多元统计

陈崇双

西南交通大学数学学院统计系

ccsmars@swjtu.edu.cn

2018-2019学年

第一节：问题背景

因子分析和主成分分析，都是多元分析中的降维方法。其思想始于1904年Charles Spearman对学生考试成绩的研究。

近年来随着电子计算机的高速发展, 因子分析已成功应用于心理学、医学、气象、地质、经济学等领域。

问题背景

例1

某公司出了一套测试试卷，包含*50*道题，涉及语言表达能力、逻辑思维能力、思想修养、生活常识、兴趣爱好等方面。对*100*名招聘人员进行测试，获得每个人关于每道题的成绩。尽管总分度量了应聘人员的综合情况，但想了解应聘人员各个方面的能力是否都适应公司的要求。

问题背景

例1

某公司出了一套测试试卷，包含50道题，涉及语言表达能力、逻辑思维能力、思想修养、生活常识、兴趣爱好等方面。对100名招聘人员进行测试，获得每个人关于每道题的成绩。尽管总分度量了应聘人员的综合情况，但想了解应聘人员各个方面的能力是否都适应公司的要求。

分析:

- 每一种能力都比较抽象，不能直接观测或度量;
- 每道题目可能都涉及到上述这些能力，一般称为公共因子;

例1

某公司出了一套测试试卷，包含50道题，涉及语言表达能力、逻辑思维能力、思想修养、生活常识、兴趣爱好等方面。对100名招聘人员进行测试，获得每个人关于每道题的成绩。尽管总分度量了应聘人员的综合情况，但想了解应聘人员各个方面的能力是否都适应公司的要求。

分析:

- 每一种能力都比较抽象，不能直接观测或度量；
- 每道题目可能都涉及到上述这些能力，一般称为公共因子；
- 不同人员在这些能力上的差异，导致了试卷上各题得分上的差异；
- 根据每个人员的每项得分，估计出他们的各项能力表现。

问题背景

因子分析的目标：将可观测变量通过潜在的公共因子来加以解释，并且希望公共因子的数目尽可能少。

问题背景

因子分析的目标：将可观测变量通过潜在的公共因子来加以解释，并且希望公共因子的数目尽可能少。

因子分析的任务：

- 通过可观测变量去估计不可观测的公共因子;

问题背景

因子分析的目标：将可观测变量通过潜在的公共因子来加以解释，并且希望公共因子的数目尽可能少。

因子分析的任务：

- 通过可观测变量去估计不可观测的公共因子;
- 若推测出公共因子存在，需解释它们的实际含义;

问题背景

因子分析的目标：将可观测变量通过潜在的公共因子来加以解释，并且希望公共因子的数目尽可能少。

因子分析的任务：

- 通过可观测变量去估计不可观测的公共因子;
- 若推测出公共因子存在，需解释它们的实际含义;
- 公共因子不能解释可观测变量所表达的部分信息，归结为特殊因子来承载，需推断其强度;

问题背景

因子分析的目标：将可观测变量通过潜在的公共因子来加以解释，并且希望公共因子的数目尽可能少。

因子分析的任务：

- 通过可观测变量去估计不可观测的公共因子;
- 若推测出公共因子存在，需解释它们的实际含义;
- 公共因子不能解释可观测变量所表达的部分信息，归结为特殊因子来承载，需推断其强度;
- 依据样品的指标值, 测算出样品在各个公共因子上的水平(也称因子得分)。

因子分析类型：

- R型因子分析: 研究变量之间的相关关系
- Q型因子分析: 研究样本之间的相关关系

总体$\mathbf{X} = (X_1, X_2, \cdots, X_p)^\mathsf{T}$的均值向量为$\mu = (\mu_1, \mu_2, \cdots, \mu_p)^\mathsf{T}$，协差阵为$\Sigma$。从中抽取$n$个样本$x_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^\mathsf{T}, i = 1, 2, \cdots, n$。

- 假设1: 每个变量都可以由m个公共因子$\mathbf{F} = (F_1, F_2, \cdots, F_m)^\mathsf{T}$和一个特殊因子线性表示；
- 假设2: 公共因子$\mathbf{F}$的各分量不相关，且均值向量为$E(\mathbf{F}) = \mathbf{0}$，协差阵为$Cov(\mathbf{F}) = I_m$；
- 假设3: 特殊因子$\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p)^\mathsf{T}$的各分量不相关，其均值向量为$E(\varepsilon) = \mathbf{0}$，协差阵为$\Sigma_\varepsilon = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2)$；
- 假设4: 特殊因子ε与公共因子$\mathbf{F}$不相关。

数学模型

$$\begin{cases} X_1 = \mu_1 + a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = \mu_2 + a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\ \ldots \\ X_p = \mu_p + a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_p \end{cases}$$

数学模型

$$
\begin{cases}
X_1 = \mu_1 + a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\
X_2 = \mu_2 + a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\
\cdots \\
X_p = \mu_p + a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_p
\end{cases}
$$

矩阵形式:

$$\boldsymbol{X} = \mu + \boldsymbol{A}\boldsymbol{F} + \boldsymbol{\varepsilon}$$

