Lavaanis näeks selline mudel välja järgnevalt:



Joonis 2.

mudel3 <- "  
# defineerime latentsed tunnused  
maj60 =~ x1 + x2 + x3  
dem60 =~ y1 + y2 + y3 + y4  
dem65 =~ y5 + y6 + y7 + y8  
  
# regressioonseosed latentsete tunnuste vahel  
dem60 ~ maj60  
dem65 ~ maj60 + dem60  
# jääkdispersioonide vahelised korrelatsioonid  
y1 ~~ y5  
y2 ~~ y4 + y6  
y3 ~~ y7  
y4 ~~ y8  
y6 ~~ y8"

Sobitame mudeli andmetele ja uurime väljundit. Kuna mudel sisaldab ka regressioonseoseid, siis kasutame sobitamiseks funktsiooni sem. (Kuigi tegelikult teevad vähemalt praeguse seisuga funktsioonid *cfa* ja sem sama asja, siis tulevikus võib see olukord muutuda.) Kas mudel sobib andmetega?

fit3 <- sem(mudel3, data=PoliticalDemocracy)

## Warning in lav\_data\_full(data = data, group = group, cluster = cluster, :  
## lavaan WARNING: unordered factor(s) with more than 2 levels detected in  
## data: x1 x2 x3 y1 y2 y3 y4 y5 y6 y7 y8

summary(fit3, fit.measures=TRUE, standardized=TRUE)

## lavaan (0.5-23.1097) converged normally after 223 iterations  
##   
## Number of observations 75  
##   
## Estimator ML  
## Minimum Function Test Statistic 34.580  
## Degrees of freedom 35  
## P-value (Chi-square) 0.488  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 542.972  
## Degrees of freedom 55  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 1.000  
## Tucker-Lewis Index (TLI) 1.001  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) NA  
## Loglikelihood unrestricted model (H1) NA  
##   
## Number of free parameters 31  
## Akaike (AIC) NA  
## Bayesian (BIC) NA  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.000  
## 90 Percent Confidence Interval 0.000 0.082  
## P-value RMSEA <= 0.05 0.751  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.049  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## maj60 =~   
## x1 1.000 18.271 0.912  
## x2 1.079 0.068 15.983 0.000 19.720 0.987  
## x3 0.801 0.067 11.923 0.000 14.632 0.873  
## dem60 =~   
## y1 1.000 5.168 0.635  
## y2 0.770 0.164 4.689 0.000 3.978 0.700  
## y3 0.772 0.167 4.617 0.000 3.988 0.720  
## y4 0.613 0.151 4.066 0.000 3.170 0.587  
## dem65 =~   
## y5 1.000 3.042 0.725  
## y6 1.115 0.207 5.392 0.000 3.391 0.675  
## y7 1.750 0.277 6.313 0.000 5.323 0.840  
## y8 1.133 0.249 4.548 0.000 3.446 0.571  
##   
## Regressions:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## dem60 ~   
## maj60 0.090 0.039 2.316 0.021 0.320 0.320  
## dem65 ~   
## maj60 0.039 0.015 2.597 0.009 0.231 0.231  
## dem60 0.490 0.092 5.305 0.000 0.832 0.832  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## .y1 ~~   
## .y5 6.348 2.762 2.298 0.022 6.348 0.349  
## .y2 ~~   
## .y4 2.687 2.217 1.212 0.226 2.687 0.151  
## .y6 7.515 2.297 3.272 0.001 7.515 0.499  
## .y3 ~~   
## .y7 1.664 2.589 0.643 0.520 1.664 0.126  
## .y4 ~~   
## .y8 6.355 2.799 2.270 0.023 6.355 0.293  
## .y6 ~~   
## .y8 1.882 2.011 0.936 0.349 1.882 0.102  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## .x1 67.509 15.010 4.498 0.000 67.509 0.168  
## .x2 10.704 11.920 0.898 0.369 10.704 0.027  
## .x3 66.654 12.724 5.238 0.000 66.654 0.237  
## .y1 39.621 7.758 5.107 0.000 39.621 0.597  
## .y2 16.489 3.517 4.689 0.000 16.489 0.510  
## .y3 14.787 3.484 4.244 0.000 14.787 0.482  
## .y4 19.146 3.631 5.272 0.000 19.146 0.656  
## .y5 8.330 1.708 4.876 0.000 8.330 0.474  
## .y6 13.736 2.631 5.221 0.000 13.736 0.544  
## .y7 11.801 3.499 3.373 0.001 11.801 0.294  
## .y8 24.606 4.399 5.594 0.000 24.606 0.674  
## maj60 333.828 65.403 5.104 0.000 1.000 1.000  
## .dem60 23.977 8.735 2.745 0.006 0.898 0.898  
## .dem65 1.216 0.795 1.530 0.126 0.131 0.131

# Modifikatsiooniindeksid

Modifikatsiooniindeksid näitavad, kui palju väheneks mudeli hii-ruut statistik mingi vaba parameetri mudelisse lisamisel. (Mida väiksem hiiruut, seda paremini mudel andmetega sobib.) Hoiatusena tuleks märkida, et indeksite puhul on tegemist umbkaudsete oletustega. Statistikaprogramm ei proovi kõigi nende parameetrite mudelisse lisamist ka tegelikult läbi, vaid kasutab nende hindamiseks mingit suhteliselt jämedat rusikareeglit.

