主成分分析）是一种工具可以用来找到一个低维度的表现方式来展示更多的变量, 主成分分析的目的是用一个低维度的图表来尽可能的解释所有的变量。

结合我们的数据表格，我们发现横向有４个与医院科室有关的变量，分别是内科，外科，儿科与妇科。纵向有23个与医疗纠纷原因有关的变量。 一共27个变量为我们提供了一个良好的数据环境来使用主成分分析。

在R里我们计算出四个科室的方差，我们可以看出外科的方差是135. 45。这个方差值比其他3个科室都要高出很多很多，这不仅有说明外科的数据离散程度大，而更说明了外科的数据和其他三个科室是在不同水平上的。

> apply(dispute, 2, var)

Medical Surgical Pediatric Gynecologic

37.877470 135.450593 3.173913 2.857708

一份数据的第一因子应该要有最大的方差，要找到第一因子，首先我们用 cov（）函数计算出 协方差矩阵， 然后我们用 eigen( ) 函数计算出协方差矩阵的特征值，在R里eigen（）函数可以同时计算出协方差矩阵的特征值和特征向量 。

dispute.cov <- cov(scaled\_df)

> dispute.eigen <- eigen(dispute.cov)

> str(dispute.eigen)

List of 2

$ values : num [1:4] 2.55 0.738 0.439 0.273

$ vectors: num [1:4, 1:4] -0.43 -0.518 -0.545 -0.499 0.798 ...

- attr(\*, "class")= chr "eigen"

任何统计软件里计算出得特征向量都比较特殊，是需要反转正负号的. 在R里也是一样，我们需要将结果乘以-1来翻转正负号，这样我们可以得到一个更符合逻辑，让人易懂的图表结果。

通过R我们可以找到第一因子与第二因子：

> phi <- -phi

> row.names(phi) <- c("Medical","Surgical","Pediatric","Gynecologic")

> colnames(phi) <- c("PC1","PC2")

> phi

PC1 PC2

Medical 0.4303044 -0.79758622

Surgical 0.5184091 -0.05858001

Pediatric 0.5446281 0.15369687

Gynecologic 0.4994700 0.58034634

第一因子与儿科和外科有较强的关联性，因为两个科室分别有着较大值：0.5446281 和 0.5184091

第二因子与儿科和妇科，尤其是妇科（0.58034634）有强烈的关联性，与内科和外科有负相关联性

接下来我们需要把两个因子代入每一个医疗纠纷（纵向变量）的原因中：

> PC1 <- as.matrix(scaled\_df)%\*%phi[,1]

> PC2 <- as.matrix(scaled\_df)%\*%phi[,2]

> PC <- data.frame(Factors = row.names(dispute), PC1,PC2)

> PC

Factors PC1 PC2

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

4 Malfunction -0.8948329 0.72729344

5 Option -0.3136041 0.43071321

6 Inspection -0.2395409 0.16559160

7 Tomography -0.4877384 0.43790386

8 Indication -1.7379918 0.32594133

9 Dose -0.8481234 0.49111366

10 Bill 1.2958187 -1.09154125

11 Hospitalization 0.5364168 -0.77482753

12 Efficiacy 2.7666265 1.31841598

13 Complication -1.2198141 0.64407808

14 Fatality -2.0786576 -0.74154044

15 Adverser -0.8400459 0.46522800

16 Miscommunication 0.8226273 -2.53234677

17 Charge -1.5405425 0.30275168

18 Attitude 0.6912792 1.31305314

19 Bureaucracy -0.6980449 -0.02510776

20 Confusion 2.2442988 -0.95620898

21 Ignorance -0.1758285 0.28008616

22 Expectation 4.7947896 0.50398118

23 Hysteria 0.6433542 -0.03177427

我们将因子1按从大到小排序，并且做出相应的中文表格

> PC1sorted <- PC[order(-PC1),]

