

# Statistics 5401

## 24. Factor Scores

Gary W. Oehlert  
School of Statistics  
313B Ford Hall  
612-625-1557  
gary@stat.umn.edu

Under the factor model, and  $x$  can be written as

$$x = \mu + \mathbf{L}f + \epsilon$$

The  $f$  values are the *factor scores* for  $x$ .

The factor scores are intrinsically unobservable. Even if we know  $\mu$ ,  $\mathbf{L}$ , and  $\Psi$  exactly, we cannot compute  $f$  and  $\epsilon$  exactly.

We can, however, estimate them.

Note, this is a little different from the standard situation. We usually estimate parameters, fixed unknown quantities in our model.

Here we are estimating  $f$  and  $\epsilon$ , which are random variables.

This is sometimes called prediction, instead of estimation.

What would we do with the factor scores if we had them?

- Sometimes the factor scores are the main items of interest. For example, responses to a large questionnaire may be thought to be driven by several factors. A political questionnaire could have factors expressing economic, social justice, criminal justice, and other interests. Intelligence tests have the IQ scale, and so on.

We want the factor scores for each item in the data set.

- Sometimes we want the factor scores to use as a dimension reduction tool.

- Sometimes we can identify structure, for example, outliers, in the factor scores.

Most methods for estimating factor scores assume that  $\mathbf{L}$  and  $\Psi$  are known exactly. Of course, we never do. None the less, we will use  $\hat{\mathbf{L}}$  and  $\hat{\Psi}$  as if they were the true values.

The Regression Method.

$x$  and  $f$  have a joint normal distribution.

$$E \left( \begin{bmatrix} x \\ f \end{bmatrix} \right) = \begin{bmatrix} \mu \\ 0 \end{bmatrix}$$
$$Var \left( \begin{bmatrix} x \\ f \end{bmatrix} \right) = \begin{bmatrix} \mathbf{L}\mathbf{L}' + \Psi & \mathbf{L} \\ \mathbf{L}' & \mathbf{I} \end{bmatrix}$$

Using the properties of the multivariate normal, we have

$$E[f|x] = \mathbf{L}'(\mathbf{L}\mathbf{L}' + \Psi)^{-1}(x - \mu)$$

or

$$\hat{f} = \mathbf{L}'(\mathbf{L}\mathbf{L}' + \Psi)^{-1}(x - \mu)$$

This has  $E[\hat{f} - f|x] = 0$  and minimizes  $V[\hat{f} - f|x]$ .

We don't know  $\hat{\mathbf{L}}$  or  $\hat{\Psi}$ , so we plug in our estimates of parameters and use

$$\hat{f} = \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1}(x - \bar{\mathbf{x}})$$

The error  $\hat{f} - f$  will not be zero, even if  $\mathbf{L}$  and  $\Psi$  are known exactly. Note that the regression parameters are

$$\mathbf{L}'(\mathbf{L}\mathbf{L}' + \Psi)^{-1} = \mathbf{L}'\Sigma^{-1}$$

and we use  $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}$  as an estimate of  $\Sigma$ .

$\mathbf{S}$  is also an estimate of  $\Sigma$ , so we can also use

$$\tilde{f} = \hat{\mathbf{L}}'\mathbf{S}^{-1}(x - \bar{x})$$

to estimate factor scores.

For fully converged maximum likelihood factor extraction,

$$\hat{f} = \tilde{f}$$

This equality does not hold for other methods of factor extraction.

The Weighted Least Squares method.

$$x - \mu = \mathbf{L}f + \epsilon$$

We could consider choosing factor scores  $\hat{f}$  to make  $\epsilon$  as small as possible. Because the components of  $\epsilon$  have different variances, we could minimize the weighted sum of squared unique factors. This leads to a (weighted) regression of  $x - \mu$  on  $\mathbf{L}$ .

As with the regression method, we need to use  $\hat{\mathbf{L}}$ ,  $\hat{\Psi}$ , and  $\bar{x}$  in place of their population values. This gives

$$\hat{f}_{wls} = (\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}(x - \bar{x})$$

It turns out that

$$\hat{f}_{wls} = (\mathbf{I} - (\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1})\hat{f}$$

When  $\hat{\Psi}$  has small diagonals,  $(\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}$  will also be small, and the two estimates will be similar.

These calculations are based on  $\hat{\mathbf{L}}$  and  $\hat{\Psi}$  for  $\mathbf{S}$ . If you use  $\mathbf{R}$ , you should standardize your  $x$ s before estimating (standardized) factor scores.

```
Cmd> readdata(" ", x1, x2, x3, x4, x5, x6, x7, x8, nation)
```

```
Read from file "~/5401/JW5data/T8-6.DAT"
```

```
Cmd> X <- hconcat(x1, x2, x3, x4, x5, x6, x7, x8)
```

```
Cmd> addmacrofile("facanal.mac")
```

```
Cmd> S <- tabs(X, covar:T)
```

```
Cmd> out <- facanal(S, 2)
```

```
Convergence in 11 iterations by criterion 2
```

```
estimated uniquenesses:
```

```
(1) 0.010003 0.031506 0.32148 0.0005488
```

```
(5) 0.0019864 0.021691 0.058682 7.3213
```

```

unrotated estimated loadings:
(1,1)      0.25683      -0.21803
(2,1)      0.51033      -0.3516
(3,1)      1.2457       -0.49974
(4,1)      0.058326     -0.010245
(5,1)      0.14935      -0.0040047
(6,1)      0.77904      0.11517
(7,1)      1.7727       0.25839
(8,1)      8.5157       2.3022
minimized ml criterion:
(1)        0.33273

