POL 345: Quantitative Analysis and Politics

Precept 1 (Week 2)

September 26, 2011

In this precept, you will learn the following new materials:

- Determining object class with `class()`
- Identifying unique categories of a factor variable with `levels()`
- Coercing object class with `as.factor()`, `as.numeric()`, and `as.character()`
- Summarizing variable categories with `table()`

1 Bias in Self-reported Turnout

Surveys are frequently used to measure political behavior including voter turnout, but some researchers are concerned about the accuracy of self-reports. In particular, they worry about possible *social desirability bias* where in post-election surveys, respondents who did not vote in an election lie because they may feel that they should have voted. Is such a bias present in the American National Election Studies (ANES)? ANES is a nation-wide survey, which have been conducted for every election since 1948. It conducts face-to-face interviewing of a nationally representative sample of adults.

The `ANES.txt` data file, available at Blackboard, contains the following variables

Variable	Description
`year`	election year
`VAP`	Voting Age Population from Census
`total`	total turnout (in 1000s)
`ANES`	turnout rate estimated from ANES
`overseas`	number of votes caste by overseas voters
`felons`	population of felony prisoners (in 1000s)
`noncit`	non-citizen population (in 1000s)

1. Load the data into **R**
2. Check the dimension of the data and also obtain a summary of the data
3. Calculate the turnout rate based on VAP. Note that for this data set, we must subtract the number of overseas voters from the total voters, since VAP does not include a count of eligible overseas voters.
4. Compute the difference between VAP and ANES estimates of turnout rate. How big is the difference on average? What is the range of the difference?

5. ANES does not interview overseas voters and prisoners. Compute the adjusted VAP turnout after subtracting the number of felony prisoners and non-citizens from the VAP and the number of ballots cast by overseas voters from the total turnout. How does this adjustment change the results obtained in the previous question?

6. Compare the adjusted VAP turnout rate with the ANES turnout rate separately for presidential elections and midterm elections. Does the bias of the ANES vary across election types?

7. *(Optional)* Divide the data into half by election years such that you subset the data into two periods. Calculate the difference between the adjusted VAP turnout rate and the ANES turnout rate separately for each year within each period. Has the bias of the ANES increased over time?

2 Object Class

In **R**, every object belongs to a certain class. Classes we have seen already include `numeric`, `character`, `function`, and `data.frame`. Knowing which class each object belongs to is sometimes important because the same function performs different operations, depending on which class the input object belongs to. For example, the `summary()` function gives a different output depending on the class of its input object.

RStudio: Identifying the Class of an Object

In **RStudio**, when you create a new object, the class of the object appears in the `Workspace` window on the top-right of **RStudio**. For example, if you import a dataset into the **R** console, the `Workspace` will show the name of that object in the left column and the class of the object (dataframe) in the right column.

- The function `class()` returns the class of an object.

```
> load("turnout.RData")
> class(turnout)

[1] "data.frame"

> class(turnout$VEP)

[1] "integer"
```

- An important class we cover here is the `factor` class, which should be used for qualitative (categorical) variables (rather than the `character` class). The function `levels()` provides a vector of unique categories for a factor variable.

```
> class(turnout$State)

[1] "factor"

> levels(turnout$State)
```

```
 [1] "Alabama"              "Alaska"               "Arizona"              "Arkansas"
 [6] "Colorado"             "Connecticut"          "Delaware"             "District of C
[11] "Georgia"              "Hawaii"               "Idaho"                "Illinois"
[16] "Iowa"                 "Kansas"               "Kentucky"             "Louisiana"
[21] "Maryland"             "Massachusetts"        "Michigan"             "Minnesota"
[26] "Missouri"             "Montana"              "Nebraska"             "Nevada"
[31] "New Jersey"           "New Mexico"           "New York"             "North Caroli
[36] "Ohio"                 "Oklahoma"             "Oregon"               "Pennsylvania"
[41] "South Carolina"       "South Dakota"         "Tennessee"            "Texas"
[46] "Utah"                 "Vermont"              "Virginia"             "Washington"
[51] "Wisconsin"            "Wyoming"
```

```
> summary(turnout$State)
```

```
           Alabama             Alaska            Arizona           Arkansas
                 8                  8                  8                  8
          Colorado        Connecticut           Delaware District of Columbia
                 8                  8                  8                  8
           Georgia             Hawaii              Idaho           Illinois
                 8                  8                  8                  8
              Iowa             Kansas           Kentucky          Louisiana
                 8                  8                  8                  8
          Maryland      Massachusetts           Michigan          Minnesota
                 8                  8                  8                  8
          Missouri            Montana           Nebraska             Nevada
                 8                  8                  8                  8
        New Jersey         New Mexico           New York     North Carolina
                 8                  8                  8                  8
              Ohio           Oklahoma             Oregon       Pennsylvania
                 8                  8                  8                  8
    South Carolina       South Dakota          Tennessee              Texas
                 8                  8                  8                  8
              Utah            Vermont           Virginia         Washington
                 8                  8                  8                  8
         Wisconsin            Wyoming
                 8                  8
```

