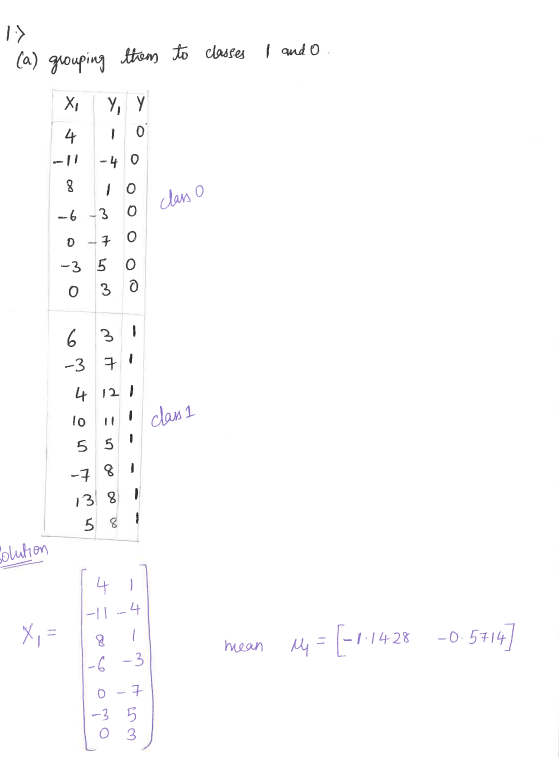
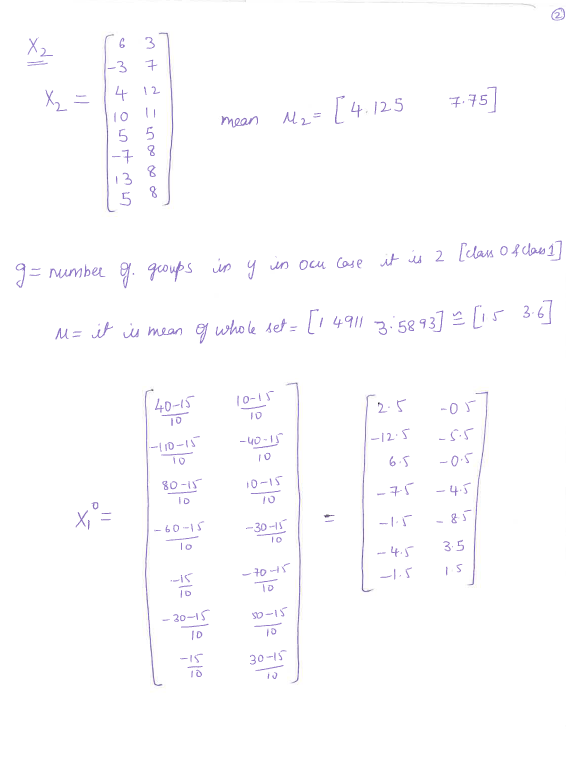
**Linga Sai Yuvesh Venketa Kotiala**

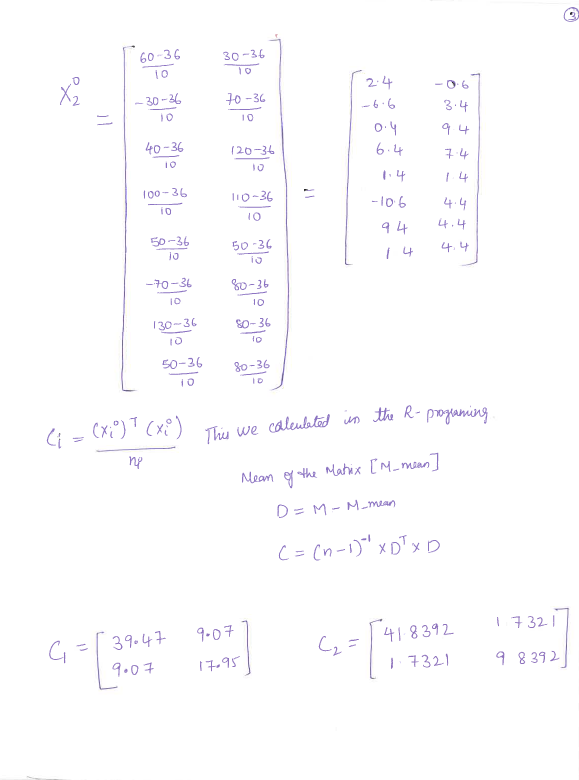
**16242113**

**ISL HW-4**

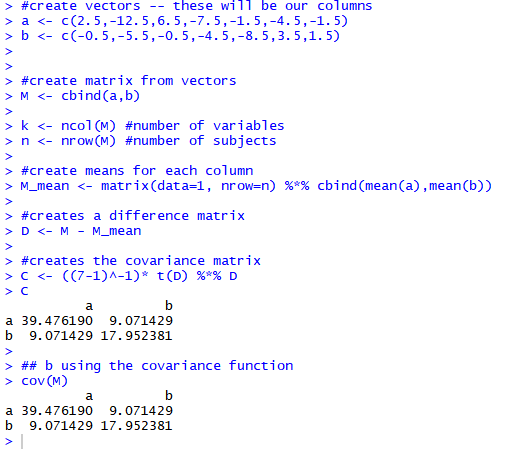
**1. (a) Use LDA to build a classier for the following data. To get full credit you must show all work and all discriminant values.**



