# hw02: R Basics

*Your name**

*06 06, 2017*

Questions in this problem set are taken from *Kleiber, C., & Zeileis, A. (2008). Applied econometrics with R. Springer Science & Business Media, p. 54.*

## Question 1

Create a square matrix, say $A$, with entries $a_{ii} = 2, ..., n-1$, $a_{11} = a_{nn} = 1$, $a_{i,i+1} = a_{i,i-1} = -1$, and $a_{ij} = 0$ elsewhere. (Where does this matrix occur in econometrics?)

```
# Write here your codes
```

## Question 2

"PARADE"" is the Sunday newspaper magazine supplementing the Sunday or weekend edition of some 500 daily newspapers in the United States of America. An important yearly feature is an article providing information on some 120–150 "randomly"" selected US citizens, indicating their profession, hometown and state, and their yearly earnings. The `Parade2005` data contain the 2005 version, amended by a variable indicating celebrity status (motivated by substantial oversampling of celebrities in these data). For the `Parade2005` data:

### a)

Determine the mean earnings in California. Explain the result.

```
# Write here your codes
```

### b)

Determine the number of individuals residing in Idaho. (What does this say about the data set?)

```
# Write here your codes
```

### c)

Determine the mean and the median earnings of celebrities. Comment.

```
# Write here your codes
```

*affiliation

### d)

Obtain boxplots of `log(earnings)` stratified by `celebrity`. Comment.

```
# Write here your codes
```

## Question 3

For the `Parade2005` data of the preceding exercise, obtain a kernel density estimate of the earnings for the full data set. It will be necessary to transform the data to logarithms (why?). Comment on the result. Be sure to try out some arguments to `density()`, in particular the plug-in bandwidth `bw`.

```
# Write here your codes
```

## Question 4

Consider the `CPS1988 data`, taken from Bierens and Ginther (2001). (These data will be used for estimating an earnings equation in the next chapter.)

### a)

Obtain scatterplots of the logarithm of the real wage (`wage`) versus `experience` and versus `education`.

```
# Write here your codes
```

### b)

In fact, `education` corresponds to years of schooling and therefore takes on only a limited number of values. Transform `education` into a factor and obtain parallel boxplots of `wage` stratified by the levels of `education`. Repeat for `experience`.

```
# Write here your codes
```

### c)

The data set contains four additional factors, `ethnicity, smsa, region, and parttime`. Obtain suitable graphical displays of log(`wage`) versus each of these factors.

```
# Write here your codes
```