KETEGARAN STATISTIK t UNTUK SATU POPULASI DAN DUA POPULASI

Tesis

Untuk memenuhi sebagian persyaratan
mencapai derajat sarjana S-2

Program Studi Matematika
Jurusan Matematika dan
Ilmu Pengetahuan Alam

Oleh :

Joko Riyono
6423/I-4/485/94

Kepada

PROGRAM PASCA SARJANA
UNIVERSITAS GADJAH MADA
1998

Tesis

KETEGARAN STATISTIK t UNTUK

SATU POPULASI DAN DUA POPULASI

dipersiapkan dan disusun oleh

Joko Riyomo

telah dipertahankan di depan Dewan Penguji

pada tanggal 24 Januari 1998

Susunan Dewan Penguji

Pembimbing Utama	Anggota Dewan Penguji Lain
Drs.H.Subanar,Ph.D. NIP.130 515 731	Prof.Dr.H.Zanzawi S. NIP.130 197 920
Pembimbing Pendamping I	Prof.Dr.Soeparna D. NIP.130 235 430
	Prof.Dr.Suryo G.M.Stats. NIP.130 367 311
Pembimbing Pendamping II	
	Dr.Sri Wahyuni,MS. NIP.131 284 231

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister

Tanggal 24 Januari 1998

Drs.Subanar,Ph.D.

Pengelola Program Studi : Matematika

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 21-Januari-1998

Joko Riyono

Tandatangan dan nama terang

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Alloh SWT atas segala rahmat dan karunianya,sehingga terselesaikannya tesis ini dengan judul "Ketegaran statistik t untuk satu populasi dan dua populasi".

Adapun tujuan penulisan tesis ini adalah untuk memenuhi sebagian persyaratan mencapai derajat sarjana S2 pada program studi Matematika jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Gajah Mada, Yogyakarta.

Pada kesempatan ini penulis juga ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Drs.H.Subanar,Ph.d. sebagai pembimbing dan Ketua Penyelenggara Program Studi S2 Matematika UGM,atas segala arahan,masukan ,dorongan serta pengorbanannya yang tulus ikhlas hingga terselesaikannya tesis ini.
2. Bapak dan Ibu Dosen Program Studi S2 Matematika UGM,yang telah banyak memberikan bekal ilmu selama masa pendidikan program studi S2 kepada penulis.
3. Prof.Dr.Ichlasul Amal,selaku Direktur Program Pasca Sarjana UGM.

4. Prof.Dr.R.Moedanton Moertedjo, selaku Rektor Universitas Trisakti yang telah memberikan kesempatan dan beasiswa kepada penulis untuk menempuh Program Studi S2 Matematika UGM.
5. Para staf administrasi Program Studi S2 UGM, khususnya pada jurusan Matematika yang banyak memberikan bantuan pada penulis selama menempuh pendidikan studi S2.
6. Ayah dan Ibu yang telah banyak memberikan dorongan dan bimbingan kepada penulis selama masa pendidikan program studi S2 hingga selesai.
7. Rekan-rekan sesama mahasiswa Program Studi S2, yang telah banyak memberikan dorongan dan masukan hingga terselesaikannya tesis ini.

Akhirnya penulis menyadari bahwa tesis ini masih jauh dari sempurna, namun kiranya dapat menjadi masukan yang baik dan bermanfaat bagi pembaca untuk studi lebih lanjut.

Yogyakarta, Desember 1997.

Penulis,

Joko Riyono.

DAFTAR ISI

	Halaman
JUDUL	i
PENGESAHAN	ii
PERNYATAAN	iii
PRAKATA	iv
DAFTAR ISI	vi
ARTI LAMBANG DAN SINGKATAN	vii
INTISARI	viii
ABSTRAK	ix
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Ruang Lingkup Masalah	4
1.3. Perumusan Masalah	5
1.4. Manfaat Penelitian	6
1.5. Tujuan Penelitian	6
1.6. Tinjauan Pustaka	7
1.7. Sistematika Penulisan	8
BAB II DESKRIPSI TEORITIS	9
2.1. Pengertian Dasar	9
2.2. Distribusi Student's t	13
2.3. Notasi "O" besar dan "o" kecil	15
BAB III KETEGARAN STATISTIK T PADA INFERENSI UNTUK MEAN SATU POPULASI.	18
3.1. Pendahuluan	18
3.2. Ekspansi Edgerworth dari Statistik T dan Hasil-hasil Yang terkait	21
3.3. Modofikasi dari Statistik T (Johnson 1978)	29
BAB IV KETEGARAN STATISTIK T PADA INFERENSI UNTUK MEAN DUA POPULASI.	36
4.1. Pendahuluan	36
4.2. Kemonotonan $\alpha(\lambda)$	39
4.3. Daerah Ketegaran	49
4.4. Prosedur Perhitungan	52
BAB V KESIMPULAN	57
RINGKASAN	59
DAFTAR PUSTAKA	96

ARTI LAMBANG DAN SINGKATAN

$\blacksquare$	= Akhir Bukti
$\mu = E(X)$	= Harga Harapan X
$Var(X)$	= Variansi X
$<$	= Lebih Kecil Dari
$\leq$	= Lebih Kecil atau sama
$>$	= Lebih besar dari
$\geq$	= Lebih besar atau sama
$\equiv$	= Identik dengan
$\in$	= Anggota himpunan
$\ni$	= Sedemikian hingga
$\forall$	= Untuk setiap
$\subset$	= Himpunan bagian
$\sim$	= Berdistribusi
$\xrightarrow{d}$	= Konvergen dalam distribusi
Φ	= Fungsi kumulatif normal standar
ϕ	= Fungsi probabilitas normal standar

KETEGARAN STATISTIK- t UNTUK SATU POPULASI DAN DUA POPULASI

Joko Riyono
Dibawah bimbingan
Drs.H.Subanar,Ph.d.

INTISARI

Uji ketegaran statistik t satu populasi dan dua populasi masing-masing untuk kasus adanya penyimpangan populasi normal dan homogenitas variansi telah dilakukan oleh Cicchitelli(1989) dan Posten, Yeh&Owen(1982). Dengan studi Monte Carlo,Cicchitelli(1989)menemukan bahwa kemiringan distribusi populasi mempengaruhi statistik t satu populasi lebih dari kurtosis , Johnson (1978) menyarankan suatu modifikasi statistik t untuk mereduksi pengaruh kemiringan ini.Hasil-hasil tersebut akan diperkuat dengan Ekspansi Edgeworth.

Posten,Yeh&Owen(1982) memberikan hasil untuk uji ketegaran statistik t dua populasi pada kasus adanya penyimpangan homogenitas variansi. Taraf ketegaran dikuantifikasi sebagai daerah ketegaran,hasilnya terbatas untuk uji dua sisi.Hasil tersebut diperluas kekasus uji satu sisi untuk masalah yang sama.

Katakunci:

Ekspansi Edgeworth,kurtosis,ketegaran,kemiringan, statistik t, daerah ketegaran.

ROBUSTNESS OF t-STATISTIC FOR THE ONE-SAMPLE AND TWO-SAMPLE

Joko Riyono
Under the Supervision
Drs.H.Subanar,Ph.d.

ABSTRACT

Cicchitelli(1989) conducted an extensive Monte Carlo study to investigate the robustness of the one sample t-statstic under nonnormal parent populations.He show that the skewness of the parent population affects the t-statistic more than the kurtosis.Johnson (1978) suggested a modification to t-statistic to reduce the effect of skweness.This paper shall addres and reinforce his emperical findings by means of Edgeworth expension of t-statistic.

Posten,Yeh, and Owen(1982),provided a quantification of the robustness level of the two-tailed two sample t-statistic, was studied under departures from the assumption of equal variance.The level of robustness was indicated by obtaining regions of robustness.In this paper,these results are extended to the one-tailed test for the same problem.

Key words:

Edgeworth expansion,kurtosis,robustness,skewness,t-statistic, regions of robustness.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Statistik t seringkali dipakai dalam membuat inferensi mean satu populasi dan inferensi untuk mean dua populasi Independen. Dalam tulisan ini akan difokuskan pada dua topik tersebut, khususnya yang menyangkut permasalahan ketegaran dari statistik t.

Dalam inferensi untuk mean satu populasi, diambil sampel Y_i , $i = 1,2, \ldots n$ dan dihitung $\overline{Y}$, S^2. Dengan asumsi populasi berdistribusi normal maka dapat diperoleh statistik T untuk inferensi mean satu populasi, yaitu :

$$T = \frac{\overline{Y} - \mu}{\left\{ S^2/n \right\}^{\frac{1}{2}}} \qquad \ldots\ldots\ldots \quad (1.1.1)$$

untuk n kecil T akan berdistribusi t dengan derajat bebas n - 1, sedang untuk n besar T akan mendekati distribusi normal standar.

Pada prakteknya, pemakai sering kesulitan menentukan berapa besar n hingga metode ini dapat digunakan atau bagaimana seandainya populasi tidak berdistribusi normal, apakah metode ini masih dapat juga digunakan.

Sedang pada inferensi untuk mean dua populasi independen, diambil sampel X_{1j} , $j = 1,2, \ldots, n_1$ dan X_{2l} , $l = 1,2,\ldots,n_2$ kemudian hitung $\overline{X}_1, \overline{X}_2, S^2$. Dengan asumsi distribusi kedua populasi normal dan variansi keduanya sama maka dapat diperoleh statistik t untuk inferensi mean dua populasi independen,yaitu :

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_{x_1} - \mu_{x_2}\right)}{\left[S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{\frac{1}{2}}} \qquad \ldots\ldots(1.1.2)$$

Sama halnya seperti pada inferensi untuk mean satu populasi, bagaimana seandainya assumsi kehomogenan variansi dilanggar apakah metode ini masih dapat digunakan. Masalah seperti ini dengan σ_1^2 dan σ_2^2 tidak diketahui ($\sigma_1^2 \neq \sigma_2^2$)disebut sebagai masalah Behren-Fisher. Meskipun untuk masalah tersebut telah ada cara pemecahannya, seperti yang diberikan oleh Neyman-Bartlett, Scheffe' ataupun Hsu, namun ketiga cara diatas masih memiliki kekurangan, terutama untuk $n_1 \neq n_2$ yaitu adanya sampel yang harus diabaikan, hasil yang masih tergantung pada urutan pengambilan sampel atau tidak adanya jaminan uji mempunyai tingkat kepercayaan pada α yang kita inginkan. Oleh karena itu pentinglah bagi pemakai untuk melihat sampel,sejauh mana perbedaan σ_1^2 dan σ_2^2 yang masih bisa ditolelir sehingga metode diatas masih dapat digunakan.

Permasalahan-permasalahan seperti diatas dimana beberapa asumsi dilanggar, didalam statistik dikenal sebagai permasalahan ketegaran (robustness) dari suatu statistik (estimator) atau suatu prosedur statistik. Menurut Hubber (1981) ketegaran adalah ketidaksensitifan terhadap sedikit penyimpangan dari asumsi. Suatu statistik atau prosedur statistik dikatakan tegar apabila ia dapat bekerja dengan baik, mempunyai bias dan variansi rendah dalam segala macam distribusi.

Studi tentang ketegaran statistik t telah banyak dilakukan, antara lain oleh Cicchitelli(1989) yang menunjuk suatu studi Monte Carlo untuk meneliti pengaruh uji inferensi mean satu populasi, dengan statistik T diatas pada populasi nonnormal. Dia memperlihatkan bahwa kemiringan yang ada dalam populasi mengakibatkan kemiringan yang ada dalam distribusi T. Johnson (1978) membuat modifikasi untuk mengoreksi kemiringaan statistik T menggunakan ekspansi Cornish-Fisher. Sedangkan Posten, Yeh dan Owen (1982) meneliti ketegaran uji t dua populasi versi dua sisi untuk kasus penyimpangan pada homogenitas variansi kedua populasi. Dalam tulisan ini nantinya akan dibahas hasil-hasil yang telah didapat Cicchitelli(1989) di atas menggunakan ekspansi Edgeworth, kemudian juga melihat penampilan modifikasi statistik T dari Johnson (1978) untuk dibandingkan dengan penampilan statistik T itu sendiri. Pada uji t dua populasi, kita akan mengembangkan hasil-hasil yang didapat oleh Posten, Yeh dan Owen (1982) untuk versi uji t satu sisi.

Tulisan ini diharapkan akan dapat memberi sumbangan yang berharga dibidang statistika inferensial. Selain itu diharapkan pembaca dapat menggunakan atau mengambil manfaat dari apa yang bisa diserap dalam tulisan ini.

1.2. Ruang Lingkup Masalah

Berdasarkan latar belakang masalah di atas, maka ruang lingkup masalah dalam penelitian ini adalah :

1. Melihat pengaruh kemiringan yang dibawa populasi pada statistik T (1.1.1) dengan ekspansi Edgeworth.
2. Mempelajari modifikasi statistik T (1.1.1.)yang dibuat Johnson untuk mengoreksi kemiringan yang dibawa oleh populasi terhadap statistik T (1.1.1).
3. Mencari batas-batas keheterogenan variansi yang bisa ditolelir sehingga metode dengan menggunakan statistik t (1.1.2) masih dapat berlaku.

1.3. Perumusan Masalah

Berdasarkan latarbelakang dan ruang lingkup masalah, masalah dalam tulisan ini dapat dirumuskan sebagai berikut :

1. Mempelajari pengaruh kemiringan populasi terhadap statistik T(1.1.1).
2. Mempelajari cara memperoleh modifikasi statistik T, yang dapat mengurangi pengaruh kemiringan yang dibawa oleh populasi terhadap statistik T (1.1.1).

3. Melihat nilai keheterogenan variansi yang masih dapat ditolelir sehingga metode dengan menggunakan statistik t (1.1.2) versi satu sisi masih berlaku.

1.4.Manfaat Penelitian

Dengan penelitian ini diharapkan dapat diambil beberapa manfaat sebagai berikut:

1. Dari segi subyektif penulis, maka tulisan ini semata-mata untuk memperkuat struktur kognitif penulis mengenai inferensi statistik.
2. Dapat memberikan sumbangan pemikiran dalam rangka penambahan wawasan mengenai inferensi statistik bagii para pemerhati masalah statistik baik yang ada didalam perguruan tinggi maupun yang ada diluar perguruan tinggi.
3. Lebih jauh lagi, informasi yang diberikan dalam tulisan ini nantinya menuntut diadakannya penelitian lebih lanjut dengan melibatkan berbagai pendekatan untuk memperoleh wawasan yang lebih luas tentang teori uji hipotesis khususnya yang menyangkut inferensi untuk mean suatu populasi.

1.5. Tujuan Penelitian

Berdasarkan rumusan masalah yang dikemukakan di atas, tujuan penulisan ini adalah untuk mempelajari :

1. Konsep teori uji hipotesis pada inferensi untuk mean suatu populasi.
2. Penemuan-penemuan Cicchitelli tentang ketegaran statistik T (1.1.1) khususnya pengaruh kemiringan (ketaksimetrisan) dari distribusi populasi asal terhadap statistik T (1.1.1).
3. Penampilan modifikasi statistik T (1.1.1) dari Johnson.
4. Batas-batas nilai keheterogenan variansi yang masih menjamin ketegaran statistik t (1.1.2) pada uji satu sisi.

1.6. Tinjauan Pustaka

Kekonservatifan uji T dipelajari oleh Gross (1976) dan Tukey & Mc. Laughlin (1963). Benjamini (1983) menunjukkan kekonservatifan uji T untuk distribusi populasi long tailed menggunakan argumen geometrik, selanjutnya pendekatan geometris diambil oleh Hotelling (1961) dan Efron (1969). Yuen & Murthy (1974) mempelajari masalah spesifik pengamatan gambar dari distribusi Student-t,mereka menabelkan titik-titik tertentu dari distribusi.

Posten, Yeh dan Owen (1982) telah memberikan suatu kuantifikasi dari tingkat ketegaran pada uji t dua sampel dibawah asumsi variansi yang sama. Tingkat ketegaran diidentifikasi dengan mendapatkan daerah ketegaran dan hasil-hasil hanya dibatasi untuk uji dua sisi.

1.7. Sistematika Penulisan

Karena tulisan ini merupakan kajian terhadap berbagai kepustakaan, maka bahan yang digunakan berupa buku, Jurnal dan lain-lain. Hasil yang diperoleh akan disusun dalam sistematika berikut :

BAB I : Pendahuluan

BAB II : Deskripsi Teoritis

BAB III : Ketegaran Statistik T pada Inferensi Untuk Mean Satu Populasi.

BAB IV : Ketegaran Statistik t Pada Inferensi Untuk Mean Dua populasi Independen.

BAB V : Kesimpulan dan Saran.

BAB II

DESKRIPSI TEORITIS

2.1. Pengertian Dasar

Uji Hipotesis adalah suatu proses percobaan untuk memutuskan kebenaran atau kesalahan pendugaan yang didasarkan pada pengamatan sampel .

Andaikan diberikan X_1, X_2, . . . , X_n sampel random dari suatu populasi berdistribusi F_θ , $\theta \in \Theta$, dengan bentuk F_θ diketahui kecuali untuk parameter θ. Dengan pengasumsian seperti ini, berikut akan diberikan beberapa definisi serta teorema yang berkaitan dengan konsep uji hipotesis.

