H. Sugiyarto, S.Si., M.Si., Ph.D.

PENGANTAR BIOSTATISTIKA

PENGANTAR BIOSTATISTIKA

Penyusun : H. Sugiyarto, M.Si., Ph.D
Layout dan Desain Cover : Rizal Redita Putra
Galang Suryaputra
Imas Abu Yazid

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI TERAPAN
UNIVERSITAS AHMAD DAHLAN

Kata Pengantar

Bismillahirrahmanirrahiim
Puji syukur Alhamdulillah penyusun panjatkan kepada Allah SWT atas berkat rahmat dan hidayah-Nya sehingga penyusun dapat menyelesaikan buku dengan judul “Pengantar Biostatistika”. Buku ini digunakan sebagai salah satu pegangan untuk mata kuliah Biostatistika.

Buku ini terdiri dari 6 bab yaitu :

1. Uji Hipotesis
2. Analisis Variansi
3. Regresi Linear Sederhana dan Korelasi
4. Regresi Linear Berganda dan Korelasi
5. Nonparametrik dan Distribusi Bebas Statistika
6. Analisis Survival

Penyusun menyadari bahwa buku ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi kesempurnaan buku ini. Akhir kata, penyusun berharap semoga buku ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan khususnya pada bidang Matematika

Yogyakarta, 2021

Penyusun

Daftar ISI

BAB I

UJI HIPOTESIS

1.1. Pendahuluan

Dalam uji hipotesis, peneliti dapat menolak atau tidak menolak (menerima) hipotesis yang diajukan. Kita akan menolak $\boldsymbol{H_0}$ apabila kenyataan yang ada berbeda secara meyakinkan atau tidak mendukung terhadap hipotesis yang diajukan. Demikian pula sebaliknya, kita akan menerima (tidak menolak) $\boldsymbol{H_0}$, jika kenyataan yang ada (data) tidak berbeda dengan hipotesis yang diajukan. Dalam menerima/menolak hipotesis tidak akan selalu benar 100%, tetapi akan selalu terdapat kesalahan (kebenaran ilmiah tidak bersifat mutlak) terutama dalam inferensi sampel terhadap populasi.

Kesalahan dalam pengambilan keputusan untuk menolak atau menerima hipotesis didasarkan pada suatu asumsi bahwa dalam ilmu pengetahuan apapun tidak ada kebenaran yang mutlak, tetapi pasti selalu ada kesalahan. Dalam uji hipotesis (uji statistik) kita jumpai adanya dua kesalahan (eror) yaitu kesalahan tipe 1 dan 2.

Kesalahan tipe 1, adalah kesalahan yang terjadi jika kita menolak $\boldsymbol{H_0}$, padahal $\boldsymbol{H_0}$ benar. Probabilitas untuk melakukan kesalahan tipe 1 ini diberi simbol $\boldsymbol{\alpha}$. Sedangkan kesalahan tipe 2 terjadi jika kita menerima (tidak menolak) $\boldsymbol{H_0}$, padahal $\boldsymbol{H_0}$ tersebut salah. Probabilitas melakukan kesalahan tipe 2 ini di beri simbol $\boldsymbol{\beta}$. Hubungan antara kesalahan 1 dan 2 ditunjukkan pada gambar berikut :

		Kondisi Sebenarnya	
		H_0 benar	H_0 salah
Keputsan	Menerima H_0	Taraf kepercayaan $1-\alpha$	Kesalahan tipe 2 β
	Menolak H_0	Kesalahan tipe 1 α	Power /Daya uji $1 - \beta$

Untuk mendapatkan keputusan yang baik, maka kedua kekeliruan tersebut harus diusahakan sekecil mungkin. Tetapi ini akan sulit dicapai, mengingat bahwa meminimalkan yang satu akan terjadi peningkatan yang lain, kecuali dengan cara memperbesar ukuran/jumlah sampel, yang pada umumnya jarang bisa dilaksanakan.

Dalam uji hipotesis diusahakan adanya keseimbangan antara kesalahan tipe 1 dan tipe 2. Artinya diusahakan pencapaian hasil pengujian hipotesis yang baik, yakni pengujian yang bersifat bahwa diantara semua pengujian yang dilakukan dengan harga α yang sama besa, ambillah sebuah kekeliruan β yang paling kecil.

Secara praktis, kekeliruan tipe 1 atau α biasanya sudah ditentukan terlebih dahulu, misalnya $\alpha = 0{,}01$ atau $\alpha = 0{,}05$. Dengan $\alpha = 0{,}05$ berarti bahwa dari tiap-tiap 100 kesimpulan yang kita buat, peluang untuk melakukan kekeliruan dengan menolak H_0 yang benar (H_0 yang seharusnya diterima) adalah sebanyak 5 kali.

Dalam statistik, yang disebut dengan hipotesis selalu diartikan sebagai hipotesis statistik atau hipotesis null (H_0). Hipotesis null (H_0) ini akan menyatakan suatu jawaban sementara bahwa keadaaan yang dibandingkan tersebut adalah tidak berbeda, atau keadaan yang dikolerasikan tersebut tidak ada hubungan didalam populasinya.

Dan supaya nampak adanya dua pilihan, hipotesis H_0 ini selalu didampingi oleh pernyataan lain yang isinya berlawanan. Pernyataan tersebut merupakan hipotesis tandingan untuk H_0, dan disebut sebagai hipotesis alternatif (H_A).

Pasangan H_0 dan H_A atau H_0 melawan H_A ini akan menentukan kriteria pengujian yang menetapkan daerah penerimaan dan daerah penolakan hipotesis. Daerah penolakan hipotesis ini sering pula dikenal dengan nama daerah kritis.

Misalkan yang akan diuji adalah suatu parameter θ (dalam penggunaannya θ ini bisa berupa rata-rata μ, proporsi π, simpangan baku σ dan sebagainya), maka akan ditemukan adanya pasangan H_0 dan H_A sebagai berikut :

1. Hipotesis mengandung pengertian “Tidak Sama”, maka pasangan H_0 dan H_A adalah :

$$H_0 : \theta = \theta_0$$

$$H_A : \theta \neq \theta_0$$

2. Hipotesis mengandung pengertian “Maksimum/lebih dari”, maka H_0 dan H_A akan berbentuk :

$$H_0 : \theta \leq \theta_0$$

$$H_A : \theta > \theta_0$$

3. Hipotesis mengandung pengertian “Minimum/kurang dari”, maka perumusan H_0 dan H_A berbentuk :

$$H_0 : \theta \geq \theta_0$$

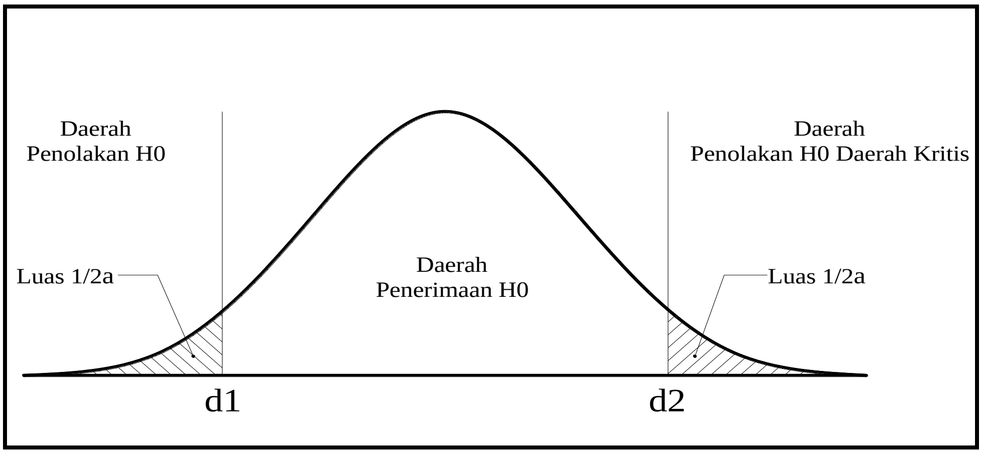
$$H_A : \theta < \theta_0$$

Dan langkah berikutnya adalah memilih teknik statistik yang akan digunakan, apakah Z, t, X^2, F atau yang lainya. Kemudian berdasarkan nilai α yang telah ditetapkan, kriteria pengujian akan dapat ditentukan.

Adapun peranan hipotesis alternatif (H_A) dalam penentuan daerah kritis (daerah penolakan H_0) adalah sebagai berikut :

a) Jika hipotesis alternatif (H_A) mempunyai rumusan <u>tidak sama</u> ($\neq$), maka dalam distribusi statistik yang digunakan, normal untuk angka Z, student untuk angka t dan seterusnya, terdapat dua daerah kritis yang masing-masing terdapat pada ujung-ujung distribusi.

   Luas daerah kritis pada tiap ujung adalah ½ α. Dan karena ada duan daerah penolakan Ho ini, maka dinamakan pengujian dua pihak (dua ekor).

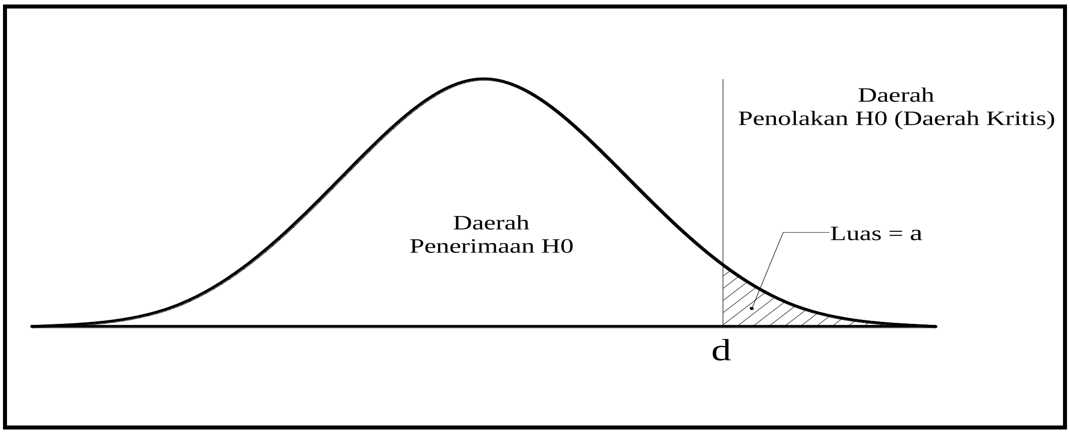


   Kedua daerah penerimaan dan penolakan H_0 tersebut dibatasi oleh bilangan $d1$ dan $d2$ yang harganya diperoleh dari daftar distribusi yang digunakan dengan peluang ralat α yang telah diterapkan.

   <u>Kriteria</u> : H_0 diterima, Jika harga statistik yang dihitung jatuh antara $d1$ dan $d2$, dan dalam hal lainnya H_0 ditolak.

b) Jika hipotesis alternatif (H_A) mempunyai rumusan lebih besar ($>$), maka dalam distribusi statistik yang digunakan terdapat sebuah daerah kritis yang letaknya diujung kanan.

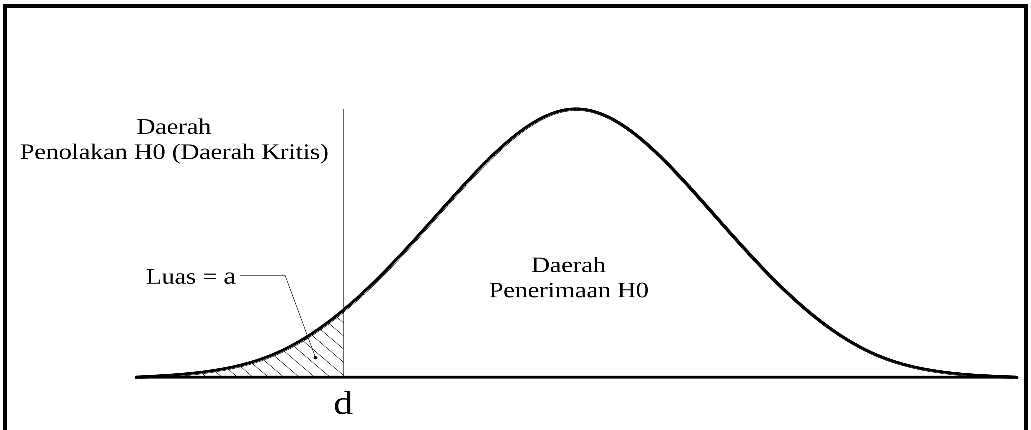
Luas daerah kritis ini adalah sama dengan α. Pengujian hipotesis ini dinamakan uji satu pihak (satu ekor) pihak kanan.



Harga d diperoleh dari daftar distribusi yang digunakan dengan peluang α yang telah ditentukan, dan menjadi batas antara daerah kritis dan daerah penerimaan H_0.

Kriteria : Tolak H_0 Jika harga statistik hasil perhitungan berdasarkan sampel lebih besar dari harga d, dan dalam hal lainya H_0 diterima.

c) Jika hipotesis alternatif (H_A) mengandung pernyataan lebih kecil $(<)$, maka daerah kritis berada di ujung kiri dari distribusi. Luas daerah ini adalah α, dan dibatasi oleh bilangan d yang diperoleh dari daftar distribusi yang bersangkutan dengan α tertentu yang telah ditetapkan. Pengujian hipotesis ini disebut pengujian satu pihak (satu ekor) pihak kiri.



Kriteria : Terima H_0, jika hasil perhitungan statistik yang diperoleh berdasarkan data penelitian lebih besar dari harga α, dan dalam hal lainya H_0 ditolak.

1.2. Uji Hipotesis : Populasi Rata-Rata Tunggal

Misalkan ada suatu populasi normal dengan rata-rata μ dan simpangan baku $\boldsymbol{\sigma}$. Dalam hal ini, akan dilakukan pengujian terhadap parameter rata-rata μ. Untuk itu, diambil sebuah sampel acak berukuran $\boldsymbol{n}$, kemudian diperoleh harga rata-rata $\overline{\boldsymbol{X}}$ dan simpangan baku $\boldsymbol{s}$. Dalam hal ini dapat dibedakan menjadi dua hal :

Populasi σ² / Variansi diketahui berdistribusi normal

Uji 2 Pihak

Hipotesisnya dirumuskan :

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0$$

Untuk menguji hipotesis ini digunakan statistik Z dengan rumus:

$$Z = \frac{X - \mu_0}{\sigma / \sqrt{n}}$$

$$\frac{\sigma}{\sqrt{n}} = Perkiraan\ standar\ eror\ dari\ rata - rata\ sampel.$$

Statistik Z ini berdistribusi normal, sehingga untuk menentukan kriteria pengujian digunakan daftar distribusi normal baku/standar.

Kriteria : H_0 diterima, jika : $-Z_{1/2(\alpha)} < Z_{hitung} < Z_{1/2(\alpha)}$ dan dalam hal lainnya H_0 ditolak. Atau $Z_{hitung} > Z_{tabel}$ maka H_0 ditolak (untuk Z_{hitung} positif) dan $Z_{hitung} < Z_{tabel}$ maka H_0 ditolak (untuk Z_{hitung} negatif).

Harga $Z_{1/2(\alpha)}$ ini diperoleh dari daftar distribusi normal baku dengan peluang $\frac{1}{2}(\alpha)$.

Kriteria p-value :

Jika p-value $\leq \alpha$ maka H_0 ditolak dan jika p-value $> \alpha$ maka H_0 gagal ditolak $(diterima)$.

Contoh 1.1. (Example 7.2.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata usia pada suatu populasi. Diambil dari 10 sampel yang memiliki rata-rata 27 tahun dengan variansi 20. Dapatkah disimpulkan bahwa rata-rata usia pada populasi berbeda dari 30 tahun?

Penyelesaian :

- **Data :** $n = 10,\ \bar{X} = 27,\ \sigma^2 = 20$ dengan $\alpha = 0.05$.

- **Asumsi :** Asumsikan sampel dimana usia di aproksimasi berdistribusi normal. Dengan asumsi variansi diketahui $\sigma^2 = 20$.

- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya sama dengan 30 (tahun) dan hipotesis alternatif rata-rata populasinya tidak sama dengan 30 (tahun). Dengan menggunakan kesalahan tipe 1 maka,

$$H_0 :\ \mu = 30$$

$$H_A :\ \mu \neq 30$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji Z.

$$Z = \frac{X - \mu_0}{\sigma/\sqrt{n}}$$

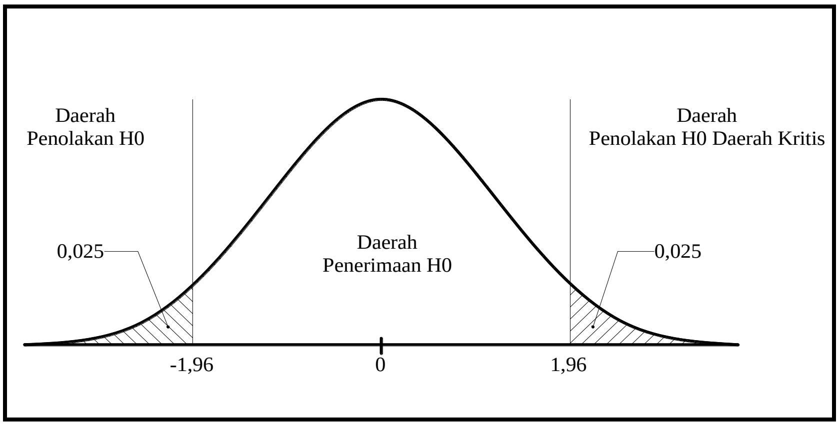
- **Distribusi pada Uji Statistik :** Distribusi sampel dan distribusi normal, kita ketahui bahwa uji statistiknya berdistribusi normal standar dengan rata-rata 0 dan variansinya 1 jika H_0 benar. Besar kemungkinan nilai pada uji statistik mempresentasikan situasi yang terjadi.

- **Kriteria Keputusan : Tabel $\boldsymbol{Z}$** $\rightarrow Z_{1/2(\alpha)} = Z_{1/2(0.05)} = Z_{0,025} = 1{,}96$**.** H_0 ditolak jika Uji Statistik $\geq 1{,}96$ atau $\leq -1{,}96$.

- **Hitung Uji Statistik :**

$$Z = \frac{27 - 30}{\sqrt{20/10}} = \frac{-3}{1.4142} = -2.12$$

- **Keputusan Statistik :**



-2.12 berada di daerah penolakan H_0 dengan $2.12 > 1.96$ atau $-2.12 < -1.96$ maka H_0 ditolak. Dengan p-value $= 0.0170$ oleh karena 2 sisi maka $0.0340 < 0.05$ maka H_0 ditolak.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa $H_A \neq 30$ diterima yakni rata-rata usia pada populasi berbeda dari 30 tahun.

Uji 1 pihak

Hipotesisnya dirumuskan :

$$H_0 : \mu = \mu_0$$

$$H_A : \mu > \mu_0 \text{ atau } H_A : \mu < \mu_0$$

Untuk menguji hipotesis ini digunakan statistik Z dengan rumus :

$$Z = \frac{X - \mu_0}{\sigma / \sqrt{n}}$$

$$\frac{\sigma}{\sqrt{n}} = Perkiraan\ standar\ eror\ dari\ rata - rata\ sampel.$$

Statistik Z ini berdistribusi normal, sehingga untuk menentukan kriteria pengujian digunakan daftar distribusi normal baku/standar.

Kriteria : H_0 ditolak, jika : $Z_{hitung} \leq Z_{(\alpha)}$.

Contoh 1.2. (Example 7.2.2 , "BIOSTATISTICS Wayne W. Daniel Edisi 10") Dengan menggunakan data pada contoh 1.1, Dapatkah disimpulkan bahwa rata-rata usia pada populasi kurang dari 30 tahun?

Penyelesaian :

- **Data :** $n = 10,\ \bar{X} = 27,\ \sigma^2 = 20$ dengan $\alpha = 0.05$.
- **Asumsi :** Asumsikan sampel dimana usia diaproksimasi dengan distribusi normal.Dengan asumsi variansi diketahui $\sigma^2 = 20$.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya lebih dari sama dengan 30 (tahun) dan hipotesis alternatif rata-rata populasinya kurang dari 30 (tahun). Dengan menggunakan kesalahan tipe 1 maka,

$$H_0 : \mu \geq 30$$

$$H_A : \mu < 30$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji Z.

$$Z = \frac{X - \mu_0}{\sigma / \sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Distribusi sampel dan distribusi normal, kita ketahui bahwa uji statistiknya berdistribusi normal standar dengan rata-rata 0 dan variansinya 1 jika H_0 benar. Besar kemungkinan nilai pada uji statistik mempresentasikan situasi yang terjadi.

- **Kriteria Keputusan : Tabel $\boldsymbol{Z} \rightarrow \boldsymbol{Z_{(\alpha)}} = \boldsymbol{Z_{(0.05)}} = \boldsymbol{Z_{0,05}} = \boldsymbol{-1,645}$.** H_0 ditolak jika Uji Statistik ≤ -1.645.

- **Hitung Uji Statistik :**

$$Z = \frac{27 - 30}{\sqrt{20/10}} = \frac{-3}{1.4142} = -2.12$$

- **Keputusan Statistik :** Oleh karena $= -2.12 < -1.645$ maka H_0 ditolak. Dengan demikian p-value $= 0.170$.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa $H_A < 30$ diterima yakni rata-rata usia pada populasi dapat disimpulkan kurang dari 30 tahun.

Populasi σ^2 / Variansi tidak diketahui berdistribusi normal

Uji 2 pihak

Hipotesisnya dirumuskan :

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0$$

Untuk menguji hipotesis ini digunakan statistik student's t dengan rumus :

$$t = \frac{X - \mu_0}{s / \sqrt{n}}$$

$$\frac{s}{\sqrt{n}} = Perkiraan\ standar\ eror\ dari\ rata - rata\ sampel.$$

Yang mana ketika H_0 benar, distribusi t-student's derajat bebasnya $n - 1$.

Kriteria : H_0 ditolak, jika : $t_{hitung} \geq t_{(\alpha)}$ atau : $t_{hitung} \leq -t_{(\alpha)}$.

Contoh 1.3. (Example 7.2.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Diberikan data sebagai berikut,

Subjek	Hari	Subjek	Hari	Subjek	Hari	Subjek	Hari
1	14	6	0	11	28	16	14
2	9	7	10	12	24	17	9
3	18	8	4	13	24		
4	26	9	8	14	2		
5	12	10	21	15	3		

Seorang peneliti ingin mengetahui rata-rata hari pada suatu populasi. Dapatkah disimpulkan bahwa rata-rata usia pada populasi tidak sama dengan 15 hari?

Penyelesaian :

- **Data :** $n = 17,\ \bar{X} = 13.2941,\ s = 8.88654$ dengan $\alpha = 0.05$.
- **Asumsi :** Asumsikan sampel dimana hari diaproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya sama dengan 15 (hari) dan hipotesis alternatif rata-rata populasinya tidak sama dengan 15 (hari).

$$H_0 :\ \mu = 15$$

$$H_A :\ \mu \neq 15$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, variansi tidak diketahui maka uji statistiknya gunakan Uji t-student's.

$$t = \frac{X - \mu_0}{s/\sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat kebebasan $n-1=17-1=16$ jika H_0 benar.

- **Kriteria Keputusan :** Dengan $\alpha = 0.05$ dengan uji 2 pihak, maka

  Tabel $t \rightarrow t_{1/2(\alpha)} = t_{1/2(0.05)} = t_{0,025} = -2.1199$. H_0 ditolak jika Uji Statistik lebih besar sama dengan 2.1199 atau lebih kecil sama dengan -2.1199.

- **Hitung Uji Statistik :**

$$t = \frac{13.2941 - 35}{8.88654\sqrt{17}} = \frac{-1.7059}{2.1553} = -0.791$$

- **Keputusan Statistik :** Dengan $-0.791 > -2.1199$ atau $0.791 < 2.1199$ maka H_0 gagal ditolak atau H_0 diterima. Dengan p-value $= 0.4403$.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa $H_0 = 15$ diterima yakni rata-rata hari pada populasi bisa sama dengan 15 hari.

Dengan menggunakan software,

1. Buat data pada **R Commander**
   #memanggil_data
   Contoh3

1	14
2	9
3	18
4	26
5	12
6	0
7	10
8	4
9	8
10	21
11	28
12	24
13	24
14	2
15	3
16	14
17	9

2. Uji t.test dengan syntax :

```
> with(contoh3, (t.test(hari, alternative='two.sided', mu=15,  conf.level=.95)))
```

Diperoleh hasil :

```
One Sample t-test

data:  hari

t = −0.79148,  df = 16,  p-value  0.4402

alternative hypothesis: true mean is not equal
to 15

95 percent confidence interval:

8.725081 17.863155

sample estimates :

mean of x
13.29412
```

Keputusan : H_0 gagal ditolak maka H_0 diterima sehingga rata-rata harinya bisa kemungkinan 15 hari.

Sampel dari populasi tidak berdistribusi normal

Jika sampel besar (≥ 30) dengan teorema limit pusat dan gunakan $Z = \frac{X-\mu_0}{\sigma/\sqrt{n}}$ sebagai uji statistiknya. Jika populasi standar deviasi/variansinya tidak diketahui, biasanya menggunakan standar deviasi sampel sebagai estimasi. Uji statistik untuk uji $H_0 : \mu = \mu_0$ adalah

$$Z = \frac{X - \mu_0}{\sigma/\sqrt{n}}$$

Ketika H_0 benar, aproksimasi distribusi berdistribusi normal ketika n besar.

Contoh 1.4. (Example 7.2.4, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata pada suatu populasi dengan 157 sampel yang memiliki rata-rata 146 dan standar deviasi 27. Dapatkah disimpulkan bahwa rata-rata pada populasi lebih dari 140?

Penyelesaian :

- **Data :** $n = 157,\ \bar{X} = 146,\ \sigma = 27$ dengan $\alpha = 0.05$.

- **Asumsi :** Asumsikan sampel di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.

- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya kurang dari sama dengan 140 dan hipotesis alternatif rata-rata populasinya lebih dari 140.

$$H_0 :\ \mu \leq 140$$

$$H_A :\ \mu > 140$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji Z.

$$Z = \frac{X - \mu_0}{\sigma/\sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Karena pada teorema limit pusat maka uji statistiknya berdistribusi normal standar dengan rata-rata 0 jika H_0 benar.

- **Kriteria Keputusan :** Tabel $Z \rightarrow Z_{(\alpha)} = Z_{(0.05)} = 1{,}645$. H_0 ditolak jika Pengujian Statistik $\geq 1{,}645$.

- **Hitung Uji Statistik :**

$$Z = \frac{146 - 140}{27\sqrt{157}} = \frac{6}{2.1548} = 2.78$$

- **Keputusan Statistik :** Oleh karena $2.78 > 1{,}645$ maka H_0 ditolak. Dengan p-value $= 1 - 0.9973 = 0.0027$ pada tabel appendix tabel D, daerah $0.0027 < 0.05$ maka daerahnya sebelah kanan 1.645.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa $H_A > 140$ diterima yakni rata-rata pada populasi dapat disimpulkan lebih dari 140.

Contoh 1.5. (Example 7.2.5 , “BIOSTATISTICS Wayne W. Daniel Edisi 10”) Diketahui data,

33.38	32.15	33.99	34.10	33.97
34.34	33. 95	33.85	34.23	32.73
33.46	34.13	34.45	34.19	34.05

Seorang peneliti ingin mengetahui rata-rata panjang pada suatu populasi. Dapatkah disimpulkan bahwa rata-rata panjang pada populasi berbeda dari 34.5 cm?

Penyelesaian :

Hipotesisnya :

$$H_0 : \mu = 34.5$$

$$H_A : \mu \neq 34.5$$

Uji statistiknya gunakan t-student’s

Dengan menggunakan software R,

1. buat data baru di **R Commander**, view data dengan syntax:

   > contoh5

   panjang

1	33.38
2	34.34
3	33.46
4	32.15
5	33.95
6	34.13
7	33.99
8	33.85
9	34.45
10	34.10
11	34.23
12	34.19
13	33.97
14	32.73
15	34.05

2. Uji t.test, syntax:

```
> with(contoh5, (t.test(panjang, alternative='two.sided', mu=34.5,  conf.level=.95)))
```

Diperoleh hasil :

One Sample t-test

data : panjang

$t = -4.3136, \; df = 14, \; p$-value = 0.0007145

alternative hypothesis : true mean is not equal to 34.5

95 percent confidence interval :

$33.44895 \;\; 34.14705$

sample estimates:

mean of x

33.798

Keputusan : p-value $= 0.0007145 < 0.05$ maka H_0 ditolak.

LATIHAN : Uji Rata-Rata Tunggal

Latihan 1.1. (Exercise 7.2.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Diambil sampel acak sederhana pada data 76 wanita di Universitas Ontario bagian barat dan Universitas McMaster dengan nilai fungsi rata-ratanya 70.7 dan standar deviasinya 14.6. Dapatkah disimpulkan bahwa rata-rata populasi pada wanita kurang dari 75?

Penyelesaian :

- **Data :** $n = 76, \; \bar{X} = 70.7, \; \sigma = 14.6$ dengan $\alpha = 0.01$.
- **Asumsi :** Asumsikan sampel dimana usia diaproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya lebih dari sama dengan 75 dan hipotesis alternatif rata-rata populasinya kurang dari 75. Dengan menggunakan kesalahan tipe 1 maka,

$$H_0 : \mu \geq 75$$

$$H_A : \mu < 75$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji Z.

$$Z = \frac{X - \mu_0}{\sigma / \sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Distribusi sampel dan distribusi normal, kita ketahui bahwa uji statistiknya berdistribusi normal standar dengan rata-rata 0 jika H_0 benar.

- **Kriteria Keputusan : Tabel $\boldsymbol{Z}$** $\rightarrow Z_{(\alpha)} = Z_{(0.01)} = 2.33$. H_0 ditolak jika Pengujian Statistik ≤ -2.33.