Defineerime sama 3-faktorilise mudeli:

HS.model <- "visual =~ x1 + x2 + x3  
textual =~ x4 + x5 + x6  
speed =~ x7 + x8 + x9"

Sobitame mudeli andmetele ja vaatame mudeli väljundit.

fit1 <- cfa(HS.model, data=HolzingerSwineford1939)  
summary(fit1, fit.measures=TRUE)

## lavaan (0.5-23.1097) converged normally after 35 iterations  
##   
## Number of observations 301  
##   
## Estimator ML  
## Minimum Function Test Statistic 85.306  
## Degrees of freedom 24  
## P-value (Chi-square) 0.000  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.931  
## Tucker-Lewis Index (TLI) 0.896  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3737.745  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 21  
## Akaike (AIC) 7517.490  
## Bayesian (BIC) 7595.339  
## Sample-size adjusted Bayesian (BIC) 7528.739  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.092  
## 90 Percent Confidence Interval 0.071 0.114  
## P-value RMSEA <= 0.05 0.001  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.065  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|)  
## visual =~   
## x1 1.000   
## x2 0.554 0.100 5.554 0.000  
## x3 0.729 0.109 6.685 0.000  
## textual =~   
## x4 1.000   
## x5 1.113 0.065 17.014 0.000  
## x6 0.926 0.055 16.703 0.000  
## speed =~   
## x7 1.000   
## x8 1.180 0.165 7.152 0.000  
## x9 1.082 0.151 7.155 0.000  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|)  
## visual ~~   
## textual 0.408 0.074 5.552 0.000  
## speed 0.262 0.056 4.660 0.000  
## textual ~~   
## speed 0.173 0.049 3.518 0.000  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|)  
## .x1 0.549 0.114 4.833 0.000  
## .x2 1.134 0.102 11.146 0.000  
## .x3 0.844 0.091 9.317 0.000  
## .x4 0.371 0.048 7.779 0.000  
## .x5 0.446 0.058 7.642 0.000  
## .x6 0.356 0.043 8.277 0.000  
## .x7 0.799 0.081 9.823 0.000  
## .x8 0.488 0.074 6.573 0.000  
## .x9 0.566 0.071 8.003 0.000  
## visual 0.809 0.145 5.564 0.000  
## textual 0.979 0.112 8.737 0.000  
## speed 0.384 0.086 4.451 0.000

Nagu näha, ei sobi mudel andmetega väga hästi. Hii-ruudu p-väärtus on alla 0.05-e, TLI ja CFI on alla 0.95-e ning RMSEA ja SRMR on üle 0.05-e. Järgnevalt vaatame mudeli modifikatsiooniindeksite tabelit ja otsime sellest üles kõige suurema indeksiga parameetri.

modindices(fit1)

