**Analisis Regresi Multivariat terhadap Dosis Cancer dengan Program Python**

**Yuni Naomi Yenusi (662016011)**

**Program Studi Matematika Fakultas Sains dan Matematika**

**Universitas Kristen Satya Wacana**

**Abstrak**

Kanker merupakan salah satu penyakit sebagai penyebab kematian. Untuk itu berbagai studi tentang cancer telah dilakukan.Kanker merupakan salah satu penyakit sebagai penyebab kematian. Untuk itu berbagai studi tentang cancer telah dilakukan pada penelitian ini memberikan analisa yang lebih teliti dengan melakukan analisa secara multivariat.**(.....)**

1. **Pendahuluan**

Kanker merupakan salah satu penyakit sebagai penyebab kematian. Untuk itu berbagai studi tentang kancer telah dilakukan. Diantaranya, penelitian dosis yang diperlukan dari medan penyinaran terhadap jaringan kanker diasumsikan dalam bentuk bidang persegi dan persegi panjang dengan beberapa ukuran selama 60 detik dimana hasil menunjukkan dosis yang aman untuk berbagai ukuran yang dikaji dengan kedua bentuk tersebut hampir sama (Sidabutar dan Setiawati,2014). Demikian pula dampak radiasi terhadap berbagai organ tubuh persekitaran jaringan yang terindikasi kanker dipelajari (....dkk,th). Kajian sensitivitas organ terhadap radiasi (OAR) dan batas dosis radiasi dirangkum dalam konsensus Jaringan Terapi Partikel Eropa (Lambrecht,dkk.,2018). Metode statistika yang digunakan dalam analisa pada penelitian tersebut adalah regresi tiap pasang variabel yang diukur. Padahal secara umum, kita dapat memahami bahwa dampak dosis radiasi terhadap jaringan sehat sekitar jaringan kanker terjadi secara simultan. Oleh karena itu, pada penelitian ini memberikan analisa yang lebih teliti dengan melakukan analisa secara multivariat.

1. **Tujuan**
2. **Dasar Teori**
3. **Regresi linear Multivariat**

Analisis regresi adalah suatu metodologi statistika untuk memprediksi nilai dari satu atau lebih variabel respon (variabel dependen) dari koleksi nilai variabel prediktor (variabel independen). Analisis ini juga dapat digunakan untuk memprediksi atau meramal pengaruh dari variabel prediktor (variabel independen) pada respon. Dalam analisis regresi, ada dua jenis variabel yaitu variabel bebas atau variabel prediktor (dinotasikan dengan X) dan variabel tak bebas atau variabel respon (dinotasikan dengan Y). Untuk melihat hubungan antara variabel respon dan sejumlah variabel prediktor secara simultan dapat digunakan analisis regresi linier dengan variabel respon diukur sekurang kurangnya dalam skala interval dam mempunyai distribusi normal.

Data yang digunakan untuk analisa regresi berupa daftar dosis radiasi terhadap jaringan kanker dan besarnya dosis radiasi terhadap jaringan organ tubuh disekitarnya. Pada penelitian ini Y dapat berupa Dosis pada salah satu bagian organ disekitar jaringan yang terkena kanker , misalnya *Brainstem* sedangkan variable X merupakan data dosis radiasi dari jaringan selain *Brainstem* (....dkk,th).ukuran Y adalah 25 sedangkan X berukuran 25 x 8. Secara umum terdapat variabel tak bebas sebanyak *p* variabel dengan asumsi data disimpan dalam matriks random *X* dengan elemen barisnya menyatakan baanyaknya observasi yaitu sebanyak *n* observasi dan kolom sebanyak *p*. Berikut ini adalah tampilan matriks *X* yang ditulis perkomponen yaitu

Model regresi linear berikut ini adalah model dengan variabel tak bebas tunggal *Y*

Dengan banyaknya data sebanyak n maka model yang memenuhi adalah sebagai berikut

Error pada setiap variabel *y* mempunyai sifat-sifat berikut

1. (konstan)
2. .

