BAB II

2.1 Dasar Perhitungan Statistik Non-Parametrik

Metode statistik inferensi(induktif) yaitu proses memperoleh informasi dari data sampel yang digunakan untuk menarik kesimpulan tentang populasi dari sampel yang dipilih.Teknik statistika induktif dibagi dua bagian besar yaitu: penaksiran parameter dan pengujian hipotesa.

Pada analisa statiatik inferensial,uji hipotesa dan estimasi sering di jumpai bahwa metode distribusi populasi tidak diketahui.Pada uji hipotesa atau estimasi rata-rata (mean) suatu populasi digunakan statistik :

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

uji hipotesa ini termasuk dalam keadaan distribusi bebas bukan non-parametrik,karena memuat parameter μ dan s.

Definisi 1 :

Statistik rank order k ,$X^{(k)}$ adalah statistik yang membawa elemen $X^{(k)}$terkecil di suatu harga k di dalam setiap pengamatan $(x_1,x_2,....x_n)$ dari $(X_1.X_2,.......X_n)$.

Jika $(X_1,X_2,......X_n)$ adalah suatu sampel acak maka $(X^{(1)} \leq X^{(2)} \leq .....X^{(n)})$ dinamakan order sampel acak di mana $X^{(1)}$ order statistik dari rank satu,elemen terkecil di dalam $(x_1,x_2,......x_n)$ dan $X^{(n)}$ order statistik dari rank n,elemen terbesar di dalam $(x_1,x_2,.......x_n)$.

Theorema limit central (pusat) :

Misalkan $x_1,x_2,......x_n$ adalah sampel acak dari suatu populasi sembarang dengan mean μ dan variansi σ^2.Maka untuk n ---> ∞variabel acak $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

mempunyai distribusi limit normal .Di katakan untuk n besar $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ mendekati distribusi normal

Bukti:

Pandang fpm variabel acak $\bar{n}(x-\mu)$ yakni :

$$M_{n(\bar{x}-\mu)}(t) = [M_{(\bar{x}-\mu)}(t)]^n$$

selanjutnya, misalkan M(t) menunjukan fpm variabel acak

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sigma/\sqrt{n}}[n(\bar{x}-\mu]$$

$M^*(t) = M_{n(\bar{x}-\mu)}(\frac{t}{\sigma/\sqrt{n}}) = [M_{(x-\mu)}(\frac{t}{\sigma/\sqrt{n}})]^n$ fpm yang ada untuk harga t dekat nol,mempunyai derivatip semua tingkat pada t=0;sehingga menggunakan theorema Taylor.

$$M_{(x-\mu)}(t) = M_{(x-\mu)}(0) + M^1_{(x-\mu)}(0)t + M^2_{(x-\mu)}(0)t^2/2! + M^3_{(x-\mu)}(\xi)t^2/3!$$

dimana ξ adalah bilangan antara 0 dan t,dan $M^3_{(\bar{x}-\mu)}$ menunjukan derivatip ketiga dari $M_{(x-\mu)}(t)$ yang dinilai pada ξ.kiata tulis :

$$\mathfrak{C}(t) = t\, M^3_{(x-\mu)}(\xi)/3! \quad (\xi \text{tergantung pada } t\ )$$

kita punya $M_{(x-\mu)}(0) = 1, M^1_{(x-\mu)}(0) = 0$ dan $M^2_{(x-\mu)}(0) = \sigma^2$

karena $E(x-\mu) = 0$ maka $M(x-\mu)^2 = \sigma^2$ maka :

$$M_{(x-\mu)}(t) = 1 + \frac{\sigma^2 t^2}{2} + t^2\, \mathfrak{C}(t) \text{ dimana :}$$

$\mathfrak{E}(t) = M^{3}_{(x-\mu)}(\xi)/3!$ ---> 0 jika t --->0 sehingga

$$M_{(x-\mu)}\left(\frac{t}{\sigma/\sqrt{n}}\right) = 1 + \frac{\sigma^2}{2}\left(\frac{t}{\sigma^2 n}\right) + \frac{t^2}{\sigma^2 n}\,\mathfrak{E}\left(\frac{t}{\sigma^2\sqrt{n}}\right) \text{ dan}$$

$$M^*(t) = \left[\, 1 + \frac{t^2}{2n} + \frac{t^2}{2}\,\mathfrak{E}\left(\frac{t}{\sigma^2\sqrt{n}}\right) \right]^n$$

kita tahu $n \longrightarrow \infty$, $M^*(t) \longrightarrow e^{t^2/2}$

berarti limit fpm $(\bar{x}-\mu)/(\sigma\sqrt{n})$ untuk $n \longrightarrow \infty$ adalah sama dengan fpm variabel acak normal standar,maka di simpulkan bahwa distribusi limit varibel $(\bar{x}-\mu)/(\sigma\sqrt{n})$ adalah normal standar.

Theorema :

Misalkan $x_1, x_2, \ldots\ldots x_n$ adalah variabel acak independen masing masing berdistribusi chi-kuadrat dengan derajat bebas $k_1, k_2, \ldots\ldots k_n$.

maka variabel acak :

$$Y = x_1 + x_2 + \ldots\ldots + x_n$$

berditribusi chi-kuadrat dengan dk $k = k_1 + k_2 + \ldots\ldots + k_n$

Bukti :

fpm variabel acak chi-kuadrat dengan derajat bebas k_i adalah $(1-2t)^{-k/2}$,jika fpm x_i adalah $M_{x_i}(t)$ maka fpm Y adalah : $M_Y(t) = M_{x_1}(t)\, M_{x_2}(t) \ldots\ldots M_{x_n}(t)$

$$= (1-2t)^{-k_1/2}\,(1-2t)^{-k_2/2} \ldots\ldots\ldots (1-2t)^{-k_n/2}$$

$$= (1-2t)^{-\sum_{i=1}^{n} k_i/2} = (1-2t)^{-k/2}$$

fpm variabel acak berdistribusi chi-kuadrat dengan derajat bebas $k = \sum_{i=1}^{n} k_i$, jika Y berdistribusi chi-kuadrat dengan derajat bebas k.

