Pacmann

Insurance

Probability Course - Sekolah Data Pacmann

By Eric Kristanto

Outline

- Introduction
- Dataset
- Descriptive Statistic Analysis
- Categorical Variables Analysis
- Continuous Variables Analysis
- Variables Correlation
- Hypothesis Testing
- Conclusion

Introduction

Introduction

Pribadi yang menjadi peserta asuransi kesehatan wajib membayar premi.

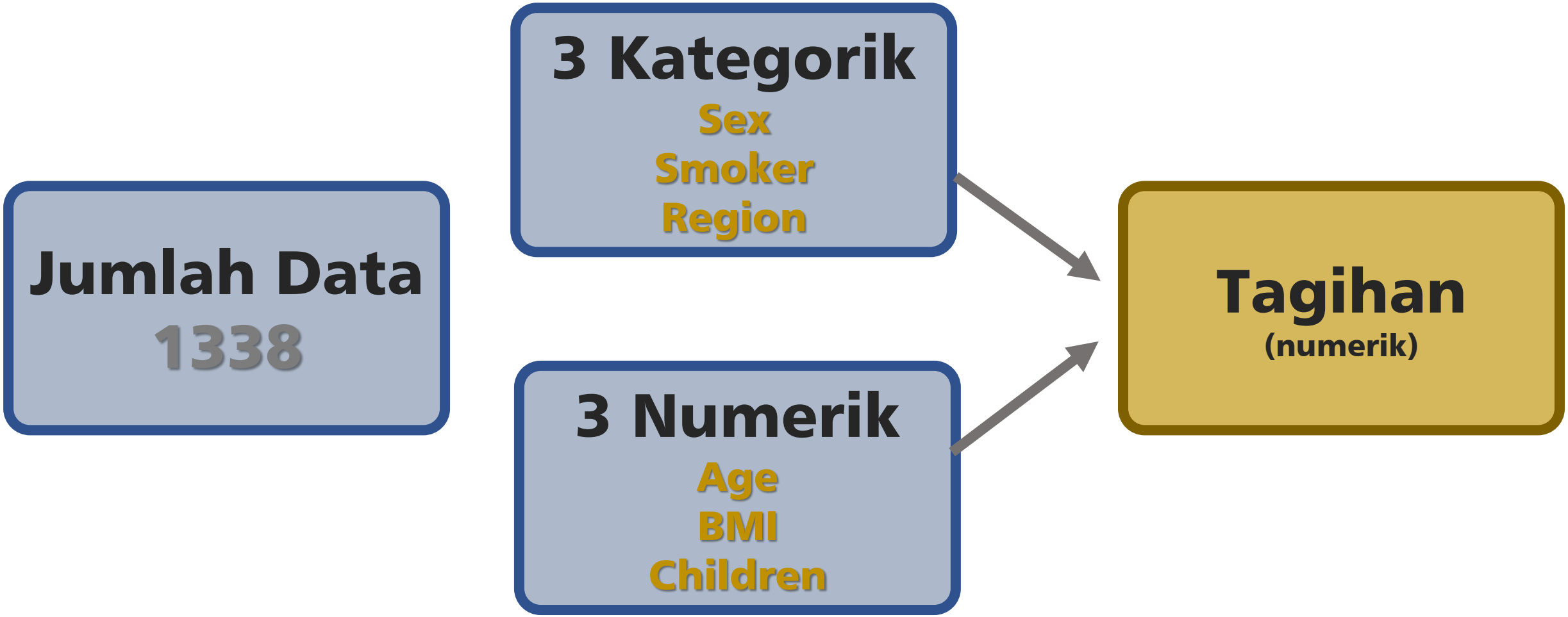
Premi akan digunakan perusahaan asuransi untuk membayar tagihan kesehatan yang diklaim pengguna

Beberapa hal yang mungkin mempengaruhi premi adalah :

- **Usia**
- **BMI**
- **Jenis Kelamin**
- **Banyak Anak**
- **Aktifitas Merokok**
- **Wilayah / Tempat tinggal**

