

## Analisis statistik: data “synthetic asthma dataset”

### **B. Pertanyaan Analisis Kompleks** (dari hasil hipotesis yang di dapatkan di modul 4)

- 1) H1 = Pasien dengan riwayat keluarga asma memiliki risiko lebih tinggi menderita asma dibandingkan pasien tanpa riwayat keluarga.
- 2) H2 = Status merokok memengaruhi tingkat kontrol asma, tetapi kesimpulan yang lebih valid memerlukan data kontrol asma yang lebih lengkap.
- 3) H3 = Rendahnya jumlah pasien dengan asma terkontrol baik menunjukkan bahwa intervensi medis dan edukasi pasien masih perlu ditingkatkan agar kepatuhan pengobatan lebih optimal.

### **C. Pemilihan & Argumentasi Uji Statisik (Nama & Penjelasan)**

#### **Pertanyaan 1 (Hipotesis 1)**

- Nama uji statistik : Uji Chi-Square of Independence ( $\chi^2$ )
- Penjelasan: Uji Chi-Square digunakan untuk mengetahui apakah terdapat hubungan atau ketergantungan antara dua variabel kategorikal. Dalam hipotesis ini, kedua variabel yaitu Family\_History (ada/tidak) dan Has\_Asthma (ya/tidak) merupakan variabel kategorikal. Dengan demikian, uji Chi-Square merupakan metode yang tepat untuk menguji apakah riwayat keluarga asma berhubungan secara signifikan dengan kemungkinan seseorang menderita asma.  
Secara teoretis, menurut Field (2013) dan Agresti (2019), **Chi-Square Test of Independence** digunakan untuk menguji hubungan dua variabel nominal atau ordinal tanpa asumsi distribusi normal. maka dari itu uji ini cocok ketika kita ingin melihat **keterkaitan antar kategori** dalam populasi besar.

#### **Pertanyaan 2 (Hipotesis 2)**

- Pada hipotesis ini, variabel yang terlibat adalah:
  - Smoking\_Status → kategorikal (Never, Former, Current)
  - Asthma\_Control\_Level → kategorikal (Well Controlled, Partly Controlled, Poorly Controlled, Unknown)
- Hipotesis kedua tidak dilakukan uji statistik karena variabel Asthma\_Control\_Level memiliki proporsi nilai Unknown yang sangat tinggi sehingga distribusi datanya tidak seimbang. Kondisi ini menyebabkan pelanggaran asumsi uji Chi-Square, yaitu frekuensi harapan minimal 5 di setiap sel tabel (Field, 2013; McHugh, 2013).

Oleh karena itu, hipotesis ini hanya dianalisis secara deskriptif melalui visualisasi data untuk menggambarkan pola hubungan antara status merokok dan tingkat kontrol asma tanpa menarik kesimpulan inferensial.

#### D. Pemaparan hasil, interpretasi dan solusi pengecekan uji asumsi

##### Analisis pertanyaan 1:

Rumusan Hipotesis:

1.  $H_0$  : Tidak ada hubungan antara riwayat keluarga dengan kemungkinan menderita asma.
2.  $H_1$  : Ada hubungan antara riwayat keluarga dengan kemungkinan menderita asma.

Metode Uji Statistik: Chi-Square Test of Independence

```
Tabel Kontingensi:
Has_Asthma      0      1
Family_History
0              6079   887
1              1488  1546

Chi-Square Statistic: 1675.0775786178124
p-value: 0.0
Degrees of Freedom: 1

Expected Frequencies:
Has_Asthma      0      1
Family_History
0              5271.1722 1694.8278
1              2295.8278  738.1722

Hasil: Tolak  $H_0 \rightarrow$  Ada hubungan signifikan antara riwayat keluarga dan asma.
```

Interpretasi Hasil:

- Nilai Chi-square = 1675.08 dengan p-value = 0.0 (lebih kecil dari 0.05) maka bisa dikatakan menunjukkan adanya hubungan yang signifikan antara Family\_history dan Has\_Asthma.
- Jadi yang berarti pasien dengan riwayat keluarga asma memiliki risiko lebih tinggi untuk menderita asma di bandingkan pasien tanpa riwayat keluarga.

