|  |
| --- |
| [Type the company name] |
| STAT 445 Assignment 6 |
| Kun Yang 301178299 |
|  |
|  |
|  |

|  |
| --- |
|  |

***Problem 9.18 Do only the maximum likelihood portion of part (d).*> (CORR<- matrix(c(1,0.4919,0.2636,0.4653,-0.2277,0.0652,**

**+ 0.4919,1,0.3127,0.3506,-0.1917,0.2045,**

**+ 0.2635,0.3127,1,0.4108, 0.0647,0.2493,**

**+ 0.4653,0.3506,0.4108,1,-0.2249,0.2293,**

**+ -0.2277,-0.1917,0.0647,-0.2249,1,-0.2144,**

**+ 0.0652,0.2045,0.2493,0.2293,-0.2144,1),ncol=6,byrow=TRUE))**

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1.0000 0.4919 0.2636 0.4653 -0.2277 0.0652

[2,] 0.4919 1.0000 0.3127 0.3506 -0.1917 0.2045

[3,] 0.2635 0.3127 1.0000 0.4108 0.0647 0.2493

[4,] 0.4653 0.3506 0.4108 1.0000 -0.2249 0.2293

[5,] -0.2277 -0.1917 0.0647 -0.2249 1.0000 -0.2144

[6,] 0.0652 0.2045 0.2493 0.2293 -0.2144 1.0000

**> (eig=eigen(CORR))**

$values

[1] 2.3549250 1.0718567 0.9842486 0.6643858 0.5003897 0.4241942

$vectors

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] -0.4753543 0.02188910 0.47987482 0.04564932 0.3578888 -0.6425051

[2,] -0.4719328 -0.01924243 0.20906455 0.70296351 -0.1771358 0.4556533

[3,] -0.3931540 -0.56061775 -0.26440849 -0.17551515 -0.5973982 -0.2713733

[4,] -0.4963538 -0.07723024 0.03226116 -0.60426514 0.3238962 0.5260914

[5,] 0.2563177 -0.80502150 0.01294351 0.21817123 0.4823355 0.0768220

[6,] -0.2910014 0.17559671 -0.80925398 0.24539324 0.3822272 -0.1524794

***###analysis with 3 factors***

**> (FA=factanal(covmat=CORR,factors=3) )**

Call:

factanal(factors = 3, covmat = CORR)

Uniquenesses:

[1] 0.005 0.666 0.005 0.607 0.683 0.720

Loadings:

Factor1 Factor2 Factor3

[1,] **0.994**

[2,] 0.466 0.211 0.268

[3,] 0.201 **0.976**

[4,] 0.429 0.316 0.330

[5,] -0.208 0.134 **-0.505**

[6,] 0.227 0.478

Factor1 Factor2 Factor3

SS loadings 1.474 1.170 0.670

Proportion Var 0.246 0.195 0.112

Cumulative Var 0.246 0.441 0.552

The degrees of freedom for the model is 0 and the fit was 0.0016

**> str(FA)**

List of 11

$ converged : logi TRUE

$ loadings : loadings [1:6, 1:3] 0.994 0.466 0.201 0.429 -0.208 ...

..- attr(\*, "dimnames")=List of 2

.. ..$ : NULL

.. ..$ : chr [1:3] "Factor1" "Factor2" "Factor3"

$ uniquenesses: num [1:6] 0.005 0.666 0.005 0.607 0.683 ...

$ correlation : num [1:6, 1:6] 1 0.492 0.264 0.465 -0.228 ...

$ criteria : Named num [1:3] 0.00157 18 18

..- attr(\*, "names")= chr [1:3] "objective" "counts.function" "counts.gradient"

$ factors : num 3

$ dof : num 0

$ method : chr "mle"

$ rotmat : num [1:3, 1:3] 0.7539 -0.6544 0.0577 0.6533 0.7561 ...

