

Домашнее задание 5: продвинутый блок

Файл с выполненным заданием необходимо загрузить на Dropbox до дедлайна, указанного на сайте.

Домашние задания, сданные после срока, оцениваются с использованием понижающих коэффициентов: опоздание в пределах часа – штраф 10% от полученной оценки, в пределах суток – штраф 20%, в пределах недели – штраф 50%. Домашние задания, сданные через неделю после указанного срока и позже, не принимаются и не оцениваются.

Если при проверке работ установлен факт нарушения академической этики, студент получает оценку «0» за данную работу. Работа студента, предоставившего свою работу для списывания, также аннулируется.

Постановка задачи

Напишите функцию `fcorrs()`, которая принимает на вход параметры

- `n_obs` (число наблюдений);
- `mean_vect` (вектор средних значений);
- `corr_vect` (вектор коэффициентов корреляции между переменными);

и выдает матрицу, столбцы которой являются нормально распределенными переменными, средние значения которых равны значениям из вектора `mean_vect`, и которые скоррелированы друг с другом в соответствии со значениями коэффициентов корреляции из вектора `corr_vect`.

Подсказка. Вам понадобится функция `mvrnorm()` из библиотеки `MASS`.

Пример. Нужно создать матрицу из трех нормально распределенных переменных V_1 , V_2 и V_3 , таких, что среднее значение V_1 равно 1, среднее значение V_2 равно 2, среднее значение V_3 равно 3 и $Corr(V_1, V_2) = 0.3$, $Corr(V_1, V_3) = 0.6$, $Corr(V_2, V_3) = 0.8$.

```
# входные данные
```

```
n = 100
my_means = c(1, 2, 3)
my_corrs <- c(0.3, 0.6, 0.8)
```

```
# функция и ее вывод
```

```
M <- fcorrs(n_obs = n, means_vect = my_means, corrs_vect = my_corrs)
```

```
head(M)
```

```
##           [,1]      [,2]      [,3]
## [1,]  1.1671382  1.1397532  2.8107346
## [2,]  3.1760701  2.1983994  4.9081009
## [3,]  2.9798006  3.7932459  4.6263823
## [4,]  1.4008963  2.5409431  3.0599134
## [5,] -1.1826554  0.1618920  0.5489593
## [6,]  0.5325931  0.2066273  1.4986835
```

```
# проверка нужных характеристик
```

```
mean(M[, 1]); mean(M[, 2]); mean(M[, 3]) # средние
```

```

## [1] 1
## [1] 2
## [1] 3

cor(M[, 1], M[, 2]); cor(M[, 1], M[, 3]); cor(M[, 2], M[, 3]) # коэффициенты корреляции

## [1] 0.3
## [1] 0.6
## [1] 0.8

shapiro.test(M[, 1]); shapiro.test(M[, 2]); shapiro.test(M[, 3]) # нормальное распределение

##
## Shapiro-Wilk normality test
##
## data:  M[, 1]
## W = 0.99562, p-value = 0.9879

##
## Shapiro-Wilk normality test
##
## data:  M[, 2]
## W = 0.98316, p-value = 0.2324

##
## Shapiro-Wilk normality test
##
## data:  M[, 3]
## W = 0.98258, p-value = 0.2099

```

Базовый вариант (на 7-8 баллов)

Написать указанную выше функцию `fcorrs()` в предположении, что для любых значений коэффициентов корреляции между переменными можно получить нужную матрицу переменных (что на самом деле не так), но считая, что с функцией может работать несознательный пользователь, который будет вытворять безобразия — указывать в векторе корреляций значения, превышающие по модулю 1. В связи с последним, в случае, если функции на вход подается вектор корреляций, содержащий значения, выходящие за рамки $[-1,1]$, функция выводит на экран сообщение “The vector of correlation coefficients contains inappropriate values.” и больше ничего не делает.

Продвинутый вариант (на 9-10 баллов)

Написать указанную выше функцию `fcorrs()`:

Во-первых, в предположении, что с функцией может работать несознательный пользователь. В связи с последним, в случае, если функции на вход подается вектор корреляций, содержащий значения, выходящие за рамки $[-1,1]$, функция выводит на экран сообщение “The vector of correlation coefficients contains inappropriate values” и больше ничего не делает.

Во-вторых, с учетом того, что не для любых значений коэффициентов корреляции между переменными можно получить нужную матрицу переменных — корреляционная матрица должна быть неотрицательно определена (определитель матрицы должен быть больше или равен 0). В случае, если созданная по заданному вектору коэффициентов корреляций корреляционная матрица отрицательно определена, функция `fcorrs()` должна делать следующее:

- выводить на экран сообщение “Matrix of correlations is not positive definite. Trying to find the nearest positive definite matrix.”;
- в качестве корреляционной матрицы брать ближайшую положительно определенную матрицу (функция `nearPD()` из библиотеки `Matrix`), выводить на экран старые (поданные пользователем на вход) значения коэффициентов корреляций с пометкой *Old values of correlation coefficients*, выводить на экран новые (из ближайшей положительно определенной матрицы) значения коэффициентов корреляций с пометкой *New values of correlation coefficients*, выводить на экран [норму Фробениуса](#) для разности исходной (неправильной) корреляционной матрицы и новой (правильной) матрицы с пометкой “Frobenius norm of matrix difference”; выдавать матрицу переменных, полученную на основе заданного пользователем вектора средних значений и ближайшей положительно определенной корреляционной матрицы.

Пример.

```
F <- fcorrs(n_obs = 100, means_vect = c(2, 5, 7), corrs_vect = c(0.3, 0.4, -0.8))
```

```
## Matrix of correlations is not positive definite.
## Trying to find the nearest positive definite matrix.
## Old values of correlation coeffs: 0.3 0.4 -0.8
## New values of correlation coeffs: 0.2843482 0.3837501 -0.7761998
## Frobenius norm of matrix difference: 0.04637858
```

```
head(F)
##           [,1]      [,2]      [,3]
## [1,]  3.0937341  6.353769  6.415080
## [2,] -0.3572004  1.919156  8.417409
## [3,]  3.5617773  4.192694  8.804377
## [4,]  0.4378003  5.801622  5.200375
## [5,]  1.2141611  5.492333  6.009009
## [6,]  2.4222829  5.823135  6.484889
```