2.1.1 **Uji Hipotesa Statistik Non-Parametrik**

Uji hipotesa merupakan proses penyimpulan dari suatu sampel dapat tidaknya pernyataan/gambaran suatu populasi di terima.Suatu hipotesa yang baik didasarkan pada sifat-sifat

1. Uji harus tak bias

$$E_P[g(x_1,x_2,...x_n)] = \theta$$

dimana $x_1,x_2,....x_n$ variabel acak distribusi P dan θ parameter yang di tentukan distribusi P,untuk setiap P $\in \mathcal{P}$

2. Uji Konsisten

$$\lim_{n \to \infty} P \{ | g_n(x_1,x_2,.....x_n)-\theta | \leq \delta \} = 1$$

dimana $x_1,x_2,......x_n$ variabel acak distribusi P dan θ parameter yang ditentukandistribusi P,untuk P $\in \mathcal{P}$, $\delta > 0$

3. Uji harus lebih efisien di banding uji yang lain.

2.2 **Dua Sampel Independent**

Uji Mann-Whitney yaitu mengenai dua sampel yang diambil dari dua populasi.Perbedaan ke dua populasi akan diamati berdasarkan sampel acak yang diambil dari populasi-populasi itu.Jika data yang diamati terdiri dari skala pengamatan oedinal,maka perbedaan antara dua populasi itu biasanya di uji berdasarkan perbedaan lokasi.Secara intuitif/mendalam uji dilakukan dengan menggabungkan dua populasi,kemudian di ranking dan sampel acak terkecil menuju terbesar,Jika terdapat ranking sama,diambil rata-ratanya.Uji statistik yang digunakan membandingkan jumlah rank sampel pertama dan jumlah rank sampel kedua setelah keduanya digabungkan.

2.2.1 Uji Mann-Whitney

Dua sampel acak $x_1, x_2, \ldots.., x_n$ untuk populsi 1 berukuran n , $y_1, y_2, \ldots..., y_m$ untuk populasi 2 berukuran m. $R(x_i)$ menyatakan rank dari x_i untuk semua i dan $R(y_j)$ menyatakan rank dari y_j untuk semua j , rank dari 1 sampai $N = m + n$

Anggapan :

1. Dua sampel adalah sampel acak dari masing-masing populasi
2. Kedua sampel saling independent
3. Paling sedikit skala pengamatan ordinal
4. Jika ada perbedaan antara fungsi distribusi populasi didalam lokasi dari distribusi tersebut,hal itu jika F(x) tidak sama dengan G(x) tetapi F(x) = G(x+c),c=konstanta.

Hipotesa

A. Ho : $E(X) = E(Y)$

H1 : $E(X) \neq E(Y)$

B. Ho : $E(X) \geq E(Y)$

H1 : $E(X) < E(Y)$

C. Ho : $E(X) \leq E(Y)$

H1 : $E(X) > E(Y)$

Uji Statistik Satu-Ekor

Definisi Lemma 1 :

$$\sum_{i=a}^{N} i = \frac{(N+a)(N-a+1)}{2} \quad \text{dan} \quad \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$$

Definisi Lemma 2 :

$$\sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N-1)}{6}$$

Jika tidak ada nilai sampel sama(tie) atau tie sedikit,maka jumlah rank populasi satu :

$$T = \sum_{i=1}^{n} R(X_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

jika terdapat banyak tie :

$$T_1 = \frac{T - \dfrac{n(N+1)}{2}}{\sqrt{\dfrac{nm}{N(N+1)} \sum_{i=1}^{N} R_i^2 - \dfrac{nm(N+1)^2}{4(N-1)}}}$$

Uji Statistik Dua-Ekor

Kuantil bawah daru T diberikan tabel A7,sedangkan kuantil atas dengan rumus :

$$W_{1-p} = n(N+1) - W_p \quad \text{atau} \quad T^1 = n(N+1) - T$$

Keputusan :

A. Uji Dua-Ekor

Ho ditolak jika $T < W_{\alpha/2}$ atau $T > W_{1-\alpha/2}$

Ho diterima jika $W_{1-\alpha/2} < T < W_{\alpha/2}$

B. Uji Satu-Ekor

Jika nilai T kecil indikasi bahwa H_1 benar.

Ho ditolak jika $T < W_\alpha$

Ho diterima jika $T \geq W_\alpha$

C. Uji Satu-Ekor

Jika nilai T besar atau T^1 kecil indikasi bahwa H_1 benar

Ho ditolak jika $T > W_{1-\alpha}$ atau $T^1 < W_\alpha$

Ho diterima jka $T \leq W_{1-\alpha}$

2.2.2 Contoh

Sebuah sekolah tinggi swasta,satu kelas berisi 48 anak laki-laki,12 anak tinggal dipedesaan dan 36 tinggal diperkotaan.Diadakan uji fisik untuk melihatbahwa umumnya anak desa lebih baik dari anak kota.

Uji Mann-Whitney

Ho : anak desa tidak cenderung lebih kuat fisiknya dari anak kota.

H_1 : anak desa cenderung lebih kuat dari anak kota.

Data yang telah diurutkan,dimana score anak desa (X) dan score anak kota (Y).

