



Probability Project Insurance

Charles Sugianto

1 Descriptive Statistics Analysis

1 Rata-rata Usia

- Hasil perhitungan Excel, menunjukkan bahwa dari 1338 data pengguna asuransi yang berusia 18-64 tahun memiliki rata-rata usia 39,21 tahun

2 Rata-rata BMI

- Hasil perhitungan Excel, menunjukkan bahwa dari 1338 data pengguna asuransi, memiliki rata-rata BMI sebagai berikut:
 - Rata-rata BMI perokok = 30,71
 - Rata-rata BMI non perokok = 30,65

1 Descriptive Statistics Analysis

3 Rata-rata tagihan

- Hasil perhitungan Excel, menunjukkan bahwa dari 1338 data pengguna asuransi:
 - Rata-rata tagihan perokok = 32.050,23
 - Rata-rata tagihan non perokok = 8.434,27

4 Variansi dan standar deviasi

- Variansi tagihan kesehatan perokok = 133.207.311,2
- Variansi tagihan kesehatan non perokok = 35.925.420,5
- Standar deviasi tagihan kesehatan perokok = 11.541,55
- Standar deviasi tagihan kesehatan non perokok = 5.993,78

1 Descriptive Statistics Analysis

5 Rata-rata tagihan BMI > 25

- Hasil perhitungan Excel, menunjukkan bahwa dari 1338 data pengguna asuransi, memiliki:
 - Rata-rata tagihan perokok BMI>25 = 35.116,91
 - Rata-rata tagihan non perokok BMI>25 = 8.629,59

1 Descriptive Statistics Analysis

Analysis

Dengan data dari 1338 orang yang memiliki rata rata usia 39,21 tahun, artinya mayoritas adalah orang dewasa. Rata-rata nilai BMI baik untuk perokok maupun non perokok 30,71 dan 30,65, artinya rata-rata orang memiliki obesitas, karena normal BMI adalah 18,5-25.

Rata-rata tagihan untuk perokok yaitu $32.050,23 \pm 11.541,55$ lebih besar dari rata-rata tagihan untuk non perokok yaitu $8.434,27 \pm 5.993,78$. Jika menggunakan indikator $BMI > 25$, rata-rata tagihan perokok lebih besar yaitu 35.116,91 dari non perokok yaitu 8.629,59. Jadi, merokok membuat tagihan kesehatan lebih besar.

2 Discrete Variable Analysis

1 Tagihan Berdasarkan Gender

- Hasil perhitungan Excel, menunjukkan bahwa dari 1338 data memiliki
 - Rata-rata tagihan perempuan = 12.569,58
 - Rata-rata tagihan laki-laki = 13.956,75
 - Tagihan tertinggi perempuan = 63.770,43
 - Tagihan tertinggi laki-laki = 62.592,87

2 Discrete Variable Analysis

2 Proposi

- Hasil perhitungan Excel, menunjukkan bahwa
 - Proposi data banyak orang terbanyak berada di Wilayah Southeast.

Wilayah	Banyak Orang	Proporsi
Northeast	324	0,242
Northwest	325	0,243
Southeast	364	0,272
Southwest	325	0,243

2 Discrete Variable Analysis

3 Proposi

- Hasil perhitungan Excel, menunjukkan bahwa
 - Proposi non perokok lebih banyak dari pada perokok.

