Principal Component Analysis on multiple group data

Application in Dublin voter data

Kyusoon Kim

October 25, 2021

Seoul National University Multiscale Lab

Table of contents

Intro

Motivation

The standard PCA assumes that each observation has the same covariance structure. In other words, it is a one-sample method.

- Population PCA

$$X = (X_1, ..., X_p)^\top \sim (\mu, \Sigma), \quad \text{W.L.O.G} \quad \mu = 0$$

$$u_k = \underset{\substack{u \in \mathbb{R}^p, \|u\|=1 \\ u^\top u_j = 0,\, j=1,...,k-1}}{\operatorname{argmax}} \operatorname{Var}(u^\top \mathbf{X}) = \underset{\substack{u \in \mathbb{R}^p, \|u\|=1 \\ u^\top u_j = 0,\, j=1,...,k-1}}{\operatorname{argmax}} u^\top \Sigma u$$

- Sample PCA

$$\mathbf{X}^\top = [x_1, ..., x_n]_{p \times n}, \quad \text{W.L.O.G} \quad \bar{X} = 0$$

$$\hat{u}_k = \underset{\substack{u \in \mathbb{R}^p, \|u\|=1 \\ u^\top u_j = 0,\, j=1,...,k-1}}{\operatorname{argmax}} \widehat{\operatorname{Var}}(u^\top X) = \underset{\substack{u \in \mathbb{R}^p, \|u\|=1 \\ u^\top u_j = 0,\, j=1,...,k-1}}{\operatorname{argmax}} u^\top \left(\frac{1}{n-1} \mathbf{X}^\top \mathbf{X}\right) u$$

Motivation

- $S = \frac{1}{n-1}\mathbf{X}^\top\mathbf{X} = \frac{1}{n-1}\sum_{i=1}^{n} x_i x_i^\top \approx \Sigma?$
- If $x_1, \ldots, x_n \sim (\mu, \Sigma)$, the approximation is plausible.
- However, if there exist multiple groups with different covariance structures?
- How can we identify the different covariance structures and implement PCA for the multigroup data?

Goals

- Identify covariance groups for a given dataset.
- Implement PCA for the multigroup data.
- Compare the result with the standard PCA.

Methodology

Geographically Weighted PCA (GWPCA)

Geographically Weighted Principal Component Analysis (GWPCA) is a localized version of PCA, introduced by Fotheringham *et al.* (2002).

- In a GWPCA, instead of the assumption $x_i \sim (\mu, \Sigma)$, we assume $x_i | (u_i, v_i) \sim (\mu(u_i, v_i), \Sigma(u_i, v_i))$ to consider local effects.
- The elements of $\mu(u, v)$ may be interpreted as smoothed trends in the variables.
- The elements of $\Sigma(u, v)$ may be interpreted as second-order trends, showing the local variability.

Geographically Weighted PCA (GWPCA)

In a standard PCA, only the global Σ would be considered. However, in a GWPCA, each local variance-covariance matrix would be considered at each location.

- Consider a distance-decay kernel weight function W_i for x_i where W_i is a diagonal matrix of geographic weights (e.g. using a bi-square function : $(W_i)_{jj} = (1 - (d_{ij}/r)^2)^2$ if $d_{ij} \leq r$, $(W_i)_{jj} = 0$ otherwise).
- Suppose $\mathbf{X}$ is locally mean centered. (Subtract locally weighted mean $\bar{x}(u_i, v_i) = \frac{\sum x_j (W_i)_{jj}}{\sum (W_i)_{jj}}$)
- $\Sigma(u_i, v_i)$ can be estimated by using W_i.
  (i.e. $\Sigma(u_i, v_i) \sim \mathbf{X}^\top W(u_i, v_i)\mathbf{X} = \sum_{j=1}^{n} W(u_i, v_i)_{jj} x_j x_j^\top$)
- $\mathbf{X}^\top W_i \mathbf{X} = \mathbf{X}^\top W(u_i, v_i)\mathbf{X} = \mathbf{LVL}^\top | (u_i, v_i)$
- **Locally Weighted PCA (LWPCA)** is the same concept in attribute space, corresponding to GWPCA in geographic space.

Geographically Weighted PCA (GWPCA)

Usefulness

- For GWPCA, analyses and interpretations all take place locally yielding p eigenvalues, p sets of component loadings and p sets of component scores at each location.
- It makes possible to analyze spatial heterogeneity, the change of local structure.
  - How data dimensionality varies spatially?
  - How the original variables influence each spatially-varying component?