X	Y	Rank	X	Y	Rank	X	Y	Rank
	1,0	1		6,2	17		11,3	33
	1,8	2	6,3		18	11,4		34
	2,1	3		6,4	19		11,8	35
	2,4	4		6,7	20,5	12,5		36
	2,6	5		6,7	20,5		12,6	37
2,7		6	7,3		22		12,7	38
	3,2	7		7,6	23	12,9		39
	3,6	8		7,9	24		14,2	40
	4,0	9		8,3	25		14,8	41,5

4,2		10	9,0		26	14,8		41,5
	5,0	11		9,1	27		15,3	43
	5,6	13		9,9	28		16,0	44
	5,6	13		10,6	30,5	16,1		45
	5,6	13		10,6	30,5		16,9	46
	5,9	15	10,6		30,5		17,7	47
	6,1	16		10,6	30,5		18,6	48

Diketahui : n = 12 , m = 37 , N = 12 + 36 = 48

Uji satu ekor

$$T = \sum_{i=1}^{12} R(x_i) = 321$$

$$\sum_{i=1}^{48} R_i^2 = 38016$$

$$T_1 = \frac{321 - 12\,\dfrac{49}{2}}{\sqrt{\dfrac{(12)(36)}{(48)(47)}(38016) - \dfrac{(12)(36)(49)^2}{4(47)}}} = 0,6431$$

Untuk α = 0,05 daerah kritis $W_{0,95}$ = 1,6449 (Tabel A1),jelas T_1 = 0,6431 < $W_{0,95}$ = 1,6449.Jadi Ho diterima,pada T_1 = 0,6431 ; α = 0,74 (tabel A_1) dan daerah penolakan $\hat{\alpha}$ = 0,26

Kesimpulan : anak desa tidak cenderung lebih kuat fisiknya dari anak kota.

Uji dua-ekor

Untuk α = 0,05 dari rumus tabel A7,

$$W_{1-\alpha} = \frac{n(N+1)}{2} + X_{1-\alpha}\sqrt{\frac{nm(N+1)}{2}}$$

$$W_{0,95} = 294 + (1,6449)(42) = 363,1 \quad \text{(kuantil atas)}$$

$$W_p = n(N+1) - W_{1-p}$$

$$W_{0,05} = 588 - 363,1 = 224,9 \quad \text{(kuantil bawah)}$$

jadi $W_{0,05} < T < W_{0,95}$ maka Ho diterima.

2.2.1 INTERVAL KEPERCAYAAN SELISIH DUA MEAN

Uji Mann - Whitney untuk interval kepercayaan selisih dua mean Ho : $E(X) = E(Y) + d$

Data terdiri dari n sampel untuk variabel X_i dan m untuk variabel Y_i maka didapat nm pasang (X_i, Y_j). k menunjukkan banyaknya pasangan dimana $X_i > Y_j$, $X_i - Y_j > 0$ dari $T = \sum_{i=1}^{n} R(X_i)$ maka jumlah rank X_s adalah $k + \frac{n(n+1)}{2}$ bila $Y_s > X_s$ maka $T = 1+2+3+........+n = \frac{n(n+1)}{2}$ (lemma 1) akibat dari adanya k pasang (X_i, Y_j) dimana $Y = \frac{m(m+1)}{2}$ lebih kecil $S = k + \frac{n(n+1)}{2}(Y<S)$ yang merupakan perluasan untuk T dari k unit, maka $T = k + \frac{n(n+1)}{2}$.

Nilai T dimana Ho : $E(X) : E(Y) + d$ diterima diberikan tabel A7 pada $W_{\alpha/2}$ maka $k = W_{\alpha/2} + \frac{n(n+1)}{2}$. Nilai d didapat dari Y_s yang mewakili Y_j sebanyal k pasang (X_i, Y_j+d) memenuhi $X_i > Y_j+d$, $X_i - Y_j > d$ jika diambil Y_s TERBESAR untuk selish $X_i - Y_j$, $Y_s > X_s$ diperoleh harga d terkecil dinamakan "selisih terkecil" dan sebaliknya paling sedikit k pasang (X_i, Y_j+d) memenuhi $X_i > Y_j+d$ atau $X_i \geq Y_j+d$ dinamakan "selisih terbesar", dari kumpulan-kumpulan harga d diperoleh interval kepercayaan selisih dua mean.

Data, terdiri dari dua sampel acak $X_1, X_2, ...X_n$ dan X variabel acak distribusi X_i ; $Y_1, Y_2, ...Y_m$ dan Y variabel acak distribusi Y_j.

Anggapan :

1. Sampel acak dari masing-masing populasi
2. Kedua sampel saling independent
3. Fungsi distribusi dua populasi sama

Metode :

$$k = W_{\alpha/2} - \frac{n(n+1)}{2}$$

dimana $W_{\alpha/2}$ diberikan tabel A_7 untuk ukuran n dan m, interval kepercayaan yang diharapkan 1-α.

Dari k pasang (X_i, Y_j) dicari selisih terbesar dan selisih terkecil, $d = X_i - Y_j$ data diurutkan dari kecil ke besar.

Perhitungan selisihterkecil sebanyak k , harga d terbesar sebagai unit bawah (L).

Perhitungan selisih terbesar sebanyak k, harga d terkecil sebagai unit atas (U)

Interval diberikan :

$$P [L \leq E(X)-E(Y) \leq U] \geq 1-\alpha$$

Contoh :

5 tukang gempur menghaluskan dengan mixer A.

5 tukang gempur batu dengan mixer B.

(X_i) Mixer A (Dalam Menit)	(Y_i) Mixer B (dalam menit)
7,3	7,4
6,9	6,8
7,2	6,9
7,8	6,7
7,2	7,1

Perhitungan :

$$K = 18 - (5)(6)/2 = 3$$

dimana $W_{0,025}$ = 18 (Tabel A7) n=5, m=5, α=0,05

Urutan sampel terkecil menuju terbesar

X_i	Y_i
6,9	6,7
7,2	6,8
7,2	6,9
7,3	7,1
7,8	7,4

Selisih terkecil (X_s- Y_s)	Selisih terbesar (X_s- Y_s)
6,9 - 7,4 = -0,5	7,8 - 6,7 = 1,1
6,9 - 7,1 = -0,2	7,8 - 6,8 = 1,0
7,2 - 7,4 = -0,2 = L	7,8 - 6,9 = 0,9 = U

Intervel Konfedensi 95% (L,U) untuk E(X) - E(Y) adalah (-0,2 ; 0,9)

$$P [-0,2 \le E(X) - E(Y) \le 0,9] \ge 1 - 0,05$$

2.3. BEBERAPA SAMPEL INDEPENDEN.

Uji Kruskal-Wallis untuk k sampel independen, dimana $k \ge 2$, seandainya ada K populasi, akan diuji bahwa k populasi itu mempunyai distribusi yang sama.

