

Analysis of Variance (ANOVA)

Dr. Budi Marpaung, ST., MT.



Pengantar

Dalam pengujian selisih rata-rata diasumsikan bahwa varians kedua populasi adalah sama.

Dalam kenyataannya asumsi itu perlu diuji atau dibuktikan terlebih dahulu kebenarannya.

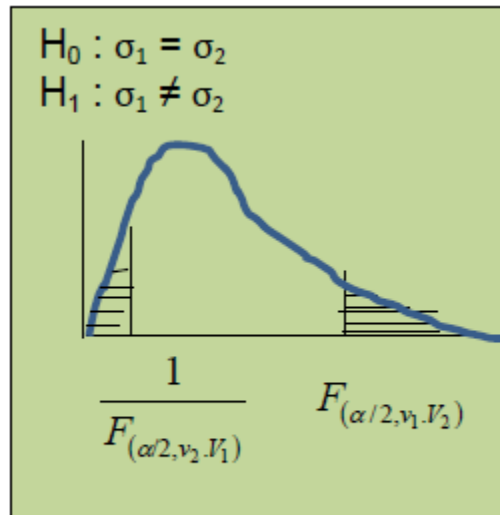
Pengujian kesamaan varians menggunakan Distribusi F, yang dikembangkan oleh R.A. Fisher pada awal dasawarsa 1920-an.

Uji Kesamaan Varians

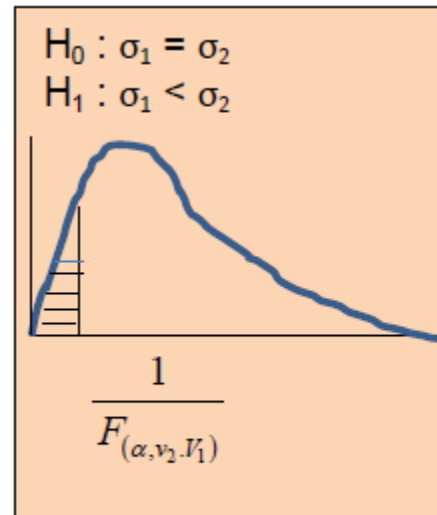
Jika dari dua populasi yang memiliki distribusi normal diambil dua sampel independen dengan ukuran n_1 dan n_2 , maka varians populasi dan diduga dengan varians sampel S_1^2 dan S_2^2 , yaitu :

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} \quad S_2^2 = \frac{\sum (x_i - \bar{x}_2)^2}{n_2 - 1} \quad F_{hitung} = \frac{S_1^2}{S_2^2}$$

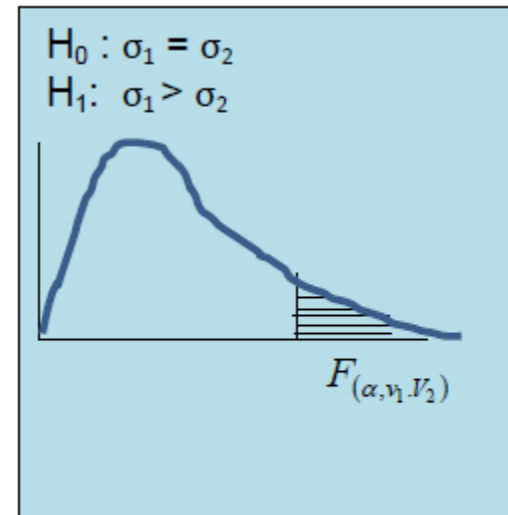
Daerah Kritis Uji Hipotesis



Dwi Sisi



Eka Sisi Kiri



Eka Sisi Kanan

Derajat kebebasan v_1 (pembilang) = $(n_1 - 1)$
Derajat kebebasan v_2 (penyebut) = $(n_2 - 1)$

Soal 1 : Uji Kesamaan Varians

Terdapat dua kelompok sampel yang berasal dari populasi yang berbeda. Sampel pertama sebanyak 17 sampel memiliki rata-rata 1200 dan simpangan baku 60. Sedangkan sampel kedua sebanyak 21 sampel memiliki rata-rata 1300 dan simpangan baku 50. Dengan data tersebut dapatkah disimpulkan bahwa varians populasi 1 lebih besar dari varians populasi 2? Gunakan tingkat kesalahan 5%.

