MODUL

PENELITIAN
UJI DIAGNOSTIK DAN SKRINING

dr. I Wayan Gede Artawan Eka Putra, M.Epid.
dr. I Made Sutarga, M.Kes.
Made Pasek Kardiwinata, S.KM., M.Kes
Ni Luh Putu Suariyani, S.KM., M.IntHlth&Dev
dr. Ni Wayan Septarini, MPH
Dr. Drh I Made Subrata, M.Erg.

Program Studi Kesehatan Masyarakat
Fakultas Kedokteran Universitas Udayana
Denpasar, 2016

DAFTAR ISI

Bagian 1 Konsep Penelitian Uji Diagnostik dan Skrining 1
1.1 Pendahuluan 1
1.2 Baku emas 3
1.3 Independensi 4
1.4 Karakteristik Test Diagnostik dan Skrining Yang Baik 4
1.5 Karakteristik Penyakit Yang Cocok Diskrining 5
Bagian 2 Menghitung Sensitifitas, Spesifisitas, LR+, LR -, NPP, NPN dan Membandingkan 2 Test Berskala Data Nominal 7
2.1 Pendahuluan 7
2.2 Tujuan Pembelajaran 7
2.3 Pokok Bahasan 7
2.4 Cara menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN 8
2.5 Aplikasi Stata 10
2.6 Interpretasi Hasil Uji Diagnose 14
2.7 Interpretasi hasil analisis perbedaan kemampuan dua test 15
Bagian 4 Analisis *Reciever operating characteristic* (ROC), penentuan titik potong dan membandingkan kemampuan 2 test atau lebih berskala kontinyu 16
4.1 Pendahuluan 16
4.2 Tujuan Pembelajaran 16
4.3 Pokok Bahasan 16
4.4 Definisi dan ukuran yang dinilai dalam analisis ROC 17
4.5 Penentuan titik potong 17
4.6 Aplikasi Stata 19
4.7 Interpretasi Hasil 22
Bagian 5 Uji Skrining Kombinasi (Multiple) 24
5.1 Pendahuluan 24
5.2 Tujuan Pembelajaran 24
5.3 Pokok Bahasan 24
5.4 Konsep Uji Skrining Kombinasi (Multiple) 25
5.5 Program Skrining 26
5.6 Pertimbangan Penggunaan Suatu Test pada Program Skrining 27

5.7 Aplikasi Stata Untuk Uji Skrining Kombinasi (Multiple) 27
5.8 Interpretasi Hasil 33
Bagian 6 Menilai Reliabilitas Suatu Test 34
6.1 Pendahuluan 34
6.2 Tujuan Pembelajaran 34
6.3 Pokok Bahasan 35
6.4 Ukuran menilai reliabilitas test dengan hasil ukur berskala data kategorikal 35
6.5 Ukuran untuk menilai reliabilitas test dengan hasil ukur berskala data kontinyu 36
6.6 Aplikasi Stata menghitung ukuran reliabilitas test berskala data kategorikal 37
6.7 Aplikasi Stata untuk menghitung ukuran reliabilitas test berskala data kontinyu 38
6.8 Interpretasi hasil analisis penilaian reliabilitas 42
Ringkasan Perintah (Command) Stata Untuk Analisis Data Penelitian Uji Diagnostik dan Skrining 44
Daftar Pustaka 45

Bagian 1 Konsep Penelitian Uji Diagnostik dan Skrining

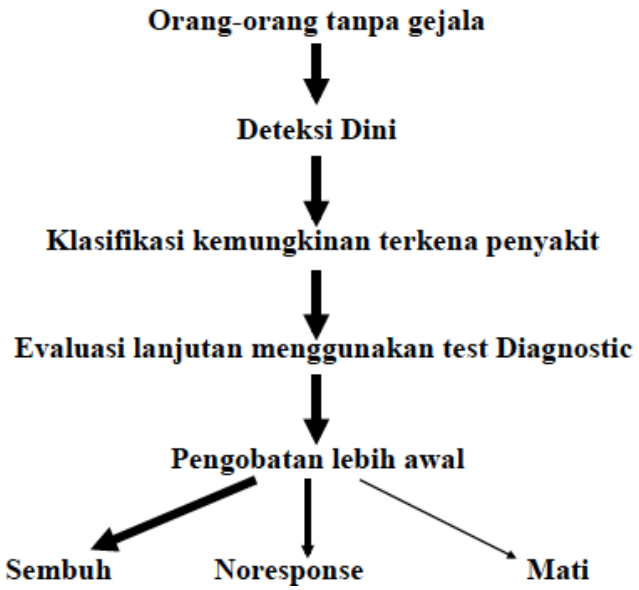
1.1 Pendahuluan

Penelitian uji diagnostik dan skrining sangat penting di dunia kedokteran dan kesehatan masyarakat. Istilah uji diagnostik dan uji skrining pada dasarnya memiliki tujuan yang sama. Uji diagnostik digunakan didunia kedokteran, dalam melakukan prosedur diagnostik sering digunakan beberapa test secara bertahap untuk meningkatkan probabilitas individu mengalami suatu penyakit tertentu. Misalnya dalam prosedur diagnostik kanker payudara pada wanita berumur lebih dari 40 tahun, dimana terlebih dahulu dilakukan mamografi dan bila hasil mamografi positif maka probabilitas wanita tersebut untuk mengalami kanker payudara menjadi lebih tinggi dibandingkan bila hasil test negatif. Bila mamografi positif, maka selanjutnya akan dilakukan pemeriksaan biopsi menggunakan *fine needle aspiration biopsy* (FNAB). Bila hasil FNAB positif maka probabilitas wanita tersebut mengalami kanker payudara menjadi lebih besar. Hasil test tersebut selanjutnya oleh klinisi dijadikan dasar untuk melakukan tindakan pembedahan dan pemeriksaan masa tumor untuk pemeriksaan patologi anatomi yang selanjutnya dipakai dasar untuk menentukan kanker atau tidak. Sebaliknya, penelitian uji skrining dipakai untuk mendapatkan suatu test yang mampu menapis atau mendeteksi secara dini kemungkinan seseorang tanpa keluhan menderita suatu penyakit tertentu pada sekelompok populasi.

Penelitian uji diagnostik dan skrining mempunyai tujuan yang sama yaitu untuk menilai validitas dan reliabilitas suatu test dalam mendeteksi kemungkinan adanya suatu penyakit secara lebih dini (deteksi dini).Validitas meliputi sensitifitas dan spesifisitas. Sensitifitas adalah kemampuan suatu test untuk menyatakan positif pada orang-orang yang sakit, sedangkan spesifisitas adalah kemampuan suatu test untuk menyatakan negatif orang-orang yang tidak sakit. Reliabilitas meliputi penilaian terhadap *bias interobserver* dan *bias intraobserver*. Bias interobserver adalah perbedaan hasil ukur antar observer, sebaliknya bias intra observer adalah perbedaan hasil ukur oleh observer yang sama pada waktu yang berbeda pada subjek yang sama. Uji reliabilitas dalam modul ini disampaikan pada bagian akhir, walaupun demikian sebaiknya dilakukan pada awal studi sebelum uji diagnosis dilaksanakan. Selain validitas dan reliabilitas juga akan dipelajari tentang efikasi suatu test jika nanti diterapkan pada suatu populasi. Efikasi meliputi nilai prediktif test positif (NPP) dan nilai prediktif test negatif (NPN). NPP

mencerminkan probabilitas sakit jika hasil test dinyatakan positif, sedangkan NPN mencerminkan probabilitas tidak sakit jika hasil testnya dinyatakan negatif.

Deteksi dini merupakan bagian dari pencegahan sekunder yang terdiri dari deteksi dini dan dikuti pengobatan tepat. Konsep dari deteksi dini adalah mendeteksi kemungkinan mengalami suatu penyakit pada orang-orang tanpa gejala. Dengan melakukan deteksi dini maka klasifikasi memungkinkan terkena suatu penyakit pada seseorang menjadi lebih awal diketahui. Pada tahap selanjutnya, orang dengan hasil deteksi dini positif dievaluasi menggunakan prosedur diagnosis untuk memastikan apakah benar mengalami penyakit sehingga terapi menjadi lebih cepat diberikan dan outcome penyakit sebagian besar berakhir dengan kesembuhan (prognosis lebih baik). Konsep ini sejalan dengan teori bahwa semakin dini suatu penyakit diketahui maka peluang sembuh sempurna semakin besar dan biaya yang diperlukan untuk penatalaksanaannya menjadi semakin murah. Konsep deteksi dini tersebut dapat dibuatkan bagan sebagai berikut:



Gambar 1. Konsep Skrining

1.1 Rancangan Penelitian Uji Diagnostik Dan Skrining

Penelitian uji diagnostik dan skrining tergolong penelitian observasional deskriptif dengan rancangan yang umum digunakan adalah *cross-sectional study*. Tergolong dalam observational deskriptif karena pada penelitian ini hanya dilakukan observasi tanpa adanya intervensi (perlakuan). Selain itu penelitian uji diagnostik dan skrining dilakukan pada satu populasi dengan tujuan menggambarkan validitas suatu

test terhadap suatu baku emas, tanpa adanya tujuan untuk mengetahui asosiasi dan hubungan sebab akibat. Rancangan cross-sectional study berarti bahwa semua variabel, termasuk test yang diuji dan *gold standar* (baku emas) diukur pada satu periode waktu yang sama. Waktu pengambian bahan (speciment) atau pengukuran penyakit yang diteliti dengan test yang diuji dan baku emas sangat penting dilakukan pada periode waktu yang sama untuk menjamin bahwa kondisi penyakit masih sama. Jika waktu pengukuran atau mengambilan bahan oleh test yang diuji dan gold standar berbeda maka adanya perbedaan hasil bukan karena kurang validnya test yang diuji melainkan karena perbedaan waktu pengukuran. Dengan kata lain, terjadi kesalahan hasil perhitungan sensitifitas, spesifisitas, NPP dan NPN karena perbedaan periode waktu pengukuran antara test yang diuji dengan baku emas. Selaian rancangan cross-sectional, case-control juga dapat dipakai sepanjang data yang diperlukan terdapat pada catatan medis pasien. Mislnya test yang diuji adalah test untuk apendisitis sedangkan baku emasnya adalah hasil PA dari bahan yang diambil pada saat apendektomi. Pada kasus ini kasusnya adalah PA positif apendik dan kontrol adalah pasien dengan PA negatif apendik. Data hasil test diambil dari catatan medik pasien bersangkutan. Selanjutnya hasil PA dan hasil test disilangkan untuk menguji validitas test. Oleh karena itu penggunaan rancangan penelitian *kohort* dan *case control* kurang tepat.

1.2 Baku emas

Isu penting lain dalam merancang penelitian uji diagnostik dan skrining adalah masalah penentuan atau pemilihan test yang digunakan sebagai baku emas. Tidak sembarang test bisa digunakan sebagai baku emas. Test yang berdasarkan beberapa literatur menunjukkan validitas lebih tinggi tetapi masih tergolong test untuk skrining tidak bisa digunakan sebagai baku emas. Baku emas adalah test yang memang sudah dipercaya kebenarannya sesuai dengan patofisiologi penyakit. Salah satu contoh dalam mendiagnosis tuberkulosis (TB) pada anak, kurang tepat kalau dilakukan penelitian uji diagnosis sistem skoring TB anak oleh dokter umum dengan dokter spesialis anak sebagai baku emas karena hasil skoring TB anak oleh dokter spesialis anak belum tentu benar. Dalam patofisiologi TB pada anak perbaikan klinis setalah pemberian terapi obat anti TB dapat digunakan sebagai baku emas. Dengan kata lain pemilihan baku emas suatu penyakit tertentu yang akan dilakukan penelitian uji diagnosis dan skrining harus sesuai dengan keilmuwan (teori) yang berlaku. Seperti baku emas untuk kanker payudara adalah hasil pemeriksaan patologi anatomi (PA).

1.3 Independensi

Selain waktu pengukuran dan pemilihan baku emas, masalah yang juga penting diperhatikan adalah orang yang mengukur hasil baku emas dan test yang diuji haruslah orang yang berbeda dan tidak boleh saling mengetahui hasil masing-masing (*blindness*). Selain itu pasien sebagai subjek penelitian untuk sementara waktu sampai semua pemeriksaan selesai dilakukan sebaiknya tidak diberitahukan hasil pemeriksaannya. Hal ini penting untuk menjamin independensi hasil pemeriksaan baik pada test yang diuji maupun pada baku emas. Bila antar pemeriksaan dilakukan oleh orang yang sama atau antar pemeriksa (pemeriksa pada test yang diuji dan baku emas) saling mengetahui hasil pada pemeriksaan lainnya maka akan ada kecendrungan untuk menyamakan hasil. Dengan kata lain akan berakibat terhadap hasil perhitungan validitas akan lebih tinggi dari nilai yang sebenarnya atau *over estimasi*.

