**Praktikum ADK pertemuan ke-3 dan 4**

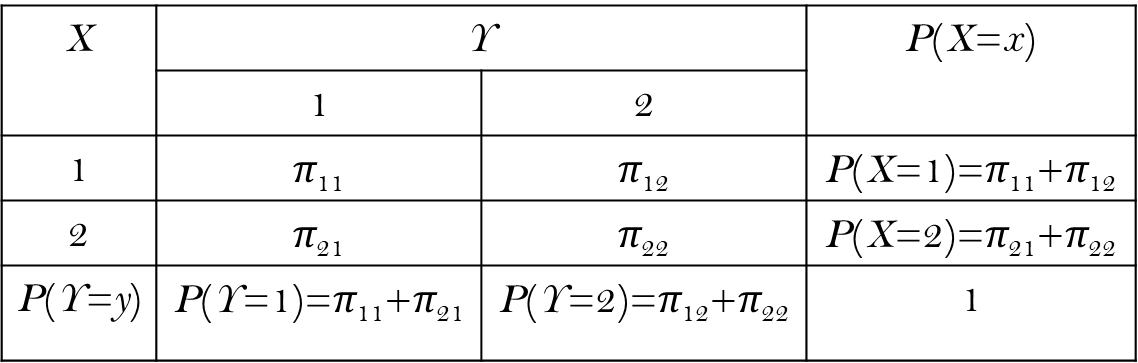
**Association between two categorical variables**

Pokok bahasan:

1. Struktur peluang tabel kontingensi
2. Sensitifitas dan spesifisitas
3. Uji beda dua proporsi
4. Relative risk
5. Rasio odds
6. Uji khi kuadrat untuk kebebasan antar peubah
7. Fisher Exact Test

**STRUKTUR PELUANG TABEL KONTINGENSI**

Tabel kontingensi memuat segugus pengamatan acak berukuran *n*, dimana pada masing-masing pengamatan dicatat dua peubah katagorik *X*(1,2) dan *Y*(1,2). Tabel kontingensi 2×2 menggambarkan **sebaran peluang bersama** {*π*11, *π*12, *π*21, *π*22}.



* Peluang *Y* bersyarat *X* ditulis *P*(*Y*=*y*|*X*=*x*) = *πy*|*x*



* Jika *nij* adalah banyaknya pengamatan yang terdapat pada baris-*i* dan kolom-*j*, penduga kemungkinan maksimum bagi peluang bersama *πij* adalah

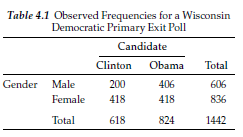


* Penduga bagi peluang bersyarat *πj|i*



Contoh 1

Misal terdapat data jenis kelamin dan pilihan kandidat presiden sebagai berikut:



Tentukan peluang bagi:

1. Yang memilih Clinton
2. Yang memilih Obama
3. Yang memilih Clinton jika ybs laki-laki
4. Yang memilih Clinton jika ybs perempuan
5. Yang memilih Obama jika ybs laki-laki
6. Yang memilih Obama jika ybs perempuan

Jawab:

1. Yang memilih Clinton=
2. Yang memilih Obama=
3. Yang memilih Clinton jika ybs laki-laki=
4. Yang memilih Clinton jika ybs perempuan=
5. Yang memilih Obama jika ybs laki-laki=
6. Yang memilih Obama jika ybs perempuan i=

Sintaks R:

#----INPUT DATA----#

pilpres<-matrix(c(200,406,418,418), nrow=2,byrow=TRUE)

colnames(pilpres)<-c("clinton","obama")

rownames(pilpres)<-c("M","F")

tabelpilpres<-as.table(pilpres)

tabelpilpres

datapilpres<-as.data.frame(pilpres)

datapilpres

#----jawab contoh 1---#

addmargins(tabelpilpres)

#a

a<-618/1442

a

#b

b<-824/1442

b

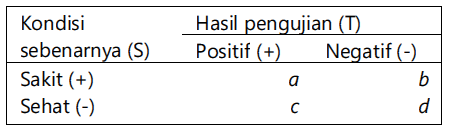
prop.table(tabelpilpres)

#c-f

prop.table(tabelpilpres,margin=1)

**SENSITIVITAS DAN SPESIFITAS**

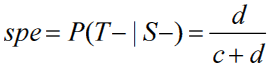
Sensitivitas dan spesifisitas merupakan salah satu alat dalam diagnosa. Awalnya, kedua statistik ini digunakan untuk melakukan diagnosa kesehatan, namun pada perkembangannya juga digunakan dalam diagnosa model-model statistika. Perhatikan tabel berikut :



