

Листок 01. Описательные статистики

Библиотека Pandas

Упражнение 1. Загрузите датасет `sleep75`.

1. вычислите размер датасета (число наблюдений & число переменных)
2. Заполните следующую таблицу со значениями переменных

index	sleep	totwrk	age	male
0				
5				
100				
700				

3. Вычислите корреляционную матрицу для следующих переменных: `sleep`, `totwrk`, `age`
4. Заполните следующую таблицу

Desc.Stat	sleep	totwrk	age	hrwage
max				
min				
mean				
median				
st.dev				
var (unbiased)				
var (biased)				
1st quartile				
3rd quartile				

Замечание: 1st/3rd квантили – 25%/75% квантили соответственно.

5. Сколько наблюдений соответствуют следующим условиям
 - (a) `sleep>3000`
 - (b) `totwrk<2000`
 - (c) `age>40`
 - (d) `age<30`
6. Сколько наблюдений с условием `totwrk=0`? Кто эти люди?
7. Есть ли в датасете пропущенные наблюдения? Сколько их?

Упражнение 2. Загрузите датасет **Electricity**.

1. вычислите размер датасета (число наблюдений & число переменных)
2. заполните следующую таблицу со значениями переменных

index	cost	q	pl	pk	pf
1					
15					
48					
87					

3. Вычислите корреляционную матрицу для следующих переменных: cost, q, pl, pk, pf
4. Заполните следующую таблицу

Desc.Stat	cost	q	pl	pk	pf
max					
min					
mean					
median					
st.dev					
var (unbiased)					
var (biased)					
1st quartile					
3rd quartile					

Замечание: 1st/3rd квантили – 25%/75% квантили соответственно.

5. Сколько наблюдения соответствуют следующим условиям
 - (a) $\text{cost} > 40$
 - (b) $q < 5000$
 - (c) $q > 4000$
 - (d) $20 < \text{cost} < 50$
6. Есть ли в датасете пропущенные наблюдения? Сколько их?

Упражнение 3. Загрузите датасет **wage2**.

1. вычислите размер датасета (число наблюдений & число переменных)
2. заполните следующую таблицу со значениями переменных

index	wage	hours	IQ	educ	exper	age
1						
25						
179						
800						

3. Вычислите корреляционную матрицу для следующих переменных: wage, hours, IQ, educ, exper
4. Заполните следующую таблицу

Desc.Stat	wage	hours	IQ	educ	exper	wage
max						
min						
mean						
median						
st.dev						
var (unbiased)						
var (biased)						
1st quartile						
3rd quartile						

Замечание: 1st/3rd квантили – 25%/75% квантили соответственно.

5. Сколько наблюдения соответствуют следующим условиям
 - (a) wage>1000
 - (b) age<40
 - (c) exper>10
 - (d) 100<IQ<130

6. Есть ли в датасете пропущенные наблюдения? Сколько их?

Упражнение 4. Загрузите датасет **Labour**. Создайте новый датасет, содержащий log-переменные из исходного датасета.

Упражнение 5. Загрузите датасет **Electricity**. Создайте новый датасет, содержащий log-переменные из исходного датасета.