1.4 Karakteristik Test Diagnostik dan Skrining Yang Baik

Test diagnostik dan skrining harus memenuhi karakteristik tertentu supaya bisa disebut sebagai test yang baik. Karakteristik atau ciri-ciri ini juga merupakan alasan atau latar belakang untuk melaksanakan suatu penelitian uji diagnostik dan skrining. Dalam membuat suatu latar belakang dilakukan penelitian uji diganostik dan skrining sangat penting menyampaikan kelebihan dari test yang dinilai validitas, reliabilitas dan efikasinya. Salah satu contoh adalah ketika inspeksi visual asam asetat (IVA) pertama kali diperkenalkan dan ingin dinilai dan dibandingkan validitasnya terhadap pap smear dengan hasil patologi anatomi sebagai baku emas. Maka penting disampaikan pada latar belakang beberapa keunggulan IVA seperti lebih murah, cepat dan mudah dikerjakan. Beberapa syarat test disebut baik adalah:

1. Ekonomis

   Test yang ekonomis berarti biaya yang diperlukan untuk melakukan suatu test diagnostik lebih murah dari yang biasa digunakan.

2. Cepat

   Test yang cepat berarti waktu yang diperlukan dari sampel diambil sampai didapatkan hasil lebih cepat dari yang biasa digunakan.

3. Mudah dikerjakan

   Mudah dikerjakan mempunyai makna suatu test tidak memerlukan suatu spesialisasi atau keahlian khusus untuk mengerjakan. Suatu test dinyatakan

mudah jika bisa dikerjakan oleh tenaga medi dengan pelatihan yang relatif singkat.

4. Bebas dari risiko dan ketidaknyamanan

   Bebas dari risiko dan ketidaknyamanan mempunyai makna suatu test tidak memerlukan tindakan invasif dan proses penerapannya sedikit atau minimal menimbulkan ketidaknyamanan pada pasien. Salah satu pengertian ketidaknyamanan adalah timbulnya rasa sakit (nyeri) saat test dikerjakan. Semakin tidak ada rasa nyeri dan tidak ada tindakan invasif maka semakin baik suatu test skrining.

5. Dapat diterima di masyarakat

   Dapat diterima dimasyarakat mempunyai arti test skrining tidak bertentangan dengan norma dan etika yang berlaku di masyarakat.

6. Valid

   Validitas suatu test menunjukkan kemampuan (ketepatan) suatu test untuk mendapatkan nilai yang sesuai dengan kenyataan yang sebenarnya. Dalam dunia kesehatan kenyataan yang sebenarnya terbagi menjadi 2, sakit dan tidak sakit. Sehingga penilaian terhadap validitas meliputi 2 aspek yaitu ketepatan menilai kondisi sakit disebut sensitifitas dan ketepatan untuk menilai kondisi tidak sakit disebut spesifisitas. Dengan kata lain sensitifitas adalah kemampuan suatu test untuk menyatakan positif orang-orang yang sakit, sedangkan spesifisitas adalah kemampuan suatu test untuk menyatakan negatif orang-orang yang tidak sakit.

7. Reliabel

   Reliabilitas suatu test menunjukkan konsistensi (kesamaan) hasil ukur bila dikerjakan lebih dari sekali terhadap pasien (subjek pengamatan) yang sama pada kondisi yang sama pula. Sumber ketidaksamaan hasil ukur (bias) dapat terjadi karena observer berbeda yang disebut dengan bias inter observer dan observer sama tapi waktu pengamatan berbeda yang disebut bias intra obserber.

1.5 Karakteristik Penyakit Yang Cocok Diskrining

Tidak semua penyakit cocok untuk didilakukan skrining, ada beberapa karkateristik penyakit yang sebaiknya dipenuhi supaya bisa dilakukan skrining. Karakteristik penyakit yang cocok diskrining meliputi:

1. Penyakit serius dan dengan konsekuensi berat

Yang dimaksud dengan penyakit serius dan dengan konsekuensi berat adalah penyakit yang mendapat perhatian di masyarakat, peningkatan kejadiannya menimbulkan masalah keresahan masyarakat. Penyakit seperti ini biasanya memiliki konsekuensi (akibat) menimbulkan kecacatan dan kematian. Beberapa contoh penyakit dengan karakteristik seperti ini adalah kanker, diabetes melitus, tekanan darah tinggi dan HIV/Aids

2. Pengobatan lebih efektif pada tahap awal

   Pengobatan lebih efektif pada tahap awal berarti probabilitas untuk sembuh, tidak mengalami kecacatan dan kematian jauh lebih tinggi jika terdeteksi dan diobati pada tahap awal (preklinis) dibandingkan tahap lanjut. Salah satu contoh penyakit lebih efektif diobati pada tahap awal adalah kanker payudara dimana bila terdeteksi dan diobati pada tahap awal akan menurunkan mortalitasnya sehingga sangat cocok dilakukan skrining. Berbeda halnya dengan kanker pankreas, walaupun ditemukan dan diobati pada tahap awal tetapi *survivor rate* dari penyakit ini tidak berbeda dengan yang ditemukan dan diobati pada tahap lanjut.

3. Penyakit dengan *detectable preclinical phase* yang lama

   Penyakit dengan *detectable preclinical phase* yang lama berarti memiliki waktu periode subklinis/preklinis (belum timbul tanda dan gejala) yang panjang, tetapi sudah terjadi perubahan di dalam tubuh (patologi anatomi) yang memungkinkan penyakit tersebut terdeteksi. Semakin lama *detectable preclinical phase* maka semakin banyak kesempatan untuk menemukan pada stadium yang lebih dini. Dengan kata lain semakin lama *detectable preclinical phase*s uatu penyakit, maka semakin baik untuk diskrining.

4. Prevalensi penyakit tinggi pada populasi

   Persyaratan ini sangat penting untuk mejamin efektifitas hasil skrining. Semakin tinggi prevalensi suatu penyakit di populasi maka kemungkinan untuk benar sakit pada orang-orang dengan hasil tes positif semakin tinggi. Dengan kata lain, jika skrining dilakukan pada populasi dengan prevalensi suatu penyakit tinggi, maka jumlah hasil positif palsu akan semakin sedikit. Hal ini akan sangat bagus karena akan menghasilkan nilai prediktif test positif (NPP) yang lebih tinggi.

Bagian 2 Menghitung Sensitifitas, Spesifisitas, LR+, LR -, NPP, NPN dan Membandingkan 2 Test Berskala Data Nominal

2.1 Pendahuluan

Pada bagian ini akan dipelajari tentang ukuran-ukuran yang digunakan untuk menilai kemampuan suatu test untuk mendeteksi adanya suatu penyakit. Ukuran yang digunakan adalah validitas yang terdiri dari sensitifitas, spesifisitas dan akurasi. Selain itu dipelajari juga tentang *likelihod ratio positive* (LR+) dan *likelihod ratio negative* (LR-). Efikasi suatu test jika diterapkan pada populasi dan apa saja yang mempengaruhi efikasi. Efikasi suatu test terdiri dari nilai prediktif test positif (NPP) dan nilai prediktif test negatif (NPN). Pada bagian ini juga akan dipelajari cara membandingkan kemampuan 2 test berskala data nominal 2 kategori dalam mempediksi suatu penyakit dengan menggunakan baku emas.

2.2 Tujuan Pembelajaran

1. Untuk meningkatkan pengetahuan dan pemahaman isu-isu penting dalam menilai kemampuan suatu test untuk mendeteksi adanya penyakit menggunakan sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN.
2. Mampu mengaplikasikan Stata untuk menghitung dan menginterprtasikan semua ukuran tersebut.
3. Mampu mengaplikasikan Stata untuk membandingkan kemampuan 2 test berskala data nominal 2 kategori dalam memprediksi suatu penyakit dengan menggunakan baku emas
4. Mampu membaca dan menginterpretasikan hasil analisis.

2.3 Pokok Bahasan

1. Cara menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN
2. Aplikasi Stata untuk menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN
3. Aplikasi Stata untuk membandingkan kemampuan 2 test berskala data nominal 2 kategori dalam mempediksi suatu penyakit dengan menggunakan baku emas
4. Cara membaca dan interpretasi sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN dan menyimpulkan perbandingan kemampuan 2 test berskala data

nominal 2 kategori dalam mempediksi suatu penyakit dengan menggunakan baku emas

2.4 Cara menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN

Dalam penelitian uji diagnostik dan skrining dimana test uji berskala data nominal 2 kategori maka hasil pengamatan disajikan dalam bentuk tabulasi silang 2 sebagai berikut:

Baku Emas

		Sakit	Tidak sakit	
Test yang dinilai	+	TP	FP	TP+FP
	-	FN	TN	FP+TN
		TP+FN	FP+TN	Total

Dalam menampilkan tabulasi silang, lazimnya hasil test yang dinilai berada pada baris (row) sedangkan baku emas berada di kolom (colomn). Berdasarkan tabulasi silang tersebut maka dapat didefinisikan terlebih dahulu arti dari tiap sel sebagai berikut:

1. TP = jumlah yang dinyatakan positif oleh test dan baku emas menyatakan sakit.
2. FP = jumlah yang dinyatakan positif oleh test tetapi baku emas menyatakan tidak sakit.
3. FN = jumlah yang dinyatakan negatif oleh test tetapi baku emas meyatakan sakit.
4. TN = jumlah yang dinyatakan negatif oleh test dan baku emas juga menyatakan tidak sakit.
5. TP+FN adalah keseluruhan jumlah orang yang sakit
6. FP+TN adalah keseluruhan jumlah yang tidak sakit
7. TP+FP adalah keseluruhan jumlah yang hasil testnya positif
8. FN+TN adalah keseluruhan jumlah yang hasil testnya negatif
9. Total adalah jumlah total sampel yang diteliti

Rumus untuk menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN adalah:

1. Sensitifitas, adalah proporsi hasil test positif diantara orang-orang yang sakit atau dapat diterjemahkan dengan rumus sebagai berikut

$$Sensitifitas = \frac{\text{TP}}{TP + FN}\ x100\%$$

Sensitifitas menunjukkan kemampuan suatu test untuk menyatakan positif orang-orang yang sakit. Semakin tinggi sensitifitas suatu test maka semakin banyak mendapatkan hasil test positif pada orang-orang yang sakit atau semakin sedikit jumlah negatif palsu.

2. Spesifisitas, adalah proporsi hasil test negatif diantara orang orang yang tidak sakit atau dapat diterjemahkan dengan rumus sebagai berikut:

$$Spesifisitas = \frac{\text{TN}}{FP + TN}\ x100\%$$

Spesifisitas menunjukkan kemampuan suatu test untuk menyatakan negatif orang-orang yang tidak sakit. Semakin tinggi spesifisitas suatu test maka semakin banyak mendapatkan hasil test negatif pada orang-orang yang tidak sakit atau semakin sedikit jumlah positif palsu.

3. Akurasi, adalah proporsi hasil test benar (true value) diantara semua yang diperiksa atau dapat diterjemahkan dengan rumus sebagai berikut:

$$Akurasi = \frac{\text{TP} + \text{TN}}{Total}\ x100\%$$

4. NPP, adalah proporsi orang yang sakit diantara hasil test positif atau dapat diterjemahkan dengan rumus sebagai berikut:

$$NPP = \frac{\text{TP}}{TP + FP}\ x100\%$$

NPP menunjukkan besarnya kemungkinan mengalami sakit pada orang-orang dengan hasil test positif. Besarnya NPP tidak hanya dipengaruhi oleh sensitifitas dan spesifisitas tetapi juga prevalensi penyakit pada populasi studi. Semakin tinggi prevalensi penyakit di populasi maka NPP akan semakin besar. Hal inilah yang mengisyaratkan pentingnya melakukan skrining pada populasi tertentu dengan prevalensi penyakit yang lebih tinggi.

5. NPN, adalah proporsi orang yang tidak sakit diantara hasil test negatif atau dapat diterjemahkan dengan rumus sebagai berikut:

$$NPN = \frac{\text{TN}}{FN + TN}\ x100\%$$

NPN menunjukkan besarnya kemungkinan **tidak** mengalami sakit pada orang-orang dengan hasil test negatif. Besarnya NPP juga dipengaruhi prevalensi penyakit pada populasi studi selain sensitifitas dan spesifisitas. Hanya saja arah

hubungannya berbeda, dimana semakin tinggi prevalensi penyakit di populasi maka NPN akan semakin kecil.

