# Common principal components

Theo Pepler
Unit for Biometry
Stellenbosch University

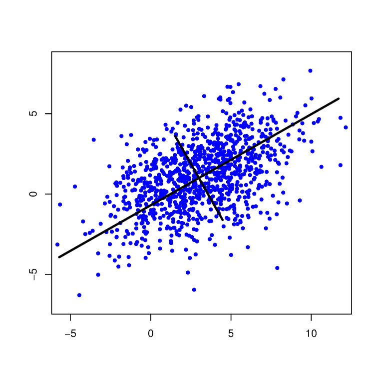
22 September 2014

## Principal component analysis (PCA)

$$\mathbf{\Sigma} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}'$$

Example:

$$\mathbf{\Sigma} = \begin{bmatrix} 0.87 & -0.49 \\ 0.49 & 0.87 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0.87 & 0.49 \\ -0.49 & 0.87 \end{bmatrix}$$



## Common principal components (CPC)

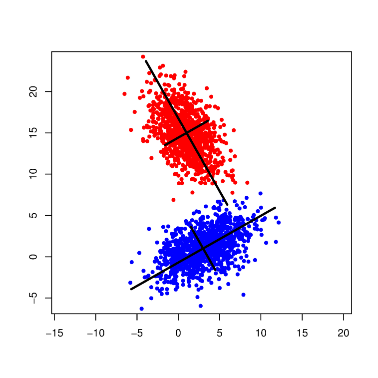
$$\mathbf{\Sigma}_1 = \mathbf{B}\mathbf{\Lambda}_1\mathbf{B}'$$

$$\mathbf{\Sigma}_2 = \mathbf{B}\mathbf{\Lambda}_2\mathbf{B}'$$

Example:

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 0.87 & -0.49 \\ 0.49 & 0.87 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0.87 & 0.49 \\ -0.49 & 0.87 \end{bmatrix}$$

$$\mathbf{\Sigma}_2 = \begin{bmatrix} 0.87 & -0.49 \\ 0.49 & 0.87 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} 0.87 & 0.49 \\ -0.49 & 0.87 \end{bmatrix}$$



## Simultaneous diagonalisation algorithms

- Flury-Gautschi (FG), (Flury and Gautschi, 1986)

$$\phi\left(\boldsymbol{L}_1, \ldots, \boldsymbol{L}_k; n_1, \ldots, n_k\right) = \prod_{i=1}^{k} \frac{\left[\det\left(\operatorname{diag}\boldsymbol{L}_i\right)\right]^{n_i}}{\left[\det\left(\boldsymbol{L}_i\right)\right]^{n_i}} \tag{1}$$

- JADE package (Cardoso and Souloumiac, 1996)

$$\min\left(\sum_{i=1}^{k}\sum_{j=1}^{p}\sum_{\substack{h=1\\h\neq j}}^{p} l_{jhk}^2\right) \tag{2}$$

- Stepwise CPC (Trendafilov, 2010)
  - estimates eigenvectors sequentially
  - ensures common eigenvectors have same rank order in all groups

## The Vermont Oxford Network (VON) data

- Birth weight (kg)
- Apgar score at 1 min (0–10)
- Apgar score at 5 mins (0–10)
- Gestational age (weeks)
- Head circumference (cm)
- Temperature (°C)

Source: Wikipedia (https://en.wikipedia.org/wiki/Neonatal_intensive_care_unit)

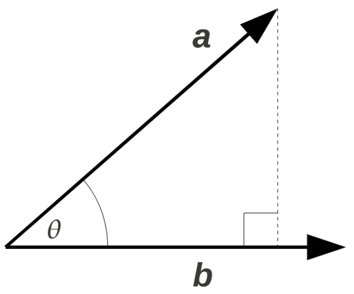
Regions:

- South Africa ($n_1 = 2921$)
- Namibia ($n_2 = 120$)