> PC1sorted

Factors PC1 PC2

22 Expectation 4.7947896 0.50398118

12 Efficiacy 2.7666265 1.31841598

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

18 Attitude 0.6912792 1.31305314

23 Hysteria 0.6433542 -0.03177427

11 Hospitalization 0.5364168 -0.77482753

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

5 Option -0.3136041 0.43071321

7 Tomography -0.4877384 0.43790386

1 Malpractise -0.6880152 -0.03697162

19 Bureaucracy -0.6980449 -0.02510776

15 Adverser -0.8400459 0.46522800

9 Dose -0.8481234 0.49111366

4 Malfunction -0.8948329 0.72729344

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

13 Complication -1.2198141 0.64407808

17 Charge -1.5405425 0.30275168

8 Indication -1.7379918 0.32594133

14 Fatality -2.0786576 -0.74154044

|  |
| --- |
| Factors PC1 |
| 22 Expectation 4.7947896 患者诊疗期望高 |
| 12 Efficiacy 2.7666265 治疗效果 |
| 20 Confusion 2.2442988 医生交代不明确 |
| 10 Bill 1.2958187 医疗费用 |
| 16 Miscommunication 0.8226273 缺乏沟通 |
| 18 Attitude 0.6912792 服务态度差 |
| 23 Hysteria 0.6433542 患者无理取闹 |
| 11 Hospitalization 0.5364168 住院天数 |
| 21 Ignorance -0.1758285 患者医疗知识匮乏 |
| 6 Inspection -0.2395409 特殊检查未告知 |
| 5 Option -0.3136041 备选诊疗方案缺失 |
| 7 Tomography -0.4877384 影像学未结合体征 |
| 1 Malpractise -0.6880152 医疗技术过失或过错 |
| 19 Bureaucracy -0.6980449 诊疗流程繁杂 |
| 15 Adverser -0.8400459 不良反应 |
| 9 Dose -0.8481234 药物调配或剂量偏差 |
| 4 Malfunction -0.8948329 医疗器械故障 |
| 2 Reoperation -1.0066844 是否二次手术 |
| 3 Consultation -1.0257465 会诊 |
| 13 Complication -1.2198141 并发症 |
| 17 Charge -1.5405425 医疗费用 |
| 8 Indication -1.7379918 手术指征、手术窗口期选择错误 |
| 14 Fatality -2.0786576 死亡 |

我们将因子2按从大到小排序，并且做出相应的中文表格

> PC2sorted <- PC[order(-PC2),]

> PC2sorted

Factors PC1 PC2

12 Efficiacy 2.7666265 1.31841598

18 Attitude 0.6912792 1.31305314

4 Malfunction -0.8948329 0.72729344

13 Complication -1.2198141 0.64407808

22 Expectation 4.7947896 0.50398118

9 Dose -0.8481234 0.49111366

15 Adverser -0.8400459 0.46522800

7 Tomography -0.4877384 0.43790386

5 Option -0.3136041 0.43071321

8 Indication -1.7379918 0.32594133

17 Charge -1.5405425 0.30275168

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

19 Bureaucracy -0.6980449 -0.02510776

23 Hysteria 0.6433542 -0.03177427

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

14 Fatality -2.0786576 -0.74154044

11 Hospitalization 0.5364168 -0.77482753

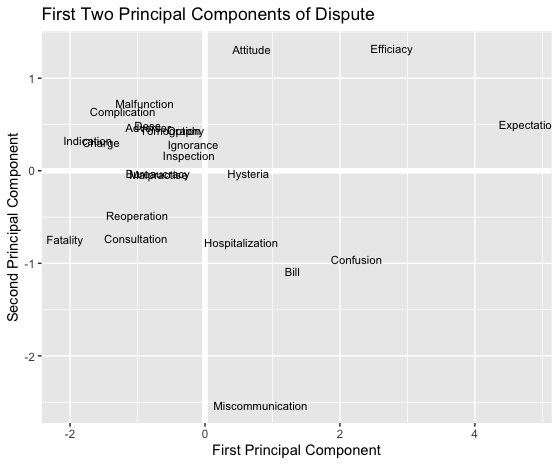
20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