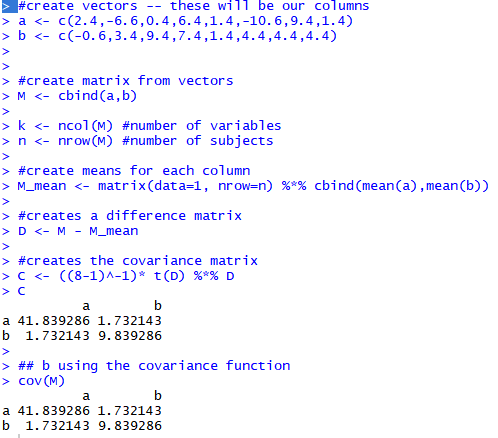


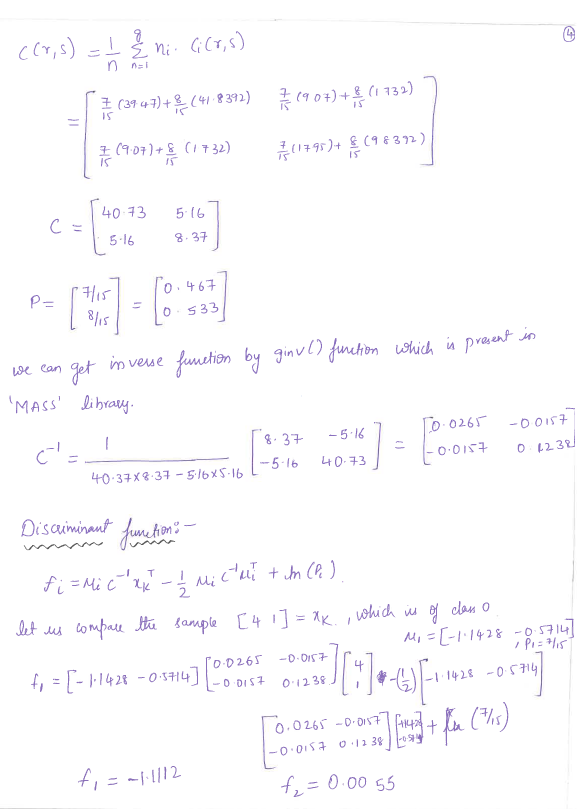


C1

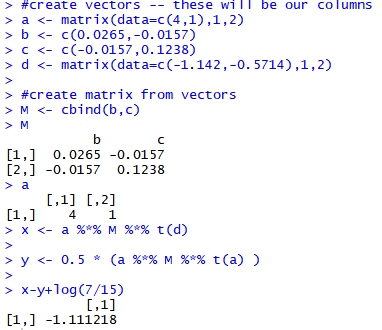


C2

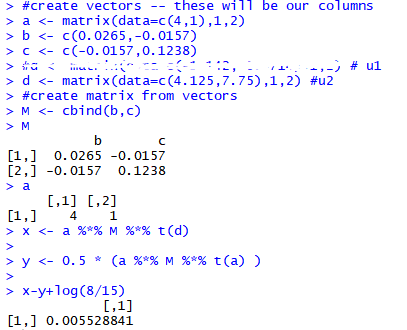




**[4,1] class0**

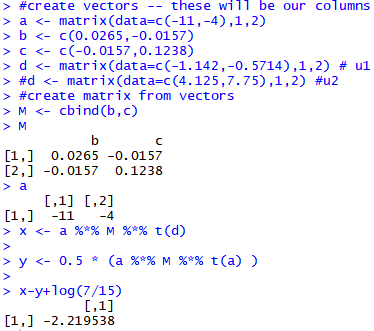


F1 = -1.1112

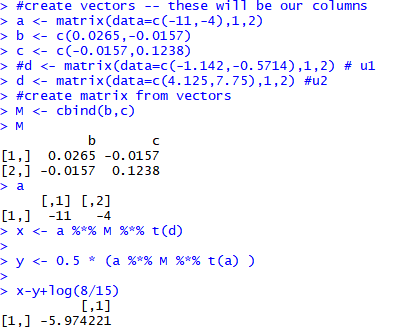


F2 = 0.0055

**[-11, -4] class0**

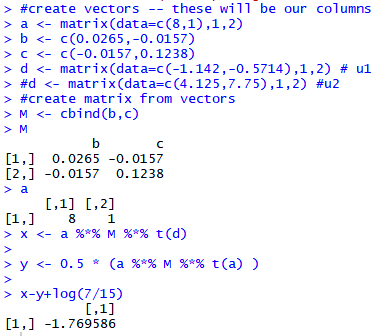


F1 = -2.2195

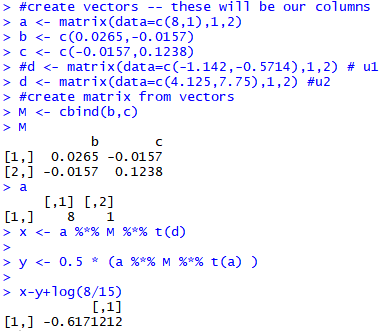


F2 = -5.9742

**[8,1] class0**

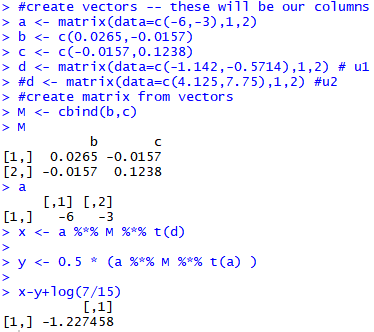


**F1=**-1.769586

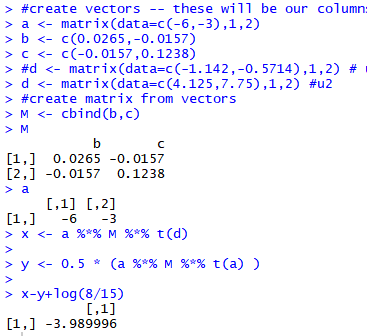


F2 = -0.6171

**[-6, -3]** **class0**

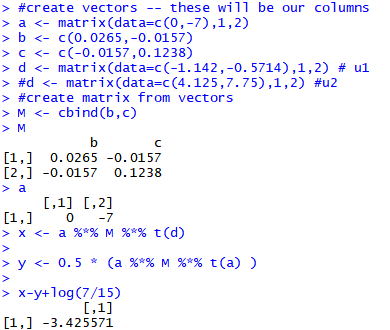


F1= -1.227458

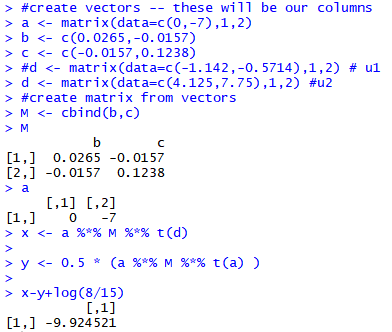


F2= -3.989996