Untuk selanjutnya proses regresi dilakukan dengan bantuan Python terhadap data yang dikaji

1. **Metode Penelitian**

Dalam pembelajaran Komputasi matematika terkait dengan penggunaan Python untuk mengolah data kami mendapatkan sebuat situs di internet yang memuat data mengenai kanker lebih tepatnya data mengenai dosis toleransi perfraksi organ yang beresiko di otak. Dengan data tersebut kemudian kami mencoba untuk memodelkannya dengan mencoba beberapa metode yang sudah kami pelajari dan melihat metode mana yang paling baik untuk menampilkan data ini pada Python.

Metode yang akan digunakan dalam penelitian ini adalah metode regresi yaitu Regresi Bivariat dimana variabel y (Chiasm & Optic Nerve ) dengan variabel x1 (Brain) dan x2 (Brainstem) dan Regresi Multivariat

Berikut ini adalah bentuk data yang kami dapat dari situs <https://www.cancerdata.org/resource/doi:10.17195/candat.2018.01.1>

**Sumber data**

Menurut litaratur telah dilakukan penelitian batasan dosis radiologi untuk organ tubuh menurut konsensus dari jejaring terapi partikel dalam neuro\_oncology di Eropa. Pada literatur tersebut ditunjukan terdapat 8 jenis variabel dimana batasan dosis yang diijinkan pada organ tubuh mengikuti formula EQD2 yaitu dengan D adalah dosis total dan d = rasio dosis antara yang diukur dengan total dosis (bernilai 1atau 2 atau 3). Hasil data pengukuran dengan formula EQD2 ditunjukan dengan tabel berikut. Pada literatur batas atas dosis untuk tiap variabel menurut EQD2

Data yang akan dianalisis adalah data yang diperoleh dari situs internet sehingga sebelum menganalisis atau mengolah data tersebut perlu kita panggil data tersebut tanpaharus mengunduh terlebih dahulu.Namun data tersebut disediakan dalam sebuahsitus berupah gambar sehingga kami tidak dapat langsung memanggil data tersebut di Python dan kemudian mengolahnya. Data tersebut tersedia untuk di unduh dalam beberapa bentuk salah satunya dalam bentuk Excel sehingga diputuskan untuk mengunduh data tersebut dalam bentuk Excel dan kemudian memanggil data terbut dengan Phyton. Berikut ini adalah tampilan data cancer.

|  |
| --- |
| https://www.cancerdata.org/sites/default/files/elfinder/pictures/2018-02-28-EPTN-Neuro-OAR-tolerance-table.png |
| **Gambar 1.** Tampilan data cancer asli |
|  |

Setelah mengunduh data cancer dapat dilihat pada gambar 2 , langkah yang kami lakukan adalah menyusun data tesebut dalam susunan yang bisa membuat data tersebut dapat mudah dipanggil di Pytoh dan memiiki tampilan yang mudah dipahami.Gambar berikut adalah tampilan data pada Excel

|  |
| --- |
|  |
| **Gambar.2** Tampilan data dalam excel |

Data tersebut kemudian diubah menjadi kumpulan data biasa tanpa mengurangi esensi dari data tersebut. Format dari data yang telah diubah dapat dilihat pada gambar 3 berikut ini.

|  |
| --- |
|  |
| **Gambar 3.** Tampilan sampel data cancer yang telah diubah formatnya |

1. **Hasil dan Pembahasan**

Data diolah dengan beberapa kasus berikut ini

1. Regresi Bivariat dimana variabel y (Chiasm & Optic Nerve ) dengan variabel x1 (Brain) dan x2 (Brainstem)
2. Regresi Multivariat

Berikut ini adalah pembahasan dari kedua kasus diatas,

**Kasus 1 Hasil Regresi Bivariat**

Data tersebut diinisialisasi sebagai sampai sesuai ururtanya seperti yang terdapat pada tabel asli data tersebut dimana adalah Brain, adalah Brainstem yang pertama atau Brainstem 1, adalah Brainstem yang kedua atau Brainsteam 2, adalah Chiasm & Optic nerve, adalah Cornea, adalah Lens, adalah Retina, dan adalah Skin.