## lhs op rhs mi epc sepc.lv sepc.all sepc.nox  
## 25 visual =~ x4 1.211 0.077 0.069 0.059 0.059  
## 26 visual =~ x5 7.441 -0.210 -0.189 -0.147 -0.147  
## 27 visual =~ x6 2.843 0.111 0.100 0.092 0.092  
## 28 visual =~ x7 18.631 -0.422 -0.380 -0.349 -0.349  
## 29 visual =~ x8 4.295 -0.210 -0.189 -0.187 -0.187  
## 30 visual =~ x9 36.411 0.577 0.519 0.515 0.515  
## 31 textual =~ x1 8.903 0.350 0.347 0.297 0.297  
## 32 textual =~ x2 0.017 -0.011 -0.011 -0.010 -0.010  
## 33 textual =~ x3 9.151 -0.272 -0.269 -0.238 -0.238  
## 34 textual =~ x7 0.098 -0.021 -0.021 -0.019 -0.019  
## 35 textual =~ x8 3.359 -0.121 -0.120 -0.118 -0.118  
## 36 textual =~ x9 4.796 0.138 0.137 0.136 0.136  
## 37 speed =~ x1 0.014 0.024 0.015 0.013 0.013  
## 38 speed =~ x2 1.580 -0.198 -0.123 -0.105 -0.105  
## 39 speed =~ x3 0.716 0.136 0.084 0.075 0.075  
## 40 speed =~ x4 0.003 -0.005 -0.003 -0.003 -0.003  
## 41 speed =~ x5 0.201 -0.044 -0.027 -0.021 -0.021  
## 42 speed =~ x6 0.273 0.044 0.027 0.025 0.025  
## 43 x1 ~~ x2 3.606 -0.184 -0.184 -0.134 -0.134  
## 44 x1 ~~ x3 0.935 -0.139 -0.139 -0.105 -0.105  
## 45 x1 ~~ x4 3.554 0.078 0.078 0.058 0.058  
## 46 x1 ~~ x5 0.522 -0.033 -0.033 -0.022 -0.022  
## 47 x1 ~~ x6 0.048 0.009 0.009 0.007 0.007  
## 48 x1 ~~ x7 5.420 -0.129 -0.129 -0.102 -0.102  
## 49 x1 ~~ x8 0.634 -0.041 -0.041 -0.035 -0.035  
## 50 x1 ~~ x9 7.335 0.138 0.138 0.117 0.117  
## 51 x2 ~~ x3 8.532 0.218 0.218 0.164 0.164  
## 52 x2 ~~ x4 0.534 -0.034 -0.034 -0.025 -0.025  
## 53 x2 ~~ x5 0.023 -0.008 -0.008 -0.005 -0.005  
## 54 x2 ~~ x6 0.785 0.039 0.039 0.031 0.031  
## 55 x2 ~~ x7 8.918 -0.183 -0.183 -0.143 -0.143  
## 56 x2 ~~ x8 0.054 -0.012 -0.012 -0.010 -0.010  
## 57 x2 ~~ x9 1.895 0.075 0.075 0.063 0.063  
## 58 x3 ~~ x4 0.142 -0.016 -0.016 -0.012 -0.012  
## 59 x3 ~~ x5 7.858 -0.130 -0.130 -0.089 -0.089  
## 60 x3 ~~ x6 1.855 0.055 0.055 0.044 0.044  
## 61 x3 ~~ x7 0.638 -0.044 -0.044 -0.036 -0.036  
## 62 x3 ~~ x8 0.059 -0.012 -0.012 -0.011 -0.011  
## 63 x3 ~~ x9 4.126 0.102 0.102 0.089 0.089  
## 64 x4 ~~ x5 2.534 0.186 0.186 0.124 0.124  
## 65 x4 ~~ x6 6.220 -0.235 -0.235 -0.185 -0.185  
## 66 x4 ~~ x7 5.920 0.098 0.098 0.078 0.078  
## 67 x4 ~~ x8 3.805 -0.069 -0.069 -0.059 -0.059  
## 68 x4 ~~ x9 0.196 -0.016 -0.016 -0.014 -0.014  
## 69 x5 ~~ x6 0.916 0.101 0.101 0.072 0.072  
## 70 x5 ~~ x7 1.233 -0.049 -0.049 -0.035 -0.035  
## 71 x5 ~~ x8 0.347 0.023 0.023 0.018 0.018  
## 72 x5 ~~ x9 0.999 0.040 0.040 0.031 0.031  
## 73 x6 ~~ x7 0.259 -0.020 -0.020 -0.017 -0.017  
## 74 x6 ~~ x8 0.275 0.018 0.018 0.016 0.016  
## 75 x6 ~~ x9 0.097 -0.011 -0.011 -0.010 -0.010  
## 76 x7 ~~ x8 34.145 0.536 0.536 0.488 0.488  
## 77 x7 ~~ x9 5.183 -0.187 -0.187 -0.170 -0.170  
## 78 x8 ~~ x9 14.946 -0.423 -0.423 -0.415 -0.415

**Modifikatsiooniindeksite** tabelis näitavad 3 esimest tulpa lavaani mudeli defineerimise tähistust kasutades, millise parameetri kohta indeks käib.  
**Tulbas lhs** (sõnadest left hand side) on parameetri kohta käiva valemi vasakpoolne muutuja ja tulbas rhs (*right hand side*) on sama valemi parempoolne muutuja.  
**Tulp op** (operator) näitab meile, millist tüüpi seosega on tegu: = on latentsele muutujale laadumine ning puhul on tegemist muutujatevahelise korrelatsiooniga.  
Parameetri modifikatsiooniindeksi väärtus on toodud **tulbas mi**. Antud juhul puudutavad tabeli 24 esimest rida seoseid, mis on mudelis juba vabade parameetritena olemas ja pole seetõttu väga huvitavad. Aga 25. real olev visual = x4 ja tulbas mi olev väärtus 1.211 tähendavad seda, et kui me lubaksime muutujal x4 laaduda ka faktorile *visual*, võiksime oodata, et mudeli hii-ruut väheneb umbes 1.2 võrra.  
Lisaks modifikatsiooniindeksile arvutatakse välja ka **parameetri oodatav väärtus** (tulp epc, *expected parameter change*) ja selle standardiseeritud väärtus (meid huvitab eelkõige tulp sepc.all). Mudelisse tasub vabade parameetritena lisade eelkõige neid, mille puhul on suured nii *mi* kui *epc* väärtused. Modifikatsiooniindeksite puhul loetakse statistiliselt oluliseks väärtusi üle 3.84. Väiksema indeksi väärtusega parameetrite mudelisse lisamine üldiselt ära ei tasu. Indeksid ei ole üksteisest sõltumatud. See tähendab, et kui tabelis on mitu olulist indeksit, siis ei saa me neid korraga mudelisse lisada ja oodata, et hii-ruut kahaneb nende indeksite summa võrra. Kui me seda teeksime, võivad erinevused oodatavate ja tegelike muutuste vahel olla väga suured. Seega tuleks indeksite poolt soovitatavaid parameetreid mudelisse lisada ühekaupa. Otsime tabelist üles kõige suurema modifikatsiooniindeksiga parameetri ja lisame selle mudelisse. Praegusel juhul on kõige suurema indeksiga parameeter 30. real, mis ütleb, et kui lubaksime muutujal x9 laaduda ka faktorile visual võiks oodata et hii-ruut kahaneb umbes 36 võrra. Teeme uue mudeli, milles see muutus on sisse viidud, sobitame andmetele ja vaatame väljundit.

