Mean-only linear model

Structure of the model

The simplest possible linear model consists of just a constant term, and the fitted value will be the sample mean:

$$Y = \beta + e$$

where:

- Y is a vector of observed values
- β is a parameter, treated as unknown but constant
- e is a vector of independent, identically distributed $N(0, \sigma_e)$ random variables
- σ_e is the standard deviation of the error or residual terms e, which are assumed to have mean zero.

Example: Estimating the Sepal.Width mean from the Iris data

Read the data:

```
data(iris)
str(iris)

## 'data.frame': 150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Select only versicolor so we can easily interpret the results:

```
vsc = iris[iris$Species=='versicolor',]

str(vsc)

## 'data.frame': 50 obs. of  5 variables:
##  $ Sepal.Length: num  7 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 ...
##  $ Sepal.Width : num  3.2 3.2 3.1 2.3 2.8 2.8 3.3 2.4 2.9 2.7 ...
##  $ Petal.Length: num  4.7 4.5 4.9 4 4.6 4.5 4.7 3.3 4.6 3.9 ...
##  $ Petal.Width : num  1.4 1.5 1.5 1.3 1.5 1.3 1.6 1 1.3 1.4 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 2 2 2 2 2 2 2 2 2 2 ...
```

```r
mean(vsc$Sepal.Width)          #get the mean Sepal.Width for versicolor
```

```
## [1] 2.77
```

```r
sd(vsc$Sepal.Width)            #get the standard deviation of the sepal width
```

```
## [1] 0.3137983
```

Now run the mean-only model.

We have to resort to some trickery as R will not allow a factor with one level or an intercept-only regression.

The trick is to create a variable called x that is all ones, and tell R the model has no intercept.

This will produce the desired mean-only model fit.

```r
Sepal.Width = vsc$Sepal.Width

x           = rep(1,length(Sepal.Width))

lm1 = lm(Sepal.Width~x-1)
summary(lm1)
```

```
##
## Call:
## lm(formula = Sepal.Width ~ x - 1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.770 -0.245  0.030  0.230  0.630
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## x  2.77000    0.04438   62.42   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3138 on 49 degrees of freedom
## Multiple R-squared:  0.9876,Adjusted R-squared:  0.9873
## F-statistic:  3896 on 1 and 49 DF,  p-value: < 2.2e-16
```

Notice that the `Coefficients: Estimate` of x in the model matches the mean of `Sepal.Width`, and the `Residual standard error:` matches the standard deviation of `Sepal.Width`

While you will probably never have occasion to use this model, it is important because more complicated models reduce to the mean-only model when the

factors or continuous predictors do not significantly improve the fit over this model.