**Tahap 1. Memanggil data di Python**

Langkah berikut yang digunakan adalah memanggil data tersebut dengan Python

|  |
| --- |
| import pandas  import numpy  data = pandas.read\_excel('D:/CANCER.xlsx')  print(data)  data.head() |

maka keluarannya akan seperti pada gambar dibawah ini

|  |
| --- |
|  |
| **Gambar 4.** Sampel Output Data Cancer |

**Tahap 2. Regresi Linear dengan Python**

Gambar 4 menunjukkan bahwa data berhasil dibaca pada Python sehingga dapat kita coba dengan metode regresi linear dengan beberapa data untuk melihat keluaran dari data tersebut

|  |
| --- |
| #program asli dari: Thomas Haslwanter  import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  # Untuk menggambar 3 D  from mpl\_toolkits.mplot3d import Axes3D  # untuk keperluan statistik statsmodels 5.0  from statsmodels.formula.api import ols  # ANOVA pada model-model linear  from statsmodels.stats.anova import anova\_lm  data = pandas.read\_excel('D:/CANCER.xlsx')  fig = plt.figure()  ax = plt.axes(projection='3d')  X = data['x1'].values[:,np.newaxis]  Y = data['x2'].values[:,np.newaxis]  Z = data['x3'].values  #data = pandas.DataFrame({'x': X, 'y': Y, 'z': Z})  # mulai regresi linear  model = ols("Z ~ X + Y", data).fit()  # Tuliskan hasil uji statistik yang dilakukan  print(model.summary())  print("Nyatakan parameter estimasi")  print(model.\_results.params)  # Nyatakan analisis variansi dari model  anova\_results = anova\_lm(model)  print('Hasil anova')  print(anova\_results)  plt.show() |

Dengan program Python akan dilihat model regresi linear antara variabel z (Chiasm & Optic Nerve ) dengan variabel x (Brain) dan y (Brainstem)

|  |
| --- |
|  |
| **Gambar 5.** Output OLS dari data Cancer |

**Tahap 3. Analisis Program**

Telah dilakukan model regresi linear antara variabel z (Chiasm & Optic Nerve ) dengan variabel x (Brain) dan y (Brainstem). Pada hasil menunjukkan bilangan kondisional terlalu besar (dalam order 104). Selain itu nilai parameter yang dihasilkan berturut-turut 0,1,0 (nilai 0 ditulisa sebagai bilangan yang sangat kecil dalam orde 10-13. Jadi hasil model linear yang sudah dipilih tidak tepat. Menurut tabel OLS hasil regresi, uji t pada ketiga variabel mempunyai *p-value*> 0.05 oleh karena itu kedua variabel tidak bersifat linear terhadap z (Chiasm & Optic Nerve ). Untuk selanjutnya perlu dilakukan pemodelan yang lebih teliti. Model yang sudah ada hanya untuk menunjukkan contoh cara melakukan regresi multivariat tanpa memperhatikan makna fisis dari kasus yang sesungguhnya.