HS.model2 <- "visual =~ x1 + x2 + x3 + x9  
textual =~ x4 + x5 + x6  
speed =~ x7 + x8 + x9"  
fit2 <- cfa(HS.model2, data=HolzingerSwineford1939)  
summary(fit2, fit.measures=TRUE)

## lavaan (0.5-23.1097) converged normally after 34 iterations  
##   
## Number of observations 301  
##   
## Estimator ML  
## Minimum Function Test Statistic 52.382  
## Degrees of freedom 23  
## P-value (Chi-square) 0.000  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.967  
## Tucker-Lewis Index (TLI) 0.948  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3721.283  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 22  
## Akaike (AIC) 7486.566  
## Bayesian (BIC) 7568.123  
## Sample-size adjusted Bayesian (BIC) 7498.351  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.065  
## 90 Percent Confidence Interval 0.042 0.089  
## P-value RMSEA <= 0.05 0.133  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.045  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|)  
## visual =~   
## x1 1.000   
## x2 0.578 0.098 5.918 0.000  
## x3 0.754 0.103 7.291 0.000  
## x9 0.437 0.081 5.367 0.000  
## textual =~   
## x4 1.000   
## x5 1.115 0.066 17.016 0.000  
## x6 0.926 0.056 16.685 0.000  
## speed =~   
## x7 1.000   
## x8 1.207 0.185 6.540 0.000  
## x9 0.675 0.112 6.037 0.000  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|)  
## visual ~~   
## textual 0.396 0.072 5.506 0.000  
## speed 0.177 0.055 3.239 0.001  
## textual ~~   
## speed 0.136 0.051 2.675 0.007  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|)  
## .x1 0.576 0.100 5.731 0.000  
## .x2 1.120 0.100 11.153 0.000  
## .x3 0.830 0.087 9.515 0.000  
## .x9 0.558 0.060 9.336 0.000  
## .x4 0.373 0.048 7.800 0.000  
## .x5 0.444 0.058 7.602 0.000  
## .x6 0.357 0.043 8.285 0.000  
## .x7 0.740 0.086 8.595 0.000  
## .x8 0.375 0.094 3.973 0.000  
## visual 0.783 0.134 5.842 0.000  
## textual 0.978 0.112 8.728 0.000  
## speed 0.444 0.097 4.567 0.000

Selle mudeli sobitusastme näitajad on läinud natuke paremaks. CFI ja SRMR on isegi ületanud selle väärtuse piiri, mida võib pidada heaks, teised näitajad aga mitte. Nüüd vaatame selle uue mudeli modifikatsiooniindekseid ja otsime sealt üles kõige suurema indeksiga parameetri.

modindices(fit2)

