

Bab X

Analysis of Variance ANOVA

Contoh penjual mobil

Penjual mobil tersebut kemudian mengambil sampel secara acak dari tiap populasi dan mendapatkan informasi sebagai berikut:

<u>Level</u>	<u>n</u>	<u>Mean</u>	<u>StDev</u>
Mid-size	28	27.101 mpg	2.629 mpg
SUV	26	20.423 mpg	2.914 mpg

Uji t-test dari kedua data diatas akana menghasilkan :

T-statistik = 8.15 P-value = ~0

Apa uji ANOVA itu?

- Uji-t untuk 2-Mean
 - Contoh: seorang penjual mobil ingin mendapatkan perbedaan antara 2 jenis kendaraan dalam hal konsumsi gas perkilo:
 - Kendaraan kecil
 - Kendaraan SUV

Bagaimana jika penjual mobil tersebut ingin menambahkan satu lagi jenis kendaraan, yaitu Pickup Trucks sebagai tambahan pada jenis SUV's dan kendaraan kecil?

<u>Level</u>	<u>n</u>	<u>Mean</u>	<u>StDev</u>
Midsize	28	27.101 mpg	2.629 mpg
SUV	26	20.423 mpg	2.914 mpg
Pickup	8	23.125 mpg	2.588 mpg

Untuk itu kita harus menggunakan uji ANOVA .

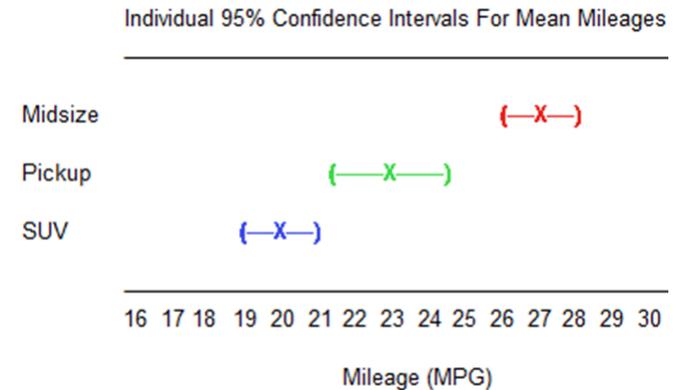
Dalam uji t untuk 2-Mean, ingin dilihat apakah

Beda antara mean 2 sample adalah signifikan.

Dalam ANOVA perbandingan dilihat pada beberapa means (lebih dari 2) dan ingin dilihat apakah:

Beda antara beberapa sampel adalah signifikan.

Perbandingan Mean



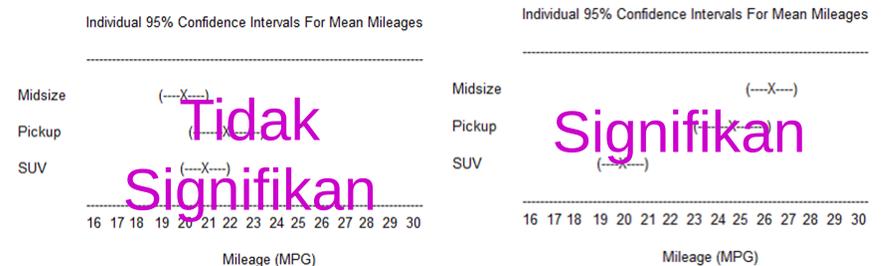
Apakah mean sample diatas terlihat berbeda signifikan?

Perbandingan Mean

Ya, bisa dilihat bahwa tidak ada selang kepercayaan yang saling tindih satu sama lain, artinya mean antar sampel berbeda secara signifikan.

Uji secara matematis diberikan oleh uji ANOVA .

Selang Kepercayaan



Bagaimana jika selang kepercayaannya berbeda? Apakah selang kepercayaan ini pertanda bahwa mean berbeda signifikan?

Uji ANOVA

$H_0: \mu_1 = \mu_2 = \mu_3$ (semua mean sama)

H_A : Tidak semua mean sama

contoh:

$H_0: \mu_{\text{Mid-size}} = \mu_{\text{SUV}} = \mu_{\text{Pickup}}$

Mean dari konsumsi gas perkilo dari kendaraan kecil, SUV dan truk adalah sama.

- H_A : Tidak semua mean dari konsumsi gas perkilo sama.