Definisi 2.1.1. (V.K Rohatgi)

Hipotesa parametrik adalah pernyataan tentang suatu parameter θ yang tidak diketahui. H_0 ,$\theta \in \Theta_0 \subset \Theta$ disebut sebagai hipotesa null sedang pernyataan H_1,$\theta \in \Theta_1 = \Theta - \Theta_0$,disebut sebagai hipotesa alternatif.

Definisi 2.1.2. (V.K. Rohatgi)

Jika $\Theta_0(\Theta_1)$ hanya berisi satu anggota, $\Theta_0(\Theta_1)$dikatakan sederhana, sebaliknya jika lebih dari satu anggota dikatakan komposit .Jadi jika suatu hipotesis sederhana, maka distribusi probabilitas X menjadi tertentu di bawah hipotesis.

Definisi 2.1.3. (Bain E.)

Daerah kritis untuk suatu uji hipotesis adalah subset dari ruang sampel yang bersesuaian dengan penolakan hipotesa null.

Definisi 2.1.4. (Bain E)

Jika H_0 merupakan hipotesa null sederhana maka probabilitas penolakan kebenaran H_0 ,yaitu α = P(TI), disebut sebagai tingkat kepercayaan dari uji. Jika H_0 merupakan komposit, maka ukuran dari uji (atau ukuran daerah kritis) adalah nilai maksimum probabilitas dari penolakan H_0 jika H_0 benar (nilai maksimum atas semua nilai parameter dibawah H_0).

Definisi 2.1.5. (Bain E.)

Fungsi Kuasa $\pi(\theta)$ uji H_0 adalah probabilitas penolakan H_0 jika nilai kebenaran parameter adalah θ.

Jadi untuk hipotesis sederhana H_0 ,$\theta = \theta_0$ versus H_1 ,$\theta = \theta_1$, diperoleh $\pi(\theta_0) = P(TI) = \alpha$ dan $\pi(\theta_1) = 1 - P(TII) = 1 - \beta$. Untuk hipotesis komposit, dikatakan H_0 , $\theta \in \Theta_0$ versus H_1 , $\theta \in \Theta - \Theta_0$,ukuran uji adalah $\alpha = \max_{\theta\in\Theta_0} \pi(\theta)$.

Definisi 2.1.6. (E.J.Dudewich)

Suatu fungsi $f_x(x)$ dikatakan simetri di sekitar μ jika $f_x(\mu + x) = f_x(\mu - x)$ untuk semua x.

Theorema 2.1.1.

Diberikan X_1, X_2, ... ,X_n sampel populasi dengan fungsi distribusi kumulatif F dan untuk $b = \frac{n-1}{n}S^2$ dengan S^2 adalah variansi sampel yaitu $S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$

maka (i) $E(b) = \frac{n-1}{n}\mu_2$.

(ii) $Var(b) = \left(n - 2 + 1/n\right)\frac{\mu_4}{n^2} + (n-1)(3-n)\frac{\mu_2^2}{n^3}$

dengan μ_i momen pusat ke i dari X.

Bukti : V.K Rohatgi hal.304 ■

Akibat 2.1.1.

Selalu berlaku :

$E(S^2) = \mu_2$ dan $Var(S^2) = \frac{\mu_4}{4} + \frac{3-n}{n(n-1)}\mu_2^2$

Bukti :Karena

$$E(b) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\mu_2$$

sehingga $E(S^2) = \mu_2$

$$Var(b) = \left(n - 2 + \frac{1}{2}\right)\frac{\mu_4}{n^2} + (n-1)(3-n)\frac{\mu_2^2}{n^3} \quad \text{sehingga}$$

$$Var\left(\frac{n-1}{n}S^2\right) = \left(\frac{n^2 - 2n + 1}{n}\right)\frac{\mu_4}{n^2} + (n-1)(3-n)\frac{\mu_2^2}{n^3}$$

$$\left(\frac{n-1}{n}\right)^2 Var(S^2) = \frac{(n-1)^2}{n}\frac{\mu_4}{n^2} + (n-1)(3-n)\frac{\mu_2^2}{n^3}$$

$$var(S^2) = \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\mu_2^2 \quad ■$$

2.2. Distribusi Student t

Seperti diketahui bahwa bentuk fungsi distribusi student t dengan derajat bebas v adalah :

$$f(t,v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}$$

Berikut ini akan kami berikan beberapa theorema yang berkaitan dengan distribusi student t di atas dan akan sering ditemui pada bab-bab berikutnya.

Theorema 2.2.1.

Jika $X_1, X_2, \ldots, X_n$ notasi sampel random $N(\mu, \sigma^2)$ maka :

1. $\bar{X}$ dan suku-suku $X_i - \bar{X}$, $i = 1, 2, \ldots, n$ independen.
2. $\bar{X}$ dan S^2 adalah independen.
3. $(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$

Bukti : Bain. E. hal 272 ■

Theorema 2.2.2.

Jika $Z \sim N(0,1)$ dan $V \sim \chi^2(v)$, dan jika Z dan V independen maka distribusi $T = \dfrac{Z}{\sqrt{V / v}}$ adalah distribusi student t dengan derajat bebas v.

Bukti : Bain E. hal. 274 ■

Theorema 2.2.3.

Jika $X_1, X_2, \ldots, X_n$ sampel random dari $N(\mu, \sigma^2)$ maka :

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$$

Bukti :

Karena

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$$

Dari theorema 2.2.1. didapat :

$V = (n - 1)\, S^2/\sigma^2 \sim \chi^2_{(n-1)}$ menggunakan Theorema 2.2.2 maka:

$$\frac{Z}{\sqrt{V/n-1}} = \frac{\sqrt{n}\left(\bar{X}-\mu\right)/\sigma}{\sqrt{(n-1)\,S^2/\sigma^2(n-1)}} = \frac{\bar{X}-\mu}{\sqrt{S^2/n}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1) \blacksquare$$

2.3. Notasi "O" Besar dan "o" Kecil.

Diberikan f dan g dua fungsi dan asumsikan bahwa $g(x) > 0$ untuk x yang cukup besar. Kita katakan bahwa $f(x)$ adalah berorder lebih besar dari $g(x)$ pada $x \to \infty$ dan kita tulis $f(x) = O(g(x))$ untuk $x \to \infty$ jika $\exists\ x_0$ dan konstanta $C > 0 \ni |f(x)| < c\ g(x)\ \forall x \geq X_0$. Jadi $f(x) = O[g(x)]$ berarti bahwa $|f(x)|/g(x)$ terbatas untuk x besar. Kita tulis $f(x) = O(1)$ untuk menyatakan f terbatas untuk x besar.

Dan mudah untuk dilihat bahwa :

(a). Jika $f_1(x) = O[g_1(x)]$, $f_2(x) = O[g_2(x)]$ maka $f_1(x) + f_2(x) = O[g_1(x) + g_2(x)]$.

(b). Jika $\alpha > 0$ adalah konstanta, $f(x) = O[\alpha g(x)]$ maka $f(x) = O[g(x)]$

(c). Jika $f_1(x) = O[g_1(x)]$ dan $f_2(x) = O[g_2(x)]$ maka $f_1(x)f_2(x) = O[g_1(x)g_2(x)]$.

Diberikan f dan g, keduanya terdefinisi dan positif untuk x besar. Kita katakan f(x) adalah berorder lebih kecil daripada g(x) untuk x besar, dan ditulis :

$f(x) = o[g(x)]$ untuk $x \to \infty$ jika $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

Kita tulis $f(x) = o(1)$ untuk menyatakan $f(x) \to o$ untuk $x \to \infty$.

Mudah dilihat bahwa :

(a). $f(x) = o[g(x)] \Rightarrow f(x) = O[g(x)]$

(b). $f(x) = o[g(x)]$, $f_2(x) = o[g_2(x)] \Rightarrow f_1(x)f_2(x) = o[g_1(x)g_2(x)]$

Bukti :

(a) Diketahui $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$

Artinya untuk sebarang $\varepsilon > 0$ $\exists$ $K = K(\varepsilon) > 0$ $\ni \forall$ $x > K$ berakibat :

$\left|\frac{f(x)}{g(x)} - o\right| < \varepsilon$ sehingga $\left|\frac{f(x)}{g(x)}\right| < \varepsilon$ karena f.g positif

maka $\left|\frac{f(x)}{g(x)}\right| = \frac{|f(x)|}{g(x)} < \varepsilon$ artinya $|f(x)| < \varepsilon\, g(x)$, $\forall x > K$

yang tidak lain menyatakan f(x) = O[g(x)] ■

contoh : f(x) adalah polinomial dalam x order n dan g(x) adalah polinomial dalam x order m dengan n < m.

(b).Diketahui $\lim\limits_{x\to\infty} \frac{f_1(x)}{g(x)} = 0$ dan $\lim\limits_{x\to\infty} \frac{f_2(x)}{g_2(x)} = 0$

maka : $\lim\limits_{x\to\infty} \frac{f_1(x)\,.\,f_2(x)}{g_1(x)\,.\,g_2(x)} = \lim\limits_{x\to\infty} \frac{f_1(x)}{g_1(x)} \,.\, \lim\limits_{x\to\infty} \frac{f_2(x)}{g_2(x)} = 0$ ■

BAB III

KETEGARAN STATISTIK T PADA INFERENSI UNTUK MEAN SATU POPULASI

3.1. Pendahuluan

Andaikan diberikan sampel random $Y_1, Y_2, \ldots, Y_n$ berukuran n dari suatu populasi G dengan mean μ yang tidak diketahui. Satu uji hipotesis yang penting H_0 , $\mu = \mu_0$,melawan satu dari tiga kemungkinan alternatif yaitu :

(a). $H_1 : \mu \neq \mu_0$

(b). $H_1 : \mu > \mu_0$ atau

(c). $H_1 : \mu < \mu_0$

Pertama diberikan : $T = \dfrac{\overline{Y} - \mu}{(S^2 / n)^{\frac{1}{2}}}$ . . . . . (3.1.1)

dan $TC = \dfrac{\overline{Y} - \mu_0}{(S^2 / n)^{\frac{1}{2}}}$ . . . . (3.1.2)

dengan $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$ dan $S^2 = \dfrac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n - 1}$, berturut-turut adalah mean sampel dan variansi sampel. Pemakai biasa menunjuk uji mengenai μ di atas menggunakan statistik TC, yang disebut sebagai statistik T satu sampel. Jika diasumsikan bahwa :

A(i) Y_i berdistribusi independen

A(ii) $G \equiv \Phi$

dimana Φ fungsi distribusi kumulatif normal standar, maka suatu hasil dalam teori statistik yang cukup baik dibuat oleh Student(1908) bahwa $T \sim t(n-1)$ adalah distribusi student t dengan derajat kebebasan (n-1). Dengan asumsi di atas, dipunyai uji pada level α :

TT: Untuk $H_0:\mu=\mu_0$ vs $H_1:\mu \neq \mu_0$, tolak H_0 jika $|TC| > t_{\alpha/2}$ ....(3.1.3)

TU: Untuk $H_0:\mu=\mu_0$ vs $H_1:\mu > \mu_0$, tolak H_0 jika $TC > t_\alpha$....... ....(3.1.4)

TL: Untuk $H_0:\mu=\mu_0$ vs $H_1:\mu < \mu_0$, tolak H_0 jika $TC < -t_\alpha$ ....(3.1.5)

dengan t_α : notasi bagian atas 100α titik persentil dari distribusi t dengan derajat bebas (n-1).

Sifat-sifat distribusi T yang disajikan pada (3.1.1) jelas tergantung pada asumsi yang diberikan pada

A(i) dan A(ii). Akan tetapi pemakai akan bertanya apakah asumsi tersebut dijumpai dan bagaimanakah pengaruh dari pelanggaran asumsi tersebut pada distribusi T.

Cicchitelli(1989) menunjukkan perluasan studi Monte Carlo untuk menyelidiki ketegaran dari statistik T satu sampel dengan distribusi populasi nonnormal. Dia menggunakan distribusi λ yang telah dikenalkan oleh Ramberg(1979) sebagai distribusi populasi. Kita akan mencoba mengkaji kembali hasil-hasil Cicchitelli tersebut dengan menggunakan ekspansi Edgeworth dari statistik T.

Untuk itu, akan kita perlukan beberapa definisi yang berkaitan dengan ekspansi Edgerworth seperti yang kita tulis berikut ini :

Definisi 3.1.1.

Untuk sebarang variabel random Y dengan $E|Y|^4 < \infty$ didefinisikan $S_k = \dfrac{\mu_3}{(\mu_2)^{3/2}}$ dan $K_u = \dfrac{\mu_4}{\mu_2^2}$. S_k adalah ukuran kemiringan dan K_u adalah ukuran kurtosis dari distribusi. $\mu_i, i = 2,3,4$ adalah Momen Pusat kedua, tiga, empat dari distribusi.

Definisi 3.1.2.

Untuk sebarang distribusi G, jika $S_k > (<) 0$ kita tunjuk G mempunyai kemiringan positif (negatif) dan jika $K_u > (<) 3$ kita tunjuk G mempunyai long (short) tailed.

Definisi 3.1.3.

$P_T(\alpha) = P_r\ (|TC| > t_{\alpha/2})$

$P_U(\alpha) = P_r\ (TC > t_\alpha)$

$P_L(\alpha) = P_r\ (TC < -t_\alpha)$

Dengan asumsi-asumsi A(i) dan A(ii), sesuai hasil student(1908) bahwa dibawah H_0, $P_T(\alpha) = P_U(\alpha) = P_L(\alpha) = \alpha$.

3.2. Ekspansi Edgeworth dari Statistik T dan Hasil-hasil yang terkait

Theorema 3.2.1 (Hall 1987)

Asumsikan $k \geq 1$, $E|Y|^{k+2} < \infty$ dan distribusi G nonsingular, maka

$F_{T_0}(y) = P_r(T_0 \leq y)$

$$= \Phi(y) + \sum_{i=1}^{k} n^{-i/2} P_i(y)\phi(y) + o(n^{-k/2}) \quad . . . (3.2.1)$$

Konvergen uniform dalam y, Dimana P_i adalah polinomial berderajat (3i-1) yang tampak dalam ekspansi Edgeworth

dari $$T_0 = n^{1/2}(\bar{Y} - \mu)\left(n^{-1}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2\right)^{-1/2}$$

$\Phi(y)$ dan $\phi(y)$ masing-masing adalah fungsi distribusi kumulatif normal standar dan fungsi densitas normal standar. Koefesien P_i adalah fungsi-fungsi dari $\mu, \sigma^2, \mu_3, \ldots, \mu_{i+2}$ dimana μ_j mewakili momen central ke-j dari G.

Untuk P_1 dan P_2 didapat :

$P_1(y) = \frac{1}{6} S_k(2y^2 + 1)$

dan

$$P_2(y) = -y\left\{\frac{1}{18} S_k^2(y^2 + 3)(y^2 - 1)\right.$$

$$\left. - \frac{1}{12}(K_u - 3)(y^2 - 3) + \frac{1}{4}(y^2 + 3)\right\}$$

dengan S_k dan K_u masing-masing ukuran kemiringan dan kurtosis dari G.

Terlihat bahwa :

$$T = \left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_0 \qquad \ldots\ldots (3.2.2)$$

Dari (3.2.1) untuk k = 2 kita peroleh :

$$P_r(T_0 \leq 0) = \frac{1}{2} + \frac{n^{-\frac{1}{2}} S_k}{6(2\pi)^{\frac{1}{2}}} + o(n^{-1}) \qquad \ldots\ldots (3.2.3)$$

Dengan asumsi kefinitan momen ke-4 dan sifat nonsingularitas dari G kita akan mencari $E(T_0)$ dan $E(T)$ sebagai berikut :

$$\begin{aligned} E(T_0) &= \int_{-\infty}^{\infty} y \; dF_{T_0}(y) \\ &= \frac{n^{\frac{1}{2}}}{6} S_k \int_{-\infty}^{\infty} y \; d\left[(2y^2 + 1)\phi(y)\right] \\ &\quad - n^{-1} \int_{-\infty}^{\infty} y d\left[\frac{1}{18} S_k^2 y(y^2 + 3)(y^2 - 1)\phi(y)\right] \\ &\quad + n^{-1} \int_{-\infty}^{\infty} y \; d\left[\left\{\frac{1}{12}(K_u - 3)y(y^2 - 3) - \frac{1}{4} y(y^2 + 3)\right\}\phi(y)\right] \\ &\quad + o(n^{-1}) \end{aligned}$$

dengan mengingat $d\phi(y) = -y\phi(y)dy$, maka :

$$E(T_0) = \frac{-n^{-\frac{1}{2}}}{2} S_k + o(n^{-1}) \qquad \ldots\ldots (3.2.4)$$

Kita dapat lihat dari (3.2.1)-(3.2.4) bahwa selama ukuran sampel naik, apapun harga kemiringan populasi, S_k maka T_0 dan juga T menjadi lebih simetri di sekitar nol.