Solusi Soal 1

1. $H_0 : \sigma_1 = \sigma_2$ atau $\sigma_1 \leq \sigma_2$

$H_1 : \sigma_1 > \sigma_2$

2. $\alpha = 0.05$

3. Statistik hitung yang digunakan adalah nilai F

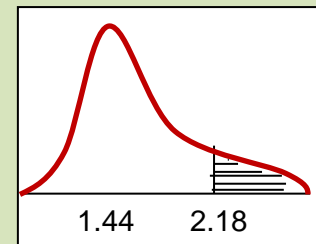
4. Daerah kritis

Dengan memperhatikan H_0 dan H_1 dan nilai ,
maka uji ini termasuk ekasisi kanan, nilai kritisnya
adalah $F_{(0.05;v1=17-1;v2=21-1)} = F_{(0.05;v1=16;v2=20)} = 2.18$

Solusi Soal 1 (Lanj....)

5. Nilai hitung

$$F_{hitung} = \frac{S_1^2}{S_2^2} = \frac{3600}{2500} = 1.44$$



6. Kesimpulan

H_0 diterima (H_1 ditolak)

Varians kedua populasi adalah sama, atau ditolak hipotesis yang menyatakan bahwa varians populasi 1 lebih besar dari varians populasi 2.

Latihan Soal 1

Terdapat dua kelompok sampel yang berasal dari populasi yang berbeda. Sampel pertama sebanyak 20 sampel memiliki rata-rata 1000 dan simpangan baku 75. Sedangkan sampel kedua sebanyak 25 sampel memiliki rata-rata 1100 dan simpangan baku 70. Dengan data tersebut dapatkah disimpulkan bahwa varians kedua populasi adalah sama? Gunakan tingkat kesalahan 5%.

Latihan Soal 2

Terdapat dua kelompok sampel yang berasal dari populasi yang berbeda, dengan data sbb.

Sampel 1: 80, 73, 75, 84, 82, 85, 87, 77, 80, 78

Sampel 2: 81, 74, 70, 80, 85, 80, 82, 75, 70, 72, 75

Dengan data tersebut dapatkah disimpulkan bahwa varians kedua populasi adalah sama? Gunakan tingkat kesalahan 5%.

Bartlett's test

Uji kesamaan lebih dari dua varians

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

against the alternative

H_1 : The variances are not all equal.

The test that we shall use, called **Bartlett's test**, is based on a statistic whose sampling distribution provides exact critical values when the sample sizes are equal. These critical values for equal sample sizes can also be used to yield highly accurate approximations to the critical values for unequal sample sizes.

First, we compute the k sample variances $s_1^2, s_2^2, \dots, s_k^2$ from samples of size n_1, n_2, \dots, n_k , with $\sum_{i=1}^k n_i = N$. Second, we combine the sample variances to give the pooled estimate

$$s_p^2 = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) s_i^2.$$

Now

$$b = \frac{[(s_1^2)^{n_1-1} (s_2^2)^{n_2-1} \dots (s_k^2)^{n_k-1}]^{1/(N-k)}}{s_p^2}$$

is a value of a random variable B having the **Bartlett distribution**. For the special case where $n_1 = n_2 = \dots = n_k = n$, we reject H_0 at the α -level of significance if

$$b < b_k(\alpha; n),$$

When the sample sizes are unequal, the null hypothesis is rejected at the α -level of significance if

$$b < b_k(\alpha; n_1, n_2, \dots, n_k),$$

where

$$b_k(\alpha; n_1, n_2, \dots, n_k) \approx \frac{n_1 b_k(\alpha; n_1) + n_2 b_k(\alpha; n_2) + \dots + n_k b_k(\alpha; n_k)}{N}.$$

As before, all the $b_k(\alpha; n_i)$ for sample sizes n_1, n_2, \dots, n_k are obtained from Table A.10.

