<table>
<tr><td colspan="3">Suplemen Responsi
ANALISIS DATA KATEGORIK (STK351)
Departemen Statistika – FMIPA IPB</td><td>Pertemuan
1</td></tr>
<tr><td>Pokok Bahasan</td><td>Sub Pokok Bahasan</td><td>Referensi</td><td>Waktu</td></tr>
<tr><td>Metode Nonparametrik</td><td>▪ Skala Pengukuran
▪ Metode Nonparameterik</td><td rowspan="2">Applied Nonparametric Statistics
Daniel (1990)</td><td rowspan="2">2 x @60 Menit</td></tr>
<tr><td>Uji Hipotesis Satu Populasi</td><td>▪ Uji Tanda
▪ Uji Peringkat Bertanda Wilcoxon
▪ Uji Binomial</td></tr>
</table>

Skala Pengukuran

Berdasarkan skala pengukuran, data digolongkan dalam empat tipe, yaitu data nominal, ordinal, interval dan rasio. Data nominal dan ordinal merupakan data kategorik, sedangkan data interval dan rasio adalah data numerik.

Skala nominal

Skala nominal merupakan skala peng ukuran paling rendah. Angka-angka yang tersaji dalam skala nominal hanya sebagai penggolongan agar dapat dibedakan saja dan tidak mengukur besaran. Sebagai contoh, dalam pengkodean jenis kelamin; kode 1 untuk laki-laki dan 0 untuk perempuan hanya untuk mem bedakan antara laki-laki dan perempuan saja dan tidak berarti nilai laki -laki lebih tinggi daripada perempuan. Jika pun dibalik, kode 1 untuk perempuan dan 0 untuk laki -laki, makna tidak akan berubah.

Skala ordinal

Skala ordinal hampir sama dengan skala no minal. Hanya saja, selain untuk membedakan, skala ordinal sudah mempunyai urutan tingkatan. Dalam skala ordinal, angka 1 memiliki nilai lebih tinggi daripada 0. Meskipun demikian, jarak antara 0 dan 1 tidak bisa dijelaskan. Contoh skala ordinal adalah ting kat kepuasan (misalnya dalam *important and performance analysis*); sangat puas (5), puas (4), cukup puas (3), tidak puas (2), dan sangat tidak puas (1). Angka -angka ini memiliki makna bahwa 2 lebih besar dari 1, 3 lebih besar dari 2 dan 1, dan seterusnya. T etapi, jarak atau selisih antara antara 1 dan 2, 2 dan 3, dan lainnya, tidak mempunyai makna apapun.

Skala interval

Pada skala interval (atau skala selang), angka-angka yang disajikan menunjukkan tingkatan dan angka yang berurutan memiliki interval (jarak) yang sama. Ciri utama skala interval adalah tidak mempunyai titik dasar (nol) mutlak sehingga operasi perbandingan tidak dapat dilakukan . Contoh skala interval adalah pada pengukuran suhu dengan standar derajat Celcius (^{0}C). Suhu 40^0 dan 20^0 memiliki selisih yang sama dengan suhu 80^0 dan 60^0 yaitu 20^0, akan tetapi suhu 40^0 tidak berarti 2 kali lebih panas dari 20^0. Demikian juga bahwa suhu 0^0 tidak berarti bahwa tidak mempunyai panas .

Skala rasio

Skala rasio merupakan skala pengukuran tertinggi. Selain dapat membedakan, menunjukkan tingkatan dan memiliki interval yang sama antar dua nilai yang berurutan, skala rasio dapat dibandingkan karena mempunyai nilai dasar (nol) multak. Contohnya adalah tinggi badan, harga barang, jumlah produksi dan lain -lain.

Skema keempat skala data ditunjukan dalam tabel berikut :

Data	Skala	Dapat dibedakan	Ada urutan tingkatan	Memiliki interval sama	Dapat dibandingkan
Kategorik	Nominal	√			
	Ordinal	√	√		
Numerik	Interval	√	√	√	
	Rasio	√	√	√	√

Mengenal jenis data penting dalam statistika karena sangat berhubungan dengan analisis statistika yang akan digunakan. Beberapa analisis statistika mensyaratkan skala data tertentu. Jika skala data tidak relevan dengan analisis yang digunakan, hasil yang diperoleh akan tidak sah.

