1. 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.

library("ISLR")  
data(Default)  
attach(Default)

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.

glm.fit = glm(default ~ income + balance, family=binomial)  
summary(glm.fit)

##   
## Call:  
## glm(formula = default ~ income + balance, family = binomial)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4725 -0.1444 -0.0574 -0.0211 3.7245   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.154e+01 4.348e-01 -26.545 < 2e-16 \*\*\*  
## income 2.081e-05 4.985e-06 4.174 2.99e-05 \*\*\*  
## balance 5.647e-03 2.274e-04 24.836 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2920.6 on 9999 degrees of freedom  
## Residual deviance: 1579.0 on 9997 degrees of freedom  
## AIC: 1585  
##   
## Number of Fisher Scoring iterations: 8

Se puede ver con claridad que tanto **income** como **balance** son variables predictoras significativas para este modelo pues p-value < .05; sin embargo, tanto la **desviación residual** como el **AIC**, son muy grandes, lo cual es un indicativo que el modelo entrenado no es el más el adecuado.

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.

##Se establece una semill para que los resultados siempre sean replicables.  
  
set.seed(13)   
boot.fn <- function(data, index) {  
 fit <- glm(default ~ income + balance, data = data, family = "binomial", subset = index)  
 return (coef(fit))  
}

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.

set.seed(13)  
library(boot)  
boot(Default, boot.fn, 1000)

##   
## ORDINARY NONPARAMETRIC BOOTSTRAP  
##   
##   
## Call:  
## boot(data = Default, statistic = boot.fn, R = 1000)  
##   
##   
## Bootstrap Statistics :  
## original bias std. error  
## t1\* -1.154047e+01 -2.509758e-02 4.106349e-01  
## t2\* 2.080898e-05 2.341013e-08 4.571344e-06  
## t3\* 5.647103e-03 1.483680e-05 2.162846e-04

Los estimados de los errores estandar generados por el boostrap son:

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

Al comparar los resultados de los errores estándar estimados para los puntos (c) y (d),vemos que son muy cercanos entre sí, por lo cual se puede decir que el modelo boostrap es eficiente.

La cercanía entre estos valores es cada vez más notoria al incrementar el número de observaciones utilizadas en el boostrap como se puede ver a continuación.

## Boostrap con 100 datos  
  
set.seed(13)  
library(boot)  
boot(Default, boot.fn, 100)

##   
## ORDINARY NONPARAMETRIC BOOTSTRAP  
##   
##   
## Call:  
## boot(data = Default, statistic = boot.fn, R = 100)  
##   
##   
## Bootstrap Statistics :  
## original bias std. error  
## t1\* -1.154047e+01 -3.496958e-02 4.214075e-01  
## t2\* 2.080898e-05 6.619300e-07 5.012127e-06  
## t3\* 5.647103e-03 9.342470e-06 2.173372e-04

## Boostrap con 5000 datos  
set.seed(13)  
library(boot)  
boot(Default, boot.fn, 5000)

##   
## ORDINARY NONPARAMETRIC BOOTSTRAP  
##   
##   
## Call:  
## boot(data = Default, statistic = boot.fn, R = 5000)  
##   
##   
## Bootstrap Statistics :  
## original bias std. error  
## t1\* -1.154047e+01 -2.999144e-02 4.343384e-01  
## t2\* 2.080898e-05 -1.531876e-09 4.795164e-06  
## t3\* 5.647103e-03 1.769232e-05 2.289305e-04