|  |
| --- |
| Factors PC2 |
| 12 Efficiacy 治疗效果 1.31841598 |
| 18 Attitude 服务态度差 1.31305314 |
| 4 Malfunction 医疗器械故障 0.72729344 |
| 13 Complication 并发症 0.64407808 |
| 22 Expectation 患者诊疗期望高 0.50398118 |
| 9 Dose 药物调配或剂量偏差 0.49111366 |
| 15 Adverser 不良反应 0.46522800 |
| 7 Tomography 影像学未结合体征 0.43790386 |
| 5 Option 备选诊疗方案缺失 0.43071321 |
| 8 Indication 手术指征、手术窗口期选择错误 0.32594133 |
| 17 Charge 违规收费 0.30275168 |
| 21 Ignorance 患者医疗知识匮乏 0.28008616 |
| 6 Inspection 特殊检查未告知 0.16559160 |
| 19 Bureaucracy 诊疗流程繁杂 -0.02510776 |
| 23 Hysteria 患者无理取闹 -0.03177427 |
| 1 Malpractise 医疗技术过失或过错 -0.03697162 |
| 2 Reoperation 是否二次手术 -0.48686032 |
| 3 Consultation 会诊 -0.72897236 |
| 14 Fatality 死亡 -0.74154044 |
| 11 Hospitalization 住院天数 -0.77482753 |
| 20 Confusion 医生交代不明确 -0.95620898 |
| 10 Bill 医疗费用 -1.09154125 |
| 16 Miscommunication 缺乏沟通 -2.53234677 |

我们现在计算出了每一个纠纷原因的第一和第二因子，将他们做图得出：



第一因子（x轴）大致指的是外科和儿科的比例，在图中我们可以看到Expectation（患者诊疗期望高），Efficiacy（治疗效果）与外科和儿科的医疗纠纷比例有强烈的正相关性，而Fatality（死亡）Indication（手术指针，窗口期选择错误）有负相关性

第二因子（y轴）大致指的是儿科和妇科的比列，尤其是妇科。 在图中我们可以看到Attitude（服务态度差）Efficiacy（治疗效果）与妇科儿科的医疗纠纷比列有强烈的正相关性，而Miscommunication（缺乏沟通）则与妇科和儿科的了了纠纷比例有强烈的负相关性

从整体看，图右斜上方的两个因素Expectation（患者诊疗期望高），Efficiacy（治疗效果）与因子1，2都有正相关性，图左斜下方的几个因素Fatality(死亡), Reoperation(二次手术), Consultation(会诊)与因子1，2都有负相关性

趋于x，y轴0处的Ignorance(患者医疗知识匮乏), Inspection(特殊检查未告知), Hysteria(无理取闹) 这几个因素并没有与科室的种类有过多的关联性

Appendix

> Medical <- c(4,7,10,2,6,7,4,1,6,19,15,7,1,6,2,21,2,5,8,18,6,18,13)

> Surgical <- c(21,16,4,6,13,6,12,9,1,15,4,33,14,7,7,28,,5,17,1,31,9,48,10)

Error in c(21, 16, 4, 6, 13, 6, 12, 9, 1, 15, 4, 33, 14, 7, 7, 28, , 5, :

> Pediatric <- c(2,1,2,3,2,4,3,1,2,4,5,7,1,0,4,3,2,3,3,5,4,7,3)

> Gynecologic <- c(1,1,1,2,3,2,2,1,3,3,2,5,2,0,1,0,1,5,2,3,2,7,4)

> dispute <- data.frame(Medical, Surgical, Pediatric, Gynecologic)

Error in data.frame(Medical, Surgical, Pediatric, Gynecologic) :

> Surgical <- c(21,16,4,6,13,6,12,9,1,15,4,33,14,7,7,28,5,17,1,31,9,48,10)

> dispute <- data.frame(Medical, Surgical, Pediatric, Gynecologic)

> dispute

Medical Surgical Pediatric Gynecologic

1 4 21 2 1

2 7 16 1 1

3 10 4 2 1

4 2 6 3 2

5 6 13 2 3

6 7 6 4 2

7 4 12 3 2

8 1 9 1 1

9 6 1 2 3

10 19 15 4 3

11 15 4 5 2

12 7 33 7 5

13 1 14 1 2

14 6 7 0 0

15 2 7 4 1

16 21 28 3 0

17 2 5 2 1

18 5 17 3 5

19 8 1 3 2

20 18 31 5 3

21 6 9 4 2

22 18 48 7 7

23 13 10 3 4

> row.names(dispute) <- c("Malpractise", "Reoperation", "Consultation", "Malfunction", "Option", "Inspection", "Tomography","Indication", "Dose", "Bill", "Hospitalization", "Efficiacy", "Complication", "Fatality", "Adverser", "Miscommunication", "Charge", "Attitude", "Bureaucracy", "Confusion", "Ignorance", "Expectation", "Hysteria")