2.3.1. UJI KRUSKAL-WALLIS

Data terdiri dari k sampel acak, besarnya sampel belum tentu sama. Seandainya sampel acak ke-i besarnya n_i, X_1; X_2, ..., X_{ni} , data-data disajikan sebagai berikut :

sampel 1	sampel 2	....	sampel k
X_{11}	X_{21}	....	X_{k1}
X_{12}	X_{22}	....	X_{k2}

$$
\begin{array}{cccc}
X_{13} & X_{23} & \cdots\cdots\cdots & X_{k3} \\
\vdots & \vdots & \cdots\cdots\cdots & \vdots \\
X_{1n_1} & X_{2n_2} & \cdots\cdots\cdots & X_{3n_k}
\end{array}
$$

Jumlah pengamatan $N = \sum_{i=1}^{k} n_i$ ....(1)

Kemudian k sampel digabungkan, di beri rank 1 sampai N, setelah diuraikan dan terkecil menuju terbesar.

Jika K (X_{ij}) untuk pengamatan X_{ij} maka

$$R_i = \sum_{j=1}^{k} R(X_{ij}) \quad , \text{ untuk } i = 1,2,\ldots,k \quad .....(2)$$

dimana R_i = Jumlah rank tiap sampel

Anggapan

1. Sampel adalah sampel dari masing-masing Populasi
2. Sampel saling independen
3. Skala pengukuran paling sedikit ordinal
4. k Populasi, fungsi distribusi yang sama atau ada populasi cenderung mempunyai harga yang lebih besar dibanding dengan populasi yng lain

Hypotesa

Ho : Semua k populasi mempunyai fungsi distribusi sama

H1 : Paling sedikit ada satu populasi cenderung mempunyai pengamatan yang lebih besar dibandingkan dengan sedikit satu populasi yang lain

atau

H1 : k Populasi tidak semuanya fungsi distribusi sama

UJI STATISTIK

$$T = \frac{1}{S^2} \left[\sum_{i=1}^{k} \frac{R_i^2}{n_i} - \frac{N(N+1)}{4} \right] \ldots\ldots\ldots\ldots\ldots\ldots (3).$$

$$S^2 = \frac{1}{N-1} \left[\sum_{\substack{\text{semua} \\ \text{rank}}} R\,(X_{ij})^2 - \frac{N(N+1)}{4} \right] \ldots\ldots\ldots\ldots (4)$$

jika tidak ada tie

$$T = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

Keputusan

Ho ditolak jika T > $X^2_{1-\alpha(k-1)}$, distribusichi-kuadrat (tabel A_2). Jika k = 3 dan $n_i \leq 5$, i = 1,2,3,... untuk harga α, dicari dari tabel A_i. Ho ditolakjika T ≥ $W\alpha$ (tabel AB)

PERBANDINGAN MULTIPLE

Digunakan jika Ho ditolak, dan populasi i dan j berbeda jika dipenuhi :

$$\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right| > t_{1-\alpha} \left[S^2 \frac{N-1-T}{N-k} \right]^{1/2} \left[\frac{1}{n_i} + \frac{1}{n_j} \right]^{1/2} \ldots\ldots\ldots (6)$$

dimana R_i dan R_j diketahui $t_{1-\alpha/2}$, dk = N - k dari distribusi t (tabel A_{25})

contoh

Empat pengamatan hasil pertumbuhan gandum per area dalam tanah dimana menggunakan empat metode

Data

1		2		3		4	
Observasi	Rank	Observasi	Rang	Observasi	Rang	Observasi	Rank
83	11	91	23	101	34	78	2
91	23	90	19,5	100	33	82	9
44	20,5	81	6,5	91	23	81	6,5
89	17	83	11	93	27	77	1
89	17	84	13,5	96	31,5	74	3
96	31,5	88	11	95	30	81	6,5
91	23	91	15	94	28,5	88	4
92	26	89	17			81	6,5
90	14,5	84	13,5				
ki :	196,5		133,0		207,0		36,5
ni :	9		10		7		8

$$T = \frac{12}{34(34+1)} \left[\frac{(196,5)^2}{9} + \frac{(153)^2}{10} + \frac{(207)^2}{7} + \frac{(38,5)^2}{8} \right] - 3(34+1)$$
$$= 25,46$$

untuk $\alpha = 0,05$, $X_{0,95} = 7,815$, dk = k-1 = 3 (tabel A_2)

jelas $T = 25,46 > X_{0,95} = 7,815$. Ho ditolak

PERBANDINGAN MULTIPLE

Populasi	$\left\| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right\|$		$2,041 \left[\frac{99,167(33-25,464)}{34-4} \right]^{1/2} \left[\frac{1}{ni} + \frac{1}{nj} \right]^{1/2}$
1 dan 2	6,533	>	4,681
1 dan 3	7,736	>	5,134
1 dan 4	17,021	>	4,950
2 dan 3	14,271	>	5,020

2 dan 4	10,480	>	4,832
3 dan 4	24,754	>	5,272

Artinya setiap pasangan berbeda

Kesimpulan ada metode dan pertumbuhan gandūm cenderung lebih baik dari yang lain

2.3.2 UJI KRUSKAL-WALLIS DENGAN TABEL KEMUNGKINAN

Untuk situasi banyak tie, dimana baris menunjukkan katagori dan kolom menunjukkan populasi

