11

Statistik Non Parametrik

1. Statistika Parametrik vs Nonparametrik

- **Statistika Parametrik :**
  - Teknik-teknik statistika yang didasarkan atas asumsi mengenai populasi yang diambil sampelnya.
  - Contoh: pada uji t diasumsikan populasi terdistribusi normal.
  - Sebutan parametrik digunakan karena pada uji t ini yang diuji adalah parameter (contoh: rata-rata populasi)
  - Membutuhkan data kuantitatif dengan level <u>interval</u> atau <u>rasio</u>
- **Statistika Nonparametrik :**
  - Cocok untuk data yang tidak memenuhi asumsi statistika parametrik atau yang berjenis kualitatif
  - Disebut juga *distribution-free statistics*
  - Didasarkan atas lebih sedikit asumsi mengenai populasi dan parameter dibandingkan dengan statistika parametrik
  - Ada yang dapat digunakan untuk data <u>nominal</u>
  - Ada yang dapat digunakan untuk data <u>ordinal</u>

2. Keuntungan dan Kekurangan Statistika Nonparametrik

- **Keuntungan :**
  - Kadang-kadang tidak ada alternatifnya pada statistika parametrik
  - Uji nonparametrik tertentu dapat digunakan untuk analisis data nominal
  - Uji nonparametrik tertentu dapat digunakan untuk analisis data ordinal
  - Proses perhitungan pada statistika nonparametrik biasanya lebih sederhana dibandingkan pada statistika parametrik, khususnya untuk sampel kecil

- **Kekurangan :**
  - Uji nonparametrik menjadi tak berguna apabila uji parametrik untuk data yang sama tersedia
  - Uji nonparametrik pada umumnya tidak tersedia secara luas dibandingkan dengan uji parametrik
  - Untuk sampel besar, perhitungan untuk statistika nonparametrik menjadi rumit

- Tiga metode uji nonparametrik pada bab ini, yaitu **Mann-Whitney, Wilcoxon, dan Rank Spearman.**

3. Mann-Whitney Test (Uji U)

- Adalah uji nonparametrik untuk membandingkan dua populasi independen (pada statistika parametrik: uji t)
- Populasi *tidak harus* terdistribusi normal (pada uji t: harus normal)
- Level data serendah-rendahnya ordinal (uji t tidak dapat)
- Hipotesis yang diuji:
  - H_0 : kedua populasi indentik
  - H_1 : kedua populasi tidak indentik
- Prosedur uji :
  a. Tetapkan satu sampel sebagai kelompok 1 dan sampel lain sebagai kelompok 2
  b. Data dari kedua kelompok disatukan dengan setiap data diberi kode asal kelompoknya
  c. Data yang telah digabungkan diberi peringkat dari 1 (nilai terkecil) sampai n
  d. Jumlah peringkat dari kelompok 1 dihitung dan diberi simbol W_1
  e. Jumlah peringkat dari kelompok 2 dihitung dan diberi simbol W_2
  f. Langkah selanjutnya, bergantung apakah sampelnya kecil atau besar

3.1. Uji U pada Sampel Kecil : $n_1 \leq 10$ dan $n_2 \leq 10$

- Hitung U_1 dan U_2

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1 \quad \text{dan} \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2$$

- *U* adalah *yang terkecil di antara U_1 dan U_2*
  Catatan: salah satu U_i saja yang perlu dihitung, sedangkan U yang satu lagi dapat dihitung dengan $U_j = n_1 n_2 - U_i$.
- Gunakan **Tabel 8** (*halaman berikut*), untuk mendapatkan nilai p untuk U yang telah dihitung.
- Untuk menggunakan **Tabel 8** tetapkan n_1 adalah yang kecil dan n_2 adalah yang besar ($n_1 < n_2$)
- Nilai p pada **Tabel 8** adalah untuk uji satu sisi. Untuk uji dua sisi, nilai p nya adalah 2 kali yang ada pada **Tabel 8**.

