Bab 1

Analisis Variansi satu faktor (Analysis Of Variance / ANOVA)

Learning Objectives

1. Design and conduct experiments involving a single
2. Understand how the anova is used to analyze the data from these experiments
3. Assess model adequacy with residual plots
4. Use multiple comparison procedures to identify specific differences between means
5. Make decisions about sample in 1 factor experiments

What is Analysis of variance (ANOVA)

→Versatile statistical tool for studying the relation between a dependent variable and one or more independent variable

→Dapat digunakan pada data yang diperoleh dari hasil eksperimen dan observasi

→ANOVA adalah suatu metoda untuk menguji hipotesis kesamaan rata-rata dari tiga atau lebih populasi

→What different ANOVA with regression ?

ANOVA → Variabel independen ; kualitatif

- *Analysis of variance* (ANOVA) digunakan untuk menyelidiki pengaruh/ efek utama dan interaksi dari variabel independen (disebut dengan "faktor")

- Pengaruh utama adalah efek langsung dari suatu variabel independen terhadap variabel dependen

- Pengaruh interaksi adalah efek bersama antar satu atau lebih variabel independen terhadap variabel dependen

- Model regresi tidak dapat meng-cover interaksi sedangkan ANOVA bisa meng-cover pengaruh interaksi

LATAR BELAKANG ANOVA

Latar belakang dikembangkan metoda ini karena ingin dilakukan UJI terhadap rata-rata populasi yg mengalami “perlakuan” yg berbeda-beda.

Pertanyaannya : apakah perbedaan rata-rata antara berbagai grup yg mengalami perlakuan berbeda tsb signifikan atau tidak.

Asumsi untuk uji ANOVA adalah:

→Populasi semuanya normal

→Standard deviasi populasi sama

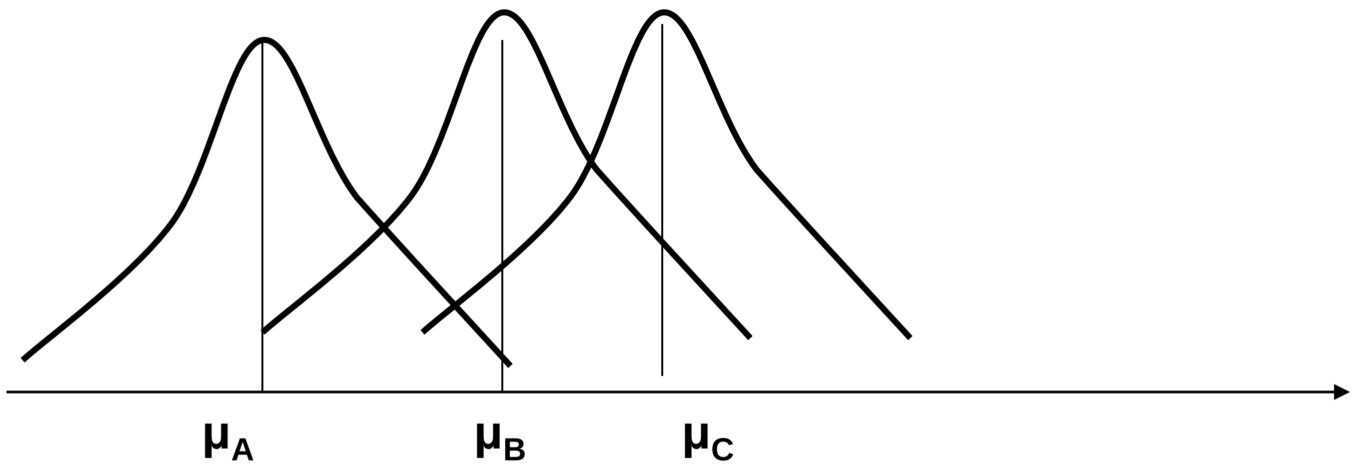
→Populasi independen

Contoh

Tiga kelompok subyek penelitian untuk menguji keakuratan alat pengukur pH digital dengan 3 model. Model yang dimaksud adalah model I, II dan III.. Data hasil penelitian adalah sebagai berikut:

I	II	III
25	17	26
11	16	20
16	18	17
26	20	26
32	10	43
25	14	46
30	19	35
17		34
		18

Ide – Uji ANOVA



Ide dasar test ANOVA adalah perbedaan rata-rata populasi ditentukan oleh dua faktor yaitu variasi data dalam 1 sampel dan variasi data antar sampel. Perbedaan rata-rata antar populasi nyata jika variasi data antar sampel besar sedangkan variasi data dalam 1 sampel kecil.