6. LR+, adalah perbandingan antara true positif (sensitivitas) dengan false positif (1 - spesifisitas) dan dinyatakan dengan rumus sebagai berikut:

$$LR+ = \frac{\text{Sensitifitas}}{(1 - Spesifisitas)}$$

LR+ digunakan untuk menilai kemampuan suatu test untuk mendeteksi adanya penyakit, semakin tinggi nilai LR+ suatu test maka semakin baik kemampuan test tersebut. Penilaian kemampuan test menggunakan LR+ lebih disukai karena lebih stabil dan nilai tertingginyanya tidak dibatasi oleh proporsi 100% seperti sensitifitas dan spesifisitas.

7. LR-, adalah perbandingan antara false negatif (1 – sensitifitas) dengan true negatif (spesifisitas) dan dinyatakan dengan rumus sebagai berikut:

$$LR- = \frac{(1 - \text{Sensitifitas})}{Spesifisitas}$$

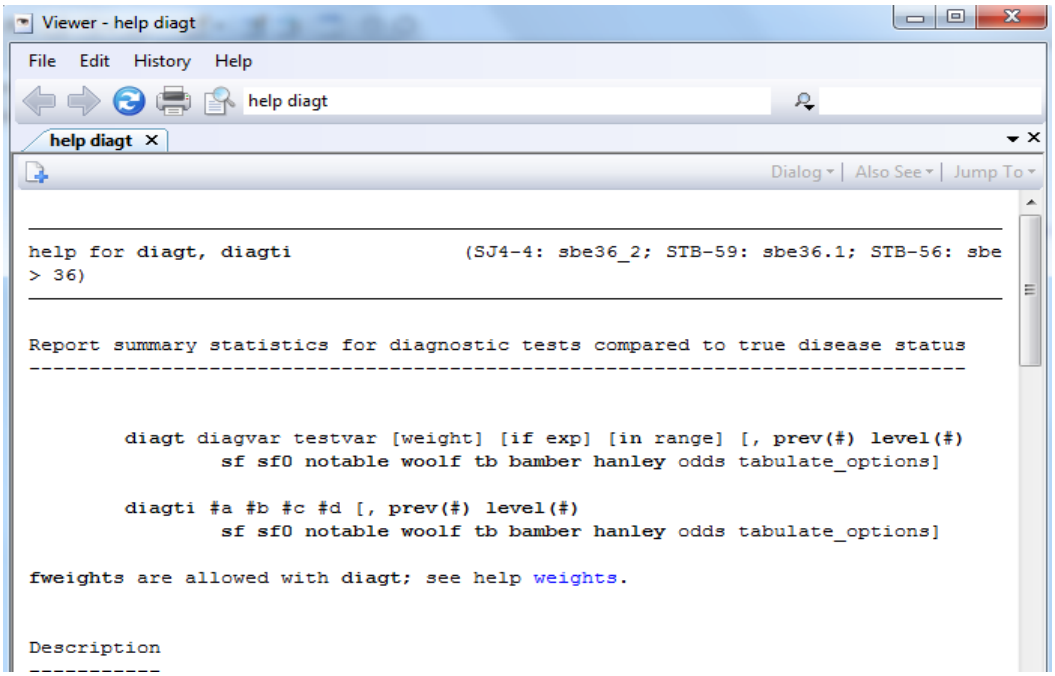
Seperti halnya LR+, LR- juga digunakan untuk menilai kemampuan suatu test untuk mendeteksi adanya penyakit tetapi dengan perbedaan cara interpretasi dimana semakin rendah nilai LR- suatu test maka semakin baik kemampuan test tersebut. Penilaian kemampuan test menggunakan LR- juga lebih stabil dan nilai terendahnya tidak terbatas.

2.5 Aplikasi Stata

Aplikasi Stata untuk menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN, maka terlebih dahulu dilihat apakah dalam fitur Stata sudah tersedia untuk command diagnositik test. Untuk mengetahui apakah fitur untuk diagnostik test sudah tersedia, maka ketik pada command window perintah sebagai berikut:

help diagt

Kemudian tekan enter, bila muncul box dialog seperti dibawah ini:



Maka itu artinya fitur Stata untuk diagnostik test sudah tersedia. Jika belum, lakukan download terlebih dahulu fitur tersebut dengan cara klik download sbe36_2 dan pastikan sebelum melakukan download komputer telah tesambung ke internet.

Untuk memulai praktikum bukalah data Skrining1.dta melui File kemudian klik Open. Maka akan terbuka data stata tentang suatu penelitian uji skrining membandingkan kemampuan test inspeksi visual asam asetat (IVA) dan test Pap Smear dengan hasil PA sebagai baku emas untuk mendiagnosis kanker serviks pada populasi berisiko. Data ini adalah kasus hipotetis dan sudah mengalami perubahan sedemikian rupa untuk kepentingan pelatihan sehingga tidak mencerminkan kondisi sebenarnya.

Untuk mencapai tujuan pembelajaran 2 dan 3 maka ikutilah langkah-langkah berikut ini:

Langkah 1.

Menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN test IVA dengan hasil PA sebagai baku emas, dengan cara mengetik pada window command sebagai berikut:

diagt hasilpa iva

Kemudian tekan enter, maka akan muncul hasil sebagai berikut:

```
. diagt hasilpa iva
```

Hasil PA	iva Pos.	Neg.	Total
Abnormal	112	32	144
Normal	60	164	224
Total	172	196	368

```
True abnormal diagnosis defined as hasilpa = 1 (labelled Positif)
```

			[95% Confidence Interval]	
Prevalence	Pr(A)	39%	34%	44.3%
Sensitivity	Pr(+|A)	77.8%	70.1%	84.3%
Specificity	Pr(-|N)	73.2%	66.9%	78.9%
ROC area	(Sens. + Spec.)/2	.755	.71	.8
Likelihood ratio (+)	Pr(+|A)/Pr(+|N)	2.9	2.3	3.67
Likelihood ratio (-)	Pr(-|A)/Pr(-|N)	.304	.221	.416
Odds ratio	LR(+)/LR(-)	9.57	5.86	15.6
Positive predictive value	Pr(A|+)	65.1%	57.5%	72.2%
Negative predictive value	Pr(N|-)	83.7%	77.7%	88.6%

Hasil analisis di atas dapat dibagi menjadi 2 bagian, bagian pertama adalah tabuasi silang 2x2 dan bagian kedua adalah hasil perhitungan sensitifitas, spesifisitas, ROC area, LR+, LR-, NPP, NPN dan Odds ratio yang merupakan perbandingan antara LR+/LR-.

Langkah 2

Menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN test Pap Smear dengan hasil PA sebagai baku emas, dengan cara mengetik pada window command sebagai berikut:

diagt hasilpa papsmear

Kemudian tekan enter, maka akan muncul hasil sebagai berikut:

```
. diagt hasilpa papsmear
```

Hasil PA	papsmear Pos.	Neg.	Total
Abnormal	120	24	144
Normal	48	176	224
Total	168	200	368

```
True abnormal diagnosis defined as hasilpa = 1 (labelled Positif)
```

			[95% Confidence Interval]	
Prevalence	Pr(A)	39%	34%	44.3%
Sensitivity	Pr(+|A)	83.3%	76.2%	89%
Specificity	Pr(-|N)	78.6%	72.6%	83.8%
ROC area	(Sens. + Spec.)/2	.81	.769	.85
Likelihood ratio (+)	Pr(+|A)/Pr(+|N)	3.89	2.99	5.05
Likelihood ratio (-)	Pr(-|A)/Pr(-|N)	.212	.146	.308
Odds ratio	LR(+)/LR(-)	18.3	10.7	31.4
Positive predictive value	Pr(A|+)	71.4%	64%	78.1%
Negative predictive value	Pr(N|-)	88%	82.7%	92.2%

Langkah 3

Untuk membandingkan kemampuan kedua test apakah kemampuannya berbeda secara bermakna maka dilakukan *compare ROC area*, dengan mengetik pada windows command sebagai berikut:

roccomp hasilpa iva papsmear

Kemudian tekan enter, maka akan muncul hasil sebagai berikut:

```
. roccomp hasilpa iva papsmear
```

	Obs	ROC Area	Std. Err.	Asymptotic Normal [95% Conf.	Interval]
iva	368	0.7550	0.0228	0.71018	0.79974
papsmear	368	0.8095	0.0208	0.76881	0.85024

```
Ho: area(iva) = area(papsmear)
    chi2(1) =     3.18       Prob>chi2 =   0.0747
```

2.6 Interpretasi Hasil Uji Diagnose

Berdasarkan hasil analisis data di atas maka dapat dibuat beberapa interpretasi sebagai berikut:

1. Prevalensi Kanker Serviks di populasi study sebesar 39%. Hasil perhitungan prevalensi antara pada langkah 1 dan langkah 2 sama karena perhitungan prevalensi berdasarkan hasil baku emas.
2. Sensitifitas test IVA sebesar 77,8% yang artinya kemampuan test IVA untuk mendapatkan hasil positif diantara wanita yang mengalami kanker serviks sebesar 77,8% atau diantara 100 wanita yang kanker serviks 78 (dibulatkan) akan dinyatakan positif oleh test IVA sedangkan sisanya negatif (negatif palsu).
3. Spesifisitas test IVA sebesar 73,2% yang artinya kemampuan test IVA untuk mendapatkan hasil negatif diantara wanita yang **tidak** kanker serviks sebesar 73,2% atau diantara 100 wanita yang **tidak** kanker serviks 73 (dibulatkan) akan dinyatakan negatif oleh test IVA sedangkan sisanya positif (positif palsu).
4. NPP test IVA pada prevalensi 39% adalah 65,1% yang artinya kemungkinan benar mengalami kanker serviks jika hasil test IVA positif sebesar 65,1% atau setiap 100 wanita di populasi studi ini yang dinyatakan positif test iva maka akan didapatkan 65 orang benar mengalami kanker serviks.

   Dalam menuliskan NPP sebaiknya selalu diawali dengan menyebutkan prevalensi penyakit yang diteliti pada populasi studi dan pada interpretasi disebut populasi studinya. **Hal ini penting karena besarnya NPP tidak hanya dipengaruhi sensitifitas dan spesifisitas tetapi juga sangat dipengaruhi oleh prevalensi penyakit di populasi studi. Semakin tinggi prevalensi maka NPP akan semakin tinggi pula.** Untuk membuktikan hal tersebut cobalah dihitung NPP test IVA jika dilakukan pada populasi dengan prevalensi kanker serviks 10% (sensitifitas dan spesifisitas tetap sama sesuai hasil analisis ini)
5. LR+ test IVA = 2,9 artinya setiap 1 hasil *false positive* test IVA akan didapatkan 3 (dibulatkan) hasil*true positive* test IVA atau setiap 10 hasil *false positive* test IVA akan didapatkan 29 hasil *true positive* test IVA. Semakin tinggi nilai LR+ maka semakin baik kemampuan suatu test untuk mendeteksi suatu penyakit.
6. LR- test IVA = 0,3 setiap 3 hasil *falsenegative* test IVA akan didapatkan 10 hasil *truenegative* test IVA. Semakin rendah nilai LR- maka semakin baik kemampuan suatu test untuk mendeteksi suatu penyakit.

Interpretasi untuk sensitifitas, spesifisitas, LR+, LR-, NPP dan NPN test Pap Smear mirip seperti interpretasi di atas hanya saja perlu disesuaikan dengan angka yang didapat pada langkah 2.

2.7 Interpretasi hasil analisis perbedaan kemampuan dua test

Selanjutnya penting disampaikan interpretasi hasil analisis perbandingan kemampuan antara test IVA dan Pap Smear untuk mendeteksi kanker serviks dengan hasil PA sebagai baku emas.

Hipotesis:

Menguji perbedaan kemampuan dua test dapat dilakukan dengan membandingkan *ROC area dari kedua test,* maka hipotesis statistik dari perbedaan tersebut dapat dinyatakan sebagai berikut:

Ho : *ROC area*IVA = *ROC area* Paps Smear

Ha : *ROC area*IVA ≠ *ROC area* Paps Smear

Berdasarkan hasil analisis pada langkah 3 di atas didapatkan luas area di bawah curve untuk test IVA adalah 0,755 (0,710 – 0,799) dan untuk test Paps Smear adalah 0,809 (0,769 – 0,850) dengan nilai p (Prob>chi2) sebesar 0,0747. Dengan menggunakan batas kemaknaan (α) 0,05, maka dapat ditarik simpulan bahwa kemampuan antara test IVA dan Paps Smear untuk mendeteksi kanker serviks dengan hasil PA sebagai baku emas tidak berbeda.

