

# 多元-09-2020270026

2020270026 王姿文

5/6/2021

## 1. 数据

- 数据叙述：2011年昆明市14个区县三大产业的产值情况
- 目标： 进行对应分析

下表为数据，以及数据的结构：

```
df <- read_excel("ex8.5.xls")
a <- df$区县名
df <- df[,-1]
rownames(df) <- a
kbl(df) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, font_size =
7)
```

	第一产业	第二产业	第三产业
五华	16897	3404530	2660319
盘龙	41404	876511	2100630
官渡	78664	2144271	3269383
西山	28926	857806	1993830
东川	37352	380106	143279
呈贡	60462	426383	360842
晋宁	136011	356203	181367
富民	68393	167256	104677
宜良	303725	298542	470218
石林	119493	139789	187173
嵩明	91599	277955	149335
禄劝	125943	106289	154621
寻甸	134828	136507	184797
安宁	88478	992527	596705

```
str(df)
```

```
## tibble[,3] [14 × 3] (S3: tbl_df/tbl/data.frame)
## $ 第一产业: num [1:14] 16897 41404 78664 28926 37352 ...
## $ 第二产业: num [1:14] 3404530 876511 2144271 857806 380106 ...
## $ 第三产业: num [1:14] 2660319 2100630 3269383 1993830 143279 ...
```

## 2. 对应分析

首先做卡方独立性检定，发现p-value<0.05，故区县与产业间不独立

```
chisq.test(df)
```

```
##  
## Pearson's Chi-squared test  
##  
## data: df  
## X-squared = 5341756, df = 26, p-value < 0.00000000000000022
```

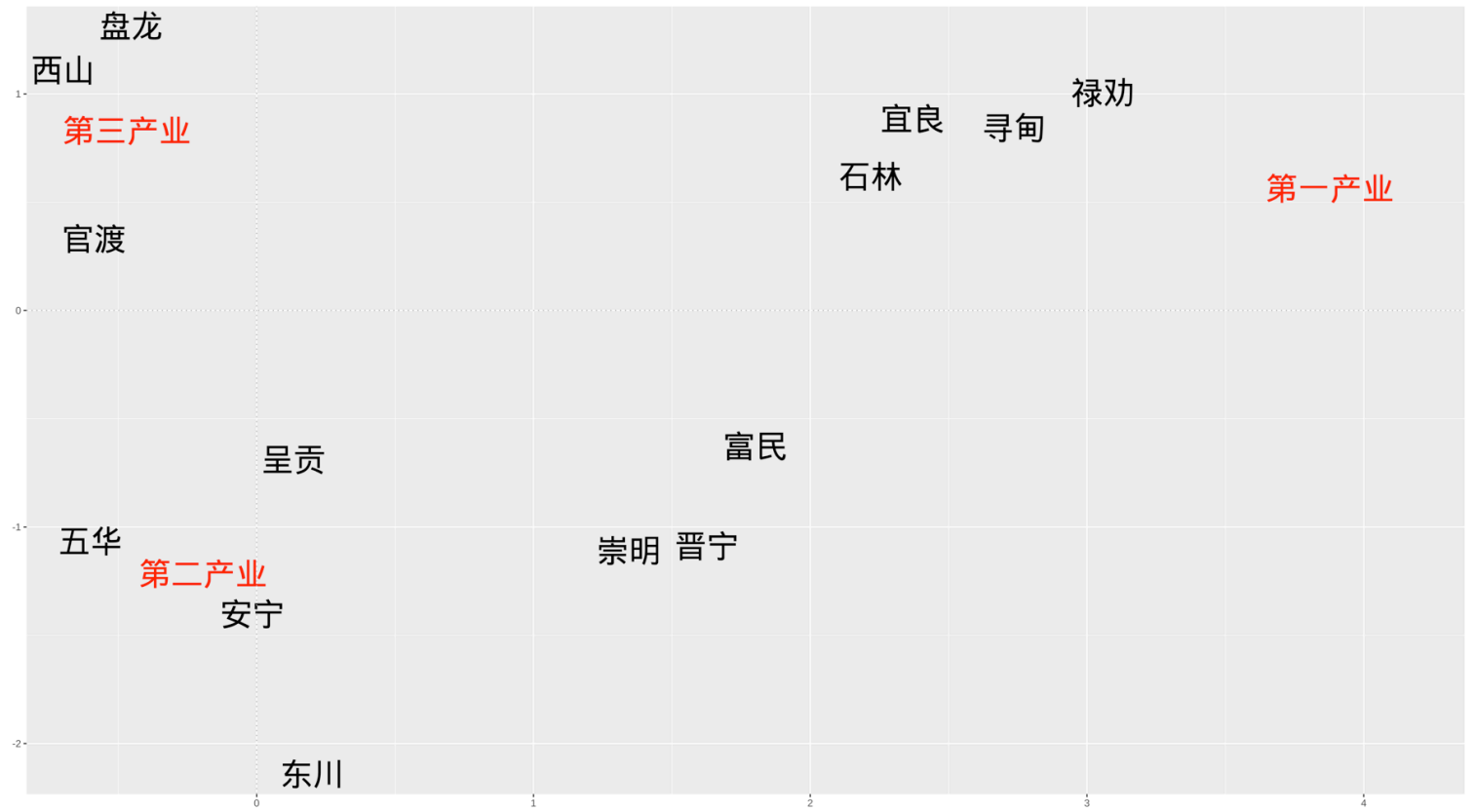
开始做对应分析，首先来看降维后的区县和产业间的典型相关分析，可以看出第一个维度的典型相关系数比第二个维度高

```
co <- corresp(df, nf=2)  
co$cor
```

```
## [1] 0.3978404 0.2452832
```