Dataset

Dataset



Descriptive Statistics Analysis

Rataan Umur

Data Umur

sample 1338

Min = 18 tahun
Max = 64 tahun

Rataan	Median
39,21	39

- Berdasarkan hasil rataan yang hampir sama dengan median, bisa dikatakan data berdasarkan umurini simetris

Rataan Umur

Rataan Umur Perokok "Laki-laki"

38.45

Rataan Umur Perokok "Perempuan"

38.61

Rataan Umur Perokok "Laki-laki" hampir **sama dengan** dari Rataan Umur Perokok "Perempuan"

Rataan BMI

Rataan BMI Total
30,66

Rataan BMI Perokok
30,71

Rataan BMI Non Perokok
30,65

- Berdasarkan hasil rataan BMI total, Perokok, dan Non Perokok hasil nya hampir sama. bisa dikatakan BMI dari Perokok dan Non Perokok adalah seimbang

Varian dari Tagihan Premi

Varian Tagihan Premi "PEROKOK"

133.207.311,21

Varian Tagihan Premi "NON PEROKOK"

35.925.420,50

Varian merupakan variasi nilai terhadap rataannya.

Varian Tagihan Premi "Perokok" **lebih besar** dari Varian Tagihan Premi "Non Perokok"

nb:
Jika ingin tahu makna nilai sebaran
Dari rata-ratanya lebih baik menggunakan standar deviasi

Rataan dari Tagihan Premi

Rataan Tagihan Premi "PEROKOK"

32.020,23

Rataan Tagihan Premi "Perokok" **lebih besar** dari Rataan Tagihan Premi "Non Perokok"

Rataan Tagihan Premi "NON PEROKOK"

8.434,27

Ratan Tagihan Premi "Perokok" hampir **4x** lebih besar daripada "Non Perokok"

Rataan dari Tagihan Premi

Rataan Tagihan Premi BMI>25 + PEROKOK

35.116,91

Rataan Tagihan Premi BMI>25 + NON PEROKOK

8.629,59

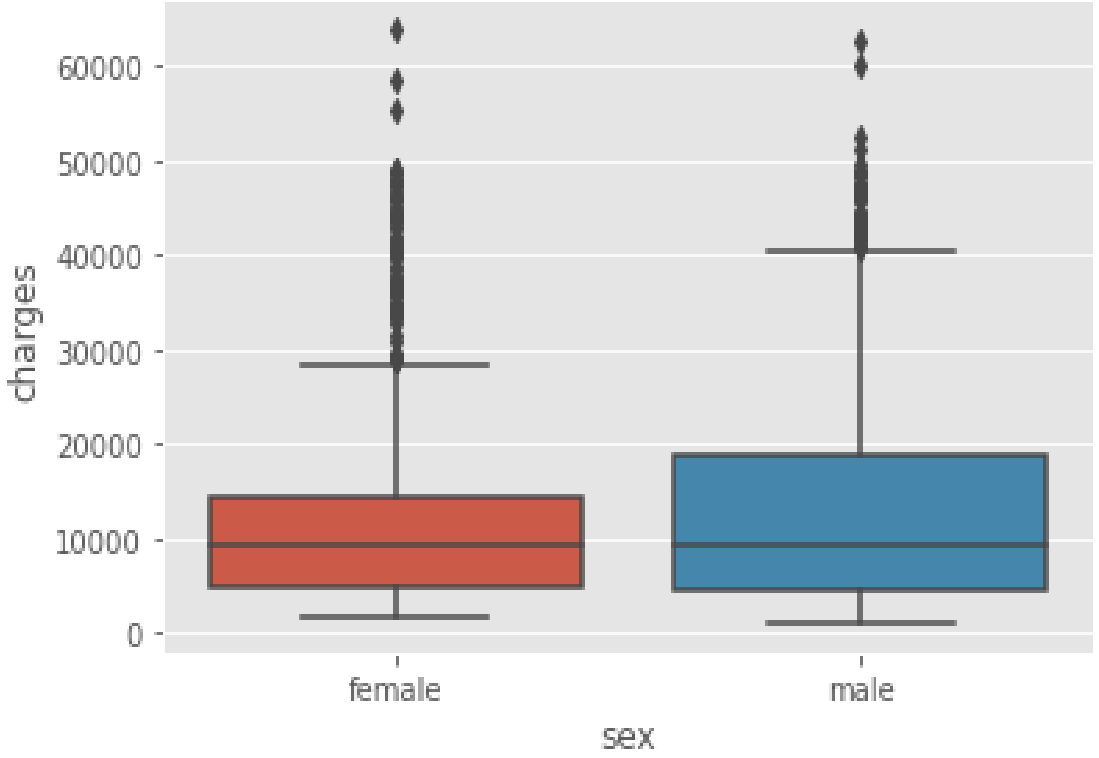
1.
Jika dibandingkan dengan slide sebelumnya, walaupun di tambah kriteria BMI>25,