Cek asumsi

- Data kategori  $\rightarrow$  sudah sesuai dengan syarat ketentuan menggunakan Chi-Square
- Frekwensi harapan (expected frequencies) semua  $> 5 \rightarrow$  maka dianggap asumsi terpenuhi
- Tidak ada sel dengan nilai 0 atau sangat kecil  $\rightarrow$  jadi aman untuk uji Chi-Square

##### Analisis Pertanyaan 3:

Rumusan Hipotesis:

1.  $H_0$  (Hipotesis Nol): Tidak ada perbedaan tingkat kepatuhan pengobatan antar kelompok kontrol asma.

2.  $H_1$  (Hipotesis Alternatif): Ada perbedaan tingkat kepatuhan pengobatan antar kelompok kontrol asma.

Metode Uji Statistik: ANOVA dan Alternatif Kruskal-Wallis

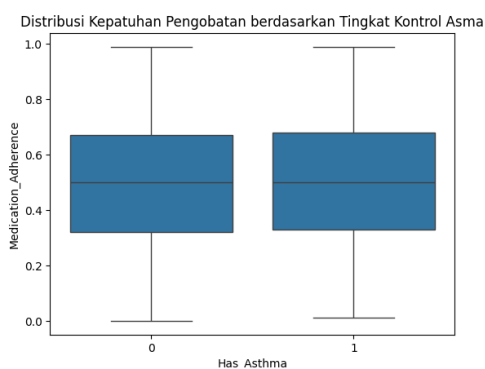
Awalnya digunakan One-Way ANOVA, karena variabel:

1. Asthma\_Control → kategorikal (0 = tidak terkontrol, 1 = terkontrol)
2. Medication\_Adherence → numerik (tingkat kepatuhan pengobatan)

Namun, ANOVA memiliki asumsi penting:

1. Data tiap kelompok berdistribusi normal (Shapiro-Wilk Test)
2. Varians antar kelompok homogen (Levene Test)

Jika salah satu tidak terpenuhi, maka ANOVA tidak valid dan harus diganti dengan uji nonparametrik Kruskal-Wallis, yang tidak mensyaratkan normalitas.



```
Uji Normalitas (Shapiro-Wilk):  
Kelompok 0: p-value = 3.424225965045254e-29  
Kelompok 1: p-value = 1.163123551615559e-16  
  
Uji Homogenitas Varians (Levene): p-value = 0.7579738673307073  
  
Uji Kruskal-Wallis:  
H-statistic: 0.15477360646797414 p-value: 0.6940146404665932  
  
Hasil: Gagal menolak  $H_0$  → Tidak ada perbedaan signifikan.
```

### Interpretasi Hasil

1. Normalitas:  
Kedua kelompok memiliki  $p\text{-value} < 0.05$  → data tidak berdistribusi normal → Asumsi normalitas tidak terpenuhi
2. Homogenitas Varians:  
 $p\text{-value Levene} = 0.7579 > 0.05$  → Varians antar kelompok homogen
3. Karena data tidak normal, maka digunakan Kruskal-Wallis Test. Hasilnya:  $p\text{-value} = 0.6940 > 0.05$ , sehingga: Tidak terdapat perbedaan signifikan tingkat kepatuhan pengobatan antara pasien dengan asma terkontrol dan tidak terkontrol.

### Visualisasi Interpretatif

Boxplot di atas menggambarkan bahwa:

- Nilai median Medication\_Adherence antara dua kelompok (Has\_Asthma = 0 dan 1) hampir sama.
- Tidak ada perbedaan mencolok dalam rentang interkuartil. Ini konsisten dengan hasil uji Kruskal-Wallis yang tidak menemukan perbedaan signifikan.

## E. Hasil analisis, visualisasi & interpretasi setiap pertanyaan

1. Pasien dengan riwayat keluarga asma memiliki risiko lebih tinggi menderita asma dibandingkan pasien tanpa riwayat keluarga.