Statistik F

- Seperti uji yang lain, uji ANOVA juga mempunyai uji statistik tersendiri
- Statistik untuk ANOVA disebut statistik F, yang kita dapatkan dari distribusi F
- Statistik F meliputi:
 - **Banyaknya sampel** yang diambil (l)
 - **Ukuran sampel** dari tiap sampel (n_1, n_2, \dots, n_l)
 - **mean** dari sampel ($\bar{x}_1, \bar{x}_2, \dots, \bar{x}_l$)
 - **standar deviasi** dari tiap sampel (s_1, s_2, \dots, s_l)

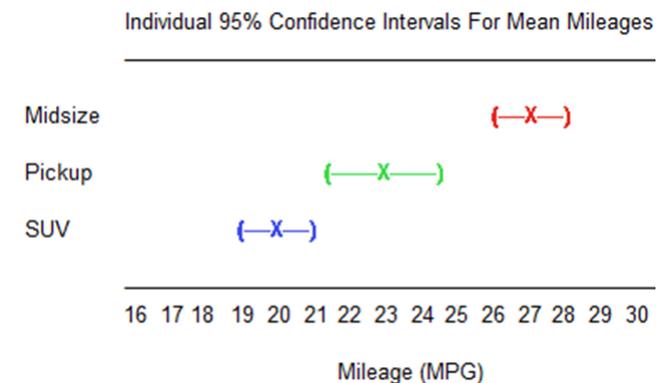
Penjelasan tentang Statistik F

Statistik F menjelaskan apakah variasi antara sampel mean signifikan

Statistik F = variasi antar mean sampel/variasi antara individu dalam sampel

Hal inilah yang akan dilihat pada saat kita melihat 95% selang kepercayaan.

Selang Kepercayaan



Dari gambar terlihat bahwa variasi antar sampel lebih dari variasi dalam sampel, maka nilai F akan besar.

Persamaan Statistik F

$$F = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \dots + n_I(\bar{x}_I - \bar{x})^2}{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_I - 1)s_I^2} \cdot \frac{I - 1}{N - I}$$

Mean (Kuadrat) points to the numerator. Standar Deviasi (Kuadrat) points to the denominator. Bobot points to the weights n_i .

Statistik F

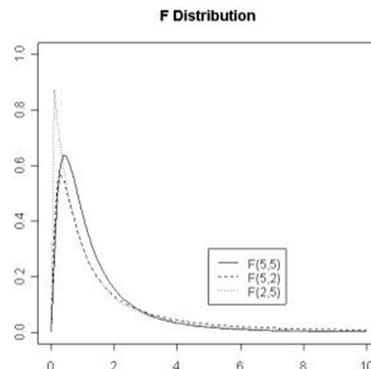
$$F = \frac{n_{Midsize}(\bar{x}_{Midsize} - \bar{x})^2 + n_{SUV}(\bar{x}_{SUV} - \bar{x})^2 + n_{Pickup}(\bar{x}_{Pickup} - \bar{x})^2}{(n_{Midsize} - 1)s_{Midsize}^2 + (n_{SUV} - 1)s_{SUV}^2 + (n_{Pickup} - 1)s_{Pickup}^2} \cdot \frac{I - 1}{N - I}$$

$$= \frac{28(27.101 - 23.786)^2 + 26(20.423 - 23.786)^2 + 8(23.125 - 23.786)^2}{(28 - 1)2.629^2 + (26 - 1)2.914^2 + (8 - 1)2.588^2} \cdot \frac{3 - 1}{62 - 3}$$

$$= 40.053$$

Derajat Kebebasan

- Uji ANOVA mempunyai 2 derajat kebebasan:
 - N-I (Total banyaknya sampel – banyaknya grup)
 - I-1 (banyaknya grup- 1)
- Contoh distribusi F dengan beberapa derajat kebebasan:



Contoh Penjual Mobil:

Data:

Level	n	Mean	StDev
Midsize	28	27.101 mpg	2.629 mpg
SUV	26	20.423 mpg	2.914 mpg
Pickup	8	23.125 mpg	2.588 mpg

F = 40.05

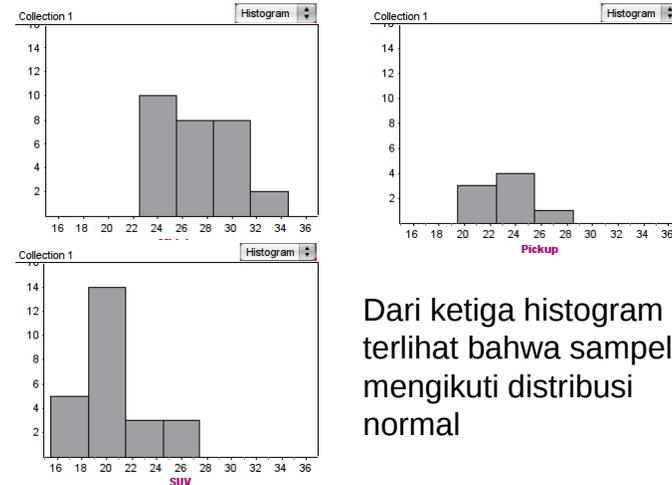
P-value = ~0 (didapat dari tabel atau keluaran perangkat lunak).