**[0, -7] class0**

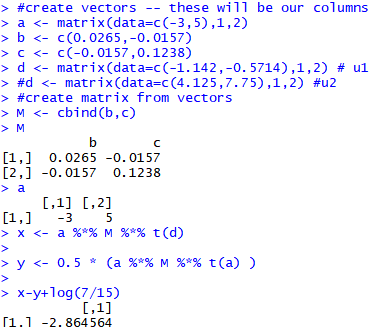


F1= -3.425571

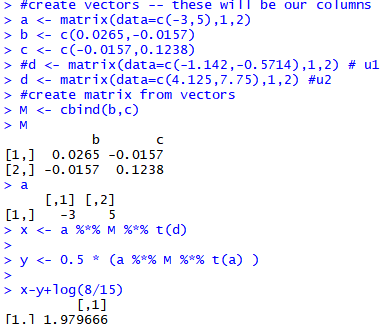


F2= -9.924521

**[-3,5] class0**

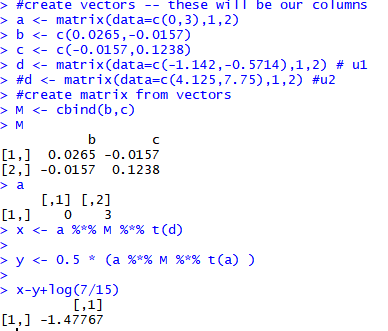


F1= -2.864564

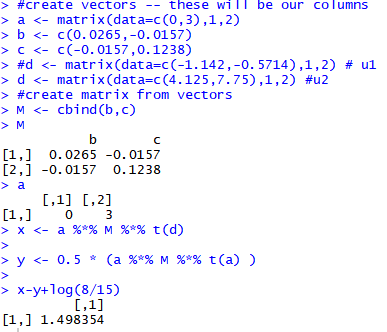


**F2=** 1.979666

**[0,3] class0**

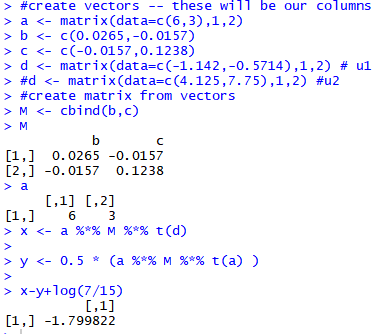


F1= -1.47767

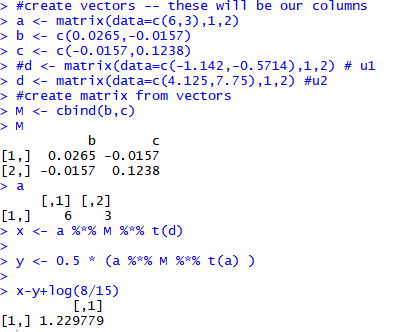


F2= 1.498354

**[6,3]** **class1**

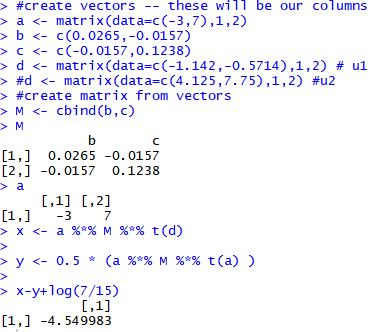


F1=-1.799822

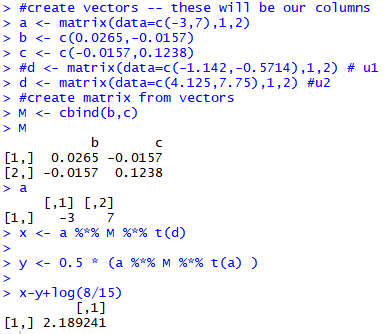


F2=1.229779

**[-3,7] class1**

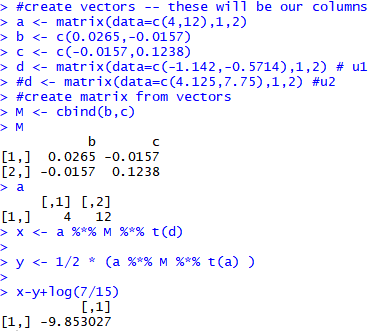


F1=-4.549983

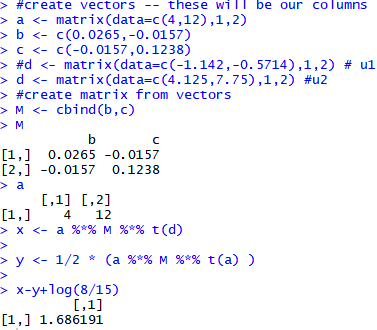


F2=2.189241

**[4,12] class1**

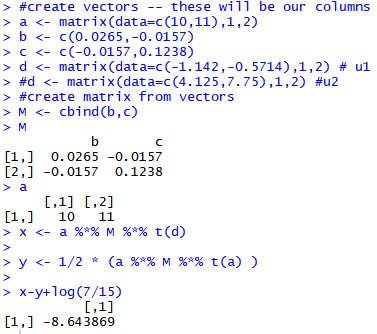