**Menentukan hubungan antar variabel**

Pada penelitian ini dilakukan beberapa tahapan penelitian diantaranya menyelidiki korelasi antar variabel. secara natural diasumsikan bahwa masing-masing organ tubuh saling terkait. Untuk itu kita menggunakan fungsi “corr” dari python dan diharapkan setiap angka pada matriks korelasi yang diperoleh jauh lebih besar dari 0.5 atau mendekati 1dan berkorelasi posotif(artinya peningkatan salah satu variabel). Hal ini ditunjukan pada hasil berikut

|  |
| --- |
| data.corr() |

Tabel. Hasil matriks korelasi antar variabel

|  |
| --- |
|  |
| **Tabel 1.** Korelasi antar Variabel |

**Tahap 4. Regresi Multivariat dengan Python**

Setelah melaukan analisis dengan regresi linear biasa antar dua variabel dapat dilihat bahwa menurut tabel OLS hasil regresi, uji t pada ketiga variabel mempunyai *p-value*> 0.05 oleh karena itu kedua variabel tidak bersifat linear terhadap z (Chiasm & Optic Nerve ). Sehingga akan dilakukan regresi multivariat terhadap data cancer dengan harapat mendapatkan hasil analisis yang lebih baik.

Berikut adalah algoritma yang digunakan dalam Python :

|  |
| --- |
| #Program 3.1  #program asli dari: Thomas Haslwanter  import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  # Untuk menggambar 3 D  from mpl\_toolkits.mplot3d import Axes3D  # untuk keperluan statistik statsmodels 5.0  from statsmodels.formula.api import ols  # ANOVA pada model-model linear  from statsmodels.stats.anova import anova\_lm  data = pandas.read\_excel('D:/CANCER.xlsx')  fig = plt.figure()  ax = plt.axes(projection='3d')  x1 = data['x1'].values[:,np.newaxis]  x3 = data['x3'].values[:,np.newaxis]  x4 = data['x4'].values[:,np.newaxis]  x5 = data['x5'].values[:,np.newaxis]  x6 = data['x6'].values[:,np.newaxis]  x7 = data['x7'].values[:,np.newaxis]  x8 = data['x8'].values[:,np.newaxis]  Z = data['x2'].values  #data = pandas.DataFrame({'x': X, 'y': Y, 'z': Z})  # mulai regresi linear  model = ols("Z ~ x1 + x3 + x4 + x5 + x6 + x7 + x8", data).fit()  # Tuliskan hasil uji statistik yang dilakukan  print(model.summary())  print("Nyatakan parameter estimasi")  print(model.\_results.params)  # Nyatakan analisis variansi dari model  anova\_results = anova\_lm(model)  print('Hasil anova')  print(anova\_results)  plt.show() |

|  |
| --- |
|  |
| **Gambar 6.**Output OLS dari data Cancer Regresi Multivariat |

**Analisis Program :**

Telah dilakukan model regresi multivariat antara variabel yang ada dalam data. Hasil menunjukkan Pada hasil menunjukkan nilai eigen terkecil yaitu 5.15e-30. Ini mungkin menunjukkan bahwa ada masalah multikolinearitas yang kuat atau bahwa matriks memiliki desain tunggal. Menurut tabel OLS hasil regresi, uji t pada ketiga variabel mempunyai *p-value*> 0.05 oleh karena itu setiap variabel tidak bersifat linear terhadap z.

1. **Kesimpulan**

Pada penelitan ini dilakukan studi tentang data kanker yang diolah dengan metode regresi dengna bantuan program python. Data diperoleh dari internet kemudian diolah dengan hasil (sebagai berikut)

Daftar Pustaka

Lambrecht,M., Daniëlle B.P. Eekers., D.B.P., Alapetite,C, Neil G. Burnet, Valentin Calugaru,Ida E.M. Coremans, PieroFossati, Morten Høyer, Johannes A. Langendijk, Alejandra Méndez Romero,Frank Paulsen, Ana Perpar, Laurette Renard, Dirk de Ruysscher, BeateTimmermann, Pavel Vitek, Damien C. Weber, Hiske L. van der Weide, Gillian A. Whitfield, Ruud Wiggenraad, Erik Roelofs,Petra Witt Nyström, Esther G.C. Troost, Radiation Dose Constraints for Organs At Risk in Neuro-Oncology; The European Particle Therapy Network Consensus., ‘‘European Particle Therapy Network” of ESTRO.,*Journal of RadiotherapyandOncology.*

