



KD 2 STATISTIKA MULTIVARIAT

UJI PERBEDAAN RATA-RATA

Review...

One-sample t-test (univariate)

■ Suppose that $x_1, \dots, x_n$ are i.i.d. $N(\mu, \sigma^2)$. Then

- ◆ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \sim N(\mu, \sigma^2/n)$
- ◆ $ns^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \sim \sigma^2 \chi_{n-1}^2$
- ◆ $\bar{x}$ and s^2 are independent

■ Suppose that the mean μ is unknown. Then we can do a one-sample t-test:

- ◆ $H_0 : \mu = \mu_0, H_a : \mu \neq \mu_0$.
- ◆ Test statistic: $t = \frac{\bar{x} - \mu_0}{s_u / \sqrt{n}}$, where s_u is sample standard deviation with $n - 1$ in the denominator.
- ◆ Under H_0 , $t \sim t_{n-1}$, i.e., it has a student t-distribution with $n - 1$ degrees of freedom.
- ◆ Compute p-value. If p-value < 0.05 , we reject the null hypothesis.

Review...

Two-sample t-test (univariate)

- Suppose that $x_1, \dots, x_n$ are i.i.d. $N(\mu_X, \sigma_X^2)$, and let $y_1, \dots, y_m$ be i.i.d. $N(\mu_Y, \sigma_Y^2)$ (independent of $x_1, \dots, x_n$).
- Suppose we want to test whether $\mu_X = \mu_Y$, under the assumption that $\sigma_X = \sigma_Y$. Then we can do a two-sample t-test:
 - ◆ $H_0 : \mu_X = \mu_Y, H_a : \mu_X \neq \mu_Y$.
 - ◆ Test statistic: $t = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2(\frac{1}{n} + \frac{1}{m})}}$, where

$$s_p^2 = \frac{1}{n + m - 2} (ns_X^2 + ms_Y^2).$$

- ◆ Under H_0 , $t \sim t_{n+m-2}$.

One sample T^2 -test (multivariate)

■ Suppose that $x_1, \dots, x_n$ are i.i.d. $N_p(\mu, \Sigma)$. Then

◆ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \sim N_p(\mu, \Sigma/n)$.

◆ $nS \sim W_p(\Sigma, n-1)$, where S is the sample covariance matrix and W_p is a **Wishart distribution**.

◆ $\bar{x}$ and S are independent.

■ Suppose that the mean μ is unknown. Then we can do a one-sample t-test:

◆ $H_0 : \mu = \mu_0, H_a : \mu \neq \mu_0$.

◆ Test statistic: $T = n(\bar{x} - \mu_0)' S_u^{-1} (\bar{x} - \mu_0)$, where S_u is the sample covariance matrix with $n-1$ in the denominator.

◆ Under H_0 , $T \sim T^2(p, n-1)$, i.e., it has **Hotelling's T^2 distribution** with parameters p and $n-1$.

Two sample T^2 -test (multivariate)

- Suppose that $x_1, \dots, x_n$ are i.i.d. $N_p(\mu_X, \Sigma_X)$, and let $y_1, \dots, y_m$ be i.i.d. $N_p(\mu_Y, \Sigma_Y)$ (independent of $x_1, \dots, x_n$).
- Suppose we want to test whether $\mu_X = \mu_Y$, under the assumption that $\Sigma_X = \Sigma_Y$. Then we can do a two-sample t-test:
 - ◆ $H_0 : \mu_X = \mu_Y$, $H_a : \mu_X \neq \mu_Y$.
 - ◆ Test statistic: $T = (nm/(n+m))(\bar{x} - \bar{y})' S_u^{-1}(\bar{x} - \bar{y})$, where

$$S_u = \frac{1}{n+m-2} (nS_1 + mS_2).$$

- ◆ Under H_0 , $T \sim T^2(p, n+m-2)$.