## lhs op rhs mi epc sepc.lv sepc.all sepc.nox  
## 26 visual =~ x4 0.904 0.066 0.059 0.051 0.051  
## 27 visual =~ x5 6.079 -0.190 -0.168 -0.131 -0.131  
## 28 visual =~ x6 2.482 0.104 0.092 0.084 0.084  
## 29 visual =~ x7 6.030 -0.241 -0.213 -0.196 -0.196  
## 30 visual =~ x8 6.030 0.291 0.257 0.254 0.254  
## 31 textual =~ x1 8.438 0.286 0.283 0.242 0.242  
## 32 textual =~ x2 0.037 -0.016 -0.016 -0.014 -0.014  
## 33 textual =~ x3 8.167 -0.236 -0.233 -0.207 -0.207  
## 34 textual =~ x9 0.096 -0.020 -0.020 -0.019 -0.019  
## 35 textual =~ x7 0.379 0.040 0.040 0.037 0.037  
## 36 textual =~ x8 0.134 -0.026 -0.026 -0.026 -0.026  
## 37 speed =~ x1 0.131 0.054 0.036 0.031 0.031  
## 38 speed =~ x2 1.859 -0.172 -0.115 -0.098 -0.098  
## 39 speed =~ x3 0.393 0.077 0.051 0.046 0.046  
## 40 speed =~ x4 0.076 -0.022 -0.014 -0.012 -0.012  
## 41 speed =~ x5 0.015 -0.010 -0.007 -0.005 -0.005  
## 42 speed =~ x6 0.167 0.031 0.020 0.019 0.019  
## 43 x1 ~~ x2 3.185 -0.152 -0.152 -0.111 -0.111  
## 44 x1 ~~ x3 0.463 -0.075 -0.075 -0.057 -0.057  
## 45 x1 ~~ x9 0.176 -0.029 -0.029 -0.024 -0.024  
## 46 x1 ~~ x4 4.148 0.083 0.083 0.061 0.061  
## 47 x1 ~~ x5 0.797 -0.040 -0.040 -0.027 -0.027  
## 48 x1 ~~ x6 0.182 0.017 0.017 0.013 0.013  
## 49 x1 ~~ x7 1.764 -0.070 -0.070 -0.055 -0.055  
## 50 x1 ~~ x8 1.725 0.067 0.067 0.057 0.057  
## 51 x2 ~~ x3 6.395 0.179 0.179 0.135 0.135  
## 52 x2 ~~ x9 0.013 0.006 0.006 0.005 0.005  
## 53 x2 ~~ x4 0.493 -0.032 -0.032 -0.024 -0.024  
## 54 x2 ~~ x5 0.052 -0.012 -0.012 -0.008 -0.008  
## 55 x2 ~~ x6 0.808 0.040 0.040 0.031 0.031  
## 56 x2 ~~ x7 7.060 -0.157 -0.157 -0.123 -0.123  
## 57 x2 ~~ x8 0.547 0.039 0.039 0.033 0.033  
## 58 x3 ~~ x9 0.349 0.033 0.033 0.029 0.029  
## 59 x3 ~~ x4 0.088 -0.012 -0.012 -0.009 -0.009  
## 60 x3 ~~ x5 8.640 -0.135 -0.135 -0.093 -0.093  
## 61 x3 ~~ x6 2.008 0.056 0.056 0.046 0.046  
## 62 x3 ~~ x7 0.115 -0.018 -0.018 -0.015 -0.015  
## 63 x3 ~~ x8 0.469 0.033 0.033 0.029 0.029  
## 64 x9 ~~ x4 0.717 -0.029 -0.029 -0.025 -0.025  
## 65 x9 ~~ x5 2.165 0.056 0.056 0.043 0.043  
## 66 x9 ~~ x6 0.786 -0.029 -0.029 -0.027 -0.027  
## 67 x9 ~~ x7 3.199 0.190 0.190 0.174 0.174  
## 68 x9 ~~ x8 2.980 -0.223 -0.223 -0.219 -0.219  
## 69 x4 ~~ x5 2.335 0.180 0.180 0.121 0.121  
## 70 x4 ~~ x6 5.249 -0.217 -0.217 -0.171 -0.171  
## 71 x4 ~~ x7 7.835 0.110 0.110 0.087 0.087  
## 72 x4 ~~ x8 4.152 -0.072 -0.072 -0.061 -0.061  
## 73 x5 ~~ x6 0.674 0.088 0.088 0.062 0.062  
## 74 x5 ~~ x7 1.653 -0.056 -0.056 -0.040 -0.040  
## 75 x5 ~~ x8 0.254 0.020 0.020 0.015 0.015  
## 76 x6 ~~ x7 0.175 -0.016 -0.016 -0.013 -0.013  
## 77 x6 ~~ x8 0.756 0.029 0.029 0.026 0.026  
## 78 x7 ~~ x8 0.096 -0.239 -0.239 -0.218 -0.218

Kõige suurem mi väärtus on real 60, mis ütleb, et lisades mudelisse muutujate *x3* ja *x5* jääkdispersioonide vahelise korrelatsiooni, võiksime oodata hii-ruudu kahanemist umbes 8.6-e võrra. Lisame jälle uue parameetri mudelisse, sobitame mudeli andmetele ja uurime väljundit.

HS.model3 <- "visual =~ x1 + x2 + x3 + x9  
textual =~ x4 + x5 + x6  
speed =~ x7 + x8 + x9  
x3 ~~ x5"  
fit3 <- cfa(HS.model3, data=HolzingerSwineford1939)  
summary(fit3, fit.measures=TRUE)