Metode Nonparametrik

Dalam inferensia statistika, dikenal dengan dua metode yaitu metode parametrik dan metode nonparametrik. Perbedaan mendasar antara keduanya terletak pada penggunaan asumsi mengenai populasi. Dalam melakukan pendugaan parameter, inferensia atau penarikan kesimpulan mengena i populasi, metode parametrik memberikan asumsi bahwa populasi menyebar menurut sebaran tertentu . Sebagai contoh, analisis ragam (ANOVA) memberikan asumsi bahwa contoh berasal dari populasi yang menyebar normal dengan ragam yang homogen. Jika asumsi ini ti dak terpenuhi, kesimpulan yang diperoleh menjadi tidak valid.

Jika asumsi yang mendasari metode parametrik tidak terpenuhi, kita dapat menggunakan metode inferensia lain yang tidak terlalu bergantung pada asumsi baku. Metode nonparametrik pada banyak kasus dapat digunakan untuk keperluan ini. Metode nonparametrik tidak membutuhkan asumsi mengenai sebaran data populasi. Karena itu, metode ini sering disebut *distribution -free method*. Statistika nonparametrik mencakup pemodelan statistika, pengujian hipotesis dan inferensia atau penarikan kesimpulan tentang populasi. Meskipun demikian, jika asumsi yang mendasari metode statistika parametrik dapat dipenuhi, penggunaan statistika nonparametrik tidak begitu disarankan.

Kelebihan metode nonparametrik antara lain : (1) asumsi yang diperlukan sangat minimum (2) pada beberapa prosedur, perhitungan dapat dilakukan dengan mudah dan cepat, (3) konsep dan metode lebih mudah dipahami dan (4) dapat diterapkan pada data dengan skala yang lebih rendah. Sedangkan kekurangan dari metode nonparametrik antara lain : (1) karena sangat sederhana dan cepat, perhitungan dalam prosedur nonparametrik terkadang dapat 'membuang' informasi dari data, (2) meskipun perhitungan sangat sederhana, prosedur nonparametrik akan sangat membosankan terutama ketika data yang digunakan berukuran besar.

Berikut beberapa contoh metode statistika parametrik dan nonparametik dalam pengujian hipotesis statistika.

Pengujian	***Metode***	
	Parametrik	***Nonparametrik***
Uji nilai tengah satu populasi	Uji-T	Uji tanda
Uji perbedaan nilai tengah dua populasi yang saling bebas	Uji-T	Uji Mann-Whitney
Uji perbedaan nilai tengah lebih dari dua populasi	Uji-F (ANOVA)	Uji Kruskal-Wallis
Uji korelasi antar dua variabel	Korelasi Pearson	Korelasi Spearman

Uji Hipotesis pada Contoh Tunggal

Metode nonparametrik pada contoh tunggal yang akan dipelajari meliputi pengujian hipotesis mengenai median dan proporsi populasi. Pengujian hipotesis mengenai median dapat dilakukan dengan dua uji, yaitu uji tanda *(sign test)* dan uji peringkat bertanda Wilcoxon *(Wilcoxon signed-rank test).* Sedangkan pengujian hipotesis mengenai proporsi populasi dapat dilakukan dengan uji binomial *(binomial test).*

Uji Tanda

Uji tanda *(sign test)* atau dalam kasus contoh tunggal secara spesifik disebut uji tanda satu contoh *(one-sample sign test)* merupakan pionir dari seluruh prosedur nonparametrik. Disebut uji tanda karena data diubah menjadi serangkaian tanda ‘plus’ (+) dan ‘minus’ (-).

Asumsi

a. Contoh acak saling bebas dengan median (M) tidak diketahui
b. Data diukur setidaknya dalam skala ordinal
c. Peubah yang diamati kontinu

Hipotesis

a. (Dua arah) : $H_0 : M = M_0$ *vs.* $H_1 : M \neq M_0$
b. (Satu arah) : $H_0 : M \leq M_0$ *vs.* $H_1 : M > M_0$
c. (Satu arah) : $H_0 : M \geq M_0$ *vs.* $H_1 : M < M_0$