Uji t (Johnson, 2007)

- ❑ Salah satu pendekatan rasional untuk membandingkan dua perlakuan adalah dengan cara menempatkan keduanya dalam unit yang sama.
- ❑ Respon yang dipasangkan dapat dianalisis dengan cara menghitung diferensinya melalui eliminasi pengaruh variasi dari unit ke unit.
- ❑ Dalam kasus univariat (respon tunggal), model umum dengan n diferensi dapat dinotasikan sebagai berikut:

$$D_j = X_{1j} - X_{2j} \quad j = 1, 2, \dots, n$$

X_{1j} = respon untuk perlakuan 1 dengan percobaan ke-j

X_{2j} = respon untuk perlakuan 2 dengan percobaan ke-j

Model ini hanya dapat menggambarkan pengaruh diferensi dari perlakuan dan mengasumsikan D_j mewakili observasi independen dari sebuah distribusi $N(\delta, \sigma_d^2)$ dengan variabel

$$t = \frac{\bar{D} - \delta}{s_d / \sqrt{n}}$$

$$\bar{D} = \frac{1}{n} \sum_{j=1}^n D_j \quad \text{dan} \quad s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (D_j - \bar{D})^2$$

berdistribusi t dengan db n-1

hipotesis $H_0 : \delta = 0$ (tidak ada diferensi rata-rata untuk perlakuan)

$H_1 : \delta \neq 0$

Pengujiannya adalah dengan membandingkan $|t|$ dengan $t_{n-1}(\alpha/2)$

KI:
Interval kepercayaan untuk diferensi rata-rata $\delta = E(X_{1j} - X_{2j})$ dinyatakan sebagai berikut:

$$\bar{d} - t_{n-1}(\alpha/2) \frac{s_d}{\sqrt{n}} \leq \delta \leq \bar{d} + t_{n-1}(\alpha/2) \frac{s_d}{\sqrt{n}}$$

Remember... MetStat 1

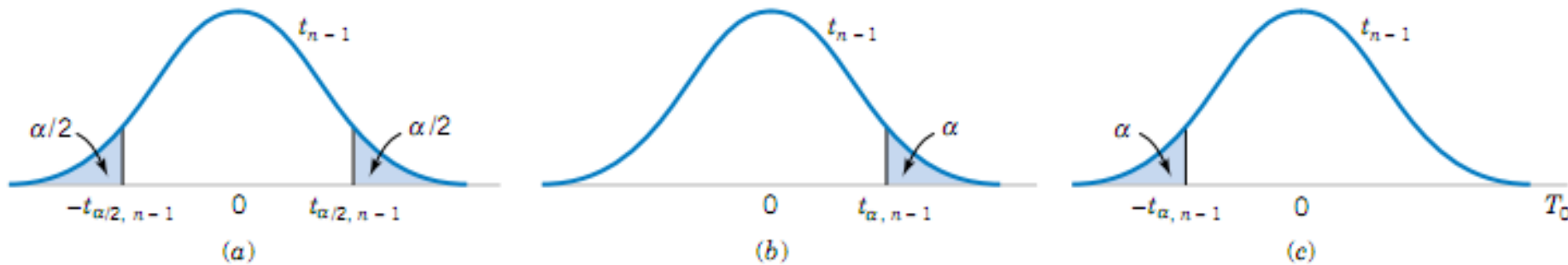


Figure 9-8 The reference distribution for $H_0: \mu = \mu_0$ with critical region for (a) $H_1: \mu \neq \mu_0$, (b) $H_1: \mu > \mu_0$, and (c) $H_1: \mu < \mu_0$.

