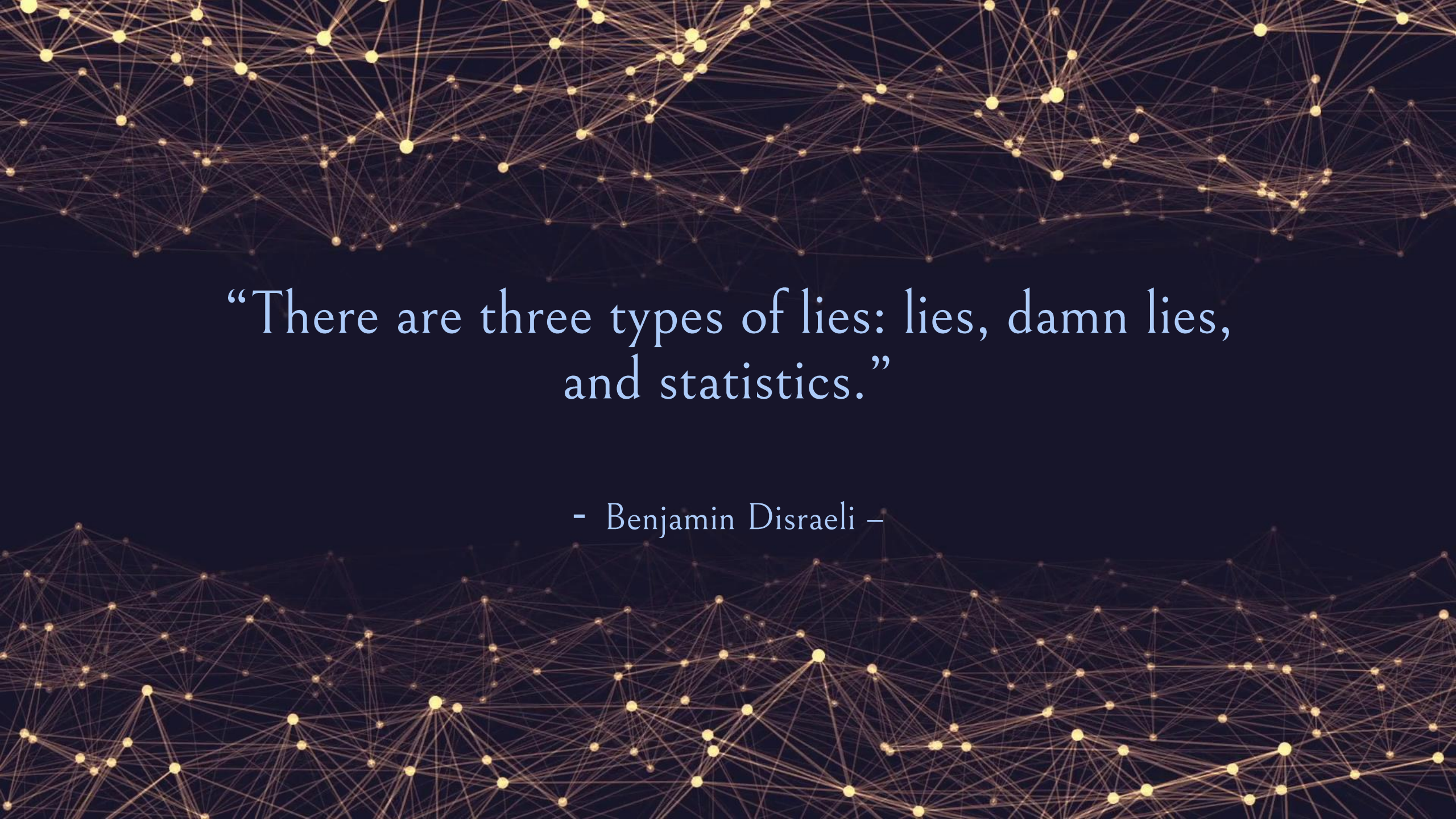




# Seberapa Benar Kebenaran Statistik?

Aditya Firman Ihsan



“There are three types of lies: lies, damn lies,  
and statistics.”

- Benjamin Disraeli -

“  
Berbicaralah dengan data!”

kata banyak orang

Tapi,  
Apa maksudnya?

Berdasarkan **data** = pasti **benar**?



Bagaimana kita tahu sesuatu itu benar?

Kebenaran elementer

disaksikan langsung (dari realita)

diasumsikan/dianggap/dipercaya benar

# Kebenaran komposit

**disimpulkan** dari pernyataan lain yang sudah diketahui benar

**diterima** dari orang yang mengatakan itu benar



Menyimpulkan sesuatu:

- Induksi dari beragam **pengamatan** spesifik
- Deduksi dari preposisi/**gagasan** umum

Bagaimana sebenarnya melakukan induksi?

Apakah

Melakukan **sebanyak mungkin** pengamatan?

Mengumpulkan data **selengkap mungkin**?

# Data

representasi **realitas**, tidak hanya berupa ukuran numerik, tapi bisa juga peristiwa (*event*)

Realita itu kompleks

+

instrumen untuk mengamatinya terbatas

=

Data yang tak pernah bisa lengkap

=

Informasi yang tidak pernah bisa pasti benar

Padahal,

Informasi yang benar selalu dibutuhkan untuk  
memutuskan sesuatu.

Yang **Pasti** adalah **Ketidakpastian**

Pada akhirnya, setiap orang selalu berpikir,  
memilih, dan bertindak dengan menerka-nerka

Manusia adalah peramal!

Tidak ada hal yang bisa diketahui secara langsung dengan kepastian 100%.

Tapi,  
kita peramal jenius.

Daripada **pasrah** pada ketidakpastian, manusia **belajar** untuk “mengendalikan” ketidakpastian itu,





Bagaimana caranya?

Jadikan ketidakpastian itu bagian dari informasi!

Itulah **statistika**

Alat untuk menyimpulkan dengan tegas lengkap  
dengan jaminan ketidakpastiannya.

~~Saya tahu bahwa A benar~~

Saya yakin sekian persen bahwa A benar

Bagaimana statistika melakukannya?