Bagian 4 Analisis *Reciever operating characteristic* (ROC), penentuan titik potong dan membandingkan kemampuan 2 test atau lebih berskala kontinyu

4.1 Pendahuluan

Pada bagian 4 akan dipelajari cara menilai kemampuan suatu test dengan hasil pengukuran test tersebut berskala kontinyu. Analisis yang digunakan untuk menilai kemapuan test dengan hasil berskala kontinyu adalah *reciever operating characteristic* (ROC) *analysis*. Berdasarkan hasil analisis tersebut kemudian ditentukan titik potong (cut of point) yang dipakai untuk menyatakan hasil test positif dan negatif.

4.2 Tujuan Pembelajaran

1. Untuk meningkatkan pengetahuan dan pemahaman tentang analisis ROC.
2. Memahami cara pembuatan kurva ROC dan menilai area under curve (AUC)
3. Memahami konsep penentuan titik potong
4. Mampu mengaplikasikan Stata dalam analisis ROC meliputi: membuat kurva ROC, menghitung AUC, menghitung sensitifitas, spesifisitas, akurasi LR+ dan LR- pada setiap nilai jika digunakan sebagai titik potong.
5. Mampu mengaplikasikan Stata untuk analisis perbandingan 2 atau lebih ROC curve.
6. Mampu membaca, menginterpretasikan hasil analisis, menentukan titik potong terbaik dan mengambil simpulan berdasarkan hasil analisis perbandingan ROC curve.

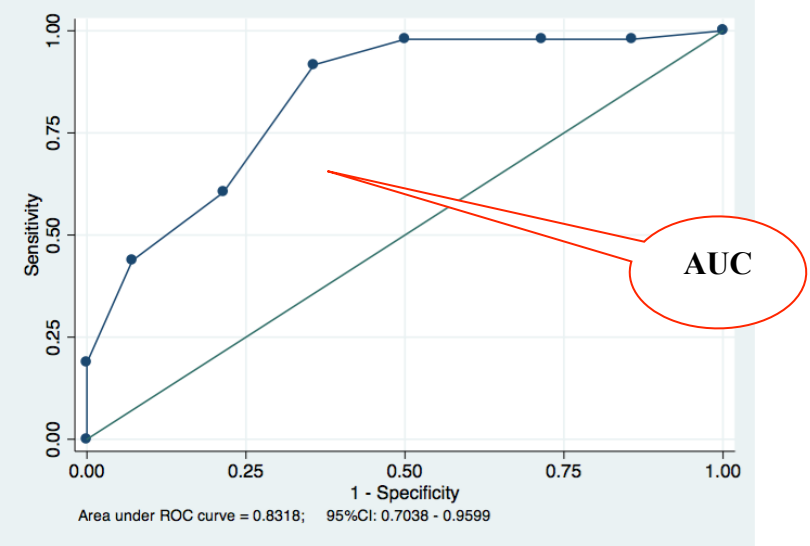
4.3 Pokok Bahasan

1. Definisi dan ukuran yang dinilai dalam analisis ROC
2. Penentuan titik potong
3. Aplikasi Stata utk analisis ROC, penentuan titik potong dan perbadingan 2 atau lebih AUC
4. Interpretasi hasil analisis ROC dan penarikan simpulan

4.4 Definisi dan ukuran yang dinilai dalam analisis ROC

Analisis ROC adalah analisis yang digunakan untuk menilai kemampuan suatu test diagnositik yang hasil pengukurannya berskala kontinyu untuk mendeteksi adanya suatu penyakit menggunakan kurva yang disebut kurva ROC. Kurva ROC adalah suatu kurva yang terbentuk dari titik koordinat antara sensitivitas yang berada di sumbu Y dan 1-spesifisitas yang berada di sumbu X. Setiap titik koordinat mewakili sensitifitas dan 1-spesifisitas yang dihasilkan oleh tiap nilai hasil pengukuran test diagnostik jika digunakan sebagai titik potong.

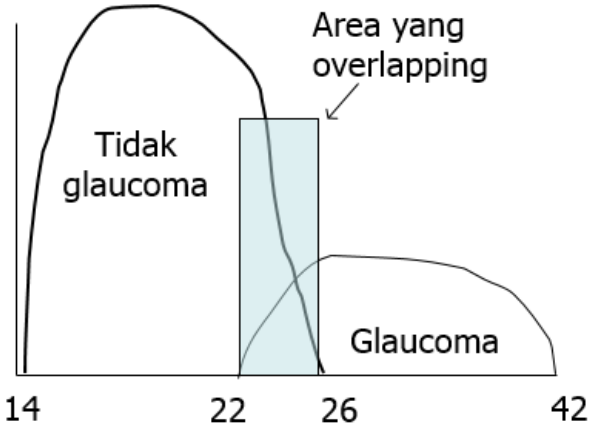
Penilaian terhadap kemampuan suatu test dilakukan dengan menggunakan luas *area under the curve* (AUC). AUC meliputi keseluruhan area di bawah curva yang terbentuk dari semua koordinat sensitifitas dan 1-spesifisitas. Nilai AUC berkisar dari 0 – 1, semakin luas AUC maka semakin baik kemampuan suatu test untuk mendeteksi suatu penyakit. Kemampuan suatu test dinyatakan baik jika AUC ≥ 0,7. Contoh gambar kurva ROC:



Gambar 2. Contoh Kurva ROC

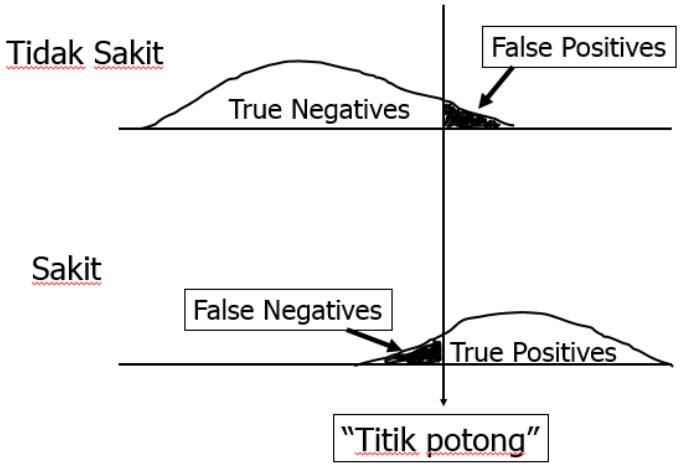
4.5 Penentuan titik potong

Selain untuk menilai kemampuan suatu test, analisis ROC juga digunakan untuk menentukan titik potong (cut of point) suatu hasil test berskala kontinyu untuk dikategorikan menjadi positif dan negatif. Penentuan titik potong merupakan bagian yang sangat penting untuk penelitian uji diagnostik dan screening karena akan menentukan nilai sensitifitas dan spesifisitas yang dihasilkan. Konsep penentuan titik potong dapat dijelaskan melalui gambar berikut ini:



Gambar 3. Sebaran Hasil Pengukuran Tekanan Intra Okular Berdasarkan Status Glaucoma

Berdasarkan Gambar 3 terlihat bahwa ada area yang overlapping antara sebaran hasil pengukuran tekanan intra okular (TIO) antara orang-orang yang glaukoma dengan yang tidak galukoma. Hal ini memerlukan kecermatan peneliti dalam menentukan titik potong karena pergesaran titik potong baik ke kanan maupun ke kiri akan berakibat pada sensitifitas dan spesifisitas test tersebut. Ilustrasi dampak pergeseran titik potong terhadap nilai sensitifitas dan spesifisitas dapat dijelaskan melalui gambar berikut ini:



Gambar 4. Ilustrasi Dampak Pergeseran Titik Potong Terhadap Sensitifitas dan Spesifisitas

Berdasarkan Gambar 4, maka dapat dijelaskan bahwa pergeseran titik potong ke kiri (ke nilai yang lebih kecil) akan mengakibat peningkatan jumlah false positive tetapi disisi lain terjadi penurunan jumlah false negatif. Hal ini akan menyebabkan sensitifitas menjadi lebih tinggi tetapi spesifisitas menjadi lebih rendah. Berbeda jika titik potong

digeser ke kanan (ke nilai yang lebih besar) maka akan mengakibatkan penurunan jumlah false positive tetapi di sisi lain terjadi peningkatan jumlah false negatif. Hal ini akan menyebabkan sensitifitas menjadi lebih rendah tetapi spesifisitas menjadi lebih tinggi.

Pada hasil analisis ROC akan diberikan perhitungan sensitiffitas, spesifisitas, *correctly classified* (akurasi), LR+ dan LR- sehingga memudahkan peneliti untuk memilih titik potong terbaik. Contoh hasil analisis ROC untuk menentukan titik potong sebagai berikut.

Detailed report of sensitivity and specificity

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 5)	100.00%	0.00%	77.42%	1.0000	
(>= 6)	97.92%	14.29%	79.03%	1.1424	0.1458
(>= 7)	97.92%	28.57%	82.26%	1.3708	0.0729
(>= 8)	97.92%	50.00%	87.10%	1.9583	0.0417
(>= 9)	91.67%	64.29%	85.48%	2.5667	0.1296
(>= 10)	60.42%	78.57%	64.52%	2.8194	0.5038
(>= 11)	43.75%	92.86%	54.84%	6.1250	0.6058
(>= 12)	18.75%	100.00%	37.10%		0.8125
(> 12)	0.00%	100.00%	22.58%		1.0000

Gambar 5. Contoh Hasil Analisis ROC dalam Menentukan Titik Potong

Ada 3 hal yang penting dipertimbangkan dalam menentukan titik potong, yaitu nilai sensitifitas dan spesifisitas yang dihasilkan, nilai correctly classified dan ratio LR+/LR- serta yang ketiga adalah tujuan penggunaan uji diagnostik tersebut. Pada contoh diatas, dilihat dari nilai sensitifitas dan spesifisitas yang dihasilkan tiap titik potong (cutpoint) maka ada 2 titik potong yang bisa digunakan untuk menyatakan postif, yaitu ≥8 atau ≥9. Selanjutnya jika mepertimbangkan ingin dipilih titik potong dengan nilai correctly classified tertinggi dan ratio LR+/LR- tertinggi maka ≥8 akan menjadi pilihan. Tetapi jika mempertimbangkan tujuan supaya hasil diagnosis lebih efektif dengan kata lain supaya menghasilkan sedikit positif palsu maka titik potong ≥9 menjadi pilihan.

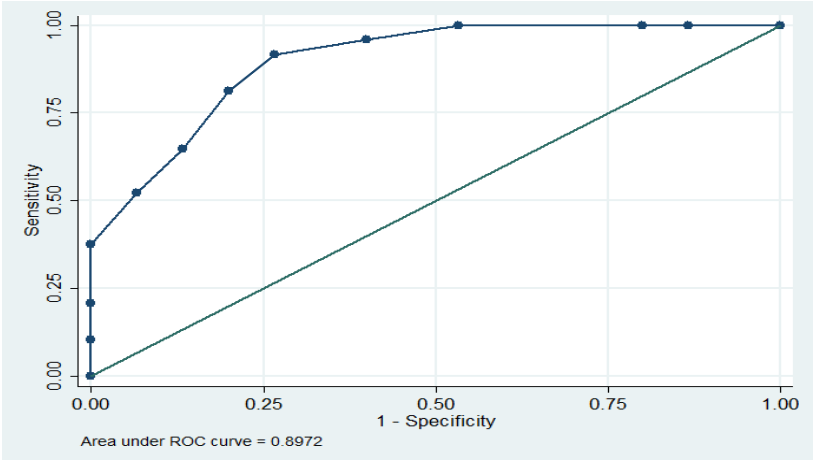
4.6 Aplikasi Stata

Dalam rangka praktek aplikasi Stata untuk analisis ROC dan penentuan titik potong maka bukalah terlebih dahulu data Diagnostik1.dta. Data ini merupakan suatu data penelitian yang bertujuan untuk menilai kemampuan skor Alvarado dan skor Tzanakis dalam mendiagnosis apendisitis akut dengan hasil PA sebagai baku emas. Data ini telah mengalami perubahan sedemikian rupa untuk kepentingan pelatihan sehingga

sudah tidak mencerminkan data aslinya. Selanjutnya lakukanlah analisis untuk menilai kemampuan skor Tzanakis dalam mendiagnosis apendisitis akut dengan hasil PA sebagai baku emas kemudian tentukanlah titik potongnya. Untuk memulai analisis tujuan tersebut maka ketiklah pada window command sebagai berikut:

roctab hasilpa skortzanakis, detail graph summary

Kemudian tekan entermaka akan diberikan hasil analisis sebagai berikut:



Dan dalam bentuk tabel sebagai berikut:

```
. roctab hasilpa skortzanakis, detail graph summary

Detailed report of sensitivity and specificity
```

