

# **Principal component analysis**

---

## **of the influencing factors of phase cost**

# Introduction

---

**Concept**

**Cases**

**Forecast and results analysis**

---

# Concept-- Basic idea

---

## □ Problem

variables —correlation  
—overlap

## □ target

processing these variables

—reducing the variable dimension  
—reflect the maximum difference

---

# Concept-- Basic idea

---

## □ Method

each of the main components of  $X$  are equivalent to a respective feature value from the covariance matrix  $\Sigma$  and the corresponding orthogonal unitized feature vector.

---

# Concept-- Basic idea

---

## □ Method

The contribution rate of the first principal component indicates the largest proportion of the comprehensive information contained within the original variables, while the remainder decreases in order of size from the original variables.

---

# Concept-- Step

---

□ normalize , Correlation coefficient matrix

Formula 1  $\Sigma = (\sigma_{ij})_{p \times p} = E[(X - E(X))(X - E(X))^T]$

Formula 2  $x_{ij}^* = \frac{x_{ij} - E(x_i)}{\sqrt{\sigma_{ii}}}$

---

# Concept-- Step

---

- to determine principal component by characteristic value

Set  $f_i = (f_{i1}, f_{i2}, \dots, f_{ip})^T \quad i = 1, 2, 3, \dots, p$

get the Formula 3:

$$\left\{ \begin{array}{l} Z_1 = f_1^T X = f_{11}x_1 + f_{12}x_2 + \dots + f_{1p}x_p \\ Z_2 = f_2^T X = f_{21}x_1 + f_{22}x_2 + \dots + f_{2p}x_p \\ \dots\dots \\ Z_p = f_p^T X = f_{p1}x_1 + f_{p2}x_2 + \dots + f_{pp}x_p \end{array} \right.$$

---

# Concept-- Step

---

## Formula 4

$$\begin{aligned} \text{var}(Z_i) &= \text{var}(f_i^T X) = f_i^T \Sigma f_i & \text{cov}(Z_i, Z_j) &= \text{cov}(f_i^T X, f_j^T X) = f_i^T \Sigma f_j \\ i &= 1, 2, 3, \dots, p & j &= 1, 2, 3, \dots, p \end{aligned}$$

## Formula 5

$$\text{cov}(Z_i, Z_j) = f_i^T \Sigma f_j = 0$$

- $Z_i = f_i^T X$  is called the  $i$ -th principal component of  $x_1, x_2, \dots, x_p$
-



# Cases

---

Table 1. Quantized values of the influencing factors in previous phase

Code	Total Cost	Capacity	Target	Cont.	Key Technic	Hard Achievements	Soft Achievements	Cycle	Personal	Test	Hard Coefficient	Evaluation
P1	500.00	7.00	2.00	11.00	4.00	8.00	3.00	3.00	0.00	1.80	2.40	5.00
P2	300.00	9.30	1.00	12.00	0.00	2.00	3.00	3.00	0.00	1.70	6.00	1.00
P3	578.78	8.00	3.00	6.00	0.00	2.00	3.00	5.00	0.00	1.50	1.90	1.00
P4	500.00	10.00	1.00	4.00	0.00	0.00	3.00	3.00	0.00	1.80	2.40	1.00
P5	200.00	8.40	1.00	5.00	5.00	2.00	3.00	4.00	11.35	0.80	3.00	1.00
P6	400.00	9.20	1.00	8.00	4.00	2.00	1.00	3.00	14.45	1.50	3.60	1.00
P7	500.00	8.80	1.00	6.00	4.00	0.00	2.50	2.50	12.30	2.60	2.40	1.00

---

# Cases

Table 2. Correlation coefficient matrix

	Total cost	Capacity	Target	Cont.	Key Technic	Tech. method	Hard Achievements	Soft Achievements	Cycle	Personal	Test	Hard Coefficient	Evaluation
Total cost	1	-.191	.603	-.145	-.344	-.084	.069	.023	.064	-.366	.604	-.623	.245
Capacity	-.191	1	-.651	-.330	-.400	.238	-.817*	-.255	-.335	.152	.113	.375	-.751
Target	.603	-.651	1	.050	-.303	-.240	.405	.303	.728	-.505	-.085	-.473	.320
Cont.	-.145	-.330	.050	1	-.031	.083	.633	-.031	-.293	-.294	.111	.646	.517
Key Technic	-.344	-.400	-.303	-.031	1	.301	.300	-.381	-.261	.754	-.110	-.214	.301
Tech. method	-.084	.238	-.240	.083	.301	1	-.047	-.968**	-.185	.580	-.141	.159	-.167
Hard Achievements	.069	-.817*	.405	.633	.300	-.047	1	.142	.021	-.321	-.156	-.062	.937**
Soft Achievements	.023	-.255	.303	-.031	-.381	-.968**	.142	1	.299	-.697	-.051	-.104	.211
Cycle	.064	-.335	.728	-.293	-.261	-.185	.021	.299	1	-.240	-.614	-.303	-.185
Personal	-.366	.152	-.505	-.294	.754	.580	-.321	-.697	-.240	1	-.043	-.044	-.350
Test	.604	.113	-.085	.111	-.110	-.141	-.156	-.051	-.614	-.043	1	-.121	.106
Hard Coefficient	-.623	.375	-.473	.646	-.214	.159	-.062	-.104	-.303	-.044	-.121	1	-.222
Evaluation	.245	-.751	.320	.517	.301	-.167	.937**	.211	-.185	-.350	.106	-.222	1