Dengan dibawa dulu memutar ke

dunia tanpa bias

bernama matematika

Entitas Riil

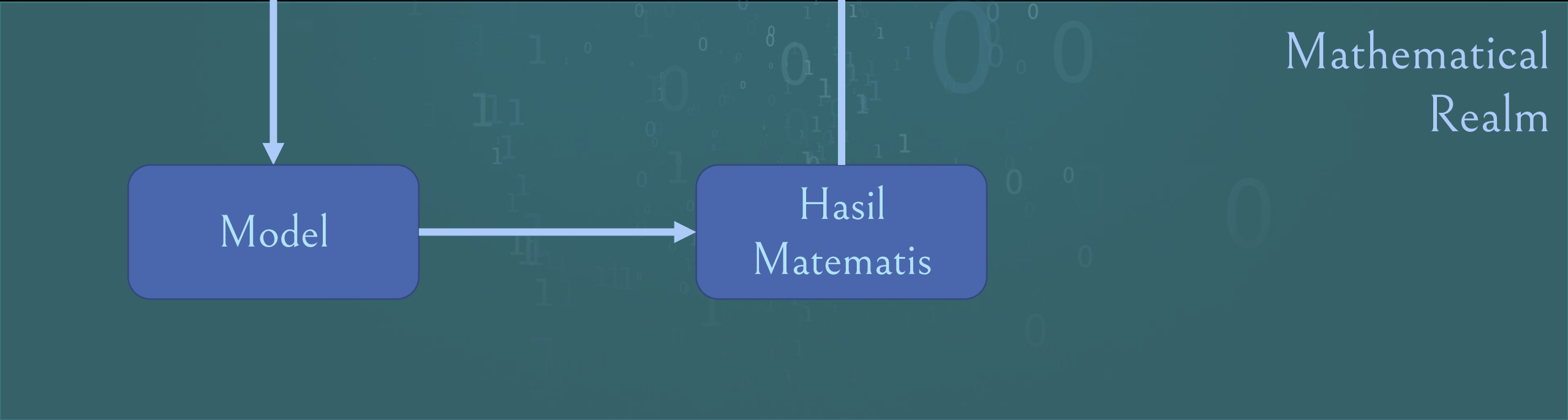
Kesimpulan

Reality

Mathematical  
Realm

Model

Hasil  
Matematis



Matematika bekerja dengan objek minim tafsir

Objek riil  $\rightarrow$  Kuantifikasi

Bagaimana cara kuantifikasi...

- kemajuan ekonomi?
- kepuasan masyarakat?
- keberhasilan pendidikan?

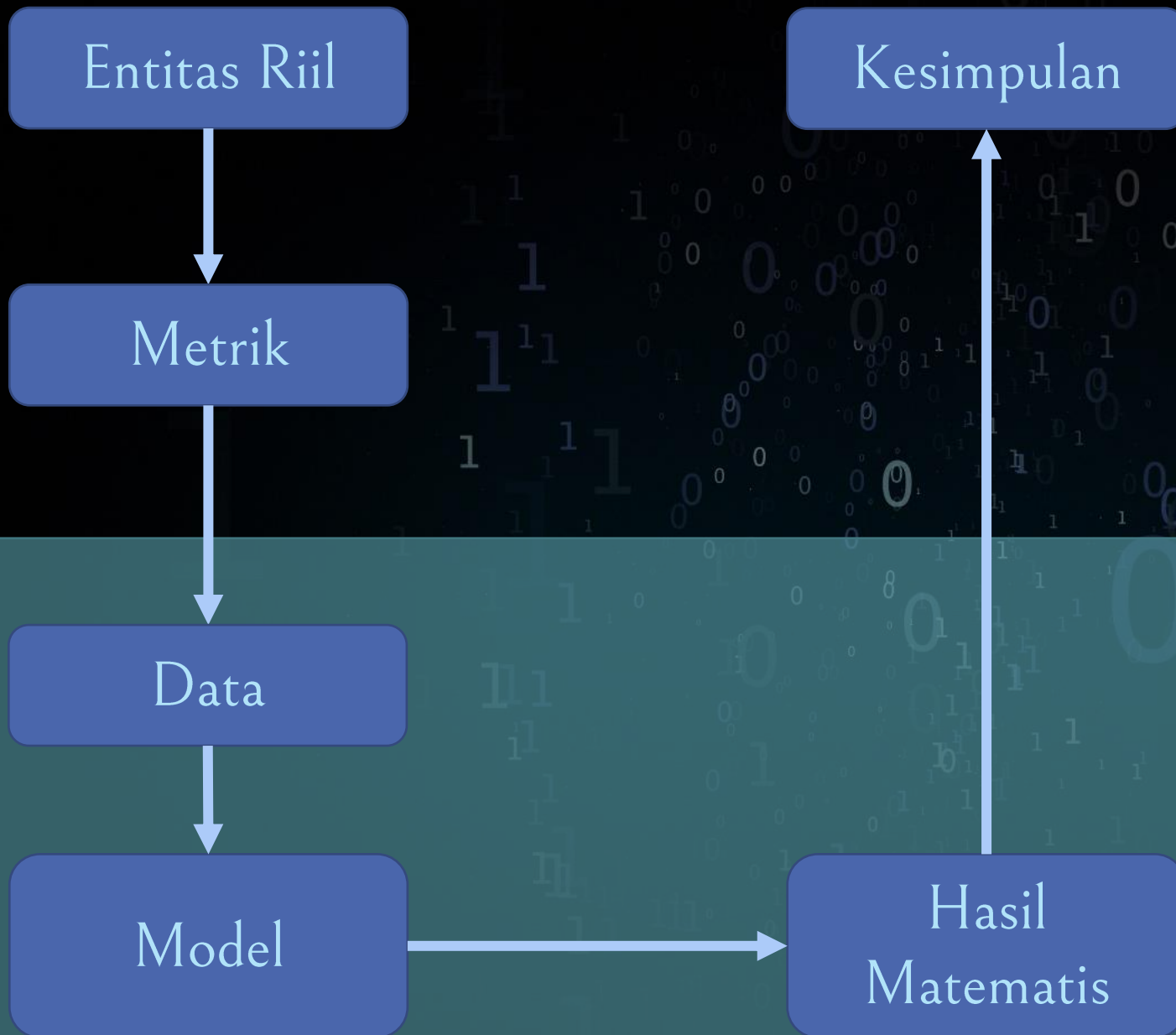
# Metrik

Ukuran buatan sebagai representasi terdekat  
konsep kualitatif



# Metrik

- Kemajuan ekonomi -> GDP
- Kepuasan masyarakat -> Indeks Kebahagiaan
  - Keberhasilan Pendidikan -> nilai ujian



Reality

Mathematical  
Realm

Sederhana?