Challenges

- Overall interpretation is difficult.
- Optimal bandwidth should be chosen.
  - Small bandwidths lead to more rapid spatial variation.
  - Large bandwidths yield closer results to the global PCA.

Greater Dublin voter turnout data

Voter turnout and characteristics of social structure data in the 322 electoral divisions (EDs) forming Greater Dublin, Ireland for the 2002 General election and the 2002 census

- `DiffAdd` : percentage of the population in each ED who are one-year migrants (i.e. moved to a different address 1 year ago)
- `LARent` : percentage of the population in each ED who are local authority renters
- `SC1`: percentage of the high social class population in each ED
- `Unempl` : percentage of the unemployed population in each ED
- `LowEduc` : percentage of the population in each ED who are with little formal education
- `Age18_24` : percentage of the age group 18-24 in each ED
- `Age25_44` : percentage of the age group 25-44 in each ED
- `Age45_64` : percentage of the age group 45-64 in each ED
- `GenEl2004` : percentage of population in each ED who voted in 2002 election

GWPCA result

Globally standardized data was used.

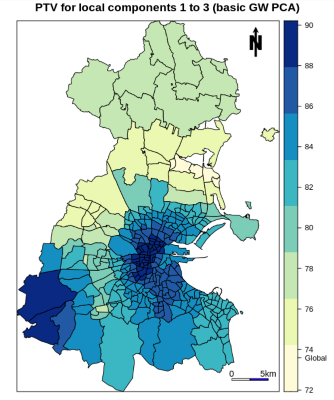

Figure 1: Proportion of total variance

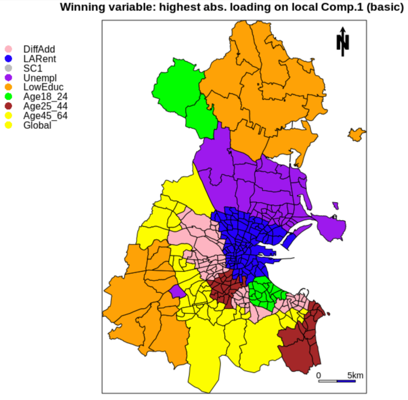

Figure 2: Winning variable

GWPCA result

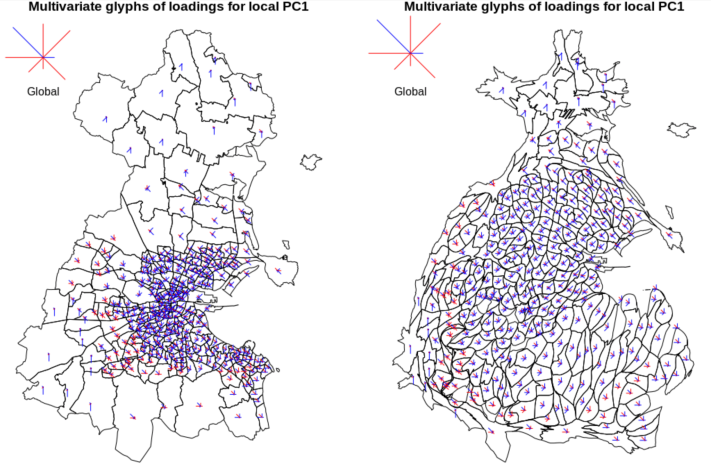

Figure 3: Multivariate glyphs of loadings for local PC1

Comparison of Global PCA and GWPCA

- Global PCA and GWPCA were implemented on the globally standardized data.
- It is known that using locally standardized data has several limitations (Harris *et al.* 2015).
- Global PC1 represents older residents (`Age45_64`) and Global PC2 affluent residents (`SC1`), which are a Greater Dublin-wide average.
- Compared to GWPCA result, it does not represents local social structure.

Comparison of Global PCA and GWPCA

- GWPCA tends to account for a higher PTV than Global PCA.
- Higher percentages are located in the south.
- Global PCA shows that `Age45_64` is dominant variable for PC1. However, there exists spatial variation of winning variable in the GWPCA result.
- The local-authority renting variable (`LARent`) plays an important part in explaining the local structure in central area and the low educational attachment (`LowEduc`) dominates in the northern and south-western area.

Dimension reduction?

Comparison between PC scores?

Principal component regression?

Let's reduce the number of location first!

Ward’s minimum variance method

- Harris *et al.* (2015) proposed to cluster the data by feeding the local loadings into Ward’s minimum variance method (Ward 1963).
- Ward’s method is a criterion applied in hierarchical cluster analysis.
- This method minimizes the total within-cluster variance by finding the pair of clusters that leads to minimum increase in total within-cluster variance after merging at each step.