Tabel 8. Fungsi Distribusi dari U

$n_2 = 8$

U	n_1							
	1	2	3	4	5	6	7	8
0	0.1111	0.0222	0.0061	…	…	…	…	0.0001
1	0.2222	0.0444	0.0121	…	…	…	…	0.0002
2	0.3333	0.0889	0.0242	…	…	…	…	0.0003
3	0.4444	…	…	…	…	…	**0.0011**	0.0005
4	0.5556	…	…	…	…	…	…	0.0009
5		…	…	…	…	…	…	…
6		…	…	…	…	…	…	…
7		…	…	…	…	…	…	…
8		…	…	…	…	…	…	…
…			…	…	…	…	…	…
…			…	…	…	…	…	…
32								0.5204

Tabel lengkap dapat dilihat pada hal. 408-415, Buku : *Statistik untuk Manajemen dan Ekonomi* oleh. N. Soemartojo (Bisa dipinjam di Perpustakaan)

- **Contoh** :

  Apakah ada perbedaan antara honor per jam pekerja kesehatan dengan pekerja pendidikan? Misalkan diambil sampel acak dari 7 pekerja kesehatan dan 8 pekerja pendidikan. Semua pekerja tersebut diwawancara dan ditanya honor perjamnya, sebagaimana tercantum di dalam tabel berikut. Lakukan pengujian Mann-Whitney untuk menentukan apakah kedua populasi berbeda di dalam penerimaan honor. Gunakan α = 5%

  Data Sampel

Pekerja Kesehatan ($)	Pekerja Pendidikan ($)
20.10	26.19
19.80	23.88
22.36	25.50
18.75	21.64
21.90	24.85
22.96	25.30
20.75	24.12
	23.45

- Karena populasi tidak dapat diasumsikan normal, maka uji t 2 sampel tidak dapat digunakan (meskipun level data adalah rasio). Jadi digunakan uji U
- H_0 : populasi honor pekerja kesehatan dan pekerja pendidikan identik
  H_1 : populasi honor pekerja kesehatan dan pekerja pendidikan tidak identik
- $n_1 = 7$ dan $n_2 = 8$
- $\alpha = 5\%$

Honor per jam	Peringkat	Kelompok
18.75	1	H
19.80	2	H
20.10	3	H
20.75	4	H
21.64	5	E
21.90	6	H
22.36	7	H
22.96	8	H
23.45	9	E
23.88	10	E
24.12	11	E
24.85	12	E
25.30	13	E
25.50	14	E
26.19	15	E

H = *Health* = Kesehatan, E = *Education* = Pendidikan

$$W_1 = 1+2+3+4+6+7+8 = 31$$

$$W_2 = 5+9+10+11+12+13+14+15 = 89$$

$$U_1 = 7*8 + \frac{7*8}{2} - 31 = 53$$

$$U_2 = 7*8 + \frac{8*9}{2} - 89 = 3$$

atau dihitung dengan 7*8 – 53 = 3

$$U = \min(U_1, U_2) = 3$$

- Dari Tabel 8 diatas, untuk $n_1 = 7$, $n_2 = 8$, dan $U = 3$, didapatkan nilai p untuk uji 1 sisi adalah 0.0011. Untuk uji 2 sisi, nilai p = 2 * 0.0011 = 0.0022. Karena nilai p < , maka tolak H_0. Artinya, populasi honor pekerja kesehatan dan pekerja pendidikan tidak identik.

  Catatan: terlihat bahwa pada umumnya pekerja pendidikan menerima honor lebih tinggi dari pada pekerja kesehatan

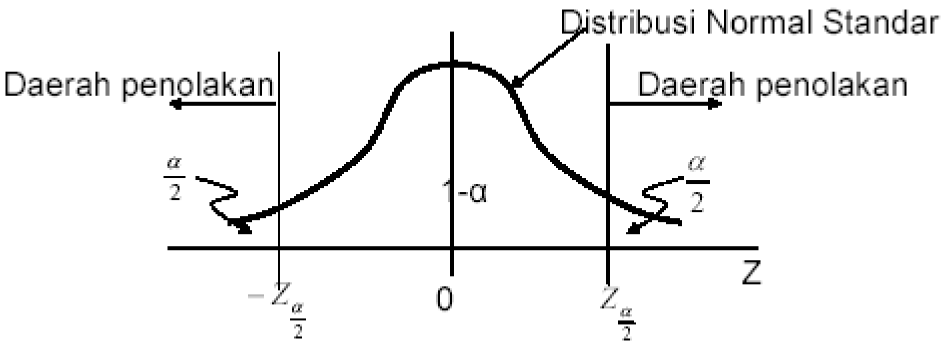
3.2. Uji U pada Sampel Besar : $n_1 > 10$ dan $n_2 > 10$

- Untuk sampel besar ($n_1 > 10$ dan $n_2 > 10$), distribusi sampling untuk U akan mendekati distribusi normal dengan rata-rata dan deviasi standar sebagai berikut:

$$\boxed{\mu_U = \frac{n_1 n_2}{2}} \quad \text{dan} \quad \boxed{\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

- H_0 : kedua populasi identik
  H_1 : kedua populasi tidak identik
- Statistik uji

$$\boxed{z = \frac{U - \mu_U}{\sigma_U}}$$



- **Contoh** :

  Apakah uang yang dibelanjakan oleh karyawan untuk makan siang ke restoran sama saja dengan yang ke warung? Untuk menguji hal ini, seorang peneliti mengumpulkan data acak dari karyawan yang makan siang ke restoran dan yang ke warung. Gunakan $\alpha = 1\%$.