> dispute

Medical Surgical Pediatric Gynecologic

Malpractise 4 21 2 1

Reoperation 7 16 1 1

Consultation 10 4 2 1

Malfunction 2 6 3 2

Option 6 13 2 3

Inspection 7 6 4 2

Tomography 4 12 3 2

Indication 1 9 1 1

Dose 6 1 2 3

Bill 19 15 4 3

Hospitalization 15 4 5 2

Efficiacy 7 33 7 5

Complication 1 14 1 2

Fatality 6 7 0 0

Adverser 2 7 4 1

Miscommunication 21 28 3 0

Charge 2 5 2 1

Attitude 5 17 3 5

Bureaucracy 8 1 3 2

Confusion 18 31 5 3

Ignorance 6 9 4 2

Expectation 18 48 7 7

Hysteria 13 10 3 4

> apply(dispute, 2, var)

Medical Surgical Pediatric Gynecologic

37.877470 135.450593 3.173913 2.857708

> scaled\_df <- apply(dispute, 2, scale)

> scaled\_df

Medical Surgical Pediatric Gynecologic

[1,] -0.67819239 0.62014007 -0.61011916 -0.7715863

[2,] -0.19074161 0.19052496 -1.17142878 -0.7715863

[3,] 0.29670917 -0.84055131 -0.61011916 -0.7715863

[4,] -1.00315958 -0.66870526 -0.04880953 -0.1800368

[5,] -0.35322520 -0.06724410 -0.61011916 0.4115127

[6,] -0.19074161 -0.66870526 0.51250009 -0.1800368

[7,] -0.67819239 -0.15316713 -0.04880953 -0.1800368

[8,] -1.16564317 -0.41093619 -1.17142878 -0.7715863

[9,] -0.35322520 -1.09832037 -0.61011916 0.4115127

[10,] 1.75906152 0.10460194 0.51250009 0.4115127

[11,] 1.10912714 -0.84055131 1.07380971 -0.1800368

[12,] -0.19074161 1.65121634 2.19642896 1.5946118

[13,] -1.16564317 0.01867892 -1.17142878 -0.1800368

[14,] -0.35322520 -0.58278224 -1.73273840 -1.3631359

[15,] -1.00315958 -0.58278224 0.51250009 -0.7715863

[16,] 2.08402870 1.22160123 -0.04880953 -1.3631359

[17,] -1.00315958 -0.75462828 -0.61011916 -0.7715863

[18,] -0.51570880 0.27644798 -0.04880953 1.5946118

[19,] -0.02825802 -1.09832037 -0.04880953 -0.1800368

[20,] 1.59657792 1.47937030 1.07380971 0.4115127

[21,] -0.35322520 -0.41093619 0.51250009 -0.1800368

[22,] 1.59657792 2.94006168 2.19642896 2.7777108

[23,] 0.78415995 -0.32501317 -0.04880953 1.0030622

> dispute.cov <- cov(scaled\_df)

> dispute.eigen <- eigen(dispute.cov)

> str(dispute.eigen)

List of 2

$ values : num [1:4] 2.55 0.738 0.439 0.273

$ vectors: num [1:4, 1:4] -0.43 -0.518 -0.545 -0.499 0.798 ...

- attr(\*, "class")= chr "eigen"

> phi <- dispute.eigen$vectors[,1:2]

> phi

[,1] [,2]

[1,] -0.4303044 0.79758622

[2,] -0.5184091 0.05858001

[3,] -0.5446281 -0.15369687

[4,] -0.4994700 -0.58034634

> phi <- -phi

> row.names(phi) <- c("Medical","Surgical","Pediatric","Gynecologic")

> colnames(phi) <- c("PC1","PC2")

> phi

PC1 PC2

Medical 0.4303044 -0.79758622

Surgical 0.5184091 -0.05858001

Pediatric 0.5446281 0.15369687

Gynecologic 0.4994700 0.58034634

> PC1 <- as.matrix(scaled\_df)%\*%phi[,1]