Contoh Soal Bartlett's test

Example 13.3: Use Bartlett's test to test the hypothesis at the 0.01 level of significance that the population variances of the four drug groups of Example 13.2 are equal.

Solution: We have the hypotheses

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2,$$

H_1 : The variances are not equal,

with $\alpha = 0.01$.

Critical region: Referring to Example 13.2, we have $n_1 = 20$, $n_2 = 9$, $n_3 = 9$, $n_4 = 7$, $N = 45$, and $k = 4$. Therefore, we reject when

$$\begin{aligned} b &< b_4(0.01; 20, 9, 9, 7) \\ &\approx \frac{(20)(0.8586) + (9)(0.6892) + (9)(0.6892) + (7)(0.6045)}{45} \\ &= 0.7513. \end{aligned}$$

Computations: First compute

$$s_1^2 = 662.862, s_2^2 = 2219.781, s_3^2 = 2168.434, s_4^2 = 946.032,$$

and then

$$\begin{aligned} s_p^2 &= \frac{(19)(662.862) + (8)(2219.781) + (8)(2168.434) + (6)(946.032)}{41} \\ &= 1301.861. \end{aligned}$$

Now

$$b = \frac{[(662.862)^{19}(2219.781)^8(2168.434)^8(946.032)^6]^{1/41}}{1301.861} = 0.8557.$$

Decision: Do not reject the hypothesis, and conclude that the population variances of the four drug groups are not significantly different. \blacksquare

Table 13.4: Serum Alkaline Phosphatase Activity Level

G-1	G-2	G-3	G-4
49.20	97.50	97.07	62.10
44.54	105.00	73.40	94.95
45.80	58.05	68.50	142.50
95.84	86.60	91.85	53.00
30.10	58.35	106.60	175.00
36.50	72.80	0.57	79.50
82.30	116.70	0.79	29.50
87.85	45.15	0.77	78.40
105.00	70.35	0.81	127.50
95.22	77.40		

Latihan Soal 3

13.6 A study measured the sorption (either absorption or adsorption) rates of three different types of organic chemical solvents. These solvents are used to clean industrial fabricated-metal parts and are potential hazardous waste. Independent samples from each type of solvent were tested, and their sorption rates were recorded as a mole percentage. (See McClave, Dietrich, and Sincich, 1997.)

Aromatics		Chloroalkanes		Esters		
1.06	0.95	1.58	1.12	0.29	0.43	0.06
0.79	0.65	1.45	0.91	0.06	0.51	0.09
0.82	1.15	0.57	0.83	0.44	0.10	0.17
0.89	1.12	1.16	0.43	0.55	0.53	0.17
1.05				0.61	0.34	0.60

Is there a significant difference in the mean sorption rates for the three solvents? Use a P -value for your conclusions. Which solvent would you use?

For the data set in Exercise 13.6, use Bartlett's test to check whether the variances are equal. Use $\alpha = 0.05$.

Analysis of Variance (1)

Distribusi Normal/ Student dapat digunakan untuk menguji apakah rata-rata **dua** populasi berbeda atau tidak berbeda

Dalam kenyataannya sering dihadapkan pada **lebih dari** dua populasi dan perlu untuk mengetahui apakah rata-rata dari semua populasi itu sama atau tidak sama.

ANOVA/ANAVA adalah metode analisa statistik yang menguji kesamaan lebih dari dua rata-rata populasi

Menggunakan distribusi F dan pengujian dilakukan selalu dalam bentuk **ekasisi kanan**.

Analysis of Variance (2)

Dikembangkan oleh **R.A Fisher**, pada awalnya digunakan untuk penelitian pertanian.