  Data Sampel

Warung ($)	Restoran ($)
2.75	4.10
3.29	4.75
4.53	3.95
3.61	3.50
3.10	4.25
4.29	4.98
2.25	5.75
2.97	4.10
4.01	2.70
3.68	3.65
3.15	5.11
2.97	4.80
4.05	6.25
3.60	3.89
	4.80
	5.50
$n_1 = 14$	$n_2 = 16$

Jawab :

- H_0 : populasi pengeluaran uang makan siang untuk karyawan yang ke warung *sama* dengan yang ke restoran
  H_1 : populasi pengeluaran uang makan siang untuk karyawan yang ke warung *tidak sama* dengan yang ke restoran
- $n_1 > 10$ dan $n_2 > 10$, maka digunakan Uji U untuk sampel besar
- $\alpha = 0.01$, apabila nilai $p < \alpha$ maka tolak H_0

Nilai	Peringkat	Kelompok
2.25	1	W
2.70	2	R
2.75	3	W
2.97	4.5	W
2.97	4.5	W
3.10	6	W
3.15	7	W
3.29	8	W
3.50	9	R
3.60	10	W
3.61	11	W
3.65	12	R
3.68	13	W
3.89	14	R
3.95	15	R

Nilai	Peringkat	Kelompok
4.01	16	W
4.05	17	W
4.10	18.5	R
4.10	18.5	R
4.25	20	R
4.29	21	W
4.53	22	W
4.75	23	R
4.80	24.5	R
4.80	24.5	R
4.98	26	R
5.11	27	R
5.50	28	R
5.75	29	R
6.25	30	R

- Jumlah peringkat yang dari kelompok W (warung) =
  W_1 = 1+3+4.5+4.5+6+7+8+10+11+13+16+17+21+22 = 144

- $$U_1 = 14*16 + \frac{14*15}{2} - 144 = 185$$
  $$U_2 = 14*16 - 185 = 39$$
  $$\mathrm{U} = \min(39,185) = 39$$
  $$\mu_{\mathrm{U}} = \frac{14*16}{2} = 112$$
  $$\sigma_U = \sqrt{\frac{14*16*31}{12}} = 24.1$$
  $$z = \frac{39-112}{24.1} = -3.03$$

Uji 2 sisi

- Nilai p untuk z = -3.03 adalah 2 * 0.0012 = 0.0024 <α tolak H_0, artinya populasi pengeluaran uang makan siang untuk karyawan yang ke warung *tidak sama* dengan yang ke restoran

4. Uji Wilcoxon untuk Data Sepadan

- Data Sepadan (*matched pairs*) :
  - Statistika Parametrik : Uji t (asumsi : populasi normal)
  - Statistika Nonparametrik : Uji Wilcoxon
- Uji Wilcoxon (seperti juga uji t) digunakan untuk menganalisis data pada 2 kelompok yang berkaitan, termasuk kasus *before- and -after* di mana orang atau objek yang sama *diamati pada 2 kondisi yang berbeda*
- Jenis data pada Wilcoxon : serendah-rendahnya ordinal
- <u>Prosedur uji</u> :

  a. n = banyaknya pasangan data
  b. Urutkan perbedaan antara kedua data (d), dari yang terkecil sampai yang terbesar, tanpa memperhatikan pakah perbedaan tersebut (+) atau (-)
  c. Jika perbedaan tersebut (-) maka peringkatnya juga diberi tanda (-)
  d. Perbedaan (d) yang bernilai 0 (apabila ada) diabaikan, dan banyak data (n) dikurangi sebanyak d yang bernilai 0
  e. Jumlahkan peringkat bertanda (-), sebut T-. Tanda (-) tidak ikut dalam penjumlahan
  f. Jumlahkan peringkat yang bertanda (+), sebut T+.
  g. Statistik uji : T = min (T- dan T+)

- Hipotesis Uji Wilcoxon :
  $H_0 : M_d = 0$ versus $H_a: M_d \neq, 0$ (*two-tailed test*)
  $H_0 : M_d = 0$ versus $H_a: M_d >, 0$ (*one-tailed test*)
  $H_0 : M_d = 0$ versus $H_a: M_d <, 0$ (*one-tailed test*)

  Catatan :

  M_d = median perbedaan antara kedua populasi
  $M_d = 0$ berarti kedua populasi identik

4.1. Uji Wilcoxon untuk Sampel Kecil (n < 15)

- Dengan *n* dan *α*, gunakan Tabel A14 (tersedia untuk *one-tailed test* dan *twotailed test :* lihat *Buku Statistika*) untuk mendapatkan T_{kritis}.
- Jika $T < T_{kritis} \rightarrow$ tolak H_0.

