

Today: Ch 9 Factor Analysis

Informal Intro: Recall the example $z_i \sim N(0,1)$

$$x_1 = z_1 + z_2$$

$$x_2 = z_1 + z_3$$

We found $\text{corr}(x_1, x_2) = .5$. But does this correlation tell us about the existence of the model with z_1 a common factor?

What if we have

$$x_1 = z_1 + z_2 + z_3$$

$$x_2 = z_1 + z_4$$

$$x_3 = z_2 + z_5$$

Does the correlation structure enable us to recover the "common factor" structure? We will return to this "simple" example after a more formal intro to Factor Analysis.

Factor Analysis: Goal is to find the underlying, but unobserved, variables (called factors) on which the measured variables depend. It turns out to be possible to infer how to estimate these factors from the observed variables. Of course, to do this, we need that the "factor model" described on p 479 is correct. We can try to see how well the model is fit by looking at residuals, but since the model is not really identifiable from the data, this has limited impact in practice. The value of a factor analysis is gauged in practice by the interesting hypotheses it suggests! Although there is a thriving field of confirmatory factor analysis, mostly pursued by psychologists and social workers, it is hard to separate the art from the science of it. We will mostly focus on exploratory factor analysis – trying to find factors and explain variability in data this way.

Model: p 279 box

Note this is for the p -variate rv X . The variability of each component is broken into two parts by $(L F + \epsilon)_i = (L F)_i + \epsilon_i$

L is $p \times m$ and F is $m \times 1$, and ϵ_i is $p \times 1$. Note that ϵ_i and F_i are assumed mean 0 and uncorrelated so there are no more parameters here except the p variances. (The covariance matrix of ϵ is assumed to be diagonal, Ψ .)

Note that we start out with (in the p -variate normal case) p means and $p(p-1)/2$ covariances, a total of $p(p+1)/2$ parameters.

The factor representation has p means, mp coefficients of L , and p variances of ϵ_i , for a total of $2p+mp$ or $p(m+2)$ parameters. The components of F are random.

So when $(p+1)/2 > m+2$, we are forcing the data into a constrained model. i.e when $m < ((p+1)/2)-2$ or in other words $m < (p-3)/2$, the factor model has fewer parameters than the full p-variate normal model.

p	m not more than	parameters in p-normal	parameters in largest m-factor model
-	-----	-----	-----
6	1	21	18
7	1	28	21
8	2	36	32
9	2	45	36
10	3	55	50

So for FA to be useful, we are looking for a fairly drastic reduction of dimensions from p to m

Like principal components, FA is most useful when

1. the number of factors is 1,2, or 3
2. we can give meaningful names to the factors.

Structure of the covariance matrix:

Note the assumptions about L and F.

F is a rv independent of ϵ . Moreover, F has means 0 and covariance = $I_{m \times m}$

The $I_{m \times m}$ assumption can be relaxed (with difficulty) but the means 0 is wlog.

And we assume ϵ is uncorrelated but can have variances $\neq 1$. i.e Ψ is diagonal

Looking at the $X = \mu + L F + \epsilon$ it may be intuited that we will have to use the correlation structure of X to estimate the split in variability between L F and ϵ . Recall that both F and ϵ are unobserved (that is, even when we have data for X).

Note also the direct results (p 480) that $Cov(X) = LL' + \Psi$ and $Cov(X,F) = L$

L is called the "factor loading" matrix since it shows how the factors are "loaded" onto the variables. Note the assumption (another assumption!) of linearity in the model.

More jargon: The variance of each of the p components of X is to be separated into a component attributable to the unobserved factors, and a component due to the unobserved "errors". The factor part of the variance is called the communality, and there is a communality h_i^2 for each component variable X_i . The "error" part ψ_i is called the "specific variance" since it has nothing in common with the "common factors" F. The variance of X_i can therefore be written as $h_i^2 + \psi_i$. Note that the communality h_i^2 must be the sum of the loadings² in the ith row of the L matrix.