- In some situations, we want to coerce a certain object into a particular class. The function `as.factor()` coerces an object into a factor, and similar operations can be done using `as.numeric()` and `as.character()` to turn an object into the numeric and character classes, respectively.

```
> state <- as.character(turnout$State)
> class(state)

[1] "character"

> summary(state) # not very useful
```

```
   Length     Class      Mode
      416 character character
```

- The function `table()` summarizes the levels of numeric, integer, factor, and character variables. Note that the `table()` function is most useful when numeric variables are coded as a relatively small number of integers.

```
> ## numeric variables
> class(turnout$VAP)

[1] "integer"

> table(turnout$VAP)[1:10] # not very useful

277261 320695 330784 331170 343957 353354 354410 366055 369147 369260
     1      1      1      1      1      1      1      1      1      1

> ## a more appropriate variable
> class(turnout$year)

[1] "integer"

> table(turnout$year)

1980 1984 1988 1992 1996 2000 2004 2008
  52   52   52   52   52   52   52   52

> ## same when coerced into a numeric variable
> turnout$year <- as.numeric(turnout$year)
> class(turnout$year)

[1] "numeric"

> table(turnout$year)

1980 1984 1988 1992 1996 2000 2004 2008
  52   52   52   52   52   52   52   52

> ## factor variable
> class(turnout$State)

[1] "factor"

> table(turnout$State)
```

```
     Alabama      Alaska     Arizona    Arkansas
           8           8           8           8
    Colorado Connecticut    Delaware  District of Columbia
           8           8           8           8
     Georgia      Hawaii       Idaho    Illinois
           8           8           8           8
        Iowa      Kansas    Kentucky   Louisiana
           8           8           8           8
    Maryland Massachusetts    Michigan   Minnesota
           8           8           8           8
    Missouri      Montana    Nebraska      Nevada
           8           8           8           8
  New Jersey   New Mexico    New York North Carolina
           8           8           8           8
        Ohio     Oklahoma      Oregon Pennsylvania
           8           8           8           8
South Carolina South Dakota   Tennessee       Texas
           8           8           8           8
        Utah      Vermont    Virginia  Washington
           8           8           8           8
   Wisconsin     Wyoming
           8           8
```

```
> ## character variable
> class(state)

[1] "character"

> table(state)

state
     Alabama      Alaska     Arizona    Arkansas
           8           8           8           8
    Colorado Connecticut    Delaware  District of Columbia
           8           8           8           8
     Georgia      Hawaii       Idaho    Illinois
           8           8           8           8
        Iowa      Kansas    Kentucky   Louisiana
           8           8           8           8
    Maryland Massachusetts    Michigan   Minnesota
           8           8           8           8
    Missouri      Montana    Nebraska      Nevada
           8           8           8           8
  New Jersey   New Mexico    New York North Carolina
           8           8           8           8
        Ohio     Oklahoma      Oregon Pennsylvania
           8           8           8           8
```

South Carolina	South Dakota	Tennessee	Texas
8	8	8	8
Utah	Vermont	Virginia	Washington
8	8	8	8
Wisconsin	Wyoming		
8	8		