- **Contoh** :

  Seorang peneliti melakukan survey mengenai biaya pemeliharaan kesehatan yang dikeluarkan oleh keluarga di kota A dan B. Peneliti tersebut mengambil enam pasang keluarga yang dipadankan secara demografis di kota A dan B. Dari keenam pasang keluarga tersebut dicatat biaya pemeliharaan kesehatan pada tahun yang lalu (dalam USD). Dengan menggunakan α =0.05, lakukan pengujian untuk menentukan apakah ada perbedaan signifikan di dalam pengeluaran biaya kesehatan di antara kedua kota tersebut.

  Data Sampel

Pasangan keluarga	A	B
1	1950	1760
2	1840	1870
3	2015	1810
4	1580	1660
5	1790	1340
6	1925	1765

Jawab :

- Karena populasi tidak dapat diasumsikan normal, maka digunakan Uji Wilcoxon (bukan uji t), meskipun datanya berlevel rasio
- H_0: $M_d = 0$ versus H_1: $M_d \neq 0$
- $\alpha = 0.05$
- n = 6 (< 15) → sampel kecil

Kel	A	B	Perbe-daan d	Pering-kat
1	1950	1760	+190	+4
2	1840	1870	-30	-1
3	2015	1810	+205	+5
4	1580	1660	-80	-2
5	1790	1340	+450	+6
6	1925	1765	+160	+3

- T+ = 4+5+6+3 = 18
- T- = 1+2 = 3
- T = min (T- dan T+) = min (18 dan 3) = 3
- n = 6, α = 0.05 → (*Two-tailed test*) T_{kritis} = 1.
  Karena T > T_{kritis} maka pertahankan H_0. Artinya tidak cukup bukti bahwa pengeluaran biaya kesehatan di kedua kota berbeda

4.2. Uji Wilcoxon untuk Sampel Besar (n >15)

- Untuk sampel besar distribusi sampling untuk T akan mendekati distribusi normal dengan rata-rata dan deviasi standar sebagai berikut:

$$\mu_T = \frac{n(n+1)}{4} \qquad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

- Statistik uji : $$z = \frac{T - \mu_T}{\sigma_T}$$

- **Contoh**:

  Sebuah perusahaan berupaya meningkatkan produktivitas dengan menerapkan kontrol kualitas. Untuk meneliti apakah penerapan kontrol kualitas tersebut memang berhasil meningkatkan produksi, diambil sampel dari 20 pekerja dan dicatat produksi dari masing-masing pekerja sebelum dan sesudah penerapan kontrol kualitas tersebut. Gunakan Uji Wilcoxon dan α = 0.01 untuk membuktikan apakah kontrol kualitas tersebut memang berhasil meningkatkan produksi.

Pekerja	Before	After	d = Before - After	Pering-kat
1	5	11	-6	-19
2	4	9	-5	-17
3	9	9	0	Hapus
4	6	8	-2	-9
5	3	5	-2	-9
6	8	7	1	+3.5
7	7	9	-2	-9
8	10	9	1	+3.5
9	3	7	-4	-14.5
10	7	9	-2	-9
11	2	6	-4	-14.5
12	5	10	-5	-17
13	4	9	-5	-17
14	5	7	-2	-9
15	8	9	-1	-3.5
16	7	6	1	+3.5
17	9	10	-1	-3.5
18	5	8	-3	-12.5
19	4	5	-1	-3.5
20	3	6	-3	-12.5

Jawab :