Untuk uji dua sisi, Daerah Kritis (DK) : Tolak H_0 jika

$$|t| > t_{n-1, \frac{\alpha}{2}} \Leftrightarrow t > t_{n-1, \frac{\alpha}{2}} \text{ atau } t < -t_{n-1, \frac{\alpha}{2}}$$

Silahkan mau menggunakan yang ruas kiri atau ruas kanan

contoh

x_1	x_2
1	1
3	4
2	7
$\bar{x}_1 = 2$	$\bar{x}_2 = 4$



X_1	X_2	D_j
1	1	0
3	4	-1
2	7	-5
		-6
		$\bar{D} = -2$

i. Susun hipotesis

$$H_0 : \delta = 0 \quad (\text{tidak ada diferensi rata-rata untuk perlakuan})$$

$$H_1 : \delta \neq 0$$

$$t = \frac{\bar{D} - \delta}{s_d / \sqrt{n}}$$

$$t = \frac{-2}{\sqrt{7} / \sqrt{3}} \Rightarrow |t| = 1,31$$

ii. Hitungan

$$s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (D_j - \bar{D})^2 = \frac{1}{2} (2^2 + 1^2 + (-3)^2) = 7$$

$$s_d = \sqrt{7}$$

$$t \text{ tabel } t_{n-1, \frac{\alpha}{2}} = t_{2, 0,025} = 4,303$$

Karena $t < t \text{ tabel}$ maka H_0 tidak ditolak

Satu sampel uji Hotelling

- Penambahan notasi dibutuhkan dalam prosedur perbandingan berpasangan pada kasus multivariat
 - Hal ini diperlukan untuk membedakan antara p respon, dua perlakuan, dan n unit eksperimen
-

X_{11j} = variabel 1 dalam perlakuan 1

X_{12j} = variabel 2 dalam perlakuan 1

⋮

X_{1pj} = variabel p dalam perlakuan 1

X_{21j} = variabel 1 dalam perlakuan 2

X_{22j} = variabel 2 dalam perlakuan 2

⋮

X_{2pj} = variabel p dalam perlakuan 2

diferensi pasangan p menjadi

$$D_{1j} = X_{11j} - X_{21j}$$

$$D_{2j} = X_{12j} - X_{22j}$$

$$\vdots \quad \quad \quad \vdots$$

$$D_{pj} = X_{1pj} - X_{2pj}$$

Persamaan *

Diberikan $D_j = [D_{1j}, D_{2j}, \dots, D_{pj}]$, $j = 1, 2, \dots, n$, dengan asumsi

$$E(D_j) = \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_p \end{bmatrix} \text{ dan } \text{Cov}(D_j) = \Sigma_d$$

Jika dilakukan penambahan, $D_1, D_2, \dots, D_n$ menjadi vektor random yang independen, inferensi tentang vektor diferensi rata-rata δ menjadi bergantung pada statistik T^2 .

Secara spesifik,

$$T^2 = n(\bar{D} - \delta)' S_d^{-1} (\bar{D} - \delta) \quad \text{dengan} \quad \bar{D} = \frac{1}{n} \sum_{j=1}^n D_j \quad \text{dan} \quad s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (D_j - \bar{D})(D_j - \bar{D})'$$

Result 6.1

Diberikan diferensi $D_1, D_2, \dots, D_n$ sampel random dari sebuah populasi $N_p(\delta, \Sigma_d)$. Maka

$$T^2 = n(\bar{D} - \delta)' S_d^{-1} (\bar{D} - \delta)$$

berdistribusi sebagai sebuah variabel random $\left[(n-1) p / (n-p) \right] F_{p, n-p}$ dan asumsi nilai δ dan Σ_d selalu bernilai benar.

Jika nilai n dan $n-p$ keduanya besar, T^2 didistribusikan sebagai sebuah variabel random

$$\chi_p^2.$$

Kondisi $\delta = 0$ ekuivalen dengan “tidak ada diferensi rata-rata antara 2 perlakuan”. Untuk variabel ke- i , $\delta_i > 0$ secara tidak langsung menyatakan bahwa perlakuan rata-rata ke-2 lebih tinggi daripada perlakuan ke-1.

