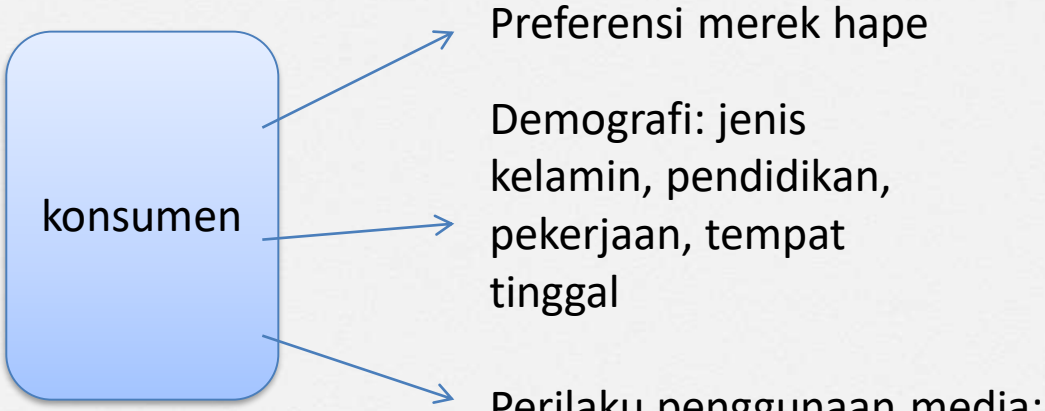
STK351
Pengantar Analisis Data Kategorik

Farit Mochamad Afendi
08128592194 – fmafendi@apps.ipb.ac.id

Tabulasi

Suatu survei dilakukan untuk mendapatkan gambaran preferensi terhadap merek hape tertentu. Dari survei ini diharapkan diperoleh gambaran profil konsumen yang cenderung memilih merek hape tertentu. Informasi ini akan sangat bermanfaat untuk penajaman *marketing campaign*, desain produk, dsb.



Tabulasi

merek Hape	Frek
A	85
B	115
C	200

Jenis Kelamin	Frek
Pria	220
Wanita	180

Jenis Kelamin	merek Hape			Total
	A	B	C	
Pria	35	25	160	220
Wanita	50	90	40	180
Total	85	115	200	400

Tabel kontingensi (Pearson)
Tabulasi silang

Struktur peluang

Jenis Kelamin	merek Hape			Total
	A	B	C	
Pria	35	25	160	220
Wanita	50	90	40	180
Total	85	115	200	400

n_{ij} n_{i+} n_{+j} n_{++}

Sebaran bersama $\pi_{ij} = n_{ij} / n$

Sebaran marjinal $\pi_{i+} = n_{i+} / n$

$\pi_{+j} = n_{+j} / n$

n acak → Y_{ij} ~ Poisson(μ_{ij})

n tetap → Y_{ij} ~ Multinomial(n,π_{ij}); $\pi_{ij} = \mu_{ij}/n$

n_i tetap → Y_{ij} ~ Multinomial(n_i,$\pi_{j|i}$); $\pi_{ij} = \mu_{ij}/\mu_i$

n_i dan n_j tetap → hipergeometrik

Struktur peluang

Row	Column 1	Column 2	Total
1	π_{11}	π_{12}	π_{1+}
	$(\pi_{1\|1})$	$(\pi_{2\|1})$	(1.0)
2	π_{21}	π_{22}	π_{2+}
	$(\pi_{1\|2})$	$(\pi_{2\|2})$	(1.0)
Total	π_{+1}	π_{+2}	1.0

Saling bebas: $\pi_{j|i} = \pi_{ij}/\pi_{i+} = (\pi_{i+}\,\pi_{+j})/\pi_{i+} = \pi_{+j}$

Tipe studi

- *Retrospective*: kasus diamati di masa kini, ditelusuri peristiwa yang terjadi di masa lalu
  - *Case control*
- *Prospective*: kondisi diamati sekarang untuk diamati dampaknya di masa depan
  - *Clinical trial* (alokasi perlakuan acak)
    - *cohort study* (alokasi perlakuan sukarela)
  - *Cross sectional study* (contoh dipilih untuk diamati perlakuan dan respon sekaligus)

Tipe studi

- *Restrospective study*
  - mengendalikan n+j,
  - menganggap frekuensi I sebagai contoh dari sebaran multinomial
- *Prospective study*
  - mengendalikan ni+,
  - menganggap frekuensi J sebagai contoh dari sebaran multinomial
- *Cross sectional study*
  - mengendalikan n,
  - menganggap frekuensi IJ sebagai contoh dari sebaran multinomial

Tipe studi

- Observational study
  - Case control
  - Cohort
  - Cross sectional
- Experimental study
  - Clinical trial

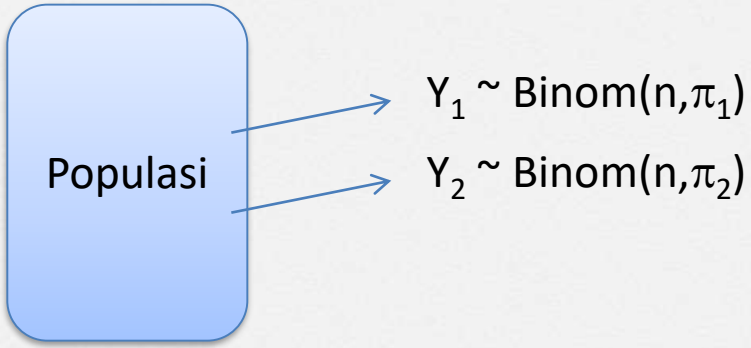
Kanker	Hasil diagnosa		Total
	Positif	Negatif	
Ya	85	15	100
Tidak	50	150	200
Total	135	165	300