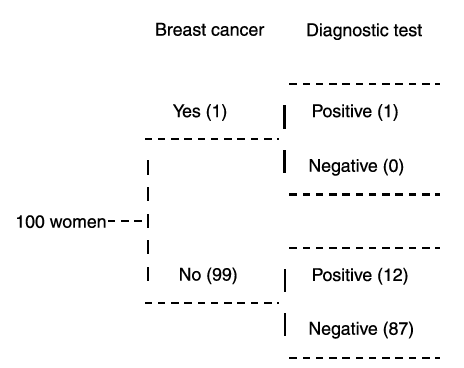
* **Sensitivitas**: peluang bahwa hasil pengujian menunjukkan bahwa seseorang positif terjangkit penyakit apabila faktanya orang tersebut memang terjangkit penyakit

****

* **Spesifisitas**: peluang bahwa hasil pengujian menunjukkan bahwa seseorang tidak terjangkit penyakit apabila faktanya orang tersebut memang tidak terjangkit penyakit



Contoh 2



Berapa nilai sensitifitas dan spesifisitas dari kasus di atas?

Jawab:

|  |  |  |  |
| --- | --- | --- | --- |
| kondisi sebenarnya | Hasil pengujian |  | total |
| positif(+) | negatif(-) |
| sakit(+) | 1 | 0 | 1 |
| sehat (-) | 12 | 87 | 99 |
| total | 13 | 87 | 100 |

Sensitivitas=

spesifisitas

sintaks R

#----INPUT DATA SAKIT----#

sakit<-matrix(c(1,12,0,87), nrow=2,byrow=TRUE)

predicted<-rownames(sakit)<-c("pos","neg")

actual<-colnames(sakit)<-c("sakit","sehat")

tabelsakit<-as.table(sakit)

tabelsakit

#---jawab contoh 2---#

install.packages("epiR")

library("epiR")

epi.tests(tabelsakit,conf.level=0.95)

**KEBEBASAN PADA TABEL KONTINGENSI**

Dua peubah (X,Y), dalam tabel kontingensi dikatakan saling bebas secara statistika apabila distribusi peluang bersyarat dari Y adalah identic untuk setiap level X. Jika kedua peubah merupakan peubah respon, maka dua peubah dinyatakan saling bebas apabila semua peluang bersama sama dengan perkalian dari peluang-peluang marginalnya. Ditulis:

untuk i=1,2,…,I dan j=1,2,…,J

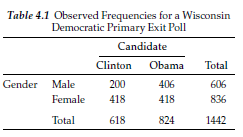
* **RESIKO RELATIF (RELATIVE RISK)**
* Resiko relatif adalah nisbah peluang sukses baris pertama pada peluang sukses baris kedua:



* Resiko relatif bernilai satu, *r*=1 menunjukkan kebebasan antara peubah baris dengan peubah kolom.
* Pada keadaan tertentu resiko relatif lebih bermakna untuk pembandingan peluang sukses. Selisih peluang 0.610−0.601 dan selisih peluang 0.010−0.001 sebesar 0.009 (meskipun dengan menunjukkan hasil uji yang berbeda), tetapi nisbah peluang 0.610/0.601 dan 0.010/0.001 adalah sangat berbeda, masing-masing adalah 1.01 dan 10.
* Inferensia untuk resiko relatif tidak sederhana

Contoh 3

Misal terdapat data jenis kelamin dan pilihan kandidat presiden sebagai berikut:



Tentukan nilai resiko relative antara pria dan wanita dari kasus tersebut. Berdasarkan hasil tersebut apakah terdapat hubungan antara jenis kelamin dan pilihan kandidat presiden?

Jawab:

P(Clinton|Male)=0.330033

P(Clinton|female)=0.5

RR=0.330033/0.5=0.660066

Artinya apabila diketahui seseorang adalah laki-laki, maka kecenderungan untuk memilih Clinton adalah 0.66 kali dari kecenderungan wanita memilih Clinton, dengan kata lain wanita lebih cenderung memilih Clinton dibandingkan laki-laki.

Berdasarkan nilai resiko relative yang diperoleh terdapat hubungan antara jenis kelamin dengan pilihan kandidat presiden, hal ini karena nilai dari resiko relatif tidak sama dengan satu.