Ini bukan masalah yang luar biasa, karena telah diketahui bahwa $T \xrightarrow{d} Z$ dimana Z adalah variabel random normal standar. Jadi untuk ukuran sampel yang besar atau kemiringan populasi yang kecil T_0 dan juga T mendekati simetri disekitar nol. Dengan cara yang sama untuk mendapatkan $E(T_0)$ kita dapat juga memperoleh :

$$E\left(T_0^2\right) = \int_{-\infty}^{\infty} y^2 dF_{T_0}(y)$$

$$= 1 + \frac{2}{n} S_k^2 + \frac{3}{n} + o(n^{-1}) \quad . \; . \; . \; . \; (3.2.5)$$

Sehingga :

$$Var(T_0) = E(T_0^2) - \left\{E(T_0^2)\right\}^2$$

$$= 1 + \frac{7}{4n} S_k^2 + \frac{3}{n} + o(n^{-1}) \quad . \; . \; . \; . \; (3.2.6)$$

Selanjutnya :

$$P_r(T \leq 0) = P_r\left(\left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_0 \leq 0\right)$$

$$= P_r(T_0 \leq 0)$$

$$= \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_k + o(n^{-1}) \quad . \; . \; . \; . \; (3.2.7)$$

$$E(T) = E\left\{\left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_0\right\}$$

$$= \left(\frac{n-1}{n}\right)^{\frac{1}{2}} E(T_0)$$

$$= (1 - \frac{1}{n})^{\frac{1}{2}} E(T_0)$$

$$= \left(1 - \frac{1}{2n} - \frac{1}{4n^2 2!} - \frac{3}{8n^3 3!} + \ldots\right)\left\{-\frac{n^{-\frac{1}{2}}}{2} S_k + o(n^{-1})\right\}$$

$$= \frac{-n^{-\frac{1}{2}}}{2} S_k + o(n^{-1}) \qquad \ldots\ldots (3.2.8)$$

$$E(T^2) = E\left(\frac{n-1}{n} T_0^2\right) = \left(\frac{n-1}{n}\right) E(T_0^2)$$

$$= \left(1 - \frac{1}{n}\right)\left(1 + \frac{2}{n} S_k^2 + \frac{3}{n} + o(n^{-1})\right)$$

$$= 1 + \frac{2}{n} S_k^2 + \frac{3}{n} - \frac{1}{n} + o(n^{-1})$$

$$= 1 + \frac{2}{n} S_k^2 + \frac{2}{n} + o(n^{-1})$$

$$Var(T) = E(T^2) - (E(T))^2$$

$$= 1 + \frac{2}{n} + \frac{7}{4n} S_k^2 + o(n^{-1}) \qquad \ldots\ldots (3.2.9)$$

Sekarang perhatikan ekspansi deret taylor dari $F_{T_0}(y)$ disekitar y = 0 didapat dari (3.2.1) yaitu :

$$F_{T_0}(y) = \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_k + \frac{y}{(2\pi)^{\frac{1}{2}}}\left(1 + \frac{S_k^2}{6n} - \frac{K_u}{4n}\right)$$

$$+ \frac{n^{-\frac{1}{2}}}{4(2\pi)^{\frac{1}{2}}} y^2 S_k + o(n^{-1}) \qquad . \; . \; . \; . \; (3.2.10)$$

dan menggunakan (3.2.2) kita dapatkan :

$$\begin{aligned} F_T(y) &= P_r(T \leq y) \\ &= P_r\left(\left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_0 \leq y\right) \\ &= P_r\left(T_0 \leq \left(\frac{n-1}{n}\right)^{\frac{1}{2}} y\right) \\ &= F_{T_0}\left(\left(\frac{n-1}{n}\right)^{\frac{1}{2}} y\right) \end{aligned}$$

$$= \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K + \frac{\left(\frac{n}{n-1}\right)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} y\left(1 + \frac{S_K{}^2}{6n} - \frac{K_u}{4n}\right)$$

$$+ \frac{n^{-\frac{1}{2}}}{4(2\pi)^{\frac{1}{2}}}\left(\frac{n}{n-1}\right) y^2 S_K + o(n^{-1})$$

$$= \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K + \frac{y}{(2\pi)^{\frac{1}{2}}}\left(1 + \frac{S_K{}^2}{6n} - \frac{K_u}{4n}\right)\left(1 + \frac{1}{2n} + \frac{3}{4n^2 2!} + \ldots\right)$$

$$+ \frac{y^2 S_K}{4(2\pi)^{\frac{1}{2}}} n^{-\frac{1}{2}}\left(1 + \frac{1}{n} + \frac{1}{n^2 2!} + \frac{1}{n^3 3!} + \ldots\right) + o(n^{-1})$$

$$F_T(y) = \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K + \frac{y}{(2\pi)^{\frac{1}{2}}}\left(1 + \frac{1}{2n} + \frac{S_K^{\,2}}{6n} - \frac{K_u}{4n}\right)$$

$$+ \frac{y^2 S_K}{4(2\pi)^{\frac{1}{2}}} n^{-\frac{1}{2}} + o(n^{-1})$$

Untuk menemukan median dari distribusi T pecahkan persamaan $F_T(y) = \frac{1}{2}$. Suatu pendekatan solusi persamaan ini adalah $y = -\frac{n^{-\frac{1}{2}}}{6} S_K$ sebab :

$$F_T\left(\frac{-n^{-\frac{1}{2}}}{6} S_K\right) = \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K$$

$$- \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K\left(1 + \frac{1}{2n} + \frac{S_K^{\,2}}{6n} - \frac{K_u}{4n}\right)$$

$$+ \frac{n^{-\frac{3}{2}}}{144(2\pi)^{\frac{1}{2}}} S_K^{\,3} + o(n^{-1})$$

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

dengan pengabaian suku $o(n^{-1})$ maka :

$$\text{Median dari } T = \frac{-n^{-\frac{1}{2}}}{6} S_K \qquad \ldots\ldots (3.2.11)$$

$$\text{Mean dari } T = \frac{-n^{-\frac{1}{2}}}{2} S_K \qquad \ldots . \quad (3.2.12)$$

Sehingga terlihat bahwa jika distribusi populasi miring kekanan yaitu $S_k > 0$ maka $\frac{-n^{-\frac{1}{2}}}{2} S_K < \frac{-n^{-\frac{1}{2}}}{6} S_K < 0$.

Jadi distribusi T miring kekiri. Sebaliknya adalah benar untuk $S_k < 0$. Hasil tersebut sesuai dengan studi awal dan penemuan Cicchitelli(1989). Sehingga kita dapat putuskan bahwa, jika distribusi populasi sangat miring dan ukuran sampel kecil maka kemiringan distribusi T tidak akan didekati dengan distribusi student-t, kecuali telah dibuat suatu koreksi kemiringan seperti yang dilakukan oleh Johnson(1978).

Dengan pengabaian suku $o(n^{-1})$ maka mean dari T < median dari T < 0 dan variansi dari T = $1 + \frac{2}{n} + \frac{7}{4n} S_k^2 (\geq 1 + \frac{2}{n}$ variansi dari suatu variabel random yang mempunyai distribusi student-t dengan derajat bebas n - 1). Jadi untuk kemiringan populasi G positif maka $P_u(\alpha) < \alpha$ dan $P_L(\alpha) > \alpha$ seperti yang diamati oleh Cicchitelli(1989).

3.3. Modifikasi dari Statistik T (Johnson 1978)

Andaikan diberikan $Y_1, Y_2, \ldots, Y_n$ adalah sampel random dari suatu distribusi G dengan mean μ, variansi σ^2 dan $\mu_3, \mu_4, \ldots$ masing-masing adalah momen central ketiga, keempat, ... dari G. Untuk sebarang variabel random Y dengan distribusi G, bentuk ekspansi Cornish-Fisher diberikan sebagai berikut :

$$CF(Y) = \mu + \sigma\xi + \frac{\mu_3}{6\sigma^2}(\xi^2 - 1) + \ldots\ldots\ldots \qquad (3.3.1)$$

dengan ξ adalah variabel normal standar.

Karena $\quad \mathrm{Var}(\overline{Y}) = \dfrac{\sigma^2}{n}$ dan

$$\mu_3(\overline{Y}) = E(\overline{Y} - \mu)^3 = \frac{1}{n^3} E\left\{\sum_{i=1}^{n}(Y_i - \mu)\right\}^3$$

$$= \frac{1}{n^3}\sum_{i=1}^{n} E(Y_i - \mu)^3 = \frac{\mu_3}{n^2}$$

Sehingga :

$$CF(\overline{Y}) = \mu + \frac{\sigma}{n^{\frac{1}{2}}}\xi + \frac{\mu_3}{6n\sigma^2}(\xi^2 - 1) + O\left(n^{-\frac{3}{2}}\right) \qquad \ldots (3.2.2)$$

Dari ekspansi di atas dapat dicatat bahwa kemiringan populasi μ_3 adalah koefisien dari suku (ξ^2-1) dan tampak dalam suku-suku lain tetapi dengan order yang lebih kecil. Kunci dalam mendapatkan modifikasi variabel T dalam pendekatan Johnson adalah mengeliminasi suku yang melibatkan μ_3 dalam variabel T pembangun yang diberikan pada bagian bawah berikut :

Diberikan variabel T pembangun

$$TJ = \frac{(\bar{Y} - \mu) + \lambda + \gamma\left\{(\bar{Y} - \mu)^2 - \frac{\sigma^2}{n}\right\}}{\left\{\frac{S^2}{n}\right\}^{\frac{1}{2}}} \qquad \ldots (3.2.3)$$

λ dalam TJ dipilih sehingga suku konstan dalam ekspansi Cornish-Fisher dari TJ berjumlah nol sehingga bias order yang lebih rendah tereleminasi γ dipilih sehingga koefesien dari suku ξ^2 dalam ekspansi Cornish-Fisher dari TJ adalah nol (dengan demikian mengeliminasi pangaruh kemiringan order yang lebih rendah). Dapat ditunjukkan bahwa $\gamma = \frac{\mu_3}{3\sigma^4}$ dan $\lambda = \frac{\mu_3}{2n\sigma^2}$ sebagai berikut :

Karena : $E(s^2) = \sigma^2$ dan $\quad var(S^2) = \frac{\mu_4}{n} + \frac{3 - n}{n(n-1)}\sigma^4$

$$var(S^2) = \frac{\mu_4}{n} + \frac{\sigma^4}{n}\left(-1 + \frac{2}{n} + \frac{2}{n^2} + \ldots\right) \approx \frac{\mu_4}{n} - \frac{\sigma^4}{n}$$

sehingga ekspansi Cornish-Fisher dari $\bar{Y}$ dan S^2, suku-suku lebih tinggi diabaikan adalah :

$$CF(\overline{Y}) = \mu + \frac{\sigma}{\sqrt{n}}\xi + \frac{\mu_3}{6n\sigma^2}(\xi^2 - 1)$$

$$CF(S^2) = \sigma^2 + \left[\frac{\mu_4 - \sigma^4}{n}\right]^{\frac{1}{2}}\eta$$

$$= \sigma^2\left\{1 + \left[\frac{\mu_4 - \sigma^4}{n\sigma^4}\right]^{\frac{1}{2}}\eta\right\}$$

dengan ξ dan η adalah variabel random normal standar. Gantikan nilai $\overline{Y}$ dan S^2 dalam (3.2.3) dengan ekspansinya, maka dengan pengabaian suku $O(n^{-1})$, ekspansi Cornish-Fisher dari TJ adalah :

$$CF(TJ) = \xi + \left(\frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{3\sqrt{n}\sigma^3}\right)\xi^2$$

$$+ \frac{\lambda\sqrt{n}}{\sigma} - \frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\mu_4 - \sigma^4}{2\sqrt{n}\sigma^2}\xi\xi^*$$

dengan ξ dan ξ^* variabel random normal standar yang independen.

Bukti :

$$TJ = \left\{(\overline{Y} - \mu) + \lambda + \gamma\left[(\overline{Y} - \mu)^2 - \left(\frac{\sigma^2}{n}\right)\right]\right\}\left(\frac{S^2}{n}\right)^{-\frac{1}{2}}$$

$$\overline{Y} - \mu = \frac{\sigma}{\sqrt{n}}\xi + \frac{\mu_3}{6n\sigma^2}\xi^2 - \frac{\mu_3}{6n\sigma^2}$$

$$(\bar{Y} - \mu)^2 = \frac{\sigma^2}{n}\xi^2 + \frac{\mu_3^2}{36n^2\sigma^4}\xi^4 + \frac{\mu_3}{36n^2\sigma^4}$$

$$+ \frac{\mu_3}{3n\sqrt{n}\sigma}\xi^3 - \frac{\mu_3^2}{18n^2\sigma^4}\xi^2 - \frac{\mu_3}{6n\sqrt{n}\sigma}\xi .$$

$$\frac{S^2}{n} = \frac{\sigma^2}{n}\left\{1 + \left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\eta\right\}$$

$$\left(\frac{S^2}{n}\right)^{-\frac{1}{2}} = \frac{\sqrt{n}}{\sigma}\left\{1 + \left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\eta\right\}^{-\frac{1}{2}}$$

$$= \frac{\sqrt{n}}{\sigma}\left\{1 - \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\eta + \frac{3}{4 2!}\frac{\mu_4 - \sigma^4}{n\sigma^4}\eta^2 + \ldots\right\}$$

sehingga:

$$CF(TJ) = \left\{\frac{\gamma\mu_3^2}{36n^2\sigma^4}\xi^4 + \frac{\gamma\mu_3}{3n\sqrt{n}\sigma}\xi^3 + \left(\frac{\mu_3}{6n\sigma^2} + \frac{\gamma\sigma^2}{n} - \frac{\gamma\mu_3^2}{18n^2\sigma^4}\right)\xi^2\right.$$

$$\left. + \left(\frac{\sigma}{\sqrt{n}} - \frac{\gamma\mu_3}{6n\sqrt{n}\sigma}\right)\xi + \frac{\gamma\mu_3^2}{36n^2\sigma^4} + \lambda - \frac{\mu_3}{6n\sigma^2} - \frac{\gamma\sigma^2}{n}\right\}\frac{\sqrt{n}}{\sigma}$$

$$\left\{1 - \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\eta + \frac{3}{4.2!}\frac{\mu_4 - \sigma^4}{\sigma^4}\eta^2 - \ldots\right\}$$

$$CF(TJ) = \left\{\frac{\gamma\mu_3^2}{36n\sqrt{n}\sigma^5}\xi^4 + \frac{\gamma\mu_3}{3n\sigma^2}\xi^3 + \left(\frac{\mu_3}{6\sqrt{n}\sigma^3} + \frac{\gamma\sigma}{\sqrt{n}} - \frac{\gamma\mu_3^2}{18n\sqrt{n}\sigma^5}\right)\xi^2 + \right.$$

$$\left.\left(1 - \frac{\gamma\mu_3}{6n\sigma^2}\right)\xi + \frac{\gamma\mu_3^2}{36n\sqrt{n}\sigma^5} + \frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\gamma\sigma}{\sqrt{n}}\right\}$$

$$\left\{1 - \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\eta + \frac{3}{4 2!} - \frac{\mu_4 - \sigma^4}{n\sigma^4}\eta^2 \ldots\right\}$$

$$CF(TJ) = \left(\frac{\mu_3}{6\sqrt{n}\sigma^3} + \frac{\gamma\sigma}{\sqrt{n}}\right)\xi^2 + \xi + \frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\gamma\sigma}{\sqrt{n}}$$
$$- \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\xi\,\eta$$

Tulis $\eta = \rho\xi + \xi^*$ dengan ρ adalah kolerasi antara $\overline{X}$ dan S^2 dan ξ^* adalah variabel random normal independen terhadap ξ. $\rho = \mu_3\left\{\sigma^2\left(\mu_4 - \sigma^4\right)\right\}^{-\frac{1}{2}}$ sehingga :

$$CF(TJ) = \xi + \left(\frac{\mu_3}{6\sqrt{n}\sigma^3} + \frac{\gamma\sigma}{\sqrt{n}}\right)\xi^2 + \frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\gamma\sigma}{\sqrt{n}}$$
$$- \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\xi\left\{\frac{\mu_3}{\left[\sigma^2\left(\mu_4 - \sigma^4\right)\right]^{\frac{1}{2}}}\xi + \xi^*\right\}$$

$$CF(TJ) = \xi + \left(\frac{\mu_3}{6\sqrt{n}\sigma^3} + \frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{2\sqrt{n}\sigma^3}\right)\xi^2 + \frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\lambda\sigma}{\sqrt{n}}$$
$$- \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\xi\xi^*$$

Dengan menyeleksi γ dan λ sehingga koefesien dari ξ^2 adalah nol juga suku konstan berjumlah nol, ekspresi hasil akan mengurangi bias, didapat :

$$\frac{\mu_3}{6\sqrt{n}\sigma^3} + \frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{2\sqrt{n}\sigma^3} = 0$$

$$-\frac{\mu_3}{3\sqrt{n}\sigma^3} + \frac{\gamma\sigma}{\sqrt{n}} = 0$$

$$\gamma = \frac{\mu_3}{3\sigma^4}$$

dan $$\frac{\gamma\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\gamma\sigma}{\sqrt{n}} = 0$$

$$\frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n}\sigma^3} - \frac{\mu_3}{3\sqrt{n}\sigma^3} = 0$$

$$\frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{2\sqrt{n}\sigma^3} = 0$$

$$\lambda = \frac{\mu_3}{2n\sigma^2}$$

Jadi $$CF(TJ) = \xi - \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}}\xi\xi^* + o\left(n^{-1}\right)$$

$$= \xi - \frac{1}{2\sqrt{n}}\left(K_u - 1\right)^{\frac{1}{2}}\xi\xi^* + o\left(n^{-1}\right) \quad ...(3.2.4)$$

dan $$TJ = \left\{\left(\bar{Y} - \mu\right) + \frac{\mu_3}{6\sigma^2 n} + \frac{\mu_3}{3\sigma^4}\left(\bar{Y} - \mu\right)^2\right\}\left(\frac{S^2}{n}\right)^{-\frac{1}{2}} \blacksquare ...(3.2.5)$$

Terlihat bahwa TJ yang diberikan oleh (3.2.5) tidak dapat dihitung dengan H : $\mu = \mu_0$, karena μ_3 dan σ^2 biasanya tidak diketahui. Johnson menyarankan mengganti μ_3 dan σ^2 dengan estimasi sampel $\hat{\mu}_3 = \sum_{i=1}^{n} \frac{\left(Y_i - \bar{Y}\right)^3}{n}$ dan variansi sampel S^2. Ekspansi Cornish-Fisher masih (3.2.4)

Untuk contoh penggunaannya uji hipotesis $H_0 : \mu = \mu_0$ melawan satu dari tiga alternatif kemungkinan, maka dipunyai uji level α sebagai berikut :

TT : untuk $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$,

tolak H_0 jika $|TJH| > t_\alpha/2$ . . . . (3.2.6)

TU : untuk $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$,

tolak H_0 jika $TJH > t_\alpha$ . . . . (3.2.7)

Tl : untuk $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$,

tolak H_0 jika $TJH < -t_\alpha$ . . . . (3.2.8)

dimana $$TJH = \frac{\left(\bar{Y} - \mu_0\right) + \hat{\mu}_3 / (6s^2n) + \left(\hat{\mu}_3 / (3s^4)\right)\left(\bar{Y} - \mu\right)^2}{\left(\frac{s^n}{n}\right)^{\frac{1}{2}}} \quad (3.2.9)$$

dan t_α notasi bagian atas 100α titik persentil dari distribusi t dengan derajat bebas (n-1).

