

Estimation of Factor Scores

Sometimes one of the purposes of factor analysis is to find the scores f_j (values) of each factor for each case. This might be to use them in further analysis, either as predictor variables or response variables. Unfortunately, it is generally impossible to actually find the factor scores, even when the factor loadings λ_{jk} and unique variances ψ_i are perfectly known. So the most you can hope for are estimates \hat{f}_j that should be close to the true f_j .

If the factors have been rotated so that each factor has a clear interpretation, estimating the scores provides a way to characterize each case (individual) in terms of the values of the common factors for that case. Even if no rotation is done, outlying values of estimated factor score vectors $\hat{\mathbf{f}}_i$ may indicate unusual individuals.

There are two common approaches to factor score estimation, the *regression* method and the *weighted least squares* method. Both are based on the factor analysis model. To simplify, I am limiting it to the case of orthogonal (uncorrelated) factors.

The factor analysis model

The setup is that you have a random sample of multivariate data satisfying the m-factor orthogonal factor analysis model

$$\mathbf{x}_i = E[\mathbf{x}_i] + \mathbf{L}\mathbf{f}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N.$$

Here $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]'$ is p by 1, $\mathbf{L} = [\lambda_{jk}]$ is p by m and $\mathbf{f}_i = [f_{i1}, \dots, f_{im}]'$ is m by 1.

In matrix terms, if $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]'$ is the data matrix, you can write \mathbf{X} as:

$$\mathbf{X} = E[\mathbf{X}] + \mathbf{F}\mathbf{L}' + \boldsymbol{\varepsilon},$$

where $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_N]'$ is an N by m matrix of factor scores. Normally $E[\mathbf{X}] = \mathbf{1}_N \boldsymbol{\mu}'$ but might have the form $E[\mathbf{X}] = \mathbf{Z}\mathbf{B}$, where \mathbf{Z} is a matrix of predictor or dummy variables, and $\mathbf{B} = [\beta_{jk}]$ is a matrix of coefficients. In the latter case, factor analysis will be based on the *residual* covariance matrix $\mathbf{S} = (\mathbf{f}_e^{-1})\mathbf{E}$ or on the correlation matrix of the residuals.

Row i of \mathbf{F} is $\mathbf{f}_i' = [f_{i1}, \dots, f_{im}]$, the vector of *scores* (values) of the m

Factor Score Estimation

common factors for the case i . To say that the model is *orthogonal* means that $V[\mathbf{f}_i] = \mathbf{I}_m$, that is, the common factors are uncorrelated.

$\mathbf{L} = [\ell_{jk}]$ is a p by m *loading matrix* where ℓ_{jk} is the loading of variable j on factor k

Row i of the N by p matrix $\boldsymbol{\epsilon}$ is $\boldsymbol{\epsilon}_i'$, the uncorrelated *unique factor* scores for case i , with $V[\boldsymbol{\epsilon}_i] = \boldsymbol{\Psi} = \text{diag}[\psi_1, \psi_2, \dots, \psi_p]$.

Whether $E[\mathbf{X}] = \mathbf{1}_N \boldsymbol{\mu}'$ or $E[\mathbf{X}] = \mathbf{ZB}$, it is important that all cases have the same variance matrix $\boldsymbol{\Sigma}$.

The methods below are for estimating factor scores when you know \mathbf{L} and $\boldsymbol{\Psi}$. In the more realistic situation when all you have are estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$, you "plug" $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ into formulas involving \mathbf{L} and $\boldsymbol{\Psi}$. The estimates turn out to be linear in the elements of \mathbf{x} .

Regression Method

The *regression method* starts from the fact that a vector $\mathbf{f} = [f_1, \dots, f_m]'$ of unknown factor scores is correlated with the observation vector $\mathbf{x} = [x_1, \dots, x_p]' = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + [\epsilon_1, \dots, \epsilon_p]'$. The joint $(p+m)$ by $(p+m)$ variance matrix of \mathbf{x} and \mathbf{f} is (assuming $V(\mathbf{f}) = \mathbf{I}_m$)

$$V\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{L} \\ \mathbf{L}' & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} & \mathbf{L} \\ \mathbf{L}' & \mathbf{I}_m \end{bmatrix}$$

When \mathbf{x} and \mathbf{f} are jointly multivariate normal, the conditional mean of \mathbf{f} given \mathbf{x} is linear in \mathbf{x} , that is, $E[\mathbf{f} | \mathbf{x}] = \boldsymbol{\beta}_{\text{reg}}'(\mathbf{x} - \boldsymbol{\mu})$, where

$$\boldsymbol{\beta}_{\text{reg}} = V[\mathbf{x}]^{-1} \text{Cov}[\mathbf{x}, \mathbf{f}] = \boldsymbol{\Sigma}^{-1} \mathbf{L} = (\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1} \mathbf{L}.$$

This is simply the matrix of coefficients for the multivariate linear regression of \mathbf{f} on \mathbf{x} .