Diberikan diferensi $d_j' = [d_{1j}, d_{2j}, \dots, d_{pj}]$, $j = 1, 2, \dots, n$ berkoresponden dengan variabel random pada persamaan $*$, uji level α dengan $H_0 : \delta = 0$ dan $H_1 : \delta \neq 0$ untuk sebuah populasi berdistribusi $N_p(\delta, \Sigma_d)$. Kriteria pengujiannya tolak H_0 jika

$$T^2 = n\bar{d}'S_d^{-1}\bar{d} > \left[\frac{(n-1)p}{n-p} \right] F_{p, n-p}(\alpha)$$

Daerah kepercayaan untuk δ

$$(\bar{d} - \delta)' S_d^{-1} (\bar{d} - \delta) \leq \left[\frac{(n-1)p}{n(n-p)} \right] F_{p,n-p}(\alpha)$$

Interval kepercayaan simultan untuk diferensi rata-rata individu δ_i dinotasikan

$$\delta_i: \bar{d}_i \pm \sqrt{\left[\frac{(n-1)p}{n-p} \right] F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_i}^2}{n}}$$

Dimana $\bar{d}_i$ elemen ke-i dari $\bar{d}$ dan $s_{d_i}^2$ adalah elemen diagonal ke-i dari S_d .

Interval kepercayaan simultan *Bonferroni* untuk diferensi rata-rata individual adalah:

$$\delta_i: \bar{d}_i \pm t_{n-1} \left(\frac{\alpha}{2p} \right) \sqrt{\frac{s_{d_i}^2}{n}}$$

contoh

Pengukuran terhadap *Biochemical Oxygen Demand* (BOD) dan *Suspended Solids* (SS) telah dilakukan, untuk sampel $n=11$, dari dua laboratories. Data ditunjukkan pada Tabel berikut.

Sampel j	Commercial Lab		State lab of Hygiene	
	$X_{11j}(\text{BOD})$	$X_{12j}(\text{SS})$	$X_{21j}(\text{BOD})$	$X_{22j}(\text{SS})$
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31
6	34	75	44	64
7	28	26	42	30
8	71	124	54	64
9	43	54	34	56
10	33	30	29	20
11	20	14	39	21

Apakah kedua laborat dari *Commercial Lab* dan *State Lab of Hygiene* setuju bahwa terdapat perbedaan rata-rata pada pengukuran BOD dan SS ?

i. Susun hipotesis $H_0 : \delta' = [\delta_1, \delta_2] = [0, 0]$

ii. Hitungan

Sampel j	Commercial Lab		State lab of Hygiene	
	$X_{11j}(\text{BOD})$	$X_{12j}(\text{SS})$	$X_{21j}(\text{BOD})$	$X_{22j}(\text{SS})$
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31
6	34	75	44	64
7	28	26	42	30
8	71	124	54	64
9	43	54	34	56
10	33	30	29	20
11	20	14	39	21

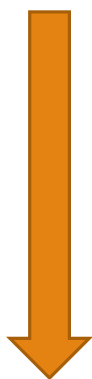


$X_{11j}(\text{BOD})$	$X_{21j}(\text{BOD})$	$d_{1j} = X_{11j} - X_{21j}$	$X_{12j}(\text{SS})$	$X_{22j}(\text{SS})$	$d_{2j} = X_{12j} - X_{22j}$
6	25	-19	27	15	12
6	28	-22	23	13	10
18	36	-18	64	22	42
8	35	-27	44	29	15
11	15	-4	30	31	-1
34	44	-10	75	64	11
28	42	-14	26	30	-4
71	54	17	124	64	60
43	34	9	54	56	-2
33	29	4	30	20	10
20	39	-19	14	21	-7