	Banyak Orang	Proporsi
Perokok	274	0,20
Non Perokok	1064	0,80

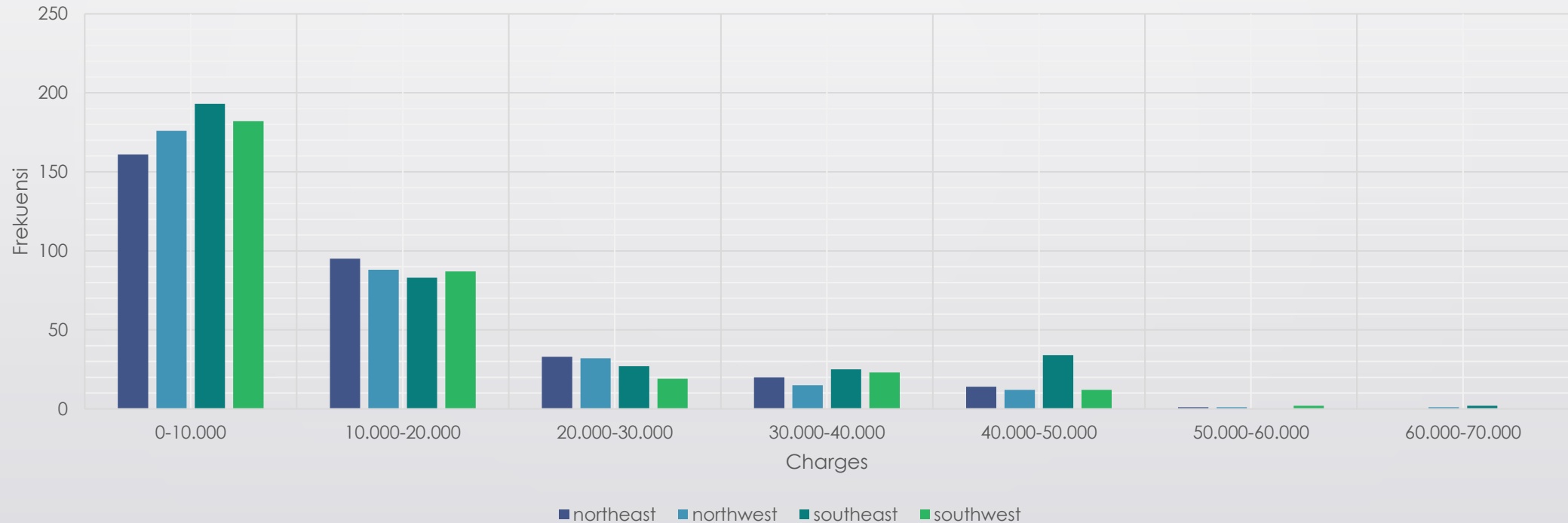
4 Peluang Jenis Kelamin Perokok

- Hasil perhitungan Excel, menunjukkan bahwa peluang laki-laki perokok lebih tinggi daripada perempuan perokok
 - $n(\text{Perempuan} \cap \text{Perokok}) = 115$
 - $n(\text{Perokok}) = 274$
 - $P(\text{Perempuan} \mid \text{Perokok}) = 115/274 = 0,42 = 42\%$
 - $P(\text{Laki-laki} \mid \text{Perokok}) = 1 - 0,42 = 0,58 = 58\%$

2 Discrete Variable Analysis

5 Distribusi Tagihan per Wilayah

- Berikut adalah hasil distribusi tagihan tiap wilayah dengan satuan data tagihan dibagi per 10.000.



2 Discrete Variable Analysis

Analysis

- Dengan data 1338 orang didapatkan suatu informasi menarik, yaitu, **tagihan laki-laki lebih sebesar 13.959,75 lebih tinggi dari pada tagihan perempuan dengan nilai 12.569,58**. Selain itu tagihan memiliki distribusi eksponensial dengan karakteristik sebagian data memiliki tagihan yang kecil.
- Proporsi non perokok memiliki nilai 80%, artinya sebagian orang termasuk dalam kategori non perokok, dan diantara perokok tersebut peluang seorang perokok laki-laki lebih besar dibandingkan perokok perempuan. Wilayah Southeast memiliki proporsi paling besar dibandingkan wilayah lainnya.

3 Continuous Variables Analysis

1 Peluang Tagihan < 16.700 Jika BMI < 25

- Hasil perhitungan excel bantuan Excel, dari 1338 data didapatkan
 - $n(\text{Charges} < 16.700 \cap \text{BMI} < 25) = 194$
 - $n(\text{BMI} < 25) = 245$
 - $P(\text{Charges} < 16.700 \mid \text{BMI} < 25) = 0,79$
 - $P(\text{Charges} > 16.700 \mid \text{BMI} < 25) = 0,21$

3 Continuous Variables Analysis

2 Peluang Tagihan > 16.700 Jika Perokok dg BMI > 25

- Hasil perhitungan excel bantuan Excel, dari 1338 data didapatkan
 - $n(\text{Charges} > 16.700 \cap \text{BMI} > 25 \ \& \ \text{Perokok}) = 215$
 - $n(\text{BMI} > 25) = 1091$
 - $P(\text{Charges} > 16.700 \mid \text{BMI} > 25 \ \& \ \text{Perokok}) = 0,98$
 - $P(\text{Charges} < 16.700 \mid \text{BMI} > 25 \ \& \ \text{Perokok}) = 0,02$