Even when you can't assume multivariate normality, $\mathbf{f}_{\text{reg}} = \boldsymbol{\beta}_{\text{reg}}'(\mathbf{x} - \boldsymbol{\mu})$ minimizes $V[\mathbf{f} - \mathbf{f}_{\text{reg}}]$ among all linear functions of \mathbf{x} . The difference $\mathbf{f} - \mathbf{f}_{\text{reg}} \equiv \mathbf{f} - \boldsymbol{\beta}_{\text{reg}}'(\mathbf{x} - E[\mathbf{x}])$ is the error incurred in estimating \mathbf{f} .

The variance matrix of an estimated score vector is

$$V[\mathbf{f}_{\text{reg}}] = \mathbf{I}_m - (\mathbf{I}_m + \boldsymbol{\Delta})^{-1}, \text{ where } \boldsymbol{\Delta} = \mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L}.$$

Factor Score Estimation

If you know \mathbf{L} and $\mathbf{\Psi}$, you can estimate the matrix \mathbf{F} of factor scores for all cases by

$$\mathbf{F}_{\text{reg}} = (\mathbf{X} - \mathbf{1}_N \boldsymbol{\mu}') \boldsymbol{\beta}_{\text{reg}} = (\mathbf{X} - \mathbf{1}_N \boldsymbol{\mu}') \boldsymbol{\Sigma}^{-1} \mathbf{L} = (\mathbf{X} - \mathbf{1}_N \boldsymbol{\mu}') (\mathbf{L} \mathbf{L}' + \mathbf{\Psi})^{-1} \mathbf{L}.$$

Since you don't know $\boldsymbol{\mu}$, \mathbf{L} , and $\mathbf{\Psi}$, you use estimates $\hat{\boldsymbol{\mu}}$, $\hat{\mathbf{L}}$, and $\hat{\mathbf{\Psi}}$. The estimated matrix of factor score coefficients is then

$$\hat{\boldsymbol{\beta}}_{\text{reg}} = \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{L}} = (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1} \hat{\mathbf{L}}.$$

The estimated mean is $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$, and the vector of estimated factor scores corresponding to response vector \mathbf{x} is

$$\hat{\mathbf{f}}_{\text{reg}} \equiv \hat{\boldsymbol{\beta}}_{\text{reg}}' (\mathbf{x} - \bar{\mathbf{x}}) = \hat{\mathbf{L}}' (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1} (\mathbf{x} - \bar{\mathbf{x}}).$$

Note that this is linear in the elements of $\mathbf{x} - \bar{\mathbf{x}}$.

You compute the entire N by m matrix $\hat{\mathbf{F}}_{\text{reg}} = [\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_N]'$ of estimated scores as

$$\hat{\mathbf{F}}_{\text{reg}} = \tilde{\mathbf{X}} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{L}} = (\mathbf{X} - \mathbf{1}_N \bar{\mathbf{x}}') (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1} \hat{\mathbf{L}}$$

where

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_N \bar{\mathbf{x}}' = \text{matrix of residuals from mean}$$

Since the sample variance matrix \mathbf{S} is also an estimate of $\boldsymbol{\Sigma}$, an alternate estimate for $\boldsymbol{\beta}_{\text{reg}}$ is $\tilde{\boldsymbol{\beta}}_{\text{reg}} \equiv \mathbf{S}^{-1} \hat{\mathbf{L}}$, with corresponding estimated factor score matrix $\tilde{\mathbf{F}} = \tilde{\mathbf{X}} \mathbf{S}^{-1} \hat{\mathbf{L}}$.