Re : $T^2 = n(\bar{D} - \delta)' S_d^{-1} (\bar{D} - \delta)$

Ditentukan

$$\bar{D} = \frac{1}{n} \sum_{j=1}^n D_j \text{ dan } s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (D_j - \bar{D})(D_j - \bar{D})'$$



$$\bar{d} = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \end{bmatrix} = \begin{bmatrix} \frac{-19 + (-22) + (-18) + \dots + 4 + (-19)}{11} \\ \frac{12 + 10 + 42 + \dots + 10 + (-7)}{11} \end{bmatrix} = \begin{bmatrix} \frac{(-103)}{11} \\ \frac{146}{11} \end{bmatrix} = \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix}$$

$d_{1j}=X_{11j}-X_{21j}$	$d_{2j}=X_{12j}-X_{22j}$
-19	12
-22	10
-18	42
-27	15
-4	-1
-10	11
-14	-4
17	60
9	-2
4	10
-19	-7
-103	146
-9,363636364	13,27272727

Re: SSCP....

d' =											
	-9,64	-12,64	-8,64	-17,64	5,36	-0,64	-4,64	26,36	18,36	13,36	-9,64
	-1,27	-3,27	28,73	1,73	-14,27	-2,27	17,27	46,73	-15,27	-3,27	-20,27

S =		199,2546	88,30908
		72,28252	418,6182

d =	
-9,64	-1,27
-12,64	-3,27
-8,64	28,73
-17,64	1,73
5,36	-14,27
-0,64	-2,27
-4,64	-17,27
26,36	46,73
18,36	-15,27
13,36	-3,27
-9,64	-20,27

dan dari persamaan $T^2 = n\bar{d}'S_d^{-1}\bar{d}$, diperoleh perhitungan berikut:

$$T^2 = 11[-9.36, \quad 13.27] \begin{bmatrix} 0.0055 & -0.0012 \\ -0.0012 & 0.0026 \end{bmatrix} \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix} = 13.6$$

iii. Keputusan

$$\begin{aligned} \text{Karena} \quad \left[(n-1)p/(n-p) \right] F_{p, n-p}(0.05) &= \left[\frac{(11-1)2}{11-2} \right] F_{2,9}(0.05) \\ &= \left[\frac{20}{9} \right] (4.26) \\ &= 9.467 \end{aligned}$$

Karena $T^2 = 13.6 > 9.467$, maka H_0 ditolak. Artinya ada diferensi rata-rata antara pengukuran dua laboratories.

KI 95%

$$\delta_i: \bar{d}_i \pm \sqrt{\left[\frac{(n-1)p}{n-p}\right] F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_i}^2}{n}}$$

$$\begin{aligned} \delta_1: \bar{d}_1 \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_1}^2}{n}} &= -9.36 \pm \sqrt{9.47} \sqrt{\frac{199.26}{11}} \\ &= -9.36 \pm 13.10 \quad \text{atau} \quad (-22.46, 3.74) \end{aligned}$$

$$\begin{aligned} \delta_2: \bar{d}_2 \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_2}^2}{n}} &= 13.27 \pm \sqrt{9.47} \sqrt{\frac{418.61}{11}} \\ &= 13.27 \pm 18.98 \quad \text{atau} \quad (-5.71, 32.25) \end{aligned}$$

Uji z, n besar → Uji Khi Kuadrat
→ kelompok Esti, Faradila, Farah

Latihan

Lakukan uji z terhadap x_1 dan x_2 untuk mengetahui ada tidaknya diferensi rata-rata biaya (*cost*) dari gasoline dan diesel truck dari data berikut:

Gasoline Truck		Diesel Truck	
x_1 (fuel)	x_2 (repair)	x_1 (fuel)	x_2 (repair)
16,44	12,43	8,5	12,26
7,19	2,7	7,42	5,13
9,92	1,35	10,28	3,32
4,24	5,78	10,16	14,72
11,2	5,05	12,79	4,17
14,25	5,78	9,6	12,72
13,5	10,98	6,47	8,89
13,32	14,25	11,35	9,95
29,11	15,09	9,15	2,94