BAB IV

KETEGARAN STATISTIK t PADA INFERENSI UNTUK MEAN DUA POPULASI INDEPENDEN

4.1. Pendahuluan

Andaikan diberikan :

$$\left.\begin{array}{l} H_0: \mu_1 = \mu_2 \text{ dan} \\ H_1: \mu_1 > \mu_2 \end{array}\right\} \quad \ldots \quad (4.1.1)$$

dengan μ_i adalah mean dari variabel random normal $X_i \sim N(\mu_i, \sigma_i^2)$ $i = 1,2$. Lebih lanjut diberikan dua sampel independen, katakanlah X_{1j} , $j = 1,2, \ldots, n_1$

$X_{2\ell}$, $\ell = 1,2, \ldots, n_2$

sering digunakan statistik

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\left\{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{\frac{1}{2}}} \quad \ldots . (4.1.2)$$

dengan $S^2 = \dfrac{\left\{\sum_j \left(X_{1j} - \overline{X}_1\right)^2 + \sum_j \left(X_{2\ell} - \overline{X}_2\right)^2\right\}}{\left(n_1 + n_2 - 2\right)}$

$$\text{dan} \qquad \overline{X} = \frac{\sum_k X_{ik}}{n_i} \quad , i = 1,2$$

Jika $\sigma_1^2 = \sigma_2^2$ maka statistik $t \sim t(n_1 + n_2 - 2)$. Dan dipunyai uji level α untuk uji yang diberikan oleh (4.1.2). Tolak H_0 jika $t \geq t_\alpha$, t_α adalah bagian atas 100α titik persentil dari variabel $t(n_1 + n_2 - 2)$.

Posten, Yeh dan Owen (1982) memberikan suatu kualifikasi dari tingkat ketegaran uji t dua sampel dibawah asumsi variansi yang sama. Tingkat ketegaran diidentifikasi dengan mendapatkan daerah ketegaran dan hasil-hasil dibatasi untuk uji dua sisi. Lewat prosedur ini, hasil tersebut memberikan suatu ukuran yang jelas dari kekuatan ketegaran untuk variansi yang heterogen pada masalah ukuran sampel yang sama atau mendekati sama. Akan diperluas hasil-hasil yang didapat,untuk uji t dua sampel versi satu sisi.

Untuk memberikan informasi kuantitatif ketegaran dari uji t dengan keheterogenan variansi, dibutuhkan tiga definisi berikut.

Andaikan : $\lambda = \dfrac{\sigma_1^2}{\sigma_2^2}$

dan $$\left.\begin{aligned}\alpha(\lambda) &= P_r\left(\text{Tolak } H_0 \,/\, \lambda, H_0\right)\\ &= P_r\left(t \geq t_\alpha \,/\, \lambda, H_0\right)\end{aligned}\right\} \quad \ldots (4.1.3)$$

Definisi 4.1.1.

Diberikan $A = \{\lambda : \lambda > 0\}$

$R(\varepsilon)$ adalah daerah ketegaran pada level α jika $R(\varepsilon) \subset A \ni \forall\ \lambda \in R(\varepsilon)$, $|\alpha(\lambda) - \alpha(1)| \leq \varepsilon$.

Definisi 4.1.2.

Union dari semua daerah ketegaran pada level α disebut daerah maksimal ketegaran pada level α dan diberi simbol $R_m(\varepsilon)$.

Definisi 4.1.3.

Uji t adalah tegar secara total pada level α jika $A = \{\lambda : \lambda > 0\}$ adalah daerah ketegaran pada level α.

Dari definisi (4.1.2) dan (4.1.3), daerah maksimal ketegaran pada level α adalah :

$$R_{a}(\varepsilon) = \{\lambda \;/\; |\alpha(\lambda) - \alpha(1)| \leq \varepsilon\} \qquad \ldots\ldots(4.1.4)$$

dengan $$\alpha(\lambda) = \int_{t_\alpha}^{\infty} P_\lambda(t)\,dt$$

dimana $P_\lambda(t)$ adalah densitas dari t jika H_0 benar.

4.2. Kemonotonan $\alpha(\lambda)$

Sifat kemonotonan $\alpha(\lambda)$ penting dalam mengevaluasi atau menghitung daerah maksimal ketegaran (4.1.4). Berikut diberikan teorema-teorema yang berkaitan dengan sifat tersebut.

Theorema 4.2.1.

Jika $\alpha(\lambda, n_1, n_2)$ menyatakan tingkat kepercayaan ekor atas uji t dua sampel dengan ukuran sampel n_1 dan n_2

maka $\alpha(\lambda : n_1 , n_2) = \alpha\left(\frac{1}{\lambda} : n_2, n_1\right)$. Maka untuk membuktikan teorema ini kita ikuti uraian dibawah ini :

Masalah perhitungan tingkat ketegaran dan daerah maksimal ketegaran didasarkan pada perhitungan $\alpha(\lambda)$ dalam (4.1.3) dan (4.1.4). Dari Yeh(1981), probabilitas ini dapat dinyatakan dalam suku-suku integral dari probabilitas pembobot student, hubungan ini adalah :

$$\alpha(\lambda; n_1, n_2) = \int_0^1 b(X; n_1, n_2) P_r\left(T^* \geq m(\lambda, X; n_1, n_2) t_\alpha\right) dx \qquad (4.2.1)$$

dimana :

$N = n_1 + n_2$, t_α adalah nilai kritis dari uji t satu sisi.

$$r = \frac{n_1}{n_2}$$

$$m(\lambda, X; n_1, n_2) = \left\{\frac{(1 + r)}{\lambda + r}\right\}^{\frac{1}{2}} \{\lambda + (1 - \lambda)x\}^{\frac{1}{2}} \qquad . . . (4.2.2)$$

$$b(x; n_1, n_2) = \frac{\Gamma\left(\frac{N - 2}{2}\right)}{\Gamma\left(\frac{n_1 - 1}{2}\right)\Gamma\left(\frac{n_2 - 1}{2}\right)} x^{\frac{(n_2 - 3)}{2}} (1 - x)^{\frac{(n_1 - 3)}{2}}$$

T^* variabel t pusat dengan derajat kebebasan N - 2 sehingga bukti theorema 4.3.1 adalah sebagai berikut dari (4.2.2) didapat:

$$m\left(\frac{1}{\lambda}, x; n_1, n_2\right) = \left\{\frac{1 + r}{\frac{1}{\lambda} + r}\right\}^{\frac{1}{2}} \left\{\frac{1}{\lambda} + \left(1 - \frac{1}{\lambda} x\right)\right\}^{\frac{1}{2}}$$

$$= \left\{\frac{1 + r}{1 + \lambda r}\right\}^{\frac{1}{2}} \{1 + (\lambda - 1)x\}^{\frac{1}{2}}$$

$$= \left\{\frac{1 + r^*}{\lambda + r^*}\right\}^{\frac{1}{2}} \{\lambda + (1 - \lambda)(1 - x)\}^{\frac{1}{2}}$$

dengan $r^* = \frac{1}{r} = \frac{n_2}{n_1}$ sehingga

$$m\left(\frac{1}{\lambda}, x; n_1, n_2\right) = m\left(\lambda(1 - x); n_2, n_1\right)$$

dan karena

$$b(1-x; n_2, n_1) = b(x ; n_1, n_2)$$

maka $\alpha\left(\frac{1}{\lambda}; n_1, n_2\right) = \int_0^1 b\left(x; n_1, n_2\right) P_r\left(T^* \geq m\left(\frac{1}{\lambda}, x; n_1, n_2\right) t\alpha\right) dx$

$$= \int_0^1 b\left(1 - x; n_2, n_1\right) P_r\left(T^* \geq m\left(\lambda, (1 - x); n_2, n_1\right) t\alpha\right) dx$$

ambil y = 1-x maka :

$$\alpha\left(\frac{1}{\lambda}; n_1, n_2\right) = \int_0^1 b(y; n_2, n_1) P_r\left(T^* \geq m(\lambda, y; n_2, n_1)t\alpha\right)dy$$
$$= \alpha(\lambda; n_2, n_1)$$

Akibat 1.2.1

Untuk $n_1 = n_2 = n$ maka $\alpha(\lambda) = \alpha\left(\frac{1}{\lambda}\right)$

Theorema 4.2.2.

Untuk ukuran sampel sama, $\alpha(\lambda)$ monoton turun untuk $\lambda \in (0,1)$ dan monoton naik untuk $\lambda \in (1,\infty)$.

Bukti :

Dari (4.2.2) terlihat

$$m(\lambda, x; n_1, n_2) = \left\{\frac{1 + r}{\lambda + r}\right\}^{\frac{1}{2}}\{\lambda + (1 - \lambda)x\}^{\frac{1}{2}}$$

maka :

$$\frac{dm}{d\lambda} = \frac{-\frac{1}{2}(1 + r)^{\frac{1}{2}}}{\lambda + r}(\lambda + r)^{\frac{1}{2}}(\lambda + (1 - \lambda)x)^{\frac{1}{2}}$$
$$+ \frac{1}{2}\left(\frac{1 + r}{\lambda + r}\right)^{\frac{1}{2}}(\lambda + (1 - \lambda)x)^{-\frac{1}{2}}(1 - x)$$

$$\frac{dm}{d\lambda} = \frac{1}{2}(1 + r)^{\frac{3}{2}}(\lambda + r)^{-\frac{3}{2}}(\lambda + (1 - \lambda)x)^{-\frac{1}{2}}$$
$$\left\{\frac{(\lambda + r)(1 - x)}{1 + r} - \frac{\lambda + (1 - \lambda)x}{1 + r}\right\}$$
$$= \frac{1}{2}(1 + r)^{\frac{3}{2}}(\lambda + r)^{-\frac{3}{2}}(\lambda + (1 - \lambda)x)^{-\frac{1}{2}}\left\{\frac{r - x(1 + r)}{1 + r}\right\}$$
$$= \frac{1}{2}(1 + r)^{\frac{3}{2}}(\lambda + r)^{-\frac{3}{2}}(\lambda + (1 - \lambda)x)^{-\frac{1}{2}}\left(\frac{r}{1 + r} - x\right)$$

dari (4.2.2)

$$\alpha(\lambda; n_1, n_2) = \int_0^1 b(x; n_1, n_2) P_r\left(T^* \geq m(\lambda, x; n_1, n_2)t_\alpha\right)dx$$

maka :

$$1-\alpha(\lambda; n_1, n_2 1) = \int_0^1 b(x; n_1, n_2) P_r\left(T^* < m(\lambda, x; n_1, n_2)t_\alpha\right)dx$$

$$= \int_0^1 b(x; n_1, n_2)\, F\left(m(\lambda, x; n_1, n_2)t_\alpha; f\right)dx \quad \ldots\ldots \quad (4.2.3)$$

dengan F(;f) fungsi distribusi variabel t dengan derajat kebebasan f = N - 2.

Sehingga :

$$\frac{d\left(1 - \alpha(\lambda; n_1, n_2)\right)}{d\lambda} = \int_0^1 b(x; n_1, n_2)\, \frac{dF\left(m(\lambda, x; n_1, n_2)t_\alpha; f\right)}{dm(\lambda, x; n_1, n_2)t_\alpha}$$
$$\frac{dm(\lambda, x; n_1, n_2)t_\alpha}{d\lambda}\, dx$$

$$\frac{-d\alpha(\lambda;\, n_1,\, n_2)}{d\lambda} = \int_0^1 b(x;\, n_1,\, n_2)\, f\big(m(\lambda,\, x;\, n_1,\, n_2)t_\alpha;\, f\big)\, \frac{dm(\lambda,\, x;\, n_1,\, n_2)}{d\lambda}\, dx$$

dengan $f(\ldots, f)$ adalah fungsi densitas variabel t dengan derajat bebas f, sehingga:

$$\frac{-d\alpha(\lambda;\, n_1,\, n_2)}{d\lambda} = \frac{t\alpha}{2}\int_0^1 \frac{b(x;\, n_1,\, n_2)\Gamma\left(\frac{f+1}{2}\right)}{\sqrt{f\pi}\;\Gamma\left(\frac{f}{2}\right)}$$

$$\left(1 + m^2(\lambda,\, x;\, n_1,\, n_2)t_\alpha^2 \;/\; f\right)^{\frac{-(f+1)}{2}}$$

$$(1 + r)^{\frac{3}{2}}(\lambda + r)^{-\frac{3}{2}}\{\lambda + (1 - \lambda)x\}^{-\frac{1}{2}}\left(\frac{r}{1 + r} - x\right)dx$$

$$\frac{-d\alpha(\lambda;\, n_1,\, n_2)}{d\lambda} = C^* t_\alpha A(\lambda)$$

dengan :

$$C^* = \frac{\Gamma\left(\frac{f+1}{2}\right)(1 + r)^{\frac{3}{2}}(\lambda + r)^{-\frac{3}{2}}}{2\sqrt{f\pi}\;\Gamma\left(\frac{n_1 - 1}{2}\right)\Gamma\left(\frac{n_2 - 1}{2}\right)}$$

$$b^*(x) = x^{\frac{(n_2-3)}{2}}(1 - x)^{\frac{(n_1-3)}{2}} \quad \text{dan}$$

$$A(\lambda) = \int_0^1 b^*(x)\left\{1 + m^2(\lambda,\, x;\, n_1,\, n_2)t_\alpha^2 \;/\; f\right\}^{-(f+1)/2}$$

jadi tanda dari $d\alpha(\lambda; n_1, n_2)$ adalah negatif tanda dari $A(\lambda)$. Untuk menentukan tanda suku akhir, tulis

$$A(\lambda) = \int_0^1 b^*(x)\, g(x)\left(\frac{r}{1+r} - x\right)dx$$

dengan

$$g(x) = \left\{1 + m^2(\lambda, x; n_1, n_2)t_\alpha^2 \;/\; f\right\}^{-(f+1)/2}\{\lambda + (1-\lambda)x\}^{-\frac{1}{2}}$$

Ambil $n_2 = n_2 = n$

$$b^*(x) = \{x(1-x)\}^{\frac{(n-3)}{2}} = b^*(1-x) \text{ maka}$$

$$A(\lambda) = \int_0^{\frac{1}{2}} b^*(x)\, g(x)\left(\frac{1}{2} - x\right)dx + \int_{\frac{1}{2}}^1 b^*(x)\, g(x)\left(\frac{1}{2} - x\right)dx$$

Ambil $x = 1 - y$ dalam integral kedua

$$A(\lambda) = \int_0^{\frac{1}{2}} b^*(x) g(x)\left(\frac{1}{2} - x\right)dx + \int_0^{\frac{1}{2}} b^*(x)\, g(1-y)\left(y - \frac{1}{2}\right)dy$$

$$= \int_0^{\frac{1}{2}} b^*(x)\{g(x) - g(1-x)\}\left(\frac{1}{2} - x\right)dx$$

$$= \int_0^{\frac{1}{2}} b^*(x)\, g(1-x)\left\{\frac{g(x)}{g(1-x)}\right\}\left(\frac{1}{2} - x\right)dx$$

Karena :

$$\frac{g(x)}{g(1-x)} =$$

$$\frac{\left\{1 + m^2(\lambda, x; n_1, n_2)t_\alpha^2 / f\right\}^{-\frac{(f+1)}{2}} \{\lambda + (1-\lambda)x\}^{-\frac{1}{2}}}{\left\{1 + m^2(\lambda, (1-x); n_1, n_2)t_\alpha^2 / f\right\}^{-\frac{(f+1)}{2}} \{\lambda + (1-\lambda)(1-x)\}^{-\frac{1}{2}}}$$

$$\frac{g(x)}{g(1-x)} = u(\lambda, x)^{\frac{(f+1)}{2}} V(\lambda, x)^{\frac{1}{2}}$$

dengan $\quad u(\lambda, x) = \dfrac{\left\{1 + m^2(\lambda, (1-x); n_1, n_2)t_\alpha^2 / f\right\}}{\left\{1 + m^2(\lambda, x; n_1, n_2)t_\alpha^2 / f\right\}}$

$$V(\lambda, x) = \frac{\lambda + (1-\lambda)(1-x)}{\lambda + (1-\lambda)x}$$

sehingga $A(\lambda) = \displaystyle\int_0^{\frac{1}{2}} b^*(x)\, g(1-x)\left\{u(\lambda, x)^{\frac{(f+1)}{2}} V(\lambda, x)^{\frac{1}{2}} - 1\right\}dx$

$A(\lambda)$ positif jika $u(\lambda,x)$ dan $v(\lambda,x)$ keduanya lebih dari satu untuk semua $x \in \left(0, \frac{1}{2}\right)$ dan $A(\lambda)$ negatif jika keduanya kurang dari satu untuk semua $x \in \left(0, \frac{1}{2}\right)$.