Populasi		1	2	3 ...	4	Total baris	$\bar{R}_1$ = rata-rata rank
katagori	1	O_{11}	O_{12}	O_{13} ...	O_{1k}	t_1	$\frac{t_1+1}{2}$
	2	O_{21}	O_{22}	O_{23} ...	O_{2k}	t_2	$t_1 + \frac{(t_1+1)}{2}$
	3	O_{31}	O_{32}	O_{33} ...	O_{3k}	t_3	$t_1 + t_2 + \frac{(t_2+1)}{2}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	c	O_{c1}	O_{c2}	O_{c3}	O_{ck}	t_c	$\sum_{i=1}^{c} t_i + \frac{(t_c+1)}{2}$
Kolom total		n_1	n_2	n_3 ...	n_k	N = total	

UJI STATISTIK

Jumlah rank

$$R_j = \sum_{i=1}^{c} O_{ij}\,\bar{R}_i \quad \ldots\ldots (9)$$

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{c} t_i \bar{R}_i^2 - \frac{N(N+1)^2}{4}\right] \quad \ldots\ldots (10)$$

$$T = \frac{1}{S^2}\left[\sum_{i=1}^{c} \frac{R_j^2}{n_j} - \frac{N(N+1)^2}{4}\right] \quad \ldots\ldots (11)$$

catatan : S^2 nilainya sama untuk pernyataan (4) dan (10)

Contoh:

Tiga bentuk membandingkan kualitas murid-muridnya, pada akhir semester untuk dilihat jika ada dari mereka menunjukkan kualitas rendah dari yang lain.

Hipotesa :

Ho : Tiga instruktur rrata-rata kualitasnya sama untuk setiap muridnya

H_1 : Ada instruktur kualitas muridnya rendah dari yang lain

Hasil Ujian

Kualitas	Instruktur 1	Instruktur 2	Instruktur 3	Total baris	Rata-rata Rank
A	4	10	6	20	10,5
B	14	6	7	27	34
C	17	7	8	34	64,5
D	6	7	6	19	91
E	2	6	1	9	105
Total Murid	43	38	28	109	

Perhitungan

$$R_j = \sum_{i=1}^{5} O_{ij} \bar{R}_i$$

R_1 = (4 x 10,5) + (14 x 34) + (17 x 64,5) + (2 x 105) = 2370,5

R_2 = (10 x 10,5) + (6 x 34) + (9 x 91) + (6 x 105) = 2156,5

R_3 = (6 X 10,5) + (7 x 34) + (6 x 91) + (1 x 105) = 1468

$$S^2 = \frac{1}{109-1} - [[(20 \times (105)^2) + (27 \times (34)^2) + (34 \times (64,5)^2 + (19 \times (91)^2) + (9 \times (105)^2) - 109 \frac{(109+1)}{4}$$

$S^2 = 941,71$

$$T = \frac{1}{941,71}\left[\frac{(2370,5)^2}{43} + \frac{(2156,5)^2}{38} + \frac{(1468)^2}{28} - \frac{(109+1)^2}{4}\right]$$

$= 0,3209$

untuk $\alpha = 0,05$, dk = 2, $X_{0,975} = 5,991$ (tabel A_2)

Jelas $X_{0,975} = 5,991 > T = 0,3209$. Jadi Ho diterima.

Kesimpulan : Tidak dapat dikatakan instruktur yang satu lebih berkualitas atau kurang berkualitas dari yang lainnya.

2.3.3. TEORY MANN WHITNEY DAN KRUSTAL WALLIS

Jika X_i dan Y_j, independen dan mempunyai distribusi sama, maka setiap aturan untuk x dan y di dalam sampel adalah mempunyai probabilitas sama. Ini yang menjadi prinsip dasar pada uji rank. Banyaknya pengamatan X_i adalah n dan Y_j adalah m, jadi besar sampel gabungan adaalah n+m. Rank sampel gabungan akan merupakan pilihan acak n bilangan bulat 1 sampai n+m. Setiap rank mempunyai kesempatan yang sama untuk dipilih sebagai rank X_i, karena harus memilih n buah rank untuk semua anggota X_i, maka distribusi S dapat dicari seperti fungsi probabilitas jumlah n bilangan bulat yang diambil secara acak tanpa pengembalian dari bilangan bulat 1,2,3,4,...,n+m ($S = \sum_{i=1}^{n} R(X_i)$) banyaknya cara memilih n bilangan bulat dari (n+m), bilangan bulat yang tersedia adalah $\binom{n+m}{n}$ dan setiap cara mempunyai probabilitas sama, terdapatlah P(s=k) sama dengan banyaknya himpunan n bilangan bulat yang jumlahnya k dibagi dengan $\binom{n+m}{n}$.

Seandainya $n = 4$ dan $m = 6$ yaitu sampelnya (X_1, X_2, X_3, X_4) dan $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6)$ kedua sampel digabung. Rank untuk sampel pertama (X_i) dapat terjadi seperti kalau dipilih 4 bilangan bulat dan 10 bilangan bulat 1,2,...,10 maka akan ada $\binom{10}{4} = \frac{10\,!}{4!6!} = 210$ macam. Jadi akan ada 210 macam harga S yang dapat terjadi. Seandainya S = 14 ini artinya jumlah rank untuk X_i dalam sampel gabungan = 14 yaitu (1,2,3,8) atau (1,2,4,7) atau (1,2,5,6) atau (2,3,4,5) disini terdapat 5 macam, sehingga $P(S=14) = \frac{5}{210}$ ternyata S terkecil adalah 10 yaitu (1,2,3,4), maka dapat dicari

$$P(S=10) = \frac{1}{210} \longrightarrow P(T=0) = \frac{1}{210} \longrightarrow P(T \leq 0) = \frac{1}{210}$$

$$P(S=11) = \frac{1}{210} \longrightarrow P(T=1) = \frac{1}{210} \longrightarrow P(T \leq 1) = \frac{2}{210}$$

$$P(S=12) = \frac{1}{210} \longrightarrow P(T+2) = \frac{1}{210} \longrightarrow P(P \leq 2) = \frac{4}{210}$$

dan seterusnya

Karena T adalah jumlah rank dari nXs, maka n dan m besar berdasarkan teorema limit central (pusat) maka ;

$$X = \frac{T - E(T)}{\sqrt{var(T)}}$$

dimana $E(T) = \dfrac{n(m+n+1)}{2}$ ....(1)