Statistik Uji

Hitung selisih nilai data dan median untuk setiap pengamatan , $X_i - M_0$. Jika hasilnya 0, abaikan pengamatan tersebut. Hitung banyaknya nilai yang bertanda ‘minus’ (S-) dan banyaknya nilai yang bertanda ‘plus’ (S+). Statistik uji tanda satu contoh (S) adalah :

a. (Hipotesis a) : $S = S' = \min(S-, S+)$
b. (Hipotesis b) : $S = S-$
c. (Hipotesis c) : $S = S+$

Kaidah Keputusan

a. (Hipotesis a) : Tolak H_0 jika $P(x \le S' \mid b(n,0.5)) \le \alpha/2$
b. (Hipotesis b) : Tolak H_0 jika $P(x \le S- \mid b(n,0.5)) \le \alpha$
c. (Hipotesis c) : Tolak H_0 jika $P(x \le S+ \mid b(\text{n},0.5)) \le \alpha$

Catatan Untuk contoh berukuran besar (yaitu sama dengan atau lebih besar dari 12) dapat didekati dengan sebaran normal menggunakan faktor koreksi 0.5 dengan rumus :

$$z = \frac{(k + 0.5) - 0.5n}{0.5\sqrt{n}}$$

Contoh :

Di bawah ini adalah waktu belajar mandiri dari tujuh mahasiswa. Ujilah apakah benar bahwa mahasiswa pada umumnyan menyediakan waktu kurang dari dua jam untuk belajar mandiri! Gunakan taraf nyata 5%.

Mahasiswa ke-	1	2	3	4	5	6	7
Lama belajar mandiri (jam)	1.5	2.1	1.7	1.8	2.2	1.1	0.8

Hipotesis : H_0 : $M \ge 2$
H_1 : $M < 2$

Statistik Uji : Sesuai dengan hipotesis di atas, statistik uji yang akan digunakan adalah S+, yaitu banyaknya selisih $X_i - M_0$ yang lebih besar dari 0. Dari data di atas, ada dua pengamatan yang selisihnya lebih besar dari 0, yaitu mahasiswa ke -2 dan ke-5, sehingga S+ = 2.

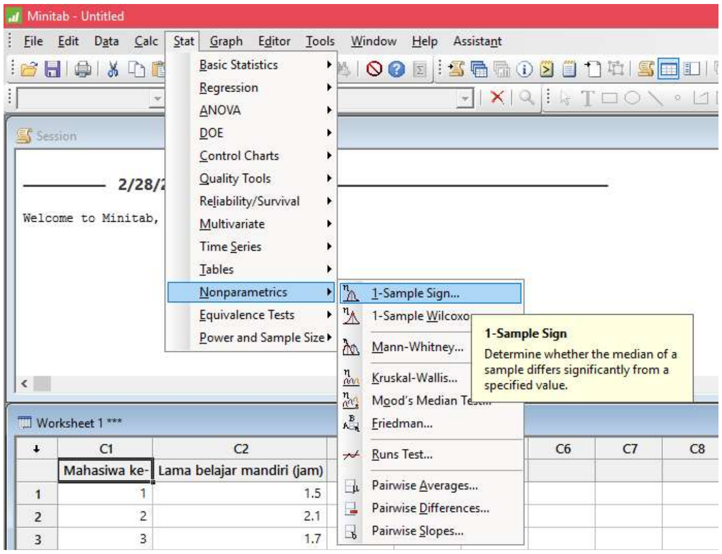
Keputusan : Dari tabel binomial (Tabel A1), diperoleh :
$P(S \le 2 \mid b(7,0.5)) = 0.0078 + 0.0547 + 0.1641 = 0.2266$
Karena nilai $P > \alpha=0.05$, maka tidak cukup bukti untuk menolak hipotesis nol.