$ n.obs : logi NA

$ call : language factanal(factors = 3, covmat = CORR)

- attr(\*, "class")= chr "factanal"

***> ###communalities***

**> L=FA$loadings**

**> (hi2=(L%\*%t(L)))**

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 0.9950001 0.4918611 0.26350000 0.4653067 -0.22769436 0.0652224

[2,] 0.4918611 0.3336753 0.31273101 0.3551980 -0.20419524 0.1870385

[3,] 0.2635000 0.3127310 0.99500005 0.4107643 0.06469249 0.2493014

[4,] 0.4653067 0.3551980 0.41076428 0.3929169 -0.21361990 0.2394891

[5,] -0.2276944 -0.2041952 0.06469249 -0.2136199 0.31692365 -0.2160465

[6,] 0.0652224 0.1870385 0.24930139 0.2394891 -0.21604647 0.2804297

***> ###specific variances***

**> (psi=diag(FA$uniquenesses))**

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 0.005 0.0000000 0.000 0.0000000 0.000000 0.0000000

[2,] 0.000 0.6663258 0.000 0.0000000 0.000000 0.0000000

[3,] 0.000 0.0000000 0.005 0.0000000 0.000000 0.0000000

[4,] 0.000 0.0000000 0.000 0.6070835 0.000000 0.0000000

[5,] 0.000 0.0000000 0.000 0.0000000 0.683076 0.0000000

[6,] 0.000 0.0000000 0.000 0.0000000 0.000000 0.7195693

***> ###residual matrix and round it to 4 decimal places***

**> RezMat=CORR-(hi2+psi)**

**> round(RezMat,4)**

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 0 0.0000 1e-04 0.0000 0.0000 0.0000

[2,] 0 0.0000 0e+00 -0.0046 0.0125 0.0175

[3,] 0 0.0000 0e+00 0.0000 0.0000 0.0000

[4,] 0 -0.0046 0e+00 0.0000 -0.0113 -0.0102

[5,] 0 0.0125 0e+00 -0.0113 0.0000 0.0016

[6,] 0 0.0175 0e+00 -0.0102 0.0016 0.0000

***Diagonal elements of the residual matrix are 0 which means that (hi2+psi) is close to the correlation matrix.***

***So m = 3 factor solution represents the observed correlations quite well.***

**Factor 1** puts significant and maximum weight on bluegill(x1). Black crappie, smallmouth bass, Largemouth bass, Walleye, Northern pike have significantly smaller weights than the weights on the bluegill group. The first factor emphasize on bluegill.

**Factor 2** puts significant and maximum weight on smallmouth bass (x3). Bluegill, Black crappie, Largemouth bass, Walleye, Northern pike have significantly smaller weights than the weights on the smallmouth bass group. So the second factor emphasize on smallmouth bass.

**Factor 3** is negatively driven by walleye. But it is not very significant.

These three factors collectively accounted for 55.2% of total variation. It represents quite small proportion of variation. But if we conduct the analysis by m=2, the cumulative variation will be even lower and the second factor is hard to interpret meaningfully. Therefore, I choose m=3.

***Analysis with 2 factors as following:***

> (FA=factanal(covmat=CORR,factors=2) )

Call:

factanal(factors = 2, covmat = CORR)

Uniquenesses:

[1] 0.453 0.600 0.005 0.581 0.829 0.914

Loadings:

Factor1 Factor2

[1,] 0.279 0.685

[2,] 0.326 0.542

[3,]  **0.997**

[4,] 0.423 0.491

[5,] -0.410

[6,] 0.253 0.148

Factor1 Factor2

SS loadings 1.424 1.195

Proportion Var 0.237 0.199

Cumulative Var 0.237 **0.437**

The degrees of freedom for the model is 4 and the fit was 0.0935

***Problem 9.19***

***Part (a) Obtain only the maximum likelihood estimates for the factor analysis model with m=2 & m=3.***

**> Data.matrix<-read.csv("Salespeopledata.csv")**

**> (CORR=cor(Data.matrix))**

SG SP NA. CT MR AR MT

SG 1.0000000 0.9260758 0.8840023 0.5720363 0.7080738 0.6744073 0.9273116

SP 0.9260758 1.0000000 0.8425232 0.5415080 0.7459097 0.4653880 0.9442960

NA. 0.8840023 0.8425232 1.0000000 0.7003630 0.6374712 0.6410886 0.8525682

CT 0.5720363 0.5415080 0.7003630 1.0000000 0.5907360 0.1469074 0.4126395

MR 0.7080738 0.7459097 0.6374712 0.5907360 1.0000000 0.3859502 0.5745533

AR 0.6744073 0.4653880 0.6410886 0.1469074 0.3859502 1.0000000 0.5663721

MT 0.9273116 0.9442960 0.8525682 0.4126395 0.5745533 0.5663721 1.0000000

**> (eig=eigen(CORR))**

$values

[1] 5.03459779 0.93351614 0.49791975 0.42124549 0.08104043 0.02034063 0.01133977

$vectors

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] -0.4336719 0.111754422 0.075488541 -0.04237344 0.6324942624 0.3365963 0.52782527