$Var(T) = \dfrac{nm(m+n+1)}{12}$ ....(2)

Jadi kuartil ke α untuk T dapat didekati dengan rumus

$$W_\alpha \cong E(T) + X_\alpha \sqrt{Var(T)}$$

$$W_\alpha = E(T) + X_\alpha \sqrt{\frac{nm(n+m+1)}{12}}$$

X_α kuartil ke α untuk variabel acak berdistribusi normal standar, dicari dari tabel normal.

Mann-Whitney menggunakan uji dua sampel (X,Y) :

Ho : $E(X) = E(Y)$	Ho : $E(X) - E(Y) = d$
Ho : $E(X) \geq E(Y)$	Ho : $E(X) = E(Y) + d$
Ho : $E(X) \leq E(Y)$	

dimana d adalah nilai khusus dari $E(X) - E(Y)$ atau kumpulan semua nilai d dimana diperoleh dari X_s dan Y_s, sehingga kita mempunyai interval konfidensi untuk $E(X) - E(Y)$ selisih dua mean.

Uji Mann-Whitney untuk dua populasi akan berkembang menjadi uji Kruskall-Wallis ynag digunakan untuk lebih dari dua populasi disini akan diperluas untuk k sampel independent dimana $k \geq 2$. seandainya ada k populasi, akan diuji bahwa k populasi itu mempunyai distribusi yang sama, dan setiap populasi diambil sampel independent. Uji Kruskall-Wallis mendasarkan pada rank setelah sampel-sampel itu digabungkan, seperti yang telah dilakukan pada uji Mann-Whitney.

Maka untuk n_i besar, R_i = Jumlah rank n_i Variabel acak, berlaku limit pusat

$$\frac{R_i - E(R_i)}{\sqrt{Var(R_i)}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

mendekati distribusi normal standar variabel acak untuk H_o benar, di mana

$$E(R_i) = \frac{n_i \ (N+1)}{2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

dan

$$Var\ (R_i) = \frac{n_i\ (N+1)(N-n_i)}{12} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

jadi :

$$\left[\frac{R_i - E(R_i)}{\sqrt{Var\ (R_i)}} \right]^2 = \frac{\{ R_i - [n_i(N+1)/2] \}^2}{n_i(N+1)(N-n_i)/12} \ldots\ldots\ldots (4)$$

adalah mendekati distribusi chi kwadrat, dk = 1.
Jika R_i Independent jumlah semua distribusi adalah

$$T' = \sum_{i=1}^{k} \frac{\{ R_i - [n_i(N+1)/2] \}^2}{n_i(N+1)(N-n_i)/12} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Pendekatan menggunakan distribusi chi kwadrat dk = k .

Jumlah dari R_{is} adalah N(N+1)/2 , tahun 1952 kruskall menunjukkan bahwa jika i di dalam batas T′ dikali dengan $(N-n_i)/N$ untuk i = 1,2,....,k didapat :

$$T = \sum_{i=1}^{k} \frac{\{ R_i - [n_i(N+1)/2]\}^2}{n_i(N+1)N/12} \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

mendekati distribusi che kuadrat dk = k - 1

2.4 UKURAN KORELASI RANK

Ukuran korelasi merupakan suatu variabel acak yang menyatakan keterkaitan dua atau lebih variabel acak. Seandainya sampel acak terdiri dari n pengamatan bivariat : (X_1,Y_1) , (X_2,Y_2) , ... , (X_n,Y_n) dan (X_i,Y_j) berdistribusi yang sama ,ukuran korelasi antara X dan Y harus memenuhi syarat :

1.Ukuran korelasi antara -1 dan 1.

2.Bila harga X makin besar berpasangan dengan Y yang juga makin besar maka korelasi dikatakan positif,atau ukuran korelasi positif.

3.Bila harga X makin besar, berpasangan dengan Y yang makin kecil maka korelasi dikatakan negatif atau ukuran korelasi negatif.

4.Bila harga X tampak berpasangan secara acak dengan Y maka ukuran korelasi mendekati nol.Hal ini terjadi bila X dan Y independent atau sering dikatakan X dan Y tidak berkorelasi.

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})\right]^{1/2}} \quad \text{atau}$$

$$= \frac{\sum_{i=1}^{n}X_iY_i-(\sum_{i=1}^{n}X_i)(\sum_{i=1}^{n}Y_i)}{\left[\left(\sum X_i^2-\frac{(\sum X_i)^2}{n}\right)\left(\sum Y_i^2-\frac{(\sum Y_i)^2}{n}\right)\right]^{1/2}} \quad \text{atau}$$

$$= \frac{\mathrm{Cov}\ (X,Y)}{\sqrt{\mathrm{Var}(X)\ \mathrm{Var}(Y)}}$$

2.4.1 KORELASI SPEARMAN , ρ (RHO)

Data terdiri dari n pengamatan bivariat (X_1,Y_1) , (X_2,Y_2), ... (X_n,Y_n). Seandainya $R(X_i)$ rank X_i sesuai dengan urutannya dari 1 sampai n dengan cara yang sama $R(Y_i)$ rank Y_i.

Jika dalam data terdapat " tie " berikan rank rata-rata kepada pengamatan yang sama .

