

Statistics Exercise

Petunjuk

- Soal ini akan menguji pemahaman tentang statistik mulai dari statistik deskriptif, uji hipotesis, dan uji korelasi.
- Gunakan JupyterLab/Jupyter Notebook untuk menuliskan code jawaban soal ini.
- Pada latihan ini, kita akan menggunakan data “**Melbourne Housing Snapshot**” yang dapat diunduh dari laman Kaggle:
<https://www.kaggle.com/dansbecker/melbourne-housing-snapshot>

Persiapan Data

```
import pandas as pd
pd.set_option('display.max_columns', None)
```

```
df = pd.read_csv('melb_data.csv')
df.head()
```

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	Bedroom2	Bathroom	Car	Landsize	BuildingArea	YearBuilt	CouncilArea	Lattit
0	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	3/12/2016	2.5	3067.0	2.0	1.0	1.0	202.0	NaN	NaN	Yarra	-37.7
1	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	4/02/2016	2.5	3067.0	2.0	1.0	0.0	156.0	79.0	1900.0	Yarra	-37.8
2	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	4/03/2017	2.5	3067.0	3.0	2.0	0.0	134.0	150.0	1900.0	Yarra	-37.8
3	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	4/03/2017	2.5	3067.0	3.0	2.0	1.0	94.0	NaN	NaN	Yarra	-37.7
4	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	4/06/2016	2.5	3067.0	3.0	1.0	2.0	120.0	142.0	2014.0	Yarra	-37.8

Nomor 1

- Tampilkan statistik deskriptif pada variabel **Rooms**, **Distance**, **Landsize**, dan **Price**!

	Rooms	Distance	Landsize	Price
count	13580.000000	13580.000000	13580.000000	1.358000e+04
mean	2.937997	10.137776	558.416127	1.075684e+06
std	0.955748	5.868725	3990.669241	6.393107e+05
min	1.000000	0.000000	0.000000	8.500000e+04
25%	2.000000	6.100000	177.000000	6.500000e+05
50%	3.000000	9.200000	440.000000	9.030000e+05
75%	3.000000	13.000000	651.000000	1.330000e+06
max	10.000000	48.100000	433014.000000	9.000000e+06

Nomor 2

- a) Berapa jumlah **Rooms** pada rumah yang harganya paling mahal?
- b) Berapa rata-rata harga rumah yang jarak ke CBD (variabel **Distance**) terdekat?
- c) Berapa rata-rata harga rumah yang luas tanahnya (variabel **Landsize**) 1 kali standar deviasi di atas rata-rata luas tanah?
- d) Berapa rentang (**range**) harga rumah (variabel **Price**) pada region **Southern Metropolitan**?

Nomor 3

Nilai outlier adalah nilai yang berada di bawah **$Q1 - (1.5 * IQR)$** atau di atas **$Q3 + (1.5 * IQR)$** .

- a) Hitung dulu Q1, Q2, dan IQR pada kolom **Price** (harga rumah)
- b) Hitung ada berapa nilai outlier pada kolom **Price**!

Nomor 4

Lakukan uji asumsi normalitas pada variabel harga rumah (**Price**) menggunakan:

- a) **Shapiro Wilk Test**
- b) **D'Agostino and Pearson's Test**

Nomor 5

Jika diasumsikan telah memenuhi syarat statistik parametrik, hitung seberapa kuat korelasi variabel **Rooms** dengan **Price**, serta arah korelasinya!

Nomor 6

- a) Dari beberapa Regionname, manakah **Regionname** yang variabel jumlah kamarnya (**Rooms**) memiliki korelasi terkuat dengan variabel harga rumah (**Price**)!
- b) Diasumsikan telah memenuhi syarat statistik parametrik.

Nomor 7

Jika diasumsikan **tidak** memenuhi syarat statistik parametrik, hitung seberapa kuat korelasi variabel **Distance** dengan **Price**, serta arah korelasinya!

Nomor 8

- a) Dari beberapa Regionname manakah **Regionname** yang variabel jumlah jarak ke CBD (**Distance**) paling tidak berkorelasi dengan variabel harga rumah (**Price**)!
- b) Serta tampilkan juga Regionname yang paling berkorelasi kedua variabel ini!
- c) Diasumsikan telah memenuhi syarat statistik parametrik.

Nomor 9

Uji hipotesis apakah metode pembelian/**Method** berkaitan dengan tipe rumah/ **Type** (dependen) atau tidak saling berkaitan (independen)!

Hypothesis:

- H_0 : The two variables are independent
- H_a : The two variables are dependent.

Nomor 10

Uji hipotesis apakah rata-rata harga rumah (**Price**) antar Regionname berbeda ataukah sama! Diasumsikan persyaratan statistik parametrik terpenuhi

Hipotesis:

- $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ (Rata-rata harga rumah antar Regionname sama)
- H_a : Setidaknya ada satu pasang Regionname yang rata-rata harga rumahnya tidak sama, tapi kita tidak tahu yang mana