F1=-9.853027

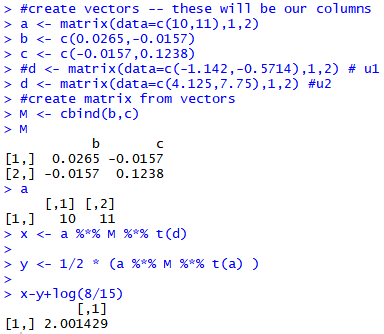


F2=1.686191

**[10,11] class1**

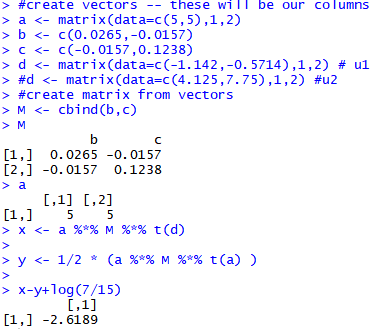


F1=-8.643869

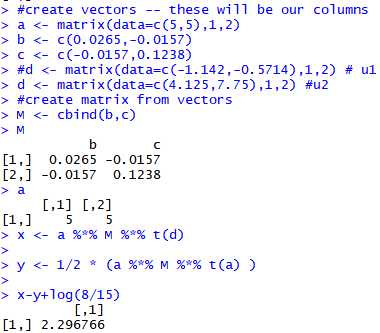


F2=2.001429

**[5,5] class1**

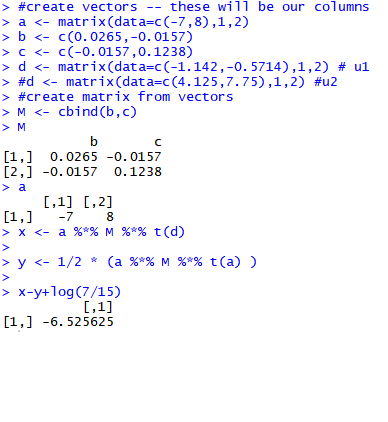


F1=-2.6189

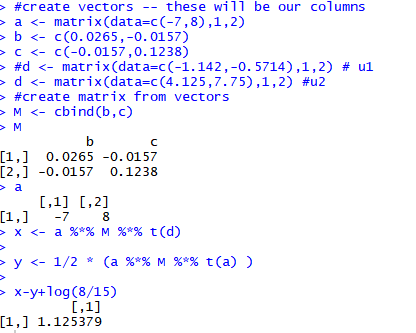


F2=2.296766

**[-7,8] class1**

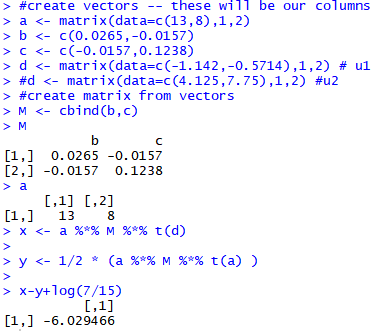


F1=-6.525625

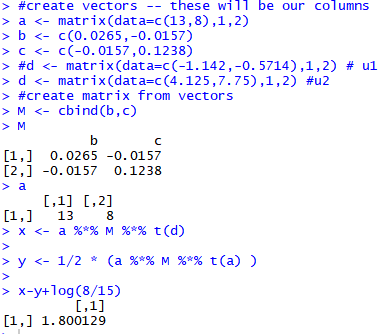


F2=1.125379

**[13,8] class1**

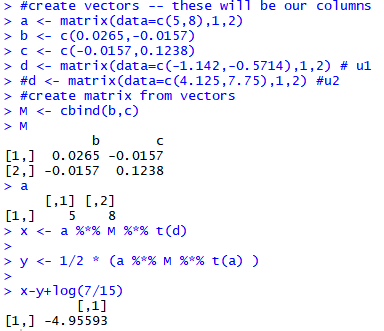


F1=-6.029466

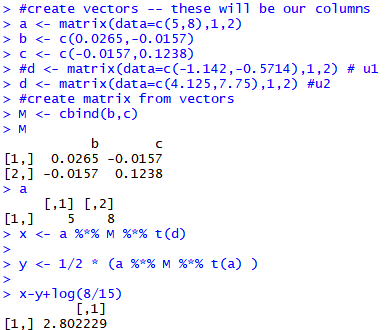


F2=1.800129

**[5,8] class1**



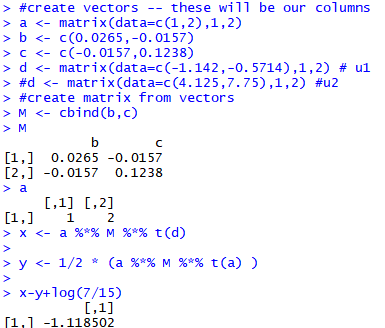
F1=-4.95593



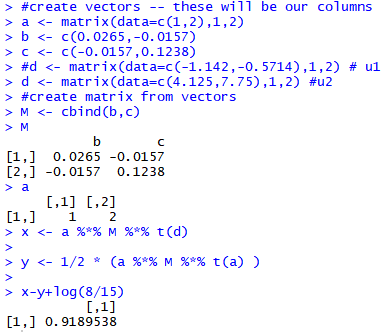
F2=2.802229

(b)