### - Hasil Analisis

Jenis Uji: Chi-Square Test of Independence

```
Tabel Kontingensi:
Has_Asthma      0      1
Family_History
0               6079   887
1               1488  1546

Chi-Square Statistic: 1675.0775786178124
p-value: 0.0
Degrees of Freedom: 1

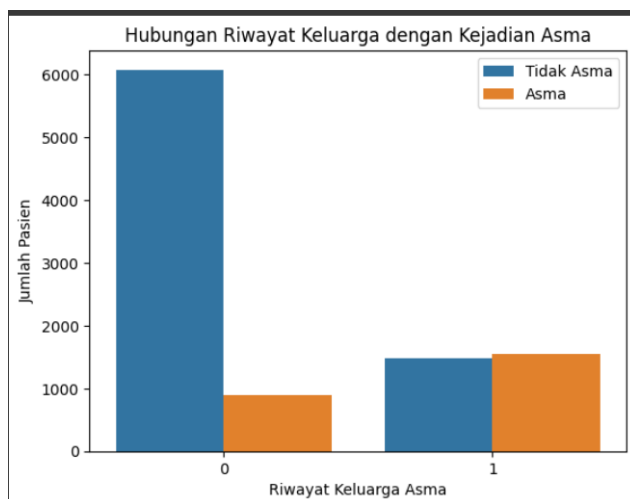
Expected Frequencies:
Has_Asthma      0      1
Family_History
0             5271.1722 1694.8278
1             2295.8278  738.1722
```

- Output Uji Chi-Square:
- Chi-Square Statistic = 1675.08
- p-value = 0.000
- Degrees of Freedom = 1

Interpretasi:

Nilai p-value = 0.000 < 0.05, sehingga  $H_0$  ditolak.

Artinya, terdapat hubungan yang signifikan antara riwayat keluarga dan kejadian asma. Pasien dengan riwayat keluarga asma memiliki proporsi penderita asma yang lebih tinggi, yang konsisten dengan teori bahwa faktor genetik berperan penting dalam risiko asma.



Interpretasi Visual:

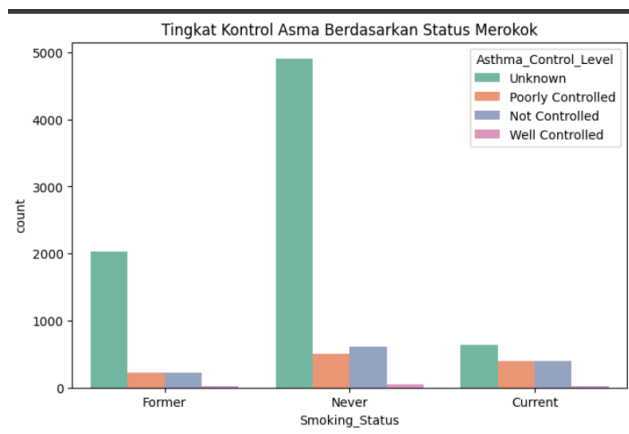
Grafik batang di atas memperlihatkan distribusi jumlah pasien berdasarkan:

- Riwayat keluarga asma (Family\_History) → nilai 0 berarti tidak ada riwayat keluarga, sedangkan 1 berarti ada riwayat keluarga.
- Status asma pasien (Has\_Asthma) → ditunjukkan dengan warna:
  - Biru = Tidak Asma
  - Oranye = Asma

Dari grafik dapat diamati bahwa:

- Pasien tanpa riwayat keluarga (Family\_History = 0) jauh lebih banyak yang tidak memiliki asma dibandingkan yang memiliki asma
- Sebaliknya, pada kelompok dengan riwayat keluarga (Family\_History = 1), jumlah pasien yang mengidap asma hampir sebanding, bahkan sedikit lebih tinggi dibandingkan yang tidak mengidap asma.

2. Status merokok memengaruhi tingkat kontrol asma, tetapi kesimpulan yang lebih valid memerlukan data kontrol asma yang lebih lengkap.