- 在第一维度上，产业的次序完全正确，反映了三个行类在对应分析图中的接近关系是完全正确的
- 第二产业和第三产业在第一维度上较为接近，然而第一产业和第三产业在第二维度上较为接近
- 区县和产业之间的关系：
  - 第一产业：宜良、石林、寻甸、禄劝，要注意的是第一产业与这四个区县远离原点且靠得近，因此关联较强
  - 第二产业：安宁、五华，区县与产业的关联强度没有第一产业来得高
  - 第三产业：关渡、西山、盘龙，区县与产业的关联强度没有第一产业来得高
  - 其他：
    - 崇明、富民、晋宁：这三个区县的产业结构十分类似，但不明显属于哪一类产业，最接近第二产业，再来才是第三产业
    - 东川：产业结构是第二产业只是关联没这么强
    - 呈贡：比起崇明、富民、晋宁，更接近第二产业和第三产业，一样不明显属于哪一类产业，最接近第二产业，再来才是第三产业，靠近原点，所以其产业结构和平均的结构相近

```
ggbiplot.corresp <- function(obj) { require(ggplot2)  
  require(ggrepel)  
  require(tibble)  
  rscore <- tibble(  
    label = rownames(obj$rscore), x = obj$rscore[,1],  
    y = obj$rscore[,2]  
  )  
  cscore <- tibble(  
    label = rownames(obj$cscore), x = obj$cscore[,1],  
    y = obj$cscore[,2]  
  )  
  p <- ggplot(mapping = aes(x = x, y = y, label=label)) +  
    geom_text_repel(data = rscore, color="black",size=10) + geom_text_repel(data = cscore,  
  color="red",size=10) + geom_hline(yintercept=0, linetype=3, col="gray") + geom_vline(xintercept=0, linetype=3, col="gray") + labs(x=NULL, y=NULL)  
  p }  
  
ggbiplot.corresp(co)
```