- **Hitung Uji Statistik :**

$$Z = \frac{70.7 - 75}{14.6\sqrt{76}} = -2.57$$

- **Keputusan Statistik :** Oleh karena $= -2.57 < 2.33$ maka H_0 ditolak.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.01$, hasil uji menunjukkan bahwa $H_A < 75$ diterima yakni rata-rata pada populasi dapat disimpulkan kurang dari 75.

Latihan 1.2. (Exercise 7.2.2 , "BIOSTATISTICS Wayne W. Daniel Edisi 10") Sebuah penelitian oleh Thienprasiddhi dkk. meneliti sampel dari 16 orang yang mengidap glaukoma dan cacat hemifield. Berikut ini adalah datanya :

$$62 \;\; 62 \;\; 68 \;\; 48 \;\; 51 \;\; 60 \;\; 51 \;\; 57 \;\; 57 \;\; 41 \;\; 62 \;\; 50 \;\; 53 \;\; 34 \;\; 62 \;\; 61$$

Dapatkah disimpulkan bahwa rata-rata usia populasi dari sampel diatas kurang dari 60 tahun?

Penyelesaian :

- **Data :** $n = 16,\ \bar{X} = 54.9375,\ \mathrm{s} = 8.8729$ dengan $\alpha = 0.05$.

- **Asumsi :** Dengan asumsi variansi tidak diketahui.

- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya lebih dari sama dengan 60 dan hipotesis alternatif rata-rata populasinya kurang dari 60.Dengan menggunakan kesalahan tipe 1 maka,

$$H_0 : \ \mu \geq 60$$

$$H_A : \ \mu < 60$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi variansi tidak diketahui maka uji statistiknya gunakan uji t.

$$t = \frac{X - \mu_0}{s/\sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat kebebasan $n - 1 = 16 - 1 = 15$ jika H_0 benar.

- **Kriteria Keputusan :** Dengan $\alpha = 0.05$ dengan uji 1 pihak maka t tabel → $t_{(\alpha)} = t_{(0.05)} = 1.7530$. H_0 ditolak jika Pengujian Statistiknya ≥ 1.7530 atau ≤ -1.7530.

- **Hitung Uji Statistik :**

$$t = \frac{54.9375 - 60}{8.8729\sqrt{16}} = -2.2822$$

- **Keputusan Statistik :** Oleh karena $= -2.2822 < -1.7530$ maka H_0 ditolak. Dengan p-value $= 0.01874$.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa $H_A < 60$ diterima yakni rata-rata usia pada populasi dapat disimpulkan kurang dari 60 tahun.

Dengan menggunakan software R, syntax :

```
>with(latihan2, (t.test(usia, alternative='less', mu=60, conf.level=.95)))
```

Diperoleh hasil :

```
One Sample t-test
data:  usia
t = −2.2822,  df = 15,  p-value = 0.01874
alternative hypothesis : true mean is less than 60
95 percent confidence interval:
    -Inf 58.82618
sample estimates:
mean of x
  54.9375
```

Keputusan : p-value $= 0.01874 < 0.05$ maka H_0 ditolak.

Latihan 1.3. (Exercise 7.2.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Tujuan penelitian oleh Luglie dkk. adalah untuk menyelidiki status sekelompok pasien yang didiagnosis menderita Talasemia mayor. Salah satu hasil akhirnya adalah indeks gigi yang rusak, hilang, dan penuh. Dalam 18 pasien, rata-ratanya adalah 10.3 dengan standar deviasinya 7.3. Apakah ini cukup membuktikan untuk menyimpulkan bahwa indeks rata-ratanya lebih besar dari 9.0?

Penyelesaian :

- **Data :** $n = 18,\ \bar{X} = 10.3,\ \mathrm{s} = 7.3$ dengan $\alpha = 0.01$.
- **Asumsi :** Asumsikan sampel dimana sampel di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya kurang dari sama dengan 9.0 dan hipotesis alternatif rata-rata populasinya lebih dari 9.0.

$$H_0 :\ \mu \leq 9.0$$

$$H_A :\ \mu > 9.0$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji t.

$$t = \frac{X - \mu_0}{{}^{\mathrm{s}}/_{\sqrt{n}}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat kebebasan $n - 1 = 18 - 1 = 17$ jika H_0 benar.
- **Kriteria Keputusan :** Dengan $\alpha = 0.01$ dengan uji 1 pihak maka table $t \to t_{(\alpha)} = t_{(0.01)} = 1.33$. H_0 ditolak jika Uji Statistik ≥ 1.33 atau ≤ -1.33.
- **Hitung Uji Statistik :**

$$t = \frac{10.3 - 9.0}{7.3\sqrt{18}} = 0.76$$

- **Keputusan Statistik :** Oleh karena $0.76 < 1.33$ maka H_0 gagal ditolak.
- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.01$, hasil uji menunjukkan bahwa rata-rata pada populasi bisa kemungkinan ≤ 9.0.

Latihan 1.4. (Exercise 7.2.4, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Sebuah penelitian memiliki catatan 25 pasien di RSJ Cronis dengan pasien rawat jalan. Rata-rata jumlah kunjungan pasien rawat jalan per pasien 4.8 dengan standar deviasinya 2. Dapatkah disimpulkan dari data tersebut bahwa rata-rata populasinya lebih dari 4 kunjungan per pasien?

Penyelesaian :

- **Data :** $n = 25,\ \bar{X} = 4.8,\ \mathrm{s} = 2$ dengan $\alpha = 0.05$.

- **Asumsi :** Asumsikan sampel dimana data di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.

- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya kurang dari sama dengan 4 dan hipotesis alternatif rata-rata populasinya lebih dari 4.

$$H_0 :\ \mu \leq 4$$

$$H_A :\ \mu > 4$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji t.

$$t = \frac{X - \mu_0}{\mathrm{s}/\sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat bebas $n - 1 = 25 - 1 = 24$ jika H_0 benar.

- **Kriteria Keputusan :** Dengan $\alpha = 0.01$ dengan uji 1 pihak maka

  t tabel → $t_{(\alpha)} = t_{(0.05)} = 1.71088$. H_0 ditolak jika Pengujian Statistik ≥ 1.71088 atau ≤ -1.71088.

- **Hitung Uji Statistik :**

$$t = \frac{4.8 - 4}{2\sqrt{25}} = 2$$

- **Keputusan Statistik :** Oleh karena $2 > 1.71088$ maka H_0 ditolak.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa rata-rata pada populasi dapat disimpulkan > 4.

Latihan 1.5. (Exercise 7.2.5, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Diketahui data $n = 49$ dengan rata-ratanya 21 dan standar deviasinya 11. Dapatkah disimpulkan dari data tersebut bahwa rata-rata populasinya kurang dari 30?

Penyelesaian :

- **Data :** $n = 49,\ \bar{X} = 21,\ \sigma = 11$ dengan $\alpha = 0.05$.
- **Asumsi :** Asumsikan sampel dimana data di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui. Data $n \geq 30$.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya lebih dari sama dengan 30 dan hipotesis alternatif rata-rata populasinya kurang dari 30.

$$H_0 :\ \mu \geq 30$$

$$H_A :\ \mu < 30$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji Z.

$$Z = \frac{X - \mu_0}{\sigma/\sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Karena pada teorema limit pusat maka uji statistiknya berdistribusi normal standar dengan rata-rata 0 jika H_0 benar.
- **Kriteria Keputusan :** Tabel $Z \rightarrow Z_{(\alpha)} = Z_{(0.05)} = 1{,}645.$ H_0 ditolak jika Pengujian Statistik $\geq 1{,}645$.
- **Hitung Uji Statistik :**

$$Z = \frac{21 - 30}{11\sqrt{49}} = 5.727$$

- **Keputusan Statistik :** Oleh karena $5.727 > 1{,}645$ maka H_0 ditolak.
- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa rata-rata pada populasi dapat disimpulkan < 30.

Latihan 1.6. (Exercise 7.2.6, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Diketahui $n = 9$ dengan rata-ratanya 6.5 dan standar deviasinya 6. Dapatkah disimpulkan dari data tersebut bahwa rata-rata populasinya lebih dari 6?

Penyelesaian :

- **Data :** $n = 9,\ \bar{X} = 6.5,\ \text{s} = 6$ dengan $\alpha = 0.05$.
- **Asumsi :** Asumsikan sampel dimana data di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya kurang dari sama dengan 6 dan hipotesis alternatif rata-rata populasinya lebih dari 6.

$$H_0 :\ \mu \leq 6$$

$$H_A :\ \mu > 6$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji t.

$$t = \frac{X - \mu_0}{\text{s}/\sqrt{n}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat kebebasan $n - 1 = 9 - 1 = 8$ jika H_0 benar.
- **Kriteria Keputusan :** Dengan $\alpha = 0.05$ dengan uji 1 pihak maka

  t tabel → $t_{(\alpha)} = t_{(0.05)} = 1.86$. H_0 ditolak jika Uji Statistik ≥ 1.86 atau ≤ -1.86.

- **Hitung Uji Statistik :**

$$t = \frac{6.5 - 6}{6\sqrt{9}} = 0.25$$

- **Keputusan Statistik :** Oleh karena $0.25 < 1.86$ maka H_0 gagal ditolak.
- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa rata-rata pada populasi dapat kemungkinan ≤ 6.

Latihan 1.7. (Exercise 7.2.7, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Diketahui data $n = 25$ dengan rata-ratanya 77 dan standar deviasinya 10. Dapatkah disimpulkan dari data tersebut bahwa rata-rata populasinya kurang dari 80?

Penyelesaian :

- **Data :** $n = 25,\ \bar{X} = 77,\ \sigma = 10$ dengan $\alpha = 0.05$.

- **Asumsi :** Asumsikan sampel dimana data di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.

- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya lebih dari sama dengan 80 dan hipotesis alternatif rata-rata populasinya kurang dari 80.

$$H_0 : \mu \geq 80$$

$$H_A : \mu < 80$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui maka uji statistiknya gunakan Uji t.

$$t = \frac{X - \mu 0}{{}^{s}/_{\sqrt{n}}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat kebebasan $n - 1 = 25 - 1 = 24$ jika H_0 benar.

- **Kriteria Keputusan :** Dengan $\alpha = 0.05$ dengan uji 1 pihak maka

  t tabel → $t_{(\alpha)} = t_{(0.05)} = 1.71088$. H_0 ditolak jika Pemgujian Statistik ≥ 1.71088 atau ≤ -1.71088.

- **Hitung Uji Statistik :**

$$t = \frac{77 - 80}{10\sqrt{25}} = -1.5$$

- **Keputusan Statistik :** Oleh karena $-1.5 > -1.71088$ maka H_0 gagal ditolak.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa rata-rata pada populasi dapat kemungkinan ≥ 80.

Latihan 1.8. (Exercise 7.2.12, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Diketahui 15 data sampel acak dengan satuan ml, berikut datanya :

$$14.0 \;\; 14.1 \;\; 14.5 \;\; 13.2 \;\; 11.2 \;\; 14.0 \;\; 14.1 \;\; 12.2$$

$$11.1 \;\; 13.7 \;\; 13.2 \;\; 16.0 \;\; 12.8 \;\; 14.4 \;\; 12.9$$

Dapatkah disimpulkan rata-ratanya tidak sama dengan 12 ml?

Penyelesaian :

- **Data :** $n = 15,\ \bar{X} = 13.42667,\ \text{s} = 1.282$ dengan $\alpha = 0.05$.

- **Asumsi :** Asumsikan sampel dimana hari di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.

- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya sama dengan $12\ (ml)$ dan hipotesis alternatif rata-rata populasinya tidak sama dengan $12\ (ml)$.

$$H_0 :\ \mu = 12$$

$$H_A :\ \mu \neq 12$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, variansi tidak diketahui maka uji statistiknya gunakan Uji t-student's.

$$t = \frac{X - \mu_0}{{}^{\text{s}}/_{\sqrt{n}}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student's dengan derajat kebebasan $n - 1 = 15 - 1 = 14$ jika H_0 benar.

- **Kriteria Keputusan :** Dengan $\alpha = 0.05$ dengan uji 2 pihak maka t tabel → $t_{{}^{1}/_{2}(\alpha)} = t_{{}^{1}/_{2}(0.05)} = t_{0,025} = 2.14479$. H_0 ditolak jika Pengujian Statistik ≥ 2.14479.

- **Hitung Uji Statistik :**

$$t = \frac{13.42667 - 12}{1.282\sqrt{15}} = 4.31$$

- **Keputusan Statistik :** Dengan $4.31 > 2.14479$, H_0 ditolak, p-value = 0.0007195.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.05$, hasil uji menunjukkan bahwa $H_A \neq 12$ diterima yakni rata-rata pada populasi dapat disimpulkan tidak sama dengan 12.

  Dengan menggunakan software R, syntax:

```
>with(latihan8, (t.test(ml, alternative='two.sided', mu=12, conf.level=.95)))
```

```
One Sample t-test
data:  ml
t = 4.31, df = 14, p-value = 0.0007195
alternative hypothesis: true mean is not equal to 12
95 percent confidence interval:
 12.71672 14.13661
```

sample estimates: mean of x 13.42667

- **Keputusan :** oleh karena $0.0007195 < 0.05$ maka H_0 ditolak.

Latihan 1.9. (Exercise 7.2.13, “BIOSTATISTICS Wayne W. Daniel Edisi 10”) Diketahui 20 data sampel acak dengan satuan liter/menit, berikut datanya:

$$132 \;\; 33 \; 91 \;\; 108 \;\; 67 \;\; 169 \;\; 54 \;\; 203 \;\; 190 \;\; 133$$

$$96 \;\; 30 \;\; 187 \;\; 21 \;\; 63 \;\; 166 \;\; 84 \;\; 110 \;\; 157 \;\; 138$$

Dapatkah disimpulkan rata-ratanya tidak sama dengan 110 liter/menit?

Penyelesaian :

- **Data :** $n = 20,\; \bar{X} = 111.6,\;$ s $= 56.30313$ dengan $\alpha = 0.01$.
- **Asumsi :** Asumsikan sampel dimana data di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.
- **Hipotesis :** Dengan Hipotesis nol rata-rata populasinya sama dengan 110 dan hipotesis alternatif rata-rata populasinya tidak sama dengan 110.

$$H_0 :\; \mu = 110$$

$$H_A :\; \mu \neq 110$$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, variansi tidak diketahui maka uji statistiknya gunakan Uji t-student’s.

$$t = \frac{X - \mu_0}{{}^{s}\!/_{\sqrt{n}}}$$

- **Distribusi pada Uji Statistik :** Distribusi yang digunakan distribusi t-student’s dengan derajat kebebasan $n - 1 = 20 - 1 = 19$ jika H_0 benar.
- **Kriteria Keputusan :** Dengan $\alpha = 0.01$ dengan uji 2 pihak maka

  t tabel → $t_{1/_{2(\alpha)}} = \; t_{1/_{2(0.01)}} = t_{0,005} = 2.84534$**.** H_0 ditolak jika Pengujian Statistik ≥ 2.84534.

- **Hitung Uji Statistik :**

$$t = \frac{111.6 - 110}{56.30313/_{20}} = 0.12709$$

- **Keputusan Statistik :** Dengan $0.12709 < 2.84534$, maka H_0 gagal ditolak dan p-value $= 0.9002$.

- **Kesimpulan :** Bahwa pada taraf $\alpha = 0.01$, hasil uji menunjukkan bahwa $H_0 = 110$ diterima yakni rata-rata pada populasi kemungkinan sama dengan 110.

Dengan menggunakan software R, syntax:

```
> with(latihan9, (t.test(liter.menit, alternative='two.sided', mu=110,  conf.level=.99)))
```

```
One Sample t-test
data :  liter.menit
t = 0.12709, df = 19, p-value = 0.9002
alternative hypothesis : true mean is not equal to 110
99 percent confidence interval:
  75.58151   147.61849
sample estimates:
mean of x
    111.6
```

- **Keputusan :** Oleh karena $0.9002 > 0.05$ maka H_0 gagal ditolak.

1.3. Uji Hipotesis : Populasi Rata-Rata Ganda

Misalkan ada 2 rata-rata populasi dengan hipotesisnya,

1. $H_0 : \mu_1 - \mu_2 = 0,\ H_A : \mu_1 - \mu_2 \neq 0$
2. $H_0 : \mu_1 - \mu_2 \geq 0,\ H_A : \mu_1 - \mu_2 < 0$
3. $H_0 : \mu_1 - \mu_2 \leq 0,\ H_A : \mu_1 - \mu_2 > 0$

Populasi σ^2 / Variansi diketahui berdistribusi normal

Dengan menggunakan uji Z maka

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

Contoh 1.6. (Example 7.3.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata 12 sampel dan 15 sampel dengan rata-rata populasi yang pertama dan kedua masing-masing $4.5\, mg/100\, ml$ dan $3.4\, mg/100ml$ yang memiliki variansi 1 dan 1.5. Dapatkah disimpulkan bahwa rata-rata pada kedua populasi berbeda?

Penyelesaian :

- **Data :** $\bar{X}_1 = 4.5\, mg/100\, ml,\ \bar{X}_2 = 3.4\, mg/100\, ml$

  $\sigma_1^2 = 1,\ \sigma_2^2 = 1.5$ dan $n_1 = 12,\ n_2 = 15$

- **Asumsi :** Asumsikan sampel dimana kedua populasi di aproksimasi dengan distribusi normal. Dengan asumsi variansi diketahui.

- **Hipotesis :**

$$H_0 :\ \mu_1 - \mu_2 = 0$$

$$H_A :\ \mu_1 - \mu_2 \neq 0$$

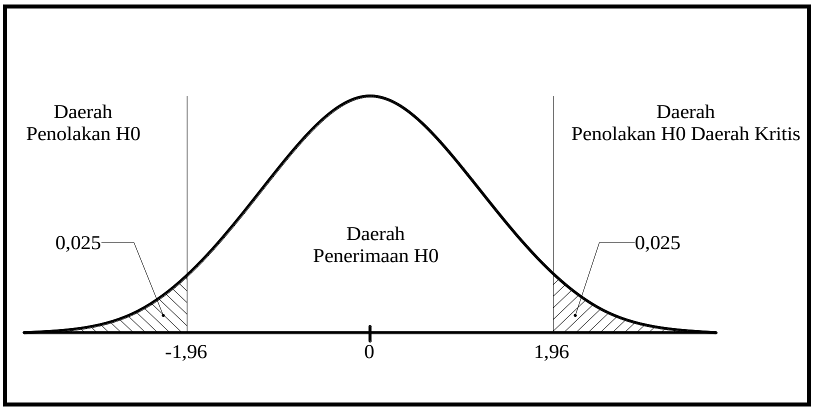
- **Distribusi pada Uji Statistik :** Distribusi sampel dan distribusi normal, kita ketahui bahwa uji statistiknya berdistribusi normal standar dengan rata-rata 0 dan variansinya 1 jika H_0 benar. Besar kemungkinan nilai pada uji statistik mempresentasikan situasi yang terjadi.

- **Kriteria Keputusan : Tabel *Z*** → $Z_{1/2(\alpha)} = Z_{1/2(0.05)} = Z_{0,025} = 1,96$**.** H_0 ditolak jika Uji Statistik $\geq 1,96$ atau $\leq -1,96$.

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji Z.

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

$$Z = \frac{(4.5\ - 3.4) - 0}{\sqrt{\frac{1}{12} + \frac{1.5}{15}}} = 2.57$$



- **Keputusan :** Oleh karena $2.57 > 1{,}96$ dengan p-value bernilai $0.0102 < 0.05$ maka H_0 ditolak sehingga selisih rata-rata kedua populasi dapat disimpulkan tidak sama dengan 0.

Populasi σ² / Variansi tidak diketahui berdistribusi normal

Populasi variansi sama

Ketika populasi variansi tidak diketahui dengan asumsi sama maka variansinya :

$$s_p{}^2 = \frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}$$

Dengan uji tes statistiknya $H_0 :\ \mu_1 = \mu_2$ dengan

$$t = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)_{\ 0}}{\sqrt{\frac{s_p{}^2}{n_1} + \frac{s_p{}^2}{n_2}}}$$

Dimana H_0 benar berdistribusi t-student's dengan derajat kebebasannya $n_1 + n_2 - 2$.

Contoh 1.7. (Example 7.3.2, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata kedua populasi.Berikut datanya,

Control	131 115 124 131 122 117 88 114 150 169
SCI	60 150 130 180 163 130 121 119 130 148

Dapatkah disimpulkan bahwa selisih rata-rata pada kedua populasi kurang dari 0?

Penyelesaian :

- **Data :**

$$\bar{X}_1 = 126.1, \qquad \bar{X}_2 = 133.1$$

$$s_1 = 21.8, s_2 = 32.2$$

$$n_1 = 10, n_2 = 10$$

- **Asumsi :** Asumsikan sampel dimana kedua populasi di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui.

- **Hipotesis :**

$$H_0 : \mu_1 - \mu_2 \geq 0$$

$$H_A : \mu_1 - \mu_2 < 0$$

- **Distribusi pada Uji Statistik :** Ketika hipotesis nol benar, gunakan uji t-student's dengan derajat bebas $n_1 + n_2 - 2$.

- **Kriteria Keputusan :** t tabel → $t_{(\alpha)} = t_{(0.05)} = -1.7341$**.** H_0 ditolak jika Pengujian Statistik > -1.7341.

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji t.

$$t = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)\ _0}{\sqrt{\frac{s_p{}^2}{n_1} + \frac{s_p{}^2}{n_2}}}$$

Dengan menghitung variansinya,

$$s_p{}^2 = \frac{9(21.8)^2 + 9(32.2)^2}{9 + 9} = 756.04$$

Maka

$$t = \frac{(126.1 - 133.1) - 0}{\sqrt{\frac{756.04}{10} + \frac{756.04}{10}}} = -0.569$$

- **Keputusan :** Oleh karena $-0.569 > -1.7341$ maka H_0 gagal ditolak dengan p-value $-1.330 < -0.569$.

- **Kesimpulan :** selisih kedua populasi lebih dari sama dengan 0.

Populasi variansi tidak sama

Ketika 2 populasi sampel acak bebas dengan variansi tidak diketahui dan asumsi tidak sama maka uji tes statistiknya $H_0 :\ \mu_1 = \mu_2$ dengan

$$t' = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)_{\ 0}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

Nilai kritis pada t' untuk level signifikan α dan 2 pihak di aproksimasi

$$t'_{1-(\alpha/2)} = \frac{(w_1t_1) + (w_2t_2)}{w_1 + w_2}$$

Dimana $w_1 = \frac{{s_1}^2}{n_1}$, $w_2 = \frac{{s_2}^2}{n_2}$, $t_1 = t_{1-(\alpha/2)}$ untuk derajat kebebasan $n_1 - 1$ dan $t_2 = t_{1-(\alpha/2)}$ untuk derajat bebas $n_2 - 1$. H_0 ditolak jika nilai hitung $t' \geq nilai\ kritis$ atau $t' \leq -nilai\ kritis$.

Untuk uji 1 pihak,

$$t'_{1-\alpha} = \frac{(w_1t_1) + (w_2t_2)}{w_1 + w_2}$$

Dimana $w_1 = \frac{{s_1}^2}{n_1}$, $w_2 = \frac{{s_2}^2}{n_2}$, $t_1 = t_{1-\alpha}$ untuk derajat bebas $n_1 - 1$ dan $t_2 = t_{1-\alpha}$ untuk derajat bebas $n_2 - 1$. H_0 ditolak jika nilai hitung $t' \geq\ nilai\ kritis$ atau $t' \leq -nilai\ kritis$.

Contoh 1.8. (Example 7.3.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata kedua populasi. Dapatkah disimpulkan bahwa selisih rata-rata pada kedua populasi tidak sama dengan 0?

Penyelesaian :

- **Data**

$$\bar{X}_1 = 19.16, \qquad \bar{X}_2 = 9.53$$

$$s_1 = 5.29, \qquad s_2 = 2.69$$

$$n_1 = 15, \qquad n_2 = 30$$

- **Asumsi :** Asumsikan sampel dimana kedua populasi di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui dan tidak sama.

- **Hipotesis :**

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 - \mu_2 \neq 0$$

- **Distribusi pada Uji Statistik :** Gunakan uji statistik distribusi t dengan variansi tidak diketahui dan tidak sama.

- **Kriteria Keputusan :** Nilai signifikansi adalah $\alpha = 0.05$, sebelum menghitung t'_{hitung} $w_1 = \frac{(5.29)^2}{15} = 1.8656,\ w_2 = \frac{(2.69)^2}{30} = 0.2412,\ t_1 = t_{1-\left(\frac{0.05}{2}\right)} = 2.1448$ untuk derajat kebebasan $n_1 - 1 = 14$ dan $t_2 = t_{1-\left(\frac{0.05}{2}\right)} = 2.0452$ untuk derajat kebebasan $n_2 - 1 = 29$ sehingga

$$t'_{1-0.05} = \frac{1.8656(2.1448) + 0.2412(2.0452)}{1.8656 + 0.2412} = 2.133$$

H_0 ditolak jika $t_{hitung} \geq 2.133$ atau $t_{hitung} \leq -2.133$.

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji t.

$$t' = \frac{(19.16 - 9.53) - 0}{\sqrt{\frac{(5.29)^2}{15} + \frac{(2.69)^2}{30}}} = 6.63$$

- **Keputusan :** Oleh karena $6.63 > 2.133$ maka H_0 ditolak dengan p-value nya < 0.05.

- **Kesimpulan :** Selisih kedua populasi tidak sama dengan 0.

Sampel dari populasi tidak berdistribusi normal

Jika sampel besar (≥ 30) dengan teorema limit pusat dan gunakan uji z sebagai uji statistiknya. Jika populasi standar deviasi/variansinya tidak diketahui, biasanya menggunakan standar deviasi sampel sebagai estimasi. Uji statistik untuk uji $H_0 : \mu = \mu_0$ adalah

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)_{\ 0}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

Ketika H_0 benar, aproksimasi distribusi berdistribusi normal ketika n besar.

Contoh 1.9. (Example 7.3.4, “BIOSTATISTICS Wayne W. Daniel Edisi 10”) Diberikan data,

$Grup$	$Rata-rata(ml/unit)$	$Ukuran\ sampel$	$Standar\ deviasi$
$Trombosit$	59.01	53	44.89
$Bukan\ trombosit$	46.61	54	34.85

Dapatkah disimpulkan bahwa rata-rata pada kedua populasi lebih dari 0?

Penyelesaian :

- **Data :**

$$\bar{X}_1 = 59.01, \qquad \bar{X}_2 = 46.61$$

$$\sigma_1 = 44.89, \qquad \sigma_2 = 34.85$$

$$n_1 = 53, \qquad n_2 = 54$$

- **Asumsi :** Asumsikan sampel dimana kedua populasi di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui dan data $\geq\ 30$.

- **Hipotesis :**

$$H_0 :\ \mu_1 - \mu_2 \leq 0$$

$$H_A :\ \mu_1 - \mu_2 > 0$$

- **Distribusi pada Uji Statistik :** Distribusi sampel dan distribusi normal, kita ketahui bahwa uji statistiknya berdistribusi normal standar dengan rata-rata 0 dan variansinya 1 jika H_0 benar. Besar kemungkinan nilai pada uji statistik mempresentasikan situasi yang terjadi.

- **Kriteria Keputusan : Tabel $\boldsymbol{Z}$ →** $Z_{(\alpha)} = Z_{(0.01)} = 2.33$**.** H_0 ditolak jika pengujian Statistik ≥ 2.33.

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi tidak diketahui dan data ≥ 30 maka uji statistiknya gunakan Uji Z.

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)_{\ 0}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

$$Z = \frac{(59.01 - 46.61) - 0}{\sqrt{\frac{44.89^2}{53} + \frac{34.85^2}{54}}} = 1.59$$

- **Keputusan :** Oleh karena $1.59 < 2.33$ dengan p-value bernilai $0.0559 > 0.01$ maka H_0 gagal ditolak sehingga selisih rata-rata kedua populasi bisa kemungkinan kurang dari sama dengan 0.

LATIHAN : Uji Rata-Rata Ganda

Latihan 1.10. (Exercise 7.3.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata BMI (indeks massa tubuh) kedua populasi. Dapatkah disimpulkan bahwa pemain Rugby lebih besar BMI nya daripada pemain multi olahraga?

Penyelesaian :

- **Data :** Dengan memisalkan pemain multi olahraga (populasi 1) dan pemain Rugby (populasi 2) sehingga

$$n_1 = 40, \qquad \bar{X}_1 = 22.41, \qquad s_1 = 1.27$$

$$n_2 = 24, \qquad \bar{X}_2 = 27.75, \qquad s_2 = 2.64$$

- **Asumsi :** Asumsikan sampel dimana kedua populasi di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui dan tidak sama.

- **Hipotesis :**

$$H_0 : \mu_1 - \mu_2 \geq 0$$

$$H_A : \mu_1 - \mu_2 < 0$$

$$atau$$

$$H_0 : \mu_1 \geq \mu_2$$

$$H_A : \mu_1 < \mu_2$$

- **Distribusi pada Uji Statistik : G**unakan uji statistik distribusi t dengan variansi tidak diketahui dan tidak sama.

- **Kriteria Keputusan :** Nilai signifikansi $\alpha = 0.01$, sebelum menghitung t'_{hitung}, $w_1 = \frac{1.27^2}{40} = 0.0403$, $t_1 = t_{1-(0.01)} = 2.7079$ untuk derajat kebebasannya $n_2 - 1 =$

39 dan $w_2 = \frac{2.64^2}{24} = 0.3030,\ t_2 = t_{1-(0.01)} = 2.8073$ untuk derajat kebebasannya $n_1 - 1 = 23$ sehingga diperoleh

$$t'_{1-0.01} = \frac{0.0403(2.7079) + 0.3030(2.8073)}{0.0403 + 0.3030} = 2.7956$$

H_0 ditolak jika $t_{hitung} \geq 2.7956$ atau $t_{hitung} \leq -2.7956$

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji t.

$$t' = \frac{(22.41 - 27.75) - 0}{\sqrt{\frac{(1.27)^2}{40} + \frac{(2.64)^2}{24}}} = -9.2856$$

- **Keputusan :** Oleh karena $-9.2856 < -2.7956$ maka H_0 ditolak dengan p-value nya < 0.005.