Banyak nilai Unknown, jadi uji formal tidak dilakukan

3. Rendahnya jumlah pasien dengan asma terkontrol baik menunjukkan bahwa intervensi medis dan edukasi pasien masih perlu ditingkatkan agar kepatuhan pengobatan lebih optimal.

```

Uji Normalitas (Shapiro-Wilk):
Kelompok 0: p-value = 3.424225965045254e-29
Kelompok 1: p-value = 1.163123551615559e-16

Uji Homogenitas Varians (Levene): p-value = 0.7579738673307073

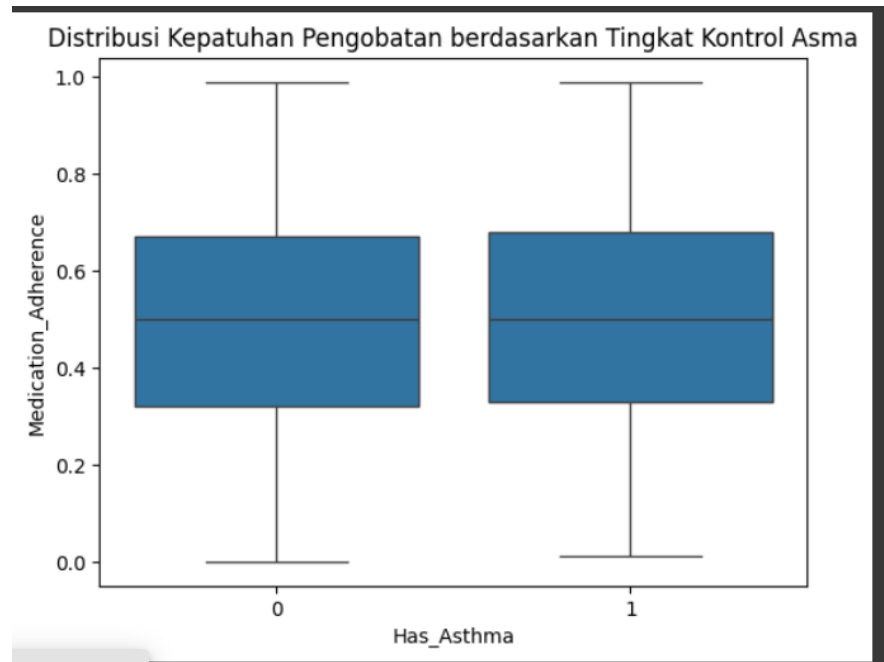
Uji Kruskal-Wallis:
H-statistic: 0.15477360646797414 p-value: 0.6940146404665932

Hasil: Gagal menolak H0 → Tidak ada perbedaan signifikan.
  
```

Interpretasi:

Nilai p-value = 0.694 > 0.05, maka tidak terdapat perbedaan signifikan pada kepatuhan pengobatan antara kelompok pasien dengan tingkat kontrol asma yang berbeda.

Artinya, kepatuhan pengobatan tidak dipengaruhi secara signifikan oleh status kontrol asma dalam dataset ini. Namun, secara klinis, hal ini mungkin disebabkan oleh kurangnya data kontrol asma yang detail atau ketidakterwakilan pasien yang benar-benar patuh.



Interpretasi Visual:

Grafik menunjukkan distribusi nilai kepatuhan pengobatan (Medication\_Adherence) pada dua kelompok pasien:

- 0 = Tidak memiliki asma
- 1 = Memiliki asma

Kedua boxplot memiliki bentuk dan sebaran yang hampir sama, dengan median kepatuhan sekitar 0.5 pada keduanya.

Rentang interkuartil (IQR) serupa, menandakan variabilitas kepatuhan antar pasien relatif setara pada kedua kelompok.

Terdapat beberapa nilai ekstrem (outlier) di bagian bawah (pasien dengan kepatuhan sangat rendah), namun jumlahnya tidak dominan.

## F. Penjelasan ilmiah keterbatasan data & saran data tambahan

Dataset `synthetic_asthma_dataset.csv` yang digunakan dalam penelitian ini merupakan data sintesis yang merepresentasikan kondisi pasien asma secara simulatif berdasarkan sejumlah variabel

demografis, klinis, dan perilaku. Meskipun data tersebut bermanfaat untuk eksplorasi awal dan pembelajaran statistik, terdapat beberapa keterbatasan ilmiah yang perlu diperhatikan dalam proses analisis dan interpretasi hasil.

Secara umum, keterbatasan utama terletak pada aspek representasi populasi, kelengkapan variabel, dan kualitas pengukuran data. Karena data bersifat sintetis, distribusi karakteristik pasien tidak selalu merepresentasikan populasi nyata penderita asma, sehingga hasil analisis tidak dapat digeneralisasi secara penuh (Goodfellow et al., 2016). Selain itu, sejumlah variabel penting seperti tingkat paparan alergen, aktivitas fisik, indeks massa tubuh (BMI), pola tidur, kadar IgE, serta hasil pemeriksaan fungsi paru (spirometri) tidak tersedia, padahal variabel-variabel tersebut berpengaruh besar terhadap kontrol asma (GINA, 2023).