> PC2 <- as.matrix(scaled\_df)%\*%phi[,2]

> PC <- data.frame(Factors = row.names(dispute), PC1,PC2)

> PC

Factors PC1 PC2

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

4 Malfunction -0.8948329 0.72729344

5 Option -0.3136041 0.43071321

6 Inspection -0.2395409 0.16559160

7 Tomography -0.4877384 0.43790386

8 Indication -1.7379918 0.32594133

9 Dose -0.8481234 0.49111366

10 Bill 1.2958187 -1.09154125

11 Hospitalization 0.5364168 -0.77482753

12 Efficiacy 2.7666265 1.31841598

13 Complication -1.2198141 0.64407808

14 Fatality -2.0786576 -0.74154044

15 Adverser -0.8400459 0.46522800

16 Miscommunication 0.8226273 -2.53234677

17 Charge -1.5405425 0.30275168

18 Attitude 0.6912792 1.31305314

19 Bureaucracy -0.6980449 -0.02510776

20 Confusion 2.2442988 -0.95620898

21 Ignorance -0.1758285 0.28008616

22 Expectation 4.7947896 0.50398118

23 Hysteria 0.6433542 -0.03177427

> PC1.sorted

[,1]

[1,] 4.7947896

[2,] 2.7666265

[3,] 2.2442988

[4,] 1.2958187

[5,] 0.8226273

[6,] 0.6912792

[7,] 0.6433542

[8,] 0.5364168

[9,] -0.1758285

[10,] -0.2395409

[11,] -0.3136041

[12,] -0.4877384

[13,] -0.6880152

[14,] -0.6980449

[15,] -0.8400459

[16,] -0.8481234

[17,] -0.8948329

[18,] -1.0066844

[19,] -1.0257465

[20,] -1.2198141

[21,] -1.5405425

[22,] -1.7379918

[23,] -2.0786576

> PC1sorted <- PC[order(-PC1),]

> PC1sorted

Factors PC1 PC2

22 Expectation 4.7947896 0.50398118

12 Efficiacy 2.7666265 1.31841598

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

18 Attitude 0.6912792 1.31305314

23 Hysteria 0.6433542 -0.03177427

11 Hospitalization 0.5364168 -0.77482753

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

5 Option -0.3136041 0.43071321

7 Tomography -0.4877384 0.43790386

1 Malpractise -0.6880152 -0.03697162

19 Bureaucracy -0.6980449 -0.02510776

15 Adverser -0.8400459 0.46522800

9 Dose -0.8481234 0.49111366

4 Malfunction -0.8948329 0.72729344

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

13 Complication -1.2198141 0.64407808

17 Charge -1.5405425 0.30275168

8 Indication -1.7379918 0.32594133

14 Fatality -2.0786576 -0.74154044

> PC2sorted <- PC[order(-PC2),]

> PC2sorted

Factors PC1 PC2

12 Efficiacy 2.7666265 1.31841598

18 Attitude 0.6912792 1.31305314

4 Malfunction -0.8948329 0.72729344

13 Complication -1.2198141 0.64407808

22 Expectation 4.7947896 0.50398118

9 Dose -0.8481234 0.49111366

15 Adverser -0.8400459 0.46522800

7 Tomography -0.4877384 0.43790386

5 Option -0.3136041 0.43071321

8 Indication -1.7379918 0.32594133

17 Charge -1.5405425 0.30275168

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

19 Bureaucracy -0.6980449 -0.02510776

23 Hysteria 0.6433542 -0.03177427

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

14 Fatality -2.0786576 -0.74154044

11 Hospitalization 0.5364168 -0.77482753

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

> library(ggplot2)

> ggplot(PC, aes(PC1, PC2)) +

+

> ggplot(PC, aes(PC1, PC2)) +

+ modelr::geom\_ref\_line(h = 0) +

+ modelr::geom\_ref\_line(v = 0) +

+ geom\_text(aes(label = Factors), size = 3) +

+ xlab("First Principal Component") +

+ ylab("Second Principal Component") +

+ ggtitle("First Two Principal Components of Dispute")