

# exercise1

WXZ

## 目录

<b>1</b>	<b>exercise1</b>	<b>1</b>
1.1	Q1: 数据的维度、观测值和覆盖年份 . . . . .	1
1.2	Q2:VEP 和 VAP 的投票率 . . . . .	3
1.3	Q3: VAP、VEP 与 ANES 的差异 . . . . .	4
1.4	Q4: ANES 在总统大选年和中期选举年与 VEP 投票率的误差	6
1.5	Q5: ANES 在上半段和下半段时间与 VEP 投票率的误差 . .	8
1.6	Q6:2008 年 vapPercentage 的调整结果 . . . . .	10

## 1 exercise1

### 1.1 Q1: 数据的维度、观测值和覆盖年份

#### 1.1.1 代码实现

```
dim(turnout) # 输出数据的维度  
summary(turnout) # 观察数据的统计量  
turnout # 观测值  
year <- turnout$year # 获取年份信息  
year.gap <- range(year) # 获取覆盖年份
```

## 1.1.2 结果

```

> dim(turnout) # 输出数据的维度
[1] 14 9
-----
> summary(turnout) # 观察数据的统计量
      year      VEP      VAP      total
Min.   :1980   Min.   :159635   Min.   :164445   Min.   : 64991
1st Qu.:1986   1st Qu.:171192   1st Qu.:178930   1st Qu.: 73179
Median :1993   Median :181140   Median :193018   Median : 89055
Mean   :1993   Mean   :182640   Mean   :194226   Mean   : 89778
3rd Qu.:2000   3rd Qu.:193353   3rd Qu.:209296   3rd Qu.:102370
Max.   :2008   Max.   :213314   Max.   :230872   Max.   :131304

      ANES      felons      noncit      overseas
Min.   :47.00   Min.   : 802   Min.   : 5756   Min.   :1803
1st Qu.:57.00   1st Qu.:1424   1st Qu.: 8592   1st Qu.:2236
Median :70.50   Median :2312   Median :11972   Median :2458
Mean   :65.79   Mean   :2177   Mean   :12229   Mean   :2746
3rd Qu.:73.75   3rd Qu.:3042   3rd Qu.:15910   3rd Qu.:2937
Max.   :78.00   Max.   :3168   Max.   :19392   Max.   :4972

      osvoters
Min.   :263
1st Qu.:263
Median :263
Mean   :263
3rd Qu.:263
Max.   :263
NA's   :13 -----
-----
> turnout # 观测值
      year      VEP      VAP      total ANES felons noncit overseas osvoters
1  1980 159635 164445 86515 71 802 5756 1803 NA
2  1982 160467 166028 67616 60 960 6641 1982 NA
3  1984 167702 173995 92653 74 1165 7482 2361 NA
4  1986 170396 177922 64991 53 1367 8362 2216 NA
5  1988 173579 181955 91595 70 1594 9280 2257 NA
6  1990 176629 186159 67859 47 1901 10239 2659 NA
7  1992 179656 190778 104405 75 2183 11447 2418 NA
8  1994 182623 195258 75106 56 2441 12497 2229 NA
9  1996 186347 200016 96263 73 2586 13601 2499 NA
10 1998 190420 205313 72537 52 2920 14988 2937 NA
11 2000 194331 210623 105375 73 3083 16218 2937 NA
12 2002 198382 215462 78382 62 3168 17237 3308 NA
13 2004 203483 220336 122295 77 3158 18068 3862 NA
14 2008 213314 230872 131304 78 3145 19392 4972 263

> year.gap
[1] 1980 2008