Kini banyak digunakan di berbagai bidang penelitian, seperti penelitian ekonomi, bisnis, *engineering*, kedokteran, farmasi, dll

ANOVA terkenal dengan Tabel ANOVA

Varians merupakan kuadrat dari simpangan baku, yang mengukur seberapa jauh data menyimpang dari rata-ratanya

Tahapan ANOVA/ANAVA

1. Menghitung semua rata-rata sampel $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$, dimana **k** adalah banyaknya sampel.

2. Menghitung semua varians sampel $(s_1^2, s_2^2, \dots, s_k^2)$, dimana

$$s = \frac{\sum (x - \bar{x})^2}{n-1}, \text{ n merupakan ukuran sampel.}$$

3. Menghitung rata-rata varians sampel = $\frac{\sum s^2}{k}$

4. Menghitung rata-rata dari rata-rata sampel = $\bar{\bar{x}} = \frac{\sum x}{k}$

5. Menghitung varians dari rata-rata sampel = $s_{\bar{x}}^2 = \frac{\sum (\bar{x} - \bar{\bar{x}})^2}{k-1}$

6. Menghitung statistik $F_{hit} = \frac{nxS_{\bar{x}}^2}{\sum S^2}$

7. Menentukan daerah kritis

Distribusi F ini memiliki derajat kebebasan pembilang (v_1) = $k - 1$, dan derajat kebebasan penyebut (v_2) = $k(n-1)$, pada tingkat kesalahan α

Tabel ANOVA One Way

Table 13.3: Analysis of Variance for the One-Way ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	SSA	$k - 1$	$s_1^2 = \frac{SSA}{k - 1}$	$\frac{s_1^2}{s^2}$
Error	SSE	$k(n - 1)$	$s^2 = \frac{SSE}{k(n - 1)}$	
Total	SST	$kn - 1$		

Soal 2 : Uji ANOVA

Sebuah perusahaan menjual sabun sejenis dalam tiga bungkus yang berlainan pada harga yang sama. Data penjualan mengikuti distribusi normal dengan varians yang sama. Penjualan selama 5 bulan diuraikan sbb.

Bungkus 1	Bungkus 2	Bungkus 3
87	78	90
83	81	91
79	79	84
81	82	82
80	80	88
$\Sigma = 410$	$\Sigma = 410$	$\Sigma = 410$

Dengan tingkat kesalahan 5% ujilah apakah rata-rata penjualan untuk tiap jenis bungkus adalah sama.

Solusi Soal 1

1. $H_0 : \mu_1 = \mu_2 = \mu_3$

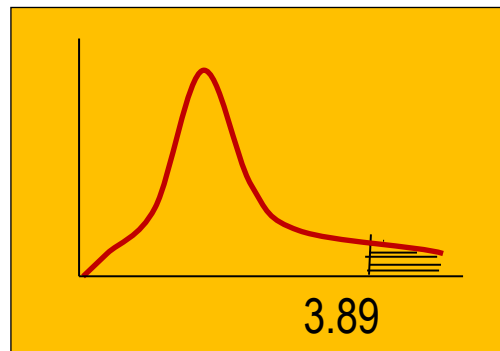
$H_1 : \mu_1, \mu_2, \mu_3$ tidak semua sama

2. $\alpha = 0.05$

3. Statistik hitung yang digunakan adalah nilai F

4. Daerah kritis

$\alpha = 0.05, v_1 = 3 - 1 = 2, \text{ dan } v_2 = 3(5-1) = 12; F_{(0.05;2;12)} = 3.89.$



Solusi Soal 1 (Lanj....)