Single Factor Analysis of Variance – Anava satu Jalan

Disebut dengan *one- way* atau *single factor analysis of variance* , **do you know why?**
→**Hanya satu faktor perlakuan yang diselidiki**
Perlakuan yang digunakan diusahakan **se-seragam mungkin**,
→*completely randomized design* (Rancangan Random Lengkap)

- Secara umum, jika *n* observasi dikenakan a perlakuan maka model linier statistik :

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases}$$

$$\Rightarrow y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases}$$

dengan $\mu_i = \mu + \tau_i$

--> rata-rata perlakuan ke-i

Jika perlakuan dipilih ttt oleh eksperimenter maka kesimpulan uji tidak bisa digeneralisasikan untuk populasi perlakuan → MODEL EFEK TETAP

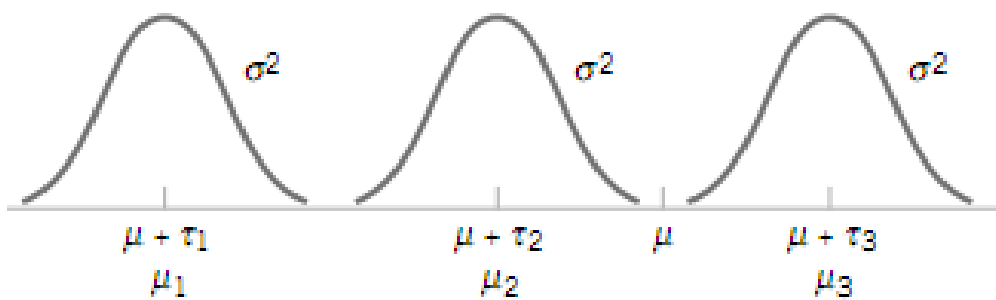
Jika perlakuan dipilih random dari populasi perlakuan oleh eksperimenter maka kesimpulan uji dapat digeneralisasikan ke seluruh populasi perlakuan → MODEL EFEK RANDOM/ *components of variance model*

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases} \Rightarrow y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases}$$

$$\text{dengan } \mu_\text{i} = \mu + \tau_i$$

$$\text{-->}\ \text{rata-rata perlakuanke-}i$$

that each treatment defines a population that has mean μ_i, consisting of the overall mean μ plus an effect τ_i that is due to that particular treatment. We will assume that the errors ϵ_{ij} are normally and independently distributed with mean zero and variance σ^2. Therefore, each treatment can be thought of as a normal population with mean μ_i and variance σ^2.



y_{ij} : observasi ke (ij)

μ : rata-rata keseluruhan perlakuan

τ_i : pengaruh/efek perlakuan ke-i

ε_{ij} : sesatan dengan asumsi NID $(0, \sigma^2)$

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases} \Rightarrow y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases}$$

$$\text{dengan } \mu_\mathrm{i} = \mu + \tau_i$$

$$\text{-->} \text{ rata-rata perlakuan ke-}i$$

Tujuan ANAVA satu jalan :

melakukan uji hipotesis tentang efek perlakuan dan mengestimasinya

Asumsi

- Sampel diambil secara random dan saling bebas (independen)
- Populasi berdistribusi berdistribusi Normal
- Populasi mempunyai kesamaan variansi

Computerized – Uji Pra Analisis

1. Normalitas

   Jika asumsi sesatan $(0, \sigma^2)$ dipenuhi maka plot normalitas nampak seperti sampel yang berasal dari distribusi normal yang berpusat ke 0 yang ditunjukkan dengan sebaran data yang cenderung membentuk garis lurus

2. Independensi

   Yaitu plot antara residual data dengan $\hat{y}_{ij}$, asumsi dipenuhi jika sebaran data cenderung tidak membentuk pola tentu dan acak

3. Homogenitas

   Yaitu plot antara residual data dengan urutan data, asumsi dipenuhi jika sebaran data cenderung tidak membentuk pola tentu dan acak

Perbedaan Asumsi Model Tetap dan Random

Model Efek Tetap

a. $\sum_{i=1}^{a} \lambda_i = 0$,

b. Sesatan diasumsikan berdistribusi Normal dan Independen atau dinotasikan $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$.

