

AI Mastery Course



**Kampus
Merdeka**
INDONESIA JAYA

orbit
FUTURE ACADEMY | Skills
For
Future
Jobs

*Artificial Intelligence
Mastery Program*

Modul 2


Introduction to everything

Bagian

Data analytics with python-
Advance statistics



Learning Objectives

- Di akhir modul ini, Anda akan dapat:
 - Memahami teori peluang
 - Memahami peluang bersyarat dan teorema bayes
 - Menjelajahi aturan empiris dan teorema chevbeys
 - Menjelajahi skor z, uji z
 - Memahami ide logis dari uji t, uji t satu sampel dan uji t berpasangan
 - Menjelajahi ANOVA dan uji chi square
- 



Agenda

Artificial Intelligence Mastery Program

01

STATUS INFERENSIAL

- Pengujian hipotesis
- Error dan selang kepercayaan
- Distribusi normal

02

UJI HIPOTESIS

- uji T
- Tes Chi Square
- ANOVA

03

KESIMPULAN

- Ulangan
- Ringkasan



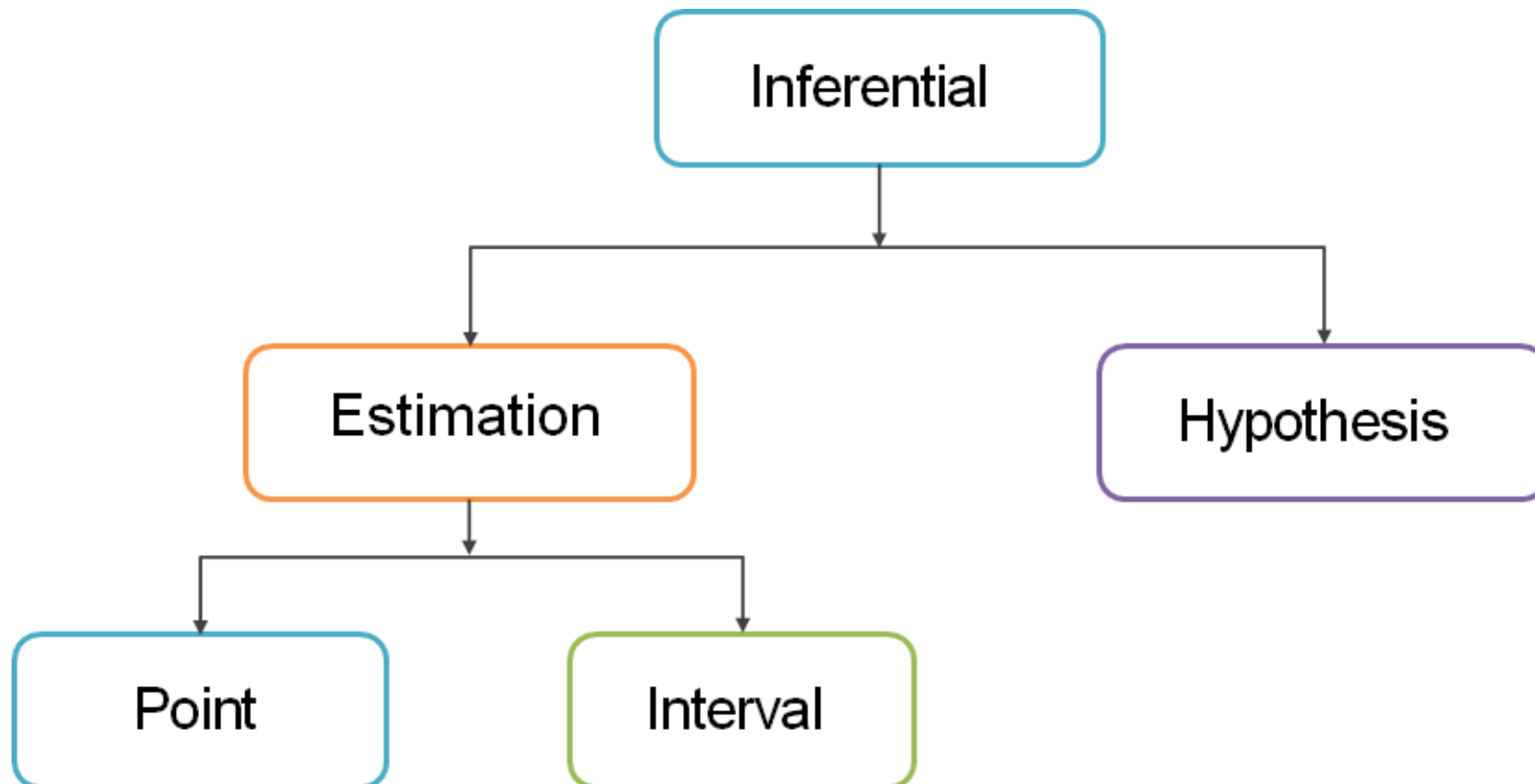
01

STATISTIK INFERENSIAL

- Pengujian hipotesis
- Kesalahan Tipe I & II
- Distribusi normal

Statistik Inferensial

Statistik Inferensial terdiri dari pengujian hipotesis dan estimasi



Hipotesis

- Hipotesis berasal dari bahasa Yunani, yaitu *Hypo* dan *Thesis*. *Hypo* berarti lemah, kurang, atau di bawah, sedangkan *Thesis* berarti teori atau pernyataan yang disajikan dengan bukti.
- Hipotesis adalah pernyataan yang masih bernilai lemah dan perlu dibuktikan

H₀ : Hipotesis Nol



Hipotesis nol sering kali merupakan klaim awal yang didasarkan pada analisis sebelumnya atau pengetahuan khusus.

H_a : Hipotesis Alternatif



Hipotesis alternatif adalah pernyataan yang digunakan peneliti untuk membuktikan suatu kebenaran berdasarkan data sampel

