Листок 01. Введение в Python

Н.В. Артамонов

4 июня 2024 г.

Содержание

1 Pandas

#**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 матриц 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 наблюдения 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?

#**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 матриц 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) cost>40
   (b) q<5000
   (c) q>4000
   (d) 20<cost<50
6. Есть ли в датасете пропущенные наблюдения? Сколько их?

#**3**. Загрузите датасет `wage2`.

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

index	wage	hour	IQ	educ	exper	age
1						
25						
179						
800						

3. Вычислите корреляционную матриц для следующих переменных: wage, hour, IQ, educ, exper

4. Заполните следующую таблицу

Desc.Stat	wage	hour	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?

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

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

2 Визуализация

#**1**. Загрузите датасет `sleep75`.

1. нарисуйте гистограммы для переменных sleep, totwrk, age, hrwage, educ

2. нарисуйте гистограмму с накопление для sleep относительно male

3. нарисуйте гистограмму с накопление для totwrk относительно south
4. нарисуйте гистограмму с накопление для totwrk относительно smsa
5. нарисуйте диаграмму рассеяния sleep vs totwrk
6. нарисуйте диаграмму рассеяния sleep vs totwrk с группировкой по male
7. нарисуйте диаграмму рассеяния sleep vs age
8. нарисуйте диаграмму рассеяния sleep vs age с группировкой по south
9. нарисуйте диаграмму рассеяния sleep vs edu
10. нарисуйте диаграмму рассеяния sleep vs edu с группировкой по smsa
11. визуализируйте корреляционную матриц для следующих переменных: sleep, totwrk, age

#**2**. Загрузите датасет `Labour`.

1. нарисуйте гистограммы для переменных output, capital, labour, wage
2. нарисуйте гистограммы для log-переменных output, capital, labour, wage
3. нарисуйте диаграммы рассеяния output vs других переменных
4. нарисуйте диаграммы рассеяния log(output) vs log других переменных
5. визуализируйте корреляционную матриц для всех переменных
6. визуализируйте корреляционную матриц для log-переменных

#**3**. Загрузите датасет `Electricity`.

1. нарисуйте гистограммы для переменных cost, q, pf, pk, pl
2. нарисуйте гистограммы для log-переменных cost, q, pf, pk, pl

3. нарисуйте диаграммы рассеяния cost vs других переменных
4. нарисуйте диаграммы рассеяния log(cost) vs log других переменных
5. визуализируйте корреляционную матриц для всех переменных
6. визуализируйте корреляционную матриц для log-переменных

#**4**. Загрузите датасет `diamonds`.

1. нарисуйте гистограммы для переменных price, carat
2. нарисуйте гистограммы для log-переменных price, carat
3. Нарисуйте гистограмму с накопление для price относительно cut
4. Нарисуйте гистограмму с накопление для carat относительно clarity
5. Нарисуйте гистограмму с накопление для log(price) относительно color
6. Нарисуйте гистограмму с накопление для log(carat) относительно color
7. нарисуйте диаграмму рассеяния price vs carat
8. нарисуйте диаграмму рассеяния log-price vs log-carat
9. нарисуйте диаграмму рассеяния log-price vs log-carat с группировкой по cut
10. нарисуйте диаграмму рассеяния log-price vs log-carat с группировкой по color
11. нарисуйте диаграмму рассеяния log-price vs log-carat с группировкой по clarity

#**5**. Загрузите датасет `Diamond`.

1. нарисуйте гистограммы для переменных price, carat
2. нарисуйте гистограммы для log-переменных price, carat
3. Нарисуйте гистограмму с накопление для price относительно certification

4. Нарисуйте гистограмму с накопление для carat относительно clarity
5. Нарисуйте гистограмму с накопление для log(price) относительно colour
6. Нарисуйте гистограмму с накопление для log(carat) относительно colour
7. нарисуйте диаграмму рассеяния price vs carat
8. нарисуйте диаграмму рассеяния log-price vs log-carat
9. нарисуйте диаграмму рассеяния log-price vs log-carat с группировкой по certification
10. нарисуйте диаграмму рассеяния log-price vs log-carat с группировкой по colour
11. нарисуйте диаграмму рассеяния log-price vs log-carat с группировкой по clarity