**[1,2]**

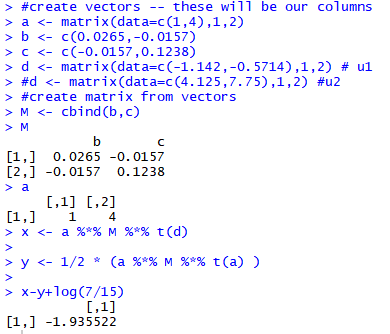


F­1= -1.118502

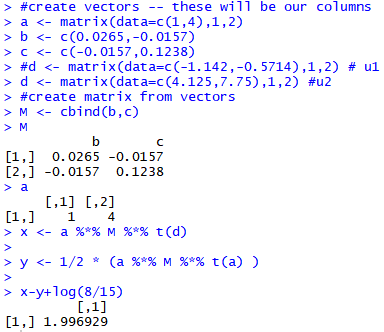


F2= 0.9189538

**[1,4]**

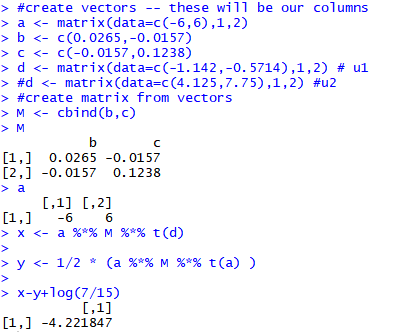


F­1= -1.935522

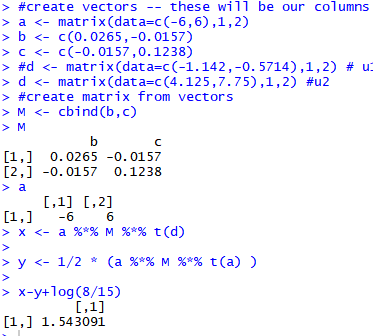


F2= 1.996929

**[-6,6]**

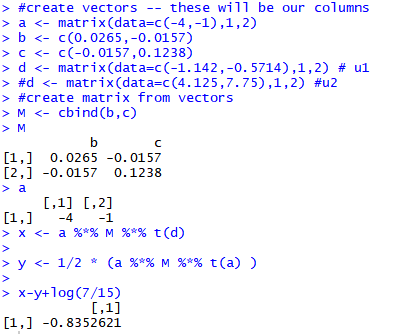


F­1= -4.221847

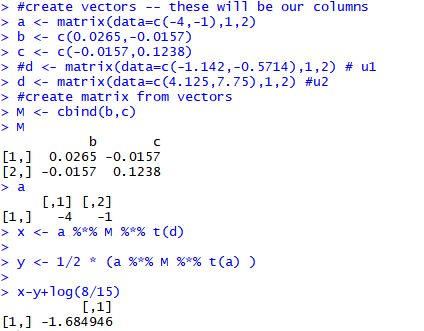


F2= 1.543091

**[-4, -1]**

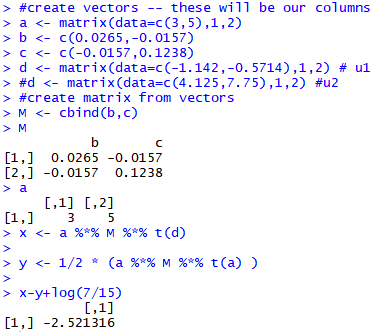


F­1= -0.8352621

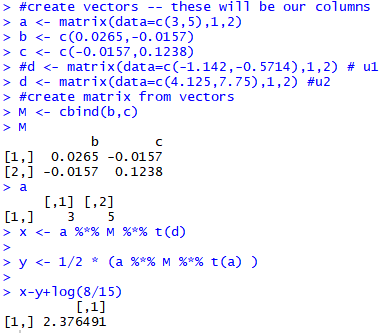


F2= -1.684946

**[3,5]**

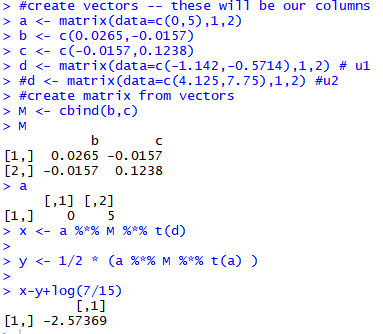


F­1= -2.521316

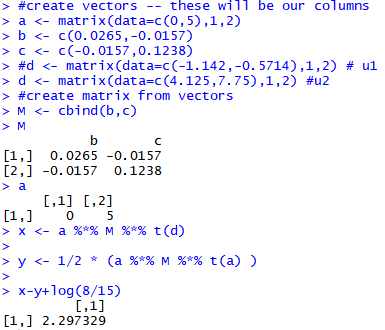


F2= 2.376491

**[0,5]**

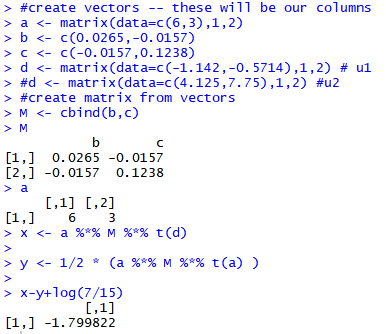


F­1= -2.57369

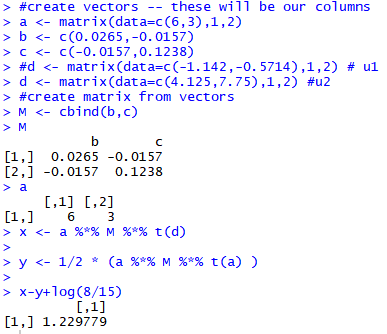


F2= 2.297329

**[6,3]**

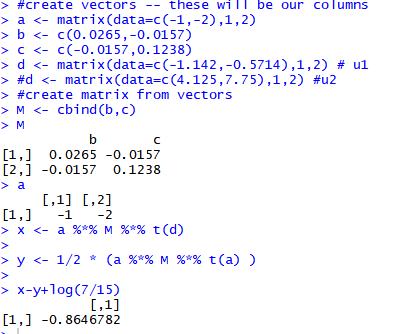


F­1= -1.799822

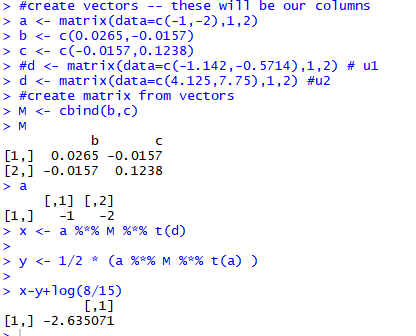


F2= 1.229779

**[-1, -2]**

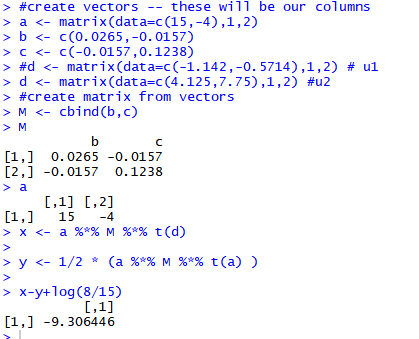


F­1= -0.8646782

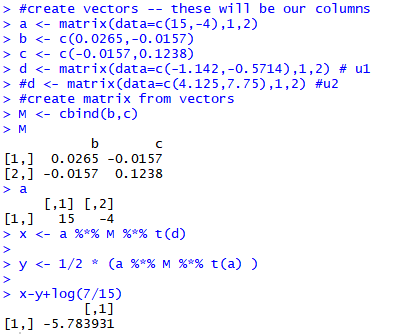


F2= -2.635071

**[15, -4]**

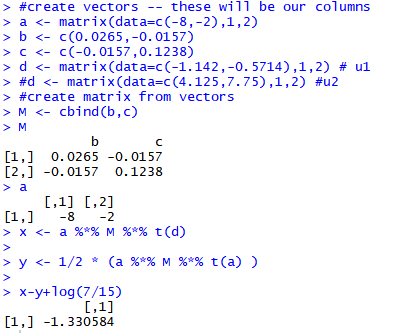


F­1= -9.306446

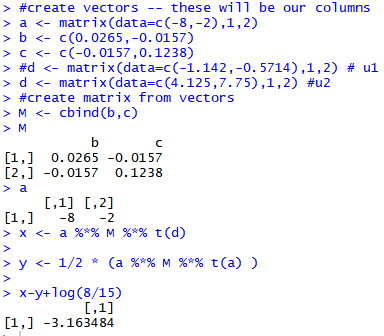


F2= -5.783931

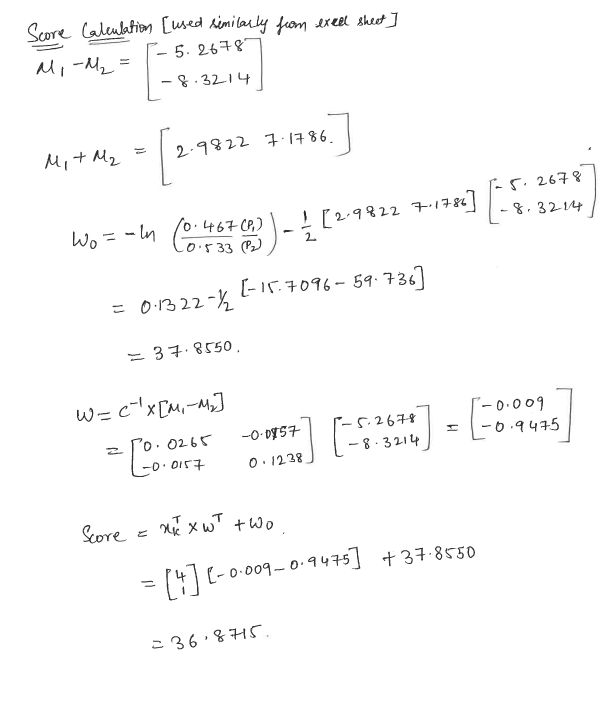
