Index of 2-L

Practical Information

WELCOME (again) to all of you...

Registration for graduate students for VHM 881 (Directed studies) is open at the Registrar's Office,

- for credit (1 credit),
- as an audit (no credits, but still listed on your transcript).

Other participants are not required to register, but you will not get any formal credit for participation.

Textbooks — added info (at webpage) for textbook suggested by Siyuan.

Today's lecture:

- gentle introduction to matrix notation and algebra (Manly: Chapter 2; TF Appendix A),
- some graphical tools to display multivariate data (Manly: Chapter 3, only 3rd edition),
- further introductions to datasets (Ibrahim, others?).

Matrix definitions

An $m \times n$ matrix $\mathbf{A}$ is a collection of mn values organized in m rows and n columns,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & & & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix}$$

- a square matrix has equally many rows and columns: $m = n$,
- a row vector, say $\mathbf{r} = (r_1\ r_2\ \ldots\ r_n)$, has only one row: $m = 1$,
- a column vector has only one column: $n = 1$,
- the transpose matrix $\mathbf{A}'$ or $\mathbf{A}^t$ results from interchanging the rows and columns (i.e., $a^t_{ij} = a_{ji}$),
  * a symmetric matrix $\mathbf{A}$ satisfies: $\mathbf{A}^t = \mathbf{A}$,
  * the transpose of a row vector is a column vector, and vice versa,
- a zero matrix $\mathbf{0}$ has all elements $= 0$,
- a one matrix $\mathbf{1}$ has all elements $= 1$,
- a square diagonal matrix $\mathbf{D}$ has only 0s outside the diagonal (i.e., $d_{ij} = 0$ when $i \neq j$),
- the unity matrix $\mathbf{I}$ is a diagonal matrix with 1s in the diagonal (i.e., $i_{ij} = 1$ when $i = j$, and $i_{ij} = 0$ otherwise),
- the trace of a square matrix is the sum of the diagonal elements,

$$\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{nn}.$$

Matrix Operations

Let $\mathbf{A}$ and $\mathbf{B}$ be matrices, and let k be a (real-valued) number, often termed a scalar, so as to distinguish from vectors and matrices.

- addition ($\mathbf{A}+\mathbf{B}$) and subtraction ($\mathbf{A}-\mathbf{B}$) of matrices occurs for each element (at its matrix position),
  - $\mathbf{A}$ and $\mathbf{B}$ must have the same dimensions,
  - e.g.: $(\mathbf{A}+\mathbf{B})_{11} = a_{11} + b_{11}$; $(\mathbf{A}-\mathbf{B})_{32} = a_{32} - b_{32}$,
- scalar multiplication by k ($k\mathbf{A}$) also occurs for each element,
- matrix multiplication ($\mathbf{A} \cdot \mathbf{B}$ or $\mathbf{A}.\mathbf{B}$ or $\mathbf{A} \times \mathbf{B}$, but most commonly just $\mathbf{AB}$) is *not* defined by multiplying pairs of matrix elements,
  - requires compatible matrix dimensions:
    no. columns of $\mathbf{A}$ must equal no. of rows of $\mathbf{B}$, say $\mathbf{A}$ $(m\times n)$ and $\mathbf{B}$ $(n\times p)$,
  - for $\mathbf{A}$ $(m\times n)$ and $\mathbf{B}$ $(n\times p)$, $\mathbf{A}.\mathbf{B}$ has dimension $(m\times p)$,
  - the (i,k)th element of $\mathbf{A}.\mathbf{B}$ is the sum
    $$(\mathbf{A}.\mathbf{B})_{ik} = \sum_j a_{ij}\, b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}\,,$$
  - it is *not* generally true that $\mathbf{A}.\mathbf{B} = \mathbf{B}.\mathbf{A}$;
    note that $\mathbf{A}.\mathbf{B}$ and $\mathbf{B}.\mathbf{A}$ are not even of the same dimensions, unless $\mathbf{A},\mathbf{B}$ are square matrices.

Matrix Inversion and Quadratic Form

Matrix inversion = the matrix version of division but more complicated than for numbers.

Definition: a square matrix $\mathbf{A}$ ($n \times n$) has inverse $\mathbf{A}^{-1}$ ($n \times n$) if the matrices satisfy: $\mathbf{A}.\mathbf{A}^{-1} = \mathbf{A}^{-1}.\mathbf{A} = \mathbf{I}$.

Example: the 2×2 matrix inverse of $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

- exists exactly when the determinant $\Delta = a_{11}a_{22} - a_{21}a_{12}$ (usually termed $|\mathbf{A}|$) is nonzero,
- can be computed as $\mathbf{A}^{-1} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} / \Delta$.