- **Kesimpulan :** Cukup membuktikan bahwa BMI pemain Rugby lebih besar daripada BMI pemain multi olahraga.

Latihan 1.11. (Exercise 7.3.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang peneliti ingin mengetahui rata-rata kedua populasi. Dengan data sebagai berikut,

Length (mm) Non-OSAS	96.80 100.70 94.55 99.65 109.15 102.75 97.70 92.10 91.90 89.50 97.00 97.70 97.00 94.55 106.45 94.55 94.05 89.45 89.85 98.20 101.00 88.25 92.60 98.25 90.85 95.25 88.80 101.40 90.55 109.80 88.95 101.05 92.60 97.00 91.95 88.95 95.75
Length (mm) OSAS	105.95 114.90 110.35 123.10 119.30 110.00 98.95 114.20 118.95 105.05 114.90 114.35 112.25 106.15 102.60 102.40 105.05 112.65 128.95 117.70 113.70 113.70 116.30 108.75 113.30 106.00 101.75

Dapatkah disimpulkan bahwa selisih rata-rata pada kedua populasi tidak sama dengan 0?

Penyelesaian :

- **Data**

$$n_1 = 37, \quad \bar{X}_1 = 95.85, \quad s_1 = 5.59$$

$$n_2 = 26, \quad \bar{X}_2 = 111.06, \quad s_2 = 6.96$$

- **Asumsi :** Asumsikan sampel dimana kedua populasi di aproksimasi dengan distribusi normal. Dengan asumsi variansi tidak diketahui dan tidak sama.

- **Hipotesis.**

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 - \mu_2 \neq 0$$

- **Distribusi pada Uji Statistik : G**unakan uji statistik distribusi t dengan variansi tidak diketahui dan tidak sama.

- **Kriteria Keputusan :** Nilai signifikansi $\alpha = 0.01$, sebelum menghitung t'_{hitung} $w_1 = \frac{(5.59)^2}{37} = 0.8445,\ w_2 = \frac{(6.96)^2}{26} = 1.8631,\ t_1 = t_{1-\left(\frac{0.01}{2}\right)} = 2.71948$ untuk derajat kebebasan $n_1 - 1 = 36$ dan $t_2 = t_{1-\left(\frac{0.01}{2}\right)} = 2.78744$ untuk derajat kebebasan $n_2 - 1 = 25$ sehingga

$$t'_{1-\left(\frac{0.01}{2}\right)} = \frac{0.8445(2.71948) + 1.8631(2.78744)}{0.8445 + 1.8631} = \frac{7.4899}{2.7076} = 2.7662$$

H_0 ditolak jika $t_{hitung} \geq 2.7662$ atau $t_{hitung} \leq -2.7662$.

- **Uji Statistik :** Uji hipotesis tentang rata-rata populasi, dengan asumsi populasi berdistribusi normal, dan variansi diketahui maka uji statistiknya gunakan Uji t.

$$t' = \frac{(95.85 - 111.06) - 0}{\sqrt{\frac{(5.59)^2}{37} + \frac{(6.96)^2}{26}}} = \frac{-15.21}{1.6455} = -9.2434$$

- **Keputusan :** Oleh karena $-9.2434 < -2.7662$, maka H_0 ditolak dengan p-value nya < 0.000000002746.

- **Kesimpulan :** selisih kedua populasi tidak sama dengan 0.

Dengan menggunakan software R diperoleh,

```
Paired t-test
data:  NonOSAS and OSAS
t = -8.9694, df = 25, p-value = 0.000000002746
alternative hypothesis: true difference in means is
not equal to 0
99 percent confidence interval:
 -19.54812 -10.27880
sample estimates:
mean of the differences
          -14.91346
```

Dengan syntax:

```
>with(Latihan3,(t.test(NonOSAS,OSAS,alternative='two.sided',conf.level=.99,paired=TRUE
)))
```

1.4. Perbandingan Pasangan (Paired Comparisons)

Alasan untuk dipasangkan : Uji statistik untuk uji hipotesis mengenai beda rata-rata populasi μ_d adalah

$$t = \frac{\bar{d} - \mu_{d_0}}{s_{\bar{d}}}$$

Dimana

$\bar{d}$: Beda rata-rata sampel

μ_{d_0}: Hipotesis beda rata-rata populasi

$s_{\bar{d}} = s_d/\sqrt{n},\ \ n$: angka pada beda sampel

s_d : standar deviasi pada beda sampel

Ketika H_0 benar, uji statistiknya berdistribusi t-student's dengan derajat kebebasan $n - 1$.

LATIHAN : Perbandingan (Paired Comparisons)

Latihan 1.12. (Exercise 7.4.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Peneliti mempelajari efek terapi reminiscene untuk depresi wanita yang lebih tua pada 15 wanita berusia 60 tahunan selama 2 bulan di fasilitas perawatan jangka panjang. Depresi diukur dengan skor GDS yang lebih tinggi menunjukkan gejala depresi lebih parah. Dapatkah Skor disimpulkan berdasarkan data bahwa subjek yang berpartisipasi dalam pengalaman terapi reminiscene rata-rata mengalami penurunan depresi dengan $\alpha = 0.01$?

pre-GDS	12	10	16	2	12	18	11	16	16	10	14	21	9	19	20
post-GDS	11	10	11	3	9	13	8	14	16	10	12	22	9	16	18

Penyelesaian :

- **Data :** Data tersebut merupakan sampel acak dari sebelum GDS dan sesudah GDS
- **Asumsi :** Asumsinya adalah bahwa perbedaan yang diamati merupakan sampel acak sederhana yang merupakan sebelum GDS dan sesudah GDS dari populasi perbedaan yang berdistribusi normal.
- **Hipotesis :** Gunakan hipotesis

$$H_0 : \mu_d \leq 0$$

$$H_A : \mu_d > 0$$

- **Uji Statistik :**

$$t = \frac{\bar{d} - \mu_d}{s_d}$$

- **Distribusi Statistik :** Distribusi statistik pada uji t adalah dengan derajat kebebasan $df = n - 1$ jika H_0 benar
- **Kriteria keputusan :** Dengan t table

$$\alpha = 0.01; \quad df = n - 1 = 15 - 1 = 14$$

$$t_\alpha(df) = t_{0.05}(14) = 2.624$$

- **Statistik hitung :**

$$\bar{d} = \frac{\sum d_i}{n} = \frac{24}{15} = 1.6$$

$$d_i^2 = (1 + 25 + 1 + 9 + 25 + 9 + 4 + 4 + 1 + 9 + 4) = 92$$

$$s_d^2 = \frac{n \sum d_i^2 - (\sum d_i)^2}{n(n-1)}$$

$$= \frac{15(92) - (24)^2}{15(14)}$$

$$= 3.828$$

$$t = \frac{1.6 - 0}{\sqrt{\frac{3.828}{15}}}$$

$$= 3.1683$$

- **Keputusan Statistik :** Diperoleh $t_{hitung} = 3.1683 > t\ tabel = 2.264$ maka H_0 ditolak
- **Kesimpulan :** Dapat disimpulkan bahwa skor rata-rata terapi reminiscence menurun pada skor depresi GDS. Hasilnya signifikan secara statistik.
- ***p*-value :** $P(t \leq 3.1683) = 0.003$.

Dengan menggunakan software R diperoleh :

```
Summary:

       PRE            POST
 Min.   : 2.00   Min.   : 3.00
 1st Qu.:10.50   1st Qu.: 9.50
 Median :14.00   Median :11.00
 Mean   :13.73   Mean   :12.13
 3rd Qu.:17.00   3rd Qu.:15.00
 Max.   :21.00   Max.   :22.00
Variansi:
F test to compare two variances
data:  GDS$PRE and GDS$POST
F = 1.1921, num df = 14, denom df = 14, p-value = 0.3734
```

```
alternative hypothesis: true ratio of variances is greater than 1
99 percent confidence interval:
        0.3224084       Inf
sample estimates:ratio of variances
      1.192118
```

Paired-test:

```
Welch Two Sample t-test
data:  GDS$PRE and GDS$POST
t = 0.90758, df = 27.787, p-value = 0.186
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 -2.75142      Inf
sample estimates:
mean of x mean of y
 13.73333  12.13333
```

Dengan Syntax:

```
>summary(GDS)
>#var test
>?var.test
>var.test(GDS$PRE,GDS$POST,alternative=c("greater"),conf.level=0.99)
> t.test(GDS$PRE,GDS$POST,alternative=c("greater"),conf.level=0.99)
```

- Kesimpulan : p-value $= 0.186 > 0.05$ maka H_0 ditolak yang artinya tidak ada perbedaan rata-rata antara Skor PRE dan POST atau rata-rata mengalami penurunan depresi.

Latihan 1.13. (Exercise 7.4.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Tujuan investigasi oleh Morley (A-17) adalah untuk mengevaluasi efektivitas analgesic dari dosis harian metadon oral pada pasien dengan sindrom nyeri neuropatik kronis. Peneliti menggunakan skala analog visual intensitas nyeri maksimum sepanjang hari. Setiap subjek mengasumsi $20\ mg$ metadon dan placebo dengan tingkat signifikan 0.05. untuk menunjukkan bahwa secara umum intensitas nyeri maksimum lebih rendah dari hari-hari ketika metadon digunakan?

$Subjek$	$Metadon$	$Placebo$	d_i	d_i^2
1	28.9	57.2	−27.4	750.76
2	73.0	69.8	3.2	10.24
3	98.6	98.2	0.4	0.16
4	58.8	62.4	−3.6	12.96
5	60.6	67.4	−6.6	43.56
6	57.2	70.6	−13.4	179.56
7	57.2	67.8	−10.6	112.36
8	89.2	95.6	−6.4	40.96
9	97.0	98.4	−1.4	1.96
10	49.8	63.2	13.4	179.56
11	37.0	63.6	−26.6	707.56

Penyelesaian :

- **Data :** Data tersebut merupakan sampel acak dari penggunaan metadon dan placebo
- **Asumsi :** Asumsinya adalah bahwa perbedaan yang diamati merupakan sampel acak sederhana yang merupakan metadon dan placebo dari populasi perbedaan yang berdistribusi normal.
- **Hipotesis :** Gunakan hipotesis satu arah

$$H_0 : \mu_d \leq 0$$

$$H_A : \mu_d > 0$$

- **Uji statistik :**

$$t = \frac{\bar{d} - \mu_d}{s_d}$$

- **Distribusi statistik :** Uji t adalah dengan derajat kebebasan $df = n - 1$ jika H_0 benar

- **Kriteria keputusan :** dengan t table

$$\alpha = 0.05; \quad df = n - 1 = 11 - 1 = 10$$

$$t_{\alpha}(df) = t_{0.05}(10) = -1.8125$$

- **Statistik Hitung :**

$$\bar{d} = \frac{\sum d_i}{n}$$

$$= \frac{-105.8}{11}$$

$$= -9.6182$$

$$d_i^2 = 2039.64$$

$$s_d^2 = \frac{n \sum d_i^2 - (\sum d_i)^2}{n(n-1)}$$

$$= \frac{11(2039.64) — (105.8)^2}{11(10)}$$

$$= 102.2036$$

$$t = \frac{-9.6182 - 0}{\sqrt{\frac{102.2036}{11}}}$$

$$= -3.1554$$

- **Keputusan statistik :** Diperoleh $t_{hitung} = -3.1554 < t_{tabel} = -1.8125$ maka H_0 ditolak

- **Kesimpulan :** Dapat disimpulkan bahwa skor rata-rata intensitas nyeri maksimum pada penggunaan metadon dan placebo. Hasilnya signifikan secara statistik

- ***p*-value :** $P(t \leq -3.1554) = 0.01$

Dengan menggunakan software R diperoleh:

```
Summary:
   metadon        placebo
 Min.   :28.9   Min.   :57.20
 1st Qu.:53.5   1st Qu.:63.40
 Median :58.8   Median :67.80
 Mean   :64.3   Mean   :74.02
 3rd Qu.:81.1   3rd Qu.:83.10
 Max.   :98.6   Max.   :98.40
Variansi:
F test to compare two variances
data:  intensitas$metadon and intensitas$placebo
F = 2.2059, num df = 10, denom df = 10, p-value = 0.1141
alternative hypothesis: true ratio of variances is greater than 1
99 percent confidence interval:
 0.4549042       Inf
sample estimates:
ratio of variances
          2.205897
Paired-test:
Welch Two Sample t-test
data:  intensitas$metadon and intensitas$placebo
t = -1.1619, df = 17.521, p-value = 0.8696
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 -31.12282       Inf
sample estimates:
mean of x mean of y
 64.30000  74.01818
```

Dengan Syntax :

```
>summary(intensitas)
>#var test
>?var.test
>var.test(intensitas$metadon,intensitas$placebo,alternative=c("greater"),conf.level=0.99)
>t.test(intensitas$metadon,intensitas$placebo,alternative=c("greater"),conf.level=0.99)
```

Kesimpulan : p-value $= 0.8696 > 0.05$ maka H_0 ditolak yang artinya tidak ada perbedaan rata-rata antara dosis harian metadon dan placebo atau rata-rata mengalami penurunan dosis.

Latihan 1.14. (Exercise 7.4.5, “BIOSTATISTICS Wayne W. Daniel Edisi 10”) Salah satu tujuan penyelidikan oleh Porcellini adalah untuk menyelidiki dampak pada jumlah sel CD4 T administrasi interleukin intermitten (IL-2) selain HAART. Tabel berikut menunjukkan jumlah CD4 T pada awal dan setelah 12 bulan memakai HAART dengan IL-2. Apakah data menunjukkan pada tingkat 0.05 perubahan yang signifikan dalam jumlah sel CD4 T?

$subjek$	$CD4\ T\ awal$	$CD4\ T\ akhir$	d_i	d_i^2
1	173	257	-84	7056
2	58	108	-50	2500
3	103	315	-212	44944
4	181	362	-181	32761
5	105	141	-36	1296
6	301	549	-248	61504
7	169	369	-200	40000

Penyelesaian :

- **Data :** Data tersebut merupakan sampel acak dari HAART dam setelah 12 bulan HAART.
- **Asumsi :** Asumsinya adalah bahwa perbedaan yang diamati merupakan sampel acak sederhana yang merupakan CD4 T awal dan CD4 T setelah 12 bulan dari populasi perbedaan berdistribusi normal.
- **Hipotesis :** Gunakan hipotesis

$$H_0 :\ \mu_d \leq 0$$

$$H_A :\ \mu_d > 0$$

- **Uji statistik :**

$$t = \frac{\bar{d} - \mu_d}{s_d}$$

- **Distribusi pada Uji Statistik :** Distribusi statistik uji t adalah dengan derajat kebebasan $df = n - 1$ jika H_0 benar

- **Kriteria Keputusan :** Dengan t tabel

$$\alpha = 0.05;\ \frac{\alpha}{2} = 0.025;\ df = n - 1 = 7 - 1 = 6$$

$$t_{\frac{\alpha}{2}}(df) = t_{0.025}(6) = -2.4469$$

- **Uji Statistik :**

$$\bar{d}\ = \frac{\sum d_i}{n} = \frac{-1011}{7} = -144.4285$$

$$d_i^2 = 190061$$

$$s_d^2 = n\frac{\sum d_i^2 - (\sum d_i)^2}{n(n-1)} = \frac{7(190061) - (-1011)^2}{7(6)} = -4.4580$$

- **Keputusan statistik :** Diperoleh $t_{hitung} = -4.4580 < t\ tabel = -2.4469$ maka H_0 ditolak

- **Kesimpulan :** Dapat disimpulkan bahwa jumlah sel CD4 T mengalami perubahan yang signifikan.

- ***p*-value :** $P(t \leq -4.4580) = 0.004$

Dengan menggunakan software R diperoleh:

```
Summary:
   AWAL          AKHIR
 Min.   : 58.0   Min.   :108.0
 1st Qu.:104.0   1st Qu.:199.0
 Median :169.0   Median :315.0
 Mean   :155.7   Mean   :300.1
 3rd Qu.:177.0   3rd Qu.:365.5
 Max.   :301.0   Max.   :549.0
Variansi:
F test to compare two variances
data:  CD4$AWAL and CD4$AKHIR
F = 0.27533, num df = 6, denom df = 6, p-value = 0.1417
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
```

```
 0.04730967 1.60235792
sample estimates:
ratio of variances
        0.2753307
```

Paired-test:

```
Welch Two Sample t-test
data:  CD4$AWAL and CD4$AKHIR
t = -2.2573, df = 9.0712, p-value = 0.9749
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -261.6139       Inf
sample estimates:
mean of x mean of y
 155.7143  300.1429
```

Dengan Syntax:

```
>summary(CD4)
>#var test
>?var.test
>var.test(CD4$AWAL,CD4$AKHIR,alternative=c("two.sided"),conf.level=0.95)
> t.test(CD4$AWAL,CD4$AKHIR,alternative=c("greater"),conf.level=0.95)
```

- Kesimpulan : p-value $= 0.9749 > 0.05$ maka H_0 ditolak yang artinya tidak ada perbedaan rata-rata antara jumlah sel CD4 Awal dan CD4 Akhir.

1.5. Uji Hipotesis : Populasi Proporsi Tunggal

Ketika sampel besar gunakan teorema limit pusat, uji statistiknya adalah

$$\boldsymbol{Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}}$$

Dimana ketika H_0 benar, distribusi di aproksimasi oleh distribusi normal standar.

LATIHAN : Populasi Proporsi Tunggal

Latihan 1.15. (Exercise 7.5.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Jacquemyn, dkk melakukan survei diantara dokter kandungan di wilayah Flanders dan memperoleh 295 tanggapan, 90 diantaranya menyatakan bahwa mereka melakukan setidaknya satu operasi Caesar sesuai permintaan setiap tahunnya. Apakah penelitian ini memberikan bukti yang cukup bagi kita yang menyimpulkan bahwa kurang dari 35% dari dokter kandungan di wilayah Flanders melakukan setidaknya 1 operasi Caesar sesuai permintaan setiap tahun. Gunakan tingkat singnifikan 0.05?

Penyelesaian :

- **Data :** Data mewakili 90 dari 295 tanggapan yang menunjukkan bahwa mereka telah melakukan setidaknya satu kali operasi Caesar sesuai permintaan setiap tahun
- **Asumsi :** Asumsinya adalah sampel sebanyak 295 tanggapan yang diambil dari populasi dan distribusi sampel proposi berdistribusi normal
- **Uji Hipotesis :** Untuk menguji hipotesis adalah bahwa dokter kandungan di wilayah Flanders melakukan minimal satu operasi Caesar sesuai permintaan setiap tahun kurang dari 35% akibatnya hipotesis satu arah

$$H_0 : p \geq 0.35$$

$$H_A : p < 0.35$$

- **Uji Statistik :**

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

- **Distribusi statistik :** Distribusi Statistik uji z berdistribusi normal dengan rata-rata nol jika H_0 benar
- **Kriteria Keputusan :** Kriteria Keputusan dengan z tabel $\alpha = 0.05$ diperoleh daerah kritisnya 1.645.
- **Statistik Hitung :**

$$\bar{p} = \frac{90}{295} = 0.3051$$

$$p_0 = 0.35$$

$$q_0 = 0.65$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

$$= \frac{0.3051 - 0.35}{\sqrt{\frac{0.35(0.65)}{n295}}}$$

$$= -1.6168$$

- **Keputusan statistik :** Keputusannya diperoleh $z_{hitung} = -1.6168 < z_{tabel} = -1.645$ maka H_0 ditolak.
- **Kesimpulan :** Dapat disimpulkan bahwa cukup bukti yang menunjukkan dokter kandungan di Flanders melakukan minimal 1 kali operasi Caesar setiap tahun sesuai permintaan kurang dari 35%. Hasil signifikan secara statistik.

1.6. Uji Hipotesis : Populasi Proporsi Ganda

Ketika $H0$ nya adalah $p_1 - p_2 = 0$, uji 2 populasi proporsi adalah sama.

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \text{ dan } \bar{q} = 1 - \bar{p}$$

Dimana x_1 dan x_2 adalah angka sampel pertama dan kedua. Pihak estimasi ini pada $p = p_1 = p_2$ gunakan perhitungan $\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}$, estimasi galat standarnya mengikuti :

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}$$

Uji statistiknya menjadi

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (\, p_1 - p_2)_0}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}}$$

Dimana diaproksimasi oleh distribusi normal standar, jika H_0 benar.

LATIHAN : Populasi Proporsi Ganda

Latihan 1.16. (Exercise 7.6.1, “BIOSTATISTICS Wayne W. Daniel Edisi 10”) Beberapa orang menggunakan wawancara telepon dari responden yang dipilih secara acak di Hongkong untuk mendapatkan informasi mengenai presepsi individu dengan kesehatan dan riwayat merokok. Diantara 1222 laki-laki saat ini, 72 melaporkan bahwa mereka memiliki kesehatan yang buruk atau sangat buruk, sementara 30 diantara 282 mantan perokok melporkan bahwa mereka memiliki kesehatan yang buruk atau sangat buruk. Apakah ini bukti yang cukup memungkinkan seseorang menyimpulkan bahwa diantara laki-laki Hongkong ada perbedaan antara perokok aktif dan mantan perokok sehubung dengan proporsi yang menganggap diri mereka memiliki kesehatan buruk dan sangat buruk? Gunakan tingkat signifikan 0.05.

Penyelesaian :

- **Data :** Mewakili 72 dari 1222 laki-laki perokok saat ini melaporkan bahwa mereka memiliki kesehatan yang buruk atau sangat buruk dan 30 dari 282 mantan perokok melaporkan bahwa mereka memiliki kesehatan yang buruk atau sangat buruk.

- **Asumsi :** Asumsinya adalah data diambil dari dua sampel acak sederhana independen yaitu perokok aktif dan mantan perokok aktif

- **Hipotesis :**

$$H_0 : p_1 = p_2$$

$$H_A : p_1 \neq p_2$$

- **Uji Statistik :**

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} \frac{\bar{p}(1-\bar{p})}{n_2}}}$$

- **Distribusi Statistik :** Distribusi Statistik uji z berdistribusi normal dengan rata-rata nol jika H_0 benar.

- **Kriteria keputusan :** Kriteria keputusan dengan z tabel

$$\alpha = 0.05$$

$$\frac{\alpha}{2} = 0.025$$

diperoleh daerah kritisnya 2.58

- **Statistik hitung :**

$$\bar{p}_1 = \frac{72}{1222} = 0.058920$$

$$\bar{p}_2 = \frac{30}{282}$$

$$= 0.106383$$

$$\bar{p} = \frac{(72 + 30)}{(1222 + 282)}$$

$$= 0.0685$$

$$\bar{q} = 1 - \bar{p}$$

$$= 1 - 0.0685$$

$$= 0.9315$$

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} \; \frac{\bar{p}(1-\bar{p})}{n_2}}}$$

$$= \frac{(0.058920 - 0.106383)}{\sqrt{\frac{(0.0685)(0.9315)}{1222} \; \frac{(0.0685)(0.9315)}{282}}}$$

$$= -2.86$$

- **Keputusan statistik :** Diperoleh $z_{hitung} = -2.86 < z\ tabel = 2.58$ maka H_0 ditolak.
- **Kesimpulan :** Dapat disimpulkan bahwa cukup bukti yang menunjukkan ada perbedaan antara perokok aktif dan mantan perokok sehubungan dengan proporsi yang menganggap diri mereka memiliki kesehatan yang buruk dan sangat buruk. Hasil signifikan secara statistik.
- ***p*-value** : p-value $= 0.004$.

1.7. Uji Hipotesis: Populasi Variansi Tunggal

Uji statistik untuk uji hipotesis mengenai populasi variansi adalah

$$X^2 = (n-1)s^2/\sigma^2$$

Dimana ketika H_0 benar, distribusi nya dengan derajat kebebasannya $n-1$.

Uji 1 pihak.

Untuk $H_A : \sigma^2 > {\sigma_0}^2$, H_0 ditolak jika $X^2 \geq x^2_{1-\alpha}$

Untuk $H_A : \sigma^2 < {\sigma_0}^2$, H_0 ditolak jika $X^2 \leq x^2_{\alpha}$

Contoh 1.10. (Example 7.7.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Tujuan studi oleh Wilkinis dkk untuk mengukur efektifitas rekombinan human growth hormon (rhGH) pada anak dengan total luas permukaan tubuh dengan luka bakar $> 40\%$. Dalam penelitian ini untuk mengetahui level faktor pertumbuhan hormon insulin (IGF-I) saat ini sebelum pemberian rhGH sampel variansi kadar IGF-I (dalam mg/ml) adalah 670.81. Dapatkah disimpulkan dari data bahwa variansi populasi tidak sama dengan 600?

Penyelesaian :

- **Asumsi :** Sampel penelitian merupakan sampel acak sederhana dari populasi anak dengan jenis tingkat IGF-I didistribusikan secara normal.
- **Hipotesis :**

$$H_0 : \sigma^2 = 600$$
$$H_1 : \sigma^2 \neq 600$$

  Jika hipotesis H_0 benar maka di distribusikan sebagai χ^2 dengan $n-1$ derajat kebebasan $\alpha = 0.05$, nilai kritis $\chi^2 = 6.262$ dan 27.408.

- **Statistik Hitung :**

$$\chi^2 = \frac{(16-1)670.81}{600} = 16.77$$

$$H_0 : 6.262 < 16.77 < 27.477$$

- **Kesimpulan :** Berdasarkan data diatas dapat disimpulkan bahwa variansi populasi tidak sama dengan 600.

1.8. Uji Hipotesis : Rasio pada 2 Populasi Variansi

Uji Rasio Variansi

Keputusan pada dua populasi variansi biasanya menggunakan uji rasio variansi, yang mana uji pada H_0 dua populasi variansinya sama ketika uji hipotesis dengan dua variansinya sama, rasio pada uji hipotesisnya sama dengan 1.

Jika hipotesis ${\sigma_1}^2 = {\sigma_2}^2$ asumsikan hipotesis benar dan dua variansi ${s_1}^2/{s_2}^2$ mengikuti distribusi F.

Contoh 1.11. (Example 7.8.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Broden membagikan teknik perbaikan menisial menggunakan spiesmen lurut kadaxer salah satu variable yang menarik adalah beban saat gagal (dalam Newton) untuk lurut yang diperbaiki dengan teknik Fast-Ftik. Kelompok 1 dengan metode jahitan vertikal, kelompok 2 setiap teknik dilapiskan pada setiap 6 spiesman SD untuk metode Fast-Ftik adalah 30.62, SD untuk metode vertikal adalah 11.37. Dapatkah disimpulkan bahwa secara umum variansi beban saat kegagalan lebih tinggi untuk Fast-Ftik dari pada vertikal?

Penyelesaian :

- **Asumsi :** Pengamatan dari setiap ab merupakan sampel independen acak urutan n diambil dari populasi yang ditentukan kombinasi tingkat tertentu dari kedua factor tersebut. Setiap populasi ab terdistribusi normal. Semua populasi memiliki variansi yang sama.

- **Hipotesis :**

$$H_0 : \alpha_i = 0$$
$$H_A : \exists\alpha_i \neq 0$$

$$H_0 : \beta_i = 0$$
$$H_1 : \exists\beta_i \neq 0$$

$$H_0 : (\alpha\beta)_i = 0$$
$$H_1 : \exists(\alpha\beta)_i \neq 0$$

- **Distribusi uji statistik :** Ketika H_0 benar dan asumsi terpenuhi, masing-masing uji statistik di distribusikan sebagai F

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

- **Statistik Hitung :**

$$VR = \frac{s_1^2}{s_2^2} = \frac{(30.62)^2}{(11.37)^2} = 7.25$$

- **Kesimpulan :** H_0 ditolak karena $7.25 > 5.05$. Variabilitas beban kegagalan lebih tinggi saat menggunakan metode Fast-Ftik dari pada jahitan.

BAB II

ANALISIS VARIANSI

2.1. Pendahuluan

Model Linear : Model stokastik adalah representasi matematis dari hubungan antar variabel.

Definisi 2.1. Hasil Variabel diwakili oleh kumpulan nilai terukur yang dihasilkan dari percobaan atau proses stokastik lainnya.

Model linear adalah model apapun yang linear dalam parameter yang menentukan model. Bentuk model linear secara umum

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_k X_{kj} + \varepsilon_j$$

Dengan
β_j : koefisien adalah model
ε_j : mewakili random error

Analisis Variansi : Didefinisikan sebagai teknik dimana variansi total yang ada dalam satu set tangga yang dipartisi menjadi dua / lebih komponen. Terkait dengan masing-masing komponen ini adalah sumber variansi tertentu, sehingga dalam analisis dapat dipastikan besarnya kontribusi masing-masing sumber ini terhadap total upah.

Aplikasi : Analisis variansi menemukan aplikasi terluasnya dalam analisis data yang berasal dari eksperimen.

2.2. Rancangan Acak Lengkap (*Complete Randomized Design*)

ANOVA Satu Arah (*One-way ANOVA*) : Jenis analisis variansi yang paling sederhana adalah yang dikenal sebagai analisis variansi satu arah dimana hanya satu sumber variansi/faktor yang diselidiki.

Analisis satu arah digunakan untuk menguji hipotesis nol bahwa tiga / lebih perlakuan sama efektifnya. Percobaan yang diperlukan untuk dirancang sedemikian rupa sehingga perlakuan yang diminati diberikan secara acak ke subjek / objek yang akan digunakan untuk mengukur keefektifan pengobatan. Untuk alasan inilah rancangan tersebut dinamakan Rancangan Acak Lengkap (RAL).