Dari semua lemma (4.2.1) dan semua lemma (4.2.2) dibawah ini, jika $\lambda \in (0,1)$, $u(\lambda,x)$ dan $v(\lambda,x)$ melebihi satu untuk setiap $x \in \left(0, \frac{1}{2}\right)$ dan jika $\lambda \in (1,\infty)$ kedua fungsi tersebut kurang dari satu untuk setiap $\lambda \in (0,1)$ dan negatif untuk setiap $\lambda \in (1,\infty)$■.

Lemma 4.2.1.

Diberikan $v(\lambda,x)$, maka untuk setiap $\lambda \in (0,1)$, $v(\lambda,x)$ lebih besar dari satu untuk setiap $x \in \left(0, \frac{1}{2}\right)$ dan untuk setiap $\lambda \in (1,\infty)$, $v(\lambda,x)$ lebih kecil dari satu untuk setiap $x \in \left(0, \frac{1}{2}\right)$.

Bukti :

$$v(\lambda, x) > 1 \Leftrightarrow \lambda + (1 - \lambda)(1 - x) > (1 - \lambda)x + \lambda$$
$$(1 - \lambda)(1 - 2x) > 0$$

Karena $\forall x \in \left(0, \frac{1}{2}\right)$, $1 - 2 > 0$ sehingga $v(\lambda,x) > 1$ untuk $1 - \lambda > 0$ atau $\lambda < 1$. Bukti untuk yang lebih kecil dari satu langkahnya sama dengan yang diatas■.

Lemma 4.2.2.

Diberikan $u(\lambda,x)$, maka $\forall x \in (0,1)$, $u(\lambda,x) > 1$, $\forall x \in \left(0, \frac{1}{2}\right)$. dan $\forall\lambda \in (1,\infty)$, $u(\lambda,x) < 1$, $\forall x \in \left(0, \frac{1}{2}\right)$.

Bukti :

$$\begin{aligned} u(\lambda, x) > 1 \Leftrightarrow m^2(\lambda, 1 - x; n_1, n_2) &> m^2(\lambda, x; n_1, n_2) \\ \lambda + (1-\lambda)(1-x) &> \lambda + (1-\lambda)x \\ (1-\lambda)(1-2x) &> 0 \end{aligned}$$

Karena $\forall x \in \left(0, \frac{1}{2}\right)$, $1-2x > 0$ sehingga $u(\lambda,x) > 1$ untuk $1 - \lambda > 0$ atau $\lambda < 1$.

Jadi untuk setiap $x \in (0,1)$ maka $A(\lambda)$ positif dan akibatnya $\frac{d\alpha}{d\lambda}$ negatif sehingga turun. Untuk setiap $\lambda \in (1-\infty)$ maka $A(\lambda)$ negatif dan akibatnya $\frac{d\alpha}{d\lambda}$ positif sehingga α naik ■.

Theorema 4.2.3.

Untuk ukuran sampel sama, daerah maksimal ketegaran tingkat ε, $R_m(\varepsilon)$ mempunyai bentuk $\left[c, \frac{1}{c}\right]$ dimana $c < 1$ yang didefinisikan dari $\alpha(c) - \alpha(1) = \varepsilon$.

Bukti :

Mengikuti langsung dari theorema(4.2.2)dan definisi (4.1.1) dan (4.1.2) ■.

4.3. Daerah Ketegaran

Ukuran sampel yang dipakai dalam studi ini (untuk $n_1 \neq n_2$) adalah $n_1 < n_2$ dengan tingkat kepercayaan nominal $\alpha(\lambda)$ yang merupakan fungsi monoton naik dalam λ pada $\lambda \in (0,\infty)$.

Daerah Maksimal ketegaran akan diberikan untuk ukuran sampel yang sama ataupun tak sama. Akan tetapi, pertama kita akan pertimbangkan kemungkinan nilai-nilai ε yang cukup kecil sehingga uji t satu sisi akan tegar secara total.

Untuk ukuran sampel sama ($n_1 = n_2 = n$) dengan sifat kemonotonan dari $\alpha(\lambda)$ berakibat bahwa untuk semua nilai dari ε yang melebihi atau sama dengan $|\alpha(0) - \alpha(1)|$ maka uji t akan tegar secara total pada level ε.

Lebih lanjut perlu untuk menghitung $\alpha(0)$ guna mengidentifikasi nilai terkecil dari ε, untuk mana uji akan tegar secara total. Sedang untuk ukuran sampel tidak sama perlu juga untuk menghitung $|\alpha(\infty) - \alpha(1)|$ dan pilih nilai terbesar dari keduanya sebagai nilai terkecil dari ε untuk mana uji akan tegar secara total.

Tabel (4.3.1) tersedia ukuran sampel $n_1 = n_2 = n, \alpha = 0{,}05$ dan $\alpha = 0{,}01$ nilai-nilai minimal ε untuk mana uji t satu sisi tegar secara total. Dari tabel 4.3.1 terlihat dengan jelas bahwa untuk ukuran sampel n = 15 dan $\alpha = 0{,}05$, uji t tegar secara total pada level 0,0055. Bahwa tak ada masalah seberapa besar perbedaan σ_1^2 dan σ_2^2 . Kebenaran tingkat kepercayaan adalah sampai 0,0055 dari level yang diinginkan sebesar 0,05. Lebih lanjut selisih dari α disini terjadi hanya jika σ_1^2 sangat berbeda dari σ_2^2 . Untuk nilai n yang lebih besar, nilai ε minimum yang menghasilkan tegar secara total akan lebih kecil. Tabel 4.3.1 ini juga dapat digunakan untuk menentukan derajat uji t satu sisi (dengan ukuran sampel sama dibawah pengabaian asumsi kehomogenan variansi).

Untuk ukuran sampel tidak sama, nilai terkecil dari ε yang memberikan ketegaran secara total untuk uji t satu sisi, ditentukan dengan evaluasi $\alpha(1)-\alpha(0)$ dan $\alpha(\infty)-(1)$ dimana nilai ε adalah nilai terbesar diantara keduanya. Adapun nilai-nilai dari $\alpha(0)$ dan $\alpha(\infty)$ diberikan dalam tabel 4.3.2 untuk ukuran sampel yang tidak sama, tetapi untuk ukuran sampel yang sama juga diberikan sebagai pembanding.

Karena tingkat kepercayaan nominal $\alpha(1)$ diketahui maka nilai ε untuk mana uji akan tegar secara total siap didapat dari tabel (4.3.2).

Selanjutnya daerah maksimal ketegaran diberikan didalam tabel (4.3.3) dan (4.3.4) masing-masing untuk $\alpha(1) = 0{,}05$, $\varepsilon = 0{,}005$, $0{,}03$, $0{,}02$ dan $\alpha(1) = 0{,}01$, $\varepsilon = 0{,}02$, $0{,}01$, $0{,}005$ pada beberapa nilai n_1 dan n_2.

4.4. Prosedur perhitungan

Dalam masalah perhitungan tingkat ketegaran dan daerah maksimal ketegaran seperti yang ditulis pada bagian awal, didasarkan pada perhitungan $\alpha(\lambda)$ dalam (4.2.1) yaitu :

$$\alpha(\lambda; n_1, n_2) = \int_0^1 b(x; n_1, n_2)\, P_r\left(T^* \geq m(\lambda, x; n_1, n_2)t_\alpha\right)dx$$

dimana untuk setiap nilai tertentu dari x, probabilitas T^* dapat dihitung dengan menggunakan hubungan antara variabel student t dengan beta untuk nilai t yang positif.

$$P_r(T^* < t) = \int_{-\infty}^{t} \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)} \frac{1}{\sqrt{(N-2)\pi}} \left(1 + \frac{x^2}{N-2}\right)^{-\left(\frac{N-1+1}{2}\right)} dx$$

$$P(T^* < t) = \frac{1}{2} + \int_0^t \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{1}{2}\right)\sqrt{n-2}} \left(1 + \frac{x^2}{N-2}\right)^{-\left(\frac{N-1}{2}\right)} dx$$

$$= \frac{1}{2} + \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^t \frac{1}{\sqrt{N-2}} \left(1 + \frac{x^2}{N-2}\right)^{-\left(\frac{N-1}{2}\right)} dx$$

dimana suku dalam integral pada persamaan terakhir dapat ditulis dengan : $\frac{1}{2}\int_0^{Y} s^{-\frac{1}{2}}(1-s)^{\frac{N-4}{2}}ds$

dimana $$y = \frac{t^2}{N - 2 + t^2}$$

menggunakan subsitusi $$\frac{x^2}{N - 2 + x^2} = s$$

sehingga:

$$P(T^* < t) = \frac{1}{2} + \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{2\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^y s^{-\frac{1}{2}}(1-s)^{\frac{N-4}{2}} ds$$

$$= \frac{1}{2} + \frac{1}{2} I_y\left(\frac{1}{2}, \frac{f}{2}\right)$$

$$= \frac{1}{2}\left(1 + I_y\left(\frac{1}{2}, \frac{f}{2}\right)\right)$$

dengan : $f = N - 2$ dan $I_y\left(\frac{1}{2}, \frac{f}{2}\right) = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^y s^{-\frac{1}{2}}(1-s)^{\frac{f-2}{2}} ds$

Tabel 4.3.1 (Diambil dari Commun.Statist.-Theory Meth.,21(8),2169-2184).

Nilai minimum dari ε untuk mana Uji-t satu sisi tegar secara total pada tingkat ε ($n_1 = n_2 = n$)

	$\alpha = 0.05$	$\alpha = 0.01$		$\alpha = 0.05$	$\alpha = 0.01$
n	ε	ε	n	ε	ε
3	.0333	.0222	15	.0055	.0036
4	.0236	.0158	20	.0041	.0026
5	.0182	.0121	25	.0032	.0021
6	.0148	.0098	30	.0027	.0017
7	.0125	.0082	40	.0020	.0013
8	.0108	.0071	50	.0016	.0010
9	.0094	.0062	100	.0008	.0005
10	.0085	.0055	∞	0	0

Tabel 4.3.2(Diambil dari Commun.Statist.-Theory Meth.,21(8),2169-2184).

Nilai dari $\alpha(0)$ dan $\alpha(\infty)$ jika $\infty(1)=0.05$

Ukuran sampel sama

n_1	n_2	$\alpha(0)$	$\alpha(\infty)$
5	5	.0682	.0682
10	10	.0585	.0585
15	15	.0555	.0555
20	20	.0541	.0541
25	25	.0532	.0532
30	30	.0527	.0527
40	40	.0520	.0520
50	50	.0516	.0516

Ukuran sampel berbeda 10%

n_1	n_2	$\alpha(0)$	$\alpha(\infty)$
9	11	.0414	.0788
18	22	.0379	.0734
27	33	.0367	.0717
36	44	.0362	.0708
45	55	.0358	.0703

Ukuran sampel berbeda 20%

n_1	n_2	$\alpha(0)$	$\alpha(\infty)$
8	12	.0276	.1027
12	18	.0276	.0980
16	24	.0247	.0958
20	30	.0242	.0945
24	36	.0233	.0926
32	48	.0233	.0926
40	60	.0231	.0920

Nilai dari $\alpha(0)$ dan $\alpha(\infty)$ jika $\infty(1)=0.01$

Ukuran sampel sama

n_1	n_2	$\alpha(0)$	$\alpha(\infty)$
5	5	.0221	.0221
10	10	.0155	.0155
15	15	.0136	.0136
20	20	.0126	.0126
25	25	.0121	.0121
30	30	.0117	.0117
40	40	.0113	.0113
50	50	.0110	.0110

Ukuran sampel berbeda 10%

n_1	n_2	$\alpha(0)$	$\alpha(\infty)$
18	22	.0068	.0214
27	33	.0062	.0201
36	44	.0059	.0195
45	55	.0057	.0191

Ukuran sampel berbeda 20%

n1	n2	$\alpha(0)$	$\alpha(\infty)$
8	12	.0046	.0395
12	18	.0037	.0356
16	24	.0033	.0338
20	30	.0031	.0327
24	36	.0029	.0321
32	48	.0027	.0312
40	60	.0026	.0307

Tabel 4.3.2(Diambil dari Commun.Statist.-Theory Meth.,21(8),2169-2184).

Nilai dari $\alpha(0)$ dan $\alpha(\infty)$ jika $\infty(1)$=0.05

Ukuran sampel sama				Ukuran sampel berbeda 10%				Ukuran sampel berbeda 20%			
n_1	n_2	$\alpha(0)$	$\alpha(\infty)$	n_1	n_2	$\alpha(0)$	$\alpha(\infty)$	n_1	n_2	$\alpha(0)$	$\alpha(\infty)$
5	5	.0682	.0682								
10	10	.0585	.0585	9	11	.0414	.0788	8	12	.0276	.1027
15	15	.0555	.0555					12	18	.0276	.0980
20	20	.0541	.0541	18	22	.0379	.0734	16	24	.0247	.0958
25	25	.0532	.0532					20	30	.0242	.0945
30	30	.0527	.0527	27	33	.0367	.0717	24	36	.0233	.0926
40	40	.0520	.0520	36	44	.0362	.0708	32	48	.0233	.0926
50	50	.0516	.0516	45	55	.0358	.0703	40	60	.0231	.0920

Nilai dari $\alpha(0)$ dan $\alpha(\infty)$ jika $\infty(1)$=0.01

Ukuran sampel sama				Ukuran sampel berbeda 10%				Ukuran sampel berbeda 20%			
n_1	n_2	$\alpha(0)$	$\alpha(\infty)$	n_1	n_2	$\alpha(0)$	$\alpha(\infty)$	n1	n2	$\alpha(0)$	$\alpha(\infty)$
5	5	.0221	.0221								
10	10	.0155	.0155					8	12	.0046	.0395
15	15	.0136	.0136					12	18	.0037	.0356
20	20	.0126	.0126	18	22	.0068	.0214	16	24	.0033	.0338
25	25	.0121	.0121					20	30	.0031	.0327
30	30	.0117	.0117	27	33	.0062	.0201	24	36	.0029	.0321
40	40	.0113	.0113	36	44	.0059	.0195	32	48	.0027	.0312
50	50	.0110	.0110	45	55	.0057	.0191	40	60	.0026	.0307

Tabel 4.3.3(Diambil dari Commun.Statist.-Theory Meth.,21(8),2169-2184).