Syarat (Kondisi)

Uji ANOVA bisa dipakai jika syarat (kondisi) dibawah ini terpenuhi:

- Sample yang diambil dari tiap populasi harus berupa **sampel yang diambil secara acak**
- Semua populasi harus mempunyai **standar deviasi yang sama** (tidak ada satu populasi yang mempunyai standar deviasi dua kali dari yang lain)
- Semua populasi harus **berdistribusi normal**

Histogram dari Sampel



Dari ketiga histogram terlihat bahwa sampel mengikuti distribusi normal

Contoh soal lain

- Seorang peneliti ingin melihat apakah nilai bahasa Inggris dari 4 sekolah terkemuka berbeda satu sama lain. Dari keempat sekolah tersebut, nilai bahasa Inggris dari siswa terdahulu diambil sebagai sampel. Hasil dari analisis sampel adalah:

hasil

<u>Sekolah</u>	<u>n</u>	<u>Mean</u>	<u>StDev</u>
A	23	4.3	0.4
B	25	3.9	0.6
C	26	4.2	0.3
D	29	3.1	0.3

Apakah terdapat perbedaan yang signifikan dari nilai bahasa Inggris keempat sekolah tersebut? (Assumsikan bahwa populasi berdistribusi normal)

Hypothesis

- $H_0: = \mu_A = \mu_B = \mu_C = \mu_D$
Mean dari nilai bahasa Inggris keempat sekolah adalah sama.
- H_A : satu dari mean nilai bahasa Inggris keempat sekolah tersebut tidak sama.

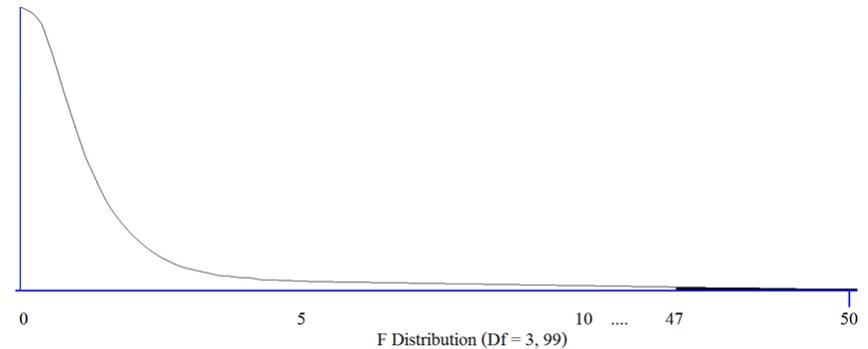
Kondisi

- Sampel acak yang diambil
- Semua standar deviasi adalah sama
 - Tidak ada standar deviasi yang dua kali standar deviasi yang lain.
- Semua populasi berdistribusi normal

Statistik F

$$F = \frac{\frac{n_{BB\&N}(\bar{x}_{BB\&N} - \bar{x})^2 + n_{RL}(\bar{x}_{RL} - \bar{x})^2 + n_{Winsor}(\bar{x}_{Winsor} - \bar{x})^2 + n_{BelHill}(\bar{x}_{BelHill} - \bar{x})^2}{I - 1}}{\frac{(n_{BB\&N} - 1)s_{BB\&N}^2 + (n_{RL} - 1)s_{RL}^2 + (n_{Winsor} - 1)s_{Winsor}^2 + (n_{BelHill} - 1)s_{BelHill}^2}{N - I}}$$
$$= \frac{23(4.3 - 3.84)^2 + 25(3.9 - 3.84)^2 + 26(4.2 - 3.84)^2 + 29(3.1 - 3.84)^2}{\frac{4 - 1}{(23 - 1) \cdot 4^2 + (25 - 1) \cdot 6^2 + (26 - 1) \cdot 3^2 + (29 - 1) \cdot 3^2}}$$
$$= 47.18$$

Kurva F



penafsiran

Karena semua syarat dipenuhi, dan berdasarkan statistik F ($df = 3,99$, $p = 0$) maka cukup bukti untuk menolak hipotesis null, maka paling tidak satu mean dari nilai bahasa Inggris keempat sekolah tidak sama.