Ukuran korelasi spearman ρ (rho) , bila tidak ada "tie" :

$$\rho = \frac{\sum_{i=1}^{n}\left[R(X_i) - \frac{n+1}{2}\right]\left[R(Y_i) - \frac{n+1}{2}\right]}{\frac{n\ (n^2 - 1)}{12}} \quad \ldots\ldots\ldots\ldots(1)$$

$$\rho = 1 - \frac{6\ \sum_{i=1}^{n}\left[\ R(X_i) - R(Y_i)\right]^2}{n\ (n^2 - 1)} = 1 - \frac{6T}{n\ (n^2 + 1)} \quad \ldots.(2)$$

dimana

$$T = \sum_{i=1}^{n}\left[\ R(X_i) - R(Y_i)\ \right]^2$$

bukti :

$$\Sigma[R(X_i) - R(Y_i)]^2 = [\ (\ R(X_i) - \frac{n+1}{2}) - (\ R(Y_i) - \frac{n+1}{2})]^2$$

$$\Sigma\ \Big\{(\ R(X_i) - \frac{n+1}{2})^2 + (\ R(Y_i) - \frac{n+1}{2}\ )^2 -$$

$$2\ [\ (\ R(X_i) - \frac{n+1}{2})\ \ R(Y_i) - \frac{n+1}{2})]\ \Big\}$$

$$\Sigma\ \ (\ R(X_i) - \frac{n+1}{2})^2 + \Sigma\ (\ R(Y_i) - \frac{n+1}{2})^2 -$$

$$2\ \ \Sigma\ [\ (\ R(X_i) - \frac{n+1}{2})\ (\ R(Y_i) - \frac{n+1}{2})]$$

$$2 \sum \left[\left(R(X_i) - \frac{n+1}{2} \right) \left(R(Y_i) - \frac{n+1}{2} \right) \right]$$

$$\sum_{i=1}^{n} R(X_i) = 1 + 2 + \ldots + n = \frac{n(n+1)}{2}, \text{ Mean : } \frac{\sum R(X_i)}{n} = \frac{n+1}{2}$$

$$\sum \left(R(X_i) - \frac{n+1}{2} \right)^2 = \sum \left(R(X_i)\right)^2 - 2 \sum R(X_i)\left(\frac{n+1}{2}\right) + \sum \left(\frac{n+1}{2}\right)$$

$$\sum \left(R(X_i)\right)^2 - 2 \left(\frac{n+1}{2} \right) \sum R(X_i) + n\left(\frac{n+1}{2} \right)^2$$

$$\frac{n(n+1)(2n+1)}{6} - 2 \frac{(n+1)}{2} \frac{n(n+1)}{2} + \frac{n(n+1)^2}{2}$$

$$\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} = \frac{n(n^2-1)}{12}$$

$$\sum \left(R(X_i) - R(Y_i) \right)^2 = \frac{n(n^2-1)}{12} + \frac{n(n^2-1)}{12} -$$

$$2 \sum \left(R(X_i) - \frac{n+1}{2}\right)\left(R(Y_i) - \frac{n+1}{2}\right)$$

$$\frac{n(n^2-1)}{6} - 2 \sum \left(R(X_i) - \frac{n+1}{2}\right) \left(R(Y_i) - \frac{n+1}{2}\right) \times \frac{6}{n(n^2-1)}$$

$$\frac{6 \sum \left(R(X_i)\ R(Y_i) \right)^2}{n(n^2-1)} = 1 - \frac{12 \sum \left(R(X_i) - \frac{n+1}{2}\right)\left(R(Y_i) - \frac{n+1}{2}\right)}{n\ (\ n^2 - 1\ )}$$

$$\frac{\sum \left(R(X_i) - \frac{n+1}{2}\right)\left(R(Y_i) - \frac{n+1}{2}\right)}{\frac{n(n^2-1)}{12}} = 1 - \frac{6 \sum \left(R(X_i) - R(Y_i)\right)}{n(n^2-1)}$$

Jika banyak "tie" digunakan

$$\rho = \frac{\sum_{i=1}^{n} R(X_i)\ R(Y_i) - n \left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n} \left(R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2 \right) \right)^{1/2} \left(\sum_{i=1}^{n} \left(R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2 \right) \right)^{1/2}}$$

HYPOTESA

A . Ho : X_i atau Y_i independen.

H1: X_i atau Y_i tidak independen, X_i dan Y_i makin besar atau X_i dan Y_i makin kecil

B . Ho : X_i dan Y_i independen (positip)

H1 : ada kecenderungan X_i makin besar dikaitkan dengan Y_i makin besar pula

C . Ho : X_i dan Y_i independen (negatip)

H1 : ada kecenderungan X_i yang makin besar dikaitkan dengan Y_i yang makin kecil

UJİ STATISTIK

A . Ho ditolak jika $\rho < W_{\alpha/2}$ atau $\rho > W_{1-\alpha/2}$

B . Ho ditolak jika $\rho > W_{1-\alpha/2}$

C . Ho ditolak jika $\rho < W_{\alpha}$

Harga-harga kwantil diberikan pada tabel A_{10}

A . Ho ditolak jika $T < W_{\alpha/2}$ atau $T > W_{1-\alpha/2}$

B . Ho ditolak jika $T \leq W_{\alpha}$

C . Ho ditolak jika $T > W_{1-\alpha}$

Harga-harga kuantil diberikan pada tabel A_{11}

Contoh :

12 pasang anak kembar diberi tes fsikologi untuk mengetahui apakah anak yang lahir tahun pertama lebih agresip dibanding yang lahir kedua.