其中$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \vdots & \vdots \\ a_{p1} & \cdots & a_{pm} \end{pmatrix}$也称为因子载荷矩阵。

性质1：因子载荷系数$a_{ij} = E(X_i F_j) = Cov(X_i, F_j)$。

数学模型

性质1：因子载荷系数$a_{ij} = E(X_i F_j) = Cov(X_i, F_j)$。

注：因子载荷系数a_{ij}为可观测指标X_i与公共因子F_j的协方差，也正比于F_j和X_i的相关系数。

数学模型

性质2：X_i的方差可分解为$D(X_i) = \sum_{j=1}^{m} a_{ij}^2 + \sigma_i^2 \triangleq h_i^2 + \sigma_i^2$。

性质2：X_i的方差可分解为$D(X_i)=\sum_{j=1}^m a_{ij}^2+\sigma_i^2 \triangleq h_i^2+\sigma_i^2$。

注1：$h_i^2=A_{(i)}^{\top}A_{(i)}$，其中$A_{(i)}$为因子载荷矩阵$A$的第$i$行元素构成的列向量。

性质2：X_i的方差可分解为$D(X_i)=\sum_{j=1}^m a_{ij}^2+\sigma_i^2 \triangleq h_i^2+\sigma_i^2$。

注1：$h_i^2=A_{(i)}^{\mathsf{T}}A_{(i)}$，其中$A_{(i)}$为因子载荷矩阵$A$的第$i$行元素构成的列向量。

注2：σ_i^2刻画了特殊因子对变量X_i的方差贡献；h_i^2刻画了全部m个公共因子对变量X_i的方差贡献，称为X_i的变量共同度。

数学模型

性质2：X_i的方差可分解为$D(X_i)=\sum_{j=1}^m a_{ij}^2+\sigma_i^2 \triangleq h_i^2+\sigma_i^2$。

注1：$h_i^2=A_{(i)}^{\mathsf{T}}A_{(i)}$，其中$A_{(i)}$为因子载荷矩阵$A$的第$i$行元素构成的列向量。

注2：σ_i^2刻画了特殊因子对变量X_i的方差贡献；h_i^2刻画了全部m个公共因子对变量X_i的方差贡献，称为X_i的变量共同度。

注3：变量共同度h_i^2越大(相对于σ_i^2 而言)，表示X_i所反映的信息可通过公共因子解释的部分越多，因子分析的效果就越好。

性质2：X_i的方差可分解为$D(X_i)=\sum_{j=1}^m a_{ij}^2+\sigma_i^2 \triangleq h_i^2+\sigma_i^2$。

注1：$h_i^2=A_{(i)}^{\mathsf{T}}A_{(i)}$，其中$A_{(i)}$为因子载荷矩阵$A$的第$i$行元素构成的列向量。

注2：σ_i^2刻画了特殊因子对变量X_i的方差贡献；h_i^2刻画了全部m个公共因子对变量X_i的方差贡献，称为X_i的变量共同度。

注3：变量共同度h_i^2越大(相对于σ_i^2 而言)，表示X_i所反映的信息可通过公共因子解释的部分越多，因子分析的效果就越好。

注4：若将特殊因子ε_i看作剩余的$p-m$个公共因子$F_{m+1}, F_{m+2}, \cdots,$ F_p的综合效果，即$\varepsilon_i=\sum_{j=m+1}^p a_{ij}F_j$，则此处的因子载荷矩阵与主成分分析中的负荷矩阵一致。

数学模型

性质3：第j个公共因子F_j对全部变量$X_1, X_2, \cdots, X_p$所提供的方差贡献总和$g_j^2 \triangleq \sum_{i=1}^{p} a_{ij}^2$, 称为$F_j$的因子重要度，衡量了第$j$个公共因子$F_j$的相对重要性。

性质3：第j个公共因子F_j对全部变量$X_1, X_2, \cdots, X_p$所提供的方差贡献总和$g_j^2 \triangleq \sum_{i=1}^{p} a_{ij}^2$, 称为$F_j$的因子重要度，衡量了第$j$个公共因子$F_j$的相对重要性。

注：$g_j^2 = A_j^\top A_j$, 其中A_j为因子载荷矩阵A的第j列元素构成的列向量。

数学模型

表: 因子载荷量

	F_1	$\dots$	F_m	$\sum_{j=1}^m a_{ij}^2$
X_1	a_{11}	$\dots$	a_{1m}	h_1^2
X_2	a_{21}	$\dots$	a_{2m}	h_2^2
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$
X_p	a_{p1}	$\dots$	a_{pm}	h_p^2
$\sum_{i=1}^p a_{ij}^2$	g_1^2	$\dots$	g_m^2	

数学模型

性质4：因子载荷矩阵不唯一。

数学模型

性质4：因子载荷矩阵不唯一。

假设公共因子$\boldsymbol{F}$对应的因子负荷矩阵是$\boldsymbol{A}$。对任何一个m阶正交矩阵$\boldsymbol{\Gamma}$ (即满足$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathsf{T}} = \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Gamma} = \boldsymbol{I}_m$), 若取$\tilde{\boldsymbol{F}} \triangleq \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}$ 作为公共因子，则