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 4)	100.00%	0.00%	76.19%	1.0000	
(>= 5)	100.00%	13.33%	79.37%	1.1538	0.0000
(>= 6)	100.00%	20.00%	80.95%	1.2500	0.0000
(>= 7)	100.00%	46.67%	87.30%	1.8750	0.0000
(>= 8)	95.83%	60.00%	87.30%	2.3958	0.0694
(>= 9)	91.67%	73.33%	87.30%	3.4375	0.1136
(>= 10)	81.25%	80.00%	80.95%	4.0625	0.2344
(>= 11)	64.58%	86.67%	69.84%	4.8437	0.4087
(>= 12)	52.08%	93.33%	61.90%	7.8125	0.5134
(>= 13)	37.50%	100.00%	52.38%		0.6250
(>= 14)	20.83%	100.00%	39.68%		0.7917
(>= 15)	10.42%	100.00%	31.75%		0.8958
(> 15)	0.00%	100.00%	23.81%		1.0000

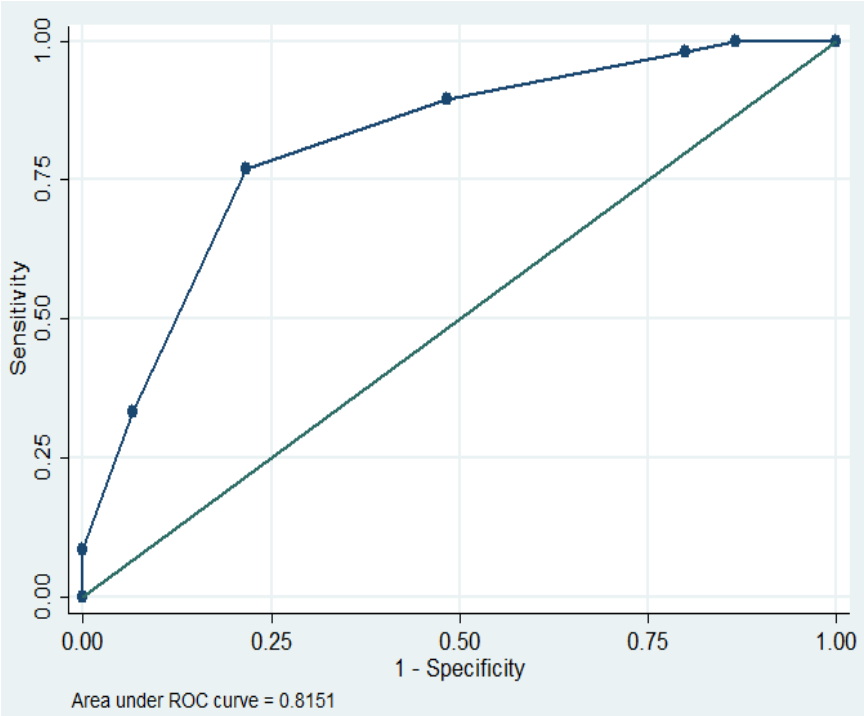
Obs	ROC Area	Std. Err.	Asymptotic Normal [95% Conf.	Interval]
252	0.8972	0.0235	0.85114	0.94330

Kemudian lakukanlah analisis untuk menilai kemampuan skor Alvarado dalam mendiagnosis apendisitis akut dengan hasil PA sebagai baku emas kemudian tentukanlah titik potong terbaik dari skor Alvarado.

Untuk memulai analisis tujuan tersebut maka ketiklah pada window command sebagai berikut:

roctab hasilpa skoralvarado, detail graph summary

Kemudian tekan enter maka akan diberikan juga hasil analisis sebagai berikut:



Dan kedua dalam bentuk tabel sebagai berikut:

```
. roctab hasilpa skoralvarado, detail graph summary

Detailed report of sensitivity and specificity
```

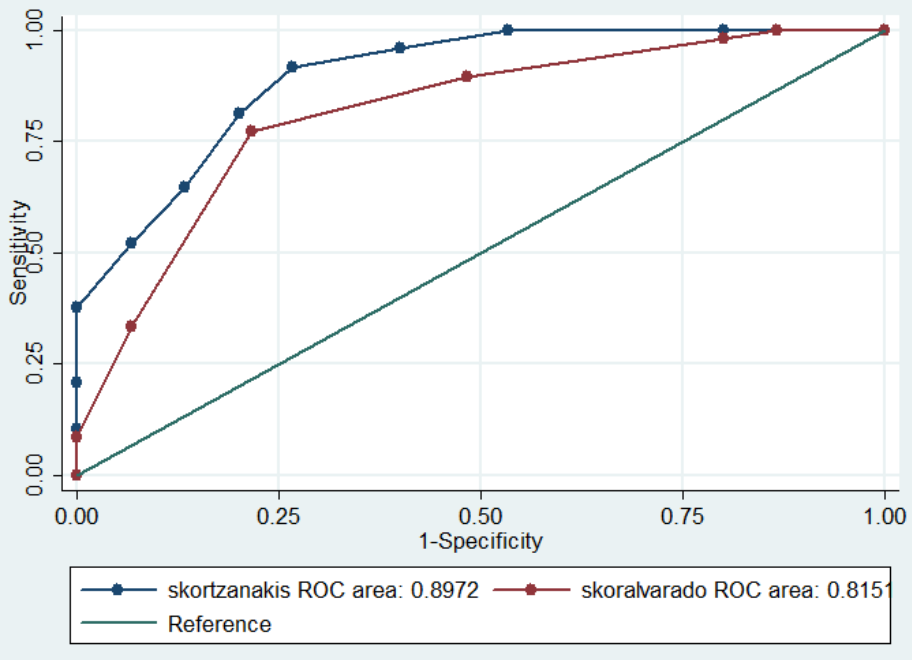
Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 4)	100.00%	0.00%	76.19%	1.0000	
(>= 5)	100.00%	13.33%	79.37%	1.1538	0.0000
(>= 6)	97.92%	20.00%	79.37%	1.2240	0.1042
(>= 7)	89.58%	51.67%	80.56%	1.8534	0.2016
(>= 8)	77.08%	78.33%	77.38%	3.5577	0.2926
(>= 9)	33.33%	93.33%	47.62%	5.0000	0.7143
(>= 10)	8.33%	100.00%	30.16%		0.9167
(> 10)	0.00%	100.00%	23.81%		1.0000

Obs	ROC Area	Std. Err.	Asymptotic Normal [95% Conf.	Interval]
252	0.8151	0.0315	0.75345	0.87676

Untuk membandingkan kemampuan kedua skor apakah berbeda dalam mendiagnosis apendisitis akut dengan hasil PA sebagai baku emas maka dilakukan prosedur*compare ROC curve*, dengan mengetik pada windows command sebagai berikut:

roccomp hasilpa skortzanakis skoralvarado, graph summary

Kemudian tekan enter, maka akan diberikan hasil sebagai berikut:



Dan kedua dalam bentuk tabel sebagai berikut:

```
. roccomp hasilpa skortzanakis skoralvarado, graph summary
```

	Obs	ROC Area	Std. Err.	Asymptotic Normal [95% Conf.	Interval]
skortzanakis	252	0.8972	0.0235	0.85114	0.94330
skoralvarado	252	0.8151	0.0315	0.75345	0.87676

```
Ho: area(skortzanakis) = area(skoralvarado)
    chi2(1) =     4.09       Prob>chi2 =    0.0433
```

4.7 Interpretasi Hasil

Berdasarkan analisis diatas maka dapat dibuat beberapa interpretasi sebagai berikut:

1. Kemampuan Skor Tzanakis dalam mendiagnosis apendisitis akut termasuk baik dengan area AUC sebesar 0,897 (89,7%) atau ≥ 70%.
2. Titik potong terbaik dari skor Tzanakis untuk mendiagnosis apendisitis akut adalah ≥ 9 yang menghasilkan sensitifitas sebesar 91,67% dan spesifisitas

sebesar 73,33% karena memiliki nilai correctly classified dan ratio LR+/LR- paling tinggi.

3. Kemampuan Skor Alvarado dalam mendiagnosis apendisitis akut termasuk baik dengan area AUC sebesar 0,815 (81,5%) atau ≥ 70%.
4. Titik potong terbaik dari skor Alavarado untuk mendiagnosis apendisitis akut adalah ≥ 8 yang menghasilkan sensitifitas sebesar 77,08% dan spesifisitas sebesar 78,33% karena pada titik tersebut menghasilkan ratio LR+/LR- tertinggi walapun bila dilihat nilai correctly classifeid nya tidak. Dalam kondisi seperti ini prioritas penilaian titik potong menggunakan ratio LR+/LR- karena lebih stabil, konsisten dan valid.

Selanjutnya penting disampaikan interpretasi hasil analisis perbandingan kemampuan antara Skor Tzanakis dan Skor Alvarado untuk mendeteksi apendisitis akut dengan hasil PA sebagai baku emas. Hasil analisis perbedaan ini tentu didasari oleh suatu hipotesis :
"Ada perbedaan kemampuan antara Skor Tzanakis dan Skor Alvarado untuk mendeteksi apendisitis akutdengan hasil PA sebagai baku emas"
Dalam menguji kemapuan suatu test yang dibandingkan adalah ROC area maka dapat ditulis kan hipotesis statistik sebagai berikut:
Ho : ROC area Skor Tzanakis = ROC area Skor Alvarado
Ha : ROC area Skor Tzanakis ≠ ROC area Skor Alvarado

5. Berdasarkan hasil analisis pada langkah 3 didapatkan nilai p (Prob>chi2) sebesar 0,0433 (< dibandingkan batas kemaknaan (α) 0,05) maka dapat ditarik simpulan bahwa kemampuan antara Skor Tzanakis dan Skor Alvarado untuk mendeteksi apendisitis akut dengan hasil PA sebagai baku emas **berbeda**.

Bagian 5 Uji Skrining Kombinasi (Multiple)

5.1 Pendahuluan

Uji skrining kombinasi adalah uji skrining yang menggunakan lebih dari satu test (multiple). Ada 2 cara penggunaan uji skrining kombinasi, pertama secara bertahap atau serial (*two stage screening*) dan yang kedua adalah secara simultan atau pararel (simultaneous screening). Masing-masing cara uji skrining kombinasi ini memiliki kelebihan dan kekurangan serta masing-masing juga akan berdampak terhadap hasil perhitungan net sensitifitas dan net spesifisitas. Untuk itu penggunaannya disesuaikan dengan kebutuhan atau tujuan program skrining. Bila ingin mendapatkan uji skrining multiple yang menghasilkan sensitifitas lebih tinggi maka dilakukan secara simultan atau pararel sedangkan bila ingin spesifisitas lebih tinggi maka dilakukan secara bertahap atau serial. Pada modul bagian 5 ini akan dibahas konsep, cara perhitungan, tujuan dan aplikasi uji skring kombinasi menggunakan Stata.

5.2 Tujuan Pembelajaran

1. Untuk meningkatkan pengetahuan dan pemahaman tentang uji skrining kombinasi.
2. Memahami penggunaan masing-masing cara uji skrining kombinasi sesuai dengan tujuan program skrining.
3. Mampu mengaplikasikan Stata dalam uji skrining kombinasi meliputi: secara bertahap (serial) dan simultan (pararel).
4. Mampu membaca, menginterpretasikan hasil analisis uji skrining kombinasi dan menentukan penggunaannya dalam program skrining.

5.3 Pokok Bahasan

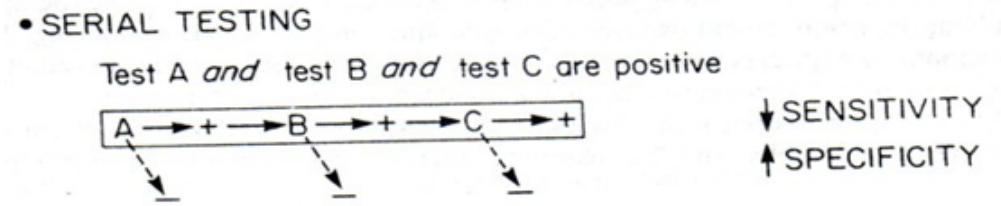
1. Konsep uji skrining kombinasi
2. Program skrining
3. Pertimbangan penggunaan suatu test pada program skrining
4. Aplikasi stata dalam melakukan analisis uji skrining kombinasi
5. Interpretasi hasil analisis uji skrining kombinasi dan menentukan penggunaannya dalam program skrining

5.4 Konsep Uji Skrining Kombinasi (Multiple)

Uji skrining kombinasi adalah uji skrining yang menggunakan lebih dari satu test (multiple) yang dilakukan dengan tujuan tertentu pada program skrining. Ada 2 cara dalam melakukan kombinasi beberapa test skrining, pertama secara bertahap atau serial (*two stage screening*) dan yang kedua adalah secara simultan atau pararel (simultaneous screening).

