

## Задача 2. Основы программирования на Питоне.

### Первая часть: создание data Frame путём генерации

Согласно теме (см. перечень тем) необходимо выполнить:

Формирование таблицы в форме словарей Python и data Frame в Pandas с сохранением в \*.xlsx и \*.csv : с именем ishodVar.\*

Количество строк таблицы N от 30 до 45

1 столбец – номер (1,2 3 4... N)

2 столбец: Фамилия Имя – заполняется последним вручную по смыслу (после розыгрыша остальных столбцов)

3 столбец – год рождения (разыгрывается по среднему и дисперсии)

4 столбец – пол (м или ж): разыгрывается по заданному соотношению вероятностей в зависимости от темы)

5 столбец – вес в кг: разыгрывается по – разному (м или ж) по среднему и дисперсии для м и ж

6 столбец – рост в см: разыгрывается по – разному (м или ж) по среднему и дисперсии для м и ж

7 столбец – 1 качественный признак из заданного набора: разыгрывается по списку вероятностей

8 столбец - 2 качественный признак из заданного набора: разыгрывается по списку вероятностей

9 столбец – количественный признак по смыслу темы (разыгрывается по среднему и дисперсии)

10 \*\*\*столбец – 3 качественный признак из заданного набора: разыгрывается по списку вероятностей (дополнительный)

Столбцы: 3 , 5 , 6 и 9 формируются после исследования на **корректность** результатов: средние значения и дисперсию надо подобрать так (для нормального – Гауссова распределения), чтобы результаты имели **здравый смысл**. Следует использовать: пакет numpy: `numpy.random.randn()` или `numpy.random.normal(scale=Rsig, loc=Rsr, size=N)`

Столбцы 4 , 7, 8 для качественных признаков формируются по списку вероятностей, учитывая, что в списке:  $\sum p_i = 1$

Каждую из этих вероятностей можно разыграть как её место в доле от 0 до 1 при `numpy.random.random()` или с использованием `numpy.random.choice()`

Вес и рост : 5 и 6 столбцы разыгрываются отдельно для лиц мужского и женского пола, после розыгрыша полов (вероятности взять в зависимости от темы( 0,5 и 0,5 или, например: 0,3 и 0,7 – по смыслу темы) После розыгрыша веса и роста результаты помещаются в столбцы 5 и 6.

Если требуется разыграть возраст для различных категорий людей: учащиеся, работающие, пенсионеры, – то это надо учитывать при выборе среднего и дисперсии.

**Функции розыгрышей желательно оформить в виде пакета с модулями, для которых провести дополнительное тестирование на корректность результатов.**

## Темы к задаче 2.

1. Спортсмены
2. Студенты
3. Сотрудники магазина
4. Разработчики ПО
5. Автомобилисты.
6. Пациенты поликлиники
7. Пользователи смартфона
8. Клиенты интернет - магазина
9. Велосипедисты
10. Старшеклассники

**Общий алгоритм:** Задаётся список номеров от 1 до N ( $N \geq 35 \div 45$ ). Разыгрывается возраст (по смыслу темы: для сотрудников, студентов с одними параметрами, для пациентов – с другими). Предварительно все розыгрыши отладить и протестировать. Далее разыгрывается пол. После этого изымаются номера одного пола и для них разыгрывается рост и вес, затем также для другого пола. После этого можно вручную заполнить фамилию и имя (или использовать методы генерации или модификации фамилий и имен. Далее разыгрываются столбцы с качественными и количественными признаками. После формирования серий – словарей (имя столбца: список) формируется dataframe и записывается в файл.

Тема

1 Спортсмены	Первый качественный признак					
	Вид спорта	атлетика	лыжи	плавание	гимнастика	хоккей
	Вероятности (м):	0,3	0,2	0,1	0,1	0,3
	Вероятности (м):	0,25	0,15	0,2	0,4	0
2 Студенты	профиль	хим технология	машиностроение	Менеджмент	Психология	Информац. Технологии
	Вероятности:	0,2	0,2	0,1	0,1	0,4
3 Сотр. Магазины	Профиль	продовольствие	Одежда	Электроника	Стройматериалы	Мебель
	Вероятности:	0,5	0,1	0,2	0,15	0,05

4	Разработчики ПО	Отделы	Web - разработчики	Web - Дизайнеры	Разработчики ядра	Системные администраторы	Тестировщики
		Вероятности:	0,3	0,15	0,05	0,3	0,2
5	Автомобилисты	Марка	BA3	Renault	VolksVagen	Hynday	Kia
		Вероятности:	0,4	0,1	0,2	0,25	0,05
6	Пациенты поликлинн.	отделение	Терапия	Хирургия	Стоматология	Эндокринология	Офтальмология
		Вероятности:	0,5	0,1	0,2	0,1	0,1
7	Пользователи смартф	Бренды	Apple	Samsung	Huawei	Xiaomi	Oppo
		Вероятности:	0,17	0,21	0,32	0,2	0,1
8	Клиенты интернет - магазина	Название	Wildberries	Ozon	Lamoda	Sitilink	Dns
		Вероятности:	0,25	0,25	0,15	0,2	0,15
9	Велосипедисты	Типы велосипедов	Складной	Дорожный	Спортивный	Туристический	Подростковый
		Вероятности:	0,15	0,25	0,15	0,35	0,1
10	Старшекласники	Направление	Гуманитарное	Естественно научное	Физико математическое	Экономическое	Информатика
		Вероятности:	0,1	0,2	0,2	0,2	0,3

ТАБЛИЦА 1 к задаче 2.