性质4：因子载荷矩阵不唯一。

假设公共因子$\boldsymbol{F}$对应的因子负荷矩阵是$\boldsymbol{A}$。对任何一个m阶正交矩阵$\boldsymbol{\Gamma}$ (即满足$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathsf{T}}=\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Gamma}=\boldsymbol{I}_m$), 若取$\tilde{\boldsymbol{F}}\triangleq\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}$ 作为公共因子，则

$$\boldsymbol{X}=\boldsymbol{A}\boldsymbol{F}+\boldsymbol{\varepsilon}=(\boldsymbol{A}\boldsymbol{\Gamma})(\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F})+\boldsymbol{\varepsilon}\triangleq\tilde{\boldsymbol{A}}\tilde{\boldsymbol{F}}+\boldsymbol{\varepsilon}$$

数学模型

性质4：因子载荷矩阵不唯一。

假设公共因子$\boldsymbol{F}$对应的因子负荷矩阵是$\boldsymbol{A}$。对任何一个m阶正交矩阵$\boldsymbol{\Gamma}$ (即满足$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathsf{T}} = \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Gamma} = \boldsymbol{I}_m$), 若取$\tilde{\boldsymbol{F}} \triangleq \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}$ 作为公共因子，则

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{F} + \boldsymbol{\varepsilon} = (\boldsymbol{A}\boldsymbol{\Gamma})(\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}) + \boldsymbol{\varepsilon} \triangleq \tilde{\boldsymbol{A}}\tilde{\boldsymbol{F}} + \boldsymbol{\varepsilon}$$

即对应的因子负荷矩阵为$\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Gamma}$。并且只要$\boldsymbol{F}, \boldsymbol{A}, \boldsymbol{\varepsilon}$满足因子分析的数学模型和基本假定，那么$\tilde{\boldsymbol{F}}, \tilde{\boldsymbol{A}}$也满足。

性质4：因子载荷矩阵不唯一。

假设公共因子$\boldsymbol{F}$对应的因子负荷矩阵是$\boldsymbol{A}$。对任何一个m阶正交矩阵$\boldsymbol{\Gamma}$ (即满足$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{I}_m$), 若取$\tilde{\boldsymbol{F}} \triangleq \boldsymbol{\Gamma}^\top\boldsymbol{F}$ 作为公共因子，则

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{F} + \boldsymbol{\varepsilon} = (\boldsymbol{A}\boldsymbol{\Gamma})(\boldsymbol{\Gamma}^\top\boldsymbol{F}) + \boldsymbol{\varepsilon} \triangleq \tilde{\boldsymbol{A}}\tilde{\boldsymbol{F}} + \boldsymbol{\varepsilon}$$

即对应的因子负荷矩阵为$\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Gamma}$。并且只要$\boldsymbol{F}, \boldsymbol{A}, \boldsymbol{\varepsilon}$满足因子分析的数学模型和基本假定，那么$\tilde{\boldsymbol{F}}, \tilde{\boldsymbol{A}}$也满足。

$$\begin{cases} E(\tilde{\boldsymbol{F}}) = E(\boldsymbol{\Gamma}^\top\boldsymbol{F}) = \boldsymbol{\Gamma}^\top E(\boldsymbol{F}) = \boldsymbol{0} \\ D(\tilde{\boldsymbol{F}}) = D(\boldsymbol{\Gamma}^\top\boldsymbol{F}) = \boldsymbol{\Gamma}^\top D(\boldsymbol{F})\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{I}_m \\ Cov(\tilde{\boldsymbol{F}}, \boldsymbol{\varepsilon}) = Cov(\boldsymbol{\Gamma}^\top\boldsymbol{F}, \boldsymbol{\varepsilon}) = \boldsymbol{\Gamma}^\top Cov(\boldsymbol{F}, \boldsymbol{\varepsilon}) = \boldsymbol{0} \end{cases}$$

数学模型

注1：$\tilde{\boldsymbol{F}} = \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}$相当于$\boldsymbol{F}$ 在m维空间作旋转变换。

数学模型

注1：$\tilde{\boldsymbol{F}} = \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}$相当于$\boldsymbol{F}$ 在m维空间作旋转变换。

注2：公共因子$\mathbf{F} = (F_1, F_2, \cdots, F_m)^{\mathsf{T}}$本不可观测, 且假设$E(\mathbf{F}) = \mathbf{0}$，$Cov(\mathbf{F}) = I_m$，则各公共因子并无主次之分。因此因子负荷矩阵不唯一。

注1：$\tilde{\boldsymbol{F}} = \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{F}$相当于$\boldsymbol{F}$ 在m维空间作旋转变换。

注2：公共因子$\mathbf{F} = (F_1, F_2, \cdots, F_m)^{\mathsf{T}}$本不可观测, 且假设$E(\mathbf{F}) = \mathbf{0}$，$Cov(\mathbf{F}) = I_m$，则各公共因子并无主次之分。因此因子负荷矩阵不唯一。基于此性质，合适的变换将便于公共因子合理解释。