Rataan premi “Perokok” hampir **4x** lebih besar daripada Non Perokok

2.
Jika dilihat lebih detail lagi, Rataan premi “Perokok” maupun “Non Perokok” sama sama meningkat ketika di beri kriteria tambahan BMI>25.
Dengan itu kita kita bisa berasumsi bahwa semakin besar BMI maka ikut sedikit menaikan premi

Categorical Variables Analysis

Gender apa dengan premi lebih tinggi ?



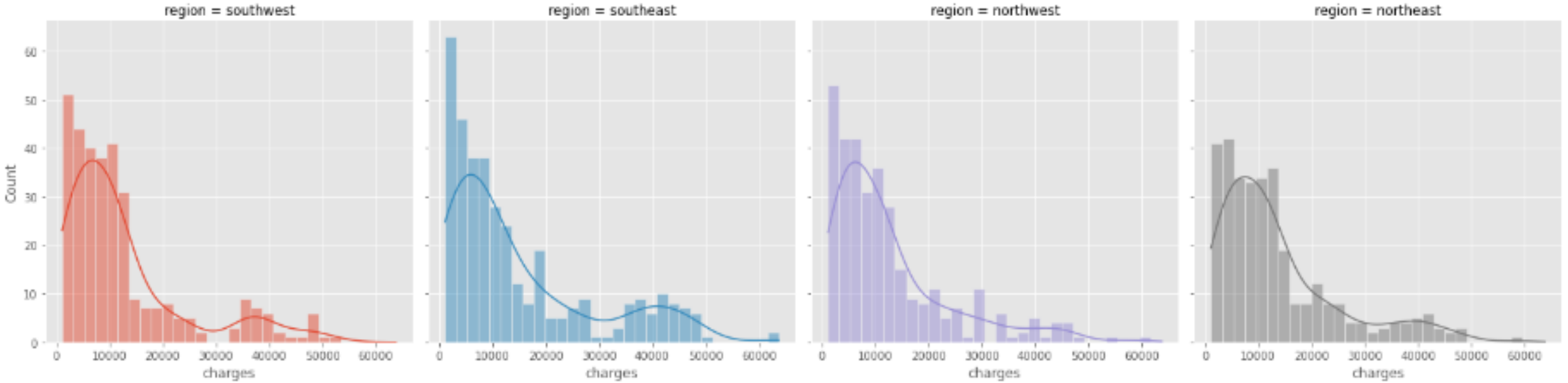
Rataan premi laki-laki : 13.956,75

Rataan premi perempuan : 12.569,58

Rataan Tagihan Premi Laki-laki **lebih besar** dari perempuan.
Namun kalau dari rentang premi dari 0-68.000, perbedaan rataan tersebut tidak terlalu signifikan. (terlihat juga pada boxplot)

Berdasarkan boxplot, distribusi premi laki-laki cenderung ke kanan atau membayar lebih tinggi

Premi di tiap region



Di semua wilayah nominal tagihan terbanyak di kisaran 0-16.000 (left skewness)

Southeast mempunyai lebih banyak orang tagihannya mendekati 0, lalu diikuti oleh region nortwest.

Southeast dan Southwest juga banyak orang yang mendapatkan tagihan di sekitaran 38.000-50.000 jika dibandingkan dengan Nortwest dan Northeast.

Proporsi orang tiap region

region	
northeast	324
northwest	325
southeast	364
southwest	325

Berdasarkan data diatas, region northeast, northwest, dan southwest mempunyai proporsi banyak orang yang hampir sama.

