

Econ 294 Assignment 1

Curtis Kephart

Winter 2015

Use R to answer the following answers.

Put solutions in an `.R` script file. Name the file by concatenating your first and last name, plus “Assignment1”. (e.g. my file would be called `CurtisKephartAssignment1.R`)

For all questions below that require some specific answer, use a `print` (or `paste` and `print`) function call in your `.R` script to report your answers.

0. Use a `print` call to report your first and last name, and student ID number.
1. Load Data. Hints, you will need different function calls for each file type, and the `.dta` file requires the `foreign` package. Depending on your OS, you may need to make use of the `url()` function.
 - From `.dta` (STATA File). Assign the name `df.dta` to the file.

https://github.com/EconomiCurtis/econ294_2015/raw/master/data/NHIS_2007_dta.dta

- From CSV. Assign the name `df.csv` to the file.

https://github.com/EconomiCurtis/econ294_2015/raw/master/data/NHIS_2007_CSV.csv

- From Tab delimited. Assign the name `df.td` to the file.

https://github.com/EconomiCurtis/econ294_2015/raw/master/data/NHIS_2007_TSV.txt

- From `.RData`. Note that `.RData` files come with names already assigned to their data structures. What is the name assigned to this `RData` file?

https://github.com/EconomiCurtis/econ294_2015/raw/master/data/NHIS_2007_RData.RData

2. Download each file above to your hard drive, how big (in KB) is each file? Which is the smallest? Besides the `.dta` file, what accounts for their variability?
3. For the object `df.rdata`, what type of data structure is this?

Apply and report `length`, `dim`, `nrow`, `ncol`, and `summary` functions.

4. Load `org_example.dta` Stata file from the URL below, and assign the name `df` to the loaded object.

https://github.com/EconomiCurtis/econ294_2015/raw/master/data/org_example.dta

Apply and report `str`. How many observations and how many variables are there?

For the variable (column) `rw` what is the `min`, `mean`, `median`, `max`, first and third quartile value? How many NAs are there?

5. Create the following vector, `v <- c(1,2,3,4,5,6,7,4,NULL,NA)`

Report `length`. Why don't the number of values in the vector match the number reported in `length`?

Report `mean` ignoring the `NA` value.

6. Matrix operations.

Create the following matrix and call it `x`.

```
1 2 3
4 5 6
7 8 9
```

Show how to find its transpose.

Find the eigenvalues and eigenvectors of `x`.

Now create the following matrix and call it `y`.

```
1 2 3
3 2 1
2 3 0
```

Find the inverse of `y`.

Multiple `y` by its inverse. From linear algebra, what is this new matrix called?

7. Create a data frame based on the follow schema and values, called `diamonds`.

carat	cut	clarity	price
5	"fair"	"SI1"	850
2	"good"	"I1"	450
0.5	"very good"	"V11"	450
1.5	"good"	"VS1"	NULL
5	"fair"	"IF"	750
NA	"Ideal"	"VVS2"	980
3	"fair"	NA	420

- What is the mean price?
- What is the mean price of cut "fair"?
- What is the mean price of cut "good", "very good", and "Ideal"?
- For diamonds with greater than 2 carats, and cut "Ideal" or "very good", what is the median price?