注1：$\tilde{F} = \Gamma^\top F$相当于$F$ 在m维空间作旋转变换。

注2：公共因子$\mathbf{F} = (F_1, F_2, \cdots, F_m)^\top$本不可观测, 且假设$E(\mathbf{F}) = \mathbf{0}$，$Cov(\mathbf{F}) = I_m$，则各公共因子并无主次之分。因此因子负荷矩阵不唯一。基于此性质，合适的变换将便于公共因子合理解释。

注3：正交旋转后，变量共同度不变，公共因子的重要度发生变化。

$$\begin{cases} \widetilde{h}_i^2 = \widetilde{A}_{(i)}^\top \widetilde{A}_{(i)} = A_{(i)}\Gamma(A_{(i)}\Gamma)^\top = A_{(i)}^\top A_{(i)} = h_i^2 \\ \widetilde{g}_j^2 = \widetilde{A}_j^\top \widetilde{A}_j = (A\Gamma_j)^\top (A\Gamma_j) = \Gamma_j^\top A^\top A\Gamma_j \end{cases}$$

其中Γ_j为Γ的第j列，只有当$\Gamma_j = \mathbf{e}_j$时，才有$A\Gamma_j = A_j$，即$\widetilde{g}_j^2 = g_j^2$。

数学模型

由于观测量纲的差异以及数量级不同造成的影响，一般将样本观测矩阵进行标准化处理，则此时的协方差阵就是原变量的相关阵。如果对相关阵进行因子分析，则

- 数学模型中$\mu_i = 0, i = 1, 2, \cdots, p$；
- 因子负荷系数a_{ij}, 也为可观测指标X_i与公共因子F_j的相关系数；
- 变量共同度$h_i^2 \leq 1, i = 1, 2, \cdots, p$。

根据模型的基本假定

$$\boldsymbol{\Sigma} = \boldsymbol{A}D(\boldsymbol{F})\boldsymbol{A}^{\top} + D(\boldsymbol{\varepsilon}) = \boldsymbol{A}\boldsymbol{A}^{\top} + \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2)$$

第三节：因子负荷矩阵估计

根据模型的基本假定

$$\boldsymbol{\Sigma} = \boldsymbol{A}D(\boldsymbol{F})\boldsymbol{A}^{\mathsf{T}} + D(\boldsymbol{\varepsilon}) = \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} + \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2)$$

可见因子负荷矩阵$\boldsymbol{A}$的估计，可归结为寻求可观测指标$\boldsymbol{X}$的协差阵$\boldsymbol{\Sigma}$进行分解。

因子负荷矩阵估计

设有n个p维可观测样品$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}$, $i = 1, 2, \cdots, n$, 分别代入模型可得$n \times p$个方程。

$$
\begin{cases}
x_{i1} = \mu_1 + a_{11}F_{i1} + a_{12}F_{i2} + \cdots + a_{1m}F_{im} + \varepsilon_{i1} \\
x_{i2} = \mu_2 + a_{21}F_{i1} + a_{22}F_{i2} + \cdots + a_{2m}F_{im} + \varepsilon_{i2} \\
\cdots \\
x_{ip} = \mu_p + a_{p1}F_{i1} + a_{p2}F_{i2} + \cdots + a_{pm}F_{im} + \varepsilon_{ip}
\end{cases}
$$

因子负荷矩阵估计

设有n个p维可观测样品$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}$, $i = 1, 2, \cdots, n$, 分别代入模型可得$n \times p$个方程。

$$\begin{cases} x_{i1} = \mu_1 + a_{11}F_{i1} + a_{12}F_{i2} + \cdots + a_{1m}F_{im} + \varepsilon_{i1} \\ x_{i2} = \mu_2 + a_{21}F_{i1} + a_{22}F_{i2} + \cdots + a_{2m}F_{im} + \varepsilon_{i2} \\ \cdots \\ x_{ip} = \mu_p + a_{p1}F_{i1} + a_{p2}F_{i2} + \cdots + a_{pm}F_{im} + \varepsilon_{ip} \end{cases}$$

其中未知参数:

- 均值参数μ_i, $i = 1, 2, \cdots, p$, 共p个;
- 因子负荷a_{ij}, $i = 1, 2, \cdots, p; j = 1, 2, \cdots, m$, 共$p \times m$个;
- 特殊因子ε_{ij}, $i = 1, 2, \cdots, n; j = 1, 2, \cdots, p$, 共$n \times p$个;
- 公共因子得分F_{ij}, $i = 1, 2, \cdots, n; j = 1, 2, \cdots, m$, 共$n \times m$个;

因子负荷矩阵估计

设有n个p维可观测样品$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}$, $i = 1, 2, \cdots, n$, 分别代入模型可得$n \times p$个方程。

$$\begin{cases} x_{i1} = \mu_1 + a_{11}F_{i1} + a_{12}F_{i2} + \cdots + a_{1m}F_{im} + \varepsilon_{i1} \\ x_{i2} = \mu_2 + a_{21}F_{i1} + a_{22}F_{i2} + \cdots + a_{2m}F_{im} + \varepsilon_{i2} \\ \cdots \\ x_{ip} = \mu_p + a_{p1}F_{i1} + a_{p2}F_{i2} + \cdots + a_{pm}F_{im} + \varepsilon_{ip} \end{cases}$$

其中未知参数:

- 均值参数μ_i, $i = 1, 2, \cdots, p$, 共p个;
- 因子负荷a_{ij}, $i = 1, 2, \cdots, p; j = 1, 2, \cdots, m$, 共$p \times m$个;
- 特殊因子ε_{ij}, $i = 1, 2, \cdots, n; j = 1, 2, \cdots, p$, 共$n \times p$个;
- 公共因子得分F_{ij}, $i = 1, 2, \cdots, n; j = 1, 2, \cdots, m$, 共$n \times m$个;

因此不可能精确解出因子负荷矩阵$\boldsymbol{A}$, 只能作适当估计。

因子负荷矩阵估计

特别地，若公共因子的个数与可观测变量的个数一样多，即$m = p$，此时从两组变量相互表达的角度，可以认为不需要特殊因子。即

$$\boldsymbol{X} = \mu + \boldsymbol{A}\boldsymbol{F}$$

因子负荷矩阵估计

特别地，若公共因子的个数与可观测变量的个数一样多，即$m = p$，此时从两组变量相互表达的角度，可以认为不需要特殊因子。即

$$\boldsymbol{X} = \mu + \boldsymbol{A}\boldsymbol{F}$$

两边同求协方差可得

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$$

因子负荷矩阵估计

特别地，若公共因子的个数与可观测变量的个数一样多，即$m=p$，此时从两组变量相互表达的角度，可以认为不需要特殊因子。即

$$\boldsymbol{X}=\mu+\boldsymbol{AF}$$

两边同求协方差可得

$$\boldsymbol{\Sigma}=\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$$

由于$\boldsymbol{\Sigma}$对称非负定，必存在正交矩阵$\boldsymbol{U}$ 使

$$\boldsymbol{\Sigma}=\boldsymbol{U}diag(\lambda_1,\lambda_2,\cdots,\lambda_p)\boldsymbol{U}^{\mathsf{T}}$$

其中$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_p\geq0$分别是$\boldsymbol{\Sigma}$ 的p 个特征根; $\boldsymbol{U}$ 的各列依次是各特征根所对应的标准正交特征向量$\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots,\boldsymbol{\xi}_p$。

因子负荷矩阵估计

由此获得$\mathbf{\Sigma}$的一种分解：

$$\mathbf{\Sigma}=(\sqrt{\lambda_1}\boldsymbol{\xi}_1,\sqrt{\lambda_2}\boldsymbol{\xi}_2,\cdots,\sqrt{\lambda_p}\boldsymbol{\xi}_p)(\sqrt{\lambda_1}\boldsymbol{\xi}_1,\sqrt{\lambda_2}\boldsymbol{\xi}_2,\cdots,\sqrt{\lambda_p}\boldsymbol{\xi}_p)^{\mathsf{T}}$$

因子负荷矩阵估计

由此获得$\boldsymbol{\Sigma}$的一种分解：

$$\boldsymbol{\Sigma} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)(\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)^{\mathsf{T}}$$

于是可取$\hat{\boldsymbol{A}} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)$。

因子负荷矩阵估计

由此获得$\boldsymbol{\Sigma}$的一种分解：

$$\boldsymbol{\Sigma} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)(\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)^{\mathsf{T}}$$

于是可取$\hat{\boldsymbol{A}} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)$。此时，第$j$个分量$F_j$的公共因子重要度$g_j = \lambda_j \boldsymbol{\xi}_j^{\mathsf{T}} \boldsymbol{\xi}_j = \lambda_j$。

由此获得$\mathbf{\Sigma}$的一种分解：

$$\mathbf{\Sigma} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)(\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)^{\mathsf{T}}$$

于是可取$\hat{\boldsymbol{A}} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_p}\boldsymbol{\xi}_p)$。此时，第$j$个分量$F_j$的公共因子重要度$g_j = \lambda_j \boldsymbol{\xi}_j^{\mathsf{T}} \boldsymbol{\xi}_j = \lambda_j$。

实际中，一般希望公共因子的个数m远小于可观测指标的个数p。从因子重要度来看，当λ_{m+1}很小时，F_m以后的公共因子就可忽略，它们对可观测指标的综合影响视为特殊因子的影响。

因子负荷矩阵估计

$$
\begin{aligned}
\boldsymbol{\Sigma} =& (\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)(\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)^{\mathsf{T}} + \varepsilon\varepsilon^{\mathsf{T}} \\
\approx& (\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)(\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)^{\mathsf{T}} + \mathrm{diag}(\sigma_1^2, \cdots, \sigma_p^2)
\end{aligned}
$$

式中$\varepsilon = (\sqrt{\lambda_{m+1}}\xi_{m+1}, \cdots, \sqrt{\lambda_p}\xi_p)^{\mathsf{T}}$,

因子负荷矩阵估计

$$\begin{aligned}\boldsymbol{\Sigma} =&(\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)(\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)^{\mathsf{T}} + \varepsilon\varepsilon^{\mathsf{T}} \\ \approx&(\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)(\sqrt{\lambda_1}\xi_1, \cdots, \sqrt{\lambda_m}\xi_m)^{\mathsf{T}} + \mathrm{diag}(\sigma_1^2, \cdots, \sigma_p^2)\end{aligned}$$

