

Математические пакеты

Домашнее задание №8 (R)

Дедлайн до 20.11 23:59

- (1) Таблица «Социологический опрос» (файл `opinion_poll.R`) (4 балла):
- (a) Создать таблицу данных «Социологический опрос» из 6 случаев и 5 переменных с помощью конструкции `data.frame()`. Имена переменных задать следующие: `name` (ФИО), `sex` (Пол), `age` (Возраст), `education` (Образование), `income` (Доход).

name	sex	age	education	income
John Davis	Male	21	Secondary	125
Angela Williams	Female	23	Higher	345
Janice Markhammer	Female	40	-	50
Bullwinkle Moose	Male	35	Secondary	399
Mary Rayburn	Female	18	Secondary	136
Greg Knox	Male	20	Higher	120

- (b) Переменные Пол и Образование представить как факторные. Проверить, что эти переменные действительно факторные любым удобным способом с соответствующим сообщением в консоли.
- (c) Добавить в таблицу новую переменную `income_20` (Доход_20), значения которой равны Доход + премия в размере 20% от Дохода.
- (d) Рассчитать среднее значение Дохода и вывести его на экран.
- (e) Используя инструменты R построить столбчатую диаграмму для Возраста и Дохода.
- (f) Добавить в таблицу новую переменную `income_S` (Доход_C), значения которой равны Доход + Доход_20.
- (g) Добавить новую переменную `income_M` (Доход_M). Необходимо найти максимальный доход и для каждого респондента определить, сколько процентов от максимального дохода составляет его доход.
- (h) Рассчитать среднее значение дохода в зависимости от пола.
- (i) Удалить столбец `income_20`.
- (2) Синусоида с погрешностью (файл `sine_err.R`) (2 балла):
- (a) Составить таблицу 1000×3 , содержащую в качестве первой переменной значения $x_i \in [0, 2\pi]$ (x_i изменяется с постоянным шагом), в качестве второй – значения $\sin x_i$ и в качестве третьей – значения $\sin x_i + X$, где $X \sim \mathcal{N}(0, 0.005)$.
- (b) Создать новый dataframe, в котором будут сохранены только первые 100 строк и последние 50 строк таблицы из п. (a).
- (c) Используя инструменты R вывести первые 5 строчек датафрейма из п. (b).
- (d) Используя инструменты R вывести последние 5 строчек датафрейма из п. (b).
- (e) Нарисовать 2 синусоиды, используя данные из п. (a).
- (3) Словарный запас (файл `vocabulary.R`) (6 баллов):
- (a) Загрузите в переменную `vocab` файл `Vocab.csv` с помощью функции `read.csv()`. Файл содержит информацию об экспериментах

по определению словарного запаса людей разного пола и образования. Он содержит следующие столбцы:

- идентификатор эксперимента
 - год эксперимента
 - пол испытуемого
 - количество лет образования
 - результат теста на словарный запас: число от 0 до 10 — количество известных испытуемому слов из списка в 10 слов.
- (b) Выполните функции `str(vocab)` и `summary(vocab)`. Убедитесь, что вы понимаете всю информацию, выведенную на экран. Убедитесь, что данные о поле испытуемых загружены в виде фактора.
- (c) Посчитайте средний словарный запас всех испытуемых, потом средний словарный запас мужчин, потом средний словарный запас женщин.
- (d) Вообще, нет смысла считать средний словарный запас для людей разного возраста. Верните вектор `vocab.means` со средним словарным запасом отдельно для каждого возраста (количества лет образования). Используйте `names(vocab.means)`, чтобы назначить имена каждому элементу вектора, это должны быть строки вида «x лет». Используйте функцию `paste()` для соединения текстов. Например, `paste(10, "лет")` создает строку «10 лет».
- (e) Постройте столбцовую диаграмму, где по горизонтали отложено количество лет образования, а по вертикали — средний словарный запас
- (f) Постройте аналогичную диаграмму, только для каждого года нужно рисовать два столбца - средний словарный запас отдельно мужчин и женщин.
- (g) С помощью `t-test` проверьте, отличается ли средний словарный запас мужчин и женщин из самого большого количества лет образования.
- (h) Как изменяется словарный запас исторически? С помощью функции `unique()` узнайте, какие года представлены в данных.
- (i) Отберите людей с нулевым количеством лет образования. Постройте график зависимости их среднего словарного запаса от года эксперимента. На графике должна отображаться линия, а не набор точек. Добавьте на график зависимости для людей со всеми остальными количествами лет образования.