Covariance structure clustering

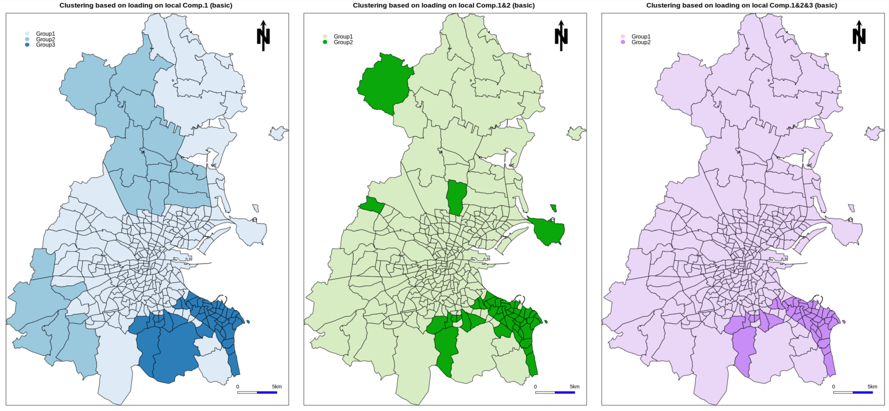

Figure 4: Ward clustering based on the local loadings

Now should we implement PCA on each group separately?

Can we say that those groups are completely unrelated?

If they are not same, are they different? How much?

Flury's Hierarchy

Compared to the univariate case, in multivariate cases, variances can have complex relationships. Flury (1988) defined possible relationships, which is called **Flury's hierarchy**.

Level	Model
1	Equality
2	Proportionality
3	CPC
4	Partial CPC (or CPC(q))
5	Heterogeneity

Table 1: Flury's Hierarchy

Flury’s Hierarchy

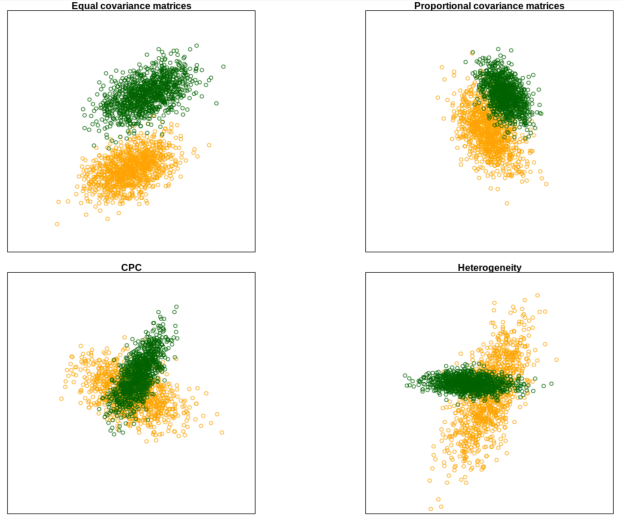

Figure 5: Flury’s Hierarchy ($p = 2$)

Common Principal Component Analysis (CPCA)

Common Principal component analysis (CPCA), introduced by Flury (1984), is a generalization of principal component analysis (PCA) to k groups.

- We have k groups and suppose that the random vectors $X^{(i)}$ of p components have zero mean and positive definite symmetric covariance matrices $\mathbf{\Sigma}_i$, $(i = 1, .., k)$.
- CPCA assumes that some rotation can diagonalize the covariance matrices simultaneously; that is, they have a common eigenvector structure but may have group-specific eigenvalues.
- Unlike the one sample PCA, no order of eigenvectors exists because the rank order of the eigenvalues may vary across groups.

Common Principal Component Analysis (CPCA)

- It might be reasonable in many practical situations. For example, if the same examination is taken by each student in many different schools, it seems plausible to assume that the basic sources of variation between students are similar in each school but differences in environments such as teaching practice and culture might place different emphases on these sources.
- A lot of biometrical examples satisfy this relationship (Flury 1984).

The hypothesis of common principal components is defined as

$$H_{cpc} : \boldsymbol{\Sigma}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}^\top,\ i = 1, \ldots k,\ \boldsymbol{B} : \text{orthonormal matrix},\ \boldsymbol{\Lambda}_i : \text{diagonal}$$

Let the p-variate random vectors $\boldsymbol{X}^{(i)}$ be independently distributed as $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $(i = 1, ..., k)$, where $\boldsymbol{\mu}_i \in \mathbb{R}^p$ and $\boldsymbol{\Sigma}_i$ are positive definite and symmetric. $\boldsymbol{S}_i$ and $\boldsymbol{N}_i = \boldsymbol{n}_i + 1$ represent sample covariance matrices and sample sizes respectively.