式中$\varepsilon = (\sqrt{\lambda_{m+1}}\xi_{m+1}, \cdots, \sqrt{\lambda_p}\xi_p)^{\mathsf{T}}$, 约等号"≈"是由于$p$阶方阵

$$(\sqrt{\lambda_{m+1}}\xi_{m+1}, \cdots, \sqrt{\lambda_p}\xi_p)(\sqrt{\lambda_{m+1}}\xi_{m+1}, \cdots, \sqrt{\lambda_p}\xi_p)^{\mathsf{T}}$$

的非对角元未必是0。

因子负荷矩阵估计

综上可得因子负荷矩阵$\boldsymbol{A}$的一种估计

$$\hat{\boldsymbol{A}} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_m}\boldsymbol{\xi}_m)$$

综上可得因子负荷矩阵$\boldsymbol{A}$的一种估计

$$\hat{\boldsymbol{A}} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_m}\boldsymbol{\xi}_m)$$

其中第j列元素$\sqrt{\lambda_j}\boldsymbol{\xi}_j$，恰好是主成分分析中第$j$主成分的系数向量$\boldsymbol{\xi}_j$的$\sqrt{\lambda_j}$倍。

因子负荷矩阵估计

综上可得因子负荷矩阵$\boldsymbol{A}$的一种估计

$$\hat{\boldsymbol{A}} = (\sqrt{\lambda_1}\boldsymbol{\xi}_1, \sqrt{\lambda_2}\boldsymbol{\xi}_2, \cdots, \sqrt{\lambda_m}\boldsymbol{\xi}_m)$$

其中第j列元素$\sqrt{\lambda_j}\boldsymbol{\xi}_j$，恰好是主成分分析中第$j$主成分的系数向量$\boldsymbol{\xi}_j$的$\sqrt{\lambda_j}$倍。上述方法也称为主成分估计法。

因子负荷矩阵估计

因子负荷矩阵$\boldsymbol{A}$的主成分估计法，可归结为求可观测变量$X_1, X_2, \cdots, X_p$的协差阵$\boldsymbol{\Sigma}$的特征根与特征向量。

因子负荷矩阵估计

因子负荷矩阵$\boldsymbol{A}$的主成分估计法，可归结为求可观测变量$X_1, X_2, \cdots, X_p$的协差阵$\boldsymbol{\Sigma}$的特征根与特征向量。

当$\boldsymbol{\Sigma}$未知时，用所得到的n个p维样本$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}, i = 1, 2, \cdots, n$估计协差阵

$$S = \frac{1}{n-1}\sum_{k=1}^{n}(\boldsymbol{x}_{(i)} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{(i)} - \bar{\boldsymbol{x}})^{\mathsf{T}}$$

其中$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{(i)}$。相关阵的估计为$\boldsymbol{R} = (r_{ij})_{p\times p}$，其中$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$。

因子负荷矩阵估计

因子分析希望公共因子的个数远小于观测指标的个数, 常用准则有

(1) 类似于主成分分析中确定主成分个数的方法，累积贡献率不低于85%，即取$m = argmin_k\{\sum_{i=1}^{k}\lambda_i/\sum_{i=1}^{p}\lambda_i \geq 85\%\}$;

(2) 根据特征根的大小，如取$m = argmax_k\{\lambda_k \geq 1\}$。

因子负荷矩阵估计

因子分析希望公共因子的个数远小于观测指标的个数, 常用准则有

(1) 类似于主成分分析中确定主成分个数的方法，累积贡献率不低于85%，即取$m = argmin_k\{\sum_{i=1}^{k}\lambda_i / \sum_{i=1}^{p}\lambda_i \geq 85\%\}$;

(2) 根据特征根的大小，如取$m = argmax_k\{\lambda_k \geq 1\}$。

在估计出$\boldsymbol{\Sigma}$和计算出$\boldsymbol{A}$以后，可用$\boldsymbol{\Sigma} - \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$的对角元素，作为各特殊因子方差$(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2)$的估计。

第四节：公共因子解释

可观测变量都有明确的实际含义，在数学上可由公共因子和特殊因子线性组合得到。公共因子的含义呢？这就是公共因子的解释。

第四节：公共因子解释

可观测变量都有明确的实际含义，在数学上可由公共因子和特殊因子线性组合得到。公共因子的含义呢？这就是公共因子的解释。

F_j作出解释的依据：因子负荷矩阵$\boldsymbol{A}$中第j列元素$\sqrt{\lambda_j}\xi_j$，分别度量了F_j 与p个可观测变量$X_1, X_2, \cdots, X_p$之间的相关信息。

可观测变量都有明确的实际含义，在数学上可由公共因子和特殊因子线性组合得到。公共因子的含义呢？这就是公共因子的解释。

F_j作出解释的依据：因子负荷矩阵$\boldsymbol{A}$中第j列元素$\sqrt{\lambda_j}\xi_j$，分别度量了F_j与p个可观测变量$X_1, X_2, \cdots, X_p$之间的相关信息。找出与F_j相关程度最强的若干个可观测变量，综合它们的含义，并对比其它变量的含义，归纳出潜在因子F_j的合理解释与命名。