Model Efek Random

a. λ_i diasumsikan berdistribusi Normal dan Independen atau dinotasikan $\lambda_i \sim \text{NID}(0, \sigma_\lambda^2)$,

b. Sesatan diasumsikan berdistribusi Normal dan Independen atau dinotasikan $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$,

c. λ_i dan ε_{ij} Independen.

Tabel data untuk percobaan faktor tunggal

Perlakuan ke-	Observasi				Total	Rata-rata
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\bullet}$	$\bar{y}_{1\bullet}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\bullet}$	$\bar{y}_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\bullet}$	$\bar{y}_{a\bullet}$
Jumlah					$y_{\bullet\bullet}$	$\bar{y}_{\bullet\bullet}$

Keterangan :

$$y_{i\bullet} = \sum_{i=1}^{n} y_{ij}, \quad y_{\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{i\bullet} = \frac{y_{i\bullet}}{n}, \quad \bar{y}_{\bullet\bullet} = \frac{y_{\bullet\bullet}}{N}, \; i = 1, \ldots, a$$

$$N = an$$

Prosedur Uji Model Efek Tetap

i. Asumsi : $\sum_{i=0}^{a} \tau_i = 0$ $\quad$ $y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,..,a \\ j = 1,..,n \end{cases}$

ii. Hipotesis:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \tau_i \neq 0 (\text{setidaknya satu } i)$$

Thus, if the null hypothesis is true, each observation consists of the overall mean μ plus a realization of the random error component ϵ_{ij}. This is equivalent to saying that all N observations are taken from a normal distribution with mean μ and variance σ^2. Therefore, if the null hypothesis is true, changing the levels of the factor has no effect on the mean response.

Prosedur ANOVA

iii. Penentuan Tabel ANAVA

Partisi Jumlah Kuadrat (JK)

$$y_{ij} - \bar{y}_{\bullet\bullet} = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet} + y_{ij} - \bar{y}_{i\bullet}$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left\{y_{ij} - \bar{y}_{\bullet\bullet}\right\}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left\{\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet} + y_{ij} - \bar{y}_{i\bullet}\right\}^2$$

$$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left\{y_{ij} - \bar{y}_{\bullet\bullet}\right\}^2}_{\mathrm{JK_T}} = \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left\{\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right\}^2}_{\mathrm{JK_P}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left\{y_{ij} - \bar{y}_{i\bullet}\right\}^2}_{\mathrm{JK_S}}$$

$$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{\bullet\bullet}\right)^2}_{\mathrm{JK_T}}$$

$$=\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)^2}_{\mathrm{JK_P}}+\underbrace{2\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)\left(y_{ij}-\bar{y}_{i\bullet}\right)}_{=0}+\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2}_{\mathrm{JK_S}}$$

Partisi JK

Beberapa definisi variasi.

1. Variasi Total

   Jumlah total kuadrat selisih data dengan rata-rata total seluruh data (overall mean)

$$\mathrm{JK_T} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left\{y_{ij} - \bar{y}_{\bullet\bullet}\right\}^2 = \sum_{i}^{a}\sum_{j}^{n} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$$

2. Variasi Antar Sampel (atau Variasi karena Perlakuan)

   Jumlah total kuadrat selisih rata-rata tiap sampel thd rata-rata total (grand mean)

$$\mathrm{JK_P} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left\{\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right\}^2 = \sum_{i=1}^{a}\frac{y_{i\bullet}^2}{n} - \frac{y_{\bullet\bullet}^2}{N}$$

TEST ANOVA – Macam Variasi

Beberapa definisi variasi.

3. Variasi Random

   Jumlah total kuadrat selisih data dengan rata-rata sampel yg terkait

$$\mathrm{JK_S} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left\{y_{ij} - \bar{y}_{i\bullet}\right\}^2 = \mathrm{JK_T} - \mathrm{JK_P}$$

The expected value of the treatment sum of squares is

$$E(SS_{\text{Treatments}}) = (a-1)\sigma^2 + n\sum_{i=1}^{a}\tau_i^2$$

and the expected value of the error sum of squares is

$$E(SS_E) = a(n-1)\sigma^2$$

iii. Tabel ANOVA 1 Jalan _ Cont..