Jangan lupa, data yang diukur tidak selalu bisa lengkap!

Data **masyarakat** Indonesia?

Data **hasil panen jagung** di Jawa Barat?

Data **cuaca** selama 1 dekade?

Data **pasien** suatu Rumah Sakit?

Melihat yang besar dari yang kecil

Perkenalkan,

**Sampel!**

Sampel < Populasi sesungguhnya

Sehingga

Kesimpulan dari Sampel < Kebenaran sesungguhnya

# Eksperimen

proses pengambilan data sampel  
(pengukuran nilai numerik atau pengamatan  
peristiwa)

Ekperimen ideal:  
Tanpa manipulasi, tanpa bias



Jaminannya?

Acak

Tidak acak

=

sampel tidak mencerminkan populasi

## Keacakan eksperimen

**Jaminan** statistik terhadap **akurasi** yang dihasilkan  
Tanpa pola yang acak, tidak ada analisa statistik

Acak berarti benar

## Sampel acak sederhana

- **Tanpa bias** (setiap elemen punya peluang yang sama untuk terpilih)
- **Independen** (pemilihan elemen yang satu tidak mempengaruhi peluang pemilihan elemen lain)

Mungkinkah?

Tidak, tapi dapat **didekati**

# Metode sampling

Cara untuk mendapatkan sampel seacak mungkin

# Metode sampling

Cara ideal:  
mengacak seluruh populasi



# Metode sampling

Sampling bertingkat

Sampling kelompok,

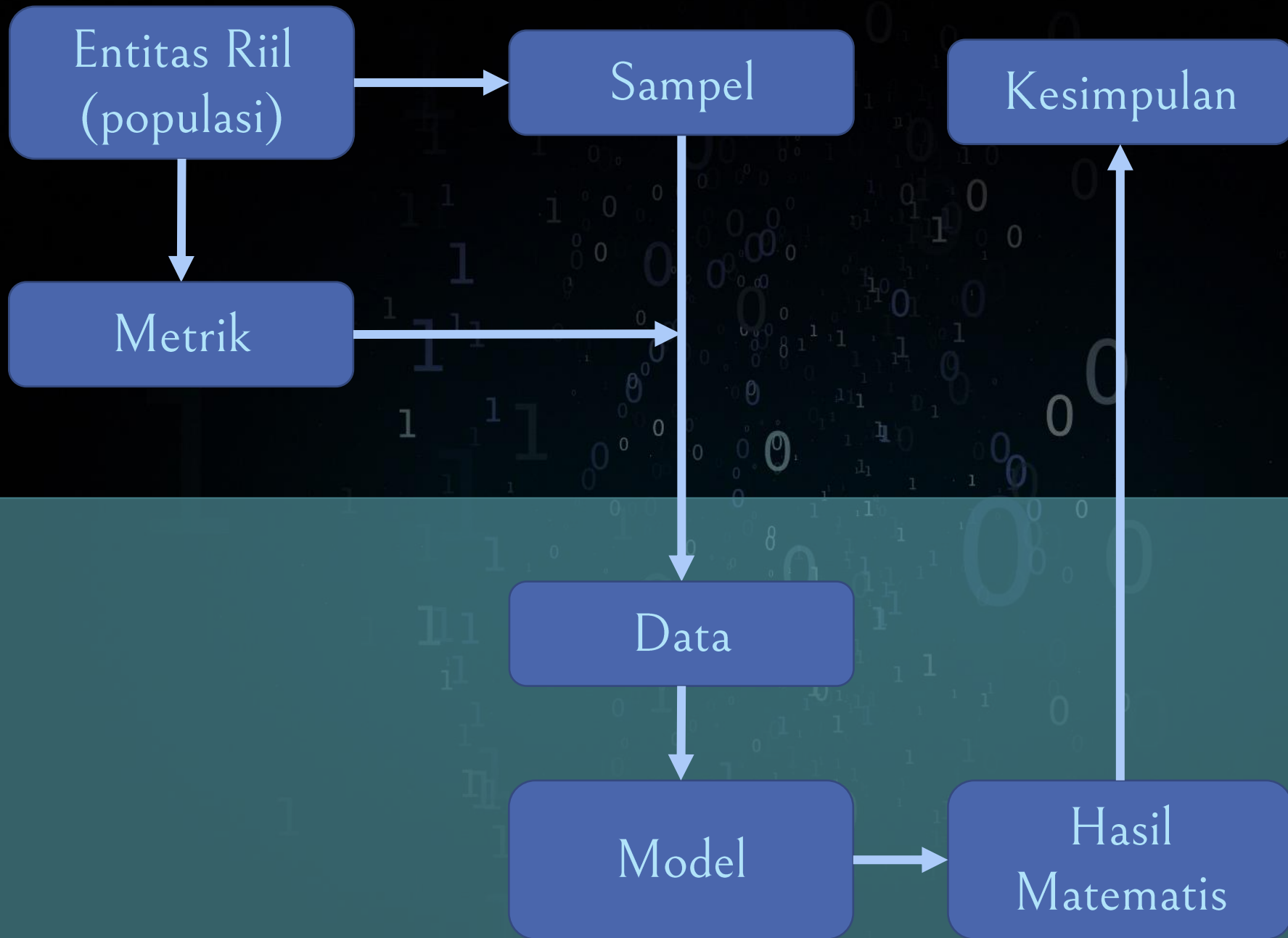
Sampling sistematis,

etc

## Metode sampling

Akan **selalu** ada **bias** yang sukar diatasi

Bagaimana memastikan responden menjawab jujur atau benar-benar paham maksud pertanyaannya?



Reality

Mathematical  
Realm

A network graph visualization consisting of numerous nodes (small yellow circles) connected by thin, light-colored lines (edges). The nodes are distributed across the frame, with some clusters and some isolated nodes. The background is dark, making the nodes and edges stand out.

*Apa yang bisa disimpulkan dari sampel?*

# Ruang Sampel

Semua **kemungkinan** hasil dari suatu eksperimen acak

# Ruang Sampel

Misal, 10 data survey untuk dua kandidat

A, B, B, A, A, A, B, A, A, B

Ruang sampelnya hanya {A, B}

# Ruang Sampel

Apakah langsung bisa disimpulkan bahwa kandidat A menang?