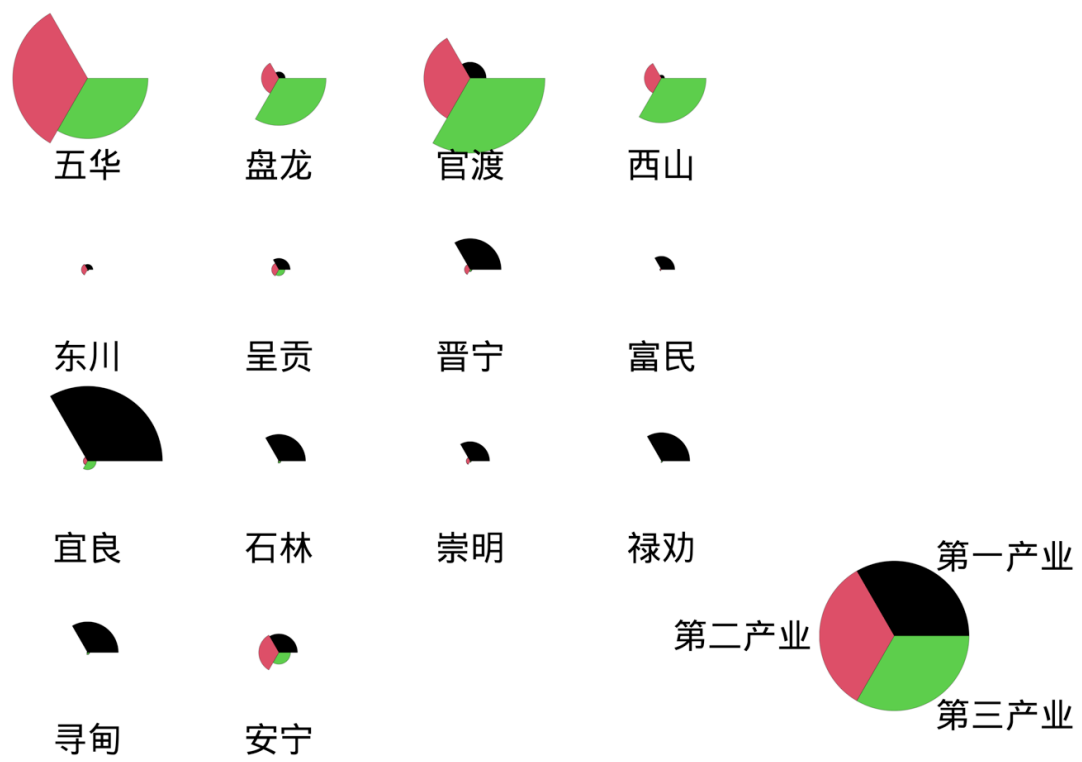


### 3. 多元图形辅助分析

#### 3.1 星图

因为归一化，所以不如对应分析结果可解释，但还是能看出不同区县内的产业占比

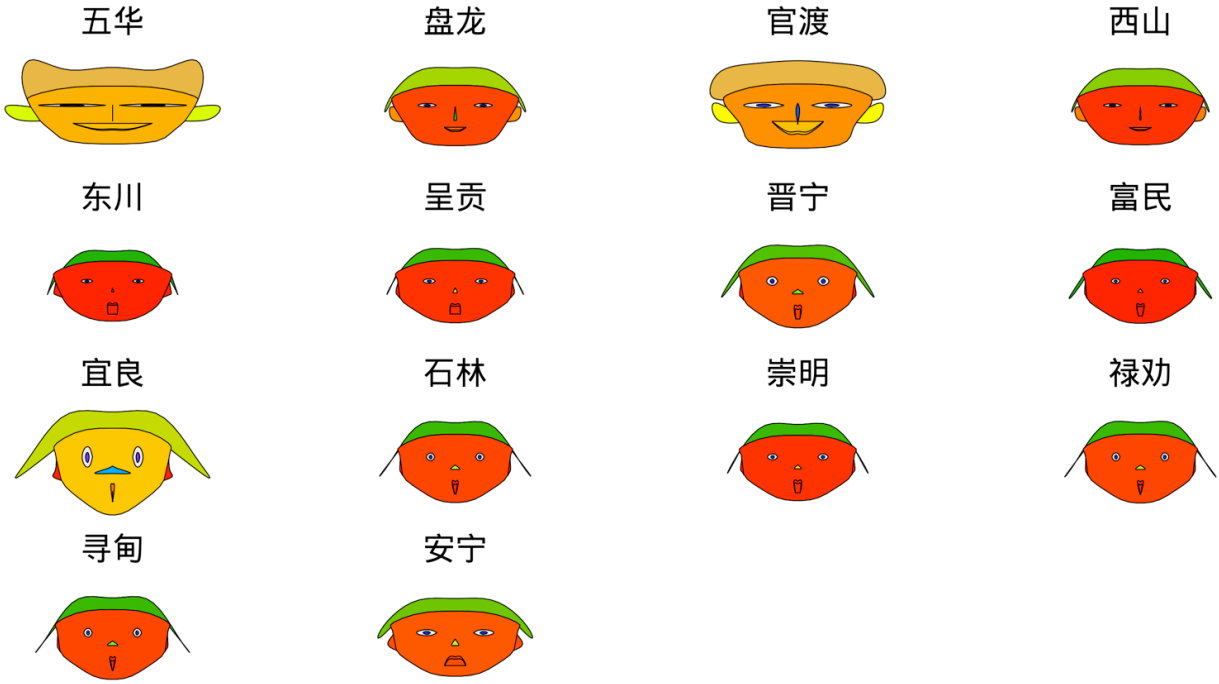
```
stars(df, len=0.9, cex=2.5, key.loc=c(12,2.5),draw.segments=TRUE,
      labels = paste(row.names(df)))
```



#### 3.2 臉譜圖

脸谱图可以明显看出哪些区县在不同的特征内的数值相近，但是无法来解释区县和产业间的关系

```
aplpack::faces(df,cex = 3.5)
```



```
## effect of variables:
## modified item      Var
## "height of face"   " 第一产业"
## "width of face"    " 第二产业"
## "structure of face" " 第三产业"
## "height of mouth"  " 第一产业"
## "width of mouth"   " 第二产业"
## "smiling"          " 第三产业"
## "height of eyes"   " 第一产业"
## "width of eyes"    " 第二产业"
## "height of hair"   " 第三产业"
## "width of hair"    " 第一产业"
## "style of hair"     " 第二产业"
## "height of nose"   " 第三产业"
## "width of nose"    " 第一产业"
## "width of ear"     " 第二产业"
## "height of ear"    " 第三产业"
```

## 4. 小結

由多元图形辅助分析可以看出，若想单看区县和产业间的结构关系，则使用对应分析是最好辨别和解释的，虽然星图和脸谱图或多或少也能解释不同的面向，然而效果都没对应分析来得好，这也展现出对应分析的优势