1. Uji Skrining secara bertahap (serial)

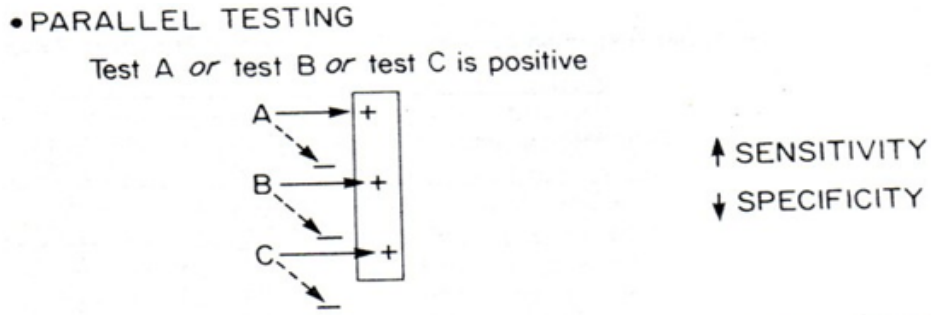
   Uji Skrining secara bertahap (serial) adalah mengunakan lebih dari satu test secara bertahap dimana untuk menyatakan seseorang positif berdasarkan hasil positif dari semua test yang digunakan. Bila hanya salah satu test menyatakan positif maka masih dinyatakan negatif. Penggunaan kombinasi dengan cara ini untuk lebih memastikan orang-orang yang dinyatakan positif dengan kata lain untuk mengurangi jumlah *false positive*. Dengan berkurangnya jumlah *false positive* maka spesifisitas akan meningkat. Tetapi disisi lain *false negative* akan meningkat sehingga sensitifitas akan menurun. Kombinasi secara bertahap (serial) dapat digambarkan melalui bagan berikut ini.



Gambar 5. Bagan Uji Skrining Kombinasi Secara Bertahap (Serial)

2. Uji Skrining kompinasi secara simultan (pararel)

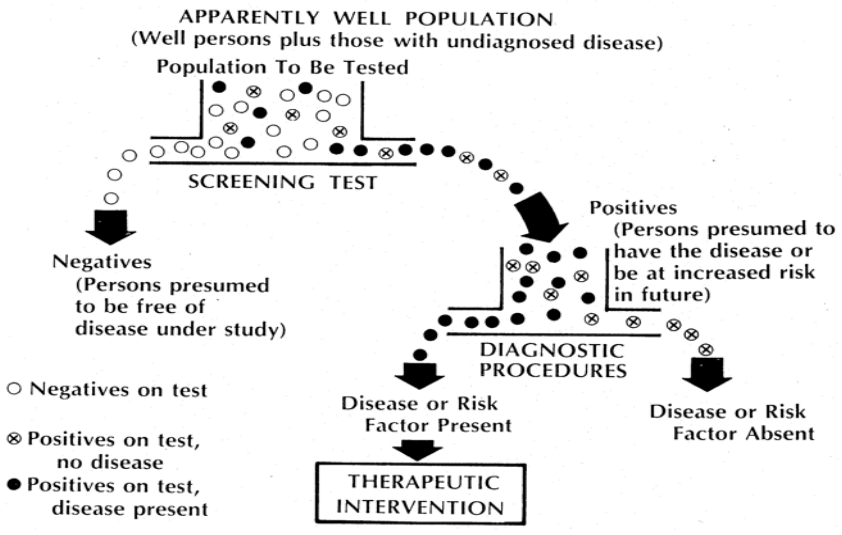
   Uji Skrining secara simultan (pararel) adalah mengunakan lebih dari satu test secara pararel dimana untuk menyatakan seseorang positif berdasarkan **salah satu** hasil positif dari beberapa test yang digunakan. Sedangkan hasil test negatif berdasarkan hasil test negatif semua test yang digunakan. Penggunaan kombinasi dengan cara ini untuk mendapatkan lebih banyak orang yang dinyatakan positif dengan kata lain untuk mengurangi jumlah *false negative*. Dengan berkurangnya jumlah *false negative*. maka sensitifitas akan meningkat. Tetapi disisi lain *false positive* akan meningkat sehingga spesifisitas akan menurun. Kombinasi secara simultan (pararel) dapat digambarkan melalui bagan berikut ini.



Gambar 6. Bagan Uji Skrining Kombinasi Secara Simultan (Pararel)

5.5 Program Skrining

Program skrining dalam tingkat pencegahan termasuk dalam pencegahan sekunder yang terdiri dari deteksi dini dan pengobatan tepat. Deteksi dini pada program skrining dilakukan pada orang-orang yang tanpa gejala (preklinis) untuk memisahkan orang yang mungkin menderita suatu penyakit (hasil skrining positif) dari orang orang yang dinyatakan tidak sakit (hasil skrining negatif). Orang-orang yang dengan hasil skrining positif akan lanjut ke prosedur diagnostik untuk memastikan ada tidaknya penyakit. Orang yang terdiagnosis sakit maka akan menjalani intervensi berupa terapi sampai sembuh. Hal ini berarti bahwa program skrining tidak hanya sebatas deteksi dini menggunkan suatu test, tetapi adalah suatu program menyeluruh dari deteksi dini, prosedur diagnostik sampai terapi pasien hingga sembuh. Pada program penanggulangan penyakit kronis, program skrining sangat penting untuk menemukan pasien secara dini, diobati kemudian disembuhkan sehingga dapat menurunkan prevalensi penyakit di populasi. Bagan program skrining dapat dijelaskan melalui Gambar 7 berikut ini.



Gambar 7. Bagan Program Skrining

5.6 Pertimbangan Penggunaan Suatu Test pada Program Skrining

Dalam melaksanakan suatu program skrining pemilihan test skrining sangat penting mempertimbangkan sensitifitas dan spesifisitasnya disesuaikan dengan tujuan dan dukungan prasarana, sarana penunjang seperti ketersediaan fasilitas, ahli untuk prosedur diagnosis dan kemampuan untuk layanan pengobatan. Test skrining yang baik tentu memiliki sensitifitas dan spesifisitas tinggi tetapi tidak semua bahkan sebagian besar test skrining tidak mempunyai kemampuan seperti itu. Kita sering dihadapkan untuk memilih antara yang sensitifitasnya tinggi atau spesifisitasnya tinggi.

Contoh bila dihadapkan suatu kondisi untuk melaksanakan program skrining untuk menurunkan prevalensi suatu penyakit tetapi dana, fasilitas dan ahli untuk prosedur diagnosis terbatas kemudian kita dihadapkan untuk memilih antara test skrining dengan sensitifitas tinggi tetapi spsifisitas sedang atau spesifisitas tinggi tetapi sensitifitas sedang. Oleh karena dana, fasilitas dan ahli untuk prosedur diagnosis terbatas maka kita ingin program skrining berjalan seefektif mungkin. Agar program berjalan efektif maka diharapkan orang yang nantinya menjalani prosedur adalah orang-orang yang tinggi kemungkinannya sakit. Dengan kata lain kita berusaha untuk meminimalisir *false positif*. Maka test skrining yang dipilih supaya *false positif* sedikit adalah test dengan spesifisitas tinggi tetapi sensitifitas sedang.

5.7 Aplikasi Stata Untuk Uji Skrining Kombinasi (Multiple)

1. **Aplikasi Stata untuk uji skrining kombinasi secara bertahap (serial)**

   Latihan aplikasi Stata untuk uji skrining kombinasi secara bertahap (serial) mengunakan data Skrining1.dta. Tugas untuk dikerjakan adalah melakukan uji skrining kombinasi test IVA dan Pap Smear secara bertahap (serial). Berapakah net sensitifitas dan spesifisitasnya?

   Untuk mengerjakan tugas ini maka bukalah terlebih dahulu data Skrining1.dta kemudian ikutilah langkah-langkah berikut ini:

   Langkah 1

   Membuat tabulasi silang antara IVA dengan Pap Smear dengan perintah:

   tab iva papsmear

   Kemudian tekan enter maka terlihat hasil sebagai berikut:

```
. tab iva papsmear
```

iva	papsmear Negatif	Positif	Total
Negatif	132	64	196
Positif	68	104	172
Total	200	168	368

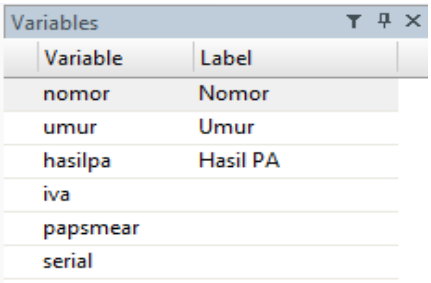
Berdasarkan tabulasi silang tersebut, jika akan dibuat test kombinasi secara serial, maka akan ada 104 orang dinyatakan positif (dinyatakan positif oleh kedua test) dan siasanya 264 orang dinyatakan negatif.

Langkah 2

Membuat variabel baru test kombinasi yang diberi nama "serial" berdasarkam kombinasi IVA dengan Pap Smear secara serial, menggunakan perintah:

gen serial = iva*papsmear

Kemudian tekan enter, maka pada windows variabel (pojok kanan atas) terlihat bertambah 1 variabel dengan nama "serial", terlihat sebagai berikut:



Ini artinya telah dibuat variabel baru bernama "serial" dari hasil perkalian nilai variabel iva dan papsmear. Oleh karena nilai variabel iva dan papsmear adalah 0 dan 1 maka jika kedua kedua variabel iva dan papsmear bernilai 1 pada nilai variabel serial menjadi 1 sedangkan kombinasi lainnya (0 dengan 0, 1 dengan 0 dan 0 dengan 1) akan menjadi 0. Untuk membuktikan hal tersebut ketiklah perintah

tab serial

Kemudian tekan enter, maka akan terlihat hasil sebagai berikut:

```
. gen serial = iva*papsmear

. tab serial
```

serial	Freq.	Percent	Cum.
0	264	71.74	71.74
1	104	28.26	100.00
Total	368	100.00	

Berdasarkan test kombinasi secara serial tersebut ada 104 dinyatakan positif (value 1) dan sisanya sebanyak 264 dinyatakan negatif (value 0). Kemudian berilah label value dengan cara menggunakan label value yang sudah ada menggunakan perintah"

lab val serial pn

kemudian tekan enter sehingga jika diulangi perintah **tab serial** akan terlihat menjadi:

```
. tab serial
```

serial	Freq.	Percent	Cum.
Negatif	264	71.74	71.74
Positif	104	28.26	100.00
Total	368	100.00	

Langkah 3

Menghitung net sensitifitas dan net spesfisitas test skrining kombinasi secara serial dengan hasil PA sebagai baku emas menggunakan perintah:

diagt hasilpa serial

Kemudian tekan enter maka akan terlihat hasil sebagai berikut:

```
. diagt hasilpa serial
```

	serial		
Hasil PA	Pos.	Neg.	Total
Abnormal	96	48	144
Normal	8	216	224
Total	104	264	368

```
True abnormal diagnosis defined as hasilpa = 1 (labelled Positif)
```

			[95% Confidence Interval]	
Prevalence	Pr(A)	39%	34%	44.3%
Sensitivity	Pr(+|A)	66.7%	58.3%	74.3%
Specificity	Pr(-|N)	96.4%	93.1%	98.4%
ROC area	(Sens. + Spec.)/2	.815	.775	.856
Likelihood ratio (+)	Pr(+|A)/Pr(+|N)	18.7	9.36	37.2
Likelihood ratio (-)	Pr(-|A)/Pr(-|N)	.346	.274	.436
Odds ratio	LR(+)/LR(-)	54	24.9	117
Positive predictive value	Pr(A|+)	92.3%	85.4%	96.6%
Negative predictive value	Pr(N|-)	81.8%	76.6%	86.3%

Bandingkanlah hasil ini dengan hasil analisis pada Bagian 3, sub 3.4 dari modul ini, dimana test IVA memiliki sensitifitas 77,8%; spesifisitas 73,2% dan Pap Smear memiliki sensitifitas 83,3%; spesifisitas 78,6%. Terlihat jumlah *false positive* sangat sedikit (8 orang) sehingga net spesifisitas meningkat menjadi 96,6% tetapi di sisi lain net sensitifitas turun menjadi 66,7%.

2. Aplikasi Stata untuk uji skrining kombinasi secara simultan (pararel)

Tugas untuk dikerjakan adalah melakukan uji skrining kombinasi test IVA dan Pap Smear secara simultan (pararel). Berapakah net sensitifitas dan spesifisitasnya?

Untuk mengerjakan tugas ini ikutilah langkah-langkah berikut ini:

Langkah 1

Membuat tabulasi silang antara IVA dengan Pap Smear dengan perintah:

tab iva papsmear

Kemudian tekan enter maka terlihat hasil sebagai berikut:

```
. tab iva papsmear
```

	papsmear		
iva	Negatif	Positif	Total
Negatif	132	64	196
Positif	68	104	172
Total	200	168	368

Berdasarkan tabulasi silang tersebut maka jika dibuat test kombinasi secara paralel akan ada 132 orang dinyatakan negatif (dinyatakan negatif oleh kedua test) dan siasanya 236 orang dinyatakan positif.