5. Nilai Hitung

$$\text{a. } \bar{x}_1 = \frac{410}{5} = 82 \quad \bar{x}_2 = \frac{400}{5} = 80 \quad \bar{x}_3 = \frac{435}{5} = 87$$

$$\text{b. } s_1^2 = \frac{(87-82)^2 + (83-82)^2 + (79-82)^2 + (81-82)^2 + (80-82)^2}{5-1} = 10$$

$$s_2^2 = \frac{(78-80)^2 + (81-80)^2 + (79-80)^2 + (82-80)^2 + (80-80)^2}{5-1} = 2.5$$

$$s_3^2 = \frac{(90-87)^2 + (91-87)^2 + (84-87)^2 + (82-87)^2 + (88-87)^2}{5-1} = 15$$

c. Rata-rata varians sampel

$$\frac{(10 + 2.5 + 15)}{3} = 9.16$$

Solusi Soal 1 (Lanj....)

5. Nilai Hitung

d. Rata-rata dari rata-rata sampel

$$\bar{x} = \frac{(82 + 80 + 87)}{3} = 83$$

e. Varians rata-rata sampel

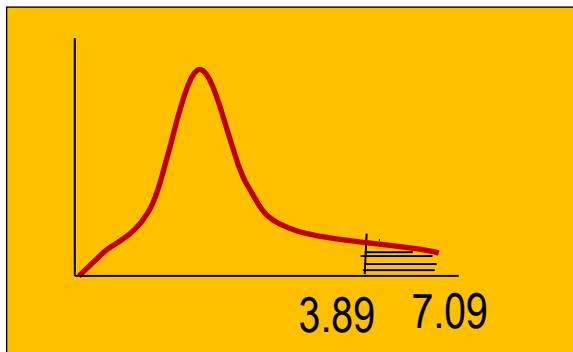
$$S_{\bar{x}}^2 = \frac{1 + 9 + 16}{3 - 1} = 13$$

f. Nilai Hitung F

$$F_{hitung} = \frac{5xS_{\bar{x}}^2}{\sum S} = \frac{5x13}{9.16} = 7.09$$

Solusi Soal 1 (Lanj....)

6. Kesimpulan



H_0 ditolak (H_1 diterima)

Rata-rata penjualan untuk masing-masing bungkus tidak semua sama

Artinya, bisa saja ada yang sama, namun tidak semuanya sama.

Latihan Soal 4

The city of Maumee comprises four districts. Chief of Police Andy North wants to determine whether there is a difference in the mean number of crimes committed among the four districts. He recorded the number of crimes reported in each district for a sample of six days. At the 0.05 significance level, can the chief of police conclude there is a difference in the mean number of crimes.

Number of Crimes			
Rec Center	Key Street	Monclova	White House
13	21	12	16
15	13	14	17
14	18	15	18
15	19	13	15
14	18	12	20
15	19	15	18

The city of Maumee comprises four districts. Chief of Police Andy North wants to determine whether there is a difference in the mean number of crimes committed among the four districts. He recorded the number of crimes reported in each district for a sample of six days. At the 0.05 significance level, can the chief of police conclude there is a difference in the mean number of crimes.

ANOVA Multi Faktor

Dalam ANOVA di atas, kita hanya berkepentingan untuk meneliti apakah penjualan sabun berbeda menurut bungkusnya.

Dalam hal ini **hanya ada satu faktor** yang mempengaruhi penjualan, yaitu bungkusnya. Analisa varians seperti ini dinamakan ANOVA 1 Faktor (*one factor/one way analysis of varians*)

Bila melibatkan beberapa faktor yang mempengaruhi, maka disebut ANOVA MULTIFAKTOR (*multiway analysis of variance*)

Soal 3 : Uji ANOVA

Ingin diteliti apakah jenis universitas dan prestasi selama kuliah mempengaruhi pendapatan pada tahun pertama dalam bekerja. Dipilih 5 universitas, dan terdapat 3 kategori prestasi dalam kuliah, yaitu terbaik, menengah, dan terendah, dengan data ,sbb.

Prestasi	Univ-1	Univ-2	Univ-3	Univ-4	Univ-5	Rata ² Sampel
Terbaik	20	18	16	14	12	16
Menengah	19	16	13	12	10	14
Terendah	18	14	10	10	8	12
Rata ² Sampel	19	16	13	12	10	$\bar{x} = 14$

Untuk tingkat kesalahan 5%, tentukan apakah (a) pendapatan sama untuk semua universitas atau tidak; (b) pendapatan sama untuk jenis prestasi atau tidak.