- Assume $\min_{1 \le i \le k} \boldsymbol{n}_i \ge p$
- $\boldsymbol{n}_i\boldsymbol{S}_i$ are independently distributed as $\boldsymbol{W}_p(\boldsymbol{n}_i, \boldsymbol{\Sigma}_i)$
- $\boldsymbol{L}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k) = \boldsymbol{C} \times \prod_{i=1}^{k} \operatorname{etr}\left(-\frac{\boldsymbol{n}_i}{2}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{S}_i\right) |\boldsymbol{\Sigma}_i|^{-\boldsymbol{n}_i/2}$

Let $\hat{\mathbf{\Sigma}}_1^{cpc}, \ldots, \hat{\mathbf{\Sigma}}_k^{cpc}$ be the MLE of $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k$ under H_{cpc} and $\boldsymbol{B} = \left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p\right)$, $\mathbf{\Lambda}_i = \mathrm{diag}\left(\lambda_{i1}, \ldots, \lambda_{ip}\right)$

- Using Lagrangian multiplier method, $\hat{\mathbf{\Sigma}}_i^{cpc} = \hat{\boldsymbol{B}}^{cpc} \hat{\mathbf{\Lambda}}_i^{cpc} \hat{\boldsymbol{B}}^{cpc\top}$ for $\hat{\boldsymbol{B}}^{cpc}, \hat{\mathbf{\Lambda}}_i^{cpc}$ satisfying

$$\hat{\boldsymbol{\beta}}_l^{cpc\top} \left(\sum_{i=1}^{k} \boldsymbol{n}_i \frac{\hat{\boldsymbol{\lambda}}_{il}^{cpc} - \hat{\boldsymbol{\lambda}}_{ij}^{cpc}}{\hat{\boldsymbol{\lambda}}_{il}^{cpc} \hat{\boldsymbol{\lambda}}_{ij}^{cpc}} \boldsymbol{S}_i \right) \hat{\boldsymbol{\beta}}_j^{cpc} = 0, \quad l, j = 1, \ldots, p, \; l \neq j$$

$$\hat{\boldsymbol{\lambda}}_{ij}^{cpc} = \hat{\boldsymbol{\beta}}_j^{cpc} \boldsymbol{S}_i \hat{\boldsymbol{\beta}}_j^{cpc}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, p$$

$$\hat{\boldsymbol{B}}^{cpc\top} \hat{\boldsymbol{B}}^{cpc} = \boldsymbol{I}_p$$

- Solution can be evaluated by **FG algorithm** (Flury and Gautschi 1986).

Hypothesis testing

- Unrestricted maximum of likelihood is obtained when $\boldsymbol{\Sigma}_i = \boldsymbol{S}_i$. The log-likelihood ratio statistic for testing H_{cpc} is

$$\boldsymbol{X}^2 = -2\log\frac{\boldsymbol{L}(\hat{\boldsymbol{\Sigma}}_1^{cpc},\ldots,\hat{\boldsymbol{\Sigma}}_k^{cpc})}{\boldsymbol{L}(\boldsymbol{S}_1,\ldots,\boldsymbol{S}_k)} = \sum_{i=1}^{k}\boldsymbol{n}_i\log\frac{|\hat{\boldsymbol{\Sigma}}_i^{cpc}|}{|\boldsymbol{S}_i|} = \sum_{i=1}^{k}\boldsymbol{n}_i\log\frac{|\mathrm{diag}\,\boldsymbol{F}_i|}{|\boldsymbol{F}_i|}$$

  where $\boldsymbol{F}_i = \hat{\boldsymbol{B}}^{cpc\top}\boldsymbol{S}_i\hat{\boldsymbol{B}}^{cpc}$.
- Under the null hypothesis, $\boldsymbol{X}^2$ is asymptotically $(\min\limits_{1\le i\le k}\boldsymbol{n}_i\to\infty)$ chi squared with $(k-1)p(p-1)/2$ degrees of freedom.
- If $\boldsymbol{X}^2 < \chi^2_{\boldsymbol{\alpha}}((k-1)p(p-1)/2)$, we can conclude the hypothesis of common principal components seems quite plausible at a given significance level $\boldsymbol{\alpha}$.
- If H_{cpc} holds, $\hat{\boldsymbol{B}}^{cpc}$ becomes a matrix of common loading vectors.