Daerah maksimal ketegaran pada tingkat ε
untuk uji-t satu sisi (tingkat kepercayaan nominal $\alpha(1)=0,01$)

$\varepsilon = 0.05$

Ukuran sampel sama			Ukuran sampel berbeda 10%			Ukuran sampel berbeda 20%		
n_1	n_2	λ range	n_1	n_2	λ range	n1	n2	λ range
5	5	0 - ∞						
10	10	0 - ∞	9	11	0 - ∞	8	12	0-60.333
15	15	0 - ∞				12	18	0 - ∞
20	20	0 - ∞	18	22	0 - ∞	16	24	0 - ∞
25	25	0 - ∞				20	30	0 - ∞
30	30	0 - ∞	27	33	0 - ∞	24	36	0 - ∞
40	40	0 - ∞	36	44	0 - ∞	32	48	0 - ∞
50	50	0 - ∞	45	55	0 - ∞	40	60	0 - ∞

$\varepsilon = 0,03$

n_1	n_2	λ range	n_1	n_2	λ range	n_1	n_2	λ range
5	5	0 - ∞						
10	10	0 - ∞	9	11	0 - ∞	8	12	0-5.160
15	15	0 - ∞				12	18	0 - 5.732
20	20	0 - ∞	18	22	0 - ∞	16	24	0 - 6.121
25	25	0 - ∞				20	30	0 - 6.403
30	30	0 - ∞	27	33	0 - ∞	24	36	0 - 6.616
40	40	0 - ∞	36	44	0 - ∞	32	48	0 - 6.692
50	50	0 - ∞	45	55	0 - ∞	40	60	0 - 7.119

$\varepsilon = 0,02$

n_1	n_2	λ range	n_1	n_2	λ range	n_1	n_2	λ range
5	5	0 - ∞						
10	10	0 - ∞	9	11	0 - 8.481	8	12	.089-2.923
15	15	0 - ∞				12	18	.135-3.030
20	20	0 - ∞	18	22	0 - 17.169	16	24	.153-3.095
25	25	0 - ∞				20	30	.164-3.137
30	30	0 - ∞	27	33	0 - 31.483	24	36	.170-3.168
40	40	0 - ∞	36	44	0 - 59.817	32	48	.178-3.209
50	50	0 - ∞	45	55	0 - 142.976	40	60	.183-3.234

BAB V

KESIMPULAN

Dari uraian didepan ,maka dapat diambil beberapa kesimpulan sebagai berikut:

Bahwa kemiringan distribusi populasi mempengaruhi kemiringan distribusi T(2.1.1) pada inferensi untuk mean satu populasi, yaitu jika distribusi populasi mempunyai kemiringan kekanan dalam hal ini $S_k > 0$ maka Mean dari $T = -(n^{-1/2}/2)S_k$ < Median dari $T = -(n^{-1/2}/6)S_k < 0$ sehingga distribusi T miring kekiri dan $P_U(\alpha) < \alpha$ serta $P_L(\alpha) > \alpha$. Sebaliknya untuk distribusi populasi mempunyai kemiringan kekiri dalam hal ini $S_k < 0$ maka Mean dari $T = -(n^{-1/2}/2)S_k$ >Median dari $T = -(n^{-1/2}/6)S_k > 0$ sehingga distribusi T miring kekanan. Jadi untuk kasus dimana distribusi populasi miring ditambah dengan ukuran sampel yang kecil, maka distribusi T(2.1.1) ini tak akan didekati dengan distribusi student-t kecuali telah dilakukan koreksi kemiringan seperti yang telah dilakukan Johnson yaitu mengganti statistik T(2.1.1) dengan statistik TJH (2.2.9). Sedangkan untuk mean dua populasi independen versi satu sisi dengan $n_1=n_2=n$,menggunakan sifat kemonotonan $\alpha(\lambda)$ berakibat bahwa untuk semua nilai ε yang melebihi atau sama dengan $\alpha(0)-\alpha(1)$ maka uji t(3.1.2) akan tegar secara total pada level ε.

Berdasarkan hasil-hasil yang dibahas pada bagian awal maka dapat pula diperluas kemasalah yang masih berkaitan dengan uji hipotesis seperti kemasalah interval konfidensi yaitu bentuk interval konfidensi untuk modifikasi statistik T dan uji hipotesis untuk inferensi mean lebih dari dua populasi.

KETEGARAN STATISTIK t UNTUK SATU POPULASI DAN DUA POPULASI

Ringkasan Tesis
Untuk memenuhi sebagian persyaratan
mencapai derajat sarjana S-2

Program Studi Matematika
Jurusan Matematika dan
Ilmu Pengetahuan Alam

Oleh :

Joko Riyono
6423/I-4/485/94

PROGRAM PASCA SARJANA
UNIVERSITAS GADJAH MADA
1998

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Statistik t seringkali dipakai dalam membuat inferensi mean satu populasi dan inferensi untuk mean dua populasi Independen. Dalam tulisan ini akan difokuskan pada dua topik tersebut, khususnya yang menyangkut permasalahan ketegaran dari statistik t.

Dalam inferensi untuk mean satu populasi, diambil sampel Y_i , $i = 1,2, \ldots n$ dan dihitung $\overline{Y}$, S^2. Dengan asumsi populasi berdistribusi normal maka dapat diperoleh statistik T untuk inferensi mean satu populasi, yaitu :

$$T = \frac{\overline{Y} - \mu}{\left\{ S^2 / n \right\}^{\frac{1}{2}}} \qquad \ldots\ldots\ldots \quad (1.1.1)$$

untuk n kecil T akan berdistribusi t dengan derajat bebas n - 1, sedang untuk n besar T akan mendekati distribusi normal standar.

Pada prakteknya, pemakai sering kesulitan menentukan berapa besar n hingga metode ini dapat digunakan atau bagaimana seandainya populasi tidak berdistribusi normal, apakah metode ini masih dapat juga digunakan.

Sedang pada inferensi untuk mean dua populasi independen, diambil sampel X_{1j} , $j = 1,2, \ldots, n_1$ dan X_{2l} , $l = 1,2, \ldots, n_2$ kemudian hitung $\bar{X}_1, \bar{X}_2, S^2$. Dengan asumsi distribusi kedua populasi normal dan variansi keduanya sama maka dapat diperoleh statistik t untuk inferensi mean dua populasi independen, yaitu :

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{x_1} - \mu_{x_2})}{\left[S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{\frac{1}{2}}} \qquad \ldots .(1.1.2)$$

Sama halnya seperti pada inferensi untuk mean satu populasi, bagaimana seandainya assumsi kehomogenan variansi dilanggar apakah metode ini masih dapat digunakan. Masalah seperti ini dengan σ_1^2 dan σ_2^2 tidak diketahui ($\sigma_1^2 \neq \sigma_2^2$)disebut sebagai masalah Behren-Fisher. Meskipun untuk masalah tersebut telah ada cara pemecahannya, seperti yang diberikan oleh Neyman-Bartlett, Scheffe' ataupun Hsu, namun ketiga cara diatas masih memiliki kekurangan, terutama untuk $n_1 \neq n_2$ yaitu adanya sampel yang harus diabaikan, hasil yang masih tergantung pada urutan pengambilan sampel atau tidak adanya jaminan uji mempunyai tingkat kepercayaan pada α yang kita inginkan. Oleh karena itu pentinglah bagi pemakai untuk melihat sampel,sejauh mana perbedaan σ_1^2 dan σ_2^2 yang masih bisa ditolelir sehingga metode diatas masih dapat digunakan.

Permasalahan-permasalahan seperti diatas dimana beberapa asumsi dilanggar, didalam statistik dikenal sebagai permasalahan ketegaran (robustness) dari suatu statistik (estimator) atau suatu prosedur statistik. Menurut Hubber (1981) ketegaran adalah ketidaksensitifan terhadap sedikit penyimpangan dari asumsi. Suatu statistik atau prosedur statistik dikatakan tegar apabila ia dapat bekerja dengan baik, mempunyai bias dan variansi rendah dalam segala macam distribusi.

Studi tentang ketegaran statistik t telah banyak dilakukan, antara lain oleh Cicchitelli(1989) yang menunjuk suatu suatu studi Monte Carlo untuk meneliti pengaruh uji inferensi mean satu populasi, dengan statistik T diatas pada populasi nonnormal. Dia memperlihatkan bahwa kemiringan yang ada dalam populasi kemiringaan statistik T menggunakan ekspansi Cornish-Fisher. Sedangkan Posten, Yeh dan Owen (1982) meneliti ketegaran uji t dua populasi versi dua sisi untuk kasus penyimpangan pada homogenitas variansi kedua populasi. Dalam tulisan ini nantinya akan dibahas hasil-hasil yang telah didapat Cicchitelli(1989) diatas menggunakan ekspansi Edgeworth, kemudian juga melihat penampilan modifikasi statistik T dari Johnson (1978) untuk dibandingkan dengan penampilan statistik T itu sendiri. Pada uji t dua populasi, kita akan mengembangkan hasil-hasil yang didapat oleh Posten, Yeh dan Owen (1982) untuk versi uji t satu sisi.

Tulisan ini diharapkan akan dapat memberi sumbangan yang berharga dibidang statistika inferensial. Selain itu diharapkan pembaca dapat menggunakan atau mengambil manfaat dari apa yang bisa diserap dalam tulisan ini.

1.2. Perumusan Masalah

Berdasarkan latar belakang dan ruang lingkup masalah, masalah dalam tulisan ini dapat dirumuskan sebagai berikut :

1. Mempelajari pengaruh kemiringan populasi terhadap statistik T(1.1.1).
2. Mempelajari cara memperoleh modifikasi statistik T, yang dapat mengurangi pengaruh kemiringan yang dibawa oleh populasi terhadap statistik T (1.1.1).
3. Melihat nilai keheterogenan variansi yang masih dapat ditolelir sehingga metode dengan menggunakan statistik t(1.1.2) versi satu sisi masih berlaku.

1.3. Tinjauan Pustaka

Kekonservatifan uji T dipelajari oleh Gross (1976) dan Tukey & Mc. Laughlin (1963).Benjamini (1983) menunjukkan kekonservatifan uji T untuk distribusi populasi long tailed menggunakan argumen geometrik, selanjutnya pendekatan geometris diambil oleh Hotelling (1961) dan Efron (1969).Yuen & Murthy (1974)mempelajari masalah spesifik pengamatan gambar dari distribusi

Student-t,mereka menabelkan titik-titik tertentu dari distribusi.

Posten, Yeh dan Owen (1982) telah memberikan suatu kuantifikasi dari tingkat ketegaran pada uji t dua sampel dibawah asumsi variansi yang sama. Tingkat ketegaran diidentifikasi dengan mendapatkan daerah ketegaran dan hasil-hasil hanya dibatasi untuk uji dua sisi.

BAB II
DESKRIPSI TEORITIS

2.1. Pengertian Dasar

Uji Hipotesis adalah suatu proses dari percobaan untuk memutuskan kebenaran atau kesalahan pendugaan yang didasarkan pada pengamatan sampel .

Andaikan diberikan X_1, X_2, . . . ,X_n sampel random dari suatu populasi berdistribusi F_θ , $\theta \in \Theta$, dimana bentuk F_θ diketahui kecuali untuk parameter θ. Dengan pengasumsian seperti ini, berikut akan kami berikan beberapa definisi serta teorema yang berkaitan dengan konsep uji hipotesis.

Definisi 2.1.1. (V.K Rohatgi)

Suatu hipotesa parametrik adalah pernyataan tentang parameter θ yang tidak diketahui. H_0 ,$\theta \in \Theta_0 \subset \Theta$ disebut sebagai hipotesa null sedang pernyataan H_1 ,$\theta \in \Theta_1 = \Theta - \Theta_0$ disebut sebagai hipotesa alternatif.

Definisi 2.1.2. (V.K. Rohatgi)

Jika $\Theta_0(\Theta_1)$ hanya berisi satu anggota, $\Theta_0(\Theta_1)$ disebut sederhana, sebaliknya jika berisi lebih dari satu anggota disebut komposit. Jadi jika suatu hipotesis sederhana, maka distribusi probabilitas X menjadi tertentu di bawah hipotesis.

Definisi 2.1.3. (Bain E.)

Daerah kritis untuk suatu uji hipotesis adalah subset dari ruang sampel yang bersesuaian dengan penolakan hipotesa null.

Definisi 2.1.4. (Bain E)

Untuk hipotesa null sederhana, H_0 maka probabilitas penolakan kebenaran H_0 yaitu α = P(TI) disebut sebagai tingkat kepercayaan dari uji. Untuk hipotesa null komposit, H_0 maka ukuran uji (atau ukuran daerah kritis) adalah nilai maksimum probabilitas dari penolakan H_0 jika H_0 benar (nilai maksimum atas semua nilai parameter di bawah H_0).

Definisi 2.1.5. (Bain E.)

Fungsi Kuasa $\pi(\theta)$ dari uji H_0 adalah probabilitas penolakan H_0 jika nilai kebenaran dari parameter adalah θ.

Jadi untuk hipotesis sederhana H_0 , $\theta = \theta_0$ versus $H_1, \theta = \theta_1$, kita punya $\pi(\theta_0) = P(TI) = \alpha$ dan $\pi(\theta_1) = 1 - P(TII) = 1 - \beta$. Untuk hipotesis komposit, dikatakan H_0 , $\theta \in \Theta_0$ versus H_1 , $\theta \in \Theta - \Theta_0$ ukuran dari uji adalah $\alpha = \max_{\theta \in \Theta_0} \pi(\theta)$

.

Definisi 2.1.6. (E.J.Dudewich)

Suatu fungsi $f_x(x)$ dikatakan simetri disekitar μ jika $f_x(\mu + x) = f_x(\mu - x)$ untuk semua x.

Theorema 2.1.1.

Diberikan $X_1, X_2, \ldots, X_n$ sampel dari populasi dengan fungsi distribusi kumulatif F dan untuk $b = \frac{n-1}{n} S^2$ dimana S^2 adalah variansi sampel yaitu $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$

maka (i) $E(b) = \frac{n-1}{n} \mu_2$

(ii) $\text{Var}(b) = \left(n - 2 + \frac{1}{n}\right) \frac{\mu_4}{n^2} + (n-1)(3-n) \frac{\mu_2^2}{n^3}$

dengan μ_i momen pusat ke i dari X.

Akibat 2.1.1.

$$E(S^2) = \mu_2 \text{ dan } Var(S^2) = \frac{\mu_4}{4} + \frac{3-n}{n(n-1)}\mu_2^2$$

2.2. Distribusi Student t

Seperti kita ketahui bahwa bentuk fungsi distribusi dari student t dengan derajat bebas v adalah :

$$f(t,v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}$$

Berikut ini akan kami berikan beberapa teorema yang berkaitan dengan distribusi student t di atas dan akan sering kita temui pada bab-bab berikutnya.

Teorema 2.2.1.

Jika $X_1, X_2, \ldots, X_n$ notasi sampel random $N(\mu, \sigma^2)$ maka :

1. $\bar{X}$ dan suku-suku $X_i - \bar{X}$, $i = 1, 2, \ldots, n$ independen.
2. $\bar{X}$ dan S^2 adalah independen.
3. $(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$

Teorema 2.2.2.

Jika $Z \sim N(0,1)$ dan $V \sim \chi^2(v)$, dan jika Z dan V independen maka distribusi $T' = \frac{Z}{\sqrt{V / v}}$ adalah distribusi student t dengan derajat bebas v.

Theorema 2.2.3.

Jika $X_1, X_2, \ldots, X_n$ sampel random dari $N(\mu, \sigma^2)$ maka :

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

2.3. Notasi "O" Besar dan "o" Kecil.

Diberikan f dan g dua fungsi dan asumsikan bahwa $g(x) > 0$ untuk x yang cukup besar. Kita katakan bahwa $f(x)$ adalah berorder lebih besar dari $g(x)$ pada $x \to \infty$ dan kita tulis $f(x) = O(g(x))$ untuk $x \to \infty$ jika $\exists$ x_0 dan konstanta $C > 0 \ni |f(x)| < c\ g(x)\ \forall x \geq X_0$. Jadi $f(x) = O[g(x)]$ berarti bahwa $|f(x)|/g(x)$ terbatas untuk x besar. Kita tulis $f(x) = O(1)$ untuk menyatakan f terbatas untuk x besar.

Dan mudah untuk dilihat bahwa :

(a). Jika $f_1(x) = O[g_1(x)]$, $f_2(x) = O[g_2(x)]$ maka $f_1(x) + f_2(x) = O[g_1(x) + g_2(x)]$.

(b). Jika $\alpha > 0$ adalah konstanta, $f(x) = O[\alpha g(x)]$ maka $f(x) = O[g(x)]$

(c). Jika $f_1(x) = O[g_1(x)]$ dan $f_2(x) = O[g_2(x)]$ maka $f_1(x)f_2(x) = O[g_1(x)g_2(x)]$.

Diberikan f dan g, keduanya terdefinisi dan positif untuk x besar. Kita katakan f(x) adalah berorder lebih kecil daripada g(x) untuk x besar, dan ditulis :

$f(x) = o[g(x)]$ untuk $x \to \infty$ jika $\lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

Kita tulis $f(x) = o(1)$ untuk menyatakan $f(x) \to o$ untuk $x \to \infty$.