Solusi : Tabel ANOVA

Sumber Variasi	Derajat Kebebasan	Jumlah Kuadrat	Rata-Rata Kuadrat	Statistik F
Universitas (Faktor A)	4 (k- 1)	SSA = 150	MSA = SSA/4 = 37.5	MSA/MSE = 70
Rangking Prestasi (Faktor B)	2 (b - 1)	SSB = 40	MSB = SSB/2 = 20	MSB/MSE = 40
Kesalahan (Galat)	8 (k-1)(b-1)	SSE = 4	MSE = SSE/8 = 0.5	
Total	14 b(k-1)	SST = 194		

$$SST = \sum_j \sum_i (x_{ij} - \bar{\bar{x}})^2 = (20 - 14)^2 + \dots + (8 - 14)^2 = 194$$

$$SSA = b \sum_j (\bar{x}_j - \bar{\bar{x}})^2 = 3((19 - 14)^2 + (16 - 14)^2 + (13 - 14)^2 + (12 - 16)^2 + (10 - 14)^2) = 150$$

$$SSB = k \sum_i (\bar{x}_i - \bar{\bar{x}})^2 = 5((16 - 14)^2 + (14 - 14)^2 + (12 - 14)^2) = 40$$

$$SSE = SST - SSA - SSB = 194 - 150 - 40 = 40$$

Kaitan jenis pendapatan dengan jenis universitas

1. $H_0 : \mu_{A1} = \mu_{A2} = \mu_{A3} = \mu_{A4} = \mu_{A5}$
 $H_1 : \mu_{A1} = \mu_{A2} = \mu_{A3} = \mu_{A4} = \mu_{A5}$ tidak semua sama

2. $\alpha = 0.05$

3. Statistik hitung yang digunakan adalah nilai F

4. Daerah kritis

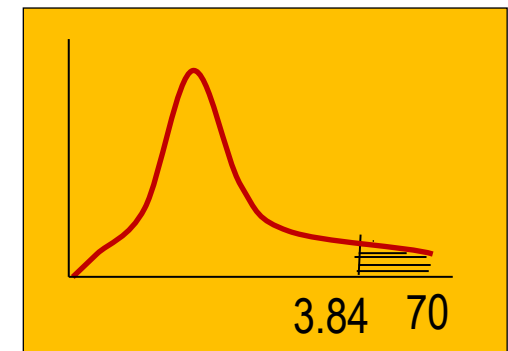
$\alpha = 0.05$, $v_1 = 5 - 1 = 4$, dan $v_2 = 4(3-1) = 8$, $F_{(0.05;4;8)} = 3.84$.

5. Nilai hitung :

$F_{\text{Hit}} = 70$ (Tabel ANOVA)

6. Kesimpulan

Nilai F_{hitung} jatuh pada daerah diarsir (daerah penolakan H_0), sehingga disimpulkan bahwa rata-rata populasi penghasilan sarjana tahun pertama kelima universitas tidak semua sama.



Kaitan jenis pendapatan dengan jenis prestasi

1. $H_0 : \mu_{B1} = \mu_{B2} = \mu_{B3}$ (semua sama)
 $H_1 : \mu_{B1} ; \mu_{B2} ; \mu_{B3}$ tidak semua sama

2. $\alpha = 0.05$

3. Statistik hitung yang digunakan adalah nilai F

4. Daerah kritis

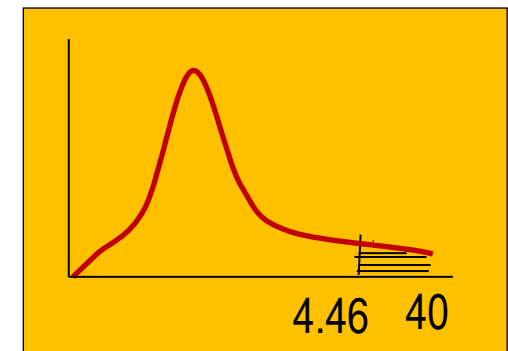
$\alpha = 0.05$, $v_1 = 3 - 1 = 2$, dan $v_2 = 2(5-1) = 8$, $F_{(0.05;2;8)} = 4.46$.