3 Continuous Variables Analysis

3 Peluang Tagihan > 16.700 Jika Perokok

- Hasil perhitungan excel bantuan Excel, dari 1338 data didapatkan
 - $n(\text{Charges} > 16.700 \cap \text{Perokok}) = 254$
 - $n(\text{Perokok}) = 274$
 - $P(\text{Charges} > 16.700 \mid \text{Perokok}) = 0,93$
 - $P(\text{Charges} < 16.700 \mid \text{Perokok}) = 0,07$

3 Continuous Variables Analysis

4 Peluang BMI > 25 Jika Tagihan > 16.700

- Hasil perhitungan excel bantuan Excel, dari 1338 data didapatkan
 - $n(\text{BMI} > 25 \cap \text{Charges} > 16.700) = 283$
 - $n(\text{Charges} > 16.700) = 334$
 - $P(\text{BMI} > 25 \mid \text{Charges} > 16.700) = 0,85$
 - $P(\text{BMI} < 25 \mid \text{Charges} > 16.700) = 0,15$

3 Continuous Variables Analysis

5 Peluang BMI > 25 Jika Tagihan >16.700

- Hasil perhitungan excel bantuan Excel, dari 1338 data didapatkan
 - $n (\text{Perokok} \cap \text{BMI} > 25 \cap \text{Charges} > 16.700) = 215$
 - $n (\text{BMI} > 25 \cap \text{Charges} > 16.700) = 283$
 - $P (\text{Perokok} \mid \text{BMI} > 25 \ \& \ \text{Charges} > 16.700) = 0,76$
 - $P (\text{Non Perokok} \mid \text{BMI} > 25 \ \& \ \text{Charges} > 16.700) = 0,24$

3 Continuous Variables Analysis

Analysis

- Jika seseorang dengan BMI<25 mendapatkan tagihan <16.700 memiliki peluang yang lebih tinggi dibandingkan dengan BMI < 25 dan tagihan >16.700, yaitu setara 79%. Seseorang perokok dengan BMI>25 mendapatkan tagihan >16.700 memiliki peluang yang lebih tinggi daripada <16.700 yaitu setara 98%.
- Namun peluang seseorang mendapatkan tagihan >16.700 dan memiliki BMI>25 lebih tinggi dari pada BMI<25, yaitu setara 85%. **Jadi, semakin besar BMI, maka semakin tinggi pula orang tersebut mendapatkan tagihan kesehatan yang besar.**
- Jadi, seorang perokok mendapatkan tagihan >16.700 lebih besar daripada tagihan <16.700, yaitu setara 93%. Sama halnya, dengan peluang seseorang perokok memiliki BMI>25 dan tagihan >16.700 lebih tinggi dibandingkan non perokok, yaitu setara 76%. **Jadi, seorang perokok memiliki peluang tagihan kesehatan yang lebih besar dari pada non perokok.**

4 Variables Correlation

Korelasi

- Koefisien korelasi usia dengan BMI = 0,11
- Koefisien korelasi BMI dengan tagihan = 0,20
- Koefisien korelasi usia dengan tagihan = 0,30
- Koefisien korelasi usia dengan jumlah anak ditanggung = 0,04
- Koefisien korelasi jumlah anak ditanggung dengan tagihan = 0,07

Kesimpulan:

Jadi, jumlah anak yang ditanggung tidak berkorelasi dengan tagihan. Selain itu, usia tidak berkorelasi dengan jumlah anak yang ditanggung maupun BMI. Korelasi antara usia dengan tagihan lebih tinggi daripada korelasi antara BMI dengan tagihan.

5 Hypothesis Testing

Uji 1: Tagihan Perokok Lebih Besar dari Tagihan Non Perokok $\alpha=0,05$

Uji untuk mengetahui variansi antara kedua populasi.