Langkah-langkah pengujian hipotesis (setelah diputuskan RAL) yaitu :

1. **Data :** Pengukuran (observasi) yang dihasilkan dari rancangan percobaan anak lengkap.
2. **Asumsi :** Model adalah representasi simbolis dari nilai tipikal kumpulan data istilah

Dalam model ini didefinisikan sebagai berikut :

1. μ : Mean dari $\forall_k$ mean populasi dan disebut grand mean
2. τ_j : Selisih mean populasi ke-j dan mean disebut efek pengobatan (*treatmen effect*)
3. ε_{ij} : Jumlah perbedaan pengukuran individu dari rata-rata populasi tempatnya berada dan disebut istilah kesalahan/eror.

Pengamatan khas dari total kumpulan data yang diteliti terdiri dari :

1. Rata-rata utama
2. Efek perlakuan
3. Istilah kesalahan yang mempresentasikan deviasi observasi dari deviasi dari grup mean

Jumlah Kuadrat Total (*Sum Square Total*) : Pertama-pertama harus mendapatkan jumlah kuadrat total. Jumlah kuadrat total adalah jumlah kuadrat penyimpanan pengamatan individu dari rata-rata semua pengamatan individu dari rata-rata semua pengamatan yang diambil bersama.

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{nj}\left(\bar{X}_{ij} - \bar{X}\right)^2$$

Jumlah Kuadrat Perlakuan (*The Within Group Sum of Squares*)

$$SSW = \sum_{j=1}^{k}\sum_{i=1}^{nj}\left(X_{ij} - \bar{X}_j\right)^2$$

Jumlah Kuadrat Galat (*The Among Groups Sum of Squares*)

$$SSA = \sum_{j=1}^{k} n_j \left(X_j - \bar{X}\right)^2$$

Estimator pertama dari $\boldsymbol{\sigma^2}$ dalam sampel apapun

$$\frac{\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)^2}{n_j - 1}$$

Menghasilkan Kuadrat Tengah Perlakuan,

$$MSW = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)^2}{\sum_{j=1}^{k}\left(n_j - 1\right)}$$

Estimator kedua dari $\boldsymbol{\sigma^2}$

$$\sigma^2 = n\,\sigma_{\bar{X}}{}^2$$

Dengan estimasi tidak bias pada $\sigma_{\bar{X}}{}^2$ dihitung dari data sampel,

$$\frac{\sum_{i=1}^{k}\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2}{k-1}$$

Menghasilkan Kuadrat Tengah Galat,

$$MSA = \frac{n\sum_{i=1}^{k}\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2}{k-1}$$

Ketika ukuran sampel tidak sama semua maka estimasinya,

$$MSA = \frac{\sum_{i=1}^{k} n_j\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2}{k-1}$$

Rasio Variansi : Membandingkan dua perkiraan ini

$$VR = \frac{among\ groups\ mean\ square}{within\ groups\ mean\ square} = \frac{MSA}{MSW} = \frac{KTG}{KTP}$$

Uji $\boldsymbol{F}$: Uji F harus memperhatikan distribusi sampel dari rasio 2 variansi sampel. Dalam bab ini dapat melihat bahwa distribusi fundamental untuk analisis variansi. Untuk alasan ini rasio yang kami tunjuk pada VR sering disebut sebagai F dan prosedur pengujiannya sering disebut $Uji\ F$. Distribusi F adalah rasio dari dua distribusi *chi-square.*

Contoh 2.1 (Example 8.2.1, “BIOSTATISTICS Wayne W. Daniel Edisi 10”) Daging buruan yang berasal dari rusa berekor dan tupai timur digunakan sebagai makanan oleh keluarga pemburu dan individu lain untuk alasan kesehatan, budaya maupun pribadi. Sebuah studi oleh David Holben yang diperoleh dari Amerika Serikat. Nilai kandungan selenium ini juga dibandingkan dengan daging sapi yang diproduksi didalam dan luar daerah yang sama. Apakah tingkat selenium berbeda diantara empat kelompok daging?

Penyelesaian :

- **Hipotesis :**

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_1 : ada\ salah\ satu\ tidak\ sama$$

- **Uji Statistik :**

$$VR = \frac{SSA}{SSW}$$

- **Distribusi Uji Statistik :** Jika H_0 benar dan asumsi terpenuhi VR mengikuti distribusi F dengan $4 - 1 = 3$ derajat kebebasan dan $194 - 4 = 140$ derajat kebebasan penyebut.

- **Kriteria Keputusan :** Misalkan bahwa $\alpha = 0.01$ nilai kritis F dari tabel uji F diperoleh < 3.95. Aturan keputusan menolak H_0 jika $VR > 3.95$

- **Uji Statistik Hitung :**

$$SST = 58009.05560$$
$$SSA = 21261.82886$$
$$SSW = 36747.22674$$

- **Keputusan statistik :** Karena F yang dihitung dari 27.00 lebih besar maka menolak H_0

- **Kesimpulan :** Kerena menolak H_0 maka dapat disimpulkan bahwa rata-rata dari keempat jenis daging tidak semuanya memiliki kandungan selenium yang sama.

- ***p*-value** : p-value $27.00 > 3.95$; $p < 0.01$

LATIHAN : RANCANGAN ACAK LENGKAP (RAL)

Latihan 2.1 (Exercise 8.2.2, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Penderita penyakit rematik (*osteoporosis*) seringkali mengalami BMD. Alendronate adalah salah satu obat yang diresepkan untuk membangun atau mencegah hilangnya BMD lebih lanjut. Holcomb dan Rothenberg mengamati 96 wanita yang menggunakan alendronate untuk menentukan apakah ada perbedaan ada dalam persentase rata-rata perubahan BMD di antara lima klasifikasi diagnosis primer yang berbeda. Kelompok 1 pasien didiagnosis dengan rheumatoid arthritis (RA). Pasien kelompok 2 adalah campuran kumpulan pasien dengan penyakit termasuk lupus, granulomatosis Wegener dan poliarteritis, dan penyakit vaskulitik lainnya (LUPUS). Pasien kelompok 3 mengalami polymyalgia rheumatica atau temporal arthritis (PMRTA). Pasien kelompok 4 menderita osteoartritis (OA) dan kelompok 5 menderita osteoporosis (O) tanpa penyakit rematik lain yang teridentifikasi dalam rekam medis. Perubahan BMD ditunjukkan pada tabel berikut.

$LUPUS$	$PMRTA$	OA	O
7,26	10,734	3,544	8,992
5,546	1,399	4,16	6,12
2,961	0,497	1,16	25,655
0,293	0,592	−0,247	2,937
8,394	3,95	5,372	15,968
2,832	0,674	6,721	5,349
−1,369	9,354	9,95	1,719
11,288	0,61	10,82	6,445
3,997	5,682	7,28	20,243
	−3,669	60605	3,29
	−7,816	7,507	
	4,563	0,163	
	−0,093	12,767	
	−0,185	3,481	
	1,302	0,917	
	5,299	15,853	
		11,146	
		−0,838	
		4,082	
		6,645	
		4,329	
		1,234	
		−2,817	
		5,075	

Penyelesaian :

Dengan menggunakan software R :

```
Syntax:
#impor data
>soal1.822<-read.table("E:/soal1.csv",header=TRUE,sep=";",na.strings="NA",dec=",",
strip.white=TRUE)
>diagnosis=soal1.822
>diagnosis
#uji normalitas shapiro.test
>by(data=diagnosis$alendronate,INDICES=diagnosis$kelompok,FUN=shapiro.test)
```

Diperoleh hasil :

```
diagnosis$kelompok: 1

        Shapiro-Wilk normality test

data:  dd[x, ]

W = 0.96891, p-value = 0.3795

---------------------------------------------------------

diagnosis$kelompok: 2

        Shapiro-Wilk normality test

data:  dd[x, ]

W = 0.98446, p-value = 0.9832

---------------------------------------------------------

diagnosis$kelompok: 3

        Shapiro-Wilk normality test

data:  dd[x, ]

W = 0.94691, p-value = 0.4424

---------------------------------------------------------

diagnosis$kelompok: 4
```

```
Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.20927, p-value = 2.196e-10
------------------------------------------------------------
diagnosis$kelompok: 5

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.85386, p-value = 0.06455
```

Menentukan Normalitas Data Secara Manual :

Dari data di atas (Lupus, PMRTA, dan OA) akan diidentifikasi secara manual untuk menentukan normalitas data tersebut.

Data Lupus

Langkah 1 : Merumuskan hipotesis

H_0 : Data berdistribusi normal

H_A : Data tidak berdistribusi normal

Langkah 2 : Menentukan nilai uji statistik

Jangkauan (J) = data terbesar – data terkecil

Banyak kelas $(k) = 1 + 3.3 \log n$

Panjang kelas $= J : k$

Selanjuatnya data diatas digunakan untuk membuat tabel berikut :

Data (dalam interval)	Nilai Tengah (x_i)	Frekuensi (f_i)	$f_i x_i$	x_i^2	$f_i x_i^2$
$(-1.369) - 1.1624$	-0.1033	2	-0.2066	0.010671	0.021342
$1.1624 - 3.6938$	2.4281	2	4.8562	5.89567	11.79134
$3.6938 - 6.2252$	4.9595	2	9.919	24.59664	49.19328
$6.2252 - 8.7566$	7.4909	1	7.4909	56.11358	56.11358
$8.7566 - 11.288$	10.0223	2	20.0446	100.4465	200.893
Jumlah		9	42.1041	187.0631	318.0125

Selanjutnya mencari rata-rata dan standar deviasi

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{42.1041}{9}$$

$$= 4.678233$$

$$SD = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

$$= \sqrt{\frac{318.0125}{9} - \left(\frac{42.1041}{9}\right)^2}$$

$$= 3.667268$$

Selanjutnya membuat dan melengkapi table berikut, sekaligus perhitungan cara melengkapinya :

Data	f_i	b_k bawah	Nilai z	b_k atas	luas tiap kelas	E_i	chi square
$-1.369\ -\ 1.1624$	2	-1.869	-1.78532 dan -0.82236	1.6624	0.1686	1.5174	0.153488
$1.1624\ -\ 3.6938$	2	0.6624	-1.09505 dan -0.1321	4.1938	0.3104	2.7936	0.225444
$3.6938\ -\ 6.2252$	2	3.1938	-0.40478 dan 0.558172	6.7252	-0.0569	-0.5121	-12.3231
$6.2252\ -\ 8.7566$	1	5.7252	0.285489 dan 1.24844	9.2566	-0.2841	-2.5569	-4.948
$8.7566\ -\ 11.288$	2	8.2566	0.975758 dan 1.938709	11.788	-0.1373	-1.2357	-8.47273
$Jumlah$							-25.3649

Tabel B

Bagaimana cara mendapatkan nilai z pada table diatas :

$$Z = \frac{batas\ kelas - \bar{X}}{SD}$$

Kita ambil batas kelas (b_k) pada baris pertama yaitu $1.6624 - (-1.869)$

Untuk batas kelas bawah (-1.869)

$$Z = \frac{batas\ kelas - \bar{X}}{SD} = \frac{-1.869 - 4.678233}{3.667268} = -1.78532$$

Untuk batas kelas bawah (1.6624)

$$Z = \frac{batas\ kelas - \bar{X}}{SD} = \frac{1.6624\ - 4.678233}{3.667268} = -0.82236$$

Lakukan seterusnya untuk nilai yang lain.

Cara mendapatkan luas tiap kelas interval pada table B diatas

z_1	z_2	$0 - z_1$	$0 - z_2$	$luas$	$0 - z$	$luas\ tiap\ kelas$
-1.78532	-0.82236	1.785316	0.822365	0.0375	0.2061	0.1686
-1.09505	-0.1321	1.095047	0.132096	0.1379	0.4483	0.3104
-0.40478	0.558172	0.404779	-0.55817	0.3446	0.2877	-0.0569
0.285489	1.24844	-0.28549	-1.24844	0.3897	0.1056	-0.2841
0.975758	1.938709	-0.97576	-1.93871	0.1635	0.0262	-0.1373

Jadi setelah memperoleh nilai z (z_1 dan z_2) kemudian cari luas dengan rumus $0 - z$ menggunakan table z.

Bagaimana mencari frekuensi yang diharapkan pada table kedua diatas?

Kita ambil luas tiap kelas interval pada baris pertama :

$$E_i = Luas\ tiap\ kelas\ interval \times n(jumlah\ responden) = 0.0375 \times 9 = 0.1686$$

Lakukan untuk baris ke-2 dan seterusnya

Pada tahap ini kita telah menyelesaikan table 3.

Langkah 3 : Menentukan taraf nyata

$$\chi^2_{tabel} = \chi^2_{(1-\alpha)(dk)}$$

Mengikuti langkah-langkah berikut :

Derajat kebebasan (df) dengan rumus

$$df = banyaknya\ kelas - 3 = 5 - 3 = 2$$

Taraf signifikansi $\alpha = 0.05$

Kita lihat pada table χ^2 untuk $\chi^2_{(0.95)(2)} = 12.5916$

Langkah 4 : Menentukan kriteria pengujian hipotesis

$$H_0\ ditolak\ jika\ \chi^2_{hitung} \geq \chi^2_{tabel}$$

$$H_0\ diterima\ jika\ \chi^2_{hitung} < \chi^2_{tabel}$$

Berdasarkan perhitungan pada table 3 diperoleh nilai

$$\chi^2_{hitung} = -25.3649$$

Dan

$$\chi^2_{tabel} = 12.5916$$

Maka H_0 $diterima$

Langkah 5 : Memberikan Kesimpulan

Karena H_0 $diterima$ artinya data Lupus pada table soal diatas memiliki distribusi normal, dan hasil ini sejalan dengan pengujian menggunakan R.

Data PMRTA

Langkah 1 : Merumuskan hipotesis

H_0 : Data berdistribusi normal
H_A : Data tidak berdistribusi normal

Langkah 2 : Menentukan nilai uji statistic

Jangkauan (J) = data terbesar – data terkecil
Banyak kelas $(k) = 1 + 3.3 \log n$
Panjang kelas $= J : k$

Selanjuatnya data diatas digunakan untuk membuat table berikut :

$Data$ $(dalam\ interval)$	$Nilai$ $Tengah$ (x_i)	$frekuensi$ (f_i)	$f_i x_i$	x_i^2	$f_i x_i^2$
$-7.816\ dan - 4.0863$		1	-5.9511521	35.41621	35.41621
$-4.0863\ dan - 0.35660$		1	-2.2214563	4.934868	4.934868
$-0.35660\ dan\ 3.3730$		8	12.0659164	2.274787	18.19829
$3.373087\ dan\ 7.10278$		4	20.9517415	27.43597	109.7439
$7.10278\ dan\ 10.8324$		2	17.9352624	80.41841	160.8368
$Jumlah$		16	42.7803119	150.4802	329.1301

Selanjutnya mencari rata-rata dan standar deviasi

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{42.7803119}{16}$$

$$= 2.6737$$

$$SD = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

$$= \sqrt{\frac{329.1301}{16} - \left(\frac{42.780311}{16}\right)^2}$$

$$= 3.663548$$

Selanjutnya buat dan lengkapi table berikut, sekaligus perhitungan cara melengkapinya :

Data	f_i	$\mathrm{b_k}$ bawah	Nilai z	$\mathrm{b_k}$ atas	luas tiap kelas	E_i	chi square
$-1.369 - 1.1624$	-4.13796	-8.316	-2.99 *dan* -1.70	-3.58	-0.0432	-0.6912	-3.5863042
$1.1624 - 3.6938$	-5.82175	-4.5863	-1.98 *dan* -0.69	0.14	-0.2212	-3.5392	0.14339164
$3.6938 - 6.2252$	-38.7135	-0.8566	-0.96 *dan* 0.32	3.87	-0.206	-3.296	3.87308746
$6.2252 - 8.7566$	-10008	2.8730	1.34 *dan* 1.345	7.60	-0.0001	-0.0016	7.60278328
$8.7566 - 11.288$	0.008077	6.60278	1.07 *dan* 2.36	11.33	0.1332	2.1312	11.3324791
Jumlah							-10056.7

Bagaimana cara mendapatkan nilai z pada table diatas :

$$Z = \frac{batas\ kelas - \bar{X}}{SD}$$

Kita ambil batas kelas (b_k) pada baris pertama yaitu

Untuk batas kelas bawah (-8.31)

$$Z = \frac{batas\ kelas - \bar{X}}{SD} = \frac{-8.31 - 2.6782}{3.667} = -2.99$$

Untuk batas kelas bawah (-3.58)

$$Z = \frac{batas\ kelas - \bar{X}}{SD} = \frac{-3.58\ - 2.67}{3.667} = -1.70$$

Lakukan seterusnya untuk nilai yang lain.

Cara mendapatkan luas tiap kelas interval pada tabel B diatas

$0 - z_1$	$0 - z_2$	$luas$ $0 - z_1$	$luas$ $0 - z_2$	$luas\ tiap\ selisih \left(\begin{matrix} luas\ 0 - z_1\ dan \\ luas\ 0 - z_2 \end{matrix} \right)$	E_i	$chi\ square$
2.999761	1.708746	0.0014	0.0446	-0.0432	-0.6912	-4.13796
1.981706	0.69069	0.0239	0.2451	-0.2212	-3.5392	-5.82175
0.96365	-0.32737	0.1685	0.3745	-0.206	-3.296	-38.7135
-1.34542	-1.34542	0.09	0.0901	-0.0001	-0.0016	-10008
-1.07246	-2.36348	0.1423	0.0091	0.1332	2.1312	0.008077
						-10056.7

Jadi setelah memperoleh nilai $z\ (z_1 \text{ dan } z_2)$ kemudian cari luas dengan rumus $0 - z$ menggunakan table z.

Bagaimana mencari frekuensi yang diharapkan pada table kedua diatas?

Kita ambil luas tiap kelas interval pada baris pertama:

$$E_i = Luas\ tiap\ kelas\ interval \times n(jumlah\ responden)$$

$$= 0.0375 \times 9$$

$$= 0.1686$$

Lakukan untuk baris ke-2 dan seterusnya

Pada tahap ini kita telah menyelesaikan tabel 3.

Langkah 3 : Menentukan taraf nyata

$$\chi^2_{tabel} = \chi^2_{(1-\alpha)(dk)}$$

Mengikuti langkah-langkah berikut :

Derajat kebebasan (df) dengan rumus

$$df = banyaknya\ kelas - 3$$

$$= 5 - 3$$

$$= 2$$

Taraf signifikansi $\alpha = 0.05$

Kita lihat pada table χ^2 untuk $\chi^2_{(0.95)(2)} = 12.5916$

Langkah 4 : Menentukan kriteria pengujian hipotesis

$$H_0\ ditolak\ jika\ \chi^2_{hitung} \geq \chi^2_{tabel}$$

$$H_0\ diterima\ jika\ \chi^2_{hitung} < \chi^2_{tabel}$$

Berdasarkan perhitungan pada tabel 3 diperoleh nilai

$$\chi^2_{hitung} = -25.3649$$

Dan

$$\chi^2_{tabel} = 22.36$$

Maka $H_0\ diterima$

Langkah 5 : Memberikan Kesimpulan

Karena $H_0\ diterima$ artinya data PMRTA pada tabel soal diatas memiliki distribusi normal, dan hasil ini sejalan dengan pengujian menggunakan R.

Analisis Uji Normalitas

Uji normalitas terlebih dahulu menggunakan Shapiro-Wilk. Perhatikan nilai p-value pada kandungan kalsium terhadap kelompok umur dengan $\alpha = 0.05$.

H_0 : Berdistribusi Normal

H_A : Tidak Berdistribusi Normal

Daerah kritis : Tolak H_0 ditolak jika p-value < 0.05

Keputusan : H_0 gagal ditolak untuk kelompok 1, 2, 3, dan 5 karena p-value > 0.05, sedangkan H_0 ditolak untuk kelompok 4.

Kesimpulan : pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa data mengikuti sebaran normal (berdistribusi normal) untuk perlakuan 1, 2, 3, dan 5.

Kemudian Uji ANOVA

```
Syntax:

>ujianova=aov(alendronate~as.factor(kelompok),data=soal1.822)
>ujianova
>summary(ujianova)
```

Diperoleh hasil :

```
Call:
   aov(formula = alendronate ~ as.factor(kelompok), data =
soal1.822)

Terms:
                as.factor(kelompok)  Residuals
Sum of Squares            114804178 3519331866
Deg. of Freedom                   4         91

Residual standard error: 6218.84
Estimated effects may be unbalanced

SUMMARY:
Df     Sum Sq  Mean Sq F value Pr(>F)
as.factor(kelompok)  4  114804178 28701044   0.742  0.566
Residuals           91 3519331866 38673977
```

ANALISIS ANOVA

Perhatikan nilai p-value pada kandungan kalsium terhadap kelompok umur dengan $\alpha = 0.05$.

H_0 : Tidak ada perbedaan (tidak ada pengaruh kelompok umur terhadap kandungan kalsium)
H_A : Terdapat perbedaan (terdapat pengaruh kelompok umur terhadap kandungan kalsium)

Daerah kritis : Tolak H_0 ditolak jika p-value < 0.05
Keputusan : H_0 gagal ditolak karena p-value $= 0.566 > 0.05$

Kesimpulan : pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa tidak ada pengaruh kelompok terhadap alendronate.

Latihan 2.2. (Exercise 8.2.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Ilich-Ernst dkk. menyelidiki asupan makanan yang mengandung kalsium di antara 113 wanita sehat usia $20 - 88$ tahun. Para peneliti membentuk empat kelompok usia sebagai berikut : Kelompok A, $20.0 - 45.9$ tahun; kelompok B, $46{,}0 - 55{,}9$ tahun; kelompok C, $56{,}0 - 65{,}9$ tahun; dan kelompok D, lebih dari 66 tahun (diukur dalam $mg/hari$). Berikut datanya :

Penyelesaian :

kandungan_kalsium	kelompok_umur
1820	1
2588	1
2670	1
1022	1
1555	1
222	1
1197	1
1249	1
1520	1
489	1
2575	1
1426	1
1846	1
1088	1
912	1
1383	1
1483	1

1723	1
727	1
1463	1
1777	1
1129	1
191	2
1098	2
644	2
136	2
1605	2
1247	2
1529	2
1422	2
445	2
990	2
489	2
2408	2
1064	2
629	2

724	3
613	3
918	3
949	3
877	3
1368	3
1692	3
697	3
849	3
1199	3
429	3
798	3
631	3
1016	3
1025	3
948	3
1085	3
775	3
1307	3
344	3
961	3
239	3
944	3
1096	3
1020	3
805	3
361	3
641	3
760	3
1652	4
1309	4
1002	4
966	4
788	4
472	4
471	4
771	4
869	4
513	4

731	4
1130	4
1034	4
1261	4
42	4
767	4
752	4
804	4
1182	4
1243	4
985	4
1295	4
1676	4
754	4
775	4
1393	4
533	4
734	4
485	4
449	4
236	4
831	4
698	4
167	4
824	4
448	4
991	4
590	4
994	4
1781	4
937	4
1022	4
1073	4
948	4
222	4
721	4
375	4
1187	4

Dengan :

Kelompok 1 = A ; Kelompok 2 = B; Kelompok 3 = C; dan Kelompok 4 = D.

```
Syntax:

#inputdata

>latihan8.2.3<-read.table("E:/latihan823.csv",header=TRUE,sep="",na.strings="NA",
dec=".", strip.white=TRUE)

>kalsium=latihan8.2.3

>by(data=kalsium$kandungan_kalsium,INDICES=kalsium$kelompok_umur,FUN=shapiro.t
est)
```

Diperoleh:

UJI NORMALITAS

```
kalsium$kelompok_umur: 1

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.95785, p-value = 0.4471

--------------------------------------------------------
kalsium$kelompok_umur: 2

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.95343, p-value = 0.6152

--------------------------------------------------------
kalsium$kelompok_umur: 3

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.9766, p-value = 0.7464

--------------------------------------------------------
kalsium$kelompok_umur: 4

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.98311, p-value = 0.7118
```

Analisis Uji Normalitas

Uji normalitas terlebih dahulu menggunakan Shapiro-Wilk. Perhatikan nilai p-value pada kandungan kalsium terhadap kelompok umur dengan $\alpha = 0.05$.

H_0 : Berdistribusi Normal

H_A : Tidak Berdistribusi Normal

Daerah kritis : Tolak H_0 ditolak jika P-Value < 0.05

Keputusan : H_0 gagal ditolak untuk semua hasil percobaan karena p-value > 0.05

Kesimpulan : pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa data mengikuti sebaran normal (berdistribusi normal) untuk semua perlakuan

Kemudian Uji ANOVA

```
Syntax:

>ujianova=aov(kandungan_kalsium~as.factor(kelompok_umur),data=latihan8.2.3)
>ujianova
>summary(ujianova)
```

Diperoleh hasil:

```
Call:
   aov(formula = kandungan_kalsium ~ 
as.factor(kelompok_umur), data = latihan8.2.3)

Terms:
                as.factor(kelompok_umur) Residuals
Sum of Squares                   5996359  23228013
Deg. of Freedom                        3       109

Residual standard error: 461.6287
Estimated effects may be unbalanced

SUMMARY:
Df   Sum Sq Mean Sq F value    Pr(>F)
as.factor(kelompok_umur)
3  5996359 1998786    9.38 0.0000144 ***
Residuals             109 23228013  213101
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 
1
```

ANALISIS ANOVA

Perhatikan nilai p-value pada kandungan kalsium terhadap kelompok umur dengan $\alpha = 0.05$.

H_0 : Tidak ada perbedaan (tidak ada pengaruh kelompok umur terhadap kandungan kalsium)

H_A : Terdapat perbedaan (terdapat pengaruh kelompok umur terhadap kandungan kalsium)

Daerah kritis : Tolak H_0 ditolak jika p-value < 0.05

Keputusan : H_0 ditolak karena p-value $= 0.0000144 < 0.05$

Kesimpulan : Pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa terdapat pengaruh kelompok umur terhadap kandungan kalsium.

Ketika kesimpulan dalam uji ANOVA adalah H_0 ditolak maka perlu dilakukan uji lanjut untuk menentukan perlakuan mana yang memiliki perbedaan. Salah satu uji lanjut, gunakan uji TUKEY HSD,

Syntax:

```
> TukeyHSD(ujianova)
```

Diperoleh hasil:

```
Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = kandungan_kalsium ~ as.factor(kelompok_umur), data = latihan8.2.3)

$`as.factor(kelompok_umur)`
          diff       lwr        upr     p adj
2-1 -455.72078 -867.4967  -43.94486 0.0238691
3-1 -583.84639 -924.3799 -243.31292 0.0001107
4-1 -596.63447 -906.7347 -286.53422 0.0000120
3-2 -128.12562 -520.0987  263.84749 0.8289312
4-2 -140.91369 -506.7580  224.93058 0.7468424
4-3  -12.78807 -296.0647  270.48859 0.9994122
```

Analisis Uji Lanjut

Pada Uji Tukey HSD, perhatikan p-value (perhatikan paling kiri) untuk keputusan.

Dengan hipotesis :

H_0 : Tidak ada perbedaan tiap pasang kelompok

H_A : Terdapat perbedaan tiap pasang kelompok

Daerah kritis : H_0 ditolak jika p-value < 0.05

Keputusan :

H_0 ditolak untuk $3-1$ (Terdapat perbedaan antara kelompok C dan A)

H_0 ditolak untuk $4-1$ (Terdapat perbedaan antara kelompok D dan A)

H_0 gagal ditolak untuk $3-2$ (Tidak ada perbedaan antara kelompok C dan B)

H_0 gagal ditolak untuk $4-2$ (Tidak ada perbedaan antara kelompok D dan B)

H_0 gagal ditolak untuk $4-3$ (Tidak ada perbedaan antara kelompok D dan C)

Kesimpulan : Pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan untuk kelompok $3-1$ dan $4-1$ terdapat perbedaan antar kelompok sedangkan untuk kelompok $3-2$, $4-2$, dan $4-3$ tidak ada perbedaan antar kelompok.

2.3. Rancangan Acak Kelompok Lengkap

Tujuan penggunaan RAKL adalah untuk mengisolasi dan hapus dari istilah kesalahan variansi yang disebabkan oleh kelompok, dengan memastikan itu cara penyelesaian akan bebas dari efek blok. Efektivitas desain tergantung pada kemampuan untuk mencapai percobaan unit kelompok yang homogen.

Keuntungan RAKL yaitu mudah dimengerti, selain itu komplikasi tertentu yang mungkin timbul dalam perjalanan penyakit percobaan mudah ditangani saat desain ini digunakan.

Tampilan data pada umumnya data hasil eksperimen memanfaatkan data rancangan acak kelompok lengkap.

Total dari kelompok ke-i : $T_{i\cdot} = \sum_{j=1}^{k} X_{ij}$

Rata-rata dari kelompok ke-i : $\bar{X}_{i\cdot} = \frac{\sum_{j=1}^{k} X_{ij}}{k} = \frac{T_{i\cdot}}{k}$

Total keseluruhan : $T_{..} = \sum_{j=1}^{k} T_{.j} = \sum_{j=1}^{k} T_{i.}$

ANOVA-2 Arah : Teknis nganalisis data dari Rancangan Acak Kelompok Lengkap disebut analisis variansi 2 arah.

