

# Rコーディングテスト練習問題

Copyright 2021 DataMix Co., Ltd.

## 本練習問題の対象者

ご自身でRのインストール及びRstudioを利用した経験、Rプログラミングに関する一定の知識があることを前提に作られています。

## 本試験に関する注意事項

- こちらの練習問題は、入学試験で出題される問題の類題です。
- 実際の入学試験は、2021年11月25日まではHackerRankというサービスで実施され、Rで関数を作成する形式の問題も出題されます。
- 2021年11月22日から行われる入学試験はExcertというサービス上で全て選択問題の形で出題される予定です。

## Q1.ベクトルの操作

以下のRコードが実行された状態で、それぞれのコードを実行した場合に得られる数字を書いてください。

```
num_list <- c(10, 20, 30, 40)
```

解答例:

2と4 が出力される場合は、半角数字と半角カンマで 2,4 と記載してください。

エラーになる場合は、errorと半角英字で記載してください。

1. num\_list
2. num\_list[2:4]
3. num\_list[c(2,4)]
4. num\_list[-1]
5. num\_list[2,4]

## Q2.データフレームの列の取り出し

以下のデータを読み込んだデータフレームdfがあります。dfの3列目の列全体を抽出するコードを全て選びなさい。

ただし、dfのindexは1から332まで連番で並んでいるものとする。

df

	id_assessment	id_student	date_submitted	is_banked	score
1	1752	248270	19	0	77
2	1752	255548	19	0	84
3	1752	256815	37	0	71
...	...	...	...	...	...
331	1753	2367155	58	0	85
332	1753	2369266	53	0	50

332 rows × 5 columns

1. head(df,n=3)
2. df[3,]
3. df[,3]
4. df\$date\_submitted
5. df[3,3]

## Q3.統計量の確認の仕方

Rで標準偏差を計算する際の関数のとして正しいものを1つ選択しなさい。

1. mean
2. median
3. var
4. sd
5. cor

## Q4.データフレームの操作

以下の処理を行うRコードを完成させてください。

datasetというフォルダの中にcsvファイルが8個あり、それらのファイルを全て読み込み一つのデータフレームにします。

次に、以下の手順でデータを確認します。

- ・行数・列数を確認する
- ・列名の一覧を確認する
- ・最初の数行を確認する
- ・要約統計量を計算する
- ・カラム "score" とカラム "date\_submitted" の散布図を描く。

最後に、カラム "score" が50点以下のレコードのみを取り出し、取り出されたデータのカラム "id\_student" と "score" だけを、1つのcsvファイルに出力します。

この処理を行う場合のRコードが以下に用意されています。

Rコードの空欄に当てはまるコードを次のコード群の中から選んでコードの番号を入力してください。

### 注意

コード群の番号は、半角数字で入力すること (例. 【1】に(2) が該当する場合、2と入力する)  
csvファイルの内容の一例は、csvデータの例を参照のこと (全てのcsvファイルのカラムは同一とする)

### コード群

```
(1)read_csv (2)read.csv (3)if (4)for (5)i in c(2:8) (6)i in 2~8 (7)i in 2:8  
(8)summary (9)count (10)cat (11)dim (12)headers (13)head  
(14)colnames (15)sum (16)barplot (17)boxplot (18)hist (19)plot  
(20)score (21)dataset (22)score, date_submitted  
(23)dataset$score, dataset$date_submitted (24)dataset["score"], dataset["date_submitted"]  
(25)subset (26)select (27)"score" <= 50 (28)score <= 50 (29)dataset$score <= 50  
(30)"id_student", "score" (31)("id_student", "score") (32)c("id_student", "score")  
(33)write_csv (34)make_csv (35)write.csv (36)make.csv (37)何も入力しない
```

### Rコード

```
# 1番目のcsvファイルを読みこむ
dataset <- 【1】("dataset/dataset1.csv")

# 2~8番目のファイルを読み込み、結合する
【2】(【3】){
  # i番目のファイル名を定義
  excel_path <- paste("dataset/dataset", i, ".csv", sep = "")

  # rbindで、読み込み済みデータと、i番目のファイルを縦に結合する
  dataset <- rbind(dataset, 【1】(excel_path))
}

# 行数・列数を確認する
【4】(dataset)

#列名の一覧を確認する
【5】(dataset)

# 最初の数行を確認する
【6】(dataset)

#要約統計量を確認する
【7】(dataset)

#カラム"score"とカラム"date_submitted"の散布図を描く
【8】(【9】)

# カラム"score"が50以下に絞り込む
under50_data <- 【10】(dataset, 【11】)

# カラム"id_student", "score"だけを取り出す
under50_data_lmt <- under50_data[【12】,【13】]

# csv形式で出力する
【14】(under50_data_lmt,"integrated_df.csv")
```

csvデータの例(dataset1.csv)

	<b>id_assessment</b>	<b>id_student</b>	<b>date_submitted</b>	<b>is_banked</b>	score
1	1752	248270	19	0	77
2	1752	255548	19	0	84
3	1752	256815	37	0	71
4	1752	258392	18	0	58
5	1752	258894	19	0	58
...	...	...	...	...	...
328	1753	2344901	54	0	45
329	1753	2358056	54	0	60
330	1753	2364471	62	0	85
331	1753	2367155	58	0	85
332	1753	2369266	53	0	50

332 rows × 5 columns

## Q5.Rコードの解釈

以下のコードの解釈として不適切なものを全て選んでください。

Rコード

```
dataset <- read.csv("dataset.csv",header=TRUE)

head(dataset,5)

ncol(dataset)

summary(dataset)

dataset[,8]<-NULL

boxplot(dataset[,1]~dataset[,5])
```

1. CSVファイルを読み込んでいる
2. データの先頭5行を確認している
3. データの行数を確認している
4. データの8列目を削除している
5. データの5列目を1列目の値ごとにグループにして箱ひげ図を描いている

## Q6.順位相関

ある学校で、入学試験の成績と入学後の成績の相関を調べる。入学試験の成績も入学後の成績も、それぞれ順位で示されている。

入学試験の順位の番号を、入学後の順位で並べたベクトルを入力した場合に、入学試験の順位と入学後の順位の相関係数を出力する関数rank\_correlationを書いてください。

答えは、小数点第5位を四捨五入して、第4位まで答えてください。

例:入学者が5人の場合

- 入学試験の成績順:1,2,3,4,5
- 入学後の順位(入試の番号を並べる):3,2,1,4,5

→この2つの順位の相関係数を調べます。

入力:nyugakugo <- c(3,2,1,4,5)

解き方のヒント:

小数点第5位を四捨五入する場合は、以下のように書きます。

```
round(0.12345, digits=4)
```

```
rank_correlation <- function(nyugakugo){  
  #ここにコードを記載してください  
}
```

```
nyugakugo <- c(3,2,1,4,5)  
rank_correlation(nyugakugo)
```