General definitions and results for $\mathbf{A}$ ($n \times n$):

- $\mathbf{A}$ has an inverse $\mathbf{A}^{-1}$ exactly when the determinant[1] $|\mathbf{A}|$ is non-zero,
- if $\mathbf{A}^{-1}$ does not exist, then $\mathbf{A}$ is called singular,
- if $\mathbf{A}^{-1}$ exists and $\mathbf{A}^{-1} = \mathbf{A}^t$, then $\mathbf{A}$ is called orthogonal.

A quadratic form Q for a symmetric matrix $\mathbf{A}$ ($n \times n$) is a linear function of a column vector $\mathbf{x} = (x_1\ x_2\ \ldots\ x_n)^t$,

$$Q = Q(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

[1] For square matrices larger than (2×2), the determinant $|\mathbf{A}|$ has an explicit but complex form, and is best calculated using mathematical/statistical software.

Mean Vectors and Covariance Matrices

Multivariate normal (MVN) distribution for a p-dimensional column vector $\mathbf{X} = (X_1\ X_2\ \ldots X_p)^t \sim \text{MVN}(\mu, \mathbf{\Sigma})$, where

* $\mu = (\mu_1\ \mu_2\ \ldots\ \mu_p)^t$ is a p-dimensional column vector of means ($\text{E}X_j = \mu_j$),
* $\mathbf{\Sigma} = (\sigma_{jk})_{jk}$ is a $(p \times p)$ positive definite[2] matrix of the pairwise covariances,[3]

has p-dimensional probability density function

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\mu)^t\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right)$$

Estimation of mean and covariance:
From i.i.d.[4] multivariate observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ we can estimate the mean vector and covariance by

* $\hat{\mu} = \bar{\mathbf{X}} = (\bar{X}_1\ \bar{X}_2 \ldots \bar{X}_p)^t$, the column vector of means for each variable,
* $\hat{\mathbf{\Sigma}} = \mathbf{S} = (s_{jk})_{jk}$, the empirical covariance matrix with the $s_{jk} = r_{jk}s_js_k$ defined in terms of the usual correlation coefficient between X_j and X_k and the standard deviations for X_j and X_k, respectively.

---

[2] A symmetric matrix $\mathbf{A}$ is positive definite if: $Q(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x} > 0$ for any non-zero $\mathbf{x}$, or equivalently if $|\mathbf{A}| > 0$.

[3] Formally, $\sigma_{jk} = \text{E}(X_jX_k) - (\text{E}X_j)(\text{E}X_k)$, and $\sigma_{jj} = \text{Var}(X_j)$; also, the correlation matrix $\mathbf{R} = (\rho_{jk})_{jk}$ gives the off-diagonal correlations $\rho_{jk} = \sigma_{jk}/(\sigma_{jj}\sigma_{kk})$.

[4] i.i.d. stands for: independent and identically distribution observations.

Eigenvalues and Eigenvectors

For a symmetric matrix $\mathbf{A}$ $(n \times n)$, a scalar λ and a column vector $\mathbf{x}$ $(n \times 1)$, we consider the equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \text{or} \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0},$$

which can also represented as the set of equations,

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= \lambda x_1, \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= \lambda x_2, \\
\vdots \quad &\quad \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= \lambda x_n.
\end{aligned}$$

The possible solutions[5] reflect fundamental properties of $\mathbf{A}$:

- the eigenvalues (or characteristic roots) λ of which there can be at most n distinct values, all of which are > 0 for a positive definite matrix, and which satisfy

$$\mathrm{tr}(\mathbf{A}) = \lambda_1 + \ldots + \lambda_n,$$

* the eigenvectors x corresponding to each λ express specific "directions" in the matrix.

Consider now these are general matrix concepts applied to a covariance matrix $(\boldsymbol{\Sigma})$:

- $\mathrm{tr}(\boldsymbol{\Sigma}) = \sigma_{11} + \ldots + \sigma_{pp} = \lambda_1 + \ldots + \lambda_n$,
  — the eigenvalues decompose the total variance. . . ,
- the eigenvectors correspond to specific directions of the variation in the data.

---

[5] Calculation of solutions λ and $\mathbf{x}$ to the eigenvalue problem are not straightforward to establish; TF work through a (2×2) example.

Graphing Multivariate Data

Fundamental challenge: getting beyond 2-d (or perhaps 3-d) scatter-plots to display relationships between variables.

Example data sets (Manly): sparrows, dogs.

Extension to 3-d plots:

- use 3rd axis (depths), and rotate points,
- or create (interpolated/smooth) surfaces for 3rd variable,

+ both ideas useful, but not a grand extension...