Common Principal Component Analysis (CPCA)

Remarks

1. If H_{cpc} holds, the dimension of the parameter space decreases from $kp + kp(p-1)/2$ to $kp + p(p-1)/2$, which leads to the smaller variance of estimates.
2. In the case k=1, CPCA reduces to the well known standard PCA.
3. No obvious fixed order of the columns of $\hat{B}^{cpc}$ need be given, since the rank order of the diagonal elements of the $\mathbf{\Lambda}_i$ is not necessarily the same for all $\mathbf{\Lambda}_i$.
4. However, if in all k groups the largest variances appear in the same CPC's, it might be able to dicard the CPC's that have small variances in all k groups simultaneously.

Common Principal Component Analysis (CPCA)

5. We can use CPC's as regressors in multiple regression even though they are not exactly uncorrelated in contrast to one-group situation.
6. The CPC model is mainly criticized because it is close to a method for simultaneous diagonalization, rather than a method for dimensionality reduction.
7. To solve this problem, Krzanowski's estimate and stepwise CPC estimate were introduced.

Krzanowski's estimate

Krzanowski (1984) suggested simple estimates of CPCs, say **Krzanowski's estimate**, which are easily obtainable.

- If H_{cpc} is true, $\boldsymbol{\Sigma}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}^\top$, $i = 1, \ldots k$. Thus $\boldsymbol{\Sigma} = \sum_{i=1}^{k} \boldsymbol{\Sigma}_i = \boldsymbol{B}(\sum_{i=1}^{k} \boldsymbol{\Lambda}_i)\boldsymbol{B}^\top = \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^\top$
- A simple estimate $\hat{\boldsymbol{B}}_{krz}^{cpc}$ can be obtained from the standard PCA of $\boldsymbol{S} = \sum_{i=1}^{k} \boldsymbol{S}_i$. Then we can get $\hat{\boldsymbol{\Lambda}}_{i,krz}^{cpc}$ and $\hat{\boldsymbol{\Sigma}}_{krz}^{cpc} = \hat{\boldsymbol{B}}_{krz}^{cpc}\hat{\boldsymbol{\Lambda}}_{i,krz}^{cpc}\hat{\boldsymbol{B}}_{krz}^{cpc\top}$
- Moreover, if H_{cpc} is true, $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{k} \frac{\boldsymbol{n_i}}{\boldsymbol{n}}\boldsymbol{\Sigma}_i = \boldsymbol{B}(\sum_{i=1}^{k} \frac{\boldsymbol{n_i}}{\boldsymbol{n}}\boldsymbol{\Lambda}_i)\boldsymbol{B}^\top = \boldsymbol{B}\boldsymbol{\Lambda}_0\boldsymbol{B}^\top$ where $\boldsymbol{n} = \sum_{i=1}^{k} \boldsymbol{n_i}$.
- Another estimate can be obtained from the standard PCA of $\boldsymbol{S}_0 = \sum_{i=1}^{k} \frac{\boldsymbol{n_i}}{\boldsymbol{n}}\boldsymbol{S}_i$.

Krzanowski's estimate

- Similar sets of estimates indicate that H_{cpc} is tenable, whereas very different sets indicate that it is not. This could be an informal test for H_{cpc}.
- It is natural to rank order the eigenvectors satisfying $\hat{\boldsymbol{\beta}}_{1,krz}^{cpc\top} \boldsymbol{S} \hat{\boldsymbol{\beta}}_{1,krz}^{cpc} \geq \ldots \geq \hat{\boldsymbol{\beta}}_{p,krz}^{cpc\top} \boldsymbol{S} \hat{\boldsymbol{\beta}}_{p,krz}^{cpc}$.
- This method is simple, intuitive and requires significantly less computational power.

Stepwise CPCA

Stepwise CPC estimation was introduced by Trendafilov (2010).

- The stepwise CPCs minimize the same objective function as the standard CPCs, but in a sequential manner.
- The standard CPC estimates minimize $\sum_{i=1}^{k} \boldsymbol{n_i} \log(\det(\operatorname{diag}(\boldsymbol{B}^\top \boldsymbol{S}_i \boldsymbol{B}))) = \sum_{i=1}^{k} \boldsymbol{n_i} \sum_{j=1}^{p} \log(\boldsymbol{\beta}_j^\top \boldsymbol{S}_i \boldsymbol{\beta}_j)$ subject to $\boldsymbol{B}^\top \boldsymbol{B} = I$.
- The stepwise CPC estimation solves this problem sequentially, *i.e.*, jth solution minimizes $\sum_{i=1}^{k} \boldsymbol{n_i} \log(\boldsymbol{\beta}^\top \boldsymbol{S}_i \boldsymbol{\beta})$ subject to $\boldsymbol{\beta}^\top \boldsymbol{\beta} = 1,\ \boldsymbol{\beta}$ is orthogonal to previous solutions.