Here is some output from the R factanal function for data generated by `mat.ex(n=100,p=8)`

```
> b=factanal(data,2,scores="regression")  
> b
```

Call:
factanal(x = data, factors = 2)

Uniquenesses:
V1 V2 V3 V4 V5 V6 V7 V8
0.570 0.643 0.581 0.534 0.005 0.470 0.545 0.643

Loadings:

	Factor1	Factor2
V1	0.570	0.324
V2	0.395	0.448
V3	0.434	0.480
V4	0.625	0.276
V5	0.234	0.970
V6	0.687	0.242
V7	0.450	0.502
V8	0.558	0.212

	Factor1	Factor2
SS loadings	2.102	1.907
Proportion Var	0.263	0.238
Cumulative Var	0.263	0.501

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 12.76 on 13 degrees of freedom.
The p-value is 0.466
> b=factanal(data,1,scores="regression")
> b

Call:

```
factanal(x = data, factors = 1)
```

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8
0.601	0.613	0.562	0.565	0.446	0.581	0.543	0.704

Loadings:

Factor1
V1 0.632
V2 0.622
V3 0.662
V4 0.660
V5 0.744
V6 0.648
V7 0.676
V8 0.544

	Factor1
SS loadings	3.384
Proportion Var	0.423

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 30.89 on 20 degrees of freedom.
The p-value is 0.0566

Comment:

We know there is only one factor in this data since we know that the "intra-class" correlation matrix (the one with all off-diagonal elements the same) was used to generate the data. But we are almost directed to use two factors! Lets try it again

p 483 If L is a solution of $\text{Cov}(X) = LL' + \Psi$, so is LT where T is orthogonal transformation (L is $p \times m$, T must be $m \times m$). Note we use "orthogonal" to mean $TT' = T'T = I$.

We try to choose T so that L is easy to interpret.

Estimation (Section 9.3) : Use standardized variables.

1. Principal Factor Method (use first m PCs, assign residual to ϵ - see box p 486. Principal Factor modification – iterate between communalities and adjusted correlation matrix – see pp 490-491)
2. Max Likelihood Method (assume normality – impose "harmless" constraint to obtain uniqueness) See p 492-493. Gives same answer for S and R (up to an orthog transformation). Is this surprising?

Example comparing 1. and 2. See Table 9.3 p 494. Also Table 9.4 p 496.

We will see that Factor Rotation can improve the MLE solution for interpretability. Section 9.4 discusses Factor Rotation.

9.5 Factor Score Estimation

We can estimate "factor scores" which give us m -tuples of estimated data for the unobserved factors. We can then plot these like we did for PCs.

Two methods:

1. Regression Method (pp 513-514) "Thomson's method"
2. Weighted Least Squares and MLE (pp 511-512) "Bartlett's" method.

Read the R help for "factanal"

Here is some output which uses the assumption that there is one factor (which we know is true for this simulated data).

```
> a=factanal(data,1,method="regression")
> a
Call:
```

```
factanal(x = data, factors = 1, method = "regression")
```

Uniquenesses:

```
V1 V2 V3 V4 V5 V6 V7 V8
0.505 0.468 0.472 0.584 0.565 0.569 0.467 0.555
```

Loadings:

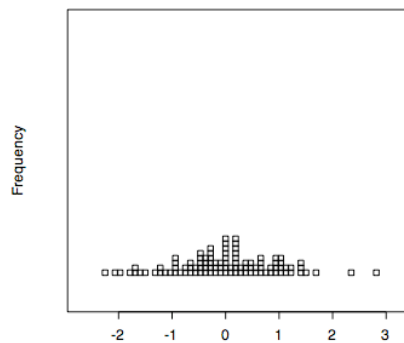
```
Factor1
V1 0.703
V2 0.730
V3 0.727
V4 0.645
V5 0.660
V6 0.657
V7 0.730
V8 0.667
```

```
Factor1
SS loadings 3.815
Proportion Var 0.477
```