When $\hat{\mathbf{\Psi}}$ and $\hat{\mathbf{L}}$ are fully converged maximum likelihood estimates, mathematics shows that $\hat{\mathbf{L}}' \mathbf{S}^{-1} = \hat{\mathbf{L}}' \hat{\boldsymbol{\Sigma}}^{-1} = \hat{\mathbf{L}}' (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1}$ and hence $\tilde{\mathbf{F}} = \hat{\mathbf{F}}_{\text{reg}}$.

Using the identity $(\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1} = \hat{\mathbf{\Psi}}^{-1} - \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{L}} (\mathbf{I}_m + \hat{\Delta})^{-1} \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1}$, where $\hat{\Delta} = \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{L}}$, another expression for $\hat{\boldsymbol{\beta}}_{\text{reg}}$ is

$$\hat{\boldsymbol{\beta}}_{\text{reg}} = \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{L}} (\mathbf{I}_m + \hat{\Delta})^{-1}.$$

Weighted least squares method

Factor scores estimated by the *weighted least squares method* are chosen in such a way as to result in small estimates $\hat{\boldsymbol{\epsilon}} = \mathbf{x} - \hat{\mathbf{L}} \hat{\mathbf{f}}$ of the unique factor scores $\boldsymbol{\epsilon}$. What is actually minimized is the weighted sum of squares $\sum_{1 \leq i \leq n} \hat{\psi}_i^{-1} \hat{\epsilon}_i^2$, using weights inversely proportional to the estimated uniquenesses $\hat{\psi}_i = \hat{V}[\epsilon_i]$.

Factor Score Estimation

The weighted least squares estimated coefficients are

$$\hat{\beta}_{LS} = \hat{\Psi}^{-1} \hat{L} \hat{\Delta}^{-1} = \hat{\beta}_{reg} (I_m + \hat{\Delta}^{-1}),$$

The matrix of estimated factor scores is

$$\hat{F}_{LS} = \tilde{X} \hat{\Psi}^{-1} \hat{L} \hat{\Delta}, \quad \tilde{X} = X - 1_N \bar{x}'$$

When $\hat{\Delta} = \hat{L}' \hat{\Psi}^{-1} \hat{L}$ is large, as will be the case when all $\hat{\psi}_i$ are small, $\hat{\beta}_{LS} \approx \hat{\beta}_{reg}$ and the two approaches lead to essentially the same estimated factor scores.

When factor analysis is based on a correlation matrix rather than a covariance matrix, you need to standardize each variable x_i ($x_i \rightarrow (x_i - \bar{x}_i)/\sqrt{s_{ii}}$) before computing scores. Alternatively, you can make a coefficient matrix $\hat{\beta}$ ($\hat{\beta}_{reg}$ or $\hat{\beta}_{LS}$) derived from a correlation matrix applicable directly to unstandardized x 's by the transformation

$$\hat{\beta} \rightarrow \text{diag}[1/\sqrt{s_{11}}, \dots, 1/\sqrt{s_{pp}}] \hat{\beta}.$$

Here is MacAnova output illustrating factor score computation for the matrix `bonedata` in file `cbbones.txt`.

```
Cmd> y <- read("","bonedata") # read from cbbones.txt
bonedata      276      6 format labels
) Bone measurements on n = 276 outbred female chickens, all in mm.
) Col. 1:  skull length
) Col. 2:  skull breadth
) Col. 3:  femur length (leg bone)
) Col. 4:  tibia length (leg bone)
) Col. 5:  humerus length (wing bone)
) Col. 6:  ulna length (wing bone)
Read from file "TP1:Stat5401:Data:cbbones.txt"

Cmd> n <- nrow(y); r <- cor(y) # sample size and correlation matrix
Cmd> print(r,format:"9.6f")
r:
      SklLngth  SklBrdth  FemLngth  TibLngth  HumLngth  UlnLngth
SklLngth  1.000000  0.583009  0.569111  0.602259  0.621119  0.602334
SklBrdth  0.583009  1.000000  0.515310  0.547599  0.583552  0.524505
FemLngth  0.569111  0.515310  1.000000  0.926105  0.877221  0.877453
TibLngth  0.602259  0.547599  0.926105  1.000000  0.873628  0.893610
HumLngth  0.621119  0.583552  0.877221  0.873628  1.000000  0.936879
UlnLngth  0.602334  0.524505  0.877453  0.893610  0.936879  1.000000

Cmd> results <- facanal(r,2,method:"mle") # 2 factor MLE extraction
Convergence in 20 iterations by criterion 2
estimated uniquenesses:
      SklLngth  SklBrdth  FemLngth  TibLngth  HumLngth  UlnLngth
0.59902      0.65349      0.12138  0.0028552  0.00015955  0.098034
```