## AIC and Chi-square methods (Flury, 1988)

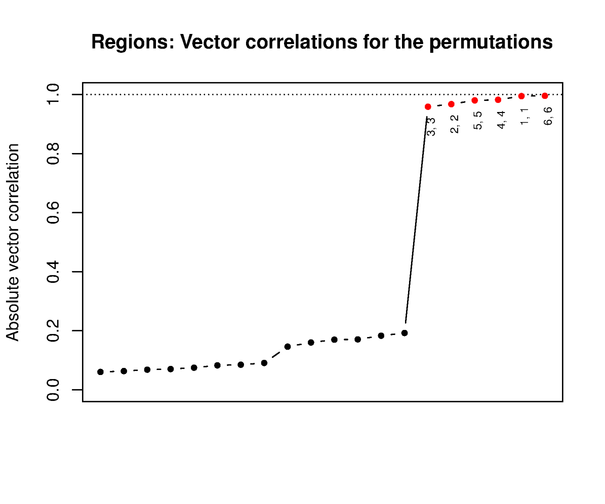
| **Model** | $\chi^2$ | **df** | $\frac{\chi^2}{\text{df}}$ | **AIC** |
|---|---|---|---|---|
| Equality | 5.99 | 1 | 5.99 | 85.77 |
| Proportionality | 10.09 | 5 | **2.02** | 81.78 |
| CPC | 2.06 | 1 | 2.06 | 81.69 |
| CPC(4) | 5.27 | 2 | 2.63 | 81.63 |
| CPC(3) | 12.87 | 3 | 4.29 | 80.37 |
| CPC(2) | 34.37 | 4 | 8.59 | 73.50 |
| CPC(1) | 15.13 | 5 | 3.03 | 47.13 |
| Heterogeneity | – | – | – | **42.00** |

## Vector correlations (Krzanowski, 1979)




$$\boldsymbol{a}'\boldsymbol{b} = \cos\theta$$

- Inspect *vector correlations* from pairwise combinations of all $p$ eigenvectors from the two groups.


Regions: Vector correlations for the permutations
Absolute vector correlation
0.0
0.2
0.4
0.6
0.8
1.0
3, 3
2, 2
5, 5
4, 4
1, 1
6, 6

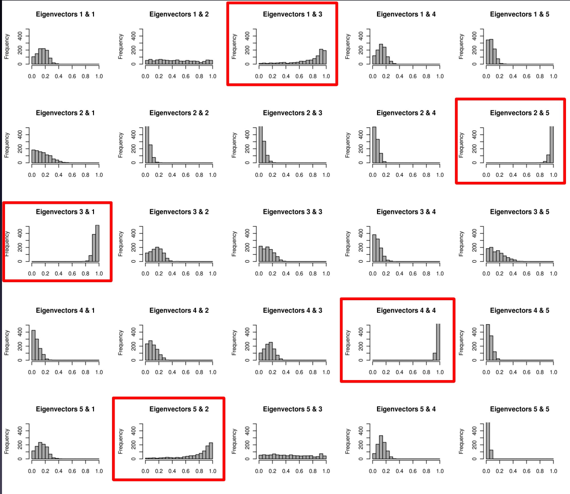
## Simulated CPC(5) data:

$k = 2$ groups

$p = 5$ variables

$n_1 = n_2 = 200$

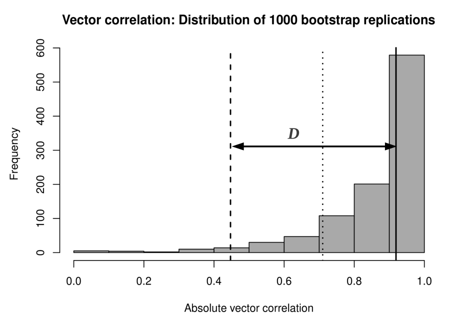
bootstrap reps $= 1000$



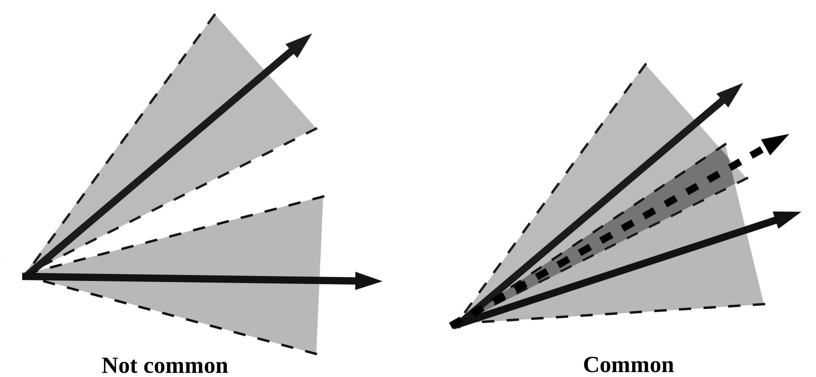
## Bootstrap vector correlation distribution (BVD)

Consider two eigenvectors to be common if:

1. median $> 0.71$
2. median $+ D \geq 1$



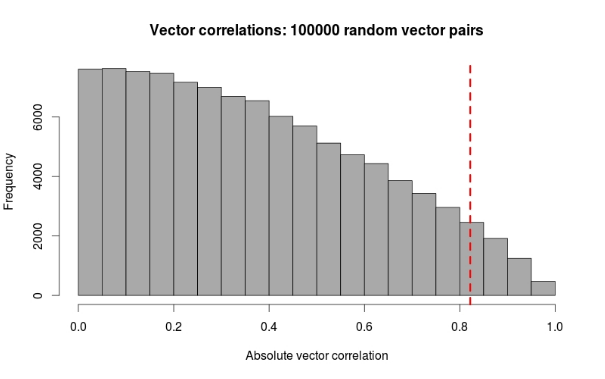
## Bootstrap confidence regions (BCR)



## Random vector correlations (RVC)

- adapted from Klingenberg and McIntyre (1998)

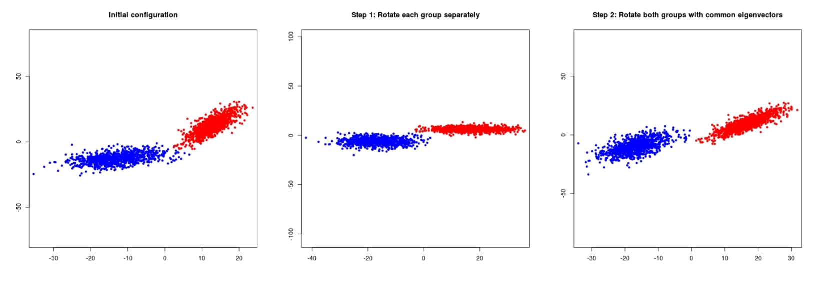
$$H_0: \text{ pair of eigenvectors are } \textit{not} \text{ common}$$



## Bootstrap hypothesis test (BootTest)

- adapted from Klingenberg (1996)

$$H_0 : \text{ pair of eigenvectors are common}$$




Twice rotated data for the $i^{th}$ group:

$$\boldsymbol{X}_i^\star = \boldsymbol{X}_i \boldsymbol{E}_i \boldsymbol{B}', \qquad i = 1, 2. \tag{3}$$

**Ensemble test**

Eigenvector pair considered equal (common) if majority vote of

- AIC
- BVD
- BCR
- RVC
- BootTest

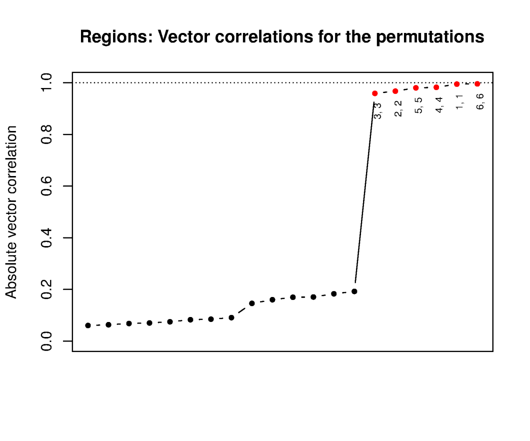
indicates it to be so.

## **Simulation results** ($p = 5$ variables)
Number of common eigenvectors correctly identified (%)

| | AIC | Chi$^2$ | BootTest | RVC | BVD | BCR | Ensemble |
|---|---|---|---|---|---|---|---|
| **Sample size** | | | | | | | |
| $n = 50$ | 33.1 | 27.0 | 26.1 | 30.0 | **33.9** | 25.6 | 32.5 |
| $n = 100$ | 34.2 | 30.7 | 26.4 | 32.2 | **36.1** | 29.4 | 35.0 |
| $n = 200$ | 43.1 | 28.1 | 33.1 | **47.2** | 44.4 | 35.3 | 46.1 |
| $n = 500$ | 43.3 | 34.8 | 46.1 | 53.3 | **56.4** | 49.4 | 54.2 |
| $n = 1000$ | 45.8 | 34.1 | 57.2 | 62.5 | 62.8 | 58.1 | **63.1** |
| **Distribution** | | | | | | | |
| Normal | 51.5 | 32.4 | 49.3 | 58.2 | **62.5** | 51.5 | 59.3 |
| Chi-squared | 43.5 | 34.2 | 39.5 | 52.0 | 51.0 | 42.7 | **52.5** |
| Multivariate $t$ | 24.7 | 26.2 | 24.5 | 25.0 | **26.7** | 24.5 | **26.7** |
| **Overall** | 39.9 | 31.0 | 37.8 | 45.1 | **46.7** | 39.6 | 46.2 |

## Application to the VON data (regions)

Ensemble test: 6 common eigenvectors



Covariance matrix estimators under the CPC model can:

- be less *biased* than when incorrectly assuming equality of the population covariance matrices, and
- be more *precise* than when incorrectly assuming that the population covariance matrices are unrelated.

