Lecture 9: S3 Objects - A Study in Linear Regression

STAT 385 - James Balamuta

June 30, 2016

On the Agenda

1. Administrative
   - HW3
   - Project Proposal
2. S3 Objects
   - What is Object-oriented programming (OOP)?
   - How are S3 Methods implemented in R?
3. The `lm` function
   - Constructing `lm()`
   - Constructing Inference via `summary.lm()`

Administrative Items

- HW3 Posted Tonight (sorry for the delay!)
  - Due on: **Wednesday, July 6th at 2:00 PM CST**.
- Project Proposals due tomorrow
  - Due on: **Friday, July 1st at 11:59 PM CST** (no extension)
- Practice Midterm posted tomorrow!
  - Available to use for practice on: **Friday, July 1st at 5:00 PM CST**
  - Make sure you do it *before* the midterm!
  - Answers to posted on **Tuesday, July 5th at 1:00 PM CST**

Programming Up Until Now...

- In the past lectures, the focus has been on the creation of modular code via **functions**.
- The goals behind creating **functions** were to:
  1. create reusable chunks of code
  2. decrease the probability of error in code
  3. make shareable code
- Now, we're going to take it one step further with **Object-oriented programming (OOP)**.

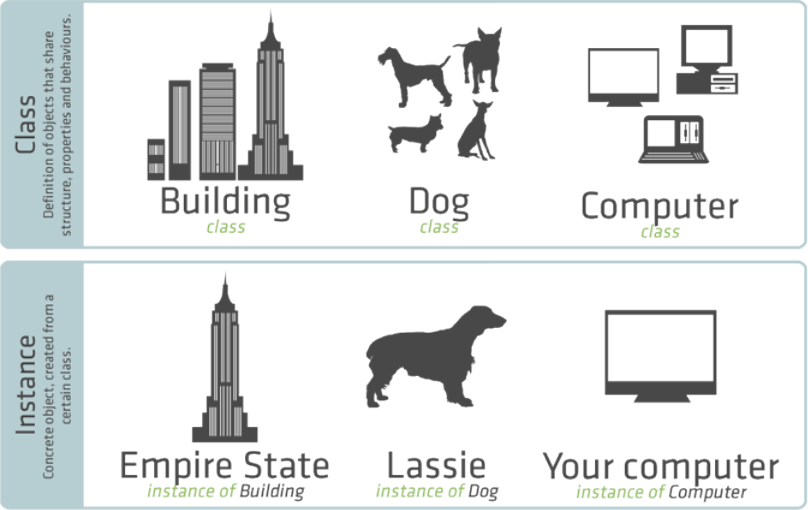
Object-oriented programming (OOP)

Definition: **Object-oriented programming (OOP)** is a programming paradigm where real world ideas can be described as a collection of items that are able to interact together.

Definitions of OOP Concepts

- Definition: **Classes** are definitions of what an **object** *is*.
  - Example: **Student** has properties of **Name**, **NetID**, **Grades**, **Address**, ...
- Definition: **Objects** are instances of a **class**. (*noun*)
  - Example: **Kevin** and **Justin** are instances of a **Student**
- Definition: **Methods** are functions that performs specific calculations on objects of a specific class. (*verb*)
  - Example: **in_class()** and **get_grade()**

Think of it as. . .



Core Tenets of OOP

- **Encapsulation:** Enables the combination of data and functions into classes
- **Polymorphism:** Functions are able to act differently across classes
- **Inheritence:** Extend a parent class by creating a child classes without copying!

OOP Concept Check

- How would an **instructor** class be defined?
- How might we abstract both so that they share a common class?

Why use OOP?

1. **Increased Modularity:**
   - Code for classes can be implemented and maintained separately from other classes.
2. **Hide Subroutines:**
   - Avoid having multiple method functions known.
3. **Code Reuse and Recycling Across Packages:**
   - Easily extend classes defined in other R packages or within the base *R* system. (e.g. Allen wrenches)
4. **Features and Debugging**:
   - Problematic class? Remove or easily revert class to an earlier version without worrying about headache across the entire system.

OOP in R

> *"To understand computations in R, two slogans are helpful:*
>
> - ***Everything that exists is an object.***
> - *Everything that happens is a function call."*
>
> *—John Chambers*

Question: *How* is everything in *R* an object?