公共因子解释

若$\boldsymbol{A}$中有些列的元素较为均衡，则难以给出公共因子的合理解释。

公共因子解释

若A中有些列的元素较为均衡，则难以给出公共因子的合理解释。

应对办法：对负荷矩阵A施以正交变换，尽量使得A的各列元素向0和1两极分化。具体说来，A的同一行元素中只有一个接近于1，其余接近于0。

公共因子解释

若A中有些列的元素较为均衡，则难以给出公共因子的合理解释。

应对办法：对负荷矩阵A施以正交变换，尽量使得A的各列元素向0和1两极分化。具体说来，A的同一行元素中只有一个接近于1，其余接近于0。换言之，每个原始变量只与某个公共因子相关较强，而与其他因子几乎不相关。

公共因子解释

若A中有些列的元素较为均衡，则难以给出公共因子的合理解释。

应对办法：对负荷矩阵A施以正交变换，尽量使得A的各列元素向0和1两极分化。具体说来，A的同一行元素中只有一个接近于1，其余接近于0。换言之，每个原始变量只与某个公共因子相关较强，而与其他因子几乎不相关。

造成的结果：一些原始变量只与某个公共因子相关，从而实现对变量进行分组。

公共因子解释

方差最大法是常用的构造方法，由H.K. Kaiser于1958年提出，主要是基于方差分析的思想。以两个因子为例进行说明。

H. F. Kaiser(1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23:187-200.

https://ci.nii.ac.jp/naid/30017108987.

公共因子解释

因子载荷矩阵和正交矩阵

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{p1} & a_{p2} \end{pmatrix}, \boldsymbol{\Gamma} = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$$

正交变换后的因子载荷矩阵

$$\boldsymbol{B} = \boldsymbol{A\Gamma} \triangleq \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{p1} & b_{p2} \end{pmatrix}$$

公共因子解释

考虑两组数据$(b_{11}^2, b_{21}^2, \cdots, b_{p1}^2), (b_{12}^2, b_{22}^2, \cdots, b_{p2}^2)$的相对方差

$$v_i = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{b_{ij}^2}{h_i^2}\right)^2 - \left(\frac{1}{p}\sum_{i=1}^{p}\frac{b_{ij}^2}{h_i^2}\right)^2, i = 1, 2$$

式中b_{ij}^2消除了b_{ij}符号的影响; 变量共同度h_i^2消除了各个变量对公共因子依赖程度的影响。然后最大化这两组数据的总方差，获得旋转角度的估计：$\widehat{\theta} = argmax(v_1 + v_2)$。

如果公共因子有m个，则每次选择两个进行旋转，一轮共$\binom{m}{2}$次，可以进行多轮，每进行一轮各列的相对方差总和会有所增加，直至改变不明显时停止。以上过程在SPSS，SAS，R等软件中都可方便完成。

因子得分

因子分析的数学模型，将每个可观测变量$(X_1, X_2, \cdots, X_p)$都通过m个潜在的公共因子$(F_1, F_2, \cdots, F_m)$和一个特殊因子来加以释：

$$X_i = \mu_i + a_{i1}F_1 + \cdots + a_{im}F_m + \varepsilon_i, \ (i = 1, 2, \cdots, p)$$

因子得分

因子分析的数学模型, 将每个可观测变量$(X_1, X_2, \cdots, X_p)$都通过m个潜在的公共因子$(F_1, F_2, \cdots, F_m)$和一个特殊因子来加以释：

$$X_i = \mu_i + a_{i1}F_1 + \cdots + a_{im}F_m + \varepsilon_i, \ (i = 1, 2, \cdots, p)$$

在估计出载荷矩阵以及明确各公共因子的含义后，希望知道每个样品$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}, i = 1, 2, \cdots, n$在各因子上的定量水平，这就是因子得分。

因子得分

记F_{ij}为第i个样品$\boldsymbol{x}_{(i)}$在因子F_j上的得分，满足：

因子得分

记F_{ij}为第i个样品$\boldsymbol{x}_{(i)}$在因子F_j上的得分，满足：

$$\begin{cases} x_{i1} = \mu_1 + a_{11}F_{i1} + \cdots + a_{1m}F_{im} + \varepsilon_{i1} \\ \cdots \\ x_{ip} = \mu_p + a_{p1}F_{i1} + \cdots + a_{pm}F_{im} + \varepsilon_{ip} \end{cases}$$

因子得分

记F_{ij}为第i个样品$\boldsymbol{x}_{(i)}$在因子F_j上的得分，满足：

$$\begin{cases} x_{i1} = \mu_1 + a_{11}F_{i1} + \cdots + a_{1m}F_{im} + \varepsilon_{i1} \\ \cdots \\ x_{ip} = \mu_p + a_{p1}F_{i1} + \cdots + a_{pm}F_{im} + \varepsilon_{ip} \end{cases}$$

其中$(x_{i1}, x_{i2}, \cdots, x_{ip})$是第$i$个样品的$p$项可观测指标; a_{ij}是估计的因子负荷; $\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{ip}$是第$i$个样品的特殊因子分量。