Stepwise CPCA

- When the FG solution of the CPC problem gives eigenvalues simultaneously decreasing in all k groups, *i.e.*, $\boldsymbol{\beta}_1^\top S_i \boldsymbol{\beta}_1 \geq \ldots \geq \boldsymbol{\beta}_p^\top S_i \boldsymbol{\beta}_p$ for all i, stepwise CPC solutions can be expressed as a solution of the problem sequentially maximizing $\sum_{i=1}^{k} \boldsymbol{n_i} \log(\boldsymbol{\beta}^\top S_i \boldsymbol{\beta})$ subject to $\boldsymbol{\beta}^\top \boldsymbol{\beta} = 1,\ \boldsymbol{\beta}$ is orthogonal to previous solutions.
- In the above situation, the set of stepwise CPC solutions is identical to the FG solutions.
- In general case, stepwise CPCs are different to the FG solutions but eigenvalues produced by stepwise CPC method tend to decrease simultaneously (or nearly) in all groups.
- Thus stepwise CPCs can be used for dimension reduction.

If H_{cpc} does not hold? What can we do?

Partial Common Principal Component Analysis (PCPCA)

Flury (1987) proposed **partial common principal component analysis (PCPCA)**, where only a proportion of the loading vectors was assumed to be shared across and the rest to be individual-specific.

- In practical applications, we are often interested in a subset of $q \leq p$ components, provided that they recover most of the variability in each of the k groups simultaneously.
- H_{cpc} may be rejected due to its inappropriateness for the components to be discarded away even though the components that we are interested in may actually be common to all groups.

Partial Common Principal Component Analysis (PCPCA)

The hypothesis of partial common principal components is defined as

$$H_{cpc}(q) : \boldsymbol{\Sigma}_i = \boldsymbol{B}^{(i)} \boldsymbol{\Lambda}_i \boldsymbol{B}^{(i)\top} \ , i = 1, \dots k,$$

where $\boldsymbol{B}^{(i)} = \left(\boldsymbol{B}_1 : \boldsymbol{B}_2^{(i)} \right)$: orthonormal matrix, $\boldsymbol{\Lambda}_i$: diagonal.

- $\boldsymbol{B}_1$ has dimension $p \times q$ and $\boldsymbol{B}_2^{(i)}$ has dimension $p \times (p - q)$.
- $H_{cpc}(p-1)$ implies H_{cpc} since all $\boldsymbol{B}^{(i)}$ are orthogonal.
- No canonical order of the columns of the $\boldsymbol{B}^{(i)}$ need be given, since the rank order of the diagonal elements of the $\boldsymbol{\Lambda}_i$ is not necessarily the same for all $\boldsymbol{\Lambda}_i$.
- In this case, similarly to the case of CPCA, we can obtain the likelihood equations. However, solving those equations are extremely laborious and no simple method has been found yet.

Hypothesis testing

Flury (1987) suggested a simple procedure that can be used to obtain approximate maximum likelihood estimates.

- The approximate log-likelihood ratio statistic for testing $H_{cpc}(q)$ is

$$\boldsymbol{X}^2_{\boldsymbol{APP}} = -2\log\frac{L(\tilde{\boldsymbol{\Sigma}}_1,\ldots,\tilde{\boldsymbol{\Sigma}}_k)}{L(\boldsymbol{S}_1,\ldots,\boldsymbol{S}_k)} = \sum_{i=1}^{k} \boldsymbol{n}_i \log\frac{|\tilde{\boldsymbol{\Sigma}}_i|}{|\boldsymbol{S}_i|} = \sum_{i=1}^{k} \boldsymbol{n}_i \log\frac{|\mathrm{diag}\,\boldsymbol{F}_i|}{|\boldsymbol{F}_i|}$$

  where $\boldsymbol{F}_i = \tilde{\boldsymbol{B}}^\top \boldsymbol{S}_i \tilde{\boldsymbol{B}}$.
- Under the null hypothesis, $\boldsymbol{X}^2_{\boldsymbol{APP}}$ is asymptotically $(\min_{1\le i\le k} \boldsymbol{n}_i \to \infty)$ chi squared with $(k-1)\{p(p-1)-(p-q)(p-q-1)\}/2$ degrees of freedom.
- If $\boldsymbol{X}^2_{\boldsymbol{APP}} < \chi^2_{\boldsymbol{\alpha}}((k-1)\{p(p-1)-(p-q)(p-q-1)\}/2)$, we can conclude the hypothesis of partial common principal components seems acceptable at a given significance level $\boldsymbol{\alpha}$.
- If H_{cpc} holds, $\tilde{\boldsymbol{B}}$ becomes a matrix of common loading vectors.