OOP in R - Answer

- In order for an **object** to exist, it must have a **class**.
- Every object in *R* returns a class when `class(x)` is called.
  - Even *functions* and *environments* have classes!

```r
class(3)              # Number
```

```
## [1] "numeric"
```

```r
class(sum)            # Function
```

```
## [1] "function"
```

```r
class(.GlobalEnv)   # Global Environment
```

```
## [1] "environment"
```

R's OOP Systems

- There are three OOP systems in *R* that differ in terms of how classes and methods are defined:
  - **S3:** Very casual/informal OO system that is used throughout *R*
  - **S4:** More formal and rigorous with class definitions.
  - **Reference classes (RC):** Very new and shiny OOP system that mimicks traditional Java and C++ message-passing OO.
- Today, we'll focus on just working within the **S3** System. Later, we'll explore **S4** and **RC**.

Detecting Object Type

- Before we begin, it's important to be able to detect the type of OOP systems begin used within the method.
- To reliable detect the OOP system, please use `ftype()` in `pryr` R Package.

```
pryr::ftype(print)
```

```
## [1] "s3"      "generic"
```

R's S3 System

- Definition: A **generic function** is used to determine the class of its arguments and select the appropriate method.
  - Examples: `summary()`, `print()`, and `plot()`
  - The `lm` class has: `summary.lm()`, `print.lm()`, and `plot.lm()`
- Generic functions have a method naming convention of: `generic.class()`
- If a class has not been defined for use in a generics, it **will** fail.
  - To avoid the failure define `generic.default()` (e.g. `summary.default`)

Generic Function

- S3 generic detectable by looking at a function's source for `UseMethod()`

```r
summary
```

```
## function (object, ...)
## UseMethod("summary")
## <bytecode: 0x7fa009077040>
## <environment: namespace:base>
```

Viewing S3 Methods Associated with Generic

```r
# All classes with a summary.*() function
methods(summary)
```

```
## [1] "summary.aov"     "summary.aovlist"
```

```r
# Methods using a particular class
methods(class='matrix')
```

```
## [1] "anyDuplicated.matrix"     "as.data.frame.matrix"
## [3] "as.raster.matrix"         "boxplot.matrix"
## [5] "coerce,ANY,matrix-method" "determinant.matrix"
```

Note: Output has been suppressed, there are considerably more usages. Try running the commands yourself!

Constructing an S3 Object

Part of S3's ability to be informal is the ease of construction.
There are two different flavors of construction:

- All in one
- The two-step

These constructions are **informal** as there is no definition of a *class*.

Constructing an S3 Object - One Step

```r
# ----- One Step S3 Construct

# Create andy student object and assign class student
andy = structure(list(), class = "student")

class(andy)                      # Check class
```

```
## [1] "student"
```

```r
str(andy)                        # Structure
```

```
##  list()
##  - attr(*, "class")= chr "student"
```

Constructing an S3 Object - Two Step by Class

```r
# ----- Two Step S3 Construct

andy = list()                 # Create object andy
                              # as a list class

class(andy) = "student"      # then set class to student

class(andy)                   # Check obj type
```

```
## [1] "student"
```

```r
str(andy)                     # Structure
```

```
##  list()
##  - attr(*, "class")= chr "student"
```

Constructing an S3 Object - Two Step by Attribute

```r
# ----- Two Step S3 Construct with Attributes

andy = list()                           # Create object andy
                                        # as a list class

attr(andy,"class") = "student"          # Set class to student

class(andy)                             # Check obj type
```

```
## [1] "student"
```

```r
str(andy)                               # Structure
```

```
##  list()
##  - attr(*, "class")= chr "student"
```

Checking for Object Status

To determine whether an object is of a specific class use `inherits(x, "class")`

```
inherits(andy, "student")
```

```
## [1] TRUE
```

```
inherits(andy, "list")
```

```
## [1] FALSE
```

Note: The list inheritance check failed as we removed that `class` definition.

Creating a New Generic

```r
# Create a role identifier

# Instructor
role.instructor = function(x){ # Instructor
  cat("Greetings and Salutations", x$fname, "\n",
      "You are an instructor for", x$course, "\n")
}

role.student = function(x){    # Student
    cat("Hey ", x$fname, "!\n",
      "Are you in ", x$course, "?\n")
}
```

Notes:

- `student` and `instructor` are the **classes**
- `role` is the method.

`UseMethod()` Properties

To create a generic function, we only need to do:

```
# Create a default case
role = function(x, ...) UseMethod("role")
```

A few notes:

- The generic function will call the first class it finds with an implementation from left to right
- If no class is found it defaults to the `person.default()` function if it is defined!
- The `...` are ellipses and they enable additional parameters to be passed through.