Namun southeast mempunyai data yang lebih besar dengan jarak 20.

Karena ada perbedaan data tersebut lebih baik membandingkan antar region berdasarkan proporsi atau peluangnya dibandingkan dengan menggunakan jumlah data

Proporsi orang berdasarkan aktivitas merokok

smoker	
no	1064
yes	274

Berdasarkan data diatas, jumlah non perokok jauh lebih tinggi dari jumlah perokok (mendekati 4x lipat).

Untuk itu kita tidak boleh membandingkan keduanya dengan menggunakan jumlah/banyak, namun lebih disarankan menggunakan proporsi atau peluangnya

Peluang laki-laki/perempuan jika diketahui merokok

smoker	sex	
no	female	547
	male	517
yes	female	115
	male	159

P(Perempuan|Rokok)

= n (perempuan dan rokok) / n (rokok)
= 115 / (115+159)
= 0.420

P(Laki-laki|Rokok)

= n (laki-laki dan rokok) / n (rokok)
= 159 / (115+159)
= 0.580

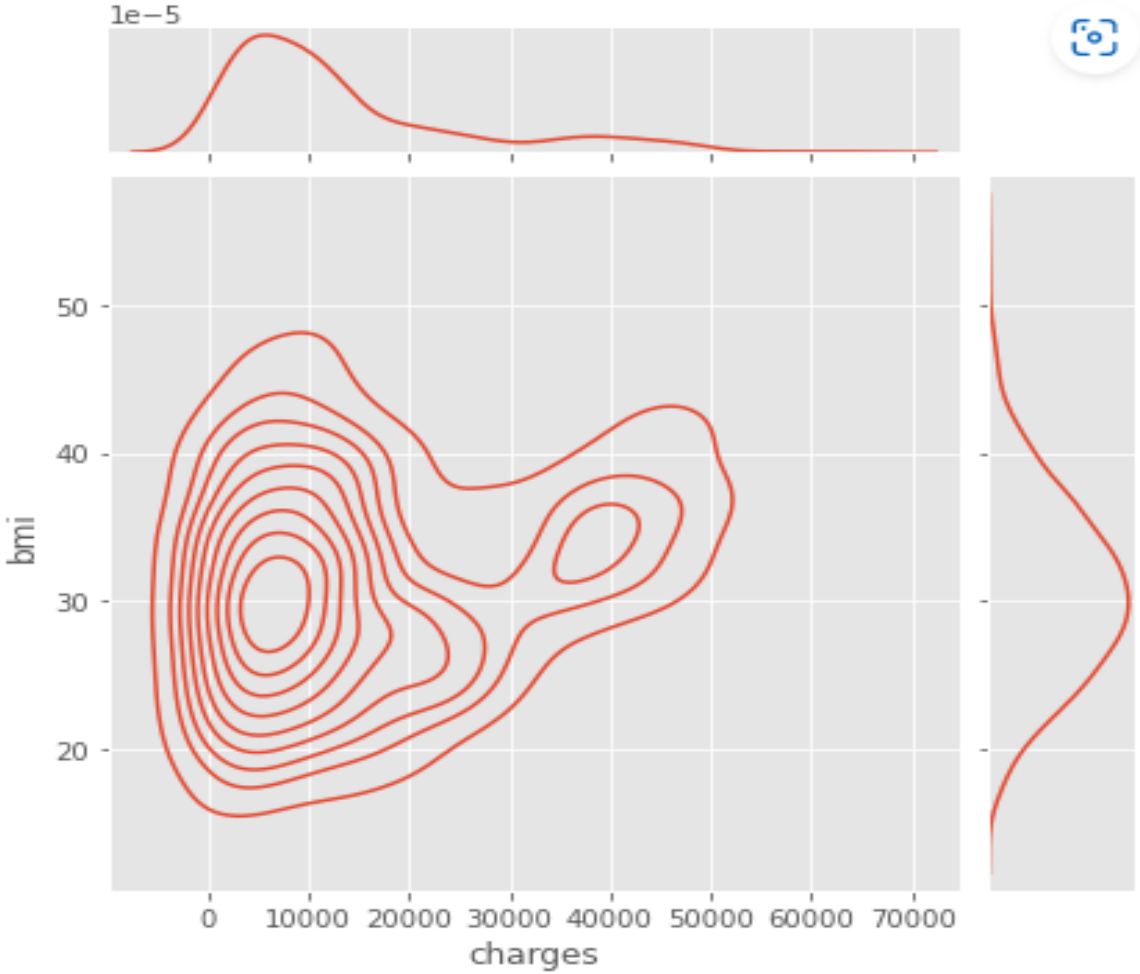
Berdasarkan slide sebelumnya,

Jika kita membandingkan dua nilai yang berbeda jauh, gunakanlah peluang atau distribusi.

Hal itu lebih adil apabila kita menjumlahkan berdasarkan nilai yang sesungguhnya

Continuous Variables Analysis

Peluang jumlah tagihan berdasarkan BMI



Bedasarkan plot di atas plot joint probability samping

Peluang paling besar terjadi saat bmi berada disekitar 30 dan charges berada disekitar 80000

Peluang bersyarat

Peluang tagihan lebih dari 16.700 jika diketahui BMI>25

P (Tagihan>16.700 | BMI>25)

= n (tagihan>16.700 dan BMI>25)
/ n (BMI>25)
= 283 / 1091
= 0.259

Peluang tagihan lebih dari 16.700 jika diketahui Perokok

P (Tagihan>16.700 | Perokok)

= n (tagihan>16.700 dan Perokok)
/ n (Perokok)
= 254 / 274
= 0.927

P (Tagihan>16.700 | Perokok) > P (Tagihan>16.700 | BMI>25)

KESIMPULAN