First idea: pairwise scatter plots ⇒ matrix or draftman's plot,

+ can show relations between a limited no. of variables,

− difficult to see similarity between observations.

Next idea: represent variable's values symbolically in displays involving multiple features, e.g. faces (Chernoff) or stars.

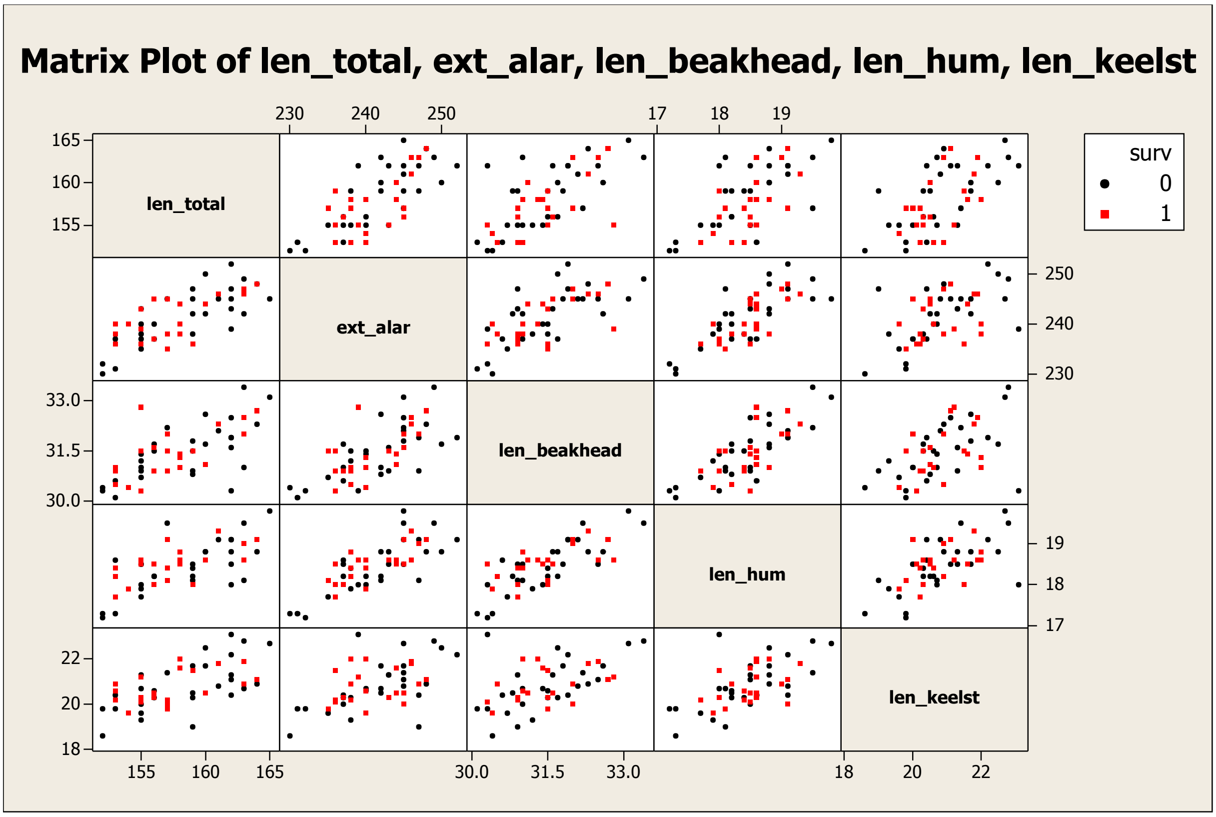
Another idea: position multiple measures per subject on x-axis at suitable (arbitrary) points and join values for same subject by line (profile plot), or position them as multiple bars beside each other (bar graph),

+ can illustrate grouping and distance between subjects,

− requires measures on comparable scales, difficult to see relation between variables.

Final comment: one key purpose of multivariate methods is to *reduce the dimension of information* across many variables into a few ("index"; Manly) variables, e.g. for the purpose of facilitating graphical exploration.