**[-8, -2]**

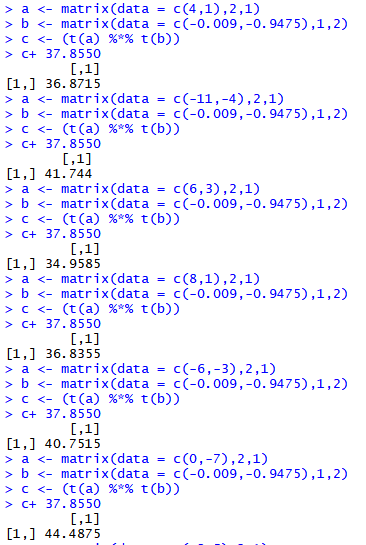


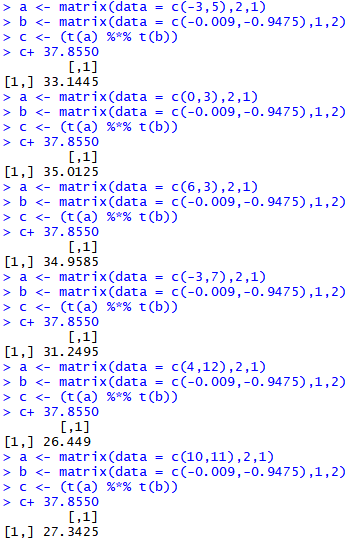
F­1= -1.330584

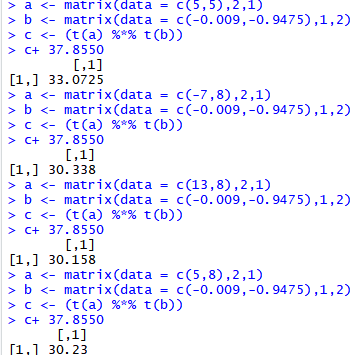


F2= -3.163484







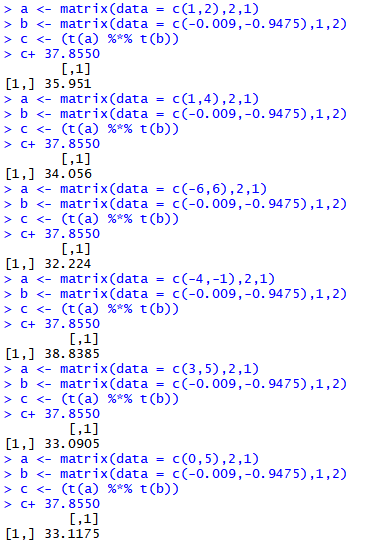


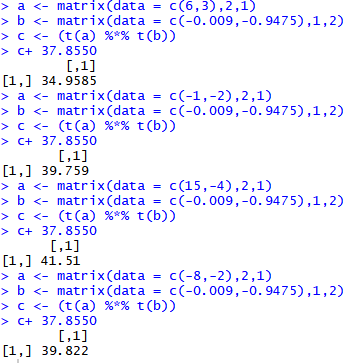
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X1 | X2 | Y | F1 | F2 | Score |
| 4 | 1 | 0 | -1.1112 | 0.0055 | 36.8715 |
| -11 | -4 | 0 | -2.2195 | -5.9742 | 41.744 |
| 8 | 1 | 0 | -1.769586 | -0.6171 | 36.8355 |
| -6 | -3 | 0 | -1.227458 | -3.989996 | 40.7515 |
| 0 | -7 | 0 | -3.425571 | -9.924521 | 44.4875 |
| -3 | 5 | 0 | -2.864564 | 1.979666 | 35.1445 |
| 0 | 3 | 0 | -1.47767 | 1.498354 | 35.0125 |
|  |  |  |  |  |  |
| 6 | 3 | 1 | -1.799822 | 1.229779 | 34.9585 |
| -3 | 7 | 1 | -4.549983 | 2.189241 | 31.2495 |
| 4 | 12 | 1 | -9.853027 | 1.686191 | 26.449 |
| 10 | 11 | 1 | -8.643869 | 2.001429 | 27.3425 |
| 5 | 5 | 1 | -2.6189 | 2.296766 | 33.0725 |
| -7 | 8 | 1 | -6.525625 | 1.125379 | 30.338 |
| 13 | 8 | 1 | -6.029466 | 1.800129 | 30.158 |
| 5 | 8 | 1 | -4.95593 | 2.802229 | 30.23 |

So by this we can assume that all the values of class 0 have the score above 35 (class0 if(score>35)) and we see that if the scores are below 35 then it is class1

