• At a Personal Computer – you must first download the package to your computer:

1. Click on "Packages" and scroll down to "Install Package(s) from CRAN..."
2. A list of possible packages to install should appear in a new window. Scroll down and click on "MASS"
3. R should proceed to install the package. If asked "Delete downloaded files (y/N)?" type "y".
4. Bring in the library by typing:
   > library(MASS)


2. Matrices in R

   (a) Generate a $10 \times 20$ matrix $A$ with $(A)_{i,j} = i + 2j$. (We use the notation $(A)_{i,j}$ to denote the element of $A$ in row $i$ and column $j$.)

   (b) Generate a $5 \times 5$ matrix $B$ with $(B)_{i,j} = (A)_{2i,10+j}$.

   (c) Generate matrix $C$ with $(C)_{i,j} = \cos((A)_{i,j})$ using the sapply function. (This problem is to make you aware of the sapply function. In practice, one would probably just run the command $C \leftarrow \cos(A)$.)

   (d) Run $C.qr \leftarrow qr(C)$, names(C.qr) gives a list of the attributes in $C.qr$. Try the following ways of accessing the rank attribute:

   i. C.qr['rank']
   ii. C.qr[['rank']]
   iii. C.qr$rank

   Explain any differences between the results of the three commands.

3. QR-Decompositions For the following matrices:

   $A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad b_1 = \begin{pmatrix} 5 \\ 3 \\ 6 \end{pmatrix}$

   $A_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \quad b_2 = \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix}$
(a) Do a $QR$-decomposition of $A$ by hand. Solve $Ax = b$ using the $QR$-decomposition of $A$ by hand.

(b) Solve $Ax = b$ using a $QR$-decomposition in R in two different ways. For the first way, use the `solve.QR` function. For the second way, do not use `solve.QR`; instead use `qr.R` and `qr.Q`. For the second way, you will need a function to perform back substitution. You can either write your own function, or take the code from the handout given out in the second week of class.

(c) Solve $Ax = b$ using the R `solve` function.

4. Linear Regression

(a) Write an R function to fit a simple linear regression with or without intercept. The function should start

```r
linear.regression <- function(x, y, intercept=TRUE)
```

Have the function return the coefficients, the fitted values, the residuals, and the residual sum-of-squares. Use the $QR$-decomposition to perform the regression.

(b) Load the MASS package by typing `library(MASS)`. The package contains a data set called `phones` that gives the number of phone calls (in millions) in Belgium between 1950 and 1973. Use your linear regression function with intercept to fit a model predicting the number of calls as a linear function of the year. Plot the raw data, the fitted values, and the regression line. You will probably want to use the `plot` and the `abline` functions.

(c) The `lqs` function is part of the MASS package. The function provides a linear regression that minimizes least median of squares:

$$\min_b \text{median} |y_i - x_i b|^2$$

You can find documentation on `lqs` by typing `help(lqs)`. Use `lqs` to find the least median of squares fit of calls as a linear function of year. Plot the raw data, the least squares fit line, and the least median of squares line. Explain your results.

(d) Would ridge regression would be an appropriate method on this data? (Why or why not).

(e) What is the difference between ridge regression and PC regression?