Langkah-langkah untuk uji hipotesis untuk RAKL :

1. **Data :** Setelah identifikasi perlakuan, kelompok, dan unit percobaan.
2. **Asumsi :** Asumsi untuk RAKL dan asumsinya mengikuti :

   Model

   X_{ij} : $\mu + \beta_i + \tau_j + \varepsilon_{ij}$, $i = 1, 2, \dots, n.\ j = 1, 2, \dots, k$

   Dengan

   X_{ij} : Nilai tipikal dari keseluruhan populasi
   μ : Konstanta yang tidak diketahui
   β_i : Efek blok, fakta bahwa unit percobaan jatuh pada kelompok ke – i
   τ_j : Efek perlakuan, fakta unit percobaan diterima pengobatan ke – j
   ε_{ij} : Komponen residual, semua sumber varians selain perlakuan dan kelompok

   Asumsi pada Model

   a. Setiap X_{ij} yang diamati merupakan sampel bebas acak berukuran 1 dari populasi kn yang diwakili.

   b. Masing-masing populasi kn ini berdistribusi normal dengan mean m_{ij} dan sama varian σ^2. ε_{ij} didistribusikan secara independen dan normal dengan mean 0 dan variansi σ^2.

   c. .

   $$\sum_{j=1}^{k} \tau_j = \sum_{i=1}^{n} \beta_i = 0$$

3. **Hipotesis :**

   H_0 : $\tau_j = 0,$ $j = 1, 2, \dots, k$

   H_A : $tidak\ semua\ \tau_j = 0$

4. **Uji Statistik :** Uji statistiknya adalah $V.R$

5. **Distribusi statistik uji :** Ketika H_0 benar dan asumsi $V.R$ terpenuhi mengikuti distribusi F

6. **Keputusan :** H_0 ditolak jika nilai hitung uji statistik $V.R \geq$ nilai kritis F

7. **Hitung Uji Statistik :** Jumlah Kuadrat Total terdiri dari 3 komponen yaitu

$$SST = SSBI + SSTr + SSE$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_{..}\right)^2$$

$$SSBI = \sum_{j=1}^{k}\sum_{i=1}^{n}(\bar{X}_{i.} - \bar{X}_{..})^2$$

$$SSTr = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2$$

$$SSE = SST - SSBI - SSTr$$

Derajat bebas untuk setiap komponen,

$$\overbrace{kn-1}^{total} = \overbrace{(n-1)}^{kelompok} + \overbrace{(k-1)}^{perlakuan} + \overbrace{(n-1)(k-1)}^{\frac{galat}{eror}}$$

Asal dari derajat bebas galat :

$$\begin{aligned}(kn-1)-(n-1)-(k-1) &= kn - n - k + 1 \\ &= n(k-1) - 1(k-1) \\ &= (n-1)(k-1)\end{aligned}$$

Table ANOVA. Tabel anova untuk RAKL ditunjukkan pada Tabel Ω

8. **Keputusan statistik :** Ketika H_0 benar, maka

$$\frac{MSTr}{MSE}$$

Adalah distribusi F dengan derajat kebebasan $((k-1), (n-1)(k-1))$.

	SS	df	MS	$V.R$
$Perlakuan$	$SSTr$	$(k-1)$	$MSTr = SSTr/(k-1)$	$MSTr/MSE$
$Kelompok$	$SSBI$	$(n-1)$	$MSBI = SSBI/(n-1)$	
$Galat$	SSE	$(n-1)\ (k-1)$	$MSE = SSE/(n-1)(k-1)$	
$Total$	SST	$(kn-1)$		

Tabel Ω

9. **Kesimpulan :** Jika menolak H_0, dapat disimpulkan H_A benar. Jika H_0 gagal ditolak maka H_0 benar.

10. **P-value.**

Contoh 8.2. (Example 8.3.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Seorang ahli terapi fisik ingin membandingkan tiga metode untuk mengajar pasiennya menggunakan metode tertentu dengan perangkat prostetik. Ia merasa bahwa kecepatan belajar akan berbeda untuk pasien yang berbeda usia dan ingin merancang eksperimen dimana pengaruh usia dapat diambil akun. Dapatkah disimpulkan bahwa dugaan dari seorang ahli terapi benar?

Penyelesaian :

- **Hipotesis :**

$$H_0 : T_j = 0$$

$$H_1 : T_j \neq 0$$

- **Uji Statistik :**

$$VR = \frac{MST_r}{MSE}$$

- **Distribusi Uji Statistik :** Ketika H_0 benar dan asumsi terpenuhi VR mengikuti distribusi F dengan 2 dan 8 derajat kebebasan dan $\alpha = 0.05$ maka

$$SST = (7-10.07)^2 + (8-10.07)^2 + \cdots + (14-10.07)^2 = 46.9335$$

$$SSB_i = 3[(8.67-10.07)^2 + (9.00-10.07)^2 + \cdots + (12.33-10.07)^2] = 24.855$$

$$SSI_r = 5[(9-10.07)^2 + (9.6-10.07)^2 + (11.6-10.07)^2] = 18.5335$$

$$SSE = 46.9335 - 24.855 - 18.5335 = 3545db(blok) = 5-1 = 4$$

Dimana

$db(total) = 3(5) - 1 = 14$
$db(perawatan) = 3 - 1 = 2$
$db(sisa) = (5 - 1)(3 - 1) = 8$

- **Keputusan statistik :** Karena rasio varian yang dihitung 30.91 lebih besar dari 4.46 maka H_0 ditolak.
- **Kesimpulan :** Dapat disimpulkan bahwa tidak semua efek pengobatan sama dengan nol, tidak semua cara pengobatan adalah sama
- ***p*-value :** $p < 0.005$.

2.4. Rancangan Pengukuran Berulang (*The Repeated Measures Design*)

Definisi 2.2. Rancangan pengukuran berulang adalah rancangan yang pengukurannya sama variabel dibuat pada setiap subjek pada dua / lebih kesempatan yang berbeda. Rancangan pengukuran berulang dimana satu faktor tambahan di masukkan ke dalam eksperimen ini disebut desain pengukuran berulang faktor tunggal.

Model

$$X_{ij} \ : \ \mu + \beta_i + \tau_j + \varepsilon_{ij}$$

$$X_{ijk} \ : \ \mu + \beta_{ij} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + Z_{ijk}$$

Dengan

$i = 1, 2, \dots, a$
$j = 1, 2, \dots, b$
$k = 1, 2, \dots, n$

Contoh 2.3. (Example 8.4.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Licciardone, dkk meneliti subjek dengan nyeri punggung bawah kroniis dan tidak spesifik. Didalam penelitian ini 18 subjek menyelesaikan kuisioner survei penelitian fisik berfungsi pada awal dan setelah 1.386 bulan. Nilai yang lebih tinggi menunjukkan fungsi fisik yang lebih baik. Tujuannya adalah untuk menentukan apakah subjek akan melaporkan peningkatan dari perbaikan minimal. Apakah ada perbedaan rata-rata nilai survei diantara empat waktu?

Penyelesaian :

- **Asumsi :** Bahwa asumsi untuk efek terapi rancangan ukuran berulang faktor tunggal terpenuhi

- **Hipotesis :**

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_1 : ada\ salah\ satu\ tidak\ sama$$

- **Derajat bebas :**

$$4 - 1 = 3$$

$$71 - 3 - 17 = 51$$

Dengan $\alpha = 0.05$ nilai kritis $F = 2.80$ maka H_0 ditolak

- $\boldsymbol{VR}$ **:** $VR = 5.50 > 2.80$

- **Kesimpulan** : Ada perbedaan dalam keempat populasi tersebut karena 5.50 lebih besar dari 4.98 nilai F untuk $\alpha = 0.05$ dan $df = 40$ maka nilai p-value < 0.005.

Contoh 2.4. (Example 8.4.2, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Pusat medis memeriksa 25 subjek dengan kanker leher dan diukur sebagai salah satu variable hasil skor kondisi mulut pasien dibagi secara acak menjadi dua kelompok perlakuan. Ini adalah pengobatan placebo dan kelompok jus lidah buaya. Kesehatan kanker diukur pada awal dan akhir pengobatan untuk mengetahui apakah ada perubahan kondisi keseharan mulut selama percobaan dan untuk melihat apakah ada perbedaan antara dua kondisi perawatan ?

Penyelesaian :

- **Asumsi :** Asumsikan unurk dua faktor pengukuran berulang percobaan terpenuhi

- **Hipotesis :**

a) $H_0 : \alpha_i = 0$
$H_1 : \alpha_i \neq 0$

b) $H_0 : \beta_j = 0$
$H_1 : \beta_j \neq 0$

c) $H_0 : (\alpha\beta)_{ij} = 0$
$H_1 : (\alpha\beta)_{ij} \neq 0$

- **Uji Statistik :**

  df waktu: $4 - 1 = 3$

  df interaksi: $(4 - 1)\ (2 - 1) = 3$

  df penyebut: $(4 - 1)\ (25 - 3) = 69$

  nilai kritis $F\ =\ 2.74$

  untuk faktor antara subjek $(2 - 1)\ \ = 1$

  derajat pembilang: $25 - 2 = 23$

  nilai kritis F menjadi 4.28 dengan $\alpha\ = 0.05$ maka H_0 **ditolak**

- **Kesimpulan :** Dapat disimpulkan bahwa tidak ada perbedaan statistik antara perawatan tetapi subjek memang mengalami perubahan dalam kondisi waktu terlepas dari perawatan yang mereka terima.

2.5. Percobaan Faktorial

Percobaan faktorial adalah percobaan dimana 2 / lebih faktor yang diselidiki secara bersamaan.

2-Faktor RAL (*Completely Randomized Design*). Berikut analisis percobaan faktorial 2-faktor RAL :

- **Data** : Hasil dari rancangan acak lengkap 2-faktor,

$$T_{ij} = \sum_{k=1}^{n} X_{ijk} \quad dan\ \bar{X}_{ij} = \frac{T_{ij}}{n}$$

- **Asumsi** : Asumsikan modal tetap dan 2 faktor sepenuhnya rancangan acak lengkap.

  Model :

$$X_{ijk} =\ \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

  Dengan

  $i = 1, 2, \dots, a$
  $j = 1, 2, \dots, b$
  $k = 1, 2, \dots, n$

Dimana

X_{ijk} : Pengamatan tipikal
m : Konstanta
α_i : Efek karena faktor A
β_j : Efek karena faktor B
$(\alpha\beta)_{ij}$: Efek karena interaksi faktor A dan B
ε_{ijk} : Galat percobaan S

Asumsi pada Model.

a. Pengamatan di setiap ab merupakan sampel bebas acak ukuran n diambil dari populasi yang ditentukan oleh kombinasi tingkat tertentu dari kedua faktor tersebut.
b. Setiap populasi ab berdistribusi normal.
c. Semua populasi memiliki variansi yang sama.

- **Hipotesis**

a. H_0 : $\alpha_i = 0$ $\quad i = 1, 2, \dots, a$

H_A : $\exists\alpha_i \neq 0$

b. H_0 : $\beta_j = 0$ $\quad j = 1, 2, \dots, b$

H_A : $\exists\beta_j \neq 0$

c. H_0 : $(\alpha\beta)_{ij} = 0$ $\quad i = 1, 2, \dots, a$

H_A : $\exists(\alpha\beta)_{ij} \neq 0$ $\quad j = 1, 2, \dots, b$

- **Uji Statistik :** Uji Statistik untuk setiap hipotesis adalah $V.R$
- **Distribusi pada uji statistik :** Ketika H_0 benar dan asumsi terpenuhi, masing-masing uji statistik menggunakan distribusi F.
- **Kriteria Keputusan :** H_0 ditolak jika nilai uji statistik $\geq$ nilai $F\ tabel$.
- **Hitung uji statistik :**

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(X_{ijk} - \bar{X}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(X_{ij} - \bar{X}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(X_{ij} - \bar{X}\right)^2$$

Atau

$$SST = SSTr + SSE$$

Jumlah kuadrat Perlakuan,

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(X_{ij} - \bar{X}\ldots\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{X}_{i..} - \bar{X}\ldots)^2 +$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\bar{X}_{.j.} - \bar{X}\ldots\right)^2 +$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}\right)^2$$

Tabel ANOVA

	SS	df	MS	$V.R$
A	SSA	$(a-1)$	$MSA = SSA/(a-1)$	MSA/MSE
B	SSB	$(b-1)$	$MSB = SSB/(b-1)$	MSB/MSE
AB	$SSAB$	$(a-1)(b-1)$	$MSAB = SSAB/(a-1)(b-1)$	MSA/MSE
$Perlakuan$	$SSTr$	$(ab-1)$		
$Galat$	SSE	$ab(n-1)$	$MSE = SSE/ab(n-1)$	
$Total$	SST		$(abn-1)$	

Tabel Φ

- **Keputusan Statistik :** Jika setiap hipotesis benar, dapat ditunjukkan rasio variansi pada tabel Φ
- **Kesimpulan** : Jika H_0 ditolak maka H_1 diterima. Jika H_0 gagal ditolak maka dapat disimpulkan H_0 bisa benar.
- $\boldsymbol{p}$**-value.**

2.6. Ringkasan :

NAMA	FORMULA
Model One-way ANOVA	$X_{ij} = \mu + \tau_j + \varepsilon_{ij}$
JKT (*Total Sum of Square*)	$SST = \sum_{j=1}^{k}\sum_{i=1}^{nj}\left(X_{ij} - \bar{X}\right)^2$
JKP (*Within-group sum of square*)	$SSW = \sum_{j=1}^{k}\sum_{i=1}^{nj}\left(X_{ij} - \bar{X}_j\right)^2$
JKG (*Among-group sum of square*)	$SSA = \sum_{j=1}^{k} n_j \left(\bar{X}_j - \bar{X}\right)^2$
KTP (*Within-group variance*)	$MSW = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)^2}{\sum_{j=1}^{k}\left(n_j - 1\right)}$
KTG (*Among-group variance I*), ukuran sampel sama	$\sigma^2 = n\,\sigma_{\bar{X}}{}^2$
KTG (*Among-group variance II*), ukuran sampel sama	$MSA = \frac{n\sum_{i=1}^{k}\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2}{k-1}$
KTG (*Among-group variance III*), ukuran sampel sama	$MSA = \frac{\sum_{i=1}^{k} n_j\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2}{k-1}$
Model Two-way ANOVA	$X_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$
JKT (*Total Sum of Square*)	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_{..}\right)^2$
JKK (*Sum of square block*)	$SSBI = \sum_{j=1}^{k}\sum_{i=1}^{n}(\bar{X}_{i.} - \bar{X}_{..})^2$
JKP (*Sum of square treatments*)	$SSTr = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2$
JKG (*Among-group sum of square*)	$SSE = SST - SSBI - SSTr$

BAB III

REGRESI LINEAR SEDERHANA DAN KORELASI

3.1. Pendahuluan

Regresi : Analisis regresi digunakan untuk menduga dan mengestimasi nilai-nilai pada data. Biasanya terdiri dari variabel Y (*dependent*) dan variabel X (*independent*).

Korelasi : Analisis korelasi mengenai hubungan antar variabel.

3.2. Model Regresi

Asumsi yang mendasari regresi linear sederhana dalam model regresi linear sederhana 2 variabel biasanya berlabel X dan Y. Huruf X digunakan untuk menunjukkan variabel yang disebut variabel bebas karena sering kali variabel itu dikendalikan oleh sang penyidik yaitu dipilih oleh penyidik dan sesuai dengan setiap nilai X yang telah dipilih sebelumnya, satu atau lebih dari variabel lain yang diperoleh dari variabel Y. Huruf Y disebut variabel terikat.

Asumsi-asumsi :

1. Nilai-nilai variabel bebas X dikatakan tetap yang berarti nilai X dipilih sebelumnya oleh penyidik sehingga dalam koleksi data mereka tidak diizinkan untuk memvariasikan dari nilai-nilai yang dipilih sebelumnya.
2. Variabel X diukur tanpa kesalahan karena tidak ada prosedur pengukuran yang sempurna yang berarti bahwa besarnya pengukuran kesalahan dalam X dapat diabaikan.
3. Untuk setiap nilai X terdapat subpopulasi nilai Y. Agar prosedur estimasi dan uji hipotesis bisa disahkan, subpopulasi harus normal.
4. Variansi pada subpopulasi Y adalah sama dinotasikan dengan σ^2.
5. Sarana dari subpopulasi terletak pada garis lurus yang sama dikenal dengan nama asumsi linear,

   $$\mu_{y|x} = \beta_0 + \beta_1 x$$

   Dimana

   $\mu_{y|x}$: Rata-rata pada subpopulasi nilai Y untuk partikular nilai X

   β_0 dan β_1 : Populasi koefisien regresi

Model Regresi Linear Sederhana :

$$y = \beta_0 + \beta_1 x + \epsilon$$

Dengan

$$\epsilon = y - (\beta_0 + \beta_1 x)$$

3.3. Persamaan Regresi Sederhana

Langkah-langkah dalam analisis regresi sederhana sebagai berikut :

1. Menentukan apakah asumsi yang mendasari hubungannya linear atau tidak dalam data yang tersedia.
2. Mendapatkan persamaan untuk baris yang paling cocok dengan data sampel.
3. Mengevaluasi persamaan untuk mendapatkan beberapa gagasan mengenai tingkat kekuatan hubungan dan kegunaan persamaan untuk dugaan dan estimasi.
4. Jika data tampaknya menyesuaikan dengan model linear, gunakan persamaan yang diperoleh dari data sampel untuk menduga dan mengestimasi.

Model :

$$y = a + bx$$

LATIHAN : Persamaan Regresi Sederhana

Latihan 3.1. (Exercise 9.3.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Plot masing-masing persamaan regresi berikut dan nyatakan apakah X dan Y terkait secara langsung atau terbalik.

a) $\hat{y} = -3 + 2x$

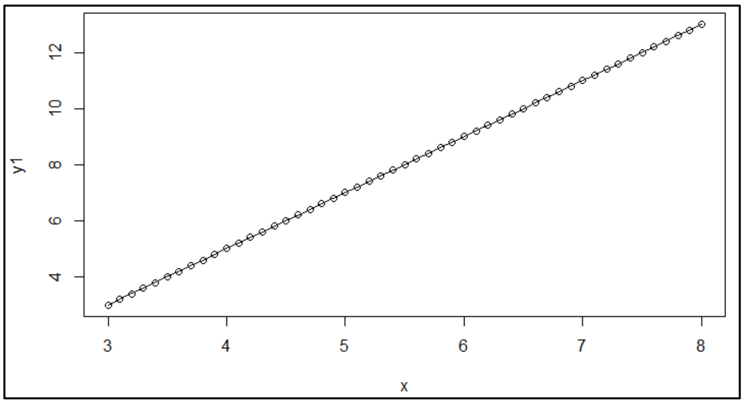
b) $\hat{y} = 3 + 0.5x$

c) $\hat{y} = 10 - 0.75x$

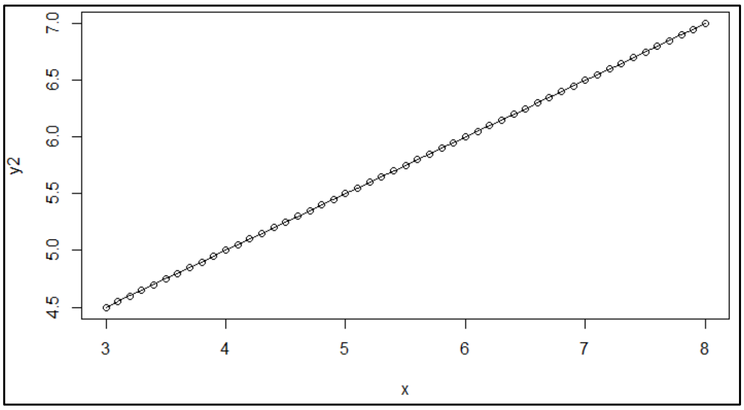
Penyelesaian :

Dengan menggunakan R diperoleh,

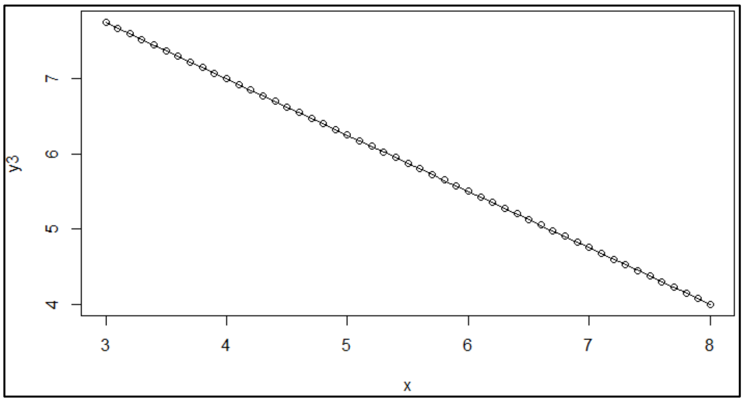
a) Terikat secara langsung semakin besar x semakin besar y



b) Terikat secara langsung semakin besar x semakin besar y



c) Tidak terikat secara langsung semakin besar x semakin kecil y



```
Syntax :
>x <- seq(from=3,to=8,by=0.1)
>y1 <- -3 + 2*x
>y2 <- 3 + 0.5*x
>y3 <- 10 - 0.75*x
>scatter.smooth(x,y1)
>scatter.smooth(x,y2)
>scatter.smooth(x,y3)
```

Latihan 3.2 (Exercise 9.3.2, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Skor berikut mewakili penilaian perawat (X) dan penilaian dokter (Y) dari kondisi 10 pasien pada saat masuk ke pusat trauma.

$$X: 18 \quad 13 \quad 18 \quad 15 \quad 10 \quad 12 \quad 8 \quad 4 \quad 7 \quad 3$$

$$Y: 23 \quad 20 \quad 18 \quad 16 \quad 14 \quad 11 \quad 10 \quad 7 \quad 6 \quad 4$$

a) Buat diagram scatter untuk data diatas

b) Plot persamaan regresi berikut pada diagram scatter dan tunjukkan mana yang menurut Anda paling sesuai dengan data. Jelaskan alasan pilihanmu.

   a. $\hat{y} = 8 + 0.5x$

   b. $\hat{y} = -10 + 2x$

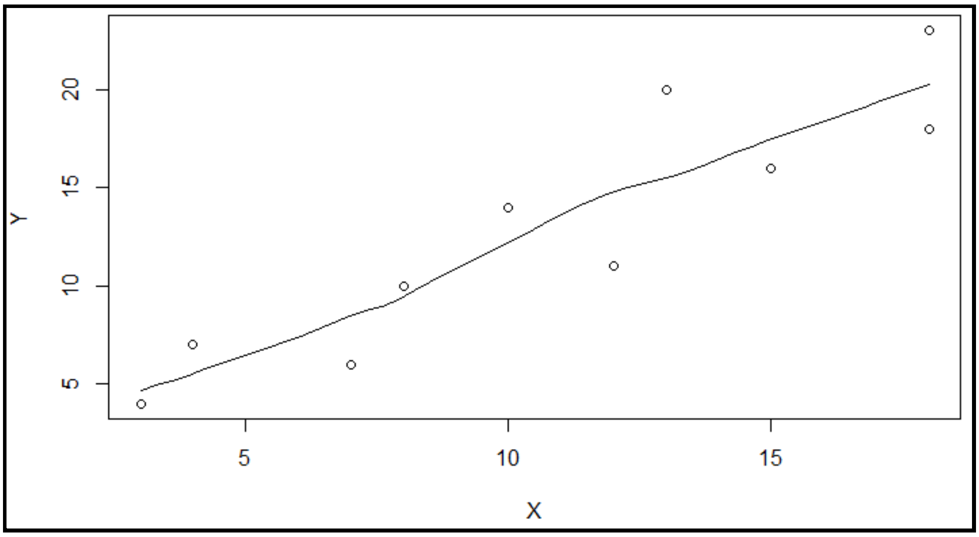
   c. $\hat{y} = 1 + 1x$

Untuk setiap latihan berikut a) gambarkan diagram scatter dan b) dapatkan persamaan regresi dan plot pada diagram scatter?

Penyelesaian :

a) Syntax :

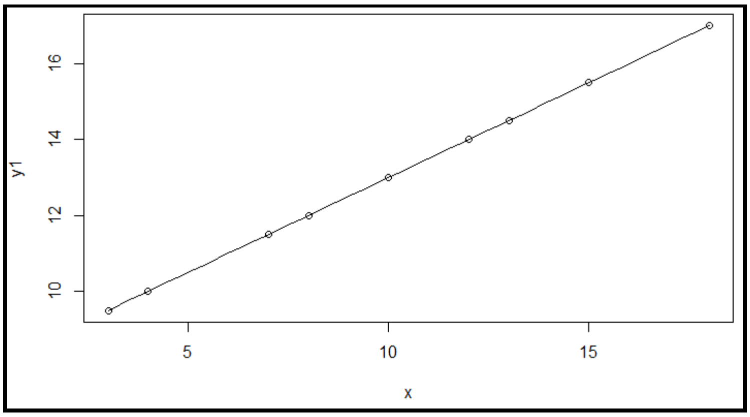
```
>latihan9.3.2<-read.table("E:/latihan932.csv",header=TRUE,sep="",na.strings="NA",
dec=".", strip.white=TRUE)
>View(latihan9.3.2)
>scatter.smooth(latihan9.3.2)
```

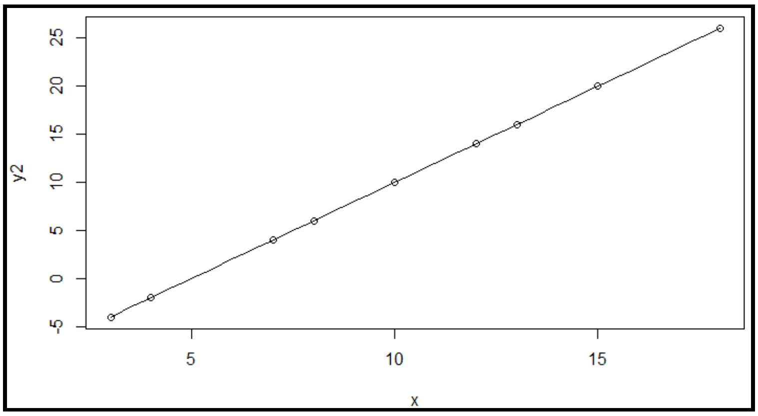


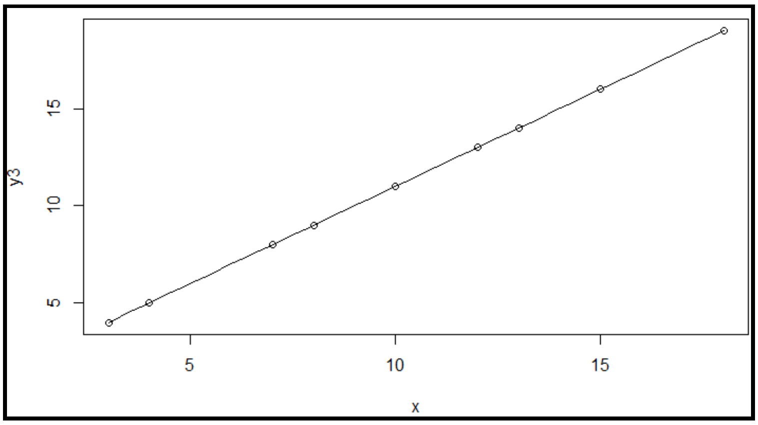
b) Persamaan regresi dan diagram berturut-turut :

```
call:
lm(formula=Y~X, data=latihan9.3.2)
coefficients:
(Intercept)          X
1.211             1.082
```

```
Syntax :
>RegModel<-lm(Y~X, data=latihan9.3.2)
>RegModel
```







Jadi diperoleh persamaan regresi $\hat{y} = 1.21 + 1.08x$. Pilihan nomor 3) adalah yang paling dekat dengan nilai hasil model regresi.

Latihan 3.3. (Exercise 9.3.3, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Methadone sering diresepkan dalam pengobatan kecanduan opioid dan nyeri kronis. Krantz dkk. mempelajari relasi antara dosis methadone dan interval QT (QTc) yang dikoreksi untuk 17 subjek yang membentuk torsade de pointes (gangguan irama jantung ventrikuler obat-obatan). QTc dihitung dari elektrokardiogram dan diukur dalam $mm/detik$. Nilai QTc yang lebih tinggi menunjukkan risiko kematian kardiovaskular yang lebih tinggi. Pertanyaan yang menarik adalah seberapa baik seseorang dapat memprediksi dan memperkirakan nilai QTc dari pengetahuan dosis methadone. Pertanyaan ini dapat dijawab dengan cara analisis regression. Karena QTc adalah variabel tentang prediksi dan estimasi, maka itu adalah variabel dependen. Dosis variable methadone, pengetahuan yang akan digunakan untuk membuat prediksi dan estimasi, adalah variable independen.

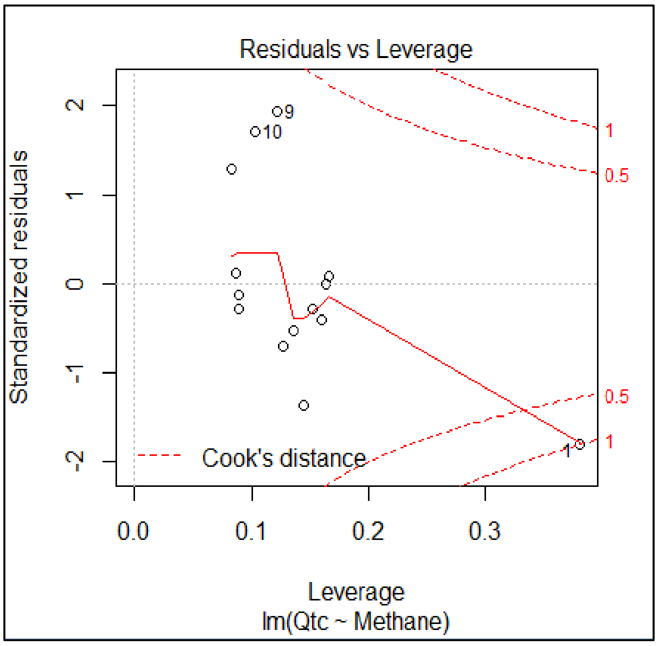
Dosis Methadone $(mg/hari)$	QTc $(mm/detik)$	Dosis Methadone $(mg/hari)$	QTc $(mm/detik)$
1000	600	650	785
550	625	600	765
97	560	660	611
90	585	270	600
85	590	680	625
126	500	540	650
300	700	600	635
110	570	330	522
65	540		

Penyelesaian :

Dengan menggunakan software R, diperoleh

```
Call:
lm(formula = Qtc ~ Methane, data = latihan933)

Coefficients:
(Intercept)      Methane
   574.1837       0.1263
```


```
Call:
lm(formula = Qtc ~ Methane, data = latihan933)

Coefficients:
(Intercept)      Methane
   574.1837       0.1263
```




```
Syntax :
> latihan933 <- read.table("E:/latihan9.3.3.csv", header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)
>RegModel1 <- lm(Qtc~Methane, data = latihan933)
>RegModel1
>plot(RegModel1)
```

Diperoleh model regresi untuk memprediksi QTc (Y) sebagai berikut : $\hat{y} = 574.18 + 0.13x$. Nilai R square : 0.1636 artinya data Dosis Methadon yang diberikan hanya mampu memprediksi secara akurat nilai Qtc sebesar 16.36 %.