## lavaan (0.5-23.1097) converged normally after 35 iterations  
##   
## Number of observations 301  
##   
## Estimator ML  
## Minimum Function Test Statistic 43.326  
## Degrees of freedom 22  
## P-value (Chi-square) 0.004  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.976  
## Tucker-Lewis Index (TLI) 0.960  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3716.755  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 23  
## Akaike (AIC) 7479.510  
## Bayesian (BIC) 7564.773  
## Sample-size adjusted Bayesian (BIC) 7491.830  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.057  
## 90 Percent Confidence Interval 0.031 0.082  
## P-value RMSEA <= 0.05 0.302  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.041  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|)  
## visual =~   
## x1 1.000   
## x2 0.598 0.100 5.991 0.000  
## x3 0.774 0.106 7.324 0.000  
## x9 0.456 0.083 5.466 0.000  
## textual =~   
## x4 1.000   
## x5 1.112 0.066 16.920 0.000  
## x6 0.933 0.056 16.701 0.000  
## speed =~   
## x7 1.000   
## x8 1.211 0.186 6.522 0.000  
## x9 0.670 0.112 5.991 0.000  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|)  
## .x3 ~~   
## .x5 -0.138 0.046 -3.005 0.003  
## visual ~~   
## textual 0.401 0.072 5.601 0.000  
## speed 0.175 0.054 3.243 0.001  
## textual ~~   
## speed 0.135 0.050 2.670 0.008  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|)  
## .x1 0.614 0.097 6.309 0.000  
## .x2 1.115 0.100 11.129 0.000  
## .x3 0.817 0.087 9.415 0.000  
## .x9 0.555 0.060 9.320 0.000  
## .x4 0.384 0.048 8.045 0.000  
## .x5 0.440 0.058 7.581 0.000  
## .x6 0.354 0.043 8.299 0.000  
## .x7 0.741 0.086 8.604 0.000  
## .x8 0.373 0.095 3.928 0.000  
## visual 0.745 0.129 5.760 0.000  
## textual 0.967 0.112 8.665 0.000  
## speed 0.442 0.097 4.557 0.000

Sobitusastmenäitajad on läinud veel natuke paremaks. Nüüd on ka TLI üle 0.95-e, kuid hii-ruut ja RMSEA veel väga head sobivust ei näita. Uurime jällegi uue mudeli modifiktatsiooniindekseid ja otsime üles suurima.

modindices(fit3)

## lhs op rhs mi epc sepc.lv sepc.all sepc.nox  
## 27 visual =~ x4 0.172 0.031 0.027 0.023 0.023  
## 28 visual =~ x5 2.175 -0.124 -0.107 -0.084 -0.084  
## 29 visual =~ x6 1.156 0.077 0.066 0.061 0.061  
## 30 visual =~ x7 6.260 -0.255 -0.220 -0.202 -0.202  
## 31 visual =~ x8 6.260 0.308 0.266 0.263 0.263  
## 32 textual =~ x1 6.408 0.258 0.254 0.218 0.218  
## 33 textual =~ x2 0.096 -0.027 -0.027 -0.023 -0.023  
## 34 textual =~ x3 4.540 -0.185 -0.181 -0.162 -0.162  
## 35 textual =~ x9 0.289 -0.036 -0.035 -0.035 -0.035  
## 36 textual =~ x7 0.357 0.039 0.039 0.036 0.036  
## 37 textual =~ x8 0.049 -0.016 -0.016 -0.016 -0.016  
## 38 speed =~ x1 0.225 0.070 0.046 0.040 0.040  
## 39 speed =~ x2 2.101 -0.184 -0.122 -0.104 -0.104  
## 40 speed =~ x3 0.330 0.070 0.047 0.042 0.042  
## 41 speed =~ x4 0.098 -0.025 -0.017 -0.014 -0.014  
## 42 speed =~ x5 0.000 -0.001 0.000 0.000 0.000  
## 43 speed =~ x6 0.106 0.024 0.016 0.015 0.015  
## 44 x1 ~~ x2 1.886 -0.112 -0.112 -0.082 -0.082  
## 45 x1 ~~ x3 1.073 -0.109 -0.109 -0.083 -0.083  
## 46 x1 ~~ x9 0.025 -0.010 -0.010 -0.009 -0.009  
## 47 x1 ~~ x4 6.108 0.103 0.103 0.076 0.076  
## 48 x1 ~~ x5 4.176 -0.098 -0.098 -0.066 -0.066  
## 49 x1 ~~ x6 0.757 0.034 0.034 0.027 0.027  
## 50 x1 ~~ x7 1.599 -0.067 -0.067 -0.053 -0.053  
## 51 x1 ~~ x8 1.617 0.065 0.065 0.055 0.055  
## 52 x2 ~~ x3 5.321 0.163 0.163 0.123 0.123  
## 53 x2 ~~ x9 0.000 -0.001 -0.001 -0.001 -0.001  
## 54 x2 ~~ x4 0.501 -0.033 -0.033 -0.024 -0.024  
## 55 x2 ~~ x5 0.074 -0.014 -0.014 -0.009 -0.009  
## 56 x2 ~~ x6 0.724 0.037 0.037 0.029 0.029  
## 57 x2 ~~ x7 7.013 -0.157 -0.157 -0.123 -0.123  
## 58 x2 ~~ x8 0.542 0.039 0.039 0.033 0.033  
## 59 x3 ~~ x9 0.347 0.033 0.033 0.029 0.029  
## 60 x3 ~~ x4 3.112 -0.083 -0.083 -0.064 -0.064  
## 61 x3 ~~ x6 0.032 0.008 0.008 0.006 0.006  
## 62 x3 ~~ x7 0.377 -0.032 -0.032 -0.027 -0.027  
## 63 x3 ~~ x8 0.605 0.037 0.037 0.033 0.033  
## 64 x9 ~~ x4 0.505 -0.025 -0.025 -0.021 -0.021  
## 65 x9 ~~ x5 1.302 0.044 0.044 0.034 0.034  
## 66 x9 ~~ x6 0.680 -0.027 -0.027 -0.025 -0.025  
## 67 x9 ~~ x7 3.686 0.203 0.203 0.185 0.185  
## 68 x9 ~~ x8 3.295 -0.234 -0.234 -0.230 -0.230  
## 69 x4 ~~ x5 1.112 0.121 0.121 0.081 0.081  
## 70 x4 ~~ x6 1.859 -0.129 -0.129 -0.101 -0.101  
## 71 x4 ~~ x7 8.015 0.112 0.112 0.089 0.089  
## 72 x4 ~~ x8 4.237 -0.073 -0.073 -0.062 -0.062  
## 73 x5 ~~ x6 0.109 0.035 0.035 0.025 0.025  
## 74 x5 ~~ x7 1.740 -0.056 -0.056 -0.041 -0.041  
## 75 x5 ~~ x8 0.386 0.024 0.024 0.018 0.018  
## 76 x6 ~~ x7 0.109 -0.012 -0.012 -0.010 -0.010  
## 77 x6 ~~ x8 0.633 0.027 0.027 0.024 0.024  
## 78 x7 ~~ x8 0.289 -0.477 -0.477 -0.434 -0.434

