

Лабораторная работа №4. Описательная статистика в Python

Контрольные вопросы

1. Какие статистические значения для датафрейма позволяет получить метод `describe()` ?
- Метод « `describe()` » используется для отображения базовой статистики для фрейма данных
 - ✓ Количество элементов с « **count** »;
 - ✓ Среднее значение с « **mean** »;
 - ✓ Стандартное отклонение с « **std** »;
 - ✓ Максимум с « **max** »;
 - ✓ Минимум с « **min** »;

```
Entrée [5]: # Code cell 5
            # Use the describe method of the new dataframe to view the table of statistics

            # SOLUTION
            df.describe()
```

Out[5]:

	weight
count	10000.000000
mean	20.499212
std	0.199874
min	19.752000
25%	20.365000
50%	20.500000
75%	20.635000
max	21.171000

2. Как добавить в датафрейм новый столбец с данными ?
- Добавление нового столбца в `dataFrame`;

```
Entrée [7]: # Code cell 7
            # Create the new column named "diff" and fill it with values.
            # Check the result.

            # SOLUTION
            df['diff'] = df['rounded'] - 20
            df.head()
```

Out[7]:

	weight	rounded	diff
0	20.440	20.44	0.44
1	20.244	20.24	0.24
2	20.549	20.55	0.55
3	20.755	20.76	0.76
4	20.740	20.74	0.74

3. Что такое частотное распределение данных ?
- Распределение частот данных можно определить как ряд данных, содержащих уникальные значения данных, и частоты появления каждого из этих данных.

```
Entrée [14]: # Code cell 14
# Rename the columns in the dataframe. Verify the result.

# SOLUTION
freq.columns = ['value', 'freq']
freq.head()
```

Out[14]:

	value	freq
0	20.48	228
1	20.50	226
2	20.43	218
3	20.46	213
4	20.45	210

4. Что делает функция `to_frame()`, почему необходимо преобразовать серию данных в датафрейм ?
 - Функция « **to_frame()** » преобразует ряд данных в `dataFrame`.
 - Необходимо преобразовать ряд данных в `dataframe`, потому что в серии данных столбец индекса устанавливается с уникальными значениями данных, а имя столбца автоматически устанавливается на 0. это необходимо изменить, чтобы дать этому столбцу название.
5. Как при выводе значения переменной в текстовой строке оставить только 2 цифры после десятичной точки, если исходное значение содержит больше цифр ?
 - Чтобы оставить только 2 цифры после запятой, необходимо использовать зарезервированное пространство, чтобы указать, куда вставлять значения переменных, и вставить в него строку « `:.2f` ». Это даст: « `{:.2f}` »

```
Entrée [11]: # Code cell 11

# Format the standard deviation result to 2 decimal points
stdstring = 'The standard deviation of the distribution is {:.2f}'.format(std)
print(stdstring)

The standard deviation of the distribution is 0.20.
```

<https://github.com/malko1278/labo-analyse-big-data.git>