MULTIVARIATE STATISTICS, Lesson 7.
Multivariate statistical methods: reduction of dimensionality

- **Principal component analysis**. Model: $\mathbf{X} = \mathbf{U}\mathbf{Y} + \mathbf{m}$,
  where $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$, $\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{D})$, where $\mathbf{U}$ is an appropriately chosen $p \times p$ orthogonal, while $\mathbf{D}$ is diagonal matrix.

  1. *Proposition*: $\mathbf{U}$ contains the eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_p$ of $\mathbf{C}$ columnwise corresponding to the eigenvalues $\lambda_1 \geq \dots \geq \lambda_p$.
  2. *Definition*: $Y_i = \mathbf{u}_i^T(\mathbf{X} - \mathbf{m})$ is the ith **principal component** with variance λ_i ($i = 1, \dots, p$).
  3. *Remark*: The variance of Y_i is λ_i, and the total variance of the principal components is equal to the total variance of the original X_i's.
  4. *Theorem*: The variance of Y_1 is the largest possible among the variances of linear combinations $\mathbf{v}^T(\mathbf{X} - \mathbf{m})$ subject to $\|\mathbf{v}\| = 1$. In general: the variance of Y_k is the largest possible among the variances of linear combinations $\mathbf{v}^T(\mathbf{X} - \mathbf{m})$ that are uncorrelated with $Y_1, \dots, Y_{k-1}$ (subject to $\|\mathbf{v}\| = 1$), $k = 2, \dots, p$.
  5. *Stronger Theorem*: The k-dimensional vector with components $(Y_1, \dots, Y_k, 0, \dots, 0)$ gives the best k-dimensional approximation of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{C})$ in the following sense: the minimum of $\mathbb{E}\|\mathbf{X} - \mathbf{A}\mathbf{X}\|$ with a $p \times p$ matrix $\mathbf{A}$ of rank k is attained by the projecton onto the k-dimensional subspace spanned by $\mathbf{u}_1, \dots, \mathbf{u}_k$ for any $k = 1, \dots, p$.
  6. *Sequential testing of hypotheses* for the number of relevant principal components. Based on the eigenvalues of the sample covariance matrix, for testing

     $$H_k \, : \, \lambda_{k+1} = \dots = \lambda_{p-1} = \lambda_p, \qquad k = 0, 1 \dots, p-1$$

     the transformed test statistic (obtained by likelihood ratio test)

     $$-2 \ln \lambda_n = n(p-k) \ln \frac{a}{g}, \qquad a = \frac{\hat{\lambda}_{k+1} + \dots + \hat{\lambda}_p}{p-k}, \quad g = (\hat{\lambda}_{k+1} \dots \hat{\lambda}_p)^{\frac{1}{p-k}}$$

     is used ($\hat{\lambda}_i$'s are the eigenvalues of the empirical covariance matrix) that for large n approximately follows $\chi^2(\frac{1}{2}(p-k+2)(p-k-1))$-distribution.

- **Factor analysis**. Model: $\mathbf{X} = \mathbf{A}\mathbf{f} + \mathbf{e} + \mathbf{m}$,
  where $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$, $\mathbf{A}$ is $p \times k$ matrix, $\mathbf{f} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$ is the **common factor** and the components of $\mathbf{e} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{D})$ are the **individual factors** of the variables with variances along the main diagonal of the diagonal matrix $\mathbf{D}$. Further, $\mathbf{f}$ and $\mathbf{e}$ are independent. If multivariate normality is not postulated, the conditions

  $$\mathbb{E}\mathbf{f} = \mathbf{0}, \quad \mathbb{E}\mathbf{f}\mathbf{f}^T = \mathbf{I}_k, \quad \mathbb{E}\mathbf{e} = \mathbf{0}, \quad \mathbb{E}\mathbf{e}\mathbf{e}^T = \mathbf{D}, \quad \mathbb{E}\mathbf{f}\mathbf{e}^T = \mathbf{0} \text{ matrix}$$

  are used. For the cordinates and variances of X_i's:

  $$X_i = \sum_{j=1}^{k} a_{ij} f_j + e_i + \mu_i, \quad c_{ii} = \sum_{j=1}^{k} a_{ij}^2 + d_{ii} \quad (= 1, \dots, p).$$

  1. *Definition*: $\sum_{j=1}^{k} a_{ij}^2$ is called *communality* of X_i ($i = 1, \dots, p$), and the entries of $\mathbf{A}$ are called *factor loadings*.

2. *Identification*: we have to solve the matrix equation $\mathbf{C} = \mathbf{A}\mathbf{A}^T + \mathbf{D}$. The solution may exist for $k \geq (2p + 1 - \sqrt{8p+1})/2$, and it is unique up to orthogonal rotation (if $\mathbf{A}$ is solution, $\mathbf{AQ}$ is also solution with any $k \times k$ ortogonal matrix $\mathbf{Q}$).
3. *ML factor analysis* (if multivariate normality is postulated): Maximize the multivariate normal likelihood function

$$-\frac{1}{2} n \ln |\mathbf{C}| - \frac{1}{2} n \mathrm{tr}\, \mathbf{C}^{-1} \hat{\mathbf{C}} + c$$

with respect to $\mathbf{A}, \mathbf{D}$ subject to $\mathbf{C} = \mathbf{A}\mathbf{A}^T + \mathbf{D}$; or equivalently, find the minimum of

$$F(\mathbf{A}, \mathbf{D}) = \ln |\mathbf{A}\mathbf{A}^T + \mathbf{D}| + \mathrm{tr}\, (\mathbf{A}\mathbf{A}^T + \mathbf{D})^{-1} \hat{\mathbf{C}}$$

that, after differentiating, gives the model equations:

$$\frac{\partial F}{\partial \mathbf{A}} = \mathbf{C}^{-1}(\mathbf{C} - \hat{\mathbf{C}})\mathbf{C}^{-1}\mathbf{A} = \mathbf{0}, \qquad \frac{\partial F}{\partial \mathbf{D}} = \mathrm{diag}\, [\mathbf{C}^{-1}(\mathbf{C} - \hat{\mathbf{C}})\mathbf{C}^{-1}] = \mathbf{0}.$$

For the solution, algorithms based on numeric iteration are at our disposal.
4. *Principal component factor analysis*: Retain the first k transformed principal components, where there is a remarkable spectral gap between the kth and $(k+1)$th eigenvalues of the sample covariance (or correlation) matrix. We use the transformation:

$$\mathbf{UY} = (\mathbf{U}\Lambda^{1/2})(\Lambda^{-1/2}\mathbf{Y}),$$

where $\mathbf{f}$ will consist of the first k components of $\Lambda^{-1/2}\mathbf{Y}$.
5. Rotation of factor loadings (VARIMAX).
6. See BMDP outputs after processing real world data.