Suurim väärtus tulbas *mi* on real 71, mis ütleb, et et lisades mudelisse muutujate *x4* ja *x7* jääkdispersioonide vahelise korrelatsiooni, võiksime oodata hii-ruudu kahanemist umbes 8 võrra. Lisame taas uue parameetri mudelisse, sobitame mudeli andmetele ja uurime väljundit.

HS.model4 <- "visual =~ x1 + x2 + x3 + x9  
textual =~ x4 + x5 + x6  
speed =~ x7 + x8 + x9  
x3 ~~ x5  
x4 ~~ x7"  
fit4 <- cfa(HS.model4, data=HolzingerSwineford1939)  
summary(fit4, fit.measures=TRUE)

## lavaan (0.5-23.1097) converged normally after 35 iterations  
##   
## Number of observations 301  
##   
## Estimator ML  
## Minimum Function Test Statistic 35.144  
## Degrees of freedom 21  
## P-value (Chi-square) 0.027  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.984  
## Tucker-Lewis Index (TLI) 0.973  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3712.664  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 24  
## Akaike (AIC) 7473.328  
## Bayesian (BIC) 7562.299  
## Sample-size adjusted Bayesian (BIC) 7486.184  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.047  
## 90 Percent Confidence Interval 0.016 0.074  
## P-value RMSEA <= 0.05 0.531  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.040  
##   
## Parameter Estimates:  
##   
## Information Expected  
## Standard Errors Standard  
##   
## Latent Variables:  
## Estimate Std.Err z-value P(>|z|)  
## visual =~   
## x1 1.000   
## x2 0.597 0.100 5.999 0.000  
## x3 0.768 0.105 7.332 0.000  
## x9 0.459 0.083 5.544 0.000  
## textual =~   
## x4 1.000   
## x5 1.112 0.065 17.057 0.000  
## x6 0.931 0.055 16.781 0.000  
## speed =~   
## x7 1.000   
## x8 1.209 0.183 6.616 0.000  
## x9 0.660 0.109 6.069 0.000  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|)  
## .x3 ~~   
## .x5 -0.140 0.046 -3.039 0.002  
## .x4 ~~   
## .x7 0.113 0.040 2.812 0.005  
## visual ~~   
## textual 0.409 0.072 5.696 0.000  
## speed 0.175 0.054 3.244 0.001  
## textual ~~   
## speed 0.126 0.051 2.476 0.013  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|)  
## .x1 0.613 0.097 6.326 0.000  
## .x2 1.116 0.100 11.139 0.000  
## .x3 0.820 0.086 9.486 0.000  
## .x9 0.555 0.059 9.361 0.000  
## .x4 0.384 0.048 8.042 0.000  
## .x5 0.438 0.058 7.602 0.000  
## .x6 0.358 0.043 8.401 0.000  
## .x7 0.742 0.087 8.581 0.000  
## .x8 0.363 0.095 3.835 0.000  
## visual 0.746 0.129 5.780 0.000  
## textual 0.967 0.111 8.727 0.000  
## speed 0.451 0.097 4.636 0.000

Sobitusastme näitajad on jällegi natuke paranenud. Nüüd on ka RMSEA alla 0.05, kuid hii-ruut pole jätkuvalt veel head taset saavutanud. Põhimõtteliselt võiksime parameetrite lisamise protsessi samal viisil jätkata. Selle kohta, kas parameetrite lisamine muutis sobitusastet oluliselt paremaks, saame arvutada ka p-väärtuse. Seda saame teha funktsiooni anova abil, millele anname ette mudelid, mida omavahel võrrelda soovime. Seda funktsiooni saame aga kasutada ainult hierarhiliste mudelite võrdlemiseks. Hierarhilised on mudelid siis, kui üks mudel sisaldub teises ja me saame ühe teisest tuletada vabade parameetrite lisamise teel.