# Cases

**Table 3. Total variance explained**

<b>Component</b>	<b>Initial Eigenvalues</b>		
	<b>Total</b>	<b>Variance %</b>	<b>Cumulative %</b>
1	4.243	32.638	32.638
2	3.132	24.094	56.732
3	2.663	20.484	77.216
4	1.357	10.438	87.654
5	1.158	8.909	96.563
<b>Component</b>	<b>Extraction Sums of Squared Loadings</b>		
	<b>Total</b>	<b>Variance %</b>	<b>Cumulative %</b>
1	4.243	32.638	32.638
2	3.132	24.094	56.732
3	2.663	20.484	77.216
4	1.357	10.438	87.654
5	1.158	8.909	96.563

Extraction Method: Principal Component Analysis

# Cases

**Table 4. Eigenvectors of component factors**

<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>
.316	.123	-.414	.763	.278
-.439	-.846	.098	.054	-.042
.867	.334	-.191	.144	.253
-.039	.524	.824	.106	.170
-.123	.389	-.519	.459	.549
-.040	-.028	.112	.023	.987
.035	.984	.072	.007	.030
.959	-.030	-.128	.090	-.195
.859	-.096	-.225	-.432	.113
-.740	-.220	-.282	.126	.420
-.211	-.011	.029	.962	-.039
-.233	-.207	.927	-.197	-.013

# Cases

**Table 5. Principal component variables and cumulative contribution rate**

<b>Principal component variables</b>	<b>Z1</b>	<b>Z2</b>	<b>Z3</b>	<b>Z4</b>	<b>Z5</b>
	-0.09207	2.20741	-0.03292	0.31892	-0.11492
	0.09182	-0.43279	2.16027	0.04059	-0.43348
	2.09327	-0.31641	-0.43478	0.17937	0.65569
	-0.60315	-0.80143	-0.91059	0.38853	-0.72401
	-0.01967	-0.04103	-0.44810	-1.85994	-0.93470
	-1.01475	-0.25152	0.00249	-0.48750	1.95207
	-0.45546	-0.36423	-0.33637	1.42002	-0.40064
<b>cumulative contribution rate</b>	<b>32.638%</b>	<b>56.732%</b>	<b>77.216%</b>	<b>87.654%</b>	<b>96.563%</b>

$$Y_{\text{total cost}} = 42.320Z_1 + 16.515Z_2 - 55.354Z_3 + 102.124Z_4 + 37.259Z_5 + 425.540$$

# Cases

---

## □ Formula 6

$$Y_{\text{total cost}} = 42.320Z_1 + 16.515Z_2 - 55.354Z_3 + 102.124Z_4 + 37.259Z_5 + 425.540$$

---

# Forecast and results analysis

---

## Table 6. Principal component variables

<b>Z1</b>	<b>Z2</b>	<b>Z3</b>	<b>Z4</b>	<b>Z5</b>
6.15	2.13	13.25	0.28	-7.32
10.41	12.1	1.48	9.22	22.28
-1.4	0.51	3.53	-3.13	-5.94
-1.77	-0.52	-2.41	-3.76	0.83
8.72	11.28	15.66	12.77	1.09
6.15	2.13	13.25	0.28	-7.32
10.41	12.1	1.48	9.22	22.28

---

# Forecast and results analysis

---

**Table 7. Forecasting values V.S. true values**

<b>Code</b>	<b>True values</b>	<b>Forecasting values</b>	<b>Errors (%)</b>
P1	1,200.00	1079.436	10.05
P2	1,200.00	1054.54	12.12
P3	1,300.00	1152.73	11.33
P4	950.00	854.79	10.02
P5	1,000.00	937.98	6.20
P6	1,000.00	873.61	12.64
P7	2,090.00	1875.43	10.27

---



# Forecast and results analysis

---

- five principal component factors
    - accumulative contribution rate of 96.56% on the original data
    - fit into the cost regression equation
-

# Forecast and results analysis

---

- got another five principal component factors for the next phase
  - calculated the cost amount of the next stage of each program.
-

# Forecast and results analysis

---

Using this method

- the size of the influencing factors had been reduced effectively
  - equation with a smaller number of unrelated new variables had been established
  - most of the information provided by the original variables were reflected
-

# Thanks!



# **True Planning Application**

---

Yin Xiaolin  
**2013-10-24**

# Contents

---

**Introduction**

**Application**

**Question and Suggestion**

---

# Introduction

---

## □ difficulty

- poor data
  - data accumulation
  - validation
-

# Application

---

PRICE company

2008 --2011

---



# Application

---

□ comprehensive support -- PRICE teams

ships   aircrafts   engines

.....

---

# Application

# -- Training

---

- from 2008 to 2012  
more than ten training programs
  - last December  
Re-development training
  - application  
ongoing: costs estimating model for an engine  
application area is expanding
-

# Application -- Follow-up Service

---

- front-line service in time
    - basic principle of the software
    - requirement of data collection
    - building product breakdown structure
    - building cost estimating model
-

# Application

---

# --Objective

- ❑ to enrich the cost estimating cases of industrial department in China
  - ❑ to establish the product cost density database reflecting the Made-In-China level
  - ❑ to provide two cross-check methods with traditional one
-

# Question and Suggestion

---

## Prospective Development

What is the developing direction of TP in the near future?

what are the study emphases of TP for project cost estimation of higher quality?

---

# Question and Suggestion

---

## Practice Cases

- if PRICE could provide us with more reference information , such as:

EBS of various armaments

practice cases of cost estimating

---

# Question and Suggestion

---

## Re-Development

- provide more operation cases of re-development
    - to help the cost estimators be familiar with the module quickly and thoroughly
-

# Thanks!