Langkah 2

Membuat variabel baru test kombinasi yang diberi nama “pararel” berdasarkam kombinasi IVA dengan Pap Smear secara pararel, menggunakan perintah:

gen pararel = iva+papsmear

Kemudian tekan enter, maka pada windows variabel (pojok kanan atas) terlihat bertambah 1 lagi variabel dengan nama “pararel”, terlihat sebagai berikut:

Variables

Variable	Label
nomor	Nomor
umur	Umur
hasilpa	Hasil PA
iva	
papsmear	
serial	
pararel	

Ini artinya telah dibuat variabel baru bernama "pararel" dari hasil penjumlahan nilai variabel iva dan papsmear. Oleh karena nilai variabel iva dan papsmear adalah 0 dan 1 maka jika kedua variabel iva dan papsmear bernilai 0 pada nilai variabel pararel menjadi 0,bila kombinasinya 1 dengan 0 dan 0 dengan 1 pada nilai variabel pararel menjadi 1 dan jika kedua variabel iva dan papsmear bernilai 1 pada nilai variabel pararel menjadi 2. Untuk membuktikan hal tersebut ketiklah perintah

tab pararel

Kemudian tekan enter, maka akan terlihat hasil sebagai berikut:

```
. tab pararel
```

pararel	Freq.	Percent	Cum.
0	132	35.87	35.87
1	132	35.87	71.74
2	104	28.26	100.00
Total	368	100.00	

Oleh karena nilai 2 dan nilai 1 akan di label sama menjadi positif maka lakukanlah prosedur recode menggunakan perintah:

recode pararel 2=1

Kemudian tekan enter dan jika diulangi perintah **tab pararel** maka terlihat menjadi:

```
. tab pararel
```

pararel	Freq.	Percent	Cum.
0	132	35.87	35.87
1	236	64.13	100.00
Total	368	100.00	

Berdasarkan test kombinasi secara pararel tersebut ada 236 orang dinyatakan positif (value 1) dan sisanya sebanyak 132 dinyatakan negatif (value 0). Kemudian berilah label value dengan cara menggunakan label value yang sudah ada menggunakan perintah"

lab val pararel pn

kemudian tekan enter sehingga jika diulangi perintah **tab serial** akan terlihat menjadi:

```
. tab pararel
```

pararel	Freq.	Percent	Cum.
Negatif	132	35.87	35.87
Positif	236	64.13	100.00
Total	368	100.00	

Langkah 3

Menghitung net sensitifitas dan net spesfisitas test skrining kombinasi secara pararel dengan hasil PA sebagai baku emas menggunakan perintah:

diagt hasilpa pararel

Kemudian tekan enter maka akan terlihat hasil sebagai berikut:

```
. diagt hasilpa pararel
```

	pararel		
Hasil PA	Pos.	Neg.	Total
Abnormal	136	8	144
Normal	100	124	224
Total	236	132	368

```
True abnormal diagnosis defined as hasilpa = 1 (labelled Positif)
```

			[95% Confidence Interval]	
Prevalence	Pr(A)	39%	34%	44.3%
Sensitivity	Pr(+|A)	94.4%	89.3%	97.6%
Specificity	Pr(-|N)	55.4%	48.6%	62%
ROC area	(Sens. + Spec.)/2	.749	.711	.787
Likelihood ratio (+)	Pr(+|A)/Pr(+|N)	2.12	1.82	2.46
Likelihood ratio (-)	Pr(-|A)/Pr(-|N)	.1	.0507	.199
Odds ratio	LR(+)/LR(-)	21.1	10	44.4
Positive predictive value	Pr(A|+)	57.6%	51%	64%
Negative predictive value	Pr(N|-)	93.9%	88.4%	97.3%

Bandingkanlah hasil ini dengan hasil analisis pada Bagian 3, sub 3.4 dari modul ini, dimana test IVA memiliki sensitifitas 77,8%; spesifisitas 73,2% dan Pap Smear memiliki sensitifitas 83,3%; spesifisitas 78,6%. Terlihat jumlah *false negative* sangat sedikit (8 orang) sehingga net sensitifitas meningkat menjadi 94,4% tetapi di sisi lain net spesifisitas turun menjadi 55,4%.

5.8 Interpretasi Hasil

Berdasarkan hasil analisis diatas, kombinasi test secara serial dan pararel, maka data dibuat beberapa interpretasi sebagai berikut:

Kombinasi test IVA dengan Pap Smear secara serial:

1. Net sensitifitas sebesar 66,7%, artinya kemampuan kombinasi test IVA dengan Pap Smear secara serial untuk menyatakan positif wanita yang kanker serviks sebesar 66,7%.
2. Net spesifisitas sebesar 96,4%, artinyakemampuan kombinasi test IVA dengan Pap Smear secara serial untuk menyatakan negatif wanita yang **tidak** kanker serviks sebesar 96,4%.
3. NPP sebesar 92,3% artinya besarnya kemungkinan wanita di populasi studi yang mendapat hasil positif berdasarkan kombinasi test IVA dengan Pap Smear secara serial untuk benar-benar mengalami kanker serviks sebesar 92,3%.

Kombinasi test IVA dengan Pap Smear secara pararel:

1. Net sensitifitas sebesar 94,4%, artinyakemampuan kombinasi test IVA dengan Pap Smear secara pararel untuk menyatakan positif wanita yang kanker serviks sebesar 94,4%.
2. Net spesifisitas sebesar 55,4%, artinyakemampuan kombinasi test IVA dengan Pap Smear secara pararel untuk menyatakan negatif wanita yang **tidak** kanker serviks sebesar 55,4%.
3. NPP sebesar 57,6% artinya besarnya kemungkinan wanita di populasi studi yang mendapat hasil positif berdasarkan kombinasi test IVA dengan Pap Smear secara pararel untuk benar-benar mengalami kanker serviks sebesar 57,6%.

Bagian 6 Menilai Reliabilitas Suatu Test

6.1 Pendahuluan

Reliabilitas adalah kemampuan suatu test untuk menghasilkan nilai yang sama dalam mengukur variabel tertentu pada individu yang sama dan kondisi yang sama pula. Sinonim dari reliabilitas yang sering digunakan adalah konsitensi, agrement dan ada juga yang menyebut sebagai *repeatability*. Reliabilitas dinilai dengan cara membandingkan lebih dari satu kali hasil ukur suatu test pada individu atau subjek yang sama dan pada kondisi yang sama. Pengukuran berulang tersebut bisa dilakukan oleh observer yang berbeda atau bisa juga oleh observer yang sama tetapi pada waktu berbeda walapun demikian kondisi subjek harus masih sama. Pengukuran berulang pada subjek yang sama ini tentu dapat menghasilkan nilai yang berbeda atau disebut sebagai bias.

Bias dapat dibagi menajdi dua jenis, bias inter observer dan bias intra observer. Bias inter observer adalah bias yang terjadi akibat dua observer menilai satu hasil test dan memberi penilaian yang berbeda. Salah satu contoh adalah perbedaan hasil pembacaan foto Rontgen yang sama oleh dua orang ahli radiologi. Bias intra observer adalah bias yang terjadi karena satu observer menilai berbeda terhadap satu hasil test dalam waktu yang berbeda. Contohnya adalah variasi yang terjadi pada pengukuran tekanan darah pada waktu yang berbeda pada seseorang dapat memberikan hasil yang berbeda.

Uji reliabilitas sebaiknya dilakukan sebelum uji diagnosis dilaksanakan. Apabila obervernya satu orang, maka uji reliabilitas dapat dilakukan secara paralel (satu kasus dilakukan pengukuran dua kali pada waktu yang berbeda oleh satu observer). Sebaliknya, bila observernya dua atau lebih, sebaiknya dilakukan secara paralel (satu kasus diukur oleh semua observer). Apabila reliabilitas rendah, maka sebelum uji diagnostik dilakukan perlu dilakukan upaya untuk meningkatkan reliablitas hasil pengukuran, misalnya melakukan pelatihan observer, merinci kriteria yang diukur atau bila terjadi disagrrement dimintakan pendapat ahli lain (second oppinion).

6.2 Tujuan Pembelajaran

1. Untuk meningkatkan pengetahuan dan pemahaman tentang penilaian reliabilitas suatu test.
2. Memahami ukuran/parameter yang digunakan menilai reliabiitas suatu test baik yang berskala data kategorikal maupun kontinyu.

3. Mampu mengaplikasikan Stata dalam menilai reliabilitas suatu test baik yang berskala data kategorikal maupun kontinyu.
4. Mampu membaca, menginterpretasikan hasil analisis dalam menilai reliabilitas suatu test.

6.3 Pokok Bahasan

1. Ukuran untuk menilai reliabilitas test dengan hasil ukur berskala data kategorikal
2. Ukuran untuk menilai reliabilitas test dengan hasil ukur berskala data kontinyu
3. Aplikasi Stata untuk menghitung ukuran reliabilitas test berskala data kategorikal
4. Aplikasi Stata untuk menghitung ukuran reliabilitas test berskala data kontinyu
5. Interpretasi hasil analisis penilaian reliabilitas

6.4 Ukuran menilai reliabilitas test dengan hasil ukur berskala data kategorikal

Dalam menilai reliabilitas suatu test dengan hasil ukur berskala kategorikal digunakan *percent agreement* dan koefesian kappa. *Percent agreement* adalah proporsi jumlah hasil penilaian yang sama antar observer diantara semua yang dinilai dalam persen. Hasil penilaian antar observer maupun intra observer dapat disajikan dalam tabulasi silang sebagai berikut:

		Observer B		
		+	-	
Observer A	+	a	b	a+b
	-	c	d	c+d
		a+c	b+d	Total

$$Percent\ agreement = \frac{a + \mathrm{d}}{Total} x100\%$$

$$koef\ kappa = \frac{(\text{observed aggrement}) - (expected\ agreement)}{1 - (expected\ \text{agreement})}$$

$$Observed\ agreement = \frac{(a + \mathrm{d})}{Total}$$

$$Expected\ agreement = \frac{\left(\frac{(a+\mathrm{b})\mathrm{x}(\mathrm{a}+\mathrm{c})}{\text{total}}\right) + \left(\frac{(c+\mathrm{d})\mathrm{x}(\mathrm{b}+\mathrm{d})}{\text{total}}\right)}{Total}$$

Interpretasi nilai Koefesien Kappa (Altman, 1991):

0.8 - 1	: sangat baik (very good)
0.6 - <0.8	: baik (good)
0.4 - <0.6	: moderate
0.2 - <0.4	: cukup (fair)
<0.2	: buruk (poor)

Untuk meningkatkan reliabilitas suatu test yang berskala data kategorikal dapat dilakukan dengan cara menyederhanakan kategori. Contohnya hasil penilaian yang sebelumnya 5 kategori kemudian disederhanakan menjadi 2 kategori saja maka hasil perhitungan precent agreement dan koefesien kappa akan menjadi lebih tinggi. Hal ini akan dibuktikan nanti padasaat latihan menggunakan Stata.

6.5 Ukuran untuk menilai reliabilitas test dengan hasil ukur berskala data kontinyu

Dalam menilai reliabilitas suatu test dengan hasil ukur berskala data kontinyu digunakan *concordance correlation coefficient (rho_c), average difference dan 95% limit of agreements. Concordance correlation coefficient (rho_c)* menunjukkan seberapa besar deviasi hasil pengukuran suatu variabel oleh dua observer. Nilai rho_c antara 0 – 1, semakin tinggi nilai rho_c maka semakin tinggi kesesuaian hasil pengukuran antar observer atau deviasinya semakin kecil. Average difference menunjukkan rerata perbedaan hasil pengukuran antar observer dan digunakan untuk menilai kecendrungan penilaian antar observer mana yang memberikan nilai lebih besar atau lebih kecil. Perhitungan average difference ini sama dengan perhitungan rerata perbedaan pada *paired t test*.

Ukuran reliabilitas yang lebih penting adalah *95% limit of agreements* yang menunjukkan variasi perbedaan hasil pengukuran antar observer pada rentang 95% data. Semakin lebar *95% limit of agreements* maka semakin lebar perbedaan hasil pengukuran antar observer. Penilaian reliabilitas menggunakan *95% limit of agreements* penting memperhatikan konteks penilaian suatu variabelsecara klinis. Contohnya dalam menilai reliabilitas antar alat dalam mengukur pH darah didapatkan *95% limit of agreements:*- 0,08 – 0,09 tergolong kurang reliabel karena secara klinis penilaian pH memang sangat ketat, rentang normalnya sangat sempit (7,35 – 7,45). Contoh lain dalam menilai volume

tumor didapatkan *95% limit of agreements:* -1,5 – 1,6 tergolong reliabel karena secara klinis rentang volume tumor memang sangat lebar dari 1 – 50 cm^3.