Petunjuk Penggunaan Software Minitab

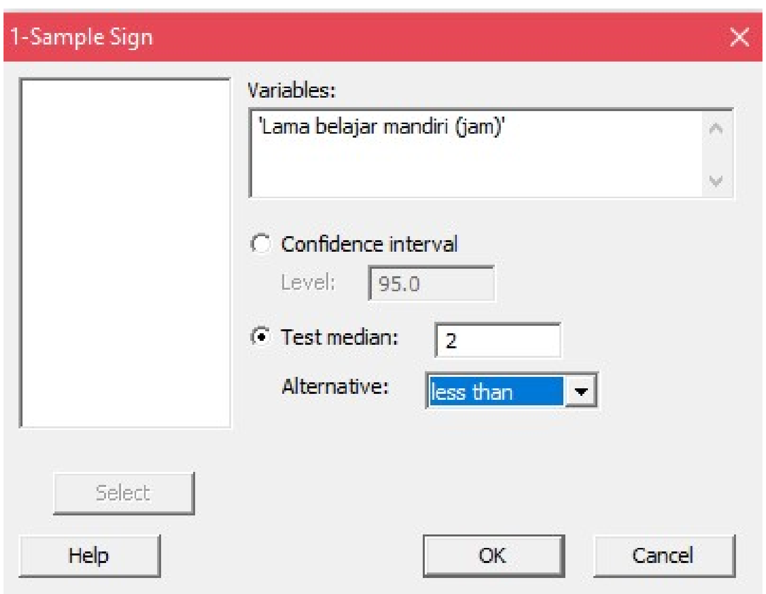
Software bisa di download di https://www.minitab.com/en-US/lp/free-statistical-software.aspx

Tahapan:

1. Entry data di worksheet session
2. Klik Stat>Nonparametrics> 1 Sample Signed



3. Isi lama belajar pada kolom "Variables", Test median ➔2, alternative ➔ less than



Output MINITAB

```
Sign Test for Median: LamaBelajar

Sign test of median =  2.000 versus < 2.000

              N  Below  Equal  Above        P  Median

LamaBelajar   7      5      0      2   0.2266   1.700
```

Uji Peringkat Bertanda Wilcoxon

Uji tanda memanfaatkan hanya tanda-tanda ‘plus’ dan ‘minus’ yang diperoleh dari selisih antara nilai pengamatan dan median pembanding, tetapi mengabaikan besarnya selisih - selisih tersebut. Wilcoxon (1945) memperkenalkan satu prosedur nonparametrik untuk menguji median yang memanfaatkan baik arah (tanda ‘plus’dan ‘minus’) maupun besar arah itu. Uji ini dikenal dengan istilah uji peringkat bertanda Wilcoxon *(Wilcoxon signed-rank test).*

Asumsi

a. Contoh acak saling bebas dengan median (M) tidak diketahui
b. Peubah yang diamati kontinu
c. Data diukur setidaknya dalam skala interval (selang)
d. Pengamatan saling bebas

Hipotesis

a. (Dua arah) : $H_0 : M = M_0$ *vs.* $H_1 : M \neq M_0$
b. (Satu arah) : $H_0 : M \leq M_0$ *vs.* $H_1 : M > M_0$
c. (Satu arah) : $H_0 : M \geq M_0$ *vs.* $H_1 : M < M_0$

Statistik Uji

Prosedur umum uji peringkat bertanda Wilcoxon adalah sebagai berikut :

1. Hitung selisih nilai data dan median untuk setiap pengamatan, $D_i = X_i – M_0$. Jika hasilnya $D_i = 0$, abaikan pengamatan tersebut.
2. Beri peringkat untuk $|D_i|$. Jika ada nilai yang sama (disebut *ties*) beri peringkat tengah *(mid-rank).*
3. Pasangkan tanda ‘plus’ dan ‘minus’ pada peringkat sesuai nilai pada langkah pertama.
4. Hitunglah : jumlah peringkat bertanda ‘plus’ (T+), dan jumlah peringkat bertanda ‘minus’ (T-).

Statistik uji yang digunakan untuk masing-masing hipotesis adalah adalah :

a. (Hipotesis a) : $T = T' = \min(T-, T+)$
b. (Hipotesis b) : $T = T-$
c. (Hipotesis c) : $T = T+$

Kaidah Keputusan

a. (Hipotesis a) : Tolak H_0 jika $T' \leq T_{n(\alpha/2)}$
b. (Hipotesis b) : Tolak H_0 jika $T- \leq T_{n(\alpha)}$
c. (Hipotesis c) : Tolak H_0 jika $T+ \leq T_{n(\alpha)}$

Catatan Untuk contoh berukuran besar dapat didekati dengan sebaran normal baku menggunakan rumus :

$$T^* = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Atau jika ada ties :

$$T^* = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}$$

Statistik uji T^* akan menyebar normal baku, $T^* \approx$ Normal (0,1)

Contoh :