## CPC estimator (Flury, 1988)

- $\boldsymbol{S}_i$ : unbiased sample covariance matrix estimator for $i^{th}$ group
- $\boldsymbol{B}$ : estimator of common eigenvector matrix

Estimator for $\boldsymbol{\Sigma}_i$ under the CPC model:

$$\boldsymbol{S}_{i(CPC)} = \boldsymbol{B}\boldsymbol{L}_i^0\boldsymbol{B}', \tag{4}$$

where

$$\boldsymbol{L}_i^0 = \text{diag}(\boldsymbol{B}'\boldsymbol{S}_i\boldsymbol{B}). \tag{5}$$

**Regularised CPC estimator**

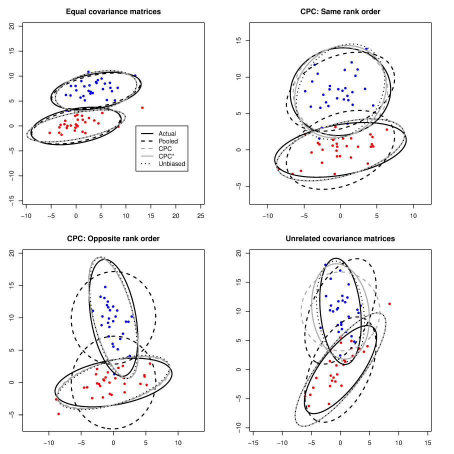
$$\boldsymbol{S}^{\star}_{i(CPC)} = \alpha_i \boldsymbol{S}_i + (1 - \alpha_i)\boldsymbol{S}_{i(CPC)}, \tag{6}$$

where $\alpha_i \in [0; 1]$ is the shrinkage intensity parameter.

Use cross-validation to find the value for $\alpha_i$ minimising a modified version of the Frobenius matrix norm on the training and validation samples.

## Covariance matrix shapes (95% confidence ellipses)

$k = 2$ populations, $p = 2$ variables



## Simulation results

Mean standardised modified Frobenius values (smaller is better):

| | Unbiased | CPC | CPC$^\star$ | Pooled |
|---|---|---|---|---|
| Full CPC | 0.269 | 0.372 | **0.192** | 0.792 |
| Half of eigenvectors common | 0.271 | 0.337 | **0.194** | 0.789 |
| Few common eigenvectors | 0.262 | 0.318 | **0.196** | 0.794 |
| Unrelated covariance matrices | 0.259 | 0.294 | **0.195** | 0.798 |

## VON data: Namibia ($n_2 = 120$)

$$S_2 = \begin{bmatrix} 0.87 & 0.62 & 0.45 & 3.02 & 3.39 & \mathbf{0.04} \\ 0.62 & 4.48 & 2.30 & 3.31 & 2.88 & \mathbf{-0.08} \\ 0.45 & 2.30 & 2.18 & 2.75 & 2.21 & \mathbf{-0.04} \\ 3.02 & 3.31 & 2.75 & 15.37 & 13.45 & \mathbf{-0.31} \\ 3.39 & 2.88 & 2.21 & 13.45 & 15.70 & \mathbf{0.05} \\ 0.04 & -0.08 & -0.04 & -0.31 & 0.05 & 0.50 \end{bmatrix}$$

$$S^{\star}_{2(\text{CPC})} = \begin{bmatrix} 0.87 & 0.57 & 0.44 & 3.16 & 3.30 & \mathbf{0.13} \\ 0.57 & 4.15 & 2.25 & 2.92 & 2.58 & \mathbf{0.07} \\ 0.44 & 2.25 & 2.31 & 2.45 & 2.02 & \mathbf{0.09} \\ 3.16 & 2.92 & 2.45 & 15.98 & 13.46 & \mathbf{0.21} \\ 3.30 & 2.58 & 2.02 & 13.46 & 15.22 & \mathbf{0.40} \\ 0.13 & 0.07 & 0.09 & 0.21 & 0.40 & 0.58 \end{bmatrix}$$