[2,] -0.4202136 -0.029287495 0.442478953 0.01075255 -0.0001182093 -0.7853424 0.09948330

[3,] -0.4210510 -0.009201975 -0.204189315 -0.32492838 -0.7010262539 0.1568114 0.39916419

[4,] -0.2942863 -0.668415809 -0.451492333 -0.30271208 0.2610080204 -0.1141710 -0.29995962

[5,] -0.3490920 -0.294944379 -0.005921773 0.84660356 -0.1742634819 0.1969091 -0.07231139

[6,] -0.2891669 0.642377957 -0.603779622 0.15367411 0.0869586057 -0.2362610 -0.22844351

[7,] -0.4074041 0.200367651 0.434039576 -0.24601320 -0.0495826418 0.3711105 -0.63622351

***###analysis with 3 factors (m=3)***

**> (FA=factanal(covmat=CORR,factors=3) )**

Call:

factanal(factors = 3, covmat = CORR)

Uniquenesses:

SG SP NA. CT MR AR MT

0.039 0.034 0.088 0.005 0.447 0.005 0.038

Loadings:

Factor1 Factor2 Factor3

SG **0.793** 0.374 0.438

SP  **0.911** 0.317 0.185

NA. **0.651** 0.544 0.438

CT 0.255  **0.964**

MR 0.542 0.465 0.207

AR 0.299 **0.950**

MT **0.917** 0.180 0.298

Factor1 Factor2 Factor3

SS loadings 3.175 1.718 1.453

Proportion Var 0.454 0.245 0.208

Cumulative Var 0.454 0.699 **0.906**

The degrees of freedom for the model is 3 and the fit was 1.4186

**> str(FA)**

List of 11

$ converged : logi TRUE

$ loadings : loadings [1:7, 1:3] 0.793 0.911 0.651 0.255 0.542 ...

..- attr(\*, "dimnames")=List of 2

.. ..$ : chr [1:7] "SG" "SP" "NA." "CT" ...

.. ..$ : chr [1:3] "Factor1" "Factor2" "Factor3"

$ uniquenesses: Named num [1:7] 0.0386 0.0345 0.0881 0.005 0.4466 ...

..- attr(\*, "names")= chr [1:7] "SG" "SP" "NA." "CT" ...

$ correlation : num [1:7, 1:7] 1 0.926 0.884 0.572 0.708 ...

..- attr(\*, "dimnames")=List of 2

.. ..$ : chr [1:7] "SG" "SP" "NA." "CT" ...

.. ..$ : chr [1:7] "SG" "SP" "NA." "CT" ...

$ criteria : Named num [1:3] 1.42 27 27

..- attr(\*, "names")= chr [1:3] "objective" "counts.function" "counts.gradient"

$ factors : num 3

$ dof : num 3

$ method : chr "mle"

$ rotmat : num [1:3, 1:3] 0.612 0.714 0.341 -0.6 0.699 ...

$ n.obs : logi NA

$ call : language factanal(factors = 3, covmat = CORR)

- attr(\*, "class")= chr "factanal"

***###communalities***

**> L=FA$loadings**

**> (hi2=(L%\*%t(L)))**

SG SP NA. CT MR AR MT

SG 0.9614284 0.9228135 0.9120898 0.5714373 0.6949357 0.6738835 0.9255320

SP 0.9228135 0.9655192 0.8471023 0.5417813 0.6799465 0.4654561 0.9481930

NA. 0.9120898 0.8471023 0.9118756 0.6991264 0.6969691 0.6402871 0.8255826

CT 0.5714373 0.5417813 0.6991264 0.9950434 0.5910466 0.1469508 0.4130097

MR 0.6949357 0.6799465 0.6969691 0.5910466 0.5533820 0.3841948 0.6425648

AR 0.6738835 0.4654561 0.6402871 0.1469508 0.3841948 0.9950317 0.5669006

MT 0.9255320 0.9481930 0.8255826 0.4130097 0.6425648 0.5669006 0.9624901

***###specific variances***

**> (psi=diag(FA$uniquenesses))**

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] 0.03857165 0.00000000 0.00000000 0.000 0.0000000 0.000 0.0000000