# Ruang Sampel

Sampel **bukan** representasi utuh populasi.  
Kita **hanya** bisa melihat **frekuensi relatifnya**.



# Frekuensi Relatif

Peluang kemunculan suatu hasil elementer  
ketika eksperimennya diulang

Ada 10 data IPK:

3.0, 3.1, 3.5, 2.5, 1.0, 3.9, 1.9, 2.6, 3.3, 3.0

Kesimpulan:

60 persen mahasiswa IPK dibawah 3?

Yang tepat:

Bila diambil sampel mahasiswa baru dari populasi, maka peluang IPK mahasiswa itu di bawah 3 adalah 0.6

# Teori Peluang

Basis **fundamental** matematika statistika

Ya, statistika berkembang dari meja judi

# Teori Peluang

Setiap anggota ruang **sampel** dipetakan ke suatu nilai **peluang** (dari 0 ke 1)

Total peluang seluruh isi ruang sampel harus sama dengan 1.

# Teori Peluang

## Contoh

Eksperimen: pelemparan koin

Ruang sampel: {gambar, angka}

Peluang gambar = peluang angka = 0.5

# Teori Peluang

Ruang sampel dibuat numerik  $\rightarrow$  **Peubah Acak**  
 $\{\text{gambar, angka}\} \rightarrow \{0,1\}$

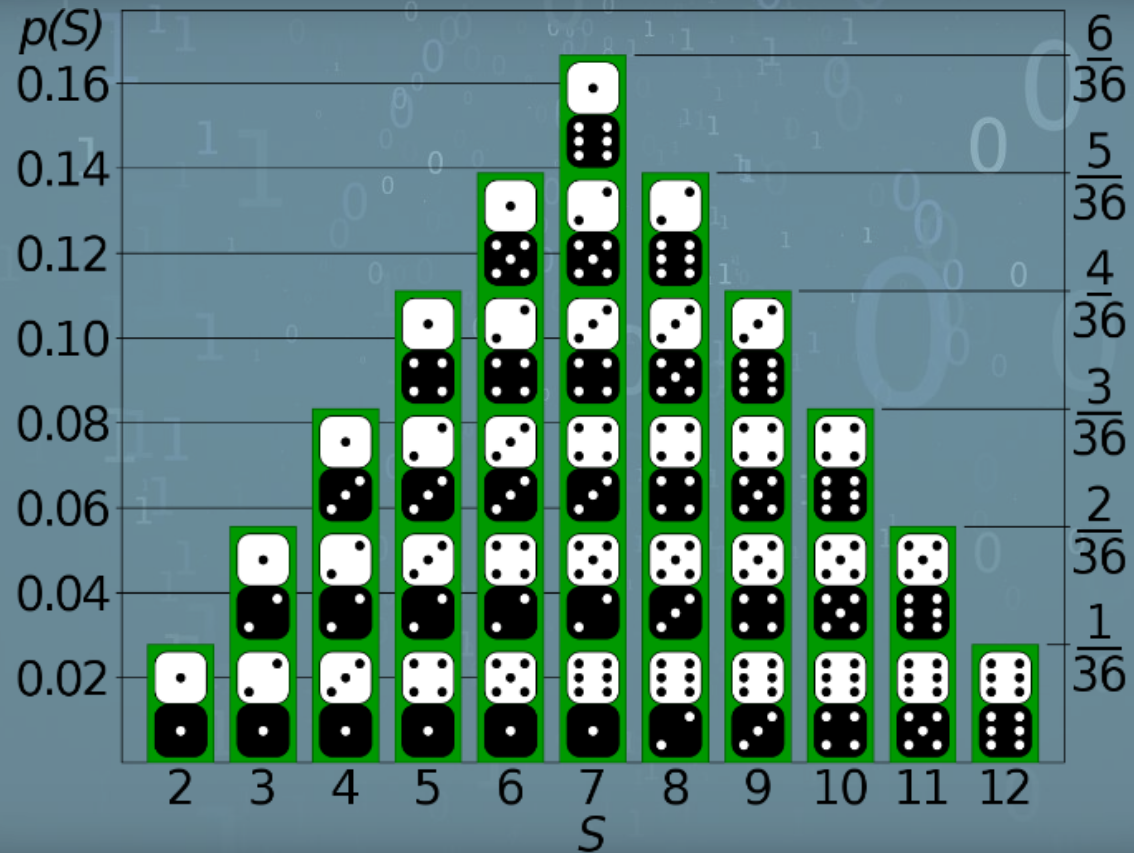
# Teori Peluang

Setiap nilai **peubah acak** dipetakan ke suatu nilai  
peluang:

## Distribusi



# Distribusi Peluang



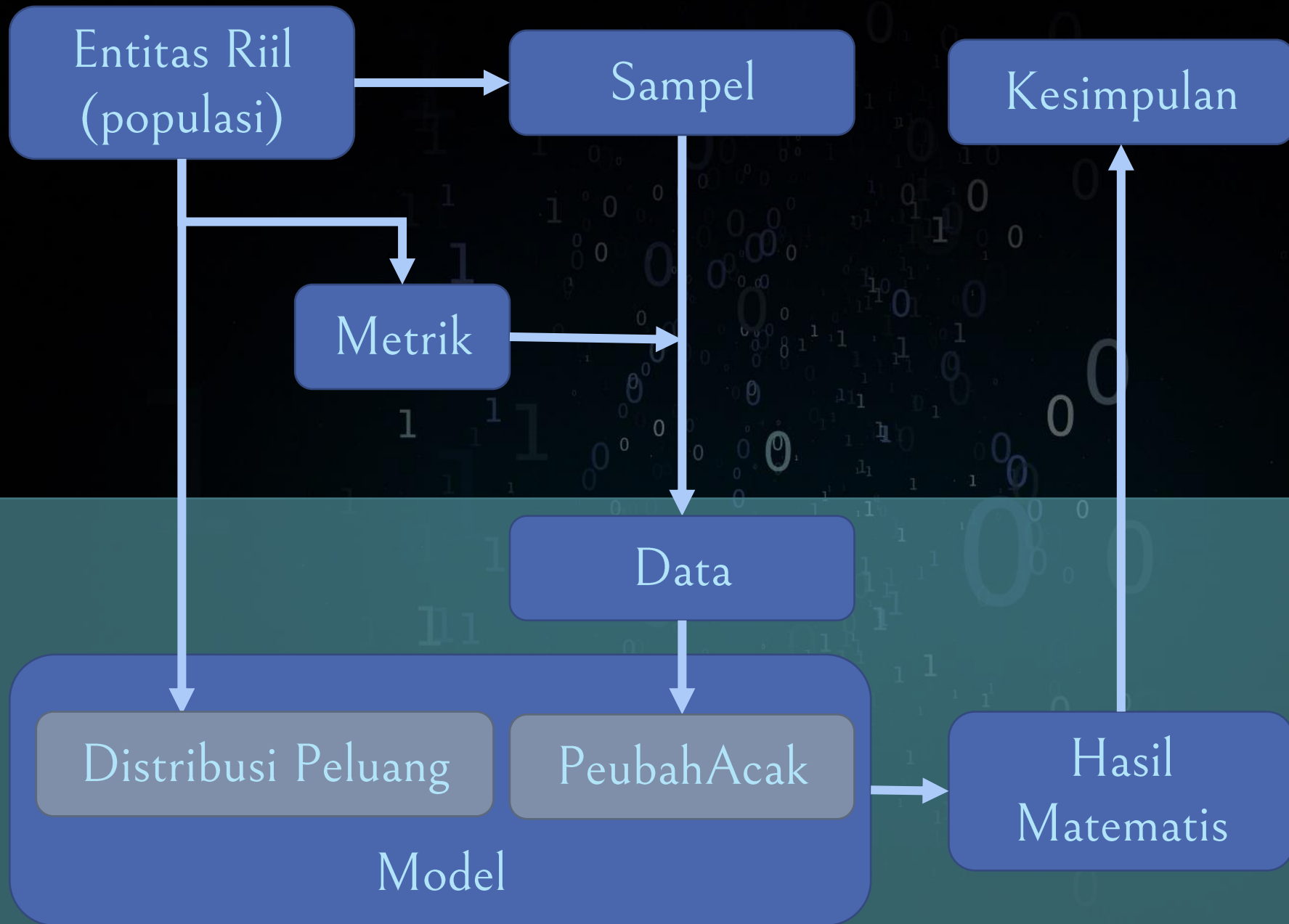
# Distribusi Peluang

Karakteristik suatu sampel  
Tidak selalu dapat dihitung  
Diasumsikan dari populasi

Distribusi peluang + Peubah acak

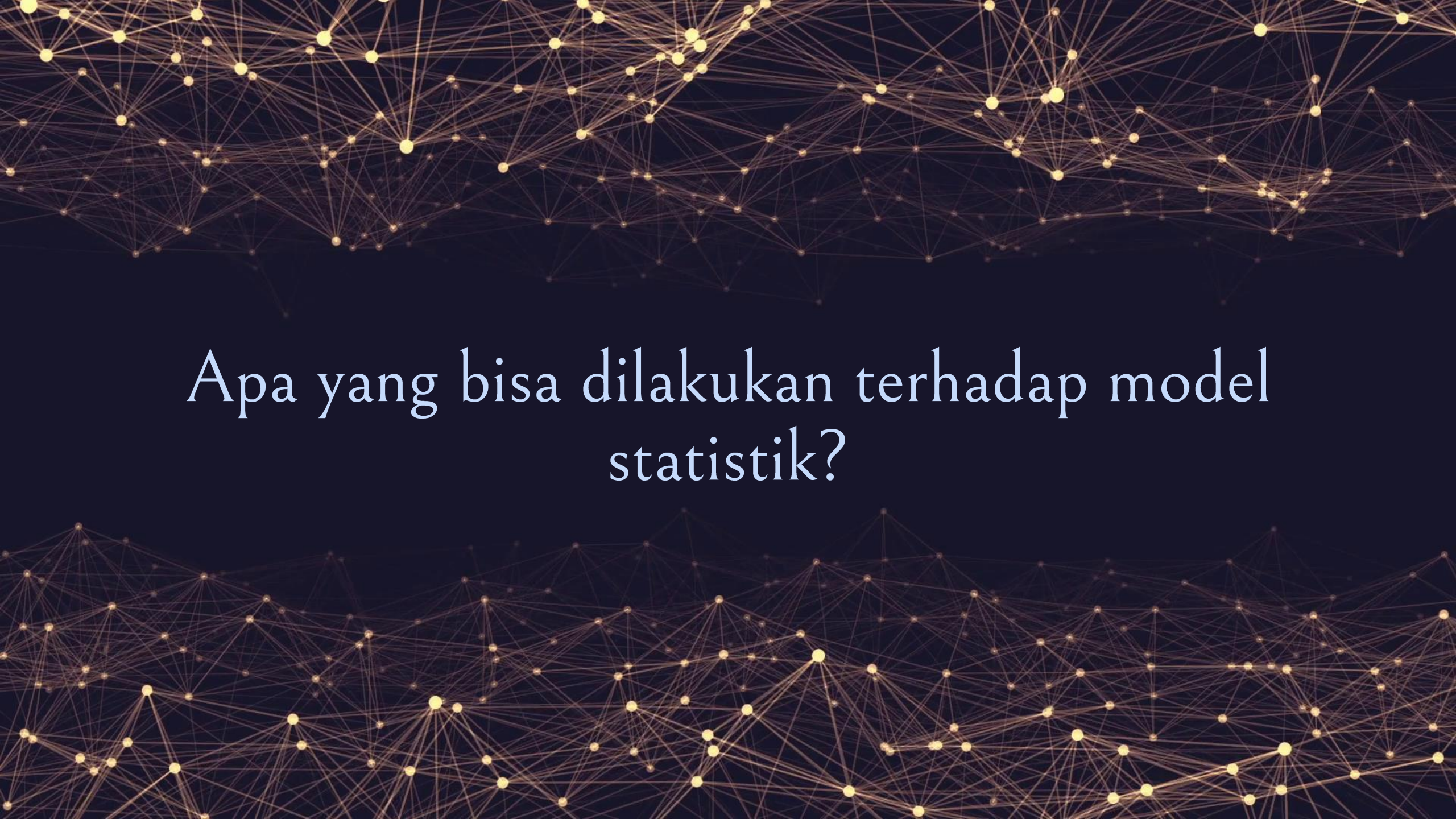
=

Model statistik



Reality

Mathematical  
Realm



Apa yang bisa dilakukan terhadap model statistik?