Factor Score Estimation

unrotated estimated loadings:

	Factor 1	Factor 2
Sk1Lngh	0.6239	0.10827
Sk1Brdth	0.58523	0.063435
FemLngh	0.8848	0.30942
TibLngh	0.88481	0.46287
HumLngh	0.99964	-0.023485
UlnLngh	0.94034	0.13315

minimized mle criterion:
(1) 0.15079

facanal() creates side effect variables LOADINGS, PSI and CRITERION which match the output.

```
Cmd> list(LOADINGS,PSI,CRITERION)
CRITERION      REAL      1
LOADINGS       REAL      6      2      (labels)
PSI            REAL      6      (labels)

Cmd> rhohat <- LOADINGS %*% LOADINGS' + dmat(PSI)
Cmd> betahat_reg <- solve(rhohat,LOADINGS) # regression method coeffs
Cmd> betahat_reg # estimated regression methodcoefficients
```

	Factor 1	Factor 2	
Sk1Lngh	0.00015889	0.0022456	Computed using rhohat as estimated correlation matrix
Sk1Brdth	0.00013662	0.001206	
FemLngh	0.0011121	0.031671	Mainly determined by x4 and x5
TibLngh	0.047277	2.0141	
HumLngh	0.95582	-1.8287	
UlnLngh	0.0014633	0.016874	

Now use r rather than rhohat as estimated correlation matrix. You get the same coefficients because you are using fully converged MLE.

```
Cmd> solve(r, LOADINGS)
```

	Factor 1	Factor 2	
Sk1Lngh	0.00015889	0.0022456	Computed using r as estimated correlation matrix
Sk1Brdth	0.00013662	0.001206	
FemLngh	0.0011121	0.031671	
TibLngh	0.047277	2.0141	
HumLngh	0.95582	-1.8287	
UlnLngh	0.0014633	0.016874	

```
Cmd> # Standardize y since model estimated from correlation matrix
Cmd> x <- standardize(y) # standardized y;s
Cmd> # The covariance matrix of these standardized variates is r
Cmd> f_reg <- x %*% betahat_reg # Compute scores by regression methods
Cmd> # Now compute weighted least squares estimates of scores
Cmd> deltahat <- LOADINGS' %*% dmat(1/PSI) %*% LOADINGS
Cmd> deltahat
```

	Factor 1	Factor 2	
Factor 1	6554	-7.8132e-15	Note the large diagonal
Factor 2	-1.1713e-14	79.492	

Factor Score Estimation

```
Cmd> betahat_ls <- dmat(1/PSI) %*% LOADINGS %*% solve(deltahat)
Cmd> betahat_ls
      Factor 1      Factor 2
(1)  0.00015891  0.0022738
(2)  0.00013664  0.0012211
(3)  0.0011122  0.032069
(4)  0.047284    2.0394
(5)  0.95596    -1.8517
(6)  0.0014635  0.017086
```

Weighted LS factor score coeffs
They are very close to reg. method
scores because the diagonal elements
of deltahat are large

```
Cmd> f_ls <- x %*% betahat_ls # compute weighted least squares scores
Cmd> list(f_reg,f_ls) # sizes of matrices of factor scores
f_ls      REAL    276    2    (labels)
f_reg     REAL    276    2    (labels)

Cmd> tabs(f_reg,covar:T)
(1,1)      0.99985    2.0739e-16
(2,1)      2.0739e-16    0.98758

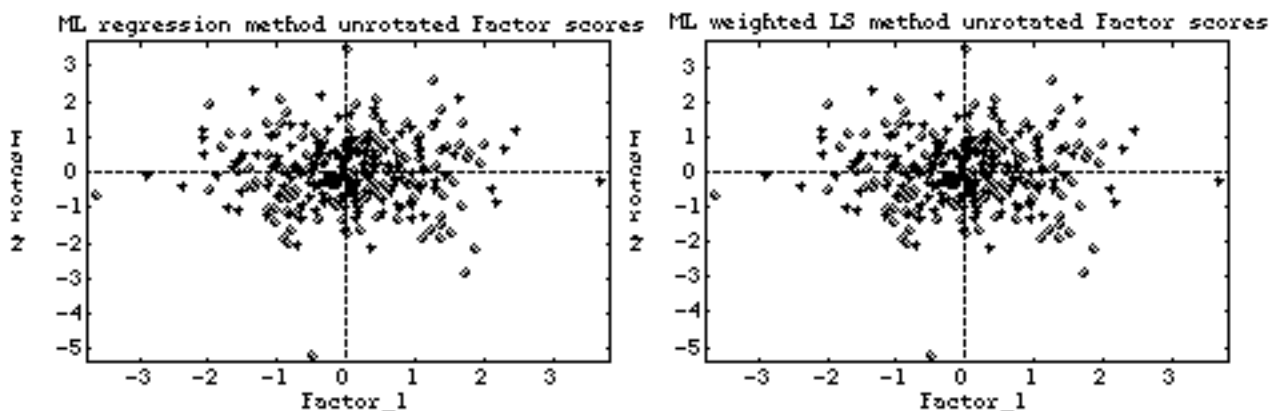
Cmd> tabs(f_ls,covar:T)
(1,1)      1.0002    1.0234e-16
(2,1)      1.0234e-16    1.0126
```

