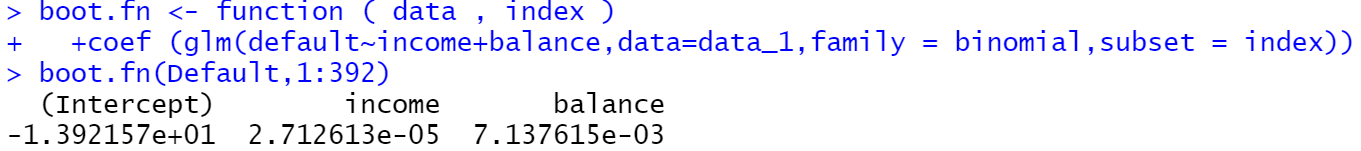
We continue to consider the use of a logistic regression model to predict the probability of default using income and balance on the Default data set. In particular, we will now compute estimates for the standard errors of the income and balance logistic regression coefficients in two different ways: (1) using the bootstrap, and (2) using the standard formula for computing the standard errors in the glm() function. Do not forget to set a random seed before beginning your analysis.

1. Using the summary() and glm() functions, determine the estimated standard errors for the coefficients associated with income and balance in a multiple logistic regression model that uses both predictors.   
   一張含有 文字 的圖片

   自動產生的描述

The standard errors for the income and balance are 4.985e-6 and 2.274e-4, respectively.

1. Write a function, boot.fn(), that takes as input the Default data set as well as an index of the observations, and that outputs the coefficient estimates for income and balance in the multiple logistic regression model.

The boot.fn() is shown below

1. Use the boot() function together with your boot.fn() function to estimate the standard errors of the logistic regression coefficients for income and balance.

一張含有 文字 的圖片

自動產生的描述

t2: income, t3: balance

1. Comment on the estimated standard errors obtained using the glm() function and using your bootstrap function.

|  |  |  |
| --- | --- | --- |
|  | Income | Balance |
| glm | 4.985e-06 | 2.274e-04 |
| Bootstrap | 8.051853e-08 | 2.010909e-05 |

The estimated standard errors are different from glm and bootstrape method and the bootstrap method provide lower estimated standard errors.

We will now consider the Boston housing data set, from the ISLR2 library.

* 1. Based on this data set, provide an estimate for the population mean of medv. Call this estimate ˆµ.

The estimate population mean is 22.53281

* 1. Provide an estimate of the standard error of ˆµ. Interpret this result. Hint: We can compute the standard error of the sample mean by dividing the sample standard deviation by the square root of the number of observations.

The estimate of standard error of ˆµ is 0.4088611

* 1. Now estimate the standard error of ˆµ using the bootstrap. How does this compare to your answer from (b)?

一張含有 文字 的圖片

自動產生的描述

The estimate is also 22.53281, but the std. error different form the answer is (b), increase form .4089 to .4107.

* 1. Based on your bootstrap estimate from (c), provide a 95 % confidence interval for the mean of medv. Compare it to the results obtained using t.test(Boston$medv). Hint: You can approximate a 95 % confidence interval using the formula [ˆµ − 2SE(ˆµ), µˆ + 2SE(ˆµ)].   
     The 95% confidence interval is

|  |  |  |
| --- | --- | --- |
|  | Upper bound | Lower bound |
| t.test | 21.72953 | 23.33608 |
| Bootstrap | 21.71148 | 23.35413 |

* 1. Based on this data set, provide an estimate, ˆµmed, for the median value of medv in the population.  
     the median of the medv is 21.2
  2. We now would like to estimate the standard error of ˆµmed. Unfortunately, there is no simple formula for computing the standard error of the median. Instead, estimate the standard error of the median using the bootstrap. Comment on your findings.  
     一張含有 文字 的圖片

     自動產生的描述

The estimated standard error is 0.3708901, we can have the 95% confidence interval is (20.45822, 21.94178)

* 1. Based on this data set, provide an estimate for the tenth percentile of medv in Boston census tracts. Call this quantity ˆµ0.1. (You can use the quantile() function.)   
     The tenth percentile of medv is 12.75
  2. Use the bootstrap to estimate the standard error of ˆµ0.1. Comment on your findings.

一張含有 文字 的圖片

自動產生的描述

The standard error of tenth percentile is .4994414. and the 95% confidence interval will be (11.75112, 13.74888)