anova(fit3, fit4)

## Chi Square Difference Test  
##   
## Df AIC BIC Chisq Chisq diff Df diff Pr(>Chisq)   
## fit4 21 7473.3 7562.3 35.144   
## fit3 22 7479.5 7564.8 43.325 8.1817 1 0.004232 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Antud juhul näeme tabeli kõige parempoolsemast tulbast, et p-väärtus on hästi väike (0.004 ehk <0.01), mis tähendab, et neljas mudel on oluliselt kolmandast parem. Mudelite indeksite alusel modifitseerimisele tuleks üldiselt läheneda väga ettevaatlikult. Mudeli märkimisväärse muutmisega kaasneb oht, et me ülesobitame mudeli just sellele konkreetsele andmestikule ja leitud mudelit ei ole hiljem võimalik üldistada teistele valimitele. Igal valimi juures on mõningad omapärasused ja kui me oma mudelit mingi valimi andmete põhjal muudame, on oht, et muutused kajastavad just neid omapärasusi ning mitte üldiselt esinevaid seoseid meid huvitavate muutujate vahel.

# Ülesanded

1. Andmefailis "suur.viisik" on 30 Suure Viisiku kitsama alaomaduse skoori ehk iga isiksuseomaduse kohta 6 alaomadust. Tehke selle andmestiku kohta 5-faktoriline kinnitava faktoranalüüusi mudel, nii nagu Suure Viisiku teooria seda ette näeb. Hinnake sellise mudeli sobivust antud andmetele. Mudeli koostamiseks on vaja teda tabelis olevate muutujate nimesid. Need saab katte funktsiooni names abil.

names(suur.viisik)

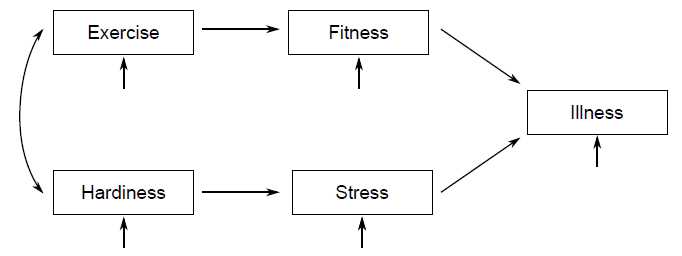
## [1] "N1" "N2" "N3" "N4" "N5" "N6" "E1" "E2" "E3" "E4" "E5" "E6" "O1" "O2"  
## [15] "O3" "O4" "O5" "O6" "A1" "A2" "A3" "A4" "A5" "A6" "C1" "C2" "C3" "C4"  
## [29] "C5" "C6"

Uurige ka mudeli parameetrite tabelit. Vaadake iga faktori puhul, millised vaadeldud muutujatest laaduvad faktorile kõige tugevamalt ja millised kõige nõrgemalt? Millised faktorid on omavahel kõige tugevamalt ja millised kõige nõrgemalt korreleeritud? Uurige jääkdispersioonide abil, milliseid vaadeldud muutujad seletab antud mudel kõige halvemini?

1. Teine ülesanne on laenatud Rex Kline'i raamatust Principles and *Practice of Structural Equation Modeling* ja kasutatavad andmed parinevad algselt uurimusest, mille autoriteks on Roth jt (1989). Maatriks nimega *illness.cov* sisaldab 5 muutuja vahelisi kovariatsioonikordajaid. Muutujad vastavad viie küsimustiku skoorile ja nende sisu on järgmine:

2.1 *exercise* - kui palju vastaja spordiga tegeleb;  
2.2 *fitness* - kui heaks hindab vastaja oma füüsilist vormi;  
2.3 *hardiness* - milliseks hindab vastaja oma vaimset sitkust;  
2.4 *stress* - milliseks hindab vastaja oma stressitaset viimase kuu jooksul;  
2.5 *illness* - mil määral on vastajal viimasel ajal esinenud erinevate haiguste füüsilisi sümptome;

Struktuurvõrrandite mudeli koostamiseks võib toorandmete asemel kasutada ka kovariatsioonimaatriksit. Maatriksi põhjal mudeli sobitamise koodirida on koodifailis juba olemas. Teie ülesandeks on koostada selle ette alloleva joonise põhjal mudeli kirjeldus. Kas antud mudel sobib andmetega?



Joonis 3.

Kas kõik mudelis esinevad seosed on statistiliselt olulised? Lisage mudelisse täiendav regressioonseos, milles sõltumatuks muutujaks on *fitness* ja sõltuvaks muutujaks *stress*. Mis juhtub mudeli sobitusastmega?