You can see that both sets of estimated scores have sample variances close to 1 and sample correlation = 0.

```
Cmd> cor(f_reg,f_ls)[run(2),-run(2)]
(1,1)      1    -4.7873e-17
(2,1)      3.822e-16    1
```

For these data, the two sets of scores are perfectly correlated with one another. Scatter plots will be identical:

```
Cmd> # Scatter plots of scores computed by both methods
Cmd> plot(Factor_1:f_reg[,1],Factor_2:f_reg[,2],symbols:"\1",\
      title:"ML regression method Factor scores")
Cmd> plot(Factor_1:f_ls[,1], Factor_2:f_ls[,2],symbols:"\1",\
      title:"ML weighted LS method Factor scores")
```

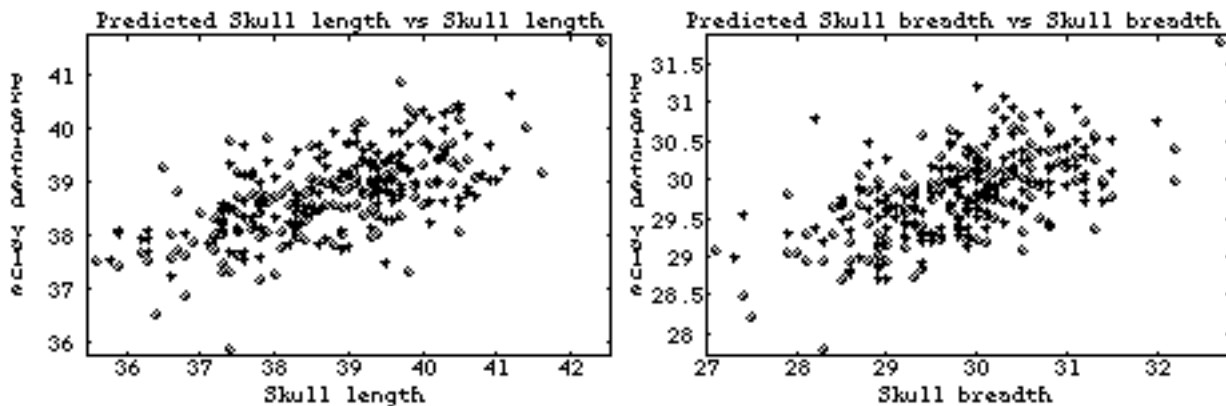


Factor Score Estimation

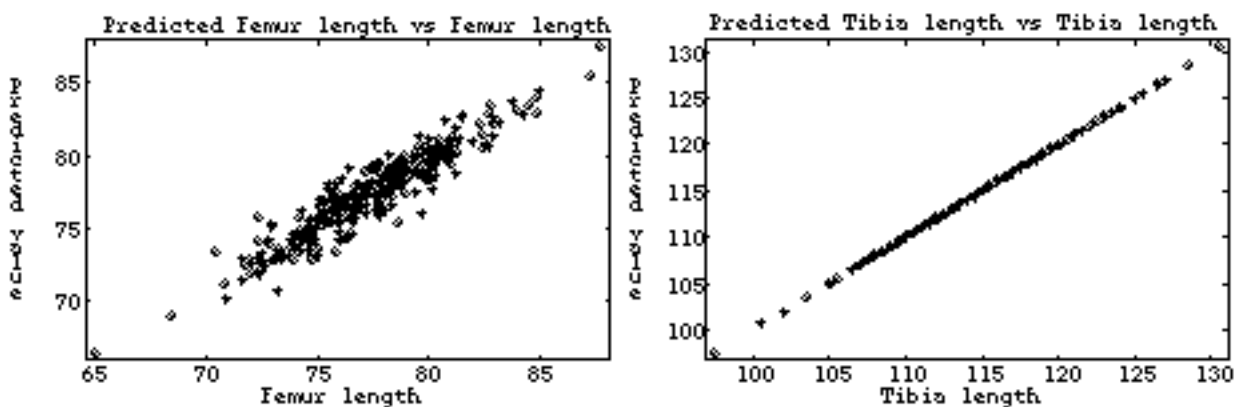
The two plots are indistinguishable. And both show an outlier.

Let's see how well the estimated factor scores predicted the actual bone sizes when weighted by the loadings and rescaled back to the original units.

```
Cmd> ybar <- describe(y,mean:T); sd <- describe(y,stddev:T)
Cmd> yhat <- ybar'+ (f_reg %*% LOADINGS') * sd'
Cmd> labels <- vector("Skull length","Skull breadth","Femur length",\