3.4. Evaluasi Persamaan Regresi

Ketika H₀ : $\boldsymbol{\beta_1} = 0$ gagal ditolak : Asumsikan hubungan populasi antara X dan Y linear maka bisa dikatakan lebih baik dari model tak linear.

Ketika H_0 : $\boldsymbol{\beta_1} = 0$ ditolak : Asumsikan tidak komitmen dengan kesalahan tipe 1, menolak H_0 pada $\beta_1 = 0$ bisa ketika mengalami kondisi sebagai berikut :

1) Hubungan kedua populasi linear
2) Data baik dengan model linear.

Koefisien determinasi : Ukuran yang objektif sering disebut Koefisien Determinasi.

3.5. Penggunaan Persamaan Regresi

Dugaan Y untuk X, ketika $\sigma^2_{y|x}$ diketahui dengan $100(1-\alpha)$ persen, estimasi interval untuk Y adalah

$$\hat{y} \pm t_{\left(1-\frac{\alpha}{2}\right)} s_{y|x} \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{\sum (x_i - \bar{x})^2}}$$

Dimana x_p adalah nilai partikular pada x dengan derajat kebebasannya $n-2$.

Estimasi Rata-rata pada Y untuk X : Interval Kepercayaan untuk $\sigma^2_{y|x}$ diketahui dengan $100(1-\alpha)$ persen maka

$$\hat{y} \pm t_{\left(1-\frac{\alpha}{2}\right)} s_{y|x} \sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{\sum (x_i - \bar{x})^2}}$$

3.6. Model Korelasi

Ketika Y dan X merupakan variabel acak maka bisa disebut model korelasi.

Distribusi Normal Bivariat : Dibawah model korelasi, X dan Y diasumsikan distribusi gabungan. Jika distribusi gabungannya adalah distribusi normal, maka dinamakan Distribusi normal bivariat.

Asumsi Korelasi :

a. Untuk setiap nilai X berdistribusi normal adalah subpopulasi nilai Y
b. Untuk setiap nilai Y berdistribusi normal adalah subpopulasi nilai X
c. Distribusi gabungan X dan Y adalah Distribusi Normal atau sering disebut dengan distribusi normal bivariat.

3.7. Koefisien Korelasi

Pada bagian 3.6 membahas mengenai distribusi normal bivariat dengan 5 parameter, σ_x, σ_y, μ_x, μ_y, dan ρ. Dari ke-4 yang pertama merupakan standar deviasi dan rata-rata pada distribusi tunggal. Sedangkan ρ untuk Populasi koefisien korelasi.

BAB IV

REGRESI LINEAR BERGANDA DAN KORELASI

4.1. Pendahuluan

Sama halnya seperti bab sebelumnya, regresi linier berganda hanya berbeda pada variabel independen, yakni lebih dari satu variabel.

4.2. Model Regresi Linear Berganda

Dalam model regresi berganda asumsikan bahwa terdapat hubungan linear antara beberapa variabel Y (*dependent*) dan k variabel bebas, X_1; X_2; $\ldots$; X_k (*independents*). Variabel bebas sering disebut sebagai variabel penjelas, karena penggunaannya dalam menjelaskan variabel Y juga disebut variabel prediktor, karena penggunaannya dalam memprediksi Y.

Asumsi : Asumsi yang mendasari analisis regresi berganda adalah sebagai berikut :

1. X_i adalah variabel tidak acak (tetap).
2. Untuk setiap kumpulan nilai X_i ada subpopulasi nilai Y. Untuk membangun interval kepercayaan tertentu dan menguji hipotesis, harus diketahui, atau harus berasumsi, bahwa subpopulasi Y terdistribusi normal. Karena kami ingin mendemonstrasikan prosedur inferensial ini
3. Varians dari subpopulasi Y semuanya sama.
4. Nilai Y bebas (*independent*) . Artinya, nilai Y yang dipilih untuk satu kumpulan nilai X tidak bergantung pada nilai Y yang dipilih pada kumpulan nilai X lainnya.

Model Persamaan : Asumsi untuk analisis regresi berganda dapat dinyatakan dengan cara yang lebih ringkas sebagai

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \epsilon_j$$

Dimana

y_j	: Nilai tipikal dari salah satu subpopulasi nilai Y
β_i	: Koefisien regresi
x_{1j} , x_{2j} , $\ldots$, x_{kj}	: Nilai tertentu dari variabel bebas X_1 ; X_2 ; $\ldots$; X_k
ϵ_j	: Variabel acak mean 0 dan varians σ^2; varians umum dari subpopulasi nilai Y.

Untuk membangun interval kepercayaan dan menguji hipotesis tentang koefisien regresi, asumsikan bahwa ϵ_j terdistribusi secara normal dan bebas. Ketika Persamaan diatas terdiri dari satu variabel terikat dan dua variabel bebas, yaitu saat model ditulis

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j$$

Jika model terdiri lebih dari dua variabel bebas, model ini digambarkan secara geometris sebagai bidang-hiper. Penyimpangan suatu titik dari bidang diwakili oleh

$$\epsilon_j = y_j - \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j}$$

4.3. Mendapatkan Persamaan Regresi Berganda

Estimasi sampel pada $\beta_0, \beta_1, \dots, \beta_k$,

$$\sum \in_j^2 = \sum \left(y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j} - \cdots - \beta_k x_{kj}\right)^2$$

diminimalkan. Kuantitas ini disebut sebagai jumlah kuadrat dari residu, juga dapat ditulis sebagai

$$\sum \in_j^2 = \sum \left(y_j - \hat{y}_j\right)^2$$

4.4. Evaluasi Persamaan Regresi Berganda

Koefisien pada Determinasi Berganda : Kuantitas ini dirujuk ke jumlah kuadrat tentang regresi atau Jumlah Kuadrat Galat (SSE). Hubungan antara tiga jumlah kuadrat dengan persamaan berikut :

$$\underbrace{\sum \left(y_j - \bar{y}\right)^2}_{SST} = \underbrace{\sum \left(\hat{y}_j - \bar{y}\right)^2}_{SSR} + \underbrace{\sum \left(y_j - \hat{y}_j\right)^2}_{SSE}$$

Dengan

SST : Jumlah Kuadrat Total

SSR : Jumlah Kuadrat Regresi

SSE : Jumlah Kuadrat Galat

Koefisien determinasi berganda diperoleh dengan membagi jumlah kuadrat yang dijelaskan dengan jumlah total kuadrat,

$$R^2_{y=1,2,\dots,k} = \frac{\sum\left(\hat{y}_j - \bar{y}\right)^2}{\sum\left(y_j - \bar{y}\right)^2} = \frac{SSR}{SST}$$

Uji Hipotesis Regresi Untuk menentukan apakah regresi secara keseluruhan signifikan (yaitu untuk menentukan apakah $R^2_{y=1,2,\dots,k}$ signifikan), Lakukan sebagai berikut :

- **Data :** Situasi penelitian dan yang dihasilkan oleh penelitian lain diperiksa untuk menentukan apakah regresi berganda merupakan teknik yang tepat untuk analisis.

- **Asumsi :** Kami berasumsi bahwa model regresi berganda dan asumsi yang mendasari seperti yang disajikan dalam Bagian 4.2 dapat diterapkan.

- **Hipotesis :** Secara umum hipotesis nol adalah $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ dan alternatifnya adalah $H_1 : tidak\ semua\ \beta_i = 0$. Dengan kata lain, hipotesis nol menyatakan bahwa semua variabel independen tidak ada nilainya di menjelaskan variasi nilai Y.

- **Uji statistic :** Statistik uji yang sesuai adalah $V.R.$, yang dihitung sebagai bagian dari analisis variansi. MSR adalah singkatan dari $mean\ square$ regresi dan MSE adalah singkatan dari $mean\ square$ tentang regresi atau sering disebut, *Mean Square Error*.

- **Distribusi statistik uji :** Ketika H_0 benar dan asumsi terpenuhi, $V.R.$ didistribusikan sebagai F dengan k dan derajat kebebasannya $n - k - 1$.

- **Kriteria Keputusan :** Tolak H_0 jika nilai yang dihitung dari $V.R.$ sama dengan lebih besar dari nilai kritis F.

- **Hitung statistik uji :** Lihat Tabel Ψ.

- **Keputusan statistic :** Tolak atau gagal untuk menolak H_0 sesuai dengan aturan keputusan.

- **Kesimpulan :** Jika H_0 ditolak, dapat disimpulkan bahwa dalam populasi dari mana sampel diambil, variabel terikat secara linier berhubungan dengan variabel bebas sebagai grup. Jika kita gagal menolak H_0, dapat disimpulkan bahwa dalam populasi dari mana sampel kami diambil, mungkin tidak ada hubungan linier antara variabel dependen dan variabel independen sebagai satu kelompok.

- **P-value :** Kami mendapatkan nilai p dari tabel distribusi F.

	SS	df	MS	$V.R$
$Regresi$	SSR	k	$MSR = SSR/k$	MSR/MSE
$Galat$	SSE	$n - k - 1$	$MSE = \frac{SSE}{n-k-1}$	
$Total$	SST	$n - 1$		

Tabel Ψ

Contoh 4.1. (Example 10.4.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Akan menguji hipotesis nol mengenai tidak adanya hubungan linear di antara tiga variabel yang dibahas dalam Contoh 10.3.1 **("BIOSTATISTICS Wayne W. Daniel Edisi 10")** : skor CDA, usia, dan tingkat pendidikan.

Penyelesaian :

- **Data :**

$$SST = 1061.36$$

$$SSR = 393.39$$

$$SSE = 667.97$$

$$R^2_{y.12} = \frac{393.39}{1061.36}$$

$$= 0.3706 \approx 0.371$$

- **Asumsi :** Asumsikan model regresi berganda terpenuhi.
- **Hipotesis :**

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : tidak\ semua\ \beta_1 = 0$$

- **Uji statistic :** Statistik uji adalah $V.R$.
- **Distribusi uji statistik :** Jika H_0 benar dan asumsi terpenuhi, statistik uji didistribusikan sebagai F dengan 2 pembilang dan 68 derajat kebebasan penyebut.
- **Kriteria keputusan :** Gunakan tingkat signifikansi $\alpha = 0.01$. H_0ditolak jika nilai yang dihitung dari $V.R. \leq 4{,}95$ (diperoleh dengan interpolasi).

- **Hitung uji statistik :** ANOVA untuk contoh ditunjukkan pada Gambar 4.3.1, di mana kita melihat bahwa nilai $V.R.$ adalah 20.02.

- **Keputusan statistik :** Karena $20.02 > 4.95$ maka H_0 ditolak.

- **Kesimpulan :** Dapat disimpulkan bahwa dalam populasi asal sampel, terdapat hubungan linear antara ketiga variabel.

- ***p*-value :** Karena $20.02 > 0.76$, nilai $p < 0.005$.

- **Kesimpulan :** Mengenai Individual $\beta_s^{'}$. Seringkali, kami ingin mengevaluasi kekuatan hubungan linear antara Y dan variabel bebas secara individual. Artinya, kita mungkin ingin menguji H_0 bahwa $\beta_i = 0$ terhadap H_A dengan $\beta_i \neq 0$ $(i = 1, 2, \dots, k)$. Validitas prosedur ini bertumpu pada asumsi yang telah dikemukakan sebelumnya : bahwa untuk setiap kombinasi nilai X_i terdapat subpopulasi yang terdistribusi secara normal dari nilai Y dengan varian σ^2.

Uji Hipotesis untuk β_i. Untuk menguji hipotesis nol bahwa β_i sama dengan beberapa nilai tertentu, katakanlah, $\beta_i = 0$, statistik t berikut dapat dihitung:

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{S_{\hat{\beta}_i}}$$

dimana derajat kebebasannya $n - k - 1$, dan $S_{\overline{\beta}_i}$ adalah simpangan $\hat{\beta}_i$. Standar deviasi dari $\bar{\beta}_i$ diberikan sebagai bagian dari keluaran dari kebanyakan paket perangkat lunak komputer yang melakukan analisis regresi.

4.5. Penggunaan Persamaan Regresi Ganda

Interval Keyakinan untuk rata-rata subpopulasi dari nilai Y mengingat nilai tertentu dari X_i Kita telah melihat bahwa interval kepercayaan $100(1 - \alpha)\%$ untuk suatu parameter dapat dibangun dengan prosedur umum penambahan dan pengurangan dari penduga kuantitas yang sama dengan faktor reliabilitas sesuai dengan $1 - \alpha$ dikalikan dengan kesalahan standar penduga. Kita juga telah melihat bahwa dalam regresi berganda penaksirnya adalah

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} + \cdots + \hat{\beta}_k x_{kj}$$

Contoh 4.2. (Example 10.5.1, "BIOSTATISTICS Wayne W. Daniel Edisi 10") Kita mengacu pada Contoh 10.3.1. **"BIOSTATISTICS Wayne W. Daniel Edisi 10".** Pertama, kami ingin membangun interval kepercayaan 95 persen untuk skor CDA rata-rata (Y) pada populasi subjek berusia 68 tahun ($X1$) yang menyelesaikan pendidikan 12 tahun ($X2$). Kedua, misalkan kita memiliki subjek yang berusia 68 tahun dan memiliki tingkat pendidikan 12 tahun. Apa yang kami prediksi untuk skor CDA subjek ini?

Penyelesaian :

Perkiraan poin dari skor CDA rata-rata adalah

$$\hat{y} = 5.494 - 0.18412(68) + 0.6108(12) = 0.3.34$$

Prediksi titik, yang sama dengan perkiraan titik yang diperoleh sebelumnya, juga

$$\hat{y} = 5.494 - 0.18412(68) + 0.6108(12) = 0.3.34$$

Selain analisis regresi, diperoleh keluaran sebagai berikut :

New Obs	Fit SE Fit	95.0% CI	95.0% PI
1	0.303 0.672	(1.038, 1.644)	(6.093, 6.699)

ditafsirkan interval ini dengan cara biasa. Pertama-tama kita melihat interval kepercayaan. Kami yakin 95 persen bahwa interval dari 1.038 hingga 1.644 mencakup rata-rata subpopulasi nilai Y untuk kombinasi nilai X_i yang ditentukan, karena parameter ini akan dimasukkan dalam sekitar 95 persen dari interval yang dapat dibangun di cara yang ditunjukkan. Sekarang pertimbangkan subjek yang berusia 68 tahun dan memiliki pendidikan 12 tahun. Kami 95 persen yakin bahwa subjek ini akan memiliki skor CDA antara 6.093 dan 6.699. Fakta bahwa P.I. lebih lebar dari C.I. seharusnya tidak mengherankan. Bagaimanapun, lebih mudah memperkirakan respons rata-rata daripada memperkirakan pengamatan individu.

4.6. Korelasi Model Ganda

Model persamaan :

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \epsilon_j$$

Dimana

y_j : Nilai tipikal dari salah satu subpopulasi nilai Y

β_i : Koefisien regresi

x_{ij} : Nilai khusus (diketahui) dari variabel acak X_i

X_i : Bukan variabel acak, tetapi dalam model korelasi ganda X_i adalah variabel acak

ϵ_j : Variabel acak mean 0 dan varians σ^2; varians umum dari subpopulasi nilai Y.

Dengan kata lain, dalam model korelasi terdapat distribusi gabungan Y dan X_i yang disebut sebagai distribusi berganda. Kita juga dapat menghitung koefisien korelasi parsial yang mengukur intensitas hubungan antara dua variable ketika pengaruh semua variabel lain telah dihilangkan.

Koefisien Korelasi Ganda : Sebagai langkah pertama dalam menganalisis hubungan antar variabel, kita melihat koefisien korelasi berganda. Koefisien korelasi ganda adalah akar kuadrat dari koefisien determinasi berganda dan, akibatnya,

$$R_{y=1,2,\ldots,k} = \sqrt{R^2_{y=1,2,\ldots,k}} \sqrt{\frac{\sum\left(\hat{y}_j - \bar{y}\right)^2}{\sum\left(y_j - \bar{y}\right)^2}} = \sqrt{\frac{SSR}{SST}}$$

Contoh 4.3. (Example 10.6.1. "BIOSTATISTICS Wayne W. Daniel Edisi 10") Wang dkk, menggunakan femur kadaver manusia dari subjek yang berusia 16 hingga 19 tahun, menyelidiki sifat-sifat ketangguhan tulang dan ukuran jaringan kolagen di dalam tulang. Dua variabel yang mengukur jaringan kolagen adalah porositas (P, dinyatakan sebagai persen) dan ukuran kekuatan tarik jaringan kolagen (S). Ukuran ketangguhan (W, Newton), adalah gaya yang dibutuhkan untuk patah tulang. 29 femur kadaver yang digunakan dalam penelitian ini bebas dari patologi terkait tulang. Analisislah sifat dan kekuatan hubungan antara ketiga variabel. Pengukuran tersebut ditunjukkan pada tabel berikut.

W	P	S
193.6	6.24	30.1
137.5	8.03	22.2
145.4	11.62	25.7
117.0	7.68	28.9
105.4	10.72	27.3
99.9	9.28	33.4
74.0	6.23	26.4
74.4	8.67	17.2
112.8	6.91	15.9
125.4	7.51	12.2
126.5	10.01	30.0
115.9	8.70	24.0
98.8	5.87	22.6
94.3	7.96	18.2
99.9	12.27	11.5
83.3	7.33	23.9
72.8	11.17	11.2
83.5	6.03	15.6
59.0	7.90	10.6
87.2	8.27	24.7
84.4	11.05	25.6
78.1	7.61	18.4
51.9	6.21	13.5
57.1	7.24	12.2
54.7	8.11	14.8
78.6	10.05	8.9
53.7	8.79	14.8
96.0	10.40	10.3
89.0	11.72	15.4

Penyelesaian :

Dengan menggunakan software R, Nilai sampel $Y, X_1,$ dan X_2, masing-masing disimpan di kolom 1 hingga 3, diperoleh

OUTPUT

```
Call:
lm(formula = W ~ P + S, data = Data10.6)

Residuals:
    Min      1Q  Median      3Q     Max 
-33.923 -19.609  -0.522  10.145  76.792 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  35.5881    29.1151   1.222  0.23255   
P             1.4518     2.7622   0.526  0.60363   
S             2.3974     0.7296   3.286  0.00291 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.41 on 26 degrees of freedom
Multiple R-squared:  0.2947,  Adjusted R-squared:  0.2405 
F-statistic: 5.433 on 2 and 26 DF,  p-value: 0.01068
```

```
Syntax:
> LinearModel.1 <- lm(W ~ P +S, data=Data10.6)
> summary(LinearModel.1)
```

Maka, persamaan kuadrat kecil adalah

$$\hat{y}_j = 35.61 + 1.451\ x_{1j} + 2.3960\ x_{2j}$$

Persamaan ini dapat digunakan untuk tujuan estimasi dan prediksi dan dapat dievaluasi dengan metode yang dibahas dalam Bagian 4.4. Berikut koefisien determinasi berganda yang diperoleh,

$$R_{y.12}^2 = 0.294$$

$$R_{y.12} = \sqrt{0.294} = 0.542$$

Korespondensi sempurna antara nilai Y yang diamati dan dihitung akan menghasilkan koefisien korelasi 1, sementara ketiadaan hubungan linier antara nilai yang diamati dan dihitung menghasilkan koefisien korelasi 0. Koefisien korelasi berganda selalu diberi tanda positif. Kita dapat menguji hipotesis nol bahwa $r_{y=1,2,\ldots,k} = 0$dengan menghitung

$$F = \frac{0.294}{1-0.294} \cdot \frac{29-2-1}{2} = 5.41$$

Korelasi Parsial Peneliti mungkin ingin mengukur kekuatan hubungan linier antara dua variabel ketika pengaruh variabel yang tersisa telah dihilangkan. Ukuran seperti itu disediakan oleh koefisien korelasi parsial

Menghitung Koefisien Korelasi Parsial Untuk tiga variabel, koefisien korelasi sederhana berikut dapat dihitung :

r_{y1}, korelasi sederhana Y dan X_1
r_{y2}, korelasi sederhana Y dan X_2
r_{y3}, korelasi sederhana Y dan X_3

Tiga variabel kasus koefisien korelasi sederhana :

1. Korelasi parsial antara Y dan X_1 setelah mengontrol pengaruh X_2 :

$$r_{y1.2} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{\left(1-r_{y2}^2\right)\left(1-r_{12}^2\right)}}$$

2. Korelasi parsial antara Y dan X_2 setelah mengontrol pengaruh X_1 :

$$r_{y2.1} = (r_{y2} - r_{y1}r_{12})/\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}$$

3. Korelasi parsial antara X_1 dan X_2 setelah mengendalikan pengaruh Y :

$$r_{12y} = (r_{y12} - r_{y1}r_{y2})/\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}$$

Uji Hipotesis mengenai Koefisien Korelasi Parsial.

Untuk $H_0 : \rho_{y1,2,\dots,k} = 0$ maka hitung,

$$t = r_{y1,2,\dots,k}\sqrt{\frac{n-k-1}{1-R_{y1,2,\dots,k}^2}}$$

Yang mana t berdistribusi student's t dengan derajat bebasnya $n - k - 1$.

Untuk $H_0 : \rho_{y1.2} = 0$ dan $H_A : \rho_{y1.2} \neq 0$ maka hitung,

$$t = 0.102\sqrt{\frac{29-2-1}{1-(0.102)^2}} = 0.523$$

Diperoleh t hitung $0.523 <$ t tabel $= 2.0555$ pada derajat bebas 26. Dengan taraf signifikansinya 0.05 (2 sisi) maka H_0 gagal ditolak.

Kesimpulan : Bisa tidak terdapat korelasi pada persamaan.

BAB V

NONPARAMETRIK DAN DISTRIBUSI BEBAS STATISTIKA

5.1. Pendahuluan

Pada bab sebelumnya membahas uji statistik dengan menggunakan metode parametrik dimana data sampel diasumsikan terdistribusi normal. Hal ini diperlukan agar pengujian menjadi valid. Distribusinya diasumsikan diketahui (seperti terdistribusi normal) dan hanya nilai parameter tertentu seperti *mean* dan standar deviasi, oleh karena itu dikatakan *parametrik.* Jika data tidak sesuai dengan asumsi yang dibuat, metode inferensi statistik non parametrik harus digunakan.Teknik *nonparametrik* membuat lebih sedikit asumsi mengenai sifat distribusi dasar. Sehingga metode ini disebut *metode bebas distribusi* (sebaran).

5.2. The Sign Test (Uji Tanda)

Uji tanda merupakan uji nonparametrik yang digunakan untuk menguji ada tidaknya perbedaan dari dua buah populasi yang saling berpasangan. Serupa dengan metode parametrik menggunakan uji-t berpasangan yang berfokus pada perbedaan nilai untuk setiap pasangan. Namun, uji statistik ini tidak membutuhkan perbedaan populasinya terdistribusi normal. Uji tanda digunakan untuk mengevaluasi menggunakan hipotesis nol bahwa dalam populasi yang mendasari perbedaan di antara pasangan, perbedaan median sama dengan 0.

Contoh 5.1. Misalkan diberikan data mengenai penyakit **fibrosis kistik (** ***Disease Cystic fibrosis*** **) yakni penyakit keturunan yang menyebabkan lendir-lendir di dalam tubuh menjadi kental dan lengket. Akan dibandingkan energi yang dikeluarkan oleh pasien pada saat istirahat dengan orang sehat yang dicocokkan dengan pasien pada karakteristik penting tertentu. D**ari 13 pasien dengan fibrosis kistik dan 13 orang sehat dicocokkan dengan pasien pada usia, jenis kelamin, tinggi, dan berat badan.

Pasangan ke-	REE($kkal/hari$)	
	CF	Kesehatan Individu
1	1153	996
2	1132	1080
3	1165	1182
4	1460	1452

5	1634	1162
6	1493	1619
7	1358	1140
8	1453	1123
9	1185	1113
10	1824	1463
11	1793	1632
12	1930	1614
13	2075	1836

Tabel Δ

Apakah median dari kedua populasi berbeda? (α = 0.05)

Penyelesaian :

- **Data :** Berikut data selisih kedua sampel :

Pasangan ke-	REE($kkal/hari$)		Beda	Tanda
	CF	Kesehatan Individu		
1	1153	996	157	$+$
2	1132	1080	52	$+$
3	1165	1182	-17	$-$
4	1460	1452	8	$+$
5	1634	1162	472	$+$
6	1493	1619	-126	$-$
7	1358	1140	218	$+$
8	1453	1123	330	$+$
9	1185	1113	72	$+$
10	1824	1463	361	$+$
11	1793	1632	161	$+$
12	1930	1614	316	$+$
13	2075	1836	239	$+$

Tabel A

Dengan menggunakan nilai-nilai diatas, hitung selisih setiap pasang. Jika perbedaannya lebih besar dari 0, pasangan diberi nilai tanda plus; jika kurang dari 0, maka diberi tanda minus. Perbedaan tepat 0 diabaikan saja (tidak memiliki pengaruh).

- **Asumsi :** Asumsi pengukuran diambil pada beda nilai pasangan adalah variabel kontinu.
- **Hipotesis :** Hipotesis nol yang diuji pada uji tanda pada dasarnya menyatakan tidak terdapat pengaruh dari kedua populasi. Dengan kata lain, probabilitas untuk memperoleh tanda positif $(+)$ sama dengan probabilitas untuk memperoleh tanda negatif $(-)$, yakni

  H_0 : Median kedua populasi sama $[P(+) = P(-)]$

  H_A : Median kedua populasi berbeda $[P(+) \neq P(-)]$

  Taraf signifikansi $\alpha = 0.05.$
- **Uji Statistik :** Uji Statistik untuk uji tanda menggunakan tanda positif.Jumlah tanda positif dalam sampel dilambangkan dengan *D.*
- **Distribusi pada Uji Statistik :** Secara ekuivalen, probabilitasnya positif adalah $\frac{1}{2}$, dan probabilitasnya negatif juga $\frac{1}{2}$. Jika tanda positif dianggap sebagai "sukses", maka tanda positif dan negatif bisa dianggap sebagai hasil dari variabel acak Bernoulli dengan probabilitas sukses $p = 0{,}5$. Jumlah total tanda plus *D* merupakan variabel acak binomial dengan parameter n dan p.Oleh karena itu, jumlah rata-rata tanda positif dalam sampel berukuran n adalah $np = \frac{n}{2}$ dan standar deviasinya :

$$\sqrt{np(1-p)} = \sqrt{\frac{n}{4}}$$

- **Kriteria Keputusan :** Jika hipotesis nol benar dan ukuran sampel n besar, z_+ mengikuti berdistribusi normal dengan mean 0 dan standar deviasi 1. Uji ini disebut uji tanda karena hanya bergantung pada tanda-tanda perbedaan yang dihitung.

$$z_+ = \frac{D - \left(\frac{n}{2}\right)}{\sqrt{n/4}}$$

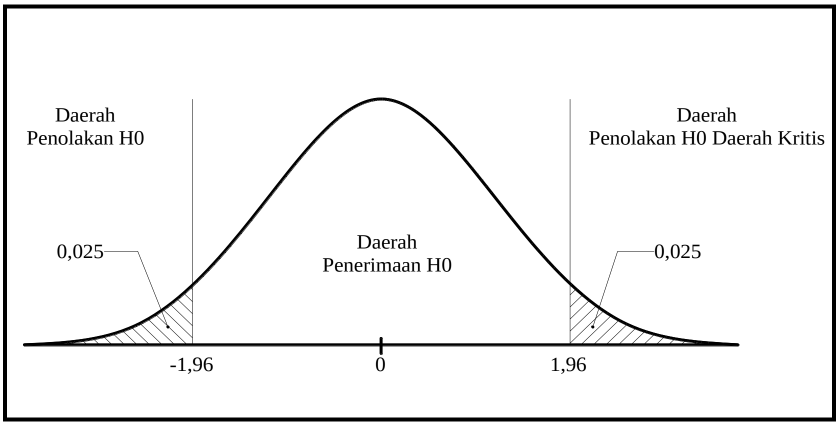
  H_0 ditolak jika $P(k \leq 2 \mid 11{,}0.5) \leq 0.05.$ atau pada distribusi normal p-value $\leq 0.05.$
- **Hitung Uji Statistik :** Pada tabel A, jumlah tanda positifnya $D = 11$ dengan

$$\frac{n}{2} = \frac{13}{2} = 6.5 \rightarrow rata - rata$$

$$\sqrt{\frac{n}{4}} = \sqrt{\frac{13}{4}} = \sqrt{3.25} = 1.80 \rightarrow standar\ deviasi.$$

Dimana

$$z_{+} = \frac{D - \left(\frac{n}{2}\right)}{\sqrt{n/4}} = \frac{11 - 6.5}{1.80} = 2.50$$



Daerah kurva dibawah normal standar pada kanan $z = 2.50$ dan pada kiri $z = -2.50$ dengan p-value $= 2\ (0.006) = 0.012$.