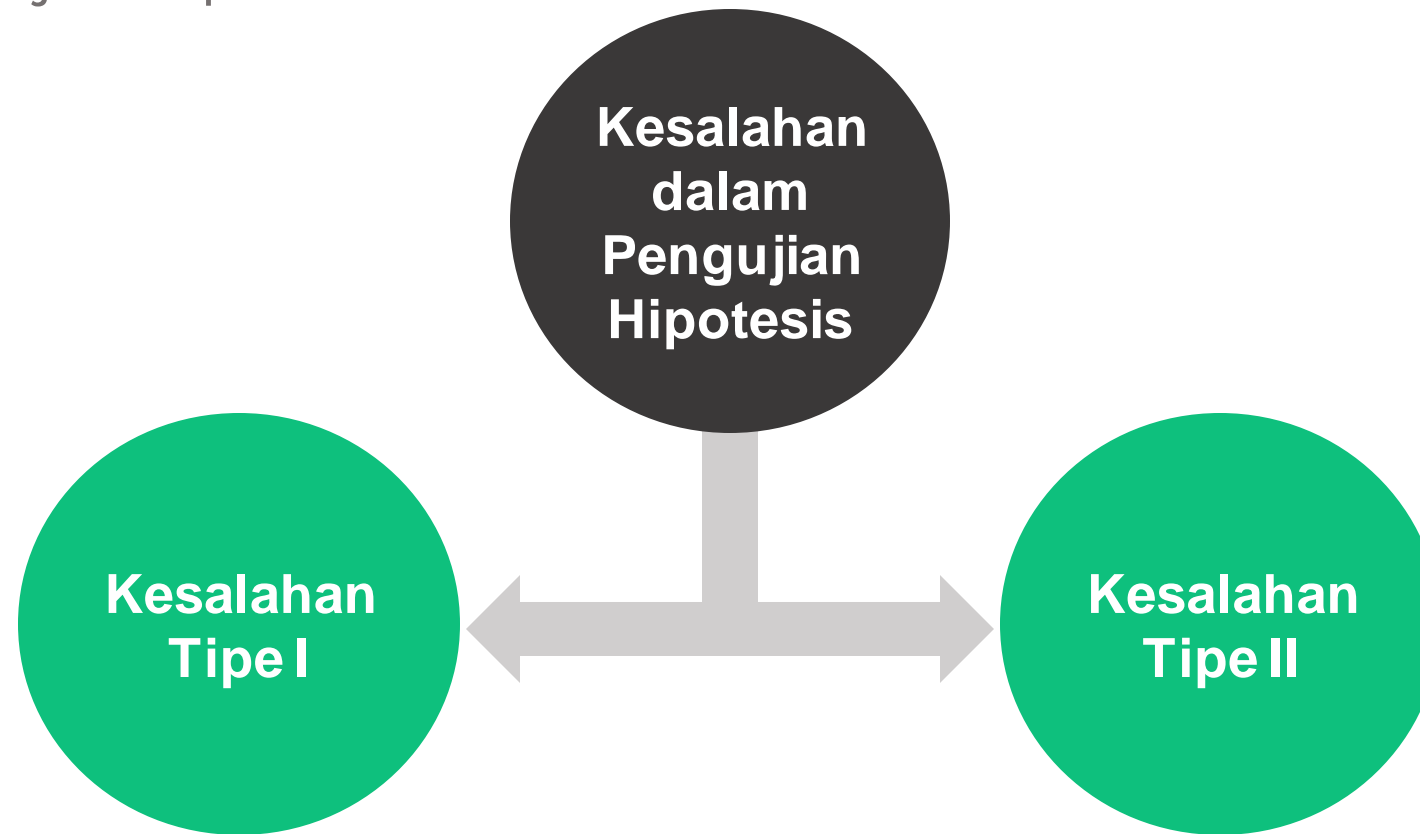
Kriteria Konstruksi Hipotesis

Sebuah hipotesis harus memenuhi kriteria berikut:

- Harus dapat diuji secara empiris, apakah itu benar atau salah.
- Harus spesifik.
- Pernyataan dalam hipotesis tidak boleh bertentangan.
- Harus menentukan variabel di mana hubungan itu akan dibangun.
- 1 hipotesis menggambarkan 1 masalah saja.

Kesalahan dalam Pengujian Hipotesis

- Dua jenis pengujian hipotesis:



Skenario Kesalahan dalam Pengujian Hipotesis

Tabel berikut merangkum skenario kesalahan dalam pengujian hipotesis.

Keputusan	H_0 benar	H_a benar
Tolak H_0	Error Tipe I	Keputusan benar
Terima H_0	Keputusan benar	Error Tipe II

Error Tipe I dan Selang Kepercayaan