```

## 1.2 Q2:VEP 和 VAP 的投票率

### 1.2.1 计算方法

$\text{vapPercentage} = \text{投票总人数} / \text{符合投票年龄的人数}$

**VAP** 的符合投票年龄的人数需要加上海外选民的数量。

$\text{vepPercentage} = \text{投票总人数} / \text{符合投票条件的人数}$

而后对两个数据值的差值进行平均值、方差、标准差、最小值、最大值等的计算

### 1.2.2 代码实现

```
# 计算总符合投票年龄的投票人数

totalVoters <- turnout$VAP + turnout$overseas

vapPercentage <- (turnout$total / totalVoters) * 100

vepPercentage <- (turnout$total / turnout$VEP) * 100

difPercentage <- vapPercentage - vepPercentage

summary(difPercentage)

sd(difPercentage)

var(difPercentage)
```

### 1.2.3 结果

```

year vapPercentage vepPercentage
1 1980      52.03972      54.19551
2 1982      40.24522      42.13701
3 1984      52.53748      55.24860
4 1986      36.07845      38.14115
5 1988      49.72260      52.76848
6 1990      35.93884      38.41895
7 1992      54.04097      58.11384
8 1994      38.03086      41.12625
9 1996      47.53376      51.65793
10 1998      34.83169      38.09316
11 2000      49.34211      54.22449
12 2002      35.82850      39.51064
13 2004      54.54777      60.10084
14 2008      55.67409      61.55433

> summary(difPercentage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.880  -4.111  -3.178  -3.493  -2.538  -1.892
> sd(difPercentage)    # 计算标准差
[1] 1.273708
> var(difPercentage)   # 计算方差
[1] 1.622331

```

## 1.3 Q3: VAP、VEP 与 ANES 的差异

### 1.3.1 计算方法

定义一个函数，用于计算两个数据的差

定义一个函数，用于计算数据的平均值、最大值、标准差和方差

调用函数

### 1.3.2 代码实现

```

DifPerCalculate <- function(x, y){
# 计算两个数据的差和它们的差之间的平均值、范围、方差和极差

# Args:

#   x: numeric

#   y: numeric

```

```
# Returns:

# 返回数据差 dif
return(x - y)
}

DifPerSummaryCal <- function(dif){
# 计算两个数据的差的平均值、最大值、最小值、标准差和方差

# Args:

# 向量 vector

# Returns:

# 返回平均值、最大值、最小值、标准差和方差

dif.mean <- mean(dif)

dif.max <- max(dif)

dif.min <- min(dif)

dif.var <- var(dif)

dif.sd <- sd(dif)

dif.summary <- c(dif.mean, dif.max, dif.min, dif.sd, dif.var)

names(dif.summary) <- c("mean", "max", "min", "sd", "var")

return(dif.summary)
}

计算 vap 与 anes 投票率的差异

difAnesVap <- DifPerCalculate(turnout$ANES, vapPercentage)

difAnesVap

计算差值的平均值、最大值、最小值、标准差和方差

DifPerSummaryCal(difAnesVap)

计算 vep 与 anes 投票率的差异
```

```
difAnesVep <- DifPerCalculate(turnout$ANES, vepPercentage)
```

```
difAnesVep
```

计算差值的平均值、最大值、最小值、标准差和方差

```
DifPerSummaryCal(difAnesVep)
```

### 1.3.3 结果

```
> difAnesvap
[1] 18.96028 19.75478 21.46252 16.92155 20.27740 11.06116 20.95903
[8] 17.96914 25.46624 17.16831 23.65789 26.17150 22.45223 22.32591

> difAnesvep
[1] 16.804491 17.862987 18.751404 14.858846 17.231520 8.581054
[7] 16.886160 14.873745 21.342072 13.906838 18.775507 22.489359
[13] 16.899156 16.445672

> DifPerSummaryCal(difAnesvep)
      mean      max      min      sd      var
16.836344 22.489359 8.581054 3.346069 11.196177

> DifPerSummaryCal(difAnesvap)
      mean      max      min      sd      var
20.329138 26.171504 11.061159 3.893322 15.157959
```