因子得分

记F_{ij}为第i个样品$\boldsymbol{x}_{(i)}$在因子F_j上的得分，满足：

$$\begin{cases} x_{i1} = \mu_1 + a_{11}F_{i1} + \cdots + a_{1m}F_{im} + \varepsilon_{i1} \\ \cdots \\ x_{ip} = \mu_p + a_{p1}F_{i1} + \cdots + a_{pm}F_{im} + \varepsilon_{ip} \end{cases}$$

其中$(x_{i1}, x_{i2}, \cdots, x_{ip})$是第$i$个样品的$p$项可观测指标; a_{ij}是估计的因子负荷; $\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{ip}$是第$i$个样品的特殊因子分量。

由于特殊因子分量未知且不可观测，所以因子得分还不能从方程组中解出，只能估计。

因子得分

回顾多元线性回归模型，

- 可观测指标$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}$视为响应变量;
- 因子负荷$\{a_{ij}, i = 1, 2, \cdots, p; j = 1, 2, \cdots, m\}$ 视为解释变量;
- 特殊因子分量$\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{ip}$视为随机波动项;
- 因子得分$\boldsymbol{F}_{(i)} = (F_{i1}, F_{i2}, \cdots, F_{im})^{\mathsf{T}}$视为回归系数。

因子得分

回顾多元线性回归模型，

- 可观测指标$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}$视为响应变量;
- 因子负荷$\{a_{ij}, i = 1, 2, \cdots, p; j = 1, 2, \cdots, m\}$ 视为解释变量;
- 特殊因子分量$\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{ip}$视为随机波动项;
- 因子得分$\boldsymbol{F}_{(i)} = (F_{i1}, F_{i2}, \cdots, F_{im})^{\mathsf{T}}$视为回归系数。

可得第i个样品的因子得分的最小二乘估计

$$\boldsymbol{F}_{(i)} = (\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{x}_{(i)}$$

因子得分

回顾多元线性回归模型，

- 可观测指标$\boldsymbol{x}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathsf{T}}$视为响应变量；
- 因子负荷$\{a_{ij}, i = 1, 2, \cdots, p; j = 1, 2, \cdots, m\}$ 视为解释变量；
- 特殊因子分量$\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{ip}$视为随机波动项；
- 因子得分$\boldsymbol{F}_{(i)} = (F_{i1}, F_{i2}, \cdots, F_{im})^{\mathsf{T}}$视为回归系数。

可得第i个样品的因子得分的最小二乘估计

$$\boldsymbol{F}_{(i)} = (\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{x}_{(i)}$$

对上式求转置，可得

$$(F_{i1}, F_{i2}, \cdots, F_{im}) = (x_{i1}, x_{i2}, \cdots, x_{ip})\boldsymbol{A}(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}$$

因子得分

记样品观测矩阵和因子得分矩阵

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, F = \begin{pmatrix} \mu_1 & F_{11} & F_{12} & \cdots & F_{1m} \\ \mu_2 & F_{21} & F_{22} & \cdots & F_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_p & F_{n1} & F_{n2} & \cdots & F_{nm} \end{pmatrix}$$

简洁表达式为

$$F = XA(A^\top A)^{-1}$$

因子分析步骤

1. 数据标准化
2. 计算相关系数矩阵
3. 计算相关系数矩阵的特征值以及特征向量
4. 确定综合因子数以及因子载荷矩阵
5. 因子旋转
6. 计算因子得分

主成分分析VS.因子分析

（1）数学模型：主成分分析本质上是一种线性变换，是将原始坐标变换到变异程度大的方向，相当于从空间上转换观看数据的角度。而因子分析本质上是从显在变量去“提练”潜在因子的过程，因子的个数m取多大是要通过一定规则确定，并且因子的形式也不是唯一确定。

主成分分析VS.因子分析

（1）数学模型：主成分分析本质上是一种线性变换，是将原始坐标变换到变异程度大的方向，相当于从空间上转换观看数据的角度。而因子分析本质上是从显在变量去“提练”潜在因子的过程，因子的个数m取多大是要通过一定规则确定，并且因子的形式也不是唯一确定。

（2）主成分分析中主成分是各变量的线性组合，因子分析中变量是各因子的线性组合。

主成分分析VS.因子分析

（1）数学模型：主成分分析本质上是一种线性变换，是将原始坐标变换到变异程度大的方向，相当于从空间上转换观看数据的角度。而因子分析本质上是从显在变量去“提练”潜在因子的过程，因子的个数m取多大是要通过一定规则确定，并且因子的形式也不是唯一确定。

（2）主成分分析中主成分是各变量的线性组合，因子分析中变量是各因子的线性组合。

（3）主成分分析中不需要有假设，因子分析则需要一些假设。

主成分分析VS.因子分析

（1）数学模型：主成分分析本质上是一种线性变换，是将原始坐标变换到变异程度大的方向，相当于从空间上转换观看数据的角度。而因子分析本质上是从显在变量去“提练”潜在因子的过程，因子的个数m取多大是要通过一定规则确定，并且因子的形式也不是唯一确定。

（2）主成分分析中主成分是各变量的线性组合，因子分析中变量是各因子的线性组合。

（3）主成分分析中不需要有假设，因子分析则需要一些假设。

（4）解释与命名。因子分析可使用旋转技术而更具优势。