**1.(b)**

**Ans.**

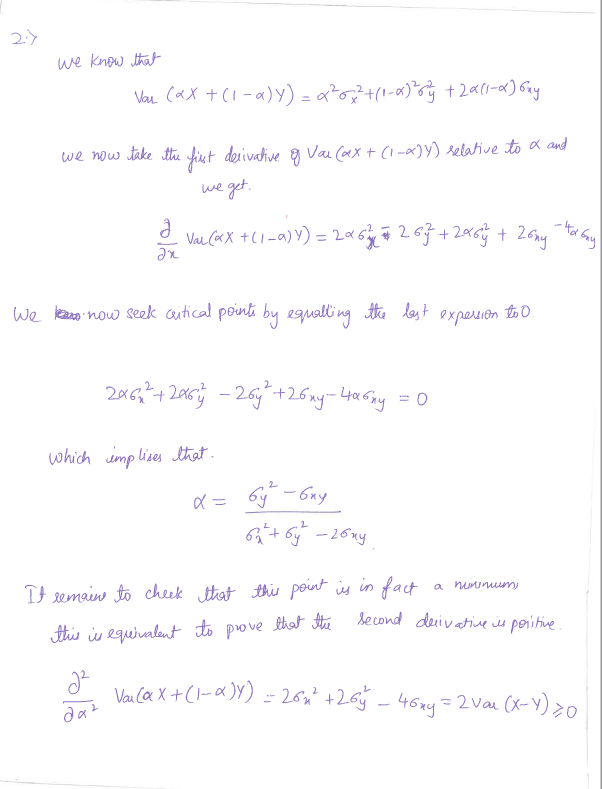




|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X1 | X2 | F1 | F2 | score | Expected Y |
| 1 | 2 | -1.118502 | 0.9189538 | 35.951 | 0 |
| 1 | 4 | -1.935522 | 1.996929 | 34.056 | 1 |
| -6 | 6 | -4.221847 | 1.543091 | 32.224 | 1 |
| -4 | -1 | -0.8352621 | -1.684946 | 38.8385 | 0 |
| 3 | 5 | -2.521316 | 2.376491 | 33.0905 | 1 |
| 0 | 5 | -2.57369 | 2.297329 | 33.1175 | 1 |
| 6 | 3 | -1.799822 | 1.229779 | 34.9585 | 1 |
| -1 | -2 | -0.8646782 | -2.635071 | 39.759 | 0 |
| 15 | -4 | -9.306446 | -5.783931 | 41.51 | 0 |
| -8 | -2 | -1.330584 | -3.163484 | 39.822 | 0 |

**2. Using basic statistical properties of the variance, as well as single variable calculus, prove that *given* by equation (5.6) in the text book does indeed minimize Var(*\_X* +(1 *α \_*)*Y* ).**

**Ans.**



**3.**  **We will now derive the probability that a given observation is part of a bootstrap sample. Suppose that we obtain a bootstrap sample from a set of n observations.**

**(a) What is the probability that the first bootstrap observation is not the jth observation from the original sample? Justify your answer.**

**Ans.**

There are n observations in the sample since bootstrap sampling draws items with replacement, we are sampling the data from the same pool with some probability every time. Those are (n-1) items in the n observations that are not in hence the probability is

(n-1)/n which can be written as :

1−1/n.

**(b) What is the probability that the second bootstrap observation is not the jth observation from the original sample?**

**Ans.**

Since we draw with replacement, it is same as the one bootstrapping sample:

1−1/n.

**(c) Argue that the probability that the jth observation is not in the bootstrap sample is**

**(1−1/n)n.**

**Ans.**

As bootstrapping sample with replacement, we have that the probability that the jth observation is not in the bootstrap sample is the product of the probabilities that each bootstrap observation is not the jth observation from the original sample

(1−1/n)⋯(1−1/n)=(1−1/n)n

as these probabilities are independant.

**(d) When n=5, what is the probability that the jth observation is in the bootstrap sample?**

**Ans.**

We have

P(jth obs in bootstrap sample)=1−(1−1/5)5=0.672.

**(e)When n=100, what is the probability that the jth observation is in the bootstrap sample ?**

**Ans**

We have

P(jth obs in bootstrap sample)=1−(1−1/100)100=0.634.

**(f) When n=10000, what is the probability that the jth observation is in the bootstrap sample ?**

**Ans.**

We have

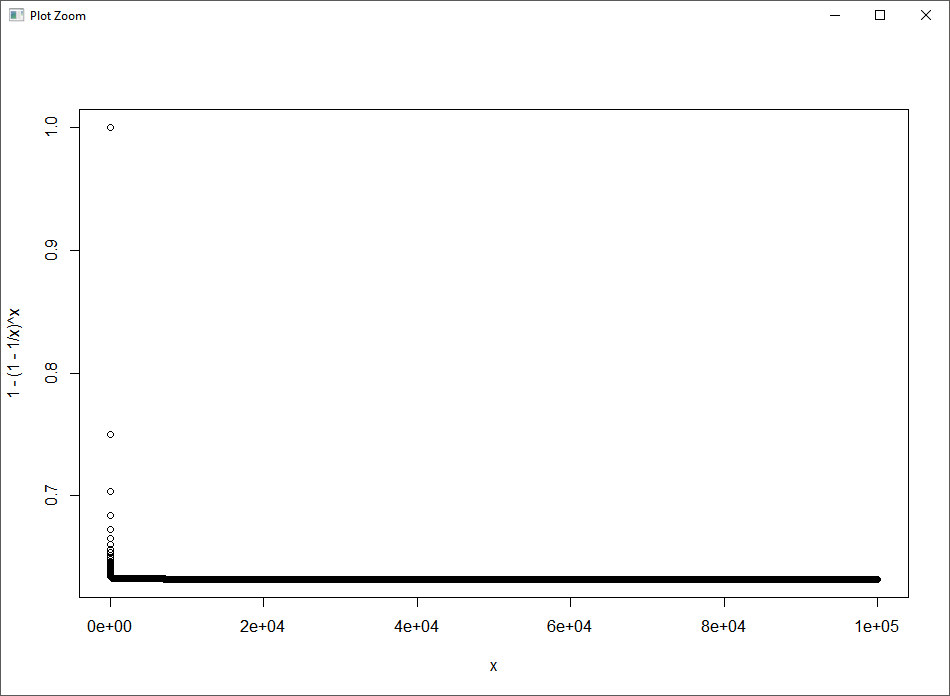
P(jth obs in bootstrap sample)=1−(1−1/10000)10000=0.632.

**(g) Create a plot that displays, for each integer value of n from 1 to 100000, the probability that the jth observation is in the bootstrap sample. Comment on what you observe.**

**Ans.**

> x <- 1:100000

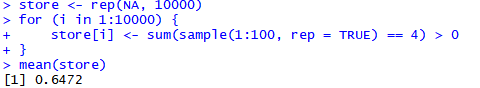
> plot(x, 1 - (1 - 1/x)^x)



We may see that the plot quickly reaches an asymptote at about *0.6320.632*.