**CPC discriminant analysis**

Allocate a new observation, $x_{\text{new}}$, to the first group if

$$-\frac{1}{2}\boldsymbol{x}'_{\text{new}}(\boldsymbol{S}^{-1}_{1(\text{CPC})}-\boldsymbol{S}^{-1}_{2(\text{CPC})})\boldsymbol{x}_{\text{new}}+(\bar{\boldsymbol{x}}'_1\boldsymbol{S}^{-1}_{1(\text{CPC})}-\bar{\boldsymbol{x}}'_2\boldsymbol{S}^{-1}_{2(\text{CPC})})\boldsymbol{x}_{\text{new}} \geq c, \tag{7}$$

where

$$c = \frac{1}{2}\ln\left(\frac{|\boldsymbol{S}_{1(\text{CPC})}|}{|\boldsymbol{S}_{2(\text{CPC})}|}\right) + \frac{1}{2}(\bar{\boldsymbol{x}}'_1\boldsymbol{S}^{-1}_{1(\text{CPC})}\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}'_2\boldsymbol{S}^{-1}_{2(\text{CPC})}\bar{\boldsymbol{x}}_2), \tag{8}$$

otherwise allocate it to the second group.

## Simulation results

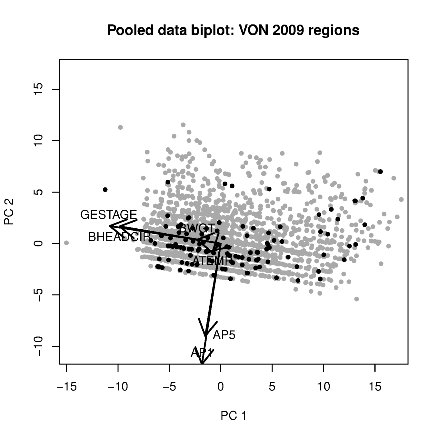
$n_1 = n_2$, $k = 2$ multivariate normal populations, $p = 10$ variables

| | | Misclassification error (%) | | | |
|---|---|---|---|---|---|
| **Structure** | $n_i$ | QDA | CPC | CPC$^\star$ | LDA |
| $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ | 50 | 37.79 | 31.65 | 31.67 | **30.68** |
| | 100 | 34.01 | 29.25 | 29.53 | **28.44** |
| | 200 | 31.27 | 28.25 | 28.35 | **27.70** |
| CPC | 50 | 22.96 | **16.50** | 16.81 | 30.43 |
| (similar | 100 | 18.12 | **14.93** | 15.08 | 28.80 |
| rank orders) | 200 | 15.89 | **14.13** | 14.26 | 27.49 |
| CPC | 50 | 3.31 | **2.15** | 2.22 | 21.55 |
| (Opposite | 100 | 2.41 | **1.95** | 1.97 | 18.31 |
| rank orders) | 200 | 1.99 | **1.84** | 1.85 | 16.56 |
| Unrelated | 50 | 8.66 | 8.14 | **6.94** | 32.94 |
| covariance | 100 | 5.85 | 7.15 | **5.57** | 30.80 |
| matrices | 200 | **4.89** | 6.95 | 4.92 | 29.76 |

## **VON data: Regions** (6 common eigenvectors)

Misclassification errors:

- QDA $= 25.2\%$
- LDA $= 25.4\%$
- CPC $= 21.2\%$
- CPC$^\star$ $= 22.9\%$

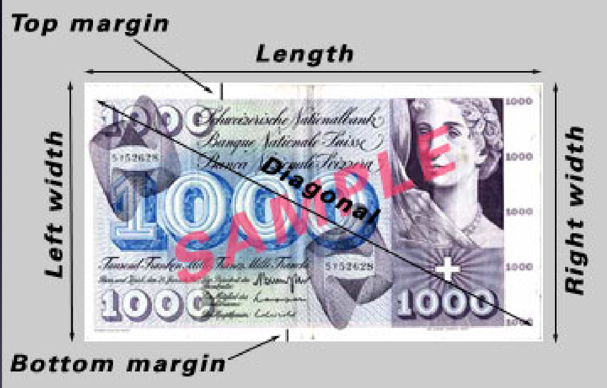


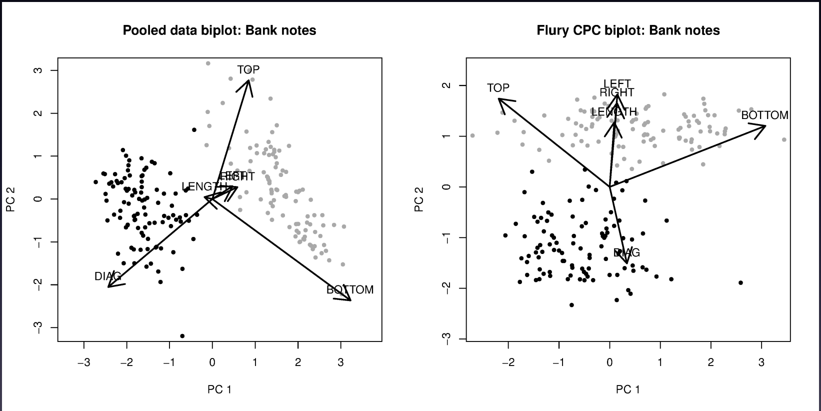
**Biplots for grouped data**

- overall quality of display
- *between-group* variation
- *within-group* variation
- representation of *variables*
    - adequacy
    - mean standard predictive error (MSPE), (Rui Alves, 2012)
- representation of *observations*
    - sample predictivities (Gower et al., 2011)

## Swiss bank notes (Flury, 1988)

Genuine notes ($n_1 = 100$), Forged notes ($n_2 = 100$)




Pooled data biplot: Bank notes
TOP
LENGTH
LEFT
RIGHT
DIAG
BOTTOM
PC 1
PC 2
Flury CPC biplot: Bank notes
TOP
LEFT
RIGHT
LENGTH
BOTTOM
DIAG
PC 1
PC 2

## Quality measures for 2D biplot of Bank Notes data

| | Overall | Within | Between | MSPE | Sample predictivities |
|---|---|---|---|---|---|
| Pooled S | 0.42 | **0.72** | 0.21 | 0.80 | 0.35 |
| Pooled data | **0.88** | 0.70 | **1.00** | **0.44** | **0.85** |
| Flury | 0.65 | 0.70 | 0.61 | 0.75 | 0.62 |
| Stepwise CPC | 0.35 | 0.71 | 0.10 | 0.75 | 0.31 |
| JADE | 0.44 | 0.71 | 0.26 | 0.79 | 0.38 |

## CPC regression

$$Y = \beta_0 + \beta_1 Z_1 + \ldots + \beta_q Z_q + \epsilon, \qquad 1 \leq q \leq p, \tag{9}$$

where $Z_j$ is the $j^{th}$ common principal component,

$$Z_j = \mathsf{b}_{j1} X_{i1} + \mathsf{b}_{j2} X_{i2} + \cdots + \mathsf{b}_{jp} X_{ip}, \qquad j = 1, \ldots, p; i = 1, \ldots, k. \tag{10}$$

Add dummy variables to design matrix to indicate group membership: Allows fitting regression models with different intercepts and/or partial slopes for the different groups.

## CPC regression: Conclusions

- CPC and PC regression provide very similar fits
- regression on full set of CPCs gives same fit as OLS regression
- CPC is covariance matrix model: not aimed at predicting a response
- PLS regression will give better results than CPC

## References

Cardoso, J.-F. and Souloumiac, A. (1996). Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164.

Diaconis, P. and Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, 248(5):116–130.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

Flury, B. (1988). *Common Principal Components and Related Multivariate Models*. Wiley, 1988.

Flury, B.N. and Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing*, 7(1):169–184.

Friedman, J.H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175.

Gower, J.C., Gardner-Lubbe, S. and Roux, N.L. (2011). *Understanding Biplots*. Wiley.

Hastie, T.J., Tibshirani, R.J. and Friedman, J.J.H. (2009). *The elements of statistical learning*. Springer-Verlag.

Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall.

Klingenberg, C.P. (1996). Multivariate allometry. *NATO ASI SERIES A LIFE SCIENCES*, 284:23–50.

Klingenberg, C.P. and McIntyre, G.S. (1998). Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*, 52(5):1363–1375.

# References (continued)

Krzanowski, W.J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74(367):703–707.

Rui Alves, M. (2012). Evaluation of the predictive power of biplot axes to automate the construction and layout of biplots based on the accuracy of direct readings from common outputs of multivariate analyses: 1. Application to principal component analysis. *Journal of Chemometrics*, 26(5):180–190.

Trendafilov, N.T. (2010). Stepwise estimation of common principal components. *Computational Statistics and Data Analysis*, 54(12):3446–3457.