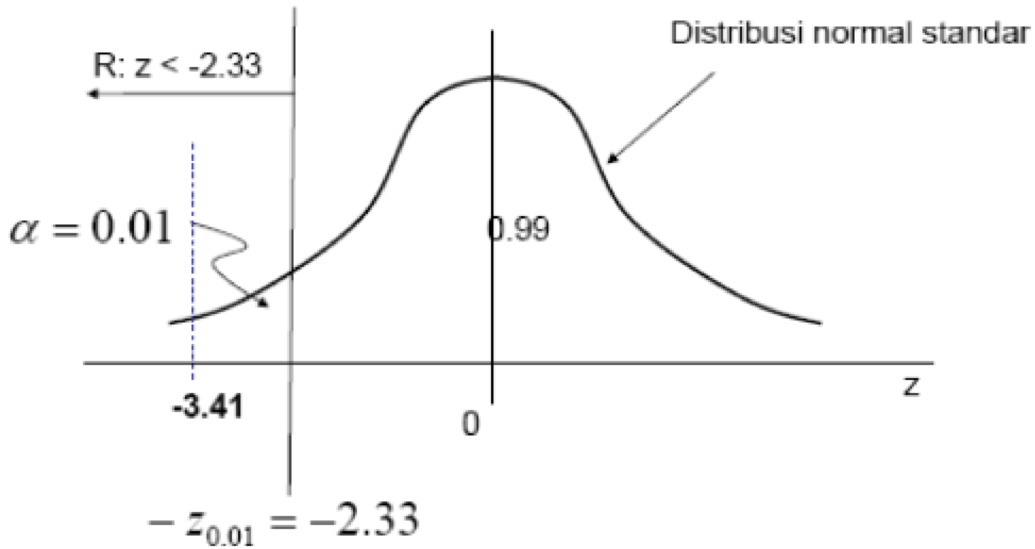
- H_0: $M_d = 0$ vs H_1: $Md < 0$
- T- = 179.5
- T+ = 10.5
- T = min(179.5, 10.5) = 10.5
- n = 19 (1 data dgn d = 0 dihapus)
- Menghitung statistik uji:

$$\mu_T = \frac{n(n+1)}{4} = \frac{19*20}{4} = 95$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{19*20*39}{24}} = 24.8$$

$$z = \frac{T-\mu_T}{\sigma_T} = \frac{10.5-95}{24.8} = -3.41$$

Dengan α = 0.01, daerah penolakan : $z < -z_{0.01} = -2.33$
Karena z terletak di daerah penolakan (-3.41 < -2.33),
maka tolak H_0. Artinya : memang benar bahwa setelah ada program kontrol kualitas, produktivitas meningkat



5. Uji Korelasi Rank Spearman

- Ukuran asosiasi antara dua variabel yang berjenis interval atau rasio digunakan: *koefisien korelasi Person*
- Untuk dua variabel berjenis ordinal, *ukuran asosiasinya* adalah *koefisien korelasi Spearman*

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

n = banyaknya pasangan data yang dicari korelasinya
d = perbedaan peringkat pada setiap pasang. Di setiap kelompok dibuat peringkatnya dari 1 sampai n.

- Interpretasi r_s sama saja dengan interpretasi r
- **Contoh 1**:

  Apakah ada hubungan kuat antara harga minyak mentah (per barrel) dan harga BBM (per galon) di pompa bensin? Untuk mengestimasi asosiasi antara kedua variabel tersebut, seorang peneliti di perusahaan minyak mengunpulkan data di sebuah kota selama 9 bulan, dan mencatat rata-rata harga di setiap bulan tersebut. Hitunglah koefisien korelasi Spearman untuk data ini.

Row	Mentah	BBM	Mentah_P	BBM_P	d	d2
1	14.60	1.05	3	1.0	2.0	4.00
2	10.58	1.06	1	2.5	-1.5	2.25
3	12.30	1.08	2	4.0	-2.0	4.00
4	15.10	1.06	4	2.5	1.5	2.25
5	18.35	1.12	5	5.0	0.0	0.00
6	22.60	1.24	6	6.0	0.0	0.00
7	28.90	1.36	8	8.0	0.0	0.00
8	31.40	1.40	9	9.0	0.0	0.00
9	26.75	1.34	7	7.0	0.0	0.00

hasil pengamatan **peringkat** **perbedaan peringkat**

$$\sum d^2 = 12.5$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6*12.5}{9(9^2-1)} = 0.89583$$


Terima Kasih
рахмат
danke
謝謝
ngiyabonga
spas
teşekkür ederim
Баярлалаа
mersi
kia ora
barka
welalin
tack
mahalo
tapadh leat
спасибо
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
kiitos
dankie
nanni
nandri
dhanyavad
thank you
gracias
хвала
asante
manana
bedankt
bayarlalaa
gràcje
hvala
mauruuru
köszönöm
akun
dankon
aćiū
obrigada
tenki
enkosi
djiere dieuf
tau
mochchakkeram
murakoze
дякую
mamnun
chokrane
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
sobodi
dèkuji
obrigado
sagolun
sukriya
kop khun krap
arigatô
takk
dakujem
trugarez
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
grazie
kam sah hamnida
rahmat
terima kasih
rahmet
ευχαριστώ
diolch
dhanyavadagalu
shukriya
merce
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
merci