Seorang dosen beranggapan bahwa median IP mahasiswa suatu kelas pada semester tertentu kurang dari 3.40. Ujilah anggapan dosen tersebut jika IP dari 10 orang mahasiswa yang diambil secara acak dari kelas tersebut adalah seperti yang tersaji dalam tabel berikut : (Gunakan taraf nyata 5%)

Mahasiswa ke-	1	2	3	4	5	6	7	8	9	10
IP	3.35	3.45	3.30	3.25	3.52	3.38	3.10	3.42	3.42	3.38

Hipotesis yang akan diuji dalam kasus ini adalah $H_0 : M \geq 4.40$ lawan $H_1 : M < 4.40$. Sesuai dengan hipotesis tersebut, maka statistik uji yang digunakan adalah T + atau jumlah peringkat selisih bertanda 'minus'. Tahapan perhitungannya adalah sebagai berikut:

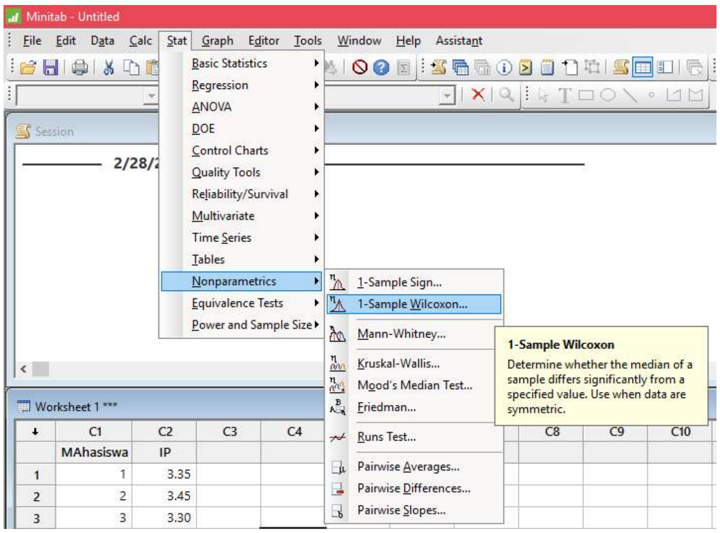
i	IP (X_i)	$D_i = X_i - 4.40$	Peringkat bagi $\lvert D_i \rvert$	Peringkat bertanda bagi $\lvert D_i \rvert$
Langkah 0		Langkah 1	Langkah 2	Langkah 3
1	3.35	–0.05	5.5	–5.5
2	3.45	0.05	5.5	5.5
3	3.30	–0.10	7	–7
4	3.25	–0.15	9	–9
5	3.52	0.12	8	8
6	3.38	–0.02	2.5	–2.5
7	3.10	–0.30	10	–10
8	3.42	0.02	2.5	2.5
9	3.42	0.02	2.5	2.5
10	3.38	–0.02	2.5	–2.5
				Langkah 4
				T - = 36.5
				T + = 18.5
				n = 10

Berdasarkan tabel di atas, diperoleh statistik uji T = T+ = 18.5. Dari tabel peringkat bertanda Wilcoxon (Tabel A3), kita peroleh $T_{10\ (0.05)}$ sekitar 11. Karena T+ lebih besar dari T_{tabel}, maka hipotesis nol <u>tidak ditolak</u> , atau dengan kata lain pernyataan dosen tersebut belum dapat dibuktikan.

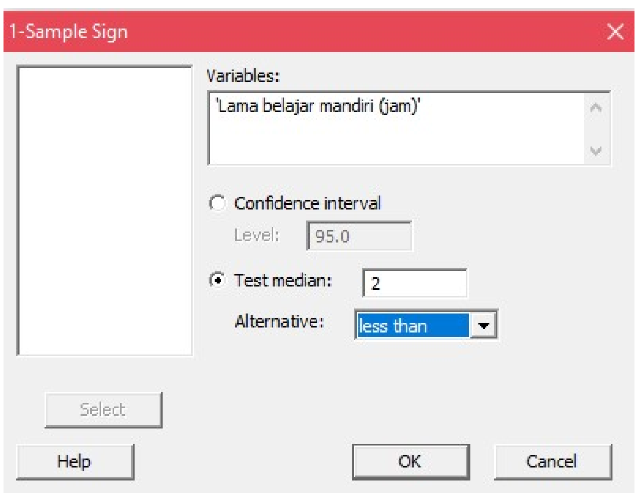
Petunjuk Penggunaan Software Minitab

Tahapan:

1. Entry data di worksheet session
2. Klik Stat>Nonparametrics>1 Sample WIlcoxon



3. Klik Stat>Nonparametrics>1 Sample WIlcoxon



Wilcoxon Signed Rank Test: IP

```
Test of median = 3.400 versus median < 3.400

          N for   Wilcoxon            Estimated
       N   Test  Statistic        P      Median

IP    10     10       18.5    0.193       3.375
```

Uji Binomial

Uji binomial *(binomial test)* merupakan prosedur nonparameterik untuk menguji hipotesis mengenai proporsi populasi.