Sparrows: Matrix Plot

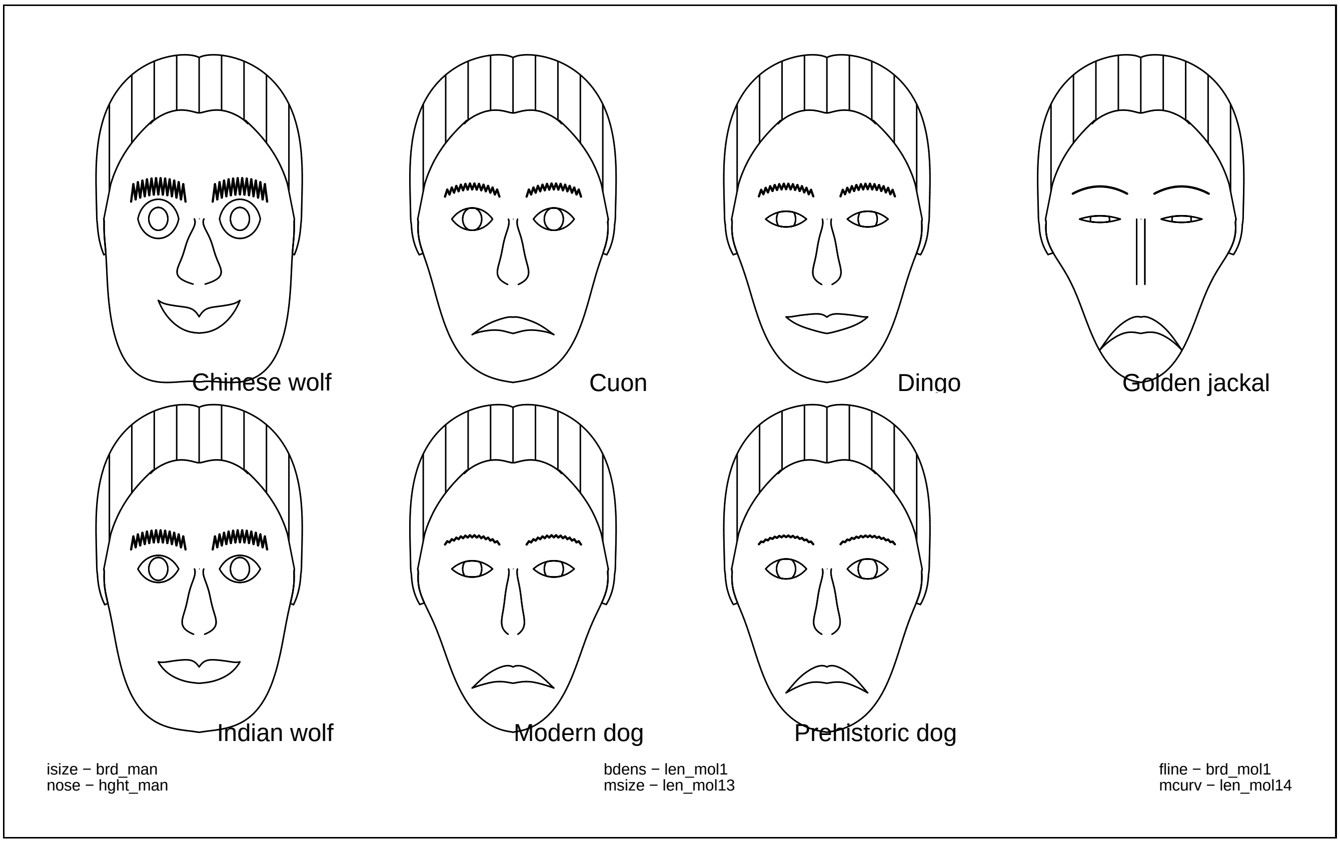


- all variables seem positively and approx. linearly associated, to similar degree,
- no obvious differences between survivors and morts,
- no apparent strong outliers.

Dogs: Chernoff plot

Idea[6]: use different facial features (up to 18) to represent variables,

- one selected feature per variable, scaled so that low/high values correspond to extreme features,
- some arbitrariness/flexibility in allocation of features to variables,
- not clear that visual impression is not affected substantially by allocation of features to variables (other symbols, e.g. stars, may be more robust).