Example Call of Generic - One Class

- An initial class `instructor` in S3 would look like so:

```
james = structure(list(fname = 'James',
                       course = "STAT385"),
                  class = "instructor")

role(james)
```

```
## Greetings and Salutations James
##  You are an instructor for STAT385
```

Example Call of Generic - Two Classes

- Here the `david` object has two class types.
- Only the first class (from left to right) will be called.

```
david = structure(list(fname = 'David',
                       course = "STAT385"),
                  class = c('student','instructor'))

role(david)
```

```
## Hey  David !
##  Are you in  STAT385 ?
```

Example Call of Generic - Unknown Class

- If we do **not**:
  1. define a `generic.*()` for a class
  2. define a `generic.default()`
- it will **error**

```
toad = structure(list(fname = 'McToady',
                          course = "STAT385"),
                 class = 'humbug')

role(toad)

## Error in UseMethod("role") :
##  no applicable method for 'role' applied
## to an object of class "list"
```

Protecting Generics with `generic.default()`

- Always protect your generic with a `generic.default()`!

```
role.default = function(x){      # Default case
  cat("I have no clue what your role is. Who are you?")
}

# Try again
role(toad)
```

```
## I have no clue what your role is. Who are you?
```

Use Inheritance!

- When assigning classes to objects, use the *inheritance* tenet!
- Write the most-specific class first and then a less specific class.

```
james = structure(list(fname = 'James',
                            course = "STAT385"),
                      class = c("instructor","list"))

print(james)
```

```
## $fname
## [1] "James"
##
## $course
## [1] "STAT385"
##
## attr(,"class")
## [1] "instructor" "list"
```

Practical Note

- Avoid calling the methods function directly.
  - Use a `generic` function to dispatch the methods to objects
  - e.g. use `summary()` instead of `summary.yourobj()`.

```
# Bad
role.instructor(james) # Not always an instructor!

# Good
role(james) # Adapts to future change!
```

Summary on S3

- Very informal and easy to work with.
- Be on your guard as it relates to class definitions.
- Define a `generic.default()` method for extra protection.

Moving Along. . . .

- Coming up next. . . **Implementing an S3 `lm` function!**
- Any questions on **Object-oriented programming**?

Understanding the Algorithm

Before we can implement an algorithm, we must understand the following:

- *What* logic is being used?
- *How* does the logic apply in a procedural form?
- *Why* is this logic present?

Thus, let's take a bit of a closer look at Multiple Linear Regression (MLR) *before* we start to implement it.

Multiple Linear Regression (MLR) Definition

Formula

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots \beta_{p-1} x_{i,p-1} + \varepsilon_i$$

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$$

Responses: $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1}$ **Errors:** $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$

Design Matrix:

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p-1} \end{pmatrix}_{n\times p}$$

Parameters:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}_{p\times 1}$$

Least Squares with Multiple Linear Regression (MLR)

Goal: Obtain the minimization of RSS.

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2$$

Errors:

$$\begin{aligned} e &= y - \hat{y} \\ &= y - X\hat{\beta} \end{aligned}$$

RSS Definition:

$$RSS = e^T e = \begin{bmatrix} e_1 & e_2 & \cdots & e_N \end{bmatrix}_{1\times 1} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}_{n\times 1}$$

$$= [e_1 \times e_1 + e_2 \times e_2 + \cdots + e_n \times e_n]_{1\times 1} = \sum_{i=1}^{n} e_i^2$$

Note: $e \neq \varepsilon$ since e is the realization of ε from the regression procedure.

Least Squares with Multiple Linear Regression (MLR)

Goal: Obtain the minimization of RSS.

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2$$

Expand RSS:

$$\begin{aligned}
RSS &= (y - X\beta)^T (y - X\beta) \\
&= \left(y^T - \beta^T X^T\right)(y - X\beta) \\
&= y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta \\
&= y^T y - \left(\beta^T X^T y\right)^T - y^T X\beta + \beta^T X^T X\beta \\
&= y^T y - y^T X\beta - y^T X\beta + \beta^T X^T X\beta \\
&= y^T y - 2\beta^T X^T y + \beta^T X^T X\beta
\end{aligned}$$

Note:

$$\beta^T_{1\times p} X^T_{p\times n} y_{n\times 1} = \left(\beta^T_{1\times p} X^T_{p\times n} y_{n\times 1}\right)^T = y^T_{1\times n} X_{n\times p} \beta_{p\times 1}$$

We are able to perform a transpose in place as the result is scalar.