Sintaks R:

prop.out <- prop.table(tabelpilpres, margin = 1)

# relative risk of male vs. female

prop.out[1,1]/prop.out[2,1]

**ODDS DAN RASIO ODDS**

* Pada suatu tabel 2×2, dengan odds pada baris-1 dan baris-2 masing-masing *odds*1= *π*1/(1− *π*1)dan *odds*2= *π*2/(1− *π*2), nisbah odds baris-1 atas odds baris-2 adalah:



* Penduga kemungkinan maksimum bagi rasio odds adalah:



* Sifat-sifat rasio odds

Pada tabel kontingensi 2×2 dengan *π*1 dan *π*2 masing-masing adalah peluang sukses pada baris-1 dan baris-2,

1. Keadaan *π*1=*π*2 menyebabkan ***odds***1=***odds***2 sehingga *θ*=1, menunjukkan kebebasan antara peubah baris dan peubah kolom. *θ*>1 menunjukkan *π*1>*π*2, dan . *θ*<1 menunjukkan *π*1<*π*2.
2. Nilai *θ* sekecil-kecilnya adalah 0 dan sebesar-besarnya adalah *infinite*, semakin jauh *θ* dari 1 semakin kuat keterkaitan antara peubah kolom pada peubah baris. Dua nilai rasio odds *θ*1dan*θ*2 menunjukkan kekuatan keterikatan yang sama besar apabila *θ*1=1/*θ*2
3. Pertukaran posisi baris atau kolom tidak menyebabkan gambaran kekuatan keterikatan baris dan kolom berubah; dengan pertukaran ini akan diperoleh *θ* baru yang nilainya sebesar 1/*θ* .
4. Nilai nisbah odds tidak perubah apabila tabel ditranspose sehingga posisi baris dan kolomnya dipertukarkan.

* Inferensi rasio odds

1. Transformasi logaritma (logaritma berbasis *e*) atas nisbah odds, dapat diperoleh sebaran yang simetris mendekati sebaran Normal; Nilai *θ* = 1 bersesuaian dengan ln*θ* = 0, menunjukkan kebebasan; Jika *θ*1=1/*θ*2 maka ln*θ*1=−ln*θ*2, menunjukkan kekuatan keterikatan yang sama tetapi berbeda arah.
2. Simpangan baku bagi statistik logaritma nisbah odds adalah:



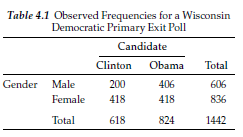
1. Selang kepercayaan (1−*α*)100% bagi ln*θ* adalah:



1. Batas bawah dan batas atas selang kepercayaan bagi *θ*  masing-masing diperoleh sebagai eksponen batas bawah dan eksponen batas atas selang kepercayaan bagi ln*θ*.

Contoh 4:

Misal terdapat data jenis kelamin dan pilihan kandidat presiden sebagai berikut:



1. Tentukan odds laki-laki
2. Tentukan odds perempuan
3. Hitung rasio odds berdasarkan poin a dan b, interpretasikan nilai rasio odds tersebut
4. Tentukan selang kepercayaan dari rasio odds yang telah Anda peroleh pada butir a
5. Berdasarkan nilai rasio odds tersebut apakah terdapat hubungan antara jenis kelamin dengan pilihan presiden?

Jawab:

1. P(C|M)=0.330033; P(O|M)=0.669967; odds(M)=0.330033/0.669967=0.492611, nilai odds ini berarti peluang laki-laki memilih Clinton 0.499 kali dari peluang laki-laki memilih Obama.
2. P(C|F)=0.5; P(O|F)=0.5; odds(F)=0.5/0.5=1, nilai odds ini berarti, peluangg wanita untuk memilih Clinton sama dengan peluang wanita untuk memilih Obama.
3. OR=0.492611/1=0.492611, artinya odds laki-laki 0.49 kali dari odds wanita.
4. S=(1/200+1/406+1/418+1/418)=0.012248
5. Berdasarkan nilai dari rasio odds terdapat hubungan antara jenis kelamin dengan pilihan presiden, karena nilai 1 (satu) tidak masuk pada selang kepercayaan .

Sintaks R:

#---jawab contoh 4---#

# odds of Male

prop.out[1,1]/prop.out[1,2]

# odds of Female

prop.out[2,1]/prop.out[2,2]

install.packages("epitools")

library("epitools")

or.out <- oddsratio(tabelpilpres, rev="b")

or.out$measure

**HUBUNGAN ANTARA RASIO ODDS DAN RESIKO RELATIF**

Perhatikan bahwa rasio odds adalah hasil kali resiko relatif dengan (1−*π*2)/(1−*π*1). Dalam keadaan *π*1 dan *π*2 yang sangat kecil, maka (1−*π*2)/(1−*π*1) π ≈ 1.