## 1.4 Q4: ANES 在总统大选年和中期选举年与 VEP 投票率的误差

### 1.4.1 计算方法

通过等差数列获取相应的总统大选年和中期选举年的 ANES 和 VEP 投票率的数据。

中期选举的年份与总统大选的年份相间隔，1980 年是总统大选年，数据缺少 2006 年的。2006 年是中期选举年。

使用 Q3 定义的函数计算差值以及差值的平均值、最大值、最小值、标准差和方差。

### 1.4.2 代码实现

中期选举年的 vepPercentage 和 ANES 数据

```
vepPercentageCenter <- vepPercentage[seq(from = 2, to=13,
by=2)]
```

```
anesCenter <- turnout$ANES[seq(from = 2, to=13, by=2)]
```

总统大选年的 vepPercentage 和 ANES 数据

```
vepPercentagePre <- c(vepPercentage[seq(from = 1, to=14,
by=2)], vepPercentage[14])
```

```
anesPre <- c(turnout$ANES[seq(from = 1, to=14, by=2)], turnout$ANES[14])
```

计算差值

```
difCenterAnesVep <- DifPerCalculate(anesCenter, vepPercentageCenter)
```

```
difPreAnesVep <- DifPerCalculate(anesPre, vepPercentagePre)
```

计算差值的平均值、最大值、最小值、均值、标准差和方差

```
DifPerSummaryCal(difCenterAnesVep)
```

```
DifPerSummaryCal(difPreAnesVep)
```

### 1.4.3 结果

```
> vepPercentageCenter
[1] 42.13701 38.14115 38.41895 41.12625 38.09316 39.51064

> anesCenter
[1] 60 53 47 56 52 62

> vepPercentagePre
[1] 54.19551 55.24860 52.76848 58.11384 51.65793 54.22449 60.10084
[8] 61.55433

> anesPre
[1] 71 74 70 75 73 73 77 78

> difCenterAnesVep
[1] 17.862987 14.858846 8.581054 14.873745 13.906838 22.489359

> difPreAnesVep
[1] 16.80449 18.75140 17.23152 16.88616 21.34207 18.77551 16.89916
[8] 16.44567

> DifPerSummaryCal(difCenterAnesVep)
      mean      max      min      sd      var
15.428805 22.489359 8.581054 4.596035 21.123537
>
> DifPerSummaryCal(difPreAnesVep)
      mean      max      min      sd      var
17.891998 21.342072 16.445672 1.653152 2.732911
```

## 1.5 Q5: ANES 在上半段和下半段时间与 VEP 投票率的误差

### 1.5.1 计算方法

通过索引获取相应的上半段和下半段时间的 ANES 和 VEP 投票率的数据。

使用 Q3 定义的函数计算差值以及差值的平均值、最大值、最小值、标准差和方差。

画图表示结果。

### 1.5.2 代码实现

获取上下半段的数据并计算差值

```
difAnesVep1 <- DifPerCalculate(turnout$ANES[1:7], vepPercentage[1:7])
```

```
difAnesVep2 <- DifPerCalculate(turnout$ANES[8:14], vepPercentage[8:14])
```

计算 difAnesVep1 和 difAnesVep2 的平均值、最大值、最小值、标准差和方差

```
DifPerSummaryCal(difAnesVep1)
```

```
DifPerSummaryCal(difAnesVep2)
```

画图

```
ID<-c(1:7)
```

```
my_frame<-data.frame(ID, difAnesVep1, difAnesVep2)
```

```
plot(difAnesVep1~ID,pch=15,col="DarkTurquoise",ylim=c(8,23),ylab="dif")
```

```
points(ID, difAnesVep1, pch=16, col="DeepPink",cex=1)
```

```
lines(difAnesVep1, col="DarkTurquoise", lty=1)
```

```
lines(difAnesVep2, col="DeepPink", lty=2)
```

```
legend("topleft", c("difAnesVep1", "difAnesVep2"), col=c("DarkTurquoise",  
"DeepPink"), text.col=c("DarkTurquoise", "DeepPink"), pch=c(15,16),lty=c(1,2))
```