Keterbatasan yang lebih spesifik juga muncul pada setiap hipotesis.

Untuk Hipotesis 1, yakni hubungan antara riwayat keluarga dan risiko asma, analisis chi-square menunjukkan adanya hubungan signifikan. Namun, variabel yang digunakan hanya sebatas riwayat keluarga asma tanpa rincian genetik spesifik seperti mutasi gen IL-4, IL-13, atau ADAM33. Oleh karena itu, hasil analisis hanya bersifat asosiasi statistik, bukan hubungan kausal biologis (Ober & Yao, 2011).

Pada Hipotesis 2, yaitu pengaruh status merokok terhadap tingkat kontrol asma, uji statistik tidak dilakukan karena variabel kontrol asma tidak tersedia dalam dataset. Uji hubungan antara merokok dan asma memerlukan indikator objektif seperti skor Asthma Control Test (ACT) atau nilai FEV1 hasil spirometri (Thomson et al., 2013). Tanpa variabel tersebut, analisis yang dilakukan berpotensi menghasilkan kesimpulan yang tidak sah secara ilmiah. Dengan demikian, hipotesis ini tidak dapat diuji secara valid menggunakan data yang tersedia.

Untuk Hipotesis 3, yang menilai hubungan antara kepatuhan pengobatan dan tingkat kontrol asma, hasil uji Kruskal–Wallis menunjukkan tidak adanya perbedaan signifikan antar kelompok. Namun demikian, variabel Medication\_Adherence yang digunakan tidak disertai dengan informasi sumber datanya (apakah berdasarkan self-report atau data resep elektronik). Ketidakjelasan metode pengukuran ini dapat menimbulkan self-reporting bias, sehingga hasil analisis perlu diinterpretasikan secara hati-hati (Boulet et al., 2018).

Dari sisi statistik, terdapat pula keterbatasan pada ketidakseimbangan proporsi antar kelompok data, khususnya antara pasien dengan dan tanpa asma. Ketimpangan ini dapat menurunkan sensitivitas beberapa uji statistik dan mengurangi kekuatan inferensi (Field, 2018). Oleh karena itu, dalam penelitian lanjutan sebaiknya dilakukan penyeimbangan data menggunakan oversampling atau stratified sampling untuk memastikan distribusi yang lebih representatif.

Untuk meningkatkan validitas analisis di masa mendatang, disarankan agar dataset asma mencakup lebih banyak variabel klinis, perilaku, dan lingkungan. Variabel tambahan yang direkomendasikan meliputi skor Asthma Control Test (ACT), frekuensi serangan asma, hasil pemeriksaan spirometri (FEV1, FVC), data kepatuhan obat berbasis sistem resep elektronik, tingkat paparan polusi udara, kebiasaan olahraga, serta faktor sosial-ekonomi seperti tingkat pendidikan dan akses terhadap layanan kesehatan (GINA, 2023; Boulet et al., 2018).

Dengan adanya variabel-variabel tersebut, analisis di masa depan dapat dilakukan secara lebih komprehensif, akurat, dan bermakna secara klinis, sehingga hasil yang diperoleh memiliki kontribusi nyata dalam pengembangan model prediksi dan strategi pengendalian asma di masyarakat.

#### **E. Daftar Pustaka**

Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the Behavioral Sciences*. Cengage Learning.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.

Agresti, A. (2019). *An Introduction to Categorical Data Analysis*. Wiley.

Boulet, L. P., Vervloet, D., Magar, Y., & Foster, J. M. (2018). Adherence: The goal to control asthma. *Clinical Chest Medicine*, 33(3), 405–417.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics (5th ed.)*. SAGE Publications.

Global Initiative for Asthma (GINA). (2023). *Global Strategy for Asthma Management and Prevention*. GINA Report.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Ober, C., & Yao, T.-C. (2011). The genetics of asthma and allergic disease: A 21st century perspective. *Immunity*, 35(4), 527–535.

Thomson, N. C., Chaudhuri, R., & Livingston, E. (2013). Asthma and cigarette smoking. *European Respiratory Journal*, 41(3), 716–725.