**Alert!**

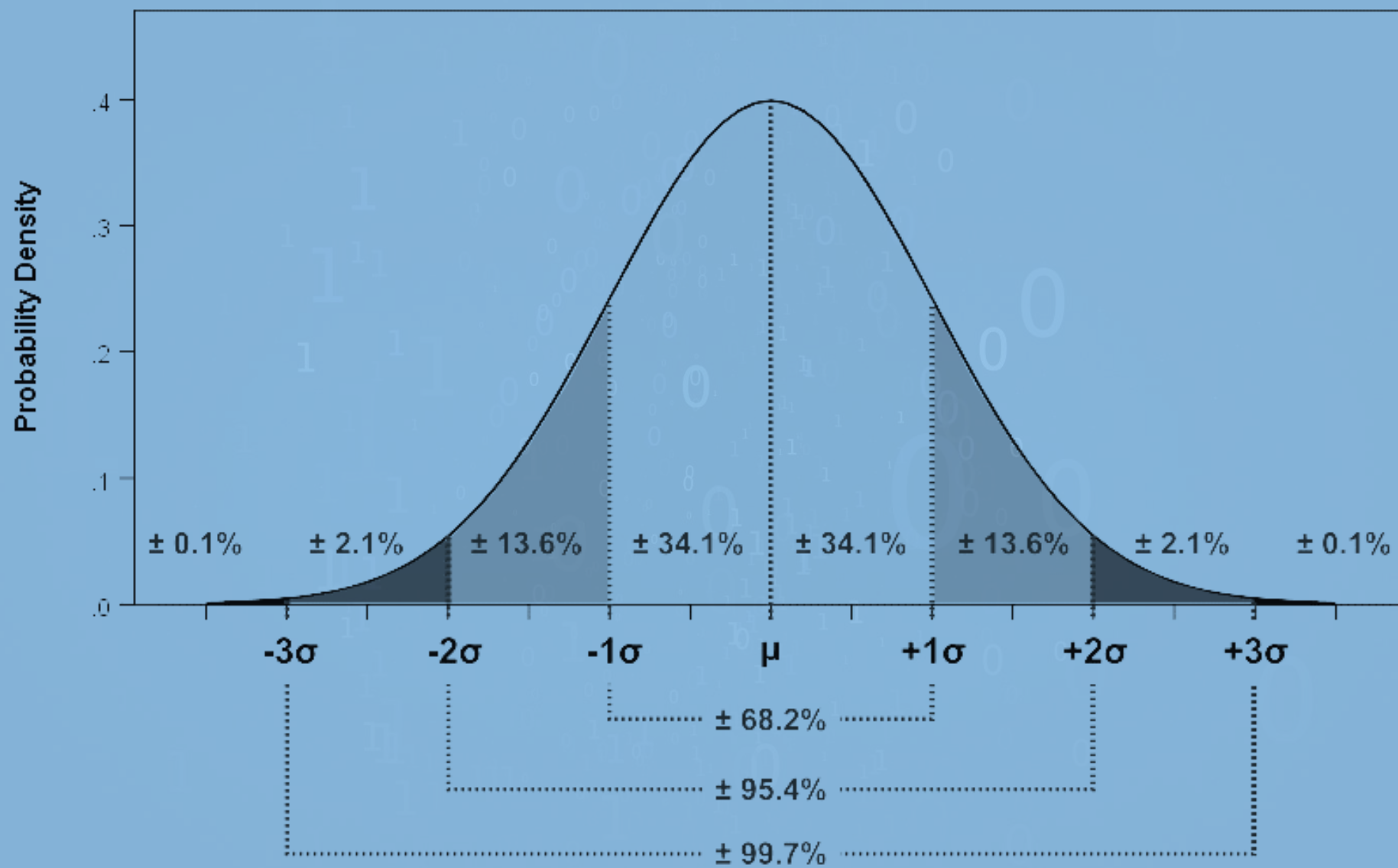
Sedikit bertemu hal teknis dulu...

# Teorema Limit Pusat

Distribusi gabungan dari sekelompok peubah acak, dalam kondisi tertentu, selalu dapat diubah menjadi berbentuk...

# Standard Normal Distribution

$\mu = 0 \mid \sigma = 1$





1.

Dari suatu data sampel,  
kita bisa hitung nilai dari variabel yang kita  
pandang

Misal data survey sebelumnya, berarti kita dapatkan 60  
persen kandidat memilih A

2.

Kita bisa hitung simpangan baku dari data sampel (bergantung banyaknya sampel)

Misal, didapatkan simpangan baku = 0.154

3.

Kita bisa tentukan apa yang mau kita atur:  
Kesalahan atau keyakinan

Misal, kita tetapkan kita ingin bisa yakin 95.4 persen dengan  
hasil survey.

4.

Gunakan grafik lonceng di atas untuk  
dapatkan kesalahannya berapa

Misal, ternyata diperoleh kesalahannya adalah 2 kali  
simpangan baku, sehingga didapatkan  $0.6 \pm 0.308$

Dalam kasus survey **10 pemilih** tadi,  
Berarti kemungkinan kandidat A menang  
sebenarnya adalah di antara  
 $29.2 - 90.8 \%$ !

Hasil statistik lengkap harus memuat 3 hal:

- Nilai perkiraan
- Margin kesalahan
- Tingkat keyakinan

The background of the image features a complex network graph. It consists of numerous small, glowing yellow nodes connected by thin, light-colored lines (edges). The nodes are distributed across the frame, with some appearing as larger, brighter hubs. The overall effect is a dense, interconnected web of points and lines, set against a dark, almost black background. The text is centered in the middle of the image.

Apakah statistika cuma itu doang?

Tentu **tidak**.