Mudah dilihat bahwa :

(a). $f(x) = o[g(x)] \Rightarrow f(x) = O[g(x)]$

(b). $f(x) = o[g(x)]$, $f_2(x) = o[g_2(x)] \Rightarrow f_1(x)f_2(x) = o[g_1(x)g_2(x)]$

BAB III

KETEGARAN STATISTIK T PADA INFERENSI UNTUK MEAN SATU POPULASI

3.1. Pendahuluan

Andaikan diberikan $Y_1, Y_2, \ldots, Y_n$ adalah sampel random berukuran n dari suatu populasi G dengan μ mean dari G yang tidak diketahui. Satu uji hipotesis yang penting $H_0 : \mu = \mu_0$ melawan satu dari tiga kemungkinan alternatif yaitu :

(a). $H_1 : \mu \neq \mu_0$

(b). $H_1 \ \mu > \mu_0$ atau

(c). $H_1 \ \mu < \mu_0$

Pertama diberikan : $T = \dfrac{\overline{Y} - \mu}{(S^2 / n)^{\frac{1}{2}}}$ . . . . . (3.1.1)

dan $TC = \dfrac{\overline{Y} - \mu_0}{(S^2 / n)}$ . . . . (3.1.2)

dimana $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$ dan $S^2 = \dfrac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n - 1}$ masing-masing adalah mean sampel dan variansi sampel. Pemakai biasa menunjuk uji mengenai μ diatas menggunakan statistik TC, yang disebut sebagai statistik T satu sampel. Jika diasumsikan bahwa :

A(i) Y_i-nya independen

A(ii) $G \equiv \Phi$

dimana Φ fungsi distribusi kumulatif normal standar, maka suatu hasil dalam teori statistik yang cukup baik dibuat oleh Student(1908) bahwa T~t(n-1) adalah distribusi student t dengan derajat kebebasan (n-1). Dengan asumsi di atas, dipunyai uji pada level α :

TT:Untuk $H_0:\mu=\mu_0$ vs $H_1:\mu \neq \mu_0$, tolak H_0 jika $|TC|>t_{\alpha/2}$ (3.1.3)

TU:Untuk $H_0:\mu=\mu_0$ vs $H_1:\mu >\mu_0$, tolak H_0 jika TC > t_α (3.1.4)

TL:Untuk $H_0:\mu=\mu_0$ vs $H_1:\mu < \mu_0$, tolak H_0 jika TC < t_α (3.1.5)

dengan t_α : notasi bagian atas 100α titik persentil dari distribusi t dengan derajat bebas (n-1).

Sifat-sifat distribusi T yang disajikan pada (3.1.1) jelas tergantung pada asumsi yang diberikan pada A(i) dan A(ii). Akan tetapi pemakai akan bertanya apakah asumsi tersebut dijumpai dan bagaimanakah pengaruh dari pelanggaran asumsi tersebut pada distribusi T.

Cicchitelli(1989) menunjukkan perluasan studi Monte Carlo untuk menyelidiki ketegaran dari statistik T satu sampel dengan distribusi populasi nonnormal. Dia menggunakan distribusi λ yang telah dikenalkan oleh Ramberg(1979) sebagai distribusi populasi. Kita akan mencoba mengkaji kembali hasil-hasil Cicchitelli tersebut dengan menggunakan ekspansi Edgeworth dari statistik T.

Untuk itu, akan kita perlukan beberapa definisi yang berkaitan dengan ekspansi Edgerworth seperti yang kita tulis berikut ini :

Definisi 3.1.1.

Untuk sebarang variabel random Y dengan $E|Y|^4 < \infty$ didefinisikan $S_k = \dfrac{\mu_3}{(\mu_2)^{3/2}}$ dan $K_u = \dfrac{\mu_4}{\mu_2^2}$. S_k adalah ukuran kemiringan dan K_u adalah ukuran kurtosis dari distribusi. $\mu_i, i = 2,3,4$ adalah Momen Pusat kedua, tiga, empat dari distribusi.

Definisi 3.1.2.

Untuk sebarang distribusi G, jika $S_k > (<) 0$ kita tunjuk G mempunyai kemiringan positif (negatif) dan jika $K_u > (<) 3$ kita tunjuk G mempunyai long (short) tailed.

Definisi 3.1.3.

$P_T(\alpha) = P_r\ (|TC| > t_{\alpha/2})$

$P_U(\alpha) = P_r\ (TC > t_\alpha)$

$P_L(\alpha) = P_r\ (TC < -t_\alpha)$

Dengan asumsi-asumsi A(i) dan A(ii), sesuai hasil student(1908) bahwa dibawah H_0, $P_T(\alpha) = P_U(\alpha) = P_L(\alpha) = \alpha$.

3.2. Ekspansi Edgeworth dari Statistik T dan Hasil-hasil yang terkait

Theorema 3.2.1 (Hall 1987)

Asumsikan $k \geq 1$, $E|Y|^{k+2} < \infty$ dan distribusi G nonsingular, maka

$F_{T_0}(y) = P_r(T_0 \leq y)$

$$= \Phi(y) + \sum_{i=1}^{k} n^{-i/2} P_i(y)\phi(y) + o(n^{-k/2}) \quad \ldots (3.2.1)$$

Konvergen uniform dalam y, Dimana P_i adalah polinomial berderajat (3i-1) yang tampak dalam ekspansi Edgeworth dari

$$T_0 = n^{1/2}(\overline{Y} - \mu)\left(n^{-1}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right)^{-1/2}$$

$\Phi(y)$ dan $\phi(y)$ masing-masing adalah fungsi distribusi kumulatif normal standar dan fungsi densitas normal standar. Koefesien P_i adalah fungsi-fungsi dari $\mu, \sigma^2, \mu_3, \ldots, \mu_{i+2}$ dimana μ_j mewakili momen central ke-j dari G.

Untuk P_1 dan P_2 didapat :

$P_1(y) = \frac{1}{6} S_k(2y^2 + 1)$ dan

$$P_2(y) = -y\left\{\frac{1}{18} S_k^2(y^2 + 3)(y^2 - 1) - \frac{1}{12}(K_u - 3)(y^2 - 3) + \frac{1}{4}(y^2 + 3)\right\}$$

dengan S_k dan K_u masing-masing ukuran kemiringan dan kurtosis dari G.

Terlihat bahwa :

$$T = \left(\frac{n - 1}{n}\right)^{\frac{1}{2}} T_0 \qquad \ldots . (3.2.2)$$

Dengan asumsi kefinitan momen ke-4 dan sifat nonsingularitas dari G kita akan mencari $E(T_0)$ dan $E(T)$ sebagai berikut :

$$E(T_0) = \int_{-\infty}^{\infty} y \; dF_{T_0}(y)$$

$$= \frac{n^{\frac{1}{2}}}{6} S_k \int_{-\infty}^{n} y \; d\left[(2y^2 + 1)\phi(y)\right]$$

$$- n^{-1} \int_{-\infty}^{\infty} y d\left[\frac{1}{18} S_k^2 y(y^2 + 3)(y^2 - 1)\phi(y)\right]$$

$$+ n^{-1} \int_{-\infty}^{\infty} y \; d\left[\left\{\frac{1}{12}\left(K_u - 3\right)y\left(y^2 - 3\right) - \frac{1}{4} y(y^2 + 3)\right\}\phi(y)\right]$$

$$+ o(n^{-1})$$

dengan mengingat $d\phi(y) = -y\phi(y)dy$, maka :

$$E(T_0) = \frac{-n^{-\frac{1}{2}}}{2} S_k + O(n^{-1}) \qquad \text{. . . . . .(3.2.4)}$$

Kita dapat lihat dari (3.2.1)-(3.2.4) bahwa selama ukuran sampel naik, apapun harga kemiringan populasi, S_k maka T_0 dan juga T menjadi lebih simetri disekitar nol. Ini bukan masalah yang luar biasa, karena telah diketahui bahwa $T \xrightarrow{d} Z$ dimana Z adalah variabel random normal standar. Jadi untuk ukuran sampel yang besar atau kemiringan populasi yang kecil T_0 dan juga T mendekati simetri disekitar nol.

Dengan cara yang sama untuk mendapatkan $E(T_0)$ kita dapat juga memperoleh :

$$E\left(T_0^2\right) = \int_{-\infty}^{\infty} y^2 dF_{T_0}(y)$$

$$= 1 + \frac{2}{n} S_k^2 + \frac{3}{n} + o(n^{-1}) \qquad . . . . (3.2.5)$$

Sehingga :

$$Var(T_0) = E(T_0^2) - \left\{E(T_0^2)\right\}^2$$

$$= 1 + \frac{7}{4n} S_k^2 + \frac{3}{n} + o(n^{-1}) \qquad . . . . (3.2.6)$$

$$E(T) = E\left(\left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_0\right)$$

$$= \left(\frac{n-1}{n}\right)^{\frac{1}{2}} E(T_0)$$

$$= (1 - \frac{1}{n})^{\frac{1}{2}} E(T_0)$$

$$= \left(1 - \frac{1}{2n} - \frac{1}{4n^2 2!} - \frac{3}{8n^3 3!} + \ldots\right)\left(-\frac{n^{-\frac{1}{2}}}{} S_k + o(n^{-1})\right)$$

$$= \frac{-n^{-\frac{1}{2}}}{2} S_k + o(n^{-1}) \qquad . . . . . (3.2.8)$$

$$E(T^2) = E\left(\frac{n-1}{n} T_0^2\right)$$

$$= \left(\frac{n-1}{n}\right) E(T_0^2)$$

$$= \left(1 - \frac{1}{n}\right)\left(1 + \frac{2}{n} S_k^2 + \frac{3}{n} + o(n^{-1})\right)$$

$$= 1 + \frac{2}{n} S_k^2 + \frac{3}{n} - \frac{1}{n} + o(n^{-1})$$

$$= 1 + \frac{2}{n} S_k^2 + \frac{2}{n} + o(n^{-1})$$

$$Var(T) = E(T^2) - (E(T))^2$$

$$= 1 + \frac{2}{n} + \frac{7}{4n} S_k^2 + o(n^{-1}) \quad \ldots . (3.2.9)$$

Sekarang perhatikan ekspansi deret taylor dari $F_{T_0}(y)$ disekitar y = 0 didapat dari (3.2.1) yaitu :

$$F_{T_0}(y) = \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_k + \frac{y}{(2\pi)^{\frac{1}{2}}}\left(1 + \frac{S_k^2}{6n} - \frac{K_u}{4n}\right)$$

$$+ \frac{n^{-\frac{1}{2}}}{4(2\pi)^{\frac{1}{2}}} y^2 S_k + o(n^{-1}) \quad \ldots . (3.2.10)$$

dan menggunakan (3.2.2) kita dapatkan :

$$
\begin{aligned}
F_T(y) &= P_r(T \le y) \\
&= P_r\left(\left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_0 \le y\right) \\
&= P_r\left(T_0 \le \left(\frac{n-1}{n}\right)^{\frac{1}{2}} y\right) \\
&= F_{T_0}\left(\left(\frac{n-1}{n}\right)^{\frac{1}{2}} y\right)
\end{aligned}
$$

$$
= \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K + \frac{\left(\frac{n}{n-1}\right)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} y\left(1 + \frac{S_K^{\ 2}}{6n} - \frac{K_u}{4n}\right)
$$

$$
+ \frac{n^{-\frac{1}{2}}}{4(2\pi)^{\frac{1}{2}}} \left(\frac{n}{n-1}\right) y^2 S_K + o(n^{-1})
$$

$$
= \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K + \frac{y}{(2\pi)^{\frac{1}{2}}} \left(1 + \frac{S_K^{\ 2}}{6n} - \frac{K_u}{4n}\right)\left(1 + \frac{1}{2n} + \frac{3}{4n^2 2!} + \ldots\right)
$$

$$
+ \frac{y^2 S_K}{4(2\pi)^{\frac{1}{2}}} n^{-\frac{1}{2}}\left(1 + \frac{1}{n} + \frac{1}{n^2 2!} + \frac{1}{n^3 3!} + \ldots\right) + o(n^{-1})
$$

$$
= \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K + \frac{y}{(2\pi)^{\frac{1}{2}}} \left(1 + \frac{1}{2n} + \frac{S_K^{\ 2}}{6n} - \frac{K_u}{4n}\right)
$$

$$
+ \frac{y^2 S_K}{4(2\pi)^{\frac{1}{2}}} n^{-\frac{1}{2}} + o(n^{-1})
$$

Untuk menemukan median dari distribusi T pecahkan persamaan $F_T(y) = \frac{1}{2}$. Suatu pendekatan solusi persamaan ini adalah $y = -\frac{n^{-\frac{1}{2}}}{6} S_K$ sebab :

$$F_T\left(\frac{-n^{-\frac{1}{2}}}{6} S_K\right) = \frac{1}{2} + \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K$$

$$- \frac{n^{-\frac{1}{2}}}{6(2\pi)^{\frac{1}{2}}} S_K\left(1 + \frac{1}{2n} + \frac{S_K^2}{6n} - \frac{K_u}{4n}\right)$$

$$+ \frac{n^{-\frac{3}{2}}}{144(2\pi)^{\frac{1}{2}}} S_K^3 + o(n^{-1})$$

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

dengan pengabaian suku $o(n^{-1})$ maka :

$$\text{Median dari } T = \frac{-n^{-\frac{1}{2}}}{6} S_K \qquad \ldots . (3.2.11)$$

$$\text{Mean dari } T = \frac{-n^{-\frac{1}{2}}}{2} S_K \qquad \ldots . (3.2.12)$$

Sehingga terlihat bahwa jika distribusi populasi miring kekanan yaitu $S_k > 0$ maka : $\frac{-n^{-\frac{1}{2}}}{2} S_K < \frac{-n^{-\frac{1}{2}}}{6} S_K < 0$.

Jadi distribusi T miring kekiri. Sebaliknya adalah benar untuk $S_k < 0$. Hasil tersebut sesuai dengan studi awal dan penemuan Cicchitelli(1989). Sehingga kita dapat putuskan bahwa, jika distribusi populasi sangat minim dan ukuran sampel kecil maka kemiringan distribusi T tidak akan didekati dengan distribusi student-t, kecuali telah dibuat suatu koreksi kemiringan seperti yang dilakukan oleh Johnson(1978).

Dengan pengabaian suku $o(n^{-1})$ maka mean dari T < median dari T < 0 dan variansi dari T = $1 + \frac{2}{n} + \frac{7}{4n} S_k^2 \Big(\geq 1 + \frac{2}{n}$ variansi dari suatu variabel random yang mempunyai distribusi student-t dengan derajat bebas $n - 1 \Big)$. Jadi untuk kemiringan populasi G positif maka $P_u(\alpha) < \alpha$ dan $P_L(\alpha) > \alpha$ seperti yang diamati oleh Cicchitelli(1989).

3.3. Modifikasi dari Statistik T (Johnson 1978)

Andaikan diberikan $Y_1, Y_2, \ldots, Y_n$ adalah sampel random dari suatu distribusi G dengan mean μ, variansi σ^2 dan μ_3, μ_4, ... masing-masing adalah momen central ketiga, keempat, ... dari G. Untuk sebarang variabel random Y dengan distribusi G, bentuk ekspansi Cornish-Fisher diberikan dengan :

$$CF(Y) = \mu + \sigma\xi + \frac{\mu_3}{6\sigma^2}(\xi^2 - 1) + \ldots \qquad (3.3.1)$$

dengan ξ adalah variabel normal standar dan

$$CF(\bar{Y}) = \mu + \frac{\sigma}{n^{\frac{1}{2}}}\xi + \frac{\mu^3}{6n\sigma^2}(\xi^2 - 1) + O\left(n^{-\frac{3}{2}}\right) \qquad \ldots (3.2.2)$$

Dari ekspansi diatas dapat dicatat bahwa kemiringan populasi μ_3 adalah koefisien dari suku (ξ^2-1) serta tampak dalam suku-suku lain dari tetapi dengan order yang lebih kecil. Kunci dalam mendapatkan modifikasi variabel T dalam pendekatan Johnson adalah mengeliminasi suku yang melibatkan μ_3 dalam variabel T pembangun yang diberikan pada bagian bawah berikut :

Diberikan variabel T pembangun

$$TJ = \frac{(\bar{Y} - \mu) + \lambda + \gamma\left\{(\bar{Y} - \mu)^2 - \frac{\sigma^2}{n}\right\}}{\left\{\frac{S^2}{n}\right\}^{\frac{1}{2}}} \qquad \ldots (3.2.3)$$

λ dalam TJ dipilih sehingga suku konstan dalam ekspansi Cornish-Fisher dari TJ berjumlah nol sehingga bias order yang lebih rendah tereleminasi γ dipilih sehingga koefesien dari suku ξ^2 dalam ekspansi Cornish-Fisher dari TJ adalah nol (dengan demikian mengeliminasi pangaruh kemiringan order yang lebih rendah).Gantikan nilai $\bar{Y}$ dan S^2 dalam (3.2.3) dengan ekspansinya, maka dengan pengabaian suku $O(n^{-1})$, ekspansi Cornish-Fisher dari TJ adalah :

$$CF(TJ) = \xi + \left(\frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{3\sqrt{n\sigma^3}}\right)\xi^2$$

$$+ \frac{\lambda\sqrt{n}}{\sigma} - \frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{6\sqrt{n\sigma^3}} - \frac{\mu_4 - \sigma^4}{2\sqrt{n\sigma^2}}\xi\xi^*$$

dengan ξ dan ξ^* variabel random normal standar yang independen.