Least Squares with Multiple Linear Regression (MLR)

Goal: Obtain the minimization of RSS.

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2$$

Take the derivative with respect to β:

$$\begin{aligned} RSS &= y^T y - 2\beta^T X^T y + \beta^T X^T X\beta \\ \frac{\partial RSS}{\partial \beta} &= -2X^T y + 2X^T X\beta \end{aligned}$$

Set equal to zero and solve:

$$\begin{aligned} 0 &= -2X^T y + 2X^T X\beta \\ 2X^T X\beta &= 2X^T y \\ X^T X\beta &= X^T y \\ \hat{\beta} &= \left(X^T X\right)^{-1} X^T y \end{aligned}$$

Mean of LS Estimator for MLR

Next up, let's take the mean of the estimator!

$$
\begin{aligned}
E\left(\hat{\beta}\right) &= E\left[\left(X^T X\right)^{-1} X^T y\right] \\
&= E\left[\left(X^T X\right)^{-1} X^T \left(X\beta + \varepsilon\right)\right] \\
&= E\left[\left(X^T X\right)^{-1} X^T X\beta + \left(X^T X\right)^{-1} X^T \varepsilon\right] \\
&= E\left[\beta + \left(X^T X\right)^{-1} X^T \varepsilon\right] \\
&= \beta + E\left[\left(X^T X\right)^{-1} X^T \varepsilon\right]
\end{aligned}
$$

Notes:

- We substituted in the definition of $y = X\beta + \varepsilon$ and then simplified the matrix
- β is a constant within the expectation and, thus, we pulled it out.

Mean of LS Estimator for MLR

$$
\begin{aligned}
E\left(\hat{\beta}\right) &= \beta + E\left[\left(X^T X\right)^{-1} X^T \varepsilon\right] \\
&= \beta + E\left[E\left[\left(X^T X\right)^{-1} X^T \varepsilon | X\right]\right] \\
&= \beta + E\left[\left(X^T X\right)^{-1} X^T \underbrace{E\left[\varepsilon | X\right]}_{=0 \text{ by model}}\right] \\
&= \beta
\end{aligned}
$$

Notes:

- Used the law of total expectation

$$E\left[X\right] = E\left[E\left[X | Y\right]\right]$$

- Showed that the estimator was *unbiased* under the assumption that the mean of the residuals is 0.

Covariance of the LS Estimator for MLR

Next, it would be helpful to know the deviations of the estimator for inference.

$$
\begin{aligned}
Cov\left(\hat{\beta}\right) &= E\left[\left(\left(X^TX\right)^{-1}X^Ty-\beta\right)\left(\left(X^TX\right)^{-1}X^Ty-\beta\right)^T\right] \\
&= E\left[\begin{array}{l}\left(\left(X^TX\right)^{-1}X^T\left(X\beta+\varepsilon\right)-\beta\right) \\ \left(\left(X^TX\right)^{-1}X^T\left(X\beta+\varepsilon\right)-\beta\right)^T\end{array}\right] \\
&= E\left[\begin{array}{l}\left(\left(X^TX\right)^{-1}X^TX\beta+\left(X^TX\right)^{-1}X^T\varepsilon-\beta\right) \\ \left(\left(X^TX\right)^{-1}X^TX\beta+\left(X^TX\right)^{-1}X^T\varepsilon-\beta\right)^T\end{array}\right] \\
&= E\left[\left(\beta+\left(X^TX\right)^{-1}X^T\varepsilon-\beta\right)\left(\beta+\left(X^TX\right)^{-1}X^T\varepsilon-\beta\right)^T\right]
\end{aligned}
$$

Covariance of the LS Estimator for MLR

$$
\begin{aligned}
Cov\left(\hat{\beta}\right) &= E\left[\left(\beta + \left(X^TX\right)^{-1}X^T\varepsilon - \beta\right)\left(\beta + \left(X^TX\right)^{-1}X^T\varepsilon - \beta\right)^T\right] \\
&= E\left[\left(\left(X^TX\right)^{-1}X^T\varepsilon\right)\left(\left(X^TX\right)^{-1}X^T\varepsilon\right)^T\right] \\
&= E\left[\left(\left(X^TX\right)^{-1}X^T\varepsilon\right)\left(\left(X^TX\right)^{-1}X^T\varepsilon\right)^T\right] \\
&= E\left[\left(X^TX\right)^{-1}X^T\varepsilon\varepsilon^T\left(X^TX\right)^{-1}X^T\right] \\
&= \left(X^TX\right)^{-1}X^TE\left[\varepsilon\varepsilon^T\right]X\left(X^TX\right)^{-1} \\
&= \left(X^TX\right)^{-1}X^T\,\mathrm{var}\left(\varepsilon\right)X\left(X^TX\right)^{-1}
\end{aligned}
$$

Note: The above calculations are useful in multiple regression paradigms with minimal modification.