6.6 Aplikasi Stata menghitung ukuran reliabilitas test berskala data kategorikal

Dalam latihan kali ini akan dipelajari aplikasi Stata untuk menghitung percent agreement dan koefesien kappa. Terlebih dahulu bukalah data Reliabilitas1.dta. Data ini adalah data penelitian tentang penilaian stadium kanker oleh 2 orang klinisi (a dan b). Penilaian stadium kanker ini terbagi menjadi 5 kategori: stadium I, II, III, IVA dan IVB. Setelah data berhasil dibuka kemudian ketiklah perintah sebagai berikut:

kap stadium_a stadium_b

Kemudian tekan enter, makan akan terlihat hasil sebagai berikut:

```
. kap stadium_a stadium_b
```

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
50.00%	21.75%	0.3610	0.0582	6.20	0.0000

Berdasarkan hasil di atas didapatkan % agreement dan koefesien kappa yang relative rendah (tergolong cukup). Salah satu penyebabnya adalah kategori stadium yang relative banyak (5 kategori). Untuk meningkatkan agreement, sederhanakanlah kategori stadium menjadi 2, stadium 1 dan 2 menjadi lower grade diberi value 1 dan stadium 3 dan 4 menjadi higer grade dan diberi value 2. Caranya dengan menjalan beberapa perintah satu persatu sebagai berikut:

gen grade_a = stadium_a

recode grade_a 1/2=1 3/5=2

gen grade_b = stadium_b

recode grade_b 1/2=1 3/5=2

Kemudian untung menghitung kembali % agreement dan koefesien kappa, ketiklah pereintah sebagai berikut:

kap grade_a grade_b

Kemudian tekan enter, maka terlihat hasil sebagai berikut:

```
. kap grade_a grade_b

             Expected
Agreement    Agreement      Kappa    Std. Err.         Z       Prob>Z
------------------------------------------------------------------------
   87.50%       59.00%     0.6951      0.1116       6.23       0.0000
```

Berdasarkan hasil di atas dapat dibuktikan bahwa % agreement dan koefesien kappa meningkat dengan menyederhanakan kategori. Pada hasil ini reliabilitas grading kanker menggunakan 2 kategori tergolong baik.

6.7 Aplikasi Stata untuk menghitung ukuran reliabilitas test berskala data kontinyu

Dalam latihan aplikasi Stata untuk menilai reliabilitas test berskala data kontinyu akan dihitung *concordance correlation coefficient (rho_c), average difference dan 95% limit of agreements*. Untuk itu bukalah terlebih dahulu data Reliabilitas2.dta. Data ini adalah data penelitian tentang pengukuran panjang, lebar, tinggi dan volume tumor melalui pemeriksaan MRI dengan kontras oleh 2 orang radiolog. Data ini telah mengalami perubahan untuk keperluan pelatihan sehingga sudah tidak mencerminkan data aslinya. Hasil pengukuran dalam satuan centi meter (cm) kemudian dibandingkan reliabilitas hasil pengukuran antar radiolog tersebut.

Tugas: Nilai lah reliabilitas pengukuran panjang, lebar, tinggi dan volume tumor melalui MRI dengan kontras oleh 2 orang radiolog tersbut

Setelah data berhasil dibuka maka untuk menilai reliabilitas pengukuran panjang tumor ketiklah perintah sebagai berikut:

concord r1_panjang r2_panjang

Kemudian tekan enter maka akan terlihat hasil sebagai berikut:

```
. concord r1_panjang r2_panjang

Concordance correlation coefficient (Lin, 1989, 2000):

 rho_c   SE(rho_c)    Obs     [   95% CI   ]      P        CI type
------------------------------------------------------------------
 0.980      0.007       22      0.966  0.994    0.000    asymptotic
                            0.959  0.990    0.000   z-transform

Pearson's r =   0.987   Pr(r = 0) = 0.000   C_b = rho_c/r =   0.993
Reduced major axis:    Slope =      1.118    Intercept =     -0.309

Difference = r1_panjang - r2_panjang

        Difference                    95% Limits Of Agreement
   Average     Std Dev.                (Bland & Altman, 1986)
------------------------------------------------------------------
    -0.041       0.211                   -0.455      0.373

Correlation between difference and mean = 0.567

Bradley-Blackwood F = 5.319 (P = 0.01405)
```

Untuk menampilkan rentang hasil pengukuran panjang oleh kedua radiolog ketiklah perintah sebagai berikut:

sum r1_panjang r2_panjang

kemudian tekan enter maka terlihat hasil sebagai berikut:

```
. sum r1_panjang r2_panjang
```

Variable	Obs	Mean	Std. Dev.	Min	Max
r1_panjang	22	2.239091	1.133922	1.01	4.3
r2_panjang	22	2.28	1.014589	1.15	4.01

Kemudian untuk menilai reliabilitas pengukuran lebar tumor ketiklah perintah sebagai berikut:

concord r1_lebar r2_lebar

Kemudian tekan enter maka akan terlihat hasil sebagai berikut:

```
. concord r1_lebar r2_lebar

Concordance correlation coefficient (Lin, 1989, 2000):

 rho_c   SE(rho_c)    Obs     [   95% CI   ]      P        CI type
------------------------------------------------------------------
 0.999      0.001       22      0.998  1.000    0.000    asymptotic
                            0.997  0.999    0.000   z-transform

Pearson's r =   0.999   Pr(r = 0) = 0.000   C_b = rho_c/r =   1.000
Reduced major axis:    Slope =      1.005    Intercept =     -0.025

Difference = r1_lebar - r2_lebar

        Difference                    95% Limits Of Agreement
   Average     Std Dev.                (Bland & Altman, 1986)
------------------------------------------------------------------
    -0.012       0.077                   -0.162       0.138

Correlation between difference and mean = 0.098

Bradley-Blackwood F = 0.350 (P = 0.70925)
```

Untuk menampilkan rentang hasil pengukuran lebar tumor oleh kedua radiolog ketiklah perintah sebagai berikut:

sum r1_lebar r2_lebar

kemudian tekan enter maka terlihat hasil sebagai berikut:

```
. sum r1_lebar r2_lebar

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    r1_lebar |        22    2.675455    1.575191      1.12       5.75
    r2_lebar |        22    2.687273    1.567668      1.22       5.85
```

Selanjutnya untuk menilai reliabilitas pengukuran tinggi tumor ketiklah perintah sebagai berikut:

concord r1_tinggi r2_tinggi

Kemudian tekan enter maka akan terlihat hasil sebagai berikut:

```
. concord r1_tinggi r2_tinggi

Concordance correlation coefficient (Lin, 1989, 2000):

 rho_c   SE(rho_c)    Obs     [    95% CI    ]      P        CI type
------------------------------------------------------------------
 0.950      0.021       22      0.908   0.992    0.000    asymptotic
                            0.886   0.979    0.000   z-transform

Pearson's r =   0.955   Pr(r = 0) = 0.000   C_b = rho_c/r =   0.995
Reduced major axis:    Slope =      1.048    Intercept =     -0.200

Difference = r1_tinggi - r2_tinggi

        Difference                    95% Limits Of Agreement
   Average     Std Dev.                (Bland & Altman, 1986)
------------------------------------------------------------------
    -0.092       0.336                   -0.751      0.567

Correlation between difference and mean = 0.154

Bradley-Blackwood F = 1.045 (P = 0.37005)
```

Untuk menampilkan rentang hasil pengukuran tinggi tumor oleh kedua radiolog ketiklah perintah sebagai berikut:

sum r1_tinggi r2_tinggi

kemudian tekan enter maka terlihat hasil sebagai berikut:

```
. sum r1_tinggi r2_tinggi

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   r1_tinggi |         22    2.167273    1.128337          1       4.08
   r2_tinggi |         22    2.259091    1.076993          1       4.05
```

Dan untuk menilai reliabilitas pengukuran volume tumor ketiklah perintah sebagai berikut:

concord r1_vol r2_vol

Kemudian tekan enter maka akan terlihat hasil sebagai berikut:

```
. concord r1_vol r2_vol

Concordance correlation coefficient (Lin, 1989, 2000):

 rho_c   SE(rho_c)    Obs     [   95% CI   ]     P        CI type
------------------------------------------------------------------
 0.998     0.001       22     0.997  1.000    0.000   asymptotic
                          0.997  0.999    0.000   z-transform

Pearson's r =   0.999   Pr(r = 0) = 0.000   C_b = rho_c/r =   0.999
Reduced major axis:   Slope =     1.035    Intercept =    -0.600

Difference = r1_vol - r2_vol

        Difference                  95% Limits Of Agreement
   Average     Std Dev.               (Bland & Altman, 1986)
------------------------------------------------------------------
     0.235       1.730                  -3.156      3.626

Correlation between difference and mean = 0.612

Bradley-Blackwood F = 6.290 (P = 0.00760)
```

Untuk menampilkan rentang hasil pengukuran volume oleh kedua radiolog ketiklah perintah sebagai berikut:

sum r1_vol r2_vol

kemudian tekan enter maka terlihat hasil sebagai berikut:

```
. sum r1_vol r2_vol
```

Variable	Obs	Mean	Std. Dev.	Min	Max
r1_vol	22	23.84046	30.98134	1.652	85.30125
r2_vol	22	23.60579	29.92309	2.27286	82.10475

6.8 Interpretasi hasil analisis penilaian reliabilitas

Berdasarkan beberapa hasil analisis diatas maka dapat dibuat interpretasi sebagai berikut:

Penilaian reliabilitas test berskala data kategorikal

1. Reliabilitas penilaian stadium tumor dengan 5 kategori tergolong cukup dengan nilai koefesien kappa hanya sebesar 0,361
2. Reliabilitas penilaian grading tumor dengan 2 kategori tergolong baik dengan nilai koefesien kappa hanya sebesar 0,695

Penilaian reliabilitas test berskala data kontinyu

3. Ada kesesuaian yang tinggi hasil pengukuran panjang, lebar, tinggi dan volume tumor menggunakan MRI dengan kontras antara radiolog 1 dengan radiolog 2 yang ditunukkan dengan nilai rho_c yang tinggi (mendekati 1).
4. Ditemukan adanya perbedaan hasil pengukuran masing masing dimensi tumor walaupun demikian rentang perbedaannya masih jauh lebih kecil dibandingkan rentang pengukurannya.
5. Reliabilitas hasil pengukuran panjang, lebar, tinggi dan volume tumor menggunakan MRI dengan kontras tergolong baik.

Ringkasan Perintah (Command) Stata Untuk Analisis Data Penelitian Uji Diagnostik dan Skrining

Menghitung sensitifitas, spesifisitas, akurasi, LR+, LR-, NPP dan NPN suatu test terhadap baku emas:

diagt varbakuemas vartest

Membandingkan kemampuan dua test:

roccomp varbakuemas vartest1 vartest2

Melakukan analisis ROC dan menentukan titik potong (test dengan skala data kontinyu):

roctab varbakuemas vartest, detail graph summary

Membandingkan 2 atau lebih *ROC curve*:

roccomp varbakuemas vartest1 vartest2, graph summary

Membuat variabel baru test kombinasi secara serial:

gen newvar = vartest1*vartest2

Membuat variabel baru test kombinasi secara pararel:

gen newvar = vartest1+vartest2

recode newvar 2=1

Menghitung *percent agreement* dan koefesien kappa (menilai reliabilitas test berskala kategorikal):

kap varobs_a varobs_b

Menghitung *concordance correlation coefficient (rho_c), average difference dan 95% limit of agreements*(menilai reliabilitas test berskala kontinyu):

concord varobs_a varobs_b

Melakukan tabulasi silang

tab var1 var2

Sumary statistik (mean, standar deviasi, nilai minimal dan maksimal

sum var1 var2 var3

Daftar Pustaka

1. Leon Gordis, Epidemiology, Fifth Edition, Elsevier Saunders, 2014,Canada.
2. Raymond S Greenberg, Stephen R Daniels, W. Dana Flanders, John William Eley and John R. Boring, Medical Epidemiology, Third Edition, Lange Medical Book, 2001, New York.
3. Bland, J. M., and D. G. Altman. 1986. Statistical methods for assessingagreement between two methods of clinical measurement. Lancet I:307-310.
4. M. Sopiyudin Dahlan, Penelitian Diagnostik: Dasar-dasar Teoritis dan Aplikasi dengan Program SPSS dan Stata, Salemba Medika, 2009, Jakarta.