- Peluang terjadinya Error Tipe I disebut juga level signifikansi (α)
- Lawan dari Peluang Error tipe I adalah Confidence Interval (CI) atau Selang Kepercayaan
- Jadi, jika selang kepercayaan 95%, maka $\alpha = 5\%$
- Bagaimana pemilihan α yang tepat? Tergantung kasus, tapi umumnya 5%
- Semakin tinggi resiko --> alpha semakin kecil

Misalnya di kasus efek obat, karena obat menyangkut nyawa seseorang, maka α maksimal 1%

- Semakin tinggi subjektifitas suatu kasus --> alpha semakin besar

Misalnya di kasus sosial, α boleh hingga 10%

P-value dan keputusan hipotesis

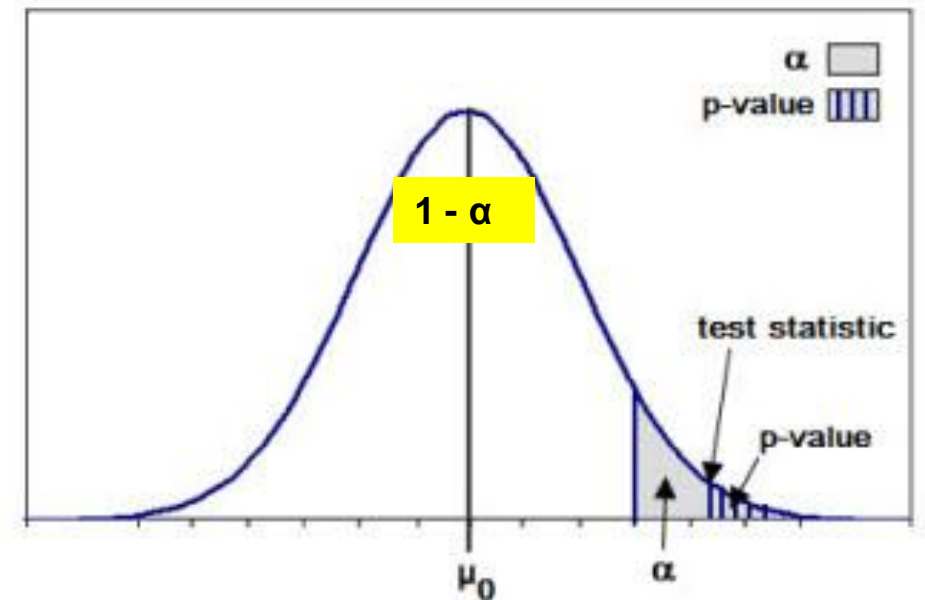
P-value adalah suatu nilai signifikansi untuk menguji hipotesis.