Partial Common Principal Component Analysis (PCPCA)

Remarks

1. X^2_{APP} is always greater than the exact log-likelihood ratio statistic X^2. Therefore rejection using X^2_{APP} does not necessarily lead to rejection of $H_{cpc}(q)$.
2. In order to obtain approximate maximum likelihood estimates described above, the columns of $\hat{B}^{cpc}$ need to be arranged in the order of high probability of being common eigenvectors and fed into the algorithm.
3. Based on the idea of Krzanowski (1979), the order can be determined by calculating geometric means of the absolute vector correlations between all pairwise combinations of eigenvectors from each group and columns of $\hat{B}^{cpc}$.

Application

Greater Dublin voter turnout data

- When we used grouped data based on loadings on local PC1, none of hypotheses were accepted and **Heterogeneity model** had the lowest AIC, which means there is no relationship between covariance groups.

Model	Hypothesis Test	AIC	No.of.CPCs
Equality	Reject	586.77	8
Proportionality	Reject	426.24	8
CPC	Reject	257.43	8
CPC(6)	Reject	258.65	6
CPC(5)	Reject	247.65	5
CPC(4)	Reject	249.34	4
CPC(3)	Reject	216.47	3
CPC(2)	Reject	177.02	2
CPC(1)	Reject	158.94	1
Heterogeneity	-	144.00	0

Table 2: Flury's Hierarchy result for clustering based on loading on local PC1

Greater Dublin voter turnout data

- When we used grouped data based on loadings on local PC1 and PC2, CPC(1) hypothesis was only accepted and **CPC(1) model** had the lowest AIC, which supports CPC(1) model.

Model	Hypothesis Test	AIC	No.of.CPCs
Equality	Reject	215.50	8
Proportionality	Reject	193.14	8
CPC	Reject	152.71	8
CPC(6)	Reject	145.04	6
CPC(5)	Reject	146.60	5
CPC(4)	Reject	138.10	4
CPC(3)	Reject	143.07	3
CPC(2)	Reject	116.36	2
CPC(1)	**Accept**	**71.14**	1
Heterogeneity	-	72.00	0

Table 3: Flury's Hierarchy result for clustering based on loading on local PC1, PC2

Greater Dublin voter turnout data

- We can compare the estimated covariance and sample covariance. The result supports CPC(1) model.

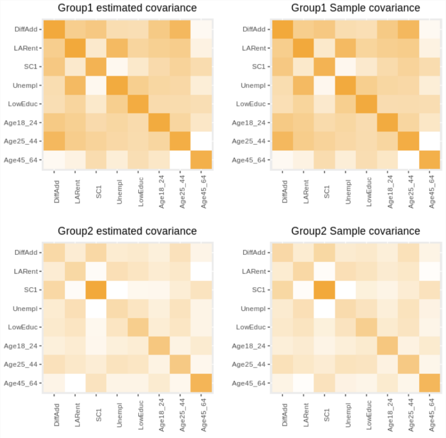

Figure 6: Comparison of covariance matrices

Greater Dublin voter turnout data

- We can compare the eigenvalues of the estimated covariance and sample covariance. CPC(1) model can be confirmed again.

Estimated		True	
Group1	Group2	Group1	Group2
3.011	1.458	3.012	1.461
2.191	1.042	2.192	1.051
1.006	0.581	1.012	0.589
0.931	0.376	0.926	0.378
0.517	0.358	0.517	0.352
0.281	0.123	0.281	0.123
0.259	0.091	0.259	0.081
0.172	0.038	0.171	0.033

Table 4: Comparison of eigenvalues

Greater Dublin voter turnout data

- When we used grouped data based on loadings on local PC1, PC2 and PC3, CPC(1) hypothesis was only accepted and **CPC(1) model** had the lowest AIC, which supports CPC(1) model.

