



بخش عملی

در این تمرین قصد داریم تا با داده ساختارها و ابزارهای محیط R آشنا شویم.

مسئله ۱.

دیتافریم iris (یکی از دیتافریم های built-in در R) را در نظر بگیرید. می خواهیم بخشی از سطرهای این دیتافریم را استخراج کنیم (توصیه می شود که از کتابخانه dplyr برای فیلتر کردن سطرها استفاده کنید):

الف

الف) سطرهایی که در شرایط زیر صدق می کنند را استخراج کنید و در دیتافریم filtered_iris ذخیره کنید.

$$Sepal.Length > 5 \bullet$$

$$Petal.Length > 3/5 \bullet$$

$$Species = virginica \bullet$$

ب

قدر مطلق تفاضل مقادیر Sepal.Length و Sepal.Width را برای هر سطر بدست آورده و در یک بردار به نام Rectangularity ذخیره کنید.

پ

بردار Rectangularity ره به عنوان ستونی جدید به دیتافریم filtered_iris اضافه کنید و نام این ستون جدید را به Sepal.Rectangularity تغییر دهید.

ت

سطرهای دیتافریم را بر حسب ستون جدید به صورت نزولی مرتب کنید و ۷ سطر ابتدایی دیتافریم را چاپ کنید

ث

می خواهیم ستونی جدید به دیتافریم اضافه کنیم که مقادیر آن از جنس Boolean هستند و مقدار آن اینگونه تعیین می شود:

$$X = \begin{cases} TRUE, & avg(Sepal.Rectangularity, Petal.Length, Petal.Width) < 5 \\ FALSE, & otherwise \end{cases}$$

ج

دیتافریم iris را تبدیل به ماتریس کرده و آن را در خروجی چاپ کنید

مسئله‌ی ۲.

دیتافریم mtcars (از دیتافریم های built-in در R) را در یک دیتافریم جدید به نام cars کپی کنید.

الف

با استفاده از دستور supply مقدار cyl هر سطر را دو برابر کنید.

ب

ب) با استفاده از تابع aggregate، میانه هر ستون را به ترتیب disp حساب کنید.

پ

ابتدا توابع زیر را پیاده سازی کنید:

- `random_shuffle(df)`: سطرهای df را بهم میریزد و برمیگرداند
 - `get_top_rows_variance(df, n, col)`: واریانس اعداد n سطر اول مربوط به ستون col را برمیگرداند
- سپس با استفاده از این توابع ۱۰۰ بار دیتافریم را در هم ریزی کرده و در مرحله i ام واریانس مقادیر ستون disp در i سطر اول را حساب کرده و در نهایت این مقادیر را در یک بردار ذخیره کنید.
- اطلاعات فایل data.csv را بخوانید و در dataframe ای به نام df ذخیره کنید:

ت

با استفاده از تابع subset، نمونه هایی که شرط زیر را دارند را استخراج کنید و مقدار source مربوط به این نمونه ها را در یک بردار ذخیره کنید.

- `sweet > ۲/۵`

ث

در صورتی که مقدار pioneer در این بردار وجود دارد، yes وگرنه no چاپ کنید.

مسئله‌ی ۳.

در این مسئله خواهیم دید که اگر چگونگی تولید نمونه از یک متغیر تصادفی نامشخص باشد، احتمال یک رویداد که روی آن تعریف میشود ممکن است خوش تعریف نباشد. یک دایره و یک مثلث متساوی الاضلاع محاط شده داخل آن در نظر بگیرید. وتری به دلخواه انتخاب می‌کنیم. می‌خواهیم احتمال اینکه طول این وتر از ضلع مثلث بیشتر باشد

را بدست آوریم. انتخاب وتر به روش های مختلفی امکان پذیر است و ما در اینجا ۳ روش را بررسی می کنیم. به ازای هر روش ابتدا به روش تئوری این احتمال خواسته شده را محاسبه کنید سپس با تکمیل کردن توابع script.R و با استفاده از شبیه سازی، این احتمال را برای هر روش محاسبه کنید:

الف

در این روش ۲ نقطه ی دلخواه بر روی محیط دایره انتخاب می شود که نشان دهنده ی دو سر وتر مدنظر هستند.

ب

در این روش یکی از شعاع های دایره به دلخواه انتخاب می شود. سپس نقطه ای دلخواه روی شعاع در نظر میگیریم. این نقطه بیانگر نقطه میانی وتر مدنظر است.

پ

در این روش یک نقطه ی دلخواه درون دایره انتخاب می شود که نشان دهنده ی نقطه ی میانی وتر است. توصیه می شود که برای هر روش یک تابع جداگانه بنویسید که نمونه گیری از وتر را انجام دهد.

نکات مهم

- فایل های R مربوط به هر سوال را به فرمت شماره سوال با پسوند R ذخیره کنید و آنها را زیپ کرده و آپلود کنید
- سال نو مبارک D:

(موفق باشید :)