Dengan menyeleksi γ dan λ sehingga koefesien dari ξ^2 adalah ncl juga suku konstan berjumlah nol, ekspresi hasil akan mengurangi bias, didapat :

$$\frac{\mu_3}{6\sqrt{n\sigma^3}} + \frac{\gamma\sigma}{\sqrt{n}} - \frac{\mu_3}{2\sqrt{n\sigma^3}} = 0$$

$$-\frac{\mu_3}{3\sqrt{n\sigma^3}} + \frac{\gamma\sigma}{\sqrt{n}} = 0$$

$$\gamma = \frac{\mu_3}{3\sigma^4}$$

dan $\dfrac{\gamma\sqrt{n}}{\sigma} - \dfrac{\mu_3}{6\sqrt{n\sigma^3}} - \dfrac{\gamma\sigma}{\sqrt{n}} = 0$

$$\frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{6\sqrt{n\sigma^3}} - \frac{\mu_3}{3\sqrt{n\sigma^3}} = 0$$

$$\frac{\lambda\sqrt{n}}{\sigma} - \frac{\mu_3}{2\sqrt{n\sigma^3}} = 0$$

$$\lambda = \frac{\mu_3}{2n\sigma^2}$$

Jadi $CF(TJ) = \xi - \frac{1}{2}\left(\frac{\mu_4 - \sigma^4}{n\sigma^4}\right)^{\frac{1}{2}} \xi\xi^* + o(n^{-1})$

$$= \xi - \frac{1}{2\sqrt{n}}\left(K_u - 1\right)^{\frac{1}{2}} \xi\xi^* + o(n^{-1}) \quad ...(3.2.4)$$

dan $$TJ = \left\{(\bar{Y} - \mu) + \frac{\mu_3}{6\sigma^2 n} + \frac{\mu_3}{3\sigma^4}(\bar{Y} - \mu)^2\right\}\left(\frac{S^2}{n}\right)^{-\frac{1}{2}} \quad ....(3.2.5)$$

Terlihat bahwa TJ yang diberikan oleh (3.2.5) tidak dapat dihitung dengan H : $\mu = \mu_0$, karena μ_3 dan σ^2 biasanya tidak diketahui. Johnson menyarankan mengganti μ_3 dan σ^2 dengan estimasi sampel $\hat{\mu}_3 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^3}{n}$ dan variansi sampel S^2. Ekspansi Cornish-Fisher masih (3.2.4) Untuk contoh penggunaannya uji hipotesis $H_0 = \mu = \mu_0$ melawan satu dari tiga alternatif kemungkinan, maka dipunyai uji level α sebagai berikut :

TT : untuk $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$,

tolak H_0 jika $|TJH| > t_\alpha/2$ ....(3.2.6)

TU : untuk $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$,

tolak H_0 jika $TJH > t_\alpha$ .... (3.2.7)

TL : untuk $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$,

tolak H_0 jika $TJH < -t_\alpha$ .... (3.2.8)

$$\text{dimana TJH} = \frac{\left(\overline{Y} - \mu_0\right) + \hat{\mu}_3 \;/\; (6s^2n) + \left(\hat{\mu}_3 \;/\; (3s^4)\right)\left(\overline{Y} - \mu\right)^2}{\left(\dfrac{s^n}{n}\right)^{\frac{1}{2}}} \qquad (3.2.9)$$

dan t_α notasi bagian atas 100α titik persentil dari distribusi t dengan derajat bebas (n-1).

BAB IV

KETEGARAN STATISTIK t PADA INFERENSI UNTUK MEAN DUA POPULASI INDEPENDEN

4.1. Pendahuluan

Andaikan diberikan :

$$\left.\begin{aligned} H_0&: \mu_1 = \mu_2 \text{ dan} \\ H_1&: \mu_1 > \mu_2 \end{aligned}\right\} \qquad \ldots \quad (4.1.1)$$

dimana μ_i adalah mean dari variabel random normal $X_i \sim N(\mu_i, \sigma_i^2)$ $i = 1,2$. Lebih lanjut diberikan dua sampel independen, katakanlah X_{1j} , $j = 1,2, \ldots ,n_1$

$X_{2\ell}$, $\ell = 1,2, \ldots ,n_2$

sering digunakan statistik

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\left\{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{\frac{1}{2}}} \qquad \ldots \ldots (4.1.2)$$

dengan $S^2 = \dfrac{\left\{\sum_j \left(X_{1j} - \overline{X}_1\right)^2 + \sum_j \left(X_{2\ell} - \overline{X}_2\right)^2\right\}}{\left(n_1 + n_2 - 2\right)}$

dan $\bar{X} = \frac{\sum_k X_{ik}}{n_i}$, i = 1,2

Jika $\sigma_1^2 = \sigma_2^2$ maka statistik $t \sim t(n_1 + n_2 - 2)$. Dan dipunyai uji level α untuk uji yang diberikan oleh (4.1.2). Tolak H_0 jika $t \geq t_\alpha$, t_α adalah bagian atas 100α titik persentil dari variabel $t(n_1 + n_2 - 2)$.

Posten, Yeh dan Owen (1982) memberikan suatu kualifikasi dari tingkat ketegaran uji t dua sampel dibawah asumsi variansi yang sama. Tingkat ketegaran diidentifikasi dengan mendapatkan daerah ketegaran dan hasil-hasil dibatasi untuk uji dua sisi. Lewat prosedur ini, hasil tersebut memberikan suatu ukuran yang jelas dari kekuatan ketegaran untuk variansi yang heterogen pada masalah ukuran sampel yang sama atau mendekati sama. Akan diperluas hasi-hasil yang didapat,untuk uji t dua sampel versi satu sisi.

Untuk memberikan informasi kuantitatif ketegaran dari uji t dengan keheterogenan variansi, dibutuhkan tiga definisi berikut.

Andaikan : $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$

$$\text{dan} \quad \left.\begin{aligned} \alpha(\lambda) &= P_r\left(\text{Tolak } H_0 \,/\, \lambda, H_0\right) \\ &= P_r\left(t \geq t_\alpha \,/\, \lambda, H_0\right) \end{aligned}\right\} \qquad \text{. . . . (4.1.3)}$$

Definisi 4.1.1.

Diberikan $A = \{\lambda : \lambda > 0\}$

$R(\varepsilon)$ adalah daerah ketegaran pada level α jika $R(\varepsilon) \subset A \ni \forall \lambda \in R(\varepsilon)$, $|\alpha(\lambda) - \alpha(1)| \leq \varepsilon$

Definisi 4.1.2.

Union dari semua daerah ketegaran pada level α disebut daerah maksimal ketegaran pada level α dan diberi simbol $R_m(\varepsilon)$.

Definisi 4.1.3.

Uji t adalah tegar secara total pada level α jika $A = \{\lambda : \lambda > 0\}$ adalah daerah ketegaran pada level α.

Dari definisi (4.1.2) dan (4.1.3), daerah maksimal ketegaran pada level α adalah :

$$R_{\alpha}(\varepsilon) = \{\lambda \;/\; |\alpha(\lambda) - \alpha(1)| \le \varepsilon\} \qquad \ldots\ldots(4.1.4)$$

dengan $\qquad \alpha(\lambda) = \int_{t_\alpha}^{\infty} P_\lambda(t)\, dt$

dimana $P_\lambda(t)$ adalah densitas dari t jika H_0 benar.

4.2. Kemonotonan $\alpha(\lambda)$

Sifat kemonotonan $\alpha(\lambda)$ penting dalam mengevaluasi atau menghitung daerah maksimal ketegaran (4.1.4). Berikut diberikan teorema-teorema yang berkaitan dengan sifat tersebut.

Theorema 4.2.1.

Jika $\alpha(\lambda, n_1, n_2)$ menyatakan tingkat kepercayaan ekor atas uji t dua sampel dengan ukuran sampel n_1 dan n_2 maka $\alpha(\lambda : n_1, n_2) = \alpha\left(\frac{1}{\lambda} : n_2, n_1\right)$. Maka untuk membuktikan teorema ini kita ikuti uraian dibawah ini :

Masalah perhitungan tingkat ketegaran dan daerah maksimal ketegaran didasarkan pada peritungan $\alpha(\lambda)$ dalam (4.1.3) dan (4.1.4). Dari Yeh(1981), probabilitas ini dapat dinyatakan dalam suku-suku integral dari probabilitas pembobot student, hubungan ini adalah :

$$\alpha(\lambda; n_1, n_2) = \int_0^1 b(X; n_1, n_2) P_r\left(T^* \geq m(\lambda, X; n_1, n_2) t_\alpha\right) dx \qquad (4.2.1)$$

dimana :

$N = n_1 + n_2$, t_α adalah nilai kritis dari uji t satu sisi.

$$r = \frac{n_1}{n_2}$$

$$m(\lambda, X; n_1, n_2) = \left\{\frac{(1 + r)}{\lambda + r}\right\}^{\frac{1}{2}} \{\lambda + (1 - \lambda)x\}^{\frac{1}{2}} \qquad \ldots (4.2.2)$$

$$b(x; n_1, n_2) = \frac{\Gamma\left(\frac{N - 2}{2}\right)}{\Gamma\left(\frac{n_1 - 1}{2}\right)\Gamma\left(\frac{n_2 - 1}{2}\right)} x^{\frac{(n_2 - 3)}{2}} (1 - x)^{\frac{(n_1 - 3)}{2}}$$

T^* variabel t pusat dengan derajat kebebasan N - 2

Akibat 1.2.1

Untuk $n_1 = n_2 = n$ maka $\alpha(\lambda) = \alpha\left(\frac{1}{\lambda}\right)$

Theorema 4.2.2.

Untuk ukuran sampel sama, $\alpha(\lambda)$ monoton turun untuk $\lambda \in (0,1)$ dan monoton naik untuk $\lambda \in (1,\infty)$.

Theorema 4.2.3.

Untuk ukuran sampel sama, daerah maksimal ketegaran tingkat ε, $R_m(\varepsilon)$ mempunyai bentuk $\left[c, \frac{1}{c}\right]$ dimana $c < 1$ yang didefinisikan dari $\alpha(c) - \alpha(1) = \varepsilon$.

4.3. Daerah Ketegaran

Ukuran sampel yang dipakai dalam studi ini (untuk $n_1 \neq n_2$) adalah $n_1 < n_2$ dengan tingkat kepercayaan nominal

$\alpha(\lambda)$ yang merupakan fungsi monoton naik dalam λ pada $\lambda \in (0,\infty)$.

Daerah Maksimal ketegaran akan diberikan untuk ukuran sampel yang sama ataupun tak sama. Akan tetapi, pertama kita akan pertimbangkan kemungkinan nilai-nilai ε yang cukup kecil sehingga uji t satu sisi akan tegar secara total.

Untuk ukuran sampel sama ($n_1 = n_2 = n$) dengan sifat kemonotonan dari $\alpha(\lambda)$ berakibat bahwa untuk semua nilai dari ε yang melebihi atau sama dengan $|\alpha(0) - \alpha(1)|$ maka uji t akan tegar secara total pada level ε.

Lebih lanjut perlu untuk menghitung $\alpha(0)$ guna mengidentifikasi nilai terkecil dari ε, untuk mana uji akan tegar secara total. Sedang untuk ukuran sampel tidak sama perlu juga untuk menghitung $|\alpha(\infty) - \alpha(1)|$ dan pilih nilai terbesar dari keduanya sebagai nilai terkecil dari ε untuk mana uji akan tegar secara total.

Tabel (4.3.1) tersedia ukuran sampel $n_1 = n_2 = n, \alpha = 0{,}05$ dan $\alpha = 0{,}01$ nilai-nilai minimal ε untuk mana uji t satu sisi tegar secara total. Dari tabel 4.3.1 terlihat dengan jelas bahwa untuk ukuran sampel $n = 15$ dan $\alpha = 0,05$, uji t tegar secara total pada level 0,0055. Bahwa tak ada masalah seberapa besar perbedaan σ_1^2 dan σ_2^2 . Kebenaran tingkat kepercayaan adalah sampai 0,0055 dari level yang diinginkan sebesar 0,05. Lebih lanjut selisih dari α disini terjadi hanya jika σ_1^2 sangat berbeda dari σ_2^2 . Untuk nilai n yang lebih besar, nilai ε minimum yang menghasilkan tegar secara total akan lebih kecil. Tabel 4.3.1 ini juga dapat digunakan untuk menentukan derajat uji t satu sisi (dengan ukuran sampel sama dibawah pengabaian asumsi kehomogenan variansi).

Untuk ukuran sampel tidak sama, nilai terkecil dari ε yang memberikan ketegaran secara total untuk uji t satu sisi, ditentukan dengan evaluasi $\alpha(1)-\alpha(0)$ dan $\alpha(\infty)-(1)$ dimana nilai ε adalah nilai terbesar diantara keduanya. Adapun nilai-nilai dari $\alpha(0)$ dan $\alpha(\infty)$ diberikan dalam tabel 4.3.2 untuk ukuran sampel yang tidak sama, tetapi untuk ukuran sampel yang sama juga diberikan sebagai pembanding.

Karena tingkat kepercayaan nominal $\alpha(1)$ diketahui maka nilai ε untuk mana uji akan tegar secara total siap didapat dari tabel (4.3.2).

Selanjutnya daerah maksimal ketegaran diberikan didalam tabel (4.3.3) dan (4.3.4) masing-masing untuk $\alpha(1) = 0,05$, $\varepsilon = 0,005$, $0,03$, $0,02$ dan $\alpha(1) = 0,01$, $\varepsilon = 0,02$, $0,01$, $0,005$ pada beberapa nilai n_1 dan n_2.

4.4. Prosedur perhitungan

Dalam masalah perhitungan tingkat ketegaran dan daerah maksimal ketegaran seperti yang ditulis pada bagian awal, didasarkan pada perhitungan $\alpha(\lambda)$ dalam (4.2.1) yaitu :

$$\alpha(\lambda; n_1, n_2) = \int_0^1 b(x; n_1, n_2)\, P_r\left(T^* \geq m(\lambda, x; n_1, n_2)t_\alpha\right)dx$$

dimana untuk setiap nilai tertentu dari x, probabilitas T^* dapat dihitung dengan menggunakan hubungan antara variabel student t dengan beta untuk nilai t yang positif.

$$P_r\left(T^* < t\right) = \int_{-\infty}^{t} \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)} \frac{1}{\sqrt{(N-2)\pi}} \left(1 + \frac{x^2}{N-2}\right)^{-\left(\frac{N-1+1}{2}\right)} dx$$

$$P\left(T^* < t\right) = \frac{1}{2} + \int_{0}^{t} \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{1}{2}\right)\sqrt{n-2}} \left(1 + \frac{x^2}{N-2}\right)^{-\left(\frac{N-1}{2}\right)} dx$$

$$= \frac{1}{2} + \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_{0}^{t} \frac{1}{\sqrt{N-2}} \left(1 + \frac{x^2}{N-2}\right)^{-\left(\frac{N-1}{2}\right)} dx$$

dimana suku dalam integral pada persamaan terakhir dapat ditulis dengan : $\frac{1}{2}\int_{0}^{y} s^{-\frac{1}{2}}(1 - s)^{\frac{N-4}{2}} ds$

dimana $y = \frac{t^2}{N-2+t^2}$ menggunakan subsitusi $\frac{x^2}{N-2+x^2} = s$

$$P\left(T^* < t\right) = \frac{1}{2} + \frac{\Gamma\left(\frac{N-2+1}{2}\right)}{2\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_{0}^{y} s^{-\frac{1}{2}}(1 - s)^{\frac{N-4}{2}} ds$$

sehingga: $= \frac{1}{2} + \frac{1}{2} I_y\left(\frac{1}{2}, \frac{f}{2}\right)$

$$= \frac{1}{2}\left(1 + I_y\left(\frac{1}{2}, \frac{f}{2}\right)\right)$$

dengan : $f = N - 2$ dan $I_y\left(\frac{1}{2}, \frac{f}{2}\right) = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_{0}^{y} s^{-\frac{1}{2}}(1 - s)^{\frac{f-2}{2}} ds$

DAFTAR PUSTAKA

Benjamini,Y.(1983),"Is T-Test Really Conservative when the Parent Distribution is Long Tailed", *J.Am. Statist. Assoc.,78,645-654.*

Dudewicz,E.J.& Mishra, S.N. (1988) , *Modern Mathematical Statistics,*John Wiley & Sons.

Engelhardt,B.(1992),*Introduction to Probability and Mathematical,*PWS-KENT Publishing Company.

Hall, P. (1992) ,*The Bootstrap and Edgeworth Expansion,* Spinger- Verlag Inc.NewYork.

Johnson , N.J. (1978), "Modified t-test and Confidence Interval for Asymmetrical Population", *J.Am. Statist.Assoc., 73,536-544.*

Lehmann,E.L.(1983),*Theory of Point Estimation,*John Wiley & Sons.

Nanayakkara, N.(1992), "On the Robustness and Johnson's Modification of the One Sample T - Statistic", *Commun. Statist.-Theory Meth.,21(11),3079-3096.*

Posten,H.O.(1992),"Robustness of the Two-Sample T-test Under Violations of the Homogeneity of Variance Assumption" , *Commun.Statist.- Theory Meth .,21(8), 2169-2184.*

Posten H.O.,Yeh H.C.&Owen D.B.(1982),"Robustness of the two sample t-test under violations of the homogeneity of variance assumption",*Commun.Statist.-TheoryMeth.,11(2),109-126.*

Rohatgi,V.K.(1976),*An Introduction to Probability Theory and Mathematical Statistics,*John Wiley & Sons.