Model	Hypothesis Test	AIC	No.of.CPCs
Equality	Reject	209.18	8
Proportionality	Reject	158.34	8
CPC	Reject	131.44	8
CPC(6)	Reject	133.44	6
CPC(5)	Reject	132.94	5
CPC(4)	Reject	126.50	4
CPC(3)	Reject	106.90	3
CPC(2)	Reject	74.05	2
CPC(1)	**Accept**	**71.82**	1
Heterogeneity	-	72.00	0

Table 5: Flury's Hierarchy result for clustering based on loading on local PC1 $\sim$ PC3

Greater Dublin voter turnout data

- We can compare the estimated covariance and sample covariance. The result supports CPC(1) model.

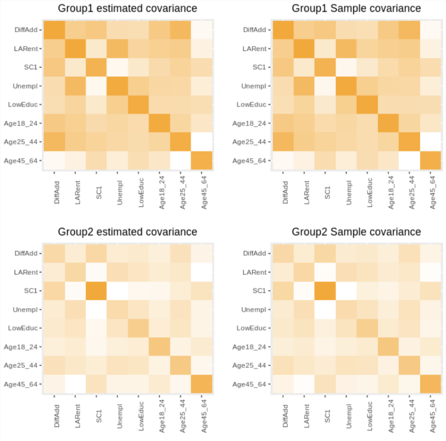

Figure 7: Comparison of covariance matrices

Greater Dublin voter turnout data

- We can compare the eigenvalues of the estimated covariance and sample covariance. CPC(1) model can be confirmed again.

Estimated		True	
Group1	Group2	Group1	Group2
2.993	1.280	2.994	1.281
2.185	1.074	2.185	1.078
0.996	0.483	0.998	0.500
0.915	0.372	0.913	0.378
0.517	0.228	0.517	0.223
0.298	0.125	0.298	0.125
0.275	0.083	0.275	0.062
0.185	0.015	0.185	0.014

Table 6: Comparison of eigenvalues

Conclusion

Summary

- Similar covariance groups can be identified for the geostatistical data via GWPCA.
- An appropriate model in Flury’s hierarchy can be choosed.
- However, the limitation of interpretation still remains.

Further studies

- We have to investigate how to interpret and use the results.
- The covariance clustering method needs to be elaborated and other methods might be considered.
- It is possible to apply the whole process on various data types. For example, river network data can be the subject of a slightly modified version by using the concept of flow direction and river distance. This method might be combined with the flow-directed PCA (Gallacher *et al.* , 2017) from the last seminar.
- Estimation methods need to be studied for the case of violation of the multivariate normality assumption.

Thank you for your attention!

Questions?

B. Flury.
Some relations between the comparison of covariance matrices and principal component analysis.
Computational Statistics & Data Analysis, 1:99–109, 1983.

B. K. Flury.
Two generalizations of the common principal component model.
Biometrika, 74:59–69, 1987.

B. N. Flury.
Common principal components in k groups.
Journal of the American Statistical Association, 79:892–898, 1984.

B. N. Flury and G. Constantine.
Algorithm as 211 : The f-g diagonalization algorithm.
Journal of the Royal Statistical Society. Series C (Applied Statistics), 34:177–183, 1985.

References ii

B. N. Flury and W. Gautschi.
An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form.
SIAM Journal on Scientific and Statistical Computing, 7:169–184, 1986.

A. S. Fotheringham, C. Brunsdon, and M. Charlton.
Geographically Weighted Regression.
John Wiley & Sons, Ltd, 2002.

I. Gollini, B. Lu, M. Charlton, C. Brunsdon, and P. Harris.
Gwmodel : An r package for exploring spatial heterogeneity using geographically weighted models.
Journal of Statistical Software, 63:1–50, 2015.

P. Harris, C. Brunsdon, and M. Charlton.
Geographically weighted principal components analysis.
International Journal of Geographical Information Science, 25:1717–1736, 2011.

P. Harris, A. Clarke, S. Juggins, C. Brunsdon, and M. Charlton.
Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set.
Geographical Analysis, 47:146–172, 2015.

W. J. Krzanowski.
Between-groups comparison of principal components.
Journal of the American Statistical Association, 74:703–707, 1979.

W. J. Krzanowski.
Between-group comparison of principal components—some sampling results.
Journal of Statistical Computation and simulation, 15:141–154, 1982.

References iv

W. J. Krzanowski.
Principal component analysis in the presence of group structure.
Journal of the Royal Statistical Society. Series C (Applied Statistics), 33:164–168, 1984.

P. T. Pepler.
The identification and application of common principal components.
PhD thesis, Stellenbosch University, December 2014.

N. T. Trendafilov.
Stepwise estimation of common principal components.
Computational Statistics & Data Analysis, 54:3446–3457, 2010.