Data agresifitas :

lahir pertama (X_i)	lahir kedua (Y_i)	$R(X_i)$	$R(Y_i)$
86	88	8	10
71	77	3,5	7
77	76	6,5	6
68	64	1	1
91	96	11,5	12
72	72	5	4,5
77	65	6,5	2,5
91	90	11,5	11
70	65	2	2,5
71	80	3,5	8
88	81	10	9
87	72	9	4,5

Perhitungan

Hypotesa

Ho : Ukuran agresifitas dari anak kembar adalah independen

H1 : Ada korelasi positip atau negatip antara ukuran agresifitas anak kembar

$$T = \sum_{i=1}^{n} [\ R(X_i) - R(Y_i)\ ]^2$$

$$= (8-10)^2+(3,5-\&)^2+(6,5-6)^2+(1-1)^2+ \ \ldots\ + (9-4,5)^2$$

$$= 75$$

$$\rho = 1 - \frac{6(75)}{12(144-1)}$$

$$= 0,7378$$

untuk α = 0,05 , n=12 , $W_{0,975}$ = 0,5804 (tabel A_{10})

jelas ρ = 0,7378 > $W_{0,975}$ = 0,5804 . Jadi Ho ditolak untuk α = 0,05 , n=12 , $W_{0,025}$ = 120 (tabel A_{11})

JELAS t = 75 < $W_{0,025}$ = 120. Jadi Ho ditolak

Kesimpulan : ada korelasi positip antara agresifitas anak kembar

2.4.2 **korelasi kendall** , τ

Ukuran korelasi Kendall τ berdasar pada order dari pengamatan, tidak harga angka-angka itu sendiri.

Data terdiri dari sampel acak bivariat sebesar n :

(X_1,Y_1) , (X_2,Y_2), ... , (X_n,Y_n) .

- Dua pengamatan konkordan, jika $X_i > X_j$ dan $Y_i > Y_j$

  contoh (1,3 ; 2,2) dan (1,6 ; 2,7)

- Dua pengamatan tak konkordan jika $X_i > X_j$ dan $Y_i > Y_j$ atau $X_i < X_j$ dan $Y_i < Y_j$

  Contoh (1,3 ; 2,2) dan (1,6 ; 1,1)

  (1 ; 3) dan (4 ; 2)

  (9 ; 4) dan (7 ; 8)

- Pasangan tidak konkordan dan tak konkordan jika $X_i = X_j$ dan $Y_i = Y_j$

  Contoh (2 , 5) dan (2 , 3)

  (4 , 3) dan (1 , 3)

  (4 , 11) dan (4 , 11)

Dari n sampel, maka akan ada $\binom{n}{2} = \frac{n(n-1)}{2}$ pasangan, sehingga terdapat hubungan :

$$N_c + N_d + \text{Banyaknya "tie"} = \frac{n(n+1)}{2}$$

dimana N_c = Banyaknya pasangan konkordan

N_d = Banyaknya pasangan tak konkordan

KORELASI KENDALL

$$\tau = \frac{N_c - N_d}{\frac{n(n-1)}{2}}$$

Bila semua pasangan konkordan maka $\tau = 1$, sebaliknya jika semua pasangan tak konkordan $\tau = -1$, $\tau = 0$ jika $N_c = N_d$

Uji Hypotesa

A. Ho : X_i dab Y_i independen (positip)

H1 : ada kecenderungan X_i makin besar dikaitkan dengan Y_i makin besar pula.

B. Ho : X_i dan Y_i independen (negatip)

H1 : ada kecenderungan X_i yang makin besar dikaitkan dengan Y_i yang makin kecil

Secara sederhana harga uji dapat dilakukan dengan menghitungkan $\frac{n(n-1)}{2}$ sehingga

$$T = N_c - N_d$$

A. Ho ditolak jika $T > W_{1-\alpha/2}$ untuk korelasi positip

B. Ho ditolak jika $T < W_{\alpha/2}$ untuk korelasi negatip

Harga-harga kwartil diberikan pada tabel A_{12}

Contoh

Data agresifitas anak kembar (X_i, Y_i).

(X_i, Y_i)		banyaknya konkurdan	banyaknya tak konkordan
(60,64)		11	0
(70,65)		9	0
(71,77)	]– tie	4	4
(71,80)		4	4

(72,72)	5	1
(77,65) ⎤	5	0
(77,76) ⎦ – tie	4	1
(86,88)	2	2
(87,72)	3	0
(88,81)	2	0
(91,90) ⎤	0	0
(91,96) ⎦ – tie	0	0
	$N_c = 49$	$N_d = 12$

Uji Hypotesa

Ho : Ukuran agrefitas dari anak kembar adalah indepeden

H1 : Ada korelasi positip dan negatip antara ukuran agresifitas anak kembar.

$$T = N_c - N_d$$
$$= 49 - 12$$
$$= 37$$

untuk $\alpha = 0{,}05$, $n=12$, $W_{0,975} = 28$ (tabel A_{12})

Jelas $T = 37 > W_{0,975} = 28$, Jadi Ho ditolak

Kesimpulan ada korelasi positip antara ukuran agresifitas anak kembar

2.4.3. KORELASI PARSIAL KENDALL'S

Dalam multivariat variabel acak (X_1, X_2, ... , X_k) terdapat antara X_1 dan X_2, antara X_2 dan X_3 dan seterusnya.

Disini ukuran korelasi, estimasi total dari suatu variabel acak secara tidak langsung dipengaruhi oleh variabel lainnya, sebab dalam multivariat, variabel acak kedua bukan hanya korelasi dengan variabel acak pertama tetapi juga dengan variabel ketiga, yang secara tak langsung membentuk estimasi total korelasi antara dua variabel acak dengan variabel acak lainnya, estimasi ini disebut "korelasi parsial".

Korelasi Parsial Kendall's

$$\tau_{ij.k} = \frac{\tau_{ij} - \tau_{ik}\,\tau_{jk}}{\sqrt{(1-\tau_{ik}^{2})\,(1-\tau_{jk}^{2})}}$$

dimana $i \neq j \neq k$