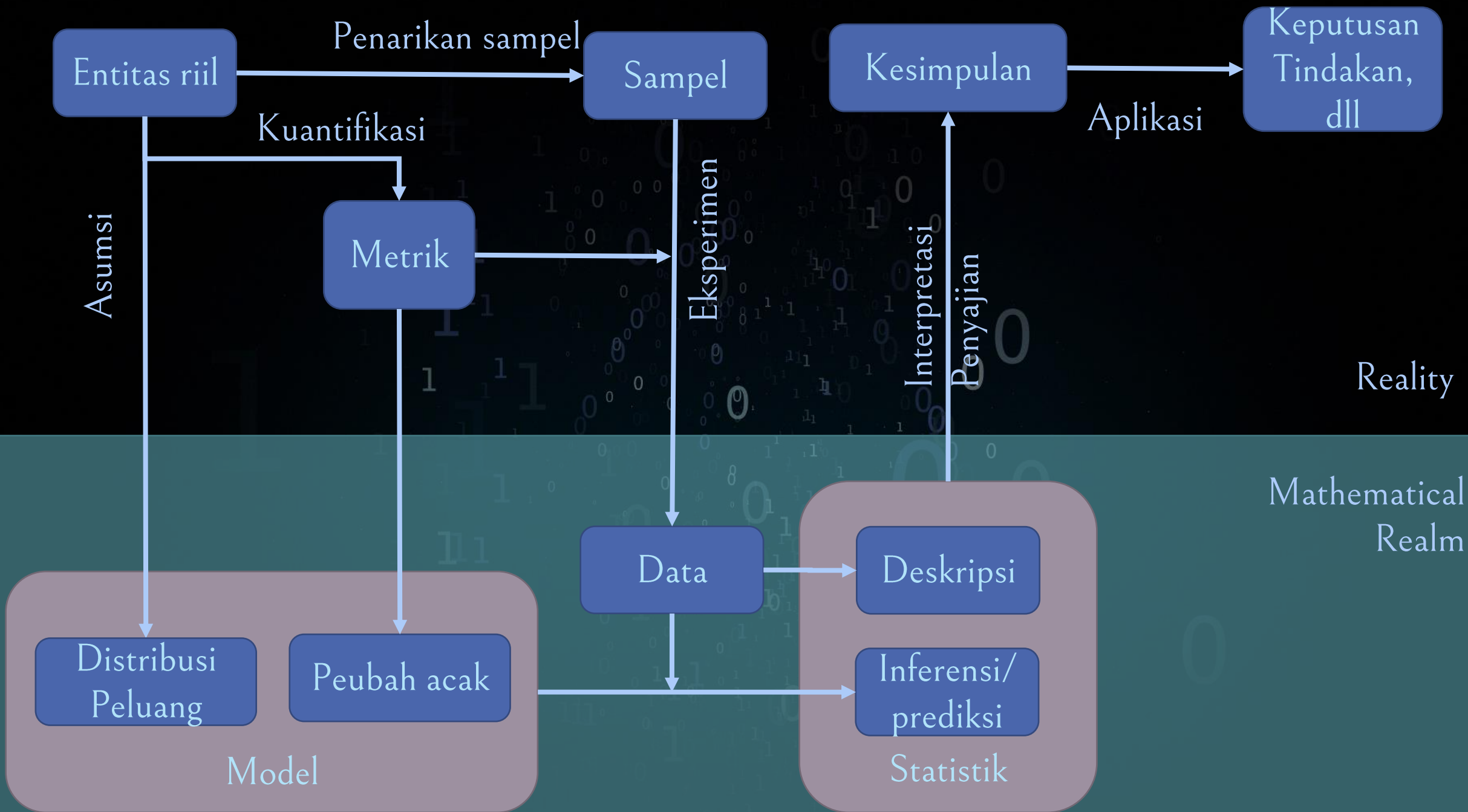
Dari perbandingan populasi, uji hipotesa, regresi, hingga aplikasinya saat ini di machine learning, semua menggunakan statistika