Kita akan lebih mudah mencari orang dengan tagihan lebih dari 16.700 diantara orang merokok dari pada mencarinya diantara orang dengan BMI lebih dari 25

Peluang bersyarat

Peluang tagihan lebih dari 16.700 jika diketahui BMI>25

P (Tagihan>16.700 | BMI>25)

= n (tagihan>16.700 dan BMI>25)
/ n (BMI>25)
= 283 / 1091
= 0.259

Peluang tagihan lebih dari 16.700 jika diketahui BMI<25

P (Tagihan>16.700 | BMI<25)

= n (tagihan>16.700 dan BMI<25)
/ n (BMI<25)
= 51 / 245
= 0.208

P (Tagihan>16.700 | BMI>25) > P (Tagihan>16.700 | BMI<25)

KESIMPULAN

Kita akan lebih mudah mencari orang dengan tagihan lebih dari 16.700 diantara orang BMI>25 dari pada mencarinya diantara orang dengan BMI < 25

Peluang bersyarat

Peluang tagihan lebih dari 16.700 jika diketahui BMI>25 dan perokok

P (Tagihan>16.700 | BMI>25, perokok)

= n (tagihan>16.700 dan BMI>25 dan perokok)
/ n (BMI>25 dan perokok)
= 215 / 219
= 0.982

Peluang tagihan lebih dari 16.700 jika diketahui BMI>25 dan non perokok

P (Tagihan>16.700 | BMI>25, non peroko)

= n (tagihan>16.700 dan BMI>25 dan non perokok)
/ n (BMI>25 dan non perokok)
= 68 / 872
= 0.078

P (Tagihan>16.700 | BMI>25,perokok) > P (Tagihan>16.700 | BMI>25,non perokok)

KESIMPULAN

Kita akan lebih mudah mencari orang dengan tagihan lebih dari 16.700 diantara orang BMI>25 dan perokok dari pada mencarinya diantara orang dengan BMI < 25 dan non perokok

Variables Correlation

Correlation

Korelasi tagihan dan bmi

```
file[['charges','bmi']].corr()
```

	charges	bmi
charges	1.000000	0.198341
bmi	0.198341	1.000000

Nilai 0.198 menggambarkan bahwa adanya hubungan antara tagihan dan BMI walaupun hubungannya rendah

Nilai positif menunjukkan bahwa semakin besar BMI maka tagihan juga semakin besar