Dalam keadaan *π*1 dan *π*2 yang sangat kecil, rasio odds praktis sama dengan, atau dapat diinterpratesikan sebagai resiko relitif.

**UJI KHI KUADRAT UNTUK KEKBEBASAN ANTAR PEUBAH**

Uji kebebasan yang biasa digunakan adalah uji Khi-Kuadrat (Chi-Square Test for Independence).

Hipotesis:

: Dua peubah saling bebas

: Dua peubah tidak saling bebas

Statistik uji:

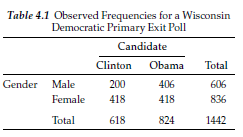
Statistic ini akan mendekati sebaran dengan derajat bebas (r-1)(c-1) dalam hal ini r adalah banyaknya baris (row) dan c adalah banyaknya kolom (column) pada tabel kontingensi.

Sedangkan d=frekuensi harapan dapat dihitung dengan formula:

Hipotesis nol akan ditolak jika nilai lebih besar dari dengan b=derajat bebas (r-1)(c-1).

Contoh 5:

Misal terdapat data jenis kelamin dan pilihan kandidat presiden sebagai berikut:



Ujilah menggunakan uji khi kuadrat, apakah peubah jenis kelamin dan pilihan presiden saling bebas atau tidak?

Jawab:

Hipotesis:

: jenis kelamin dan pilihan presiden saling bebas

: jenis kelamin dan pilihan presiden tidak saling bebas

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| observation | |  |  |  | expectation | |  |
| Gender | Candidate | |  |  | Gender | Candidate | |
| Clinton | Obama | total |  | Clinton | Obama |
| Male | 200 | 406 | 606 |  | Male | 259.7143 | 346.2857 |
| Female | 418 | 418 | 836 |  | Female | 358.2857 | 477.7143 |
| total | 618 | 824 | 1442 |  |  |  |  |

perhitungan nilai harapan:

Statistik uji:

Wilayah kritis:

Tolak jika

Keputusan: tolak artinya jenis kelamin dan pilihan presiden tidak saling bebas.

Sintaks R:

#----jawab contoh 5----#

chisq.test(tabelpilpres,correct=FALSE)

#fisher.test(tabelpilpres)

#install.packages("DescTools")

#library("DescTools")

#GTest(tabelpilpres)

**FISHER EXACT TEST**

Selang kepercayaan dan pengujian yang dilakukan sejauh ini digunakan untuk contoh berukuran besar. Semakin besar ukuran contoh maka dan akan menghampiri sebaran khi kuadrat. Akan tetapi, jika ukuran contoh kecul, inferensia menggunakan sebaran exact lebih tepat diabndingkan dengan hampiran contoh besar. Pada tabel 2x2, kebebasan dua peubah ditandai dengan nilai rasio odds sama dengan satu. Pada tabel ini untuk jumlah baris dan kolom marginal tertentu, frekuensi pada sel pertama () menentukan frekuensi pada ketiga sel lainnya. ketika nilai rasio odds sama dengan satu, peluang untuk nilai dinyatakan oleh:

Yang merupakan peluang hipergeometrik.

Hipotesis:

: dua peubah saling bebas

: dua peubah tidak saling bebas

Tolak jika nilai-p yang merupakan peluang hipergeometri sebelah kanan bahwa lebih besar atau sama dengan frekuensi teramati lebih besar dari taraf nyata yang ditentukan.

Contoh 6:

Misal terdapat data jenis kelamin dan pilihan kandidat presiden sebagai berikut:

|  |  |  |  |
| --- | --- | --- | --- |
| Gender | Candidate | |  |
| Clinton | Obama | total |
| Male | 3 | 1 | 4 |
| Female | 1 | 3 | 4 |
| total | 4 | 4 | 8 |

Ujilah menggunakan uji fisher, apakah peubah jenis kelamin dan pilihan presiden saling bebas atau tidak?

Jawab:

Fisher exact test

Hipotesis:

: jenis kelamin dan pilihan presiden saling bebas

: jenis kelamin dan pilihan presiden tidak saling bebas

Statistic uji

Wilayah kritis:

Tolak jika

Keputusan: tidak tolak , artinya jenis kelamin dan pilihan presiden saling bebas

Sintaks R:

#----jawab contoh 6----#

pilpres2<-matrix(c(3,1,1,3), nrow=2,byrow=TRUE)

colnames(pilpres2)<-c("clinton","obama")

rownames(pilpres2)<-c("M","F")

tabelpilpres2<-as.table(pilpres2)

tabelpilpres2

fisher.test(tabelpilpres2)