5. Nilai hitung :

$F_{Hit} = 40$ (Tabel ANOVA)

6. Kesimpulan

Nilai F_{hitung} jatuh pada daerah diarsir (daerah penolakan H_0), sehingga disimpulkan bahwa rata-rata populasi penghasilan sarjana tahun pertama ketiga jenis prestasi tidak semua sama.



Latihan Soal 5

Tabel berikut ini menunjukkan banyaknya pemilih masing-masing kontestan dan kelompok umur yang diambil secara random.

Kelompok Umur	Partai A	Partai B	Partai C
Muda	16	15	14
Dewasa	24	26	28
Tua	8	10	12

Dengan tingkat kesalahan 5%, ujilah hipotesis berikut ini:

- Populasi pemilih ketiga partai adalah sama
- populasi pemilih untuk ketiga kelompok umur adalah sama

Latihan Soal 6

Sebuah penelitian tentang tingkat pendapatan dan hubungannya dengan kunjungan ke Puncak untuk rekreasi. Tabel berikut ini menunjukkan banyaknya orang yang berangkat ke Puncak menurut tingkat pendapatan dan frekuensi.

Pendapatan	Ke Puncak		
	Tidak Pernah	Jarang	Sering
Rendah	77	13	8
Sedang	145	58	27
Tinggi	21	32	19

Dengan tingkat kesalahan 5%, penelitian tersebut memiliki dua kesimpulan. Pertama, populasi yang berangkat ke Puncak adalah sama untuk setiap jenis pendapatan. Kedua, populasi yang berangkat ke Puncak adalah sama untuk setiap kategori frekuensi kunjungan. Apakah Anda setuju dengan dua poin kesimpulan peneliti tersebut ..?

Tugas 8 (Buku Walpole, at al., 2012)

1. Soal No 13.42 Halaman 555
2. Soal No 13.45 Halaman 555 (hanya (a) dan (b) saja)
3. Soal No.13.1 Halaman 518
4. Soal No 13.2 Halaman 518
5. Slide berikutnya

Dikumpulkan paling lambat **Jumat, 29 November 2024**, pukul 19.59.59 Wib ke email:
budi.marpaung@ukrida.ac.id

Tugas 5a Nomor 5

Suatu penelitian akan dilakukan untuk membandingkan pengaruh jenis media pembelajaran yang digunakan guru terhadap hasil belajar siswa kelas 2 SMA khusus untuk pokok bahasan peluang. Jenis media yang dimaksudkan adalah cetak, audio, visual dan berbasis komputer. Untuk keperluan tersebut telah dipilih secara acak 12 kelas, namun setelah dilakukan tes kemampuan awal ternyata kelas-kelas tersebut dapat digolongkan menjadi 3 kelompok (kategori kemampuan awal rendah, kategori kemampuan awal sedang, kategori kemampuan awal tinggi). Masing-masing kelompok mendapatkan perlakuan 4 jenis media tersebut. Setelah pembelajaran selesai, semua kelas mendapat tes dengan soal dan waktu yang sama. Berikut adalah data tentang rata-rata nilai tes siswa dari keduabelas kelas yang digunakan dalam penelitian.

Tugas 5a Nomor 5 (Lanj...)

Kemampuan Awal	Jenis Media			
	Cetak	Audio	Visual	Berbasis Komputer
Rendah	8.1	6.5	7.4	8.4
Sedang	8.9	6.8	6.0	7.4
Tinggi	7.7	5.9	5.9	9.4

Apa saja yang dapat Anda simpulkan dari data di atas? Gunakan $\alpha = 0,05$.

SEKIAN