- $H_0: \sigma_a = \sigma_b$
 - $H_1: \sigma_a \neq \sigma_b$
 - $F = 3,71 > F_{crit} = 1,17$
- Tolak $H_0 = F > 1,17$ (F=3,71)

$$T = \frac{X_1 - X_2}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$
$$df = \frac{[(S_1^2/n_1) + (S_2^2/n_2)]^2}{[(S_1^2/n_1)^2 / (n_1-1)] + [(S_2^2/n_2)^2 / (n_2-1)]}$$

Uji T, karena standar deviasi populasi tidak diketahui.

- $H_0: \mu_a \leq \mu_b$
- $H_1: \mu_a > \mu_b$
- $df = 311,85$
- $t = 32,75 > t_{crit} = 1,65$
- Tolak $H_0 = t > 1,65$ (t=32,75)

Jadi, tagihan perokok lebih besar dari tagihan non perokok

Rumus perhitungan Excel terlampir di medium & github

5 Hypothesis Testing

Uji 2: Tagihan BMI>25 Lebih Besar dari Tagihan BMI<25 $\alpha=0,05$

Uji untuk mengetahui variansi antara kedua populasi.

- $H_0: \sigma_a = \sigma_b$
- $H_1: \sigma_a \neq \sigma_b$
- $F = 2,91 > F_{crit} = 1,19$
- Tolak $H_0 = F > 1,19$ ($F=2,91$).

Uji T, karena standar deviasi populasi tidak diketahui.

- $H_0: \mu_a \leq \mu_b$
- $H_1: \mu_a > \mu_b$
- $df = 311,85$
- $t = 4,31 > t_{crit} = 1,65$
- Tolak $H_0 = t > 1,65$ ($t=4,31$)

Jadi, tagihan BMI>25 lebih besar dari tagihan BMI<25.

Rumus perhitungan Excel terlampir di medium & github

$$T = \frac{X_1 - X_2}{\sqrt{(S_1^2/n1) + (S_2^2/n2)}}$$
$$df = \frac{[(S_1^2/n1) + (S_2^2/n2)]^2}{[(S_1^2/n1)^2 / (n1-1)] + [(S_2^2/n2)^2 / (n2-1)]}$$

5 Hypothesis Testing

Uji 3: Tagihan Laki-laki Lebih Besar dari Tagihan Perempuan $\alpha=0,05$

Uji untuk mengetahui variansi antara kedua populasi.

- $H_0: \sigma_a = \sigma_b$
- $H_1: \sigma_a \neq \sigma_b$
- $F = 1,36 < F_{crit} = 1,14$
- Tolak $H_0 = F > 1,14$ ($F=0,74$)

Uji T, karena standar deviasi populasi tidak diketahui.

- $H_0: \mu_a \leq \mu_b$
- $H_1: \mu_a > \mu_b$
- $df = 1336$
- $t = 2,1 > t_{crit} = 1,65$
- Tolak $H_0 = t > 1,65$ ($t=2,1$)

Jadi, tagihan laki-laki lebih besar dari perempuan

$$T = \frac{X_1 - X_2}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$
$$df = \frac{[(S_1^2/n_1) + (S_2^2/n_2)]^2}{[(S_1^2/n_1)^2 / (n_1-1)] + [(S_2^2/n_2)^2 / (n_2-1)]}$$

Rumus perhitungan Excel terlampir di medium & github



Conclusion

- Semakin tua umur seseorang, maka semakin besar pula tagihannya.
- Jenis kelamin, usia, BMI, dan perokok/non perokok memiliki pengaruh terhadap besarnya tagihan kesehatan.
- Semakin besar BMI, maka semakin besar pula tagihannya.
- Tagihan perokok akan cenderung memiliki tagihan lebih besar daripada tagihan non perokok.
- Tagihan laki-laki akan cenderung memiliki tagihan lebih besar daripada perempuan.



References

- Bertsekas, D. and Tsitsiklis, J.N., 2008. *Introduction to probability* (Vol. 1). Athena Scientific.
- Hayter, A.J., 2012. *Probability and statistics for engineers and scientists*. Cengage Learning.
- RI, K., 2019. Kategori Batas Ambang Indeks Massa Tubuh (IMT) untuk Indonesia.
- Stein, C., 1945. A two-sample test for a linear hypothesis whose power is independent of the variance. *The Annals of Mathematical Statistics*, 16(3), pp.243-258.