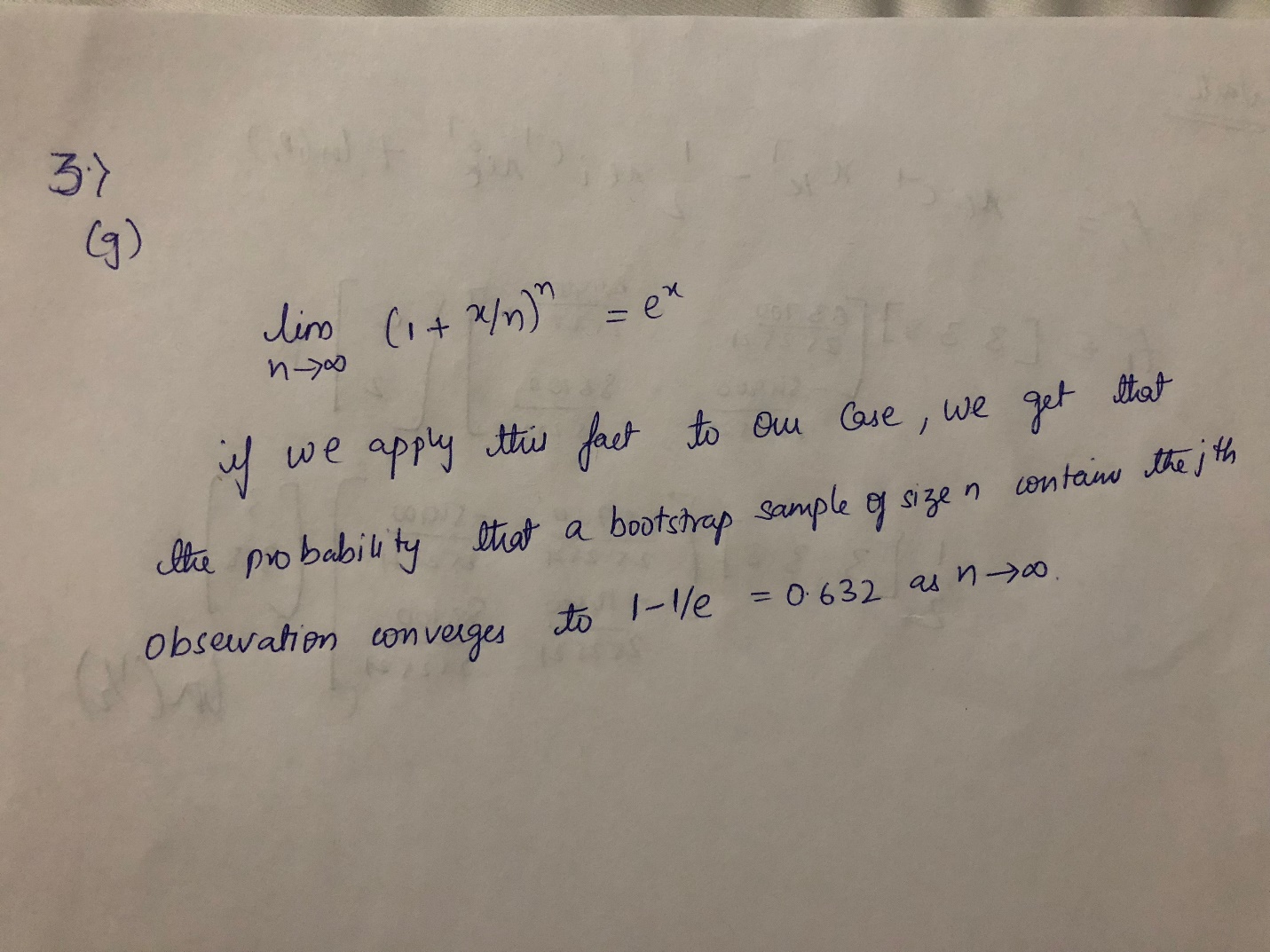
**(h) We will now investigate numerically the probability that a bootstrap sample of size n=100 contains the jth observation. Here j=4. We repeatedly create bootstrap samples, and each time we record whether or not the fourth observation is contained in the bootstrap sample.**

**Ans.**



Comment on the results obtained.

*A known fact from calculus tells us that*



**4. We now review k-fold cross-validation.**

**(a) Explain how k-fold cross-validation is implemented.**

**Ans.**

The k-fold cross validation is implemented by taking the *n* observations and randomly splitting it into *k* non-overlapping groups of length of (approximately) *n/k*. These groups act as a validation set, and the remainder (of length *n−n/k*) acts as a training set. The test error is then estimated by averaging the *k* resulting MSE estimates.

**(b) What are the advantages and disadvantages of k-fold cross validation relative to:**

**i. The validation set approach?**

**Ans.**

The validation set approach has two main drawbacks compared to k-fold cross-validation. First, the validation estimate of the test error rate can be highly variable (depending on precisely which observations are included in the training set and which observations are included in the validation set). Second, only a subset of the observations is used to fit the model. Since statistical methods tend to perform worse when trained on fewer observations, this suggests that the validation set error rate may tend to overestimate the test error rate for the model fit on the entire data set.

**ii. LOOCV?**

**Ans.**

The LOOCV cross-validation approach is a special case of k-fold cross-validation in which *k=n*. This approach has two drawbacks compared to k-fold cross-validation. First, it requires fitting the potentially computationally expensive model *n* times compared to k-fold cross-validation which requires the model to be fitted only *k* times. Second, the LOOCV cross-validation approach may give approximately unbiased estimates of the test error, since each training set contains *n−1* observations; however, this approach has higher variance than k-fold cross-validation (since we are averaging the outputs of *n* fitted models trained on an almost identical set of observations, these outputs are highly correlated, and the mean of highly correlated quantities has higher variance than less correlated ones). So, there is a bias-variance trade-off associated with the choice of *k* in k-fold cross-validation; typically using *k=5* or *k=10* yield test error rate estimates that suffer neither from excessively high bias nor from very high variance.

**5. Suppose that we use some statistical learning method to make a prediction for the response Y for a value of the predictor X. Carefully describe how we might estimate the standard deviation of our prediction.**

**Ans.**

We may estimate the standard deviation of our prediction by using the bootstrap method. In this case, rather than obtaining new independent data sets from the population and fitting our model on those data sets, we instead obtain repeated random samples from the original data set. In this case, we perform sampling with replacement *BB* times and then find the corresponding estimates and the standard deviation of those *B*.