Asumsi

a. Contoh acak biner (nilainya berupa kejadian 'sukses' dan 'gagal') berukuran *n*. rasion antara banyaknya kejadian 'sukses' (*S*) dan banyaknya pengamatan (*n*) adalah proporsi contoh.
b. Contoh acak saling bebas.
c. Peluang sukses tetap.

Hipotesis

d. (Dua arah) : $H_0 : p = p_0$ *vs.* $H_1 : p \neq p_0$
e. (Satu arah) : $H_0 : p \leq p_0$ *vs.* $H_1 : p > p_0$
f. (Satu arah) : $H_0 : p \geq p_0$ *vs.* $H_1 : p < p_0$

Statistik Uji

S = banyaknya kejadian 'sukses'

Kaidah Keputusan

d. (Hipotesis a) : Tolak H_0 jika $S \leq s_1$ atau $S > s_2$
Dimana $P(x \leq s_1) = \alpha/2$ dan $P(x > s_2) = \alpha/2$
e. (Hipotesis b) : Tolak H_0 jika $S > s$
Dimana $P(x > s) = \alpha$
f. (Hipotesis c) : Tolak H_0 jika $S \leq s$
Dimana $P(x \leq s) = \alpha$

Catatan Untuk contoh dengan *n* besar dan *p* tidak terlalu dekat dengan 0 atau 1, kita dapat menggunakan statistik uji :

$$s = np_0 + z\sqrt{np_0(1-p_0)}$$, dimana $z \approx$ Normal (0,1)

Contoh :

Seorang pejabat mengatakan bahwa di daerahnya keluarga yang mempunyai anak lebih dari hanya ada 30% di antara seluruh populasi. Dalam sebuah survei ditemukan enam dari 15 keluarga yang diambil secara acak mempunyai anak lebih dari dua. Bagaimana pendapat anda mengenai pernyataan pejabat tersebut? Gunakan taraf nyata 10%.

<u>Jawab</u> : Hipotesis yang diuji adalah H_0 : p = 0.30 lawan H_1 : $p \neq 0.30$.

n = 15

Statistik uji S = 6.

Berdasarkan tabel binomial (A1), diperoleh s_1= 1 dan s_2=7. Berdasarkan kaidah keputusan, statistik uji S=6 berada pada wilayah penerimaan H_0, sehingga kita dapat berpendapat bahwa pernyataan pejabat tersebut belum dapat diragukan.

Ringkasan

Prosedur uji tanda, uji peringkat bertanda Wilcoxon dan uji binomial.

Uji	H_1	Statistik Uji	Kaidah Penolakan H_0	Hampiran sebaran normal untuk n besar
Uji tanda	$M \neq M_0$ $M > M_0$ $M < M_0$	$S = \min(S-, S+)$ S = S- S = S+	$P(x \leq S \mid b(n,0.5)) \leq \alpha/2$ $P(x \leq S \mid b(n,0.5)) \leq \alpha$ $P(x \leq S \mid b(n,0.5)) \leq \alpha$	$z = \frac{(k+0.5)-0.5n}{0.5\sqrt{n}}$
Uji peringkat bertanda Wilcoxon	$M \neq M_0$ $M > M_0$ $M < M_0$	$T = \min(T-, T+)$ T = T- T = T+	$T \leq T_{n(\alpha/2)}$ $T \leq T_{n(\alpha)}$ $T \leq T_{n(\alpha)}$	$T^* = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ atau $T^* = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}$
Uji binomial	$p \neq p_0$ $p > p_0$ $p < p_0$	S = banyaknya kejadian	$S \leq s_1$ atau $S > s_2$ $S > s$ $S \leq s$	$s = np_0 + z\sqrt{np_0(1-p_0)}$