[2,] 0.00000000 0.03448071 0.00000000 0.000 0.0000000 0.000 0.0000000

[3,] 0.00000000 0.00000000 0.08812176 0.000 0.0000000 0.000 0.0000000

[4,] 0.00000000 0.00000000 0.00000000 0.005 0.0000000 0.000 0.0000000

[5,] 0.00000000 0.00000000 0.00000000 0.000 0.4466205 0.000 0.0000000

[6,] 0.00000000 0.00000000 0.00000000 0.000 0.0000000 0.005 0.0000000

[7,] 0.00000000 0.00000000 0.00000000 0.000 0.0000000 0.000 0.0375098

***###residual matrix and round it to 4 decimal places***

**> RezMat=CORR-(hi2+psi)**

**> round(RezMat,4)**

SG SP NA. CT MR AR MT

SG 0.0000 0.0033 -0.0281 0.0006 0.0131 0.0005 0.0018

SP 0.0033 0.0000 -0.0046 -0.0003 0.0660 -0.0001 -0.0039

NA. -0.0281 -0.0046 0.0000 0.0012 -0.0595 0.0008 0.0270

CT 0.0006 -0.0003 0.0012 0.0000 -0.0003 0.0000 -0.0004

MR 0.0131 0.0660 -0.0595 -0.0003 0.0000 0.0018 -0.0680

AR 0.0005 -0.0001 0.0008 0.0000 0.0018 0.0000 -0.0005

MT 0.0018 -0.0039 0.0270 -0.0004 -0.0680 -0.0005 0.0000

***Diagonal elements of the residual matrix are 0 which means that (hi2+psi) is close to the correlation matrix.***

***So m = 3 factor solution represents the observed correlations quite well.***

***###analysis with 2 factors***

**> (FA=factanal(covmat=CORR,factors=2) )**

Call:

factanal(factors = 2, covmat = CORR)

Uniquenesses:

SG SP NA. CT MR AR MT

0.069 0.070 0.123 0.005 0.474 0.614 0.029

Loadings:

Factor1 Factor2

SG **0.852** 0.452

SP **0.868** 0.419

NA. **0.717** 0.602

CT 0.148 **0.987**

MR 0.501 0.525

AR 0.619

MT  **0.946** 0.277

Factor1 Factor2

SS loadings 3.545 2.071

Proportion Var 0.506 0.296

Cumulative Var 0.506  **0.802**

The degrees of freedom for the model is 8 and the fit was 2.6337

**> str(FA)**

List of 11

$ converged : logi TRUE

$ loadings : loadings [1:7, 1:2] 0.852 0.868 0.717 0.148 0.501 ...

..- attr(\*, "dimnames")=List of 2

.. ..$ : chr [1:7] "SG" "SP" "NA." "CT" ...

.. ..$ : chr [1:2] "Factor1" "Factor2"

$ uniquenesses: Named num [1:7] 0.0692 0.0704 0.1233 0.005 0.4736 ...

..- attr(\*, "names")= chr [1:7] "SG" "SP" "NA." "CT" ...

$ correlation : num [1:7, 1:7] 1 0.926 0.884 0.572 0.708 ...

..- attr(\*, "dimnames")=List of 2

.. ..$ : chr [1:7] "SG" "SP" "NA." "CT" ...

.. ..$ : chr [1:7] "SG" "SP" "NA." "CT" ...

$ criteria : Named num [1:3] 2.63 33 33

..- attr(\*, "names")= chr [1:3] "objective" "counts.function" "counts.gradient"

$ factors : num 2

$ dof : num 8

$ method : chr "mle"

$ rotmat : num [1:2, 1:2] 0.95 -0.312 0.312 0.95

$ n.obs : logi NA

$ call : language factanal(factors = 2, covmat = CORR)

- attr(\*, "class")= chr "factanal"

***###communalities***

**> L=FA$loadings**

**> (hi2=(L%\*%t(L)))**

SG SP NA. CT MR AR MT

SG 0.9308083 0.9295188 0.8834712 0.5720627 0.6642317 0.5543372 0.9311883

SP 0.9295188 0.9296182 0.8749729 0.5413952 0.6547850 0.5624008 0.9373111

NA. 0.8834712 0.8749729 0.8766896 0.6996404 0.6751663 0.4798309 0.8449575

CT 0.5720627 0.5413952 0.6996404 0.9950121 0.5918666 0.1504777 0.4126431

MR 0.6642317 0.6547850 0.6751663 0.5918666 0.5264121 0.3412919 0.6189370

AR 0.5543372 0.5624008 0.4798309 0.1504777 0.3412919 0.3863622 0.6017601

MT 0.9311883 0.9373111 0.8449575 0.4126431 0.6189370 0.6017601 0.9711829

***###specific variances***

**> (psi=diag(FA$uniquenesses))**

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] 0.0691916 0.00000000 0.0000000 0.000 0.0000000 0.0000000 0.00000000