- **Keputusan :** Berdasarkan kurva diatas dan nilai p-value maka p-value $= 0.012 \leq 0.05$ yang artinya H_0 ditolak.

- **Kesimpulan :** Dapat disimpulkan bahwa energi yang dikeluarkan pada saat istirahat lebih tinggi diantara orang-orang dengan fibrosis kistik daripada diantara orang-orang sehat. Hal ini dapat disebabkan oleh sejumlah faktor, termasuk perbedaan dalam metabolisme dan usaha yang dibutuhkan untuk bernapas.

- ***p*-value :** Nilai p-value untuk uji ini adalah 0.012.

Jika ukuran sampel n kecil, kurang dari sekitar 20, statistik uji z_{+} tidak selalu dapat diasumsikan memiliki distribusi normal standar.Dalam hal ini, menggunakan pro-Prosedur untuk mengevaluasi H_0. Ingatlah bahwa di bawah hipotesis nol, D adalah variabel acak binomial dengan parameter n dan $p = \frac{1}{2}$. Oleh karena itu, dapat menggunakan distribusi binomial.

$$P(D \geq 11) = P(D = 11) + P(D = 12) + P(D = 13)$$

$$= \binom{13}{11}(0.5)^{11}(0.5)^{13-11} + \binom{13}{11}(0.5)^{12}(0.5)^{13-12} +$$

$$\binom{13}{11}(0.5)^{13}(0.5)^{13-13}$$

$$= 0.0095 + 0.0016 + 0.0001 = 0.0112$$

$P - value = 2(0.0112) = 0.0224\ \leq 0.05$. Yang artinya H_0 ditolak.

Dengan menggunakan software R maka,

```
> binom.test(2,13,0.5,alternative="two.sided")
```

```
Exact binomial test


data:  2 and 13
number of successes = 2, number of trials = 13, p-value = 0.02246
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.01920667 0.45447106
sample estimates:
probability of success
          0.1538462
```

Keterangan :

- Nilai 2 berarti jumlah tanda paling sedikit, yakni jumlah tanda negatif.
- Nilai 14 berarti jumlah seluruh tanda, yakni tanda positif dan tanda negatif.
- *Two.sided* berarti pengujian dua arah.
- Perhitungan nilai probabilitas kumulatif $X = 2$ dihitung dengan rumus binomial.

Perhatikan bahwa karena nilai probabilitas kumulatif untuk $X = 2$, yakni $0.02246 < 0.05$, maka hipotesis nol ditolak dan hipotesis alternatif diterima. Ini berarti pernyataan terdapat perbedaan yang signifikan (secara statistik) mengenai energi yang dikeluarkan pada saat istirahat lebih tinggi diantara orang-orang dengan fibrosis kistik daripada diantara orang-orang sehat.Hal ini dapat disebabkan oleh sejumlah faktor, termasuk perbedaan dalam metabolisme dan usaha yang dibutuhkan untuk bernapas.

5.3. The Wilcoxon Signed-Rank Test (Uji Peringkat Bertanda Wilcoxon)

Uji peringkat bertanda Wilcoxon dikembangkan oleh Frank Wilcoxon. Uji peringkat bertanda Wilcoxon dan uji tanda sama-sama menguji dua buah populasi berpasangan. Pada uji tanda hanya memperhatikan arah (*direction*) dari selisih untuk setiap pasangan nilai data, sedangkan pada uji Wilcoxon, selain memperhatikan arah (tanda positif $+$ atau tanda negatif $-$) dari selisih untuk setiap pasangan nilai data, juga mengukur jarak atau besar (*magnitude*) dari selisih untuk setiap pasangan nilai data. Oleh karena itu, uji peringkat bertanda Wilcoxon lebih banyak memberikan informasi dibandingkan uji tanda.

Contoh 5.2. Seorang peneliti ingin menyelidiki penggunaan obat amilorida sebagai terapi untuk pasien dengan fibrosis kistik. Hal ini diyakini bahwa obat ini dapat membantu untuk meningkatkan aliran udara di paru-paru dan dengan demikian menunda hilangnya fungsi paru yang sering dikaitkan dengan penyakitnya.Kapasitas vital paksa (FVC) adalah volume udara yang dapat dikeluarkan seseorang dari paru-paru dalam 6 detik; Ia ingin membandingkan penurunan FVC yang terjadi selama periode 25 minggu pengobatan dengan obat untuk pengurangan yang terjadi pada pasien yang sama selama periode waktu yang sama selama pengobatan dengan plasebo. Berikut data Kapasitas vital paksa (FVC),

Subjek	Reduksi FVC (ml)	
	Placebo	Obat
1	224	213
2	80	95
3	75	33
4	541	440
5	74	-32
6	85	-28
7	293	445
8	-23	-178
9	525	367
10	-38	140
11	508	323
12	255	10
13	525	65
14	1023	343

Tabel Γ

Apakah median dari kedua populasi berbeda? (α = 0.05)

Penyelesaian :

- **Data :** Berikut data selisih kedua sampel :

Subjek	Reduksi FVC (ml)		Beda	Peringkat	Tanda Peringkat	
	Placebo	Obat				
1	224	213	11	1	1	
2	80	95	-15	2		-2
3	75	33	42	3	3	
4	541	440	101	4	4	
5	74	-32	106	5	56	
6	85	-28	113	6	7	
7	293	445	-152	7		-7
8	-23	-178	155	8	8	
9	525	367	158	9	9	
10	-38	140	-178	10		-10
11	508	323	185	11	11	
12	255	10	245	12	12	
13	525	65	460	13	13	
14	1023	343	680	14	14	
					86	-19

Tabel Σ

Dengan memilih sampel acak dari n pasang pengamatan.Tabel Σ menunjukkan ukuran laporan pengurangan/reduksi FVC untuk sampel dari 14 pasien dengan fibrosis kistik.

- **Asumsi :** Asumsi pengukuran diambil pada beda nilai pasangan adalah variabel kontinu.

- **Hipotesis :** Hipotesis nol yang diuji pada uji tanda pada dasarnya menyatakan tidak terdapat pengaruh dari kedua populasi. Dengan kata lain, probabilitas untuk memperoleh tanda positif $(+)$ sama dengan probabilitas untuk memperoleh tanda negatif $(-)$, yakni

H_0 : median kedua populasi sama $[P(+) = P(-)]$

H_1 : median kedua populasi berbeda $[P(+) \neq P(-)]$

Taraf signifikansi $\alpha = 0.05$.

- **Uji Statistik :** Hitung perbedaan untuk setiap pasangan pengamatan dan peringkat nilai absolutnya dari yang terkecil hingga terbesar. Perbedaan 0 tidak memiliki peringkat (dihilangkan dari analisis), dan ukuran sampel dikurangi 1 untuk masing-masing pasangan dieliminasi. Pengamatan terikat diberi peringkat rata-rata. Jika dua terkecil Perbedaan keduanya mengambil nilai 11, misalnya, maka masing-masing akan menerima peringkat $(1 + 2)/2 = 1{,}5$. Tetapkan setiap peringkat dengan tanda positif atau negatif. Uji Statistik untuk uji tanda menggunakan jumlah peringkat positif dan jumlah peringkat negatif dalam sampel. Dengan mengabaikan tanda-tandanya, nyatakan jumlah yang lebih kecil dengan T.

- **Distribusi pada Uji Statistik :** Di bawah hipotesis nol bahwa median populasi yang mendasari perbedaan adalah sama dengan 0, Dengan harapan, sampel memiliki jumlah positif dan peringkat negatif yang sama. Selain itu, jumlah peringkat positif harus sebanding besarnya jumlah peringkat negatif. Peneliti mengevaluasi hipotesis ini dengan memperhatikan statistik.

- **Kriteria Keputusan :** Jika hipotesis nol benar dan ukuran sampel n besar, z_+ mengikuti berdistribusi normal dengan mean 0 dan standar deviasi 1. Uji ini disebut uji tanda karena hanya bergantung pada tanda-tanda perbedaan yang dihitung.

$$z_T = \frac{T - \mu_T}{\sigma_T}$$

Dimana

$$\mu_T = \frac{n(n+1)}{4} \rightarrow jumlah\ mean\ peringkat$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} \rightarrow standar\ deviasi$$

H_0 ditolak jika distribusi normal p-value ≤ 0.05.

- **Hitung Uji Statistik :** Pada tabel Σ, jumlah peringkat positif sebanyak 86 dan jumlah peringkat negatif sebanyak $|-19| = 19$ sehingga

$$\mu_T = \frac{n(n+1)}{4}$$

$$= \frac{14(14+1)}{4} = 52.5$$

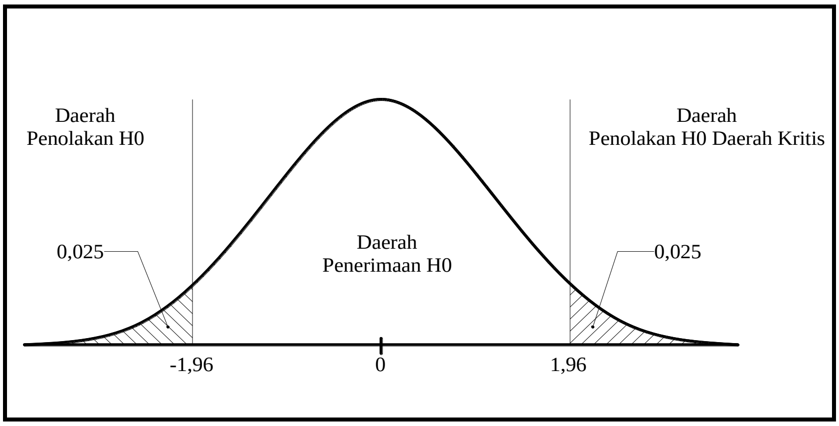
$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$= \sqrt{\frac{14(14+1)(2(14)+1)}{24}} = 15.93$$

Dengan distribusi normalnya,

$$z_T = \frac{T - \mu_T}{\sigma_T}$$

$$= \frac{19 - 52.5}{15.93} = -2.10.$$



Daerah kurva dibawah normal standar pada kanan $z = 2.10$ dan pada kiri $z = -2.10$ dengan p-value $= 2(0.018) = 0.036$.

- **Keputusan :** Berdasarkan kurva diatas dan nilai p-value maka p-value $= 0.036 \leq 0.05$ yang artinya H_0 ditolak.

- **Kesimpulan :** Dapat disimpulkan sebagian besar menunjukkan bahwa pengurangan kapasitas vital paksa lebih besar selama pengobatan dengan plasebo dibandingkan selama pengobatan dengan obat. Dengan kata lain, penggunaan obat tidak mengurangi hilangnya fungsi paru.

- **p-value :** Nilai p-value untuk uji ini adalah 0.036.

Jika n kecil, statistik uji Zr tidak dapat diasumsikan mengikuti distribusi normal standar. Dalam hal ini, tersedia tabel A.6 untuk menentukan apakah harus menolak hipotesis nol. Tabel A.6 di Lampiran A (*BIOSTATISTICS Wayne W. Daniel Edisi 10*) menampilkan fungsi distribusi jumlah yang lebih kecil dari barisan T untuk sampel dengan ukuran n kurang dari atau sama dengan 12. Kemungkinan nilai-nilai T, diwakili oleh T_0, terdaftar di sisi kiri bawah tabel; ukuran sampel ditampilkan di bagian atas. Untuk setiap kombinasi T_0 dan n, dengan memasukkan entri pada tabel adalah probabilitas T kurang dari atau sama dengan T_0. Jika $n = 8$, misalnya, probabilitas bahwa T kurang dari atau sama dengan 5 adalah $0{,}0391$. Ini adalah nilai p dari satu sisi uji hipotesis. Nilai p dari uji dua sisi yang sesuai kira-kira $2\ (0{,}0391) = 0{,}0782$.

Dengan menggunakan software R,

1. Impor data di **R Commander**

```
>contoh2 <- read.table("E:/contoh_wilcoxon.csv", header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)

>X1<-contoh2$Placebo

>X2<-contoh2$Obat
```

2. **Uji wilcoxon**

```
>wilcox.test(X1,X2,paired=TRUE,correct=FALSE)
```

```
Wilcoxon signed rank test


data:  X1 and X2
V = 86, p-value = 0.03528
alternative hypothesis: true location shift is not equal to 0
```

Berdasarkan hasil diatas, diketahui nilai *p-value* (sebelum koreksi) adalah 0.036, di mana nilai tersebut merupakan nilai probabilitas kumulatif nilai normal standar Z, yakni $Z = -2{,}10$.Karena nilai probabilitas $0{,}036 < 0.05$, maka hipotesis nol ditolak dan hipotesis alternatif diterima.Hal ini berarti pernyataan terdapat perbedaan yang signifikan secara statistika mengenai pengurangan kapasitas vital paksa selama pengobatan dengan plasebo dan selama pengobatan dengan obat. Dengan kata lain, penggunaan obat tidak mengurangi hilangnya fungsi paru.

5.4. The Wilcoxon Rank Sum Test (Uji Jumlah Peringkat Wilcoxon)

Uji *Wilcoxon rank sum test* digunakan untuk membandingkan dua sampel yang telah diambil populasi independen.Akibatnya, ini adalah pasangan nonparametrik dari dua sampel t-test. Berbeda dengan uji-t, ini tidak mengharuskan populasi yang mendasarinya terdistribusi normal atau variansnya sama.Anggap bahwa distribusi memiliki bentuk umum yang sama.Uji jumlah peringkat Wilcoxon mengevaluasi hipotesis nol bahwa median dari dua populasi adalah identik.

Contoh 5.3. Sebuah distribusi skor usia mental yang dinormalisasi untuk dua populasi anak penderita fenilketonuria (PKU). Individu dengan gangguan ini tidak mampu memetabolisme protein fenilalanin. Telah disarankan bahwa tingkat yang lebih tinggi fenilalanin serum meningkatkan kemungkinan anak mengalami defisiensi mental. Para anggota kelompok pertama memiliki rata-rata kadar fenilalanin serum harian di bawah 10,0 mg/dl; kelompok kedua memiliki kadar rata-rata di atas 10,0 mg/dl. Seorang peneliti ingin membandingkan skor usia mental normal untuk dua populasi anak-anak. Namun, ia tidak ingin berasumsi bahwa skor usia mental yang dinormalisasi didistribusikan secara normal pasien dengan gangguan ini.

Berikut datanya,

Low exposure ($< 10.0\ mg/dl$)		High exposure ($\geq 10.0\ mg/dl$)	
nMA(mos)	Rank	nMA(mos)	Rank
34.5	2.0	28.0	1.0
37.5	6.0	35.0	3.0
39.5	7.0	37.0	4.5
40.0	8.0	37.0	4.5
45.5	11.5	43.5	9.0
47.0	14.5	44.0	10.0
47.0	14.5	45.5	11.5
47.5	16.0	46.0	13.0
48.7	19.5	48.0	17.0
49.0	21.0	48.3	18.0
51.0	23.0	48.7	19.5
51.0	23.0	51.0	23.0
52.0	25.5	52.0	25.5
53.0	28.0	53.0	28.0
54.0	31.5	53.0	28.0

<table>
<tr><td>54.0</td><td>31.5</td><td>54.0</td><td>31.5</td></tr>
<tr><td>55.0</td><td>34.5</td><td>54.0</td><td>31.5</td></tr>
<tr><td>56.5</td><td>36.0</td><td>55.0</td><td>34.5</td></tr>
<tr><td>57.0</td><td>37.0</td><td rowspan="4"></td><td rowspan="4">313.0</td></tr>
<tr><td>58.5</td><td>38.5</td></tr>
<tr><td>58.5</td><td>38.5</td></tr>
<tr><td></td><td>467.0</td></tr>
</table>

Tabel $\boldsymbol{\beta}$

Apakah median dari kedua populasi berbeda? ($\alpha = 0.05$)

Penyelesaian :

- **Data :** Independen sampel acak dari masing-masing populasi yang diminati. Tabel β menampilkan sampel diambil dari dua populasi anak dengan phenylketonuria (PKU); terdapat 21 anak dengan eksposur rendah dan 18 anak dengan eksposur tinggi.

- **Asumsi :** Asumsi pengukuran diambil pada beda nilai pasangan adalah variabel kontinu.

- **Hipotesis :** Hipotesis nol yang diuji pada dasarnya menyatakan tidak terdapat pengaruh dari kedua populasi. Dengan kata lain, probabilitas untuk memperoleh tanda positif $(+)$ sama dengan probabilitas untuk memperoleh tanda negatif $(-)$, yakni

  H_0 : Median kedua populasi sama $[P(+) = P(-)]$

  H_1 : Median kedua populasi berbeda $[P(+) \neq P(-)]$

  Taraf signifikansi $\alpha = 0.05$.

- **Uji Statistik :** Gabungkan dua sampel menjadi satu kelompok besar, urutkan pengamatan dari yang terkecil ke terbesar dan tetapkan peringkat keberapa. Jika ada pengamatan terikat, peneliti menetapkan peringkat rata-rata untuk semua pengukuran dengan nilai yang sama. Perhatikan dua anak dalam sampel skor usia mental normal 37.0 bulan. Karena pengamatan ini adalah yang keempat dan kelima dalam daftar urutan 39 pengukuran, Sehingga menetapkan peringkat rata-rata $\frac{4+5}{2} = 4{,}5$. Demikian pula, tiga subjek memiliki skor usia mental yang dinormalisasi 51.0 bulan; Pengamatan ini masing-masing menerima jawaban $\frac{22+23+24}{3} = 23$.

- **Distribusi pada Uji Statistik :** Di bawah hipotesis nol bahwa median populasi yang mendasari perbedaan adalah sama dengan 0. Peneliti mengevaluasi hipotesis ini dengan mempertimbangkan statistik.

- **Kriteria Keputusan :** Jika hipotesis nol benar dan ukuran sampel n besar, z_+ mengikuti berdistribusi normal dengan mean 0 dan standar deviasi 1. Uji ini disebut uji tanda karena hanya bergantung pada tanda-tanda perbedaan yang dihitung.

$$z_w = \frac{W - \mu_w}{\sigma_w}$$

Dimana

$$\mu_w = \frac{n_s(n_s + n_L + 1)}{2} \rightarrow jumlah\ mean\ peringkat$$

$$\sigma_w = \sqrt{\frac{n_s n_L + 1(n_s + n_L + 1)}{12}} \rightarrow standar\ deviasi$$

H_0 ditolak jika distribusi normal p-value ≤ 0.05.

- **Hitung Uji Statistik :** Dalam pengujian ini adalah menemukan jumlah peringkat yang sesuai untuk masing-masing sampel asli. Jumlah yang lebih kecil dari kedua jumlah tersebut dilambangkan dengan W. Di bawah hipotesis nol bahwa populasi yang mendasari memiliki median yang identik, diharapkan peringkat untuk didistribusikan secara acak antara dua kelompok. Oleh karena itu, peringkat rata-rata untuk setiap sampel harus kira-kira sama. Uji hipotesis ini dengan dihitung dengan uji statistik

$$\mu_w = \frac{n_s(n_s + n_L + 1)}{2}$$

$$= \frac{18(18 + 21 + 1)}{2}$$

$$= 360$$

Dan

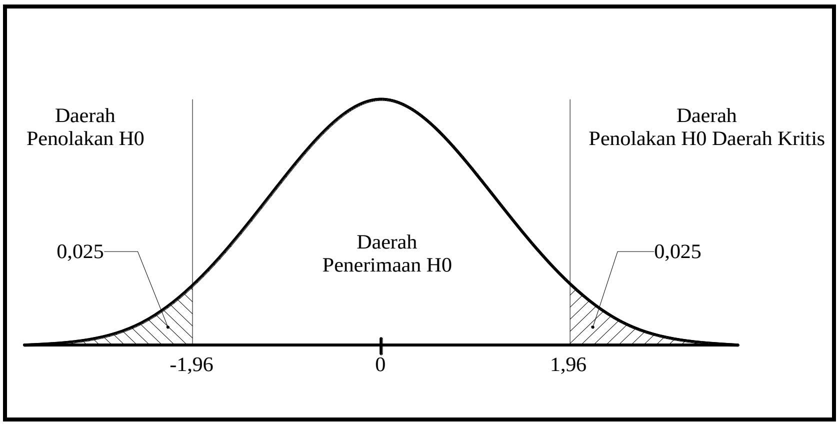
$$\sigma_w = \sqrt{\frac{n_s n_L(n_s + n_L + 1)}{12}}$$

$$= \sqrt{\frac{18(21)(18 + 21 + 1)}{12}} = 35.5$$

Dengan distribusi normalnya,

$$z_w = \frac{W - \mu_w}{\sigma_w}$$

$$= \frac{313 - 360}{35.5} = -1.32$$



Daerah kurva dibawah normal standar terletak pada daerah penerimaan H_0 dengan p-value $= 2(0.093) = 0.186$.

- **Keputusan :** Berdasarkan kurva diatas dan nilai p-value maka p-value $= 0.186 > 0.05$ yang artinya H_0 gagal ditolak.

- **Kesimpulan :** Sampel tidak memberikan bukti perbedaan median usia mental yang dinormalisasi skor untuk dua populasi; anak-anak dengan paparan serum fenilalanin yang tinggi mengalami mencapai tingkat fungsi mental yang sebanding dengan tingkat untuk anak-anak eksposur rendah.

- ***p*-value :** Nilai p-value untuk uji ini adalah 0.186.

Jika n_S dan n_L sangat kecil, Z_w tidak selalu dapat diasumsikan mengikuti distribusi standar normal. Tabel A.7 di Lampiran A (BIOSTATISTICS Wayne W. Daniel Edisi 10) menunjukkan fungsi distribusi dari jumlah yang lebih kecil dari barisan W untuk ukuran sampel $n5$ dan nL yang masing-masing kurang dari atau sama dengan 10. Nilai yang mungkin dari W, dilambangkan dengan W_0, ditampilkan di sisi kiri t-table; nilai n_1 dan n_2 muncul di atas. Dalam kasus ini, n_2 adalah yang lebih besar dari dua sampel ukuran n_S dan $n_{L,}$ dan paling kecil n_1. Untuk setiap kombinasi pada W_0, n_1, n_2, entri pada tabel probabilitas $W \leq W_0$. Misalkan $n_S = 5$ dan $n_L = 4$ dimana $n_1 = 4$ dan $n_2 = 5$, probabilitas $W \leq 13$ adalah 0.0556 yang memiliki p-value $= 2(0.0556) = \ 0.1112$ (dua arah).

5.5. Keuntungan dan Kerugian dari Metode Nonparametrik

Teknik nonparametrik memiliki beberapa keunggulan dibandingkan metode tradisional inferensi statistik. Salah satu keuntungannya adalah ia tidak memasukkan semua karakteristik asumsi batasan uji parametrik. Metode ini tidak mengharuskan populasi yang mendasarinya didistribusikan secara normal. Selain itu, karena uji nonparametrik berhubungan dengan peringkat daripada dengan nilai pengamatan yang sebenarnya, maka metode nonparametrik dapat dilakukan dengan relatif cepat untuk sampel kecil. Penggunaan rank (peringkat) membuat teknik nonparametrik kurang sensitif terhadap kesalahan pengukuran daripada uji tradisional. Karena tidak masuk akal untuk menghitung rata-rata atau simpangan baku untuk nilai ordinal, uji parametrik biasanya tidak sesuai digunakan untuk itu.

Metode nonparametrik juga memiliki sejumlah kelemahan. Jika asumsi yang mendasari uji parametrik terpenuhi, uji nonparametric kurang kuat daripada teknik parametrik yang sebanding. Ini berarti bahwa jika hipoujiis null palsu, uji nonparametric akan membutuhkan sampel yang lebih besar untuk memberikan bukti yang cukup untuk menolaknya. Jika data berasal dari populasi normal yang mendasarinya, kekuatan uji Wilcoxon adalah sekitar 95% dari uji-t; jika uji t membutuhkan 19 pengamatan untuk mencapai kekuatan tertentu, uji Wilcoxon membutuhkan 20 pengamatan untuk mencapai kekuatan yang sama. Selain itu, hipoujiis yang diuji oleh teknik nonparametric cenderung kurang spesifik daripada yang diuji dengan metode tradisional. Karena mereka mengandalkan peringkat daripada pada nilai-nilai aktual dari pengamatan, uji nonparametric tidak menggunakan semua yang diketahui tentang distribusi. Ini, tentu saja, menganggap bahwa informasi (asumsi) kami tentang populasi yang mendasarinya benar. Akhirnya, jika proporsi besar sebagian pengamatan diikat, maka σ_T dan σ_W melebihi estimasi standar simpangan baku T dan W masing-masing. Untuk mengkompensasi hal ini, istilah koreksi harus ditambahkan ke perhitungan.

5.6. Aplikasi Lebih Lanjut

Sebuah penelitian dilakukan untuk menyelidiki penggunaan *extracorporeal membrane oxygenation (ECMO)* sistem mekanis untuk mengoksigenasi darah-dalam pengobatan bayi baru lahir dengan gagal pernapasan neonatal. Diperkirakan bahwa penggunaan prosedur ini dapat mengurangi output ventrikel kiri bayi, sehingga mengurangi jumlah darah yang dipompa ke tubuh. Dengan demikian, kami ingin membandingkan dimensi ventrikel kiri sebelum dan selama penggunaan ECMO. Kami tidak bersedia berasumsi bahwa populasi perbedaan dalam dimensi ventrikel kiri biasanya didistribusikan; oleh karena itu, kami menggunakan uji peringkat ditandatangani Wilcoxon untuk mengevaluasi hipoujiis null bahwa perbedaan median sama dengan 0.

LATIHAN NONPARAMETRIK

1. Bagaimana uji nonparametrik berbeda dari yang parametrik?

   Penyelesaian :

   Berikut tabel perbedaan keduanya :

Statistik Parametrik	Statistik Nonparametrik
Memerlukan asumsi distribusi dari data yang digunakan. Biasanya distribusi data yang diperlukan adalah distribusi normal.	Tidak memerlukan asumsi distribusi sehingga sebaran data bebas.
Memerlukan jenis data bersifat metrik (kuantitatif). Bisa dikatakan juga bahwa data yang digunakan hanya data dalam bentuk data interval atau rasio.	Memerlukan data metrik (kuantitaif) dan nonmetrik (kualitatif). Atau bisa dikatakan bahwa data yang digunakan bisa dalam bentuk data nominal, ordinal, interval atau rasio.
Biasanya jumlah data yang digunakan lebih atau sama dengan 30 (data berukuran besar), sebab data yang lebih atau sama dengan 30 diasumsikan mengikuti teorema limit pusat (*central limit teorema*).	Biasanya jumlah data yang digunakan kurang dari 30.

2. Apa kelebihan dan kekurangan menggunakan uji tanda untuk menganalisis pengamatan berpasangan?

   Penyelesaian :

   Apabila skala pengukuran yang dipakai begitu kecil sehingga data mentah tidak menyediakan informasi tambahan yang cukup maka uji tanda mungkin merupakan uji yang terbaik untuk membuat penyimpulan berlandaskan data tersebut. Tetapi apabila data yang diamati mengandung informasi yang lebih dari itu (cukup banyak) maka uji tanda bukan lagi pilihan yang terbaik karena tidak dapat memanfaatkan informasi dengan baik.

3. Bagaimana uji Wilcoxon signed-rank lebih meningkat dari uji tanda?

   Penyelesaian:

   Uji Wilcoxon signed-rank merupakan penyempurnaan dari uji tanda. Uji Wilcoxon ini hampir sama dengan Uji Tanda tetapi besarnya selisih nilai angka antara positif dan negatif diperhitungkan, dan digunakan untuk menguji hipotesis komparatif 2 sampel berpasangan.Uji Wilcoxon lebih peka daripada uji tanda dalam menentukan perbedaan antara rataan populasi dan karena itu perlu dibahas secara mendalam. Jika sampel berpasangan lebih besar dari 25, maka distribusinya dianggap akan mendekati distribusi normal. Untuk itu digunakan Z sebagai Uji Statistiknya

4. Bagaimana asumsi uji Wilcoxon rank sum test berbeda dari uji t dua sampel yang mendasarinya?

   Penyelesaian :

   Pada uji Wilcoxon skala pengukurannya minimal ordinal. dan tidak butuh asumsi normalitas. Inilah yang membedakan dengan uji t berpasangan. disini ada dua keadaan dalam menggunakan wilcoxon. Pertama. ketika data yang digunakan ordinal maka pakai wilcoxon. kasus kedua ketika datanya tuh interval atau rasio maka pertama kali lihat dulu apakah normal atau tidak. kalau normal pakai uji t berpasangan dan jika tidak normal baru pakai wilcoxon. Beberapa peneliti juga mengatakan ketika data yang digunakan lebih dari 25, ada juga yang mengatakan lebih dari 30. maka pakai uji t berpasangan. alasannya dengan data yang 30 (dikatakan sampel besar) itu akan mendekati data normal.

5. Apa kelebihan dan kekurangan menggunakan peringkat daripada pengukuran berkelanjutan untuk melakukan uji hipotesis?

   Penyelesaian :

   Kelebihan untuk data statistic nonparametric akan dihasilkan uji yang sesuai dengan jenisnya. Kelemahannya untuk data berapasangan dengan serupa.

6. Misalkan Anda tertarik untuk memeriksa efek transisi dari sirkulasi janin ke postnatal di antara bayi prematur. Untuk masing-masing dari 14 bayi baru lahir yang sehat, tingkat pernapasan diukur pada dua waktu yang berbeda-sekali ketika bayi berusia kurang dari 15 hari, dan sekali lagi ketika dia berusia lebih dari 25 hari.