Covariance of the LS Estimator for MLR

$$
\begin{aligned}
Cov\left(\hat{\beta}\right) &= \left(X^T X\right)^{-1} X^T \operatorname{var}(\varepsilon) X \left(X^T X\right)^{-1} \\
&= \left(X^T X\right)^{-1} X^T \left(\sigma^2 I_N\right) X \left(X^T X\right)^{-1} \\
&= \sigma^2 \left(X^T X\right)^{-1} X^T (I) X \left(X^T X\right)^{-1} \\
&= \sigma^2 \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} \\
&= \sigma^2 \left(X^T X\right)^{-1}
\end{aligned}
$$

Note: Under Least Squares, we assume that

$$\operatorname{var}(\varepsilon) = \sigma^2 I_n$$

Multiple Linear Regression (MLR)

Solutions:

$$\hat{\beta}_{p\times 1} = \left(X^T X\right)^{-1}_{p\times p} X^T_{p\times n} y_{n\times 1}$$

$$E\left(\hat{\beta}\right) = \beta_{p\times 1}$$

$$Cov\left(\hat{\beta}\right) = \sigma^2 \left(X^T X\right)^{-1}_{p\times p}$$

Writing a my_lm() function - Part 1

To begin, we start with the basic definition for a generic method.

```
my_lm = function(x, ...) UseMethod("my_lm")
```

Writing a my_lm() function - Part 2

Now, let's implement the my_lm default method.

```
my_lm.default = function(x, y, ...){

  # Obtain the QR Decomposition of X
  # Not a good approach for rank-deficient matrices
  qr_x = qr(x)

  # Compute the Beta_hat = (X^T X)^(-1) X^T y estimator
  beta_hat =  solve.qr(qr_x, y)

  # Compute the Degrees of Freedom
  df = nrow(x) - ncol(x)    # n - p
```

Writing a my_lm() function - Part 3

```
  # Compute the Standard Deviation of the Residuals
  sigma2 <- sum((y - x %*% beta_hat) ^ 2) / df

  # Compute the Covariance Matrix
  # Cov(Beta_hat) = sigma^2 * (X^T X)^(-1)
  cov_mat = sigma2 * chol2inv(qr_x$qr)

  # Make name symmetric in covariance matrix
  rownames(cov_mat) = colnames(x)
  colnames(cov_mat) = colnames(x)

  # Return a list
  return(structure(list(coefs = beta_hat,
                        cov_mat = cov_mat,
                        sigma = sqrt(sigma2),
                        df = df),
                   class = "my_lm"))
}
```

Writing a print.my_lm() function - Part 4

```
print.my_lm = function(x, ...){
  cat("\nCoefficients:\n")
  print(x$coefs)
}
```

Writing a summary.my_lm() function - Part 5

```
# Note that summary(object, ...) instead of summary(x, ...)
summary.my_lm = function(object, ...){

  estimate = coef(object) # Store coeffecents
  sterr = sqrt(diag(object$cov_mat)) # STD Error
  t_test = estimate / sterr # T Test value
  pval = 2*pt(-abs(t_test), df=object$df)

  out = structure(list(mat =
                         cbind(estimate,
                               sterr,
                               t_test,
                               pval)),
                  class = "summary.my_lm")
  return(out)
}
```

Comparing Print Output

```r
# Ours
my_lm(x = cbind(1, mtcars$disp), y = mtcars$mpg)
```

```
##
## Coefficients:
## [1] 29.59985476 -0.04121512
```

```r
# Base R
lm(mpg~disp, data = mtcars)
```

```
##
## Call:
## lm(formula = mpg ~ disp, data = mtcars)
##
## Coefficients:
## (Intercept)         disp
##    29.59985     -0.04122
```

Writing a print.summary.my_lm() function - Part 5

- We can control how the summary generic function should look on print via print.summary.my_lm.
- Here we make use of the printCoefmat() functionality.

```
# Note that print(x,...)!!
print.summary.my_lm = function(x, ...) {

  m = x$mat
  colnames(m) = c("Estimate", "Std.Err", "T value", "Pr(>t
  printCoefmat(x$mat,
               P.value = TRUE,
               has.Pvalue = TRUE)
}
```