Pengambilan keputusan ini :

- ❑ Jika $p\text{-value} \leq \alpha$, maka H_0 ditolak
- ❑ Jika $p\text{-value} > \alpha$, maka H_0 diterima

Misalkan $\alpha=0.05$, artinya kamu **mentolerir adanya peluang terjadinya error tipe I sebesar 5%**.

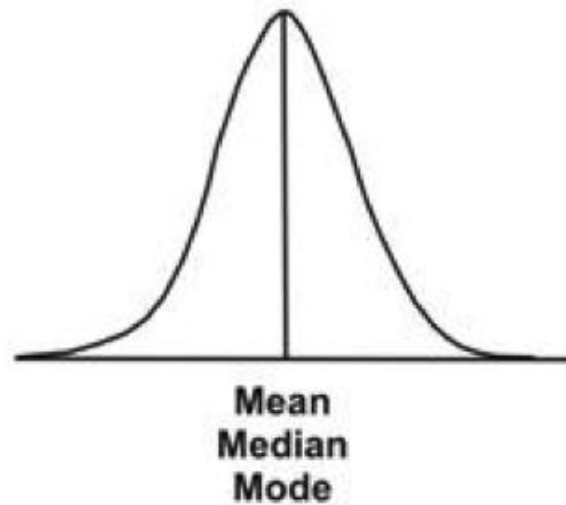
Saat misalnya kamu dapati nilai $p\text{-value} = 0.03$, maka kamu menolak H_0 dengan **menerima adanya kemungkinan kesalahan membuat keputusan sebesar 3%**



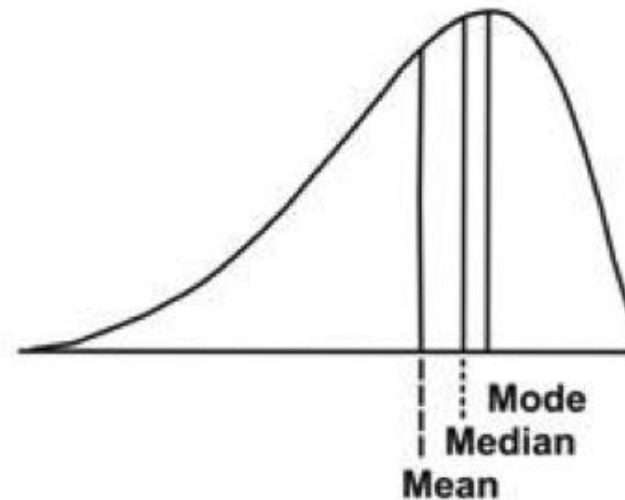
Distribusi normal

- Distribusi Normal adalah distribusi data dimana nilai modus, mean, dan mediannya mendekati nilai yang sama