"tibia length","Humerus length","Ulna length")
Cmd> for(i,run(2)){ Skull length and breadth
  plot(y[,i], yhat[,i], symbols="\1",\
  xlab:labels[i],ylab:"Predicted value",title:\
  paste("Predicted",labels[i],"vs",labels[i]))
}
```

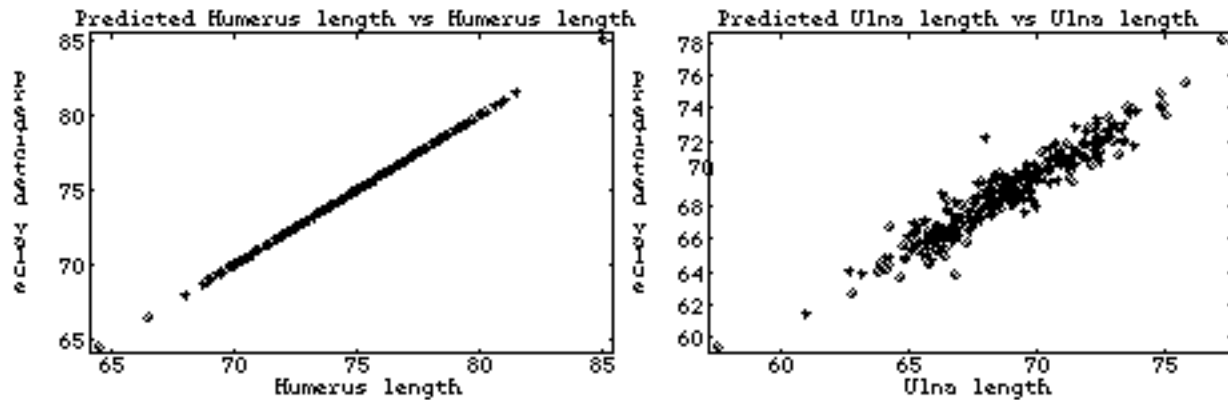


```
Cmd> for(i,run(3,4)){ Femur and Tibia
  plot(y[,i], yhat[,i], symbols="\1",\
  xlab:labels[i],ylab:"Predicted value",title:\
  paste("Predicted",labels[i],"vs",labels[i]))
}
```



Factor Score Estimation

```
Cmd> for(i,run(5,6)){ Humerus and Ulna
  plot(y[,i], yhat[,i], symbols="\1",\
  xlab:labels[i],ylab:"Predicted value",title:\
  paste("Predicted",labels[i],"vs",labels[i]))}
```



It appears that both tibia length (x_4) and humerus length (x_5) are almost perfectly predicted by the two factors. This is exactly what you should expect because $\hat{\psi}_4 = 0.00286$ and $\hat{\psi}_5 = 0.00016$. Moreover, from the coefficients used to compute the factor scores (`betahat_reg` and `betahat_ls`, see above), it is clear the most important variables in computing these scores are x_4 and x_5 .