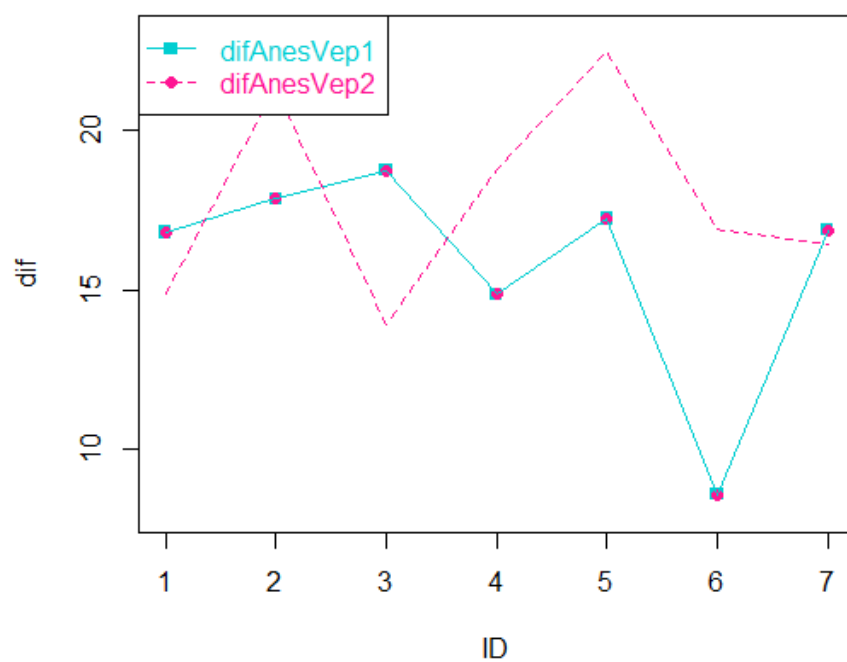


## 1.5.3 结果

```
> difAnesVep1
[1] 16.804491 17.862987 18.751404 14.858846 17.231520 8.581054 16.886160

> difAnesVep2
[1] 14.87375 21.34207 13.90684 18.77551 22.48936 16.89916 16.44567

> DifPerSummaryCal(difAnesVep1)
      mean      max      min      sd      var
15.853781 18.751404 8.581054 3.420122 11.697234
>
> DifPerSummaryCal(difAnesVep2)
      mean      max      min      sd      var
17.818907 22.489359 13.906838 3.210682 10.308479
```



## 1.6 Q6:2008 年 vapPercentage 的调整结果

### 1.6.1 计算方法

通过减去非公民和犯罪的人数获取调整后的 2008 年 VAP（符合投票年龄）的人数。

通过减去海外选民的投票数获取调整后的投票总人数。

$\text{vapPercentage} = \text{调整后的投票总人数} / \text{调整后的符合投票年龄的人数}$

通过 Q3 的函数对调整后的 2008 年 vapPercentage 与原来的值、2008 年 vepPercentage、2008 年 ANES 的值进行差异计算。

### 1.6.2 代码实现

获取 VAP2008 年调整后的人数

```
vap2008New <- turnout$VAP[14] - turnout$felons[14] - turnout$noncit[14]
```

计算 VAP2008 年调整后的投票率

```
vapPercentage2008New <- ((turnout$total[14] - turnout$osvoters[14])  
/ vap2008New) * 100
```

计算差值

```
dif2008 <- data.frame(DifPerCalculate(vapPercentage2008New,vapPercentage2008Old),  
DifPerCalculate(vapPercentage2008New,vepPercentage2008),  
DifPerCalculate(vapPercentage2008New,anes2008) )  
names(dif2008) <- c("NewOld", "NewVep", "NewAnes")
```

### 1.6.3 结果

```
      vapPercentage2008New vapPercentage2008Old vepPercentage2008 anes2008  
1      63.02542      55.67409      61.55433      78  
  
> dif2008  
      NewOld NewVep NewAnes  
1 7.351326 1.471087 -14.97458
```