Berdistribusi normal



Tidak berdistribusi normal

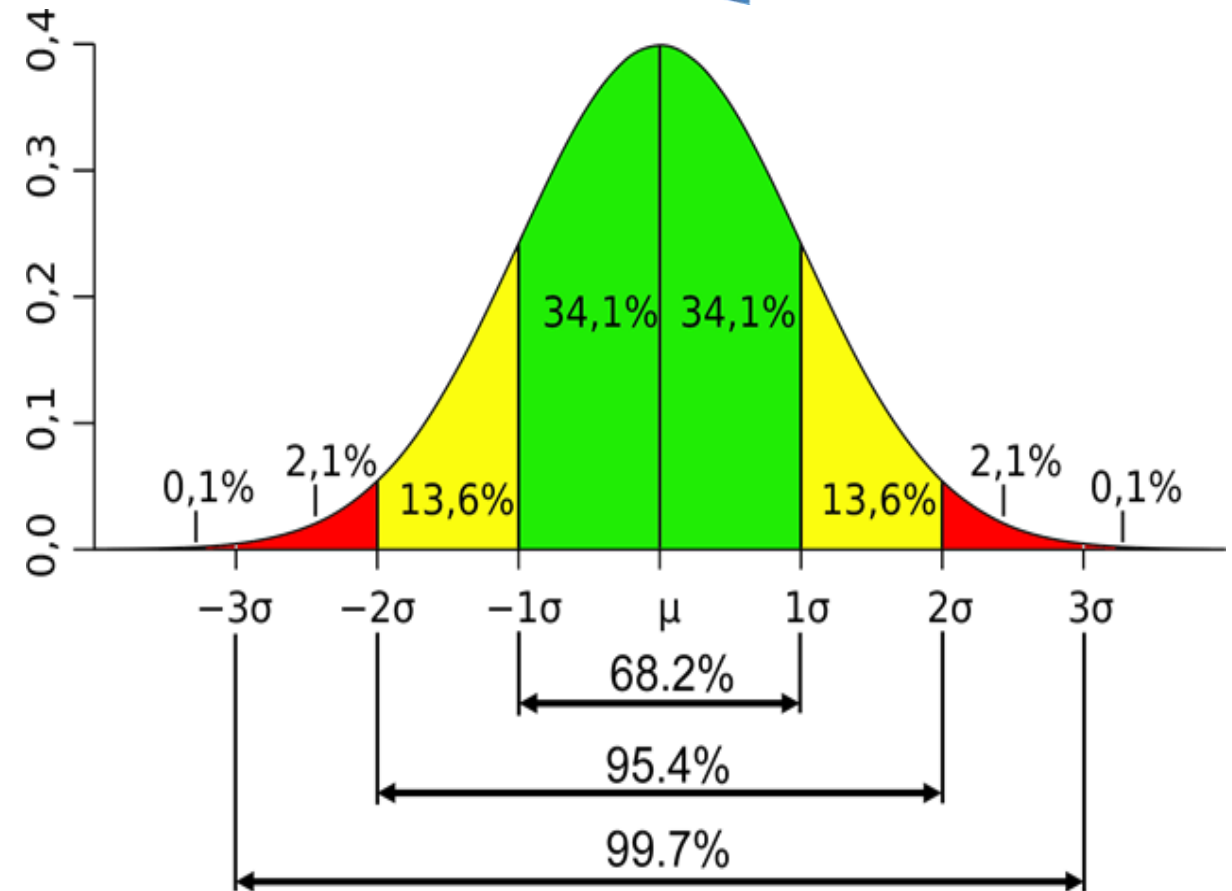


Sifat-sifat Distribusi Normal

Sifat-sifat distribusi normal adalah sebagai berikut:

- Kurva normal simetris terhadap mean
- Mean berada di tengah dan membagi area menjadi dua (bentuk histogram seperti lonceng)
- 68% data tersebar merata di sekitar mean \pm standar deviasinya,

95% tersebar di mean ± 2 *st. deviasi, dan 99% tersebar di mean ± 3 *standar deviasi



Mengapa perlu tahu distribusi normal?

- Beberapa uji statistik mengasumsikan data berdistribusi normal
- Cara menguji normalitas data dengan visualisasi :

Cek apakah histogram membentuk lonceng dengan mean pas di tengah atau tidak

- Cara menguji normalitas data dengan uji Kolmogorov-Smirnov :
- Di python, uji ini ada di library scipy :

```
from scipy import stats  
z_score = stats.zscore(data)  
stats.kstest(z_score, "norm")
```

Ho : data berdistribusi normal

p-value > alpha (misalnya 0.05) maka Ho diterima, data berdistribusi normal



02

UJI HIPOTESIS

- uji T
- Tes Chi Square
- ANOVA

UJI BEDA

- Uji beda (comparative analysis) adalah uji statistik untuk menguji pengaruh suatu perlakuan atau perbedaan dari beberapa populasi
- Contoh : pengaruh wfh terhadap kesehatan mental karyawan. Maka kita menganalisis perbedaan kesehatan mental karyawan saat wfo dan wfh
- Contoh lainnya, analisa pemerataan pendidikan di Indonesia. Maka kita membandingkan nilai UN siswa SMA dari tiap provinsi, lalu menganalisa apakah perbedaannya signifikan atau tidak
- Jika data berdistribusi normal, maka statistik yang digunakan adalah parametrik, jika tidak normal maka pilih statistik nonparametrik

Statistik Parametrik – Uji-T



- Asumsi penggunaan Uji-T adalah data harus berdistribusi normal
- Uji-t digunakan untuk membandingkan 2 kelompok, terdapat 2 jenis kasus:

1. Sampel Berhubungan (Paired Sample T-Test)

Yaitu datanya bersumber dari sampel yang sama, misalnya perbedaan kepuasan mahasiswa saat belajar online dengan ms teams dan zoom (apakah tools belajar online mempengaruhi kepuasan mahasiswa?)