Subjek	Tingkat Respirasi (napas/menit)	
	Waktu 1	Waktu
1	62	46
2	35	42
3	38	40
4	80	42
5	48	36
6	48	46
7	68	45
8	26	40
9	48	42
10	27	40
11	43	46
12	67	31
13	52	44
14	88	48

a) Menggunakan uji tanda, evaluasi hipotesis null bahwa perbedaan median dalam tingkat pernapasan untuk dua kali sama dengan 0?

b) Mengevaluasi hipotesis yang sama menggunakan uji peringkat bertanda Wilcoxon?

c) Apakah Anda mencapai kesimpulan yang sama dalam setiap kasus?

Penyelesaian :

a) Uji Tanda

Script

```
library(readxl)
Latihan7 <- read_excel("E:/Latihan7.xlsx")
View(Latihan7)
X <- Latihan7$X
Y <- Latihan7$Y
Y-X
binom.test(5,14,0.5,alternative="two.sided")
```

Console

```
> View(Latihan7)
> X <- Latihan7$X
> Y <- Latihan7$Y
> Y-X
 [1] -16   7   2 -38 -12  -2 -23  14  -6  13
[11]   3 -36  -8 -40
> binom.test(5,14,0.5,alternative="two.sided")

	Exact binomial test

data:  5 and 14
number of successes = 5, number of
trials = 14, p-value = 0.424
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1275984 0.6486199
sample estimates:
probability of success 
             0.3571429
```

Tanda positif : 5

Tanda negative : 9

Nol : 0

Observasi relevan : 14

b) Uji Wilcoxon

Script

```
X <- Latihan7$X
Y <- Latihan7$Y
X
Y
wilcox.test(X, Y, paired=TRUE)
```

Console

```
> X <- Latihan7$X
> Y <- Latihan7$Y
> X
 [1] 62 35 38 80 48 48 68 26 48 27 43 67 52
[14] 88
> Y
 [1] 46 42 40 42 36 46 45 40 42 40 46 31 44
[14] 48
> wilcox.test(X, Y, paired=TRUE)

	Wilcoxon signed rank test with
	continuity correction

data:  X and Y
V = 78.5, p-value = 0.1093
alternative hypothesis: true location shift is not equal to 0
```

c) Kesimpulan

1. Uji Tanda

   Diketahui nilai probabilitas kumulatif dari X = 5 atau p-value = 0,424. Perhatikan bahwa karena nilai probabilitas kumulatif untuk X = 4, yakni 0,424, lebih besar dari nilai tingkat signifikansi, yakni 0,05, maka hipotesis nol diterima dan hipotesis alternatif ditolak. Artinya pernyataan terdapat perbedaan (secara statistic) antara waktu 1 dan 2 tidak dapat diterima pada tingkat signifikansi 5%

2. Wilcoxon Sign-Rank

   Perhatikan bahwa berdasarkan nilai V = 78.5, di mana nilai tersebut merupakan nilai statistik dari uji Wilcoxon. Karena nilai statistik dari uji Wilcoxon, yakni 78.5 lebih besar dari nilai kritis Wilcoxon, yakni 21, maka hipotesis nol diterima dan hipotesis alternatif ditolak. Ini berarti pernyataan mengenai "terdapat perbedaan yang signifikan (secara statistika) mengenai berat badan, sebelum mengkonsumsi obat penambah berat badan merek XYZ dan sesudah mengkonsumsi obat penambah berat badan 89 merek XYZ selama satu minggu" dapat diterima pada tingkat signifikansi 5%.

BAB VI

ANALISIS SURVIVAL

6.1. Pendahuluan

Analisis survival (analisis kelangsungan hidup) adalah prosedur statistika untuk menganalisis data dengan waktu sampai terjadinya suatu peristiwa tertentu (*time until an event occurs*) sebagai variabel respons.

"Peristiwa tertentu" tersebut dalam analisis survival lazimnya disebut sebagai '**kegagalan**' (*failure*), yang dapat berupa :

- Kematian pada penderita penyakit fatal
- Eksaserbasi ulang pada penderita penyakit kronis dengan remisieksasebasi yang semula ada dalam fase remisi
- Kekambuhan pada eks-pecandu narkotika sehabis menjalani rehabilitasi dll.

Misalkan T adalah variabel random non-negatif yang menyatakan waktu sampai dengan terjadinya kegagalan, maka $f(t)$ menyatakan fungsi densitas probabilitasnya dan $F(t) = P(T < t)$ menyatakan fungsi distribusi kumulatifnya.

Komplemen fungsi distribusi kumulatif $F(t)$ adalah fungsi survival $S(t)$, yaitu probabilitas bahwa subjek survive lebih lama daripada waktu t atau probabilitas bahwa variabel random T melebihi waktu t :

$$S(t) = P(T > t)$$

$$S(t) = 1 - P(T < t)$$

$$S(t) = 1 - F(t)$$

Fungsi densitas $f(t)$ adalah turunan pertama $F(t)$ terhadap t, yaitu:

$$f(t) = \frac{dF(t)}{dt} = \frac{d\{1 - S(t)\}}{dt} = -S'(t)$$

Fungsi hazard $h(t)$ yang juga dikenal sebagai "laju kegagalan bersyarat" (*conditional failure rate*) atau 'kekuatan mortalitas' (*force of mortality*), adalah probabilitas terjadinya kegagalan pada suatu interval waktu, dengan syarat subjek tersebut survive sampai dengan awal interval, dibagi lebar interval, yaitu :

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t} = \frac{f(t)}{S(t)}$$

Fungsi hazard bukan merupakan probabilitas, melainkan berupa rate (kelajuan), sehingga fungsi hazard disebut juga sebagai rate hazard. Jika kegagalan yang dipelajari adalah kematian, maka rate hazard sama dengan rate mortalitas (*mortality rate*).

Fungsi hazard kumulatif $H(t)$ adalah :

$$H(t) = \int_0^t h(u)du$$

$$= \int_0^t \frac{f(u)}{S(u)} du$$

$$= \int_0^t \frac{1}{S(u)} \left\{ \frac{d}{du} S(u) \right\} du$$

$$= -\ln\{S(t)\}$$

Interpretasi fungsi hazard kumulatif yaitu :

a) Ukuran risiko total yang terakumulasi sampai dengan waktu *t*, dan
b) Risiko kumulatif dan probabilitas survival berbanding terbalik.

Dengan demikian maka diperoleh:

$$S(t) = \exp\{-H(t)\}$$

$$F(t) = 1 - \exp\{-H(t)\}$$

$$f(t) = h(t) \exp\{-H(t)\}$$

6.2. Tabel Kehidupan (life table)

Tabel kehidupan (*life table*) adalah tabel yang digunakan untuk melakukan perhitungan probabilitas survive dan risiko kematian pada tiap interval waktu tertentu. Tabel kehidupan dapat dibuat untuk keseluruhan himpunan subjek penelitian dan dapat dirinci lagi menjadi tabel kehidupan untuk masing-masing subkelompoknya.

Secara teoretis, risiko dalam konteks data survival adalah :

$$Risk = P(t < T < t + \Delta t | T > t)$$

Risk merupakan probabilitas bersyarat, yaitu probabilitas subjek untuk mengalami kegagalan dalam suatu interval waktu dengan syarat ia survive sampai dengan awal interval tersebut. Estimasinya dari data sampel adalah:

$$Risk = \frac{Jumlah\ kegagalan\ dalam\ suatu\ interval\ waktu}{Jumlah\ subjek\ survive\ pada\ awal\ interval\ tersebut}$$

Rate secara teoritis dalam konteks data survival adalah :

$$Rate = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}$$

Estimasi rate dari data sampel adalah jumlah kegagalan dalam suatu interval waktu dibagi jumlah waktu pengamatan terhadap seluruh subjek dalam penelitian tersebut. Jumlah waktu pengamatan terhadap seluruh subjek ini dinamakan sebagai *"person-time"*, sehingga estimasi rate adalah :

$$Rate = \frac{\text{Jumlah kegagalan dalam suatu interval waktu}}{\text{Jumlah person} - \text{time}}$$

Bentuk umum tabel kehidupan untuk data survival diperlihatkan pada Tabel E berikut :

i	Δt	PT_i	d_i	$\widehat{Rate_I}$	$\widehat{Risk_\iota}$	$\widehat{Risk_{kum}}$	$\hat{S}(t_i)$	$\hat{S}_{kum}$
1	(0,1)	...	...	...	...	...	...	...
2	(1,2)	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
n	$(n-1, n)$	...	...	...	...	...	...	...
$Total$	$(0, n)$	...	...	...	...	...	...	...

Tabel E

Keterangan :

i : Satuan waktu pengamatan ke-i, $i = 1,2,3, \dots, n$
Δt : Interval waktu ke-i : $(t_0, t_1), (t_1, t_2), \dots, (t_{n-1}, t_n)$
PT_i : Person-time untuk interval waktu ke- i
d_i : Jumlah kegagalan pada interval waktu ke- i
$\widehat{Rate_I}$: Estimasi rate pada interval waktu ke- i

$\widehat{Risk_i}$: Estimasi risiko pada interval waktu ke- i
$\widehat{Risk_{kum}}$: Estimasi risiko kumulatif pada interval waktu (t_0, t_i)
$\hat{S}(t_i)$: Estimasi fungsi survival pada interval waktu ke- i
$\hat{S}_{kum}$: Estimasi fungsi survival kumulatif pada interval waktu (t_0,t_i)

Perhatikan bahwa walaupun salah satu hasil utama dari analisis survival adalah rate, estimasi akhir tetap diperlukan untuk risiko kegagalan.

6.3. Metode Kaplan-Meier

Estimasi Kaplan-Meier dikenal juga sebagai estimasi limit-produk (*product-limit*). Karakteristik estimasi Kaplan-Meier yaitu :

- Interval waktu seringkali tidak sama, perhitungan dilakukan setiap ada kegagalan.
- Fungsi survival untuk suatu interval waktu adalah proporsi jumlah subjek survive pada awal interval dikurangi jumlah kegagalan dalam interval tersebut:

$$\hat{S}(t_i) = \frac{N_{i-1} - d_i}{N_{i-1}}$$

Jumlah sensoring ataupun *withdrawal* tidak diperhitungkan dalam estimasi Kaplan-Meier. Estimasi Kaplan-Meier lazimnya digunakan untuk langsung mengestimasi fungsi survival.

Fungsi Kaplan-Meier yaitu untuk mengestimasi fungsi survival $S(t)$, Kaplan Meier divisualisasikan dalam bentuk kurva.kurvanya semakin panjang waktu t maka $S(t)$ semakin mendekati 0 begitu pula dengan visualisasi kurva Kaplan Meier.

Contoh :

Sebelum melakukan analisis terdapat dua package yang harus diinstal terlebih dahulu yaitu :

“survival” : untuk komputasi analisis survival

```
Console:
>install.packages(c(“survival”))

> library("survival")

> data("lung") #memanggil data kanker paru-paru

> head(lung) #menampilkan 6 data teratas

inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
```

1	3	306	2	74	1	1	90	100	1175	NA
2	3	455	2	68	1	0	90	90	1225	15
3	3	1010	1	56	1	0	90	90	NA	15
4	5	210	2	57	1	1	90	60	1150	11
5	1	883	2	60	1	0	100	90	NA	0
6	12	1022	1	74	1	1	50	80	513	0

Penjelasan variabel dalam dataset :

- `Inst` : kode institusi
- `Time` : waktu survival dalam satuan hari
- `Status` : status penyensoran (1 = tersensor, 2 = meninggal)
- `Age` : usia pasien dalam satuan tahun
- `Sex` : jenis kelamin (1 = laki-laki, 2 = perempuan)
- `Ph.ecog` : skala yang digunakan dokter untuk mengetahui level kondisi pasien (0 = baik, 5 = meninggal)
- `Ph.karno` : skala karnofsky yang dinilai dari dokter (0= buruk,100=baik)
- `Pat.karno` : skala karnofsky dari pasien
- `Meal.cal` : banyak kalori yang dikonsumsi
- `Wt.loss` : banyak berat badan berkurang dalam kurun waktu 6 bulan terakhir

Fungsi **`survfit()`** dalam package survival dapat digunakan untuk menghitung estimasi $\boldsymbol{S(t)S(t)}$. Sebagai contoh kita akan menghitung estimasi $\boldsymbol{S(t)S(t)}$ berdasarkan jenis kelamin, sehingga syntax yang dituliskan adalah :

```
> fit <- survfit(Surv(time, status) ~ sex, data = lung)

> print(fit) #ringkasan dari perhitungan kurva survival

Call: survfit(formula = Surv(time, status) ~ sex, data = lung)
```

```
        n events median 0.95LCL 0.95UCL
sex=1 138    112    270     212     310
sex=2  90     53    426     348     550
> summary(fit) #output lengkap
Call: survfit(formula = Surv(time, status) ~ sex, data = lung)
                sex=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   11    138       3   0.9783  0.0124       0.9542        1.000
   12    135       1   0.9710  0.0143       0.9434        0.999
   13    134       2   0.9565  0.0174       0.9231        0.991
   15    132       1   0.9493  0.0187       0.9134        0.987
   26    131       1   0.9420  0.0199       0.9038        0.982
   30    130       1   0.9348  0.0210       0.8945        0.977
   31    129       1   0.9275  0.0221       0.8853        0.972
   53    128       2   0.9130  0.0240       0.8672        0.961
   54    126       1   0.9058  0.0249       0.8583        0.956
   59    125       1   0.8986  0.0257       0.8496        0.950
   60    124       1   0.8913  0.0265       0.8409        0.945
   65    123       2   0.8768  0.0280       0.8237        0.933
   71    121       1   0.8696  0.0287       0.8152        0.928
   81    120       1   0.8623  0.0293       0.8067        0.922
   88    119       2   0.8478  0.0306       0.7900        0.910
   92    117       1   0.8406  0.0312       0.7817        0.904
   93    116       1   0.8333  0.0317       0.7734        0.898
```

95	115	1	0.8261	0.0323	0.7652	0.892
105	114	1	0.8188	0.0328	0.7570	0.886
107	113	1	0.8116	0.0333	0.7489	0.880
110	112	1	0.8043	0.0338	0.7408	0.873
116	111	1	0.7971	0.0342	0.7328	0.867
118	110	1	0.7899	0.0347	0.7247	0.861
131	109	1	0.7826	0.0351	0.7167	0.855
132	108	2	0.7681	0.0359	0.7008	0.842
135	106	1	0.7609	0.0363	0.6929	0.835
142	105	1	0.7536	0.0367	0.6851	0.829
144	104	1	0.7464	0.0370	0.6772	0.823
147	103	1	0.7391	0.0374	0.6694	0.816
156	102	2	0.7246	0.0380	0.6538	0.803
163	100	3	0.7029	0.0389	0.6306	0.783
166	97	1	0.6957	0.0392	0.6230	0.777
170	96	1	0.6884	0.0394	0.6153	0.770
175	94	1	0.6811	0.0397	0.6076	0.763
176	93	1	0.6738	0.0399	0.5999	0.757
177	92	1	0.6664	0.0402	0.5922	0.750
179	91	2	0.6518	0.0406	0.5769	0.736
180	89	1	0.6445	0.0408	0.5693	0.730
181	88	2	0.6298	0.0412	0.5541	0.716
183	86	1	0.6225	0.0413	0.5466	0.709
189	83	1	0.6150	0.0415	0.5388	0.702

197	80	1	0.6073	0.0417	0.5309	0.695
202	78	1	0.5995	0.0419	0.5228	0.687
207	77	1	0.5917	0.0420	0.5148	0.680
210	76	1	0.5839	0.0422	0.5068	0.673
212	75	1	0.5762	0.0424	0.4988	0.665
218	74	1	0.5684	0.0425	0.4909	0.658
222	72	1	0.5605	0.0426	0.4829	0.651
223	70	1	0.5525	0.0428	0.4747	0.643
229	67	1	0.5442	0.0429	0.4663	0.635
230	66	1	0.5360	0.0431	0.4579	0.627
239	64	1	0.5276	0.0432	0.4494	0.619
246	63	1	0.5192	0.0433	0.4409	0.611
267	61	1	0.5107	0.0434	0.4323	0.603
269	60	1	0.5022	0.0435	0.4238	0.595
270	59	1	0.4937	0.0436	0.4152	0.587
283	57	1	0.4850	0.0437	0.4065	0.579
284	56	1	0.4764	0.0438	0.3979	0.570
285	54	1	0.4676	0.0438	0.3891	0.562
286	53	1	0.4587	0.0439	0.3803	0.553
288	52	1	0.4499	0.0439	0.3716	0.545
291	51	1	0.4411	0.0439	0.3629	0.536
301	48	1	0.4319	0.0440	0.3538	0.527
303	46	1	0.4225	0.0440	0.3445	0.518
306	44	1	0.4129	0.0440	0.3350	0.509

310	43	1	0.4033	0.0441	0.3256	0.500
320	42	1	0.3937	0.0440	0.3162	0.490
329	41	1	0.3841	0.0440	0.3069	0.481
337	40	1	0.3745	0.0439	0.2976	0.471
353	39	2	0.3553	0.0437	0.2791	0.452
363	37	1	0.3457	0.0436	0.2700	0.443
364	36	1	0.3361	0.0434	0.2609	0.433
371	35	1	0.3265	0.0432	0.2519	0.423
387	34	1	0.3169	0.0430	0.2429	0.413
390	33	1	0.3073	0.0428	0.2339	0.404
394	32	1	0.2977	0.0425	0.2250	0.394
428	29	1	0.2874	0.0423	0.2155	0.383
429	28	1	0.2771	0.0420	0.2060	0.373
442	27	1	0.2669	0.0417	0.1965	0.362
455	25	1	0.2562	0.0413	0.1868	0.351
457	24	1	0.2455	0.0410	0.1770	0.341
460	22	1	0.2344	0.0406	0.1669	0.329
477	21	1	0.2232	0.0402	0.1569	0.318
519	20	1	0.2121	0.0397	0.1469	0.306
524	19	1	0.2009	0.0391	0.1371	0.294
533	18	1	0.1897	0.0385	0.1275	0.282
558	17	1	0.1786	0.0378	0.1179	0.270
567	16	1	0.1674	0.0371	0.1085	0.258
574	15	1	0.1562	0.0362	0.0992	0.246

583	14	1	0.1451	0.0353	0.0900	0.234
613	13	1	0.1339	0.0343	0.0810	0.221
624	12	1	0.1228	0.0332	0.0722	0.209
643	11	1	0.1116	0.0320	0.0636	0.196
655	10	1	0.1004	0.0307	0.0552	0.183
689	9	1	0.0893	0.0293	0.0470	0.170
707	8	1	0.0781	0.0276	0.0390	0.156
791	7	1	0.0670	0.0259	0.0314	0.143
814	5	1	0.0536	0.0239	0.0223	0.128
883	3	1	0.0357	0.0216	0.0109	0.117

sex=2

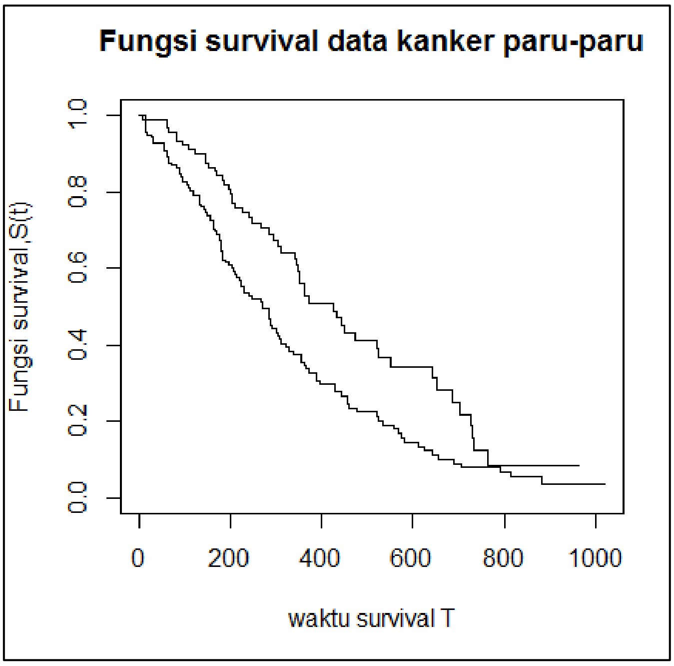
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
5	90	1	0.9889	0.0110	0.9675	1.000
60	89	1	0.9778	0.0155	0.9478	1.000
61	88	1	0.9667	0.0189	0.9303	1.000
62	87	1	0.9556	0.0217	0.9139	0.999
79	86	1	0.9444	0.0241	0.8983	0.993
81	85	1	0.9333	0.0263	0.8832	0.986
95	83	1	0.9221	0.0283	0.8683	0.979
107	81	1	0.9107	0.0301	0.8535	0.972
122	80	1	0.8993	0.0318	0.8390	0.964
145	79	2	0.8766	0.0349	0.8108	0.948
153	77	1	0.8652	0.0362	0.7970	0.939
166	76	1	0.8538	0.0375	0.7834	0.931

167	75	1	0.8424	0.0387	0.7699	0.922
182	71	1	0.8305	0.0399	0.7559	0.913
186	70	1	0.8187	0.0411	0.7420	0.903
194	68	1	0.8066	0.0422	0.7280	0.894
199	67	1	0.7946	0.0432	0.7142	0.884
201	66	2	0.7705	0.0452	0.6869	0.864
208	62	1	0.7581	0.0461	0.6729	0.854
226	59	1	0.7452	0.0471	0.6584	0.843
239	57	1	0.7322	0.0480	0.6438	0.833
245	54	1	0.7186	0.0490	0.6287	0.821
268	51	1	0.7045	0.0501	0.6129	0.810
285	47	1	0.6895	0.0512	0.5962	0.798
293	45	1	0.6742	0.0523	0.5791	0.785
305	43	1	0.6585	0.0534	0.5618	0.772
310	42	1	0.6428	0.0544	0.5447	0.759
340	39	1	0.6264	0.0554	0.5267	0.745
345	38	1	0.6099	0.0563	0.5089	0.731
348	37	1	0.5934	0.0572	0.4913	0.717
350	36	1	0.5769	0.0579	0.4739	0.702
351	35	1	0.5604	0.0586	0.4566	0.688
361	33	1	0.5434	0.0592	0.4390	0.673
363	32	1	0.5265	0.0597	0.4215	0.658
371	30	1	0.5089	0.0603	0.4035	0.642
426	26	1	0.4893	0.0610	0.3832	0.625

```
  433    25      1   0.4698  0.0617      0.3632        0.608
  444    24      1   0.4502  0.0621      0.3435        0.590
  450    23      1   0.4306  0.0624      0.3241        0.572
  473    22      1   0.4110  0.0626      0.3050        0.554
  520    19      1   0.3894  0.0629      0.2837        0.534
  524    18      1   0.3678  0.0630      0.2628        0.515
  550    15      1   0.3433  0.0634      0.2390        0.493
  641    11      1   0.3121  0.0649      0.2076        0.469
  654    10      1   0.2808  0.0655      0.1778        0.443
  687     9      1   0.2496  0.0652      0.1496        0.417
  705     8      1   0.2184  0.0641      0.1229        0.388
  728     7      1   0.1872  0.0621      0.0978        0.359
  731     6      1   0.1560  0.0590      0.0743        0.328
  735     5      1   0.1248  0.0549      0.0527        0.295
  765     3      1   0.0832  0.0499      0.0257        0.270


> summary(fit)$table #ringkasan pendek
      records n.max n.start events   *rmean *se(rmean) median 0.95LCL 0.95UCL
sex=1     138   138     138    112 325.0663   22.59845    270     212     310
sex=2      90    90      90     53 458.2757   33.78530    426     348     550
```

Plot

```
> plot(fit,main="Fungsi survival data kanker paru-paru",xlab="waktu survival T",ylab="Fungsi survival,S(t)")
```



6.4. Metode Log-Rank

Uji log-rank adalah uji statistik untuk memperbandingkan dua atau lebih fungsi survival, baik dalam tabel kehidupan ataupun grafik kurvanya.

Uji log-rank adalah uji khi-kuadrat untuk sampel besar, yang memperbandingkan frekuensi sel *observed* dengan *expected* untuk seluruh kategori interval waktu.Hipotesis nol yang diuji adalah H_0 : Tidak ada perbedaan antar fungsi survival. Statistik pengujinya adalah:

$$x^2 = \sum_{i}^{\#Grup} \frac{(O_i - E_i)^2}{E_i}$$

yang berdistribusi khi-kuadrat dengan derajat bebas G – 1; G menyatakan jumlah grup perbandingan.

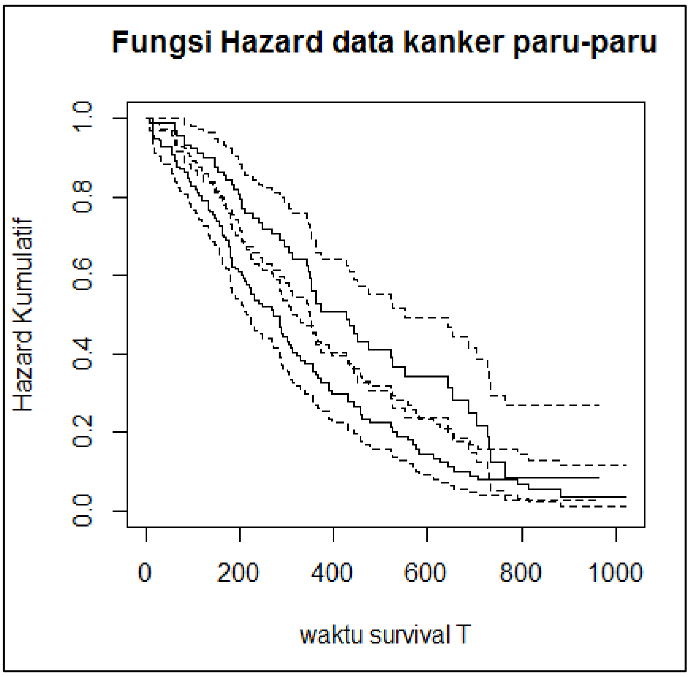
Apabila hanya ingin melakukan uji log rank tanpa visualisasi kaplan-meier maka menggunakan fungsi `survdiff()` pada package survival. Sebagai contoh yaitu syntax berikut untuk uji log rank berdasarkan jenis kelamin :

```
> surv_diff <- survdiff(Surv(time, status) ~ sex, data = lung)
> surv_diff
Call:
survdiff(formula = Surv(time, status) ~ sex, data = lung)

        N Observed Expected (O-E)^2/E (O-E)^2/V
sex=1 138      112     91.6      4.55      10.3
sex=2  90       53     73.4      5.68      10.3

 Chisq= 10.3  on 1 degrees of freedom, p= 0.001
```

Sama halnya dengan penjelasan sebelumnya didapatkan kesimpulan bahwa terdapat perbedaan yang signifikan pada fungsi survival kedua jenis kelamin karena didapatkan hasil pvalue=(0.001) < α=(0.05).

Plot

```
> plot(fit,main="Fungsi Hazard data kanker paru-paru",xlab="waktu survival T",ylab="Hazard Kumulatif",conf.int=TRUE,legend.labs=c("Male","Female",ggtheme=theme_bw(),fun="cumhaz")))
```



Kurva $h(t)$ tidak harus dimulai dari satu dan bergerak menuju nol, tetapi kurva $h(t)$ dapat dimulai dari nilai berapapun dengan syarat $h(t) = 0$ dan dapat bergerak ke atas maupun ke bawah terhadap waktu t.

Daftar Pustaka

Cross, C. L., & Daniel, W. W. (2013). *Biostatistics A Foundation for Analysis in the Health Sciences*. Wiley.

Pagano, M., & Gauvreau, K. (2018). *Principle of Biostatistics*. CRC Press.

Ross, Sheldon M. 2010. *Introduction to Probability Models* (Tenth Edition). California: Elsevier

Wackerly, Dennis D, William Mendenhall III, Richard L Scheaffer. 2008. *Mathematical Statistics with Applications* (Seventh Edition). Florida: Duxbury Press

Soong, T.T. 1988. *Fundamentals of Probability and Statistics for Engineers.* New York: John Wiley and Sons

Ott R lyman, Michael Longnecker. 2015. *An Introduction to Statistics Methods & Data Analysis* (Seventh Edition). Texas: Cengage Learning

Devore, Jay L, Kenneth N Berk. 2012. *Modern Mathematical Statitics with Applications* (Second Edition), New York: Springer

Roussas, G. 2003. *An Introduction to Probability and Statistical Inference.* San Diego: Academic Press

Ross, Sheldon M. 2000. *Introduction to Probability and Statistics for Engineers And Scientist.* San Diego: Harcourt Aca

Taylor, L D. 1974. *Probability and Mathematical Statistics.* New York: Harper & Row

Rinaman, W. C. 1993. *Foundations of Probability and Statistics.* New York: Saunders College Publishing.

Meyer, P. L. 1970. *Introductory Probability and Statistical Applications.* Reading: Addison W

Casella, G., and R. L. Berger. 2002. *Statitical Inference, (Second Edition)*. Pacific Grove, California: Duxbury.

Cramer, H. 1973. *The Elements of Probability Theory and Some of Its Applications, (Second Edition).* Huntington, New York: Krieger

Hogg, R. V., A. T. Craig, and J. W. Mckean. 2005. *Introduction to Mathematical Statistics, (Sixth Edition).* Upper Saddle River, New Jersey: Pearson Prentice Hall

Lindgren, B. W. 1993. *Statistical Theory, (Fourth Edition).* Boca Raton, Florida: Chapman and Hall/CRC.

Miller, I,. and M. Miller. 2003. *John E. Freud's Mathematical Statistics with Applicatins, (Seventh Edition).* Upper Saddle River, New Jersey: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D.Boes. 1974. *Introduction to the Theory of Statistics, (Third Edition).* New York: McGraw-Hill

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, New York: John Wiley.

Parzen, Emanuel. 1960. *Modern Probability Theory and Its Applications*, New York: John Wiley.