Тема

		Второй качественный признак			
1 Спортсмены	Уровень	разряд	к.м.с	м.с	
	Вероятности (м):	0,7	0,2	0,1	
	Вероятности (ж):	0,75	0,15	0,1	
2 Студенты	Уровень	Бакал 1-2 курс	Бакал 3-4 курс	Магистратура	
	Вероятности:	0,49	0,45	0,06	
3 Сотр. Магазины	Уровень	Продавец	Старш продавец	Зав секции	
	Вероятности:	0,6	0,35	0,05	
4 Разработчики ПО	Уровень	Junior	Middle	Senior	
	Вероятности:	0,5	0,45	0,05	
5 Автомобилисты	Категория	Эконом	Средний	Престиж	
	Вероятности:	0,75	0,2	0,05	
6 Пациенты поликлинн.					
	Соц статус	Работающие	Учащиеся	Пенсионеры	

7	Пользователи смартф	Вероятности:	0,5	0,1	0,4		
		Категория	Эконом	Средний	Престиж		
		Вероятности:	0,7	0,25	0,05		
8	Клиенты интернет - магазина	Использование	Иногда	Часто	Постоянно		
		Вероятности:	0,6	0,3	0,1		
		Использование	Иногда	Часто	Постоянно		
9	Велосипедисты	Вероятности:	0,3	0,55	0,15		
		Использование	Иногда	Часто	Постоянно		
10	Старшеклассники	Мотивация	Низкая	Средняя	Высокая		
		Вероятности:	0,1	0,2	0,7		

ТАБЛИЦА 2 к задаче 2.

Тема

Количественный признак по теме						
1	Спортсмены	Среднее	Среднеквадратичное отклонение			
		Значения(м):	12,5	4,5		
		Значения(ж):	12,5	4,5		
		Среднее	Среднеквадратичное отклонение			
2	Студенты	Число членов семьи	Среднее	Среднеквадратичное отклонение		
		Значения:	3,5	1,8		

3	Сотр. Магазины	Закредитованность	Среднее	Среднеквадратичное отклонение		
		Значения:	150000	80000		
4	Разработчики ПО	Душевой доход семьи	Среднее	Среднеквадратичное отклонение		
		Значения:	25000	12000		
5	Автомобилисты	Закредитованность	Среднее	Среднеквадратичное отклонение		
		Значения:	150000	80000		
6	Пациенты поликлиник.	Душевой доход семьи	Среднее	Среднеквадратичное отклонение		
		Значения:	25000	12000		
7	Пользователи смартф	Душевой доход семьи	Среднее	Среднеквадратичное отклонение		
		Значения:	25000	12000		
8	Клиенты интернет - магазина	Уровень дохода на клиента	Среднее	Среднеквадратичное отклонение		
		Значения:	35000	16000		
9	Велосипедисты	Душевой доход семьи	Среднее	Среднеквадратичное отклонение		
		Значения:	25000	12000		
10	Старшеклассники	Душевой доход семьи	Среднее	Среднеквадратичное отклонение		
		Значения:	25000	12000		

ТАБЛИЦА 3 к задаче 2.

## 2. Вторая часть. Работа со списками, кортежами, множествами и словарями

2.1. Прочитать data Set (data Frame) из файла предварительно переименовав его с lshodVar.\* на WorkVar.\* (Это нужно для того чтобы сохранить “первоисточник”).

Вывести его на экран.

Под **ЗАПИСЬЮ** далее будем понимать строку таблицы. Результаты выводить:

- A. На экран
- B. В текстовый файл
- C. В Excel или его аналог
  - Выделить серии.
  - Сортировать записи по возрастанию и убыванию в сериях.
  - Фильтровать записи по признакам, входящим в выражения как данной, так и других серий.
  - Формировать множества по усмотрению студента и демонстрировать операцию с ними
  - Реализовывать фильтрацию с сортировкой по усмотрению студента (не менее 5 – 6 вариантов)

2.2. Сформировать новые словари:



- С ключами из качественных признаков соответствующих серий и значениями, включающими кортежи из остальных серий.
- С ключами из кортежей с 2 качественными признаками соответствующих серий и значениями, включающими кортежи из остальных серий
- Значениями в словарях являются записи

2.3. **Фильтровать** словари по ограничениям с качественными и количественными признаками в кортежах значений, в том числе используя их комбинации:

- По полу
- Возрасту
- Росту
- Весу

2.4. Фильтровать словари по ограничениям с качественными и количественными признаками в кортежах значений характерных для заданной темы

2.5. Использовать пакет из модулей с функциями для решения выше поставленных задач

2.6. Вывод результатов рекомендуется осуществлять: на консоль, в текстовые файлы, используя append – 'а', в Excel

Использовать модули: numpy, pandas, openpyxl

Можно использовать также \*\*\* модули:

```
import matplotlib.pyplot as plt
import scipy.stats as sts
import copy
from copy import deepcopy
```