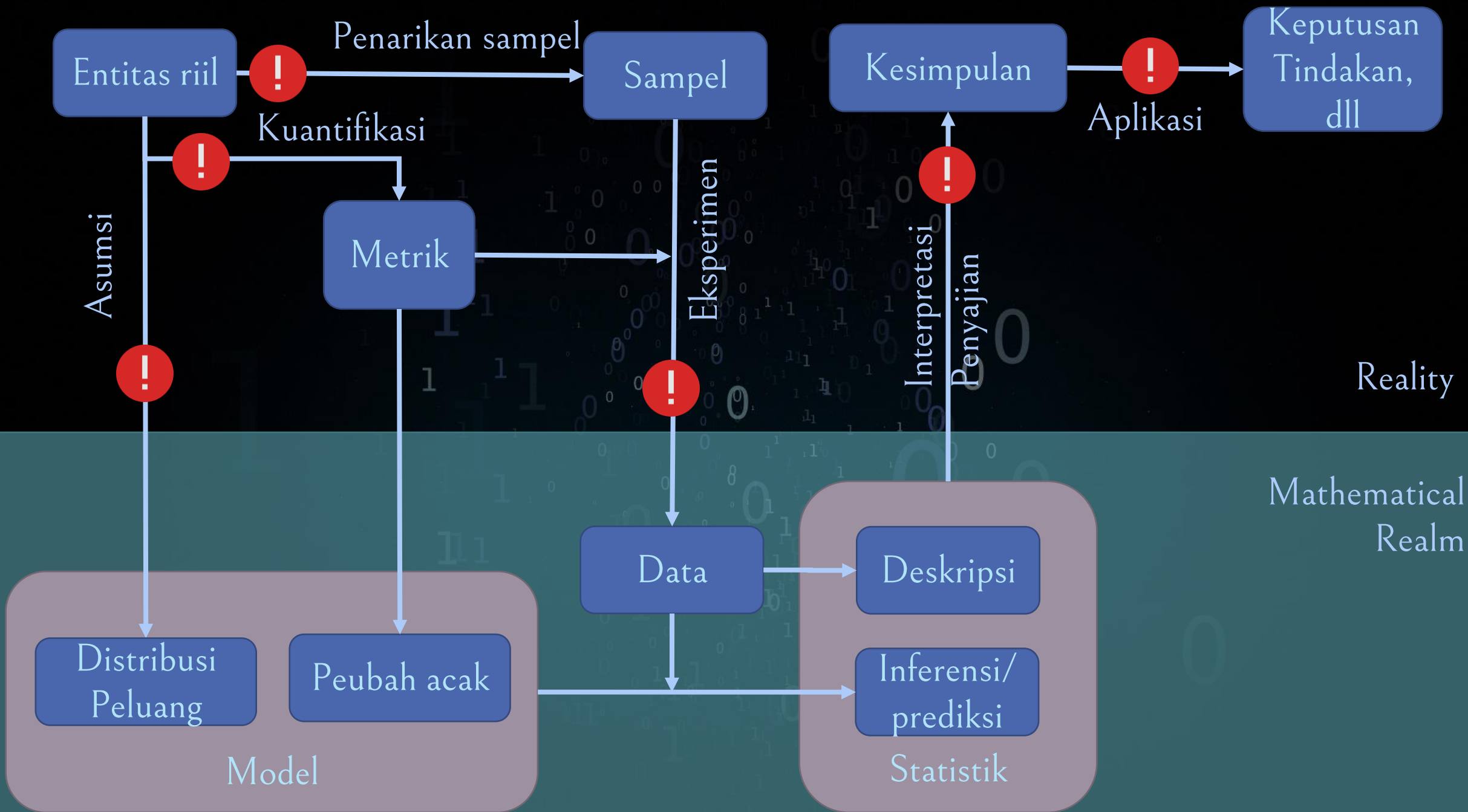


Tapi semuanya berdiri di atas konsep yang sama,

Konsep peluang, konsep keacakan,

Konsep bahwa kita bisa mengukur  
ketidakpastian





# Potensi **bias** statistik

- Asumsi model
- Formulasi metrik (Kuantifikasi)
  - Penarikan Sampel
  - Pengambilan data dari sampel
- Interpretasi dan penyajian kesimpulan statistik

## 1. **Bias** Asumsi

Realita mengandung banyak variable, sehingga perlu disederhanakan

## 1. Bias Asumsi

Hal seperti independensi setiap eksperimen tidak punya jaminan selain diasumsikan

## 2. Bias Metrik

Angka = label terurut

Tapi, label dari apa?

## 2. Bias Metrik

Nilai Ujian Pilihan Ganda → kemampuan memilih suatu pilihan yang tepat dari pertanyaan yang diberikan, yang diatur sesuai target kurikulum.



## 2. Bias Metrik

Apakah ketepatan memilih pilihan jawaban merepresentasikan keberhasilan sekolah?

### 3. Bias sampel

Ekstraksi data dari realitas tidak mungkin efisien 100 persen

## 4. **Bias** penyajian

Apa yang disampaikan ke publik tidak selalu utuh, apalagi jika memementingkan keterbacaan cepat



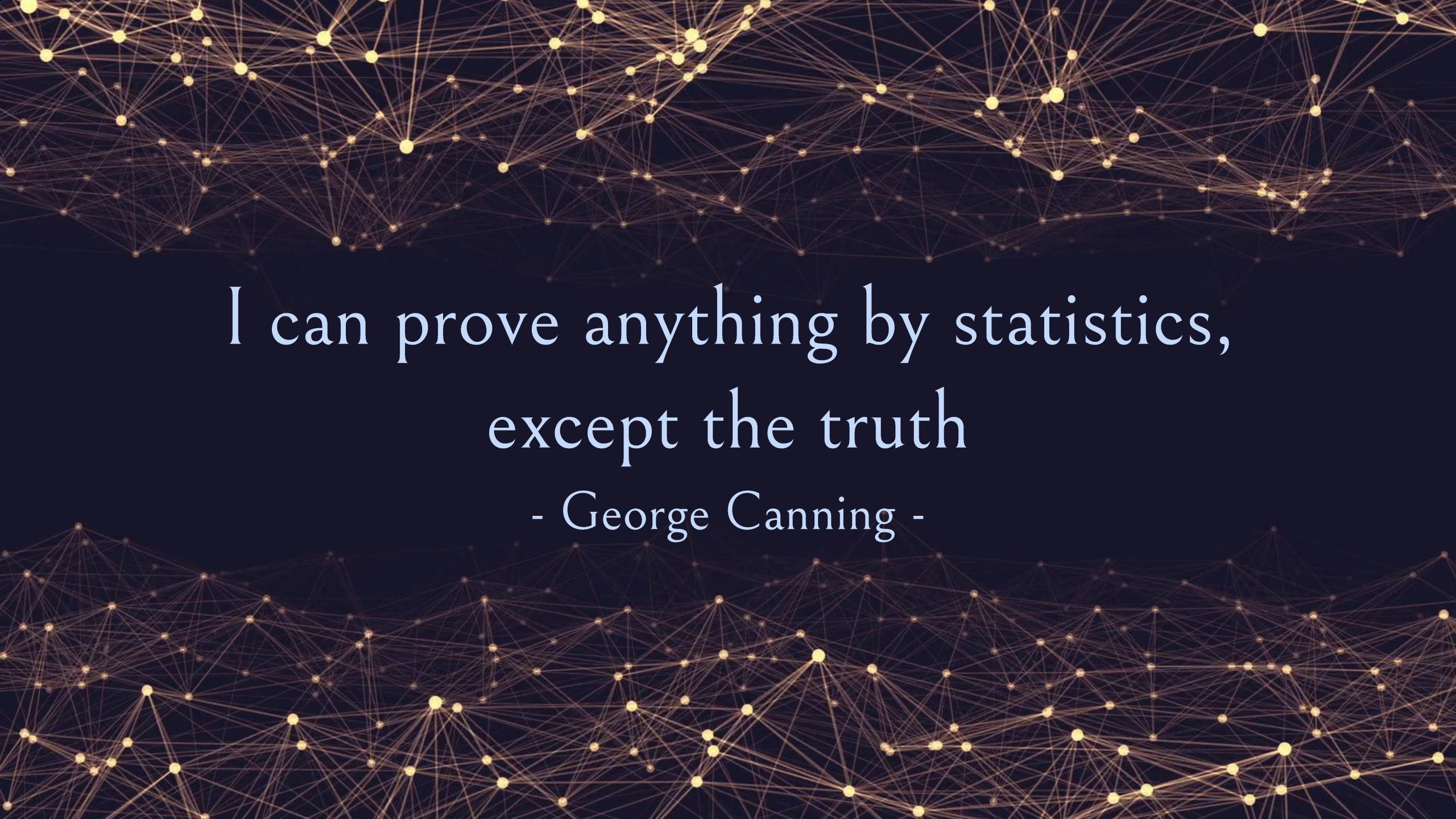
Jadi,  
Seberapa benar kebenaran statistik?



I can prove anything by statistics,

...

- George Canning -



I can prove anything by statistics,  
except the truth

- George Canning -