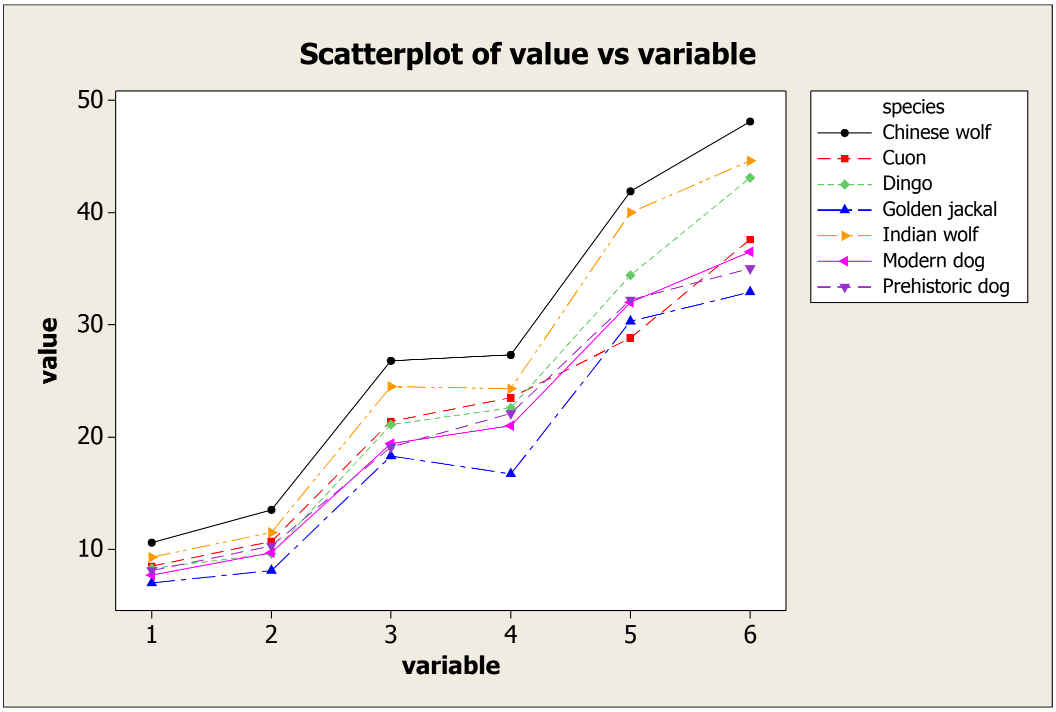
- plot[7] shows closeness of prehistoric and modern dog, also their distance to the Chinese wolf.

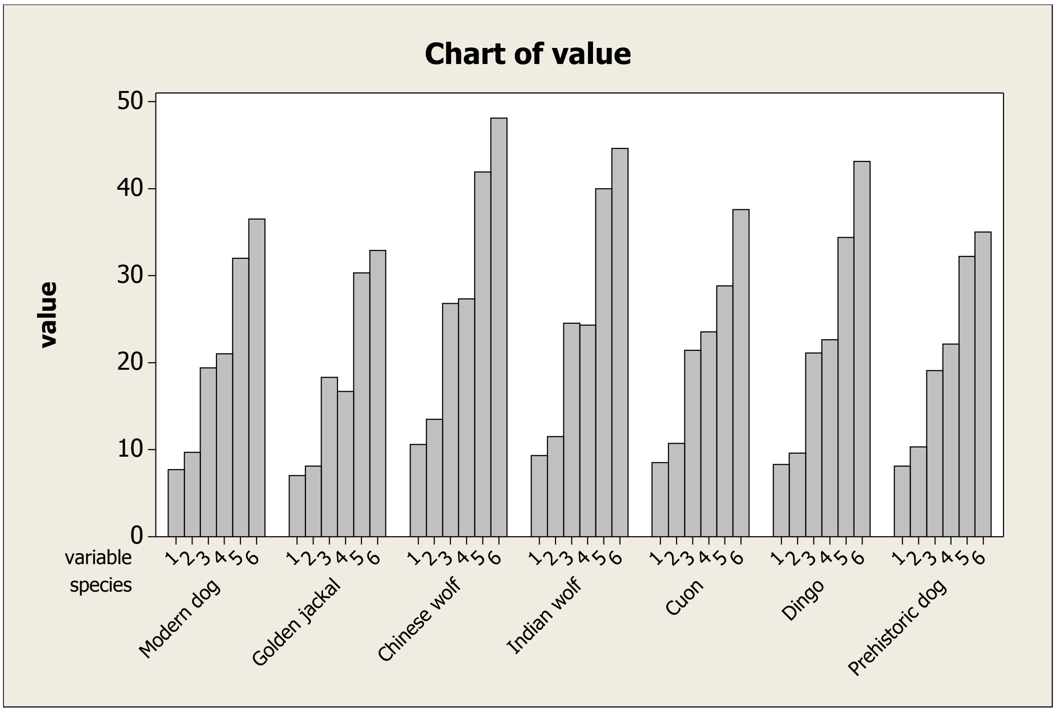
---

[6] Attributed to Chernoff H (1973), *J. Amer. Statist. Assoc.* **68**, 361-368.

[7] Done with Stata `chernoff` add-on package, using the features: `isize` (eye size), `nose` (nose line), `bdens` (brow density), `msize` (mouth size), `mcurv` (mouth curvature) and `fline` (face line).

Dogs: Profile Plot and Bar Graph





- both plots show closeness of prehistoric and modern dog, and their differences from the Chinese wolf.