Korelasi tagihan dan umur

```
file[['charges','age']].corr()
```

	charges	age
charges	1.000000	0.299008
age	0.299008	1.000000

Nilai 0.299 menggambarkan bahwa adanya hubungan antara tagihan dan umur walaupun hubungannya cukup rendah

Nilai positif menunjukkan bahwa semakin tua umur maka tagihan juga semakin besar

Korelasi tagihan dan aktivitas rokok

```
file['smoker'] = file['smoker'].astype('category')
file['smoker'].cat.categories = [0,1]
file['smoker'] = file['smoker'].astype('float')
file[['charges','smoker']].corr()
```

	charges	smoker
charges	1.000000	0.787251
smoker	0.787251	1.000000

Nilai 0.787 menggambarkan bahwa adanya hubungan antara tagihan dan aktivitas rokok dengan hubungan yang tinggi

Nilai positif menunjukkan bahwa semakin orang tersebut merokok maka tagihan juga semakin besar

Correlation

Korelasi tagihan dan jenis kelamin

	charges	sex
charges	1.000000	0.057292
sex	0.057292	1.000000

Korelasi tagihan dan banyak anak

	charges	children
charges	1.000000	0.067998
children	0.067998	1.000000

Nilai 0.057 dan 0.068 yang **mendekati 0** menggambarkan bahwa **tidak adanya hubungan** antara tagihan dan jenis kelamin serta antara tagihan dan banyak anak

Hypothesis Testing

Uji Hipotesis :

Tagihan kesehatan perokok lebih tinggi daripada tagihan kesehatan non perokok

- Alpha = 0.05
- Uji t test karena mau menguji rataan dengan SD populasi tidak diketahui
- upper tail test/ uji pihak kanan, karena hipotesis sebagai berikut:
- $H_o : \mu_{rokok} = \mu_{non\ rokok}$
  $H_1 : \mu_{rokok} > \mu_{non\ rokok}$

Hasil uji:

```
Statistics = 32.7519, p-value = 0.0000000000
```

Kesimpulan :

Maka tolak Ho, sehingga cukup bukti untuk mengatakan bahwa
Tagihan premi perokok lebih tinggi dari pada tagihan premi non perokok

Uji Hipotesis :

Tagihan premi dengan BMI>25 lebih tinggi daripada tagihan premi dengan BMI<25

- Alpha = 0.05
- Uji t test karena mau menguji rataan dengan SD populasi tidak diketahui
- upper tail test/ uji pihak kanan, karena hipotesis sebagai berikut:
- $H_o : \mu_{bmi>25} = \mu_{bmi<25}$
  $H_1 : \mu_{bmi>25} > \mu_{bmi<25}$

Hasil uji:

```
Statistic = 5.9299, p-value = 0.0000000025
```

Kesimpulan :

Maka tolak Ho, sehingga cukup bukti untuk mengatakan bahwa
Tagihan premi pada orang bmi>25 lebih tinggi dari pada tagihan premi pada orang bmi>25

Uji Hipotesis :

BMI laki-laki dan perempuan sama

- Alpha = 0.05
- Uji t test karena mau menguji rataan dengan SD populasi tidak diketahui
- two tail test/ uji dua pihak, karena hipotesis sebagai berikut:
- $H_o : \mu_L = \mu_P$
  $H_1 : \mu_L \neq \mu_P$

Hasil uji:

```
Statistics = 1.697028 , p-value = 0.089924
```

Kesimpulan :

Maka gagal tolak Ho, sehingga cukup bukti untuk mengatakan bahwa BMI laki laki = BMI perempuan

Conclusion

Conclusion

- Dari Uji Hypotesis,
  - Tagihan premi perokok **≥** tagihan premi non perokok
  - Tagihan premi orang dengan BMI di atas 25 **≥** tagihan premi orang dengan BMI di atas 25
  - BMI laki-laki **=** BMI perempuan

- Dari Korelasi,
  - ada hubungan yang **tinggi** antara tagihan dan aktivitas merokok
  - ada hubungan yang **rendah** antara tagihan dan BMI serta antara tagihan dan umur
  - tidak ada hubungan antara tagihan dan jenis kelamin serta antara tagihan dan banyak anak