```

```
Cmd> L <- out$loadings
```

```
Cmd> Psi <- dmat(out$psihat)
```

```

Cmd> L
(1,1)      0.25683      -0.21803
(2,1)      0.51033      -0.3516
(3,1)      1.2457       -0.49974
(4,1)      0.058326     -0.010245
(5,1)      0.14935      -0.0040047
(6,1)      0.77904      0.11517
(7,1)      1.7727       0.25839
(8,1)      8.5157       2.3022

```

```
Cmd> Lr <- rotation(L)
```

```

Cmd> Lr
(1,1)      0.19027      -0.27803
(2,1)      0.39958      -0.47371
(3,1)      1.0698       -0.81047
(4,1)      0.053562     -0.025259
(5,1)      0.14301      -0.043237
(6,1)      0.78184      -0.094295
(7,1)      1.7781       -0.2181
(8,1)      8.8214      -0.024332

```

```
Cmd> Xc <- X - tabs(X,mean:T)'
```

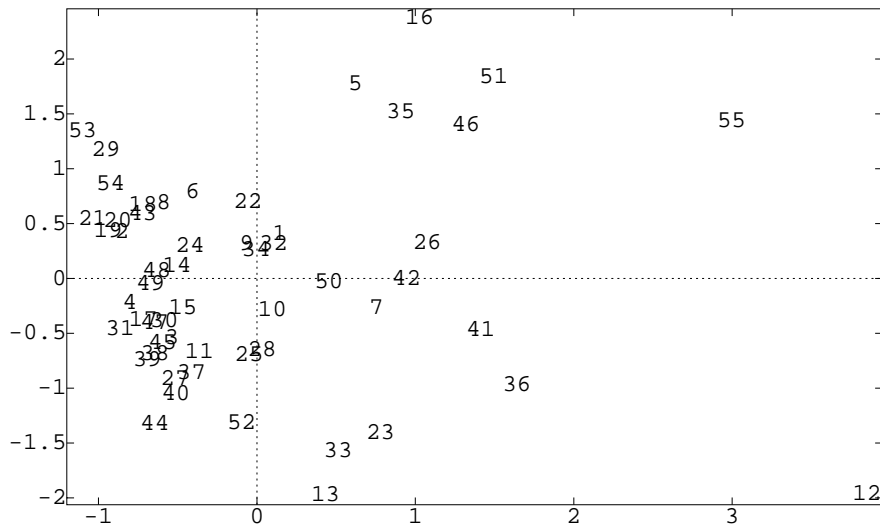
```
Cmd> fhat <- L'%*%solve(L%*%L'+Psi)%*%Xc'
```

```
Cmd> fhat <- fhat'
```

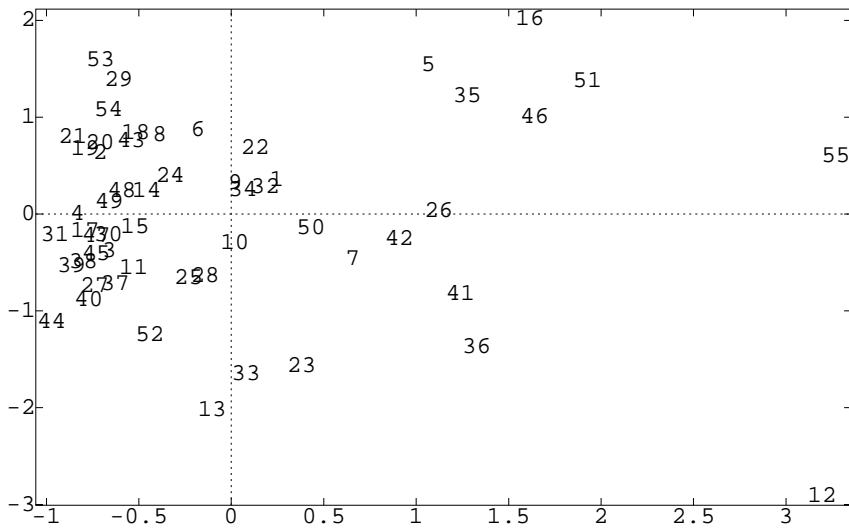
```
Cmd> fhatr <- Lr'%*%solve(Lr%*%Lr'+Psi)%*%Xc'
```

```
Cmd> fhatr <- fhatr'
```

```
Cmd> chplot(fhat[,1],fhat[,2])
```



```
Cmd> chplot(fhatr[,1],fhatr[,2])
```



```
Cmd> fhatls <- solve(L'%%solve(Psi)%%L)\  
%%L'%%solve(Psi)%%Xc'
```

```
Cmd> fhatls <- fhatls'
```

```
Cmd> chplot(fhatls[,1],fhatls[,2])
```