Sumber Variansi	JK	db	RK	Fo
Perlakuan	JKP	a-1	RKP=JKP/(a-1)	Fp=RKP/RKS
Sesatan	JKS	a(n-1)	RKS=JKS/a(n-1)	
Total	JKT	an-1		

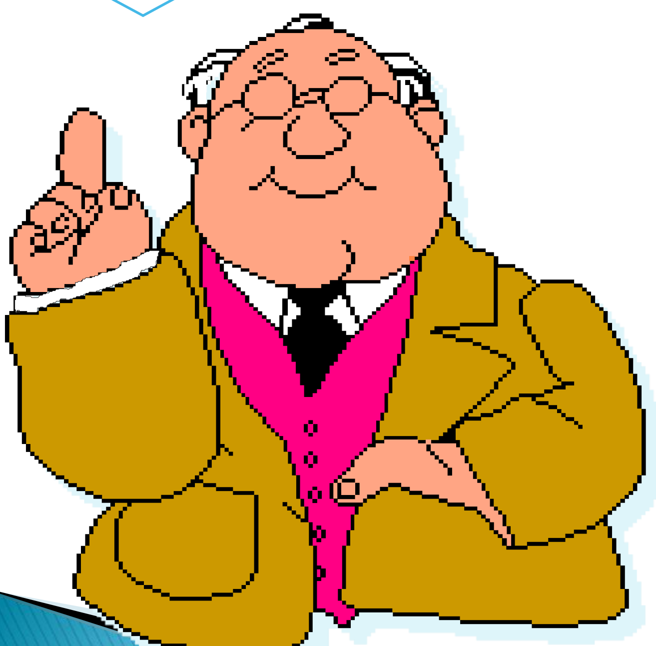
iv. Daerah Kritis

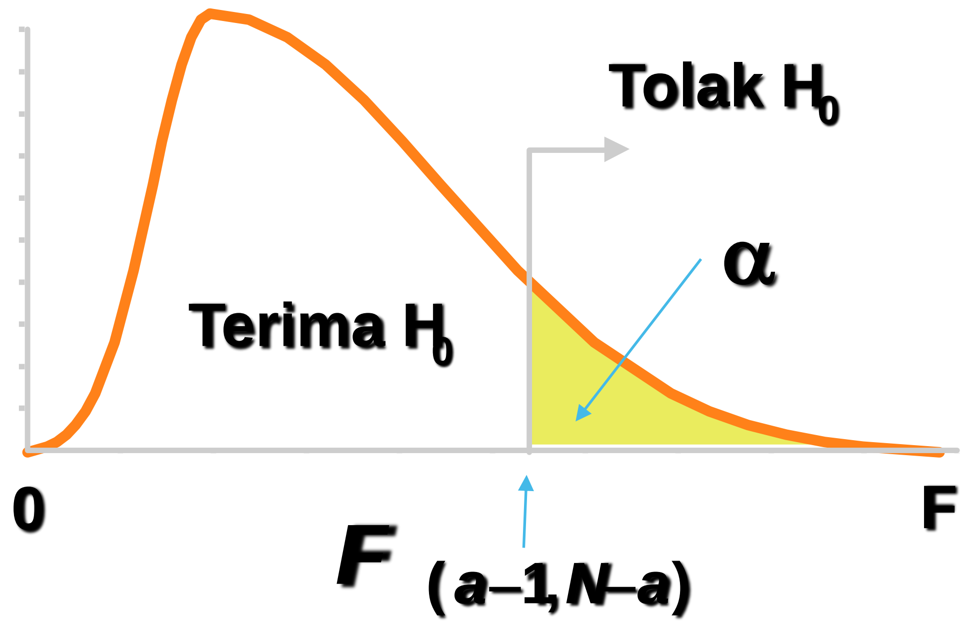
$$\text{Tolak } H_0 \text{ jika } Fp > F_{db(perlakuan),db(sesatan)}$$

$$\text{Tolak } H_0 \text{ jika } Fp > F_{(a-1),(a(n-1))}$$

One-Way ANOVA F-Test Critical Value

If means are equal, $F = MST / MSE \approx 1$. Only reject large F!





Always One-Tail!