[2,] 0.0000000 0.07038038 0.0000000 0.000 0.0000000 0.0000000 0.00000000

[3,] 0.0000000 0.00000000 0.1233088 0.000 0.0000000 0.0000000 0.00000000

[4,] 0.0000000 0.00000000 0.0000000 0.005 0.0000000 0.0000000 0.00000000

[5,] 0.0000000 0.00000000 0.0000000 0.000 0.4735849 0.0000000 0.00000000

[6,] 0.0000000 0.00000000 0.0000000 0.000 0.0000000 0.6136386 0.00000000

[7,] 0.0000000 0.00000000 0.0000000 0.000 0.0000000 0.0000000 0.02881701

***###residual matrix and round it to 4 decimal places***

**> RezMat=CORR-(hi2+psi)**

**> round(RezMat,4)**

SG SP NA. CT MR AR MT

SG 0.0000 -0.0034 0.0005 0.0000 0.0438 0.1201 -0.0039

SP -0.0034 0.0000 -0.0324 0.0001 0.0911 -0.0970 0.0070

NA. 0.0005 -0.0324 0.0000 0.0007 -0.0377 0.1613 0.0076

CT 0.0000 0.0001 0.0007 0.0000 -0.0011 -0.0036 0.0000

MR 0.0438 0.0911 -0.0377 -0.0011 0.0000 0.0447 -0.0444

AR 0.1201 -0.0970 0.1613 -0.0036 0.0447 0.0000 -0.0354

MT -0.0039 0.0070 0.0076 0.0000 -0.0444 -0.0354 0.0000

***Diagonal elements of the residual matrix are 0 which means that (hi2+psi) is close to the correlation matrix, so m = 2 factor solution represents the observed correlations quite well.***

***Part (b) Interpret these estimates where possible.***

***For analysis with 3 factors (m=3)***

**Factor 1** puts maximum weight on mathematics(x7) test and it places large weights on sales growth (x1) and profitability of sales (x2) and new account sales(x3). Thus, the first factor can be described as the salesperson’s performance (sales growth, profitability of sales, new account sales) with one’s mathematical ability.

**Factor 2** puts significant and maximum weight on creativity test. Thus, the second factor can be described as creative ability.

**Factor 3** puts significant and maximum weight on abstract reasoning test. Thus, the third factor can be described as abstract reasoning ability

These three factors collectively accounted for 90.6% of total variation.

***For analysis with 2 factors (m=2)***

**Factor 1** puts maximum weight on mathematics(x7) test and it places large weights on sales growth (x1) and profitability of sales (x2) and new account sales(x3). Thus, the first factor can be described as the salesperson’s performance (sales growth, profitability of sales, new account sales) with one’s mathematical ability.

**Factor 2** puts significant and maximum weight on creativity test. Thus, the second factor can be described as creative ability.

These three factors collectively accounted for 80.2% of total variation.

**Part (d) As stated. I.e., perform the specified tests and, in light of the results of these tests and the interpretations in Part (b), draw an appropriate conclusion on the best value for m.**

***For analysis with 3 factors (m=3)***

**> (A<-det(hi2+psi))**

**.**

[1] 7.727395e-05

**> (B<-det(CORR))**

[1] 1.842687e-05

**=3**

**> (C<-50-1-(2\*7+4\*3+5)/6)**

[1] 43.83333

**with df=3 α=0.01**

**> (Test<-C\*log(A/B))**

[1] 62.83714

So we reject for m=3.

***For analysis with 2 factors (m=2)***

**.**

**> (A<-det(hi2+psi))**

[1] 0.0002572145

**=8**

**> (B<-det(CORR))**

[1] 1.842687e-05

**> (C<-50-1-(2\*7+4\*2+5)/6)**

**with df=8 α=0.01**

[1] 44.5

**> (Test<-C\*log(A/B))**

[1] 117.3065

So we reject for m=2.

**Therefore, we can say neither of the m = 2, m= 3 models can fit the criterion.**

**But p-value are closer to α when m=3. And in light of part (b), I prefer m=3, because all of the 3 factors can be interpreted meaningfully. When m=3, diagonal elements of the residual matrix are 0. These three factors collectively accounted for 90.6% of total variation.**