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 26.69 on 20 degrees of freedom.
The p-value is 0.144

To get the scores

```
> my.dotplot(a$scores)
```



```
> factanal(data,1,method="Bartlett")
```

Call:

```
>a=factanal(x = data, factors = 1, method = "Bartlett")
```

```
>a
```

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8
0.505	0.468	0.472	0.584	0.565	0.569	0.467	0.555

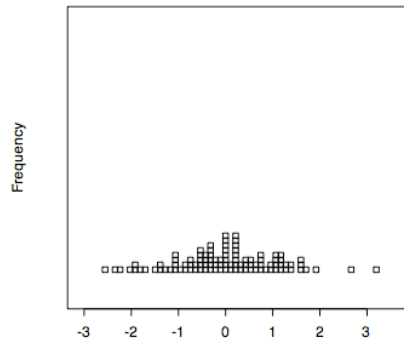
Loadings:

Factor1
V1 0.703
V2 0.730
V3 0.727
V4 0.645
V5 0.660
V6 0.657
V7 0.730
V8 0.667

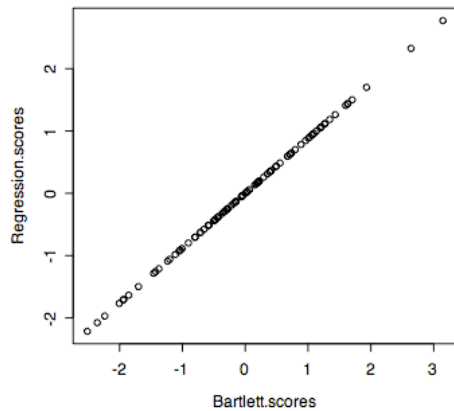
Factor1
SS loadings 3.815
Proportion Var 0.477

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 26.69 on 20 degrees of freedom.
The p-value is 0.144

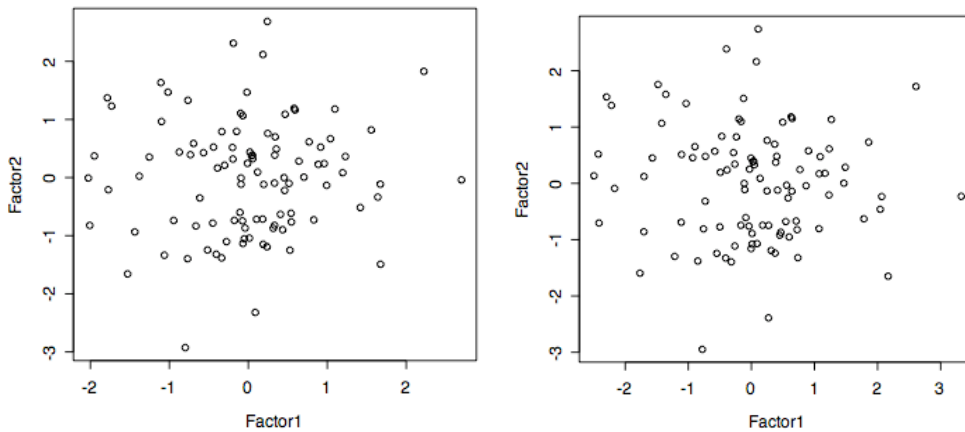
```
>  
> my.dotplot(a$scores)
```



How do the two methods of producing scores compare?



Lets see what scores look like for $m=2$ factors (we know second one is completely error)



left is regression method. right is Bartlett method.

Suggestion: These are **not** to hand in.

1. Pick a data set from Ch 9 that interests you. Perform a FA on it including the estimation of factor scores. Try to interpret the result. Talk to me if you need any clarification of the outcome.
2. Return to the introductory example with x_i and z_i . How much identifiability of the model does the correlation matrix allow?