

## نکات مهم:

- برای تحویل تمارین گزارشی با استفاده از RMarkdown ایجاد نموده که در آن به صورت کامل صورت سوال، کد مربوطه، خروجی کد و توضیحات مربوطه آورده شود و سپس در سامانه آموزش مجازی بارگذاری شود. برای بارگذاری، ابتدا فایل word خروجی RMarkdown را تهیه کرده و سپس آن را pdf نموده و فایل pdf را به همراه فایل Rmd در سامانه بارگذاری کنید. صورت سوال را می توانید به صورت یک عکس در گزارش بیاورید.
- دقت کنید که حتما برای پاسخ به تمامی قسمت های سوالات کدی نوشته و پاسخ را به صورت خروجی کد بیاورید.

۱. از تابع paste برای ایجاد بردارهای حرفی زیر به طول ۳۰ استفاده کنید:
  - ("label 1", "label 2", ....., "label 30") (دقت کنید که یک فاصله تکی (single space) بین label و اعداد وجود دارد)
  - ("fn1", "fn2", ..., "fn30") (دقت کنید که هیچ فاصله ای بین fn و اعداد وجود ندارد)

۲. ابتدا کدهای زیر را به منظور ایجاد دو بردار به طول ۲۵۰ از متغیرهای صحیح تصادفی که با جایگذاری از بین اعداد ۰، ۱، ۲، ...، ۹۹۹ انتخاب می شوند، اجرا کنید.

```
set.seed(50)
```

```
xVec = sample(0:999, 250, replace=T)
```

```
yVec = sample(0:999, 250, replace=T)
```

فرض کنید بردار xVec با نماد  $\mathbf{X}=(x_1, x_2, \dots, x_n)$  و بردار yVec با نماد  $\mathbf{Y}=(y_1, y_2, \dots, y_n)$  نمایش داده شوند.

- بردار  $(y_2 - x_1, \dots, y_n - x_{n-1})$  را ایجاد کنید.
  - بردار  $\left(\frac{\sin(y_1)}{\cos(x_2)}, \frac{\sin(y_2)}{\cos(x_3)}, \dots, \frac{\sin(y_{n-1})}{\cos(x_n)}\right)$  را ایجاد کنید.
  - بردار  $(x_1 + 2x_2 - x_3, x_2 + 2x_3 - x_4, \dots, x_{n-2} + 2x_{n-1} - x_n)$  را ایجاد کنید.
  - $\sum_{i=1}^{n-1} \frac{e^{-x_{i+1}}}{x_{i+1}+10}$  را محاسبه کنید. (از تابع sum استفاده کنید)
- (بدون استفاده از حلقه ها و با استفاده از بردارها موارد بالا را حل کنید).
- مقادیری را در بردار yVec مشخص کنید که مقدارشان از ۶۰۰ بزرگتر است. اندیس این مقادیر را در بردار yVec پیدا کنید و مقادیر متناظرشان را در بردار xVec پیدا کنید.

## تمرین سری اول – تحلیل آماری کاربردی

۳. ابتدا کدهای زیر را به منظور ایجاد یک ماتریس  $10 \times 6$  از متغیرهای صحیح تصادفی که با جایگذاری از بین اعداد ۱، ۲، ...، ۱۰ انتخاب می شوند، اجرا کنید.

```
set.seed(75)
```

```
aMat = matrix(sample(10, size=60, replace=T), nr=6)
```

- تعداد عناصری را در هر سطر ماتریس پیدا کنید که بزرگتر از ۴ هستند.
- کدام سطرها شامل دقیقاً ۲ بار تکرار عدد ۷ هستند؟

۴. مجموعه داده cereal را از پکیج liver گرفته (ابتدا پکیج را نصب و سپس آن را فراخوانی کنید) و به سوالات زیر پاسخ دهید:

- ۱۰ سطر اول مجموعه داده را فراخوانی کنید.
- متغیرهای مختلف مجموعه داده و نوع آن ها را مشخص کنید.
- یک متغیر جدید تحت عنوان totalcarb به مجموعه داده اضافه کنید. این متغیر را جمع دو متغیر carbo و sugars در نظر بگیرید.
- نوع چه تعداد از غلات (cereals) در این مجموعه داده گرم است؟ (type="Hot")
- کدام یک از غلات بیشترین پروتئین (protein) را دارد؟ نام (name) و تولید کننده (Manuf) آن را مشخص کنید.
- چه تعداد سازنده منحصر بفرد (unique manufacturers) در این مجموعه وجود دارد؟ (راهنمایی: استفاده از دستور unique)
- زیرمجموعه ای از این مجموعه داده را پیدا کنید که کمتر از ۸۰ کالری (calories) و بیش از ۲۰ واحد ویتامین (vitamins) دارند؟ کالری غلات این زیرمجموعه را با علامت ؟ به عنوان مقدار گم شده جایگزین کنید. زیر مجموعه حاصل را فقط با در نظر گرفتن متغیرهای calories، name، vitamins و sugars نمایش دهید.
- چه تعداد از غلات چربی (fat) بیشتر از ۲ دارند؟ چربی این غلات را با علامت ؟ به عنوان مقدار گم شده جایگزین کنید.
- نام متغیر Manufacturer را به Producer تغییر دهید.
- متغیر shelf را از مجموعه داده حذف کنید
- مجموعه داده تغییر داده شده (با اضافه کردن یک متغیر جدید، جایگزین کردن برخی مقادیر با علامت ؟، تغییر نام متغیر Manuf و حذف متغیر shelf) را در قالب یک فایل csv جدید ذخیره کنید. (دستور write.csv)
- تعداد مقادیر گم شده در هر ستون مجموعه داده تغییر داده شده را به تفکیک Producer مشخص کنید. برای این منظور، ابتدا یک تابع تعریف کنید که تعداد مقادیر گم شده در هر ستون را شمارش می کند و سپس با استفاده

## تمرین سری اول – تحلیل آماری کاربردی

از یک حلقه و استفاده از دستور supply، تعداد داده های گمشده در هر ستون را به تفکیک Producer مشخص کنید. (تشیویی: به جای استفاده از حلقه از تابع by استفاده کنید)

- در مجموعه داده تغییر داده شده، یک متغیر تحت عنوان anymiss اضافه کنید که دارای دو مقدار miss و nomiss است. در صورتی این متغیر مقدار miss اتخاذ می کند که در مشاهده مربوطه مقدار گمشده ای وجود داشته باشد.

۵. تابع پیوسته زیر را در نظر بگیرید:

$$f(x) = \begin{cases} x^2 + 2x + 3 & \text{if } x < 0 \\ x + 3 & \text{if } 0 \leq x < 2 \\ x^2 + 4x - 7 & \text{if } x \geq 2 \end{cases}$$

یک تابع در R با نام tmpFn بنویسید که یک بردار xVec گرفته و برای تک تک عناصر آن مقدار تابع f(x) را در قالب یک بردار برگرداند.

۶. فرض کنید  $x_0=1$  و  $x_1=2$  و

$$x_j = x_{j-1} + \frac{2}{x_{j-1}} \quad \text{for } j = 2, 3, \dots$$

تابعی تحت عنوان testLoop بنویسید که یک مقدار n گرفته و n-1 مقدار اول توالی  $\{x_j\}_{j \geq 0}$  را برگرداند. (به عبارت دیگر، مقادیر  $x_0, x_1, x_2, \dots, x_{n-2}$  را برگرداند)

۷. در مباحث سری زمانی، برای یک بردار  $(x_1, \dots, x_n)$ ، مقدار خودهمبستگی نمونه از وقفه k (the sample autocorrelation of lag k) از طریق رابطه زیر محاسبه می شود:

$$r_k = \frac{\sum_{i=k+1}^n (x_i - \bar{x})(x_{i-k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

به عنوان مثال،

$$r_1 = \frac{\sum_{i=2}^n (x_i - \bar{x})(x_{i-1} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{(x_2 - \bar{x})(x_1 - \bar{x}) + \dots + (x_n - \bar{x})(x_{n-1} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- تابعی تحت عنوان tmpFn2 بنویسید که یک بردار xVec را گرفته و یک متغیر لیست شامل مقادیر  $r_1$  و  $r_2$  را برگرداند. به طور خاص، از این تابع استفاده کرده و مقادیر  $r_1$  و  $r_2$  را برای بردار (2, 5, 8, ..., 53, 56) محاسبه کنید.

## تمرین سری اول – تحلیل آماری کاربردی

---

- **(تشویقی)** تابع قبلی را تعمیم داده به صورتی که دو ورودی داشته باشد: یکی بردار  $xVec$  و دیگری یک مقدار صحیح  $k$  که می تواند از ۱ تا  $n-1$  مقدار بگیرد ( $n$  طول بردار  $xVec$  است). این تابع باید برداری از مقادیر  $r_1, r_2, \dots, r_k$  را برگرداند. مقدار پیش فرض  $k$  را در تابع برابر ۲ در نظر بگیرید. (راهنمایی: برای حل ساده تر می توان از خانواده توابع `apply` مانند `sapply` استفاده کرد)
- 

موفق باشید – شادمان