2. Sampel independent (Independent Sample T-Test)

Yaitu datanya bersumber dari sampel yang berbeda, misalnya analisis perbedaan kemampuan matematika siswa sekolah swasta dan negeri

Statistik Parametrik – ANOVA

- Analisis of Variance (ANOVA) digunakan untuk membandingkan lebih dari 2 kelompok
- Asumsi dari ANOVA adalah data harus berdistribusi normal dan variansnya harus homogen
- Pengujian homogen ini bisa dilakukan dengan uji Bartlet atau Levene. Kedua uji ini tersedia di library scipy

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Keterangan:
s²: Varian s : Standar deviasi

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Statistik Nonparametrik

- Jika asumsi pada statistik parametrik tidak terpenuhi bukan berarti kita tidak bisa menganalisa data, kita bisa menggunakan statistik nonparametrik :

	Data	Asumsi	2 kelompok berhubungan	2 kelompok independen	> 2 kelompok
Parametrik	Numerik	Normalitas	Paired Sample T-test	Independent Sample T-test	ANOVA
Non-Parametrik	Numerik	-	Wilcoxon	Mann-Whitney	Kruskal Wallis, Friedman
	Ordinal	-	Wilcoxon	Mann-Whitney	Kruskal Wallis, Friedman
	Nominal	-	Mc Nemar	Chi-Square	Chi-Square



03

KESIMPULAN

- Ulangan
- Ringkasan

Ringkasan

- Uji Hipotesis
- P-value
- Selang kepercayaan
- H_0 dan H_a
- Distribusi Normal
- Uji-T
- ANOVA
- Statistik nonparametrik

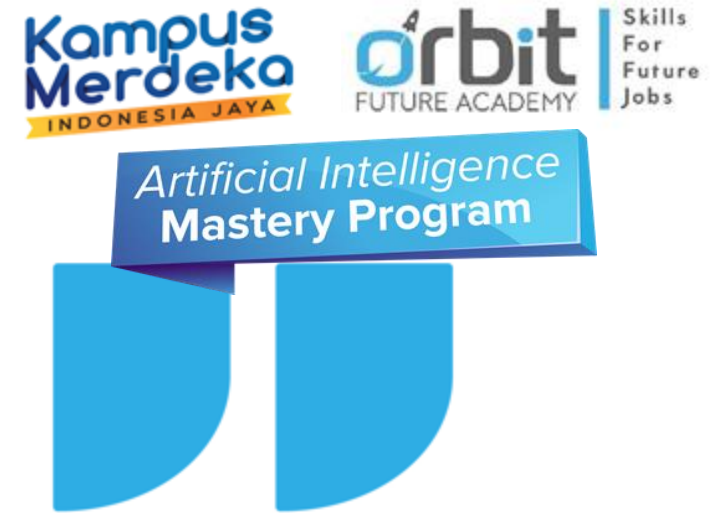


Review

Pertanyaan

Jika seseorang ingin mengetahui perbedaan kondisi kesehatan mental mahasiswa sebelum dan saat pandemi, yang manakah uji yang sesuai?

- A. ANOVA – uji f
- B. uji T sampel independent
- C. uji T sampel berhubungan
- D. Tes chi kuadrat



Review

Pertanyaan

Jika seseorang ingin mengetahui perbedaan kondisi kesehatan mental mahasiswa sebelum dan saat pandemik, yang manakah uji yang sesuai?

- A. ANOVA – uji f
- B. uji T sampel independent
- C. uji T sampel berhubungan**
- D. Tes chi kuadrat

Jawaban : C, karena walaupun datanya ada 2 (sebelum dan saat pandemi) tapi bersumber dari sampel yang sama



**Kampus
Merdeka**
INDONESIA JAYA

orbit
FUTURE ACADEMY

Skills
For
Future
Jobs

*Artificial Intelligence
Mastery Program*

TERIMA KASIH

Orbit Future Academy

PT Orbit Ventura Indonesia
Center of Excellence (Jakarta Selatan)
Gedung Veteran RI, Lt.15
Unit Z15-002, Plaza Semanggi
Jl. Jenderal Sudirman Kav.50, Jakarta
12930, Indonesia

- 📍 Jakarta Selatan/Pusat
- 📍 Jakarta Barat/BSD
- 📍 Kota Bandung
- 📍 Kab. Bandung
- 📍 Jawa Barat

Hubungi Kami

Director of Sales & Partnership
ira@orbitventura.com
+62 858-9187-7388

Social Media

 Orbit Future Academy

 OrbitFutureAcademy

 @OrbitFutureAcademyIn1

 Orbit Future Academy