Kepekaan (*sensitivity*) → kemampuan mendeteksi yang sakit

85/100 = 0.85

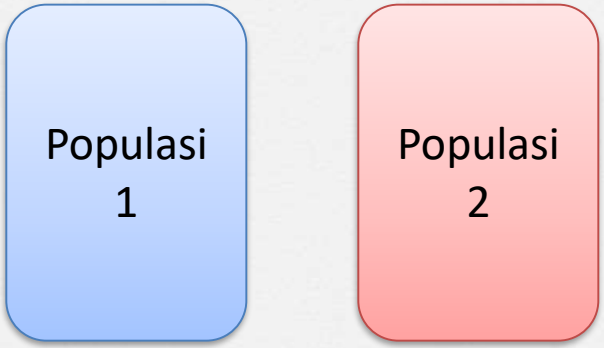
Kekhususan (*specificity*) → kemampuan mendeteksi yang tidak sakit

150/200 = 0.75



Pengujian asosiasi: $\pi_{ij} = \pi_i \pi_j$

$E_{ij} = \pi_i \pi_j n$
$= (n_i/n)(n_j/n) n$
$= (n_i n_j/n)$



$Y \sim Binom(n,\pi_1)$ $Y \sim Binom(n,\pi_2)$

Pengujian kehomogenan: $\pi_1 = \pi_2 = \pi$

$E_{ij} = n_i \pi$
$= n_i n_j/n$

Uji Khi Kuadrat

$$\chi^2 = \sum_{i=1}^{b}\sum_{j=1}^{k}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} \sim \chi^2_{(db=bk-1)}$$

$$p = P\left(\chi^2 > \chi^2_{(db=bk-1)}\right)$$

O_{ij} = besarnya frekuensi teramati
E_{ij} = besarnya frekuensi di bawah H_0
b = banyaknya baris
k = banyaknya kolom

Jenis Kelamin	merek Hape			Total
	A	B	C	
Pria	35 (46.75)	25 (63.25)	160 (110)	220
Wanita	50 (38.25)	90 (51.75)	40 (90)	180
Total	85	115	200	400

H0: Tidak ada asosiasi antara jenis kelamin dan preferensi merek HP
H1: Ada asosiasi antara jenis kelamin dan preferensi merek HP

E11 = (220 . 85)/400 = 46.75
E21 = (180 . 85)/400 = 38.25

$$\chi^2 = \frac{(35-46.75)^2}{46.75} + \ldots + \frac{(40-90)^2}{90} = 108.471$$

$$p = P(\chi^2_{db=4} > 108.471) = 0.000$$

Tolak H0 → ada asosiasi antara keduanya

Beda Proporsi

During the early 1950s, polio rates in the U.S. were above 25,000 annually; in 1952 and 1953, the U.S. experienced an outbreak of 58,000 and 35,000 polio cases, respectively, up from a typical number of some 20,000 a year, with deaths in those years numbering 3,200 and 1,400.

"Polio pioneers"—some of the many children who took part in trials of poliomyelitis vaccine

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1114166/

The first effective polio vaccine was developed in 1952 by Jonas Salk and a team at the University of Pittsburgh that included Julius Youngner, Byron Bennett, L. James Lewis, and Lorraine Friedman, which required years of subsequent testing.

Beda Proporsi

$$H_0 : \pi_1 = \pi_2$$
$$H_0 : \pi_1 \neq \pi_2$$

- Populasi 1: mendapat vaksin
- Populasi 2: mendapat plasebo
- n_1 dan n_2: banyaknya contoh dari populasi 1 dan 2
- x_1 dan x_2: banyaknya kasus dari contoh populasi 1 dan 2
- π_1 dan π_2: proporsi populasi 1 dan 2
- p_1 dan p_2: proporsi contoh populasi 1 dan 2

$$z = \frac{(p_1 - p_2)}{SE}$$

$$p_1 = \frac{x_1}{n_1} \qquad p_2 = \frac{x_2}{n_2}$$

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Beda Proporsi

Double blind experiment

Study group	Study population	Total Polio Case
Vaccinated	200745	57
Placebo	201229	142


```
Test and CI for Two Proportions

Sample    X       N  Sample p
1        57  200745  0.000284
2       142  201229  0.000706

Difference = p (1) - p (2)
Estimate for difference:  -0.000421721
95% CI for difference:  (-0.000559175, -0.000284267)
Test for difference = 0 (vs not = 0):  Z = -6.01  P-Value = 0.000

Fisher's exact test: P-Value = 0.000
```

Perbandingan proporsi

Jenis kelamin	merek hape		Total
	A	B	
Pria	35	185	220
Wanita	50	130	180
Total	85	315	400

Resiko relatif

P(A|Pria) = 35/220 = 0.16

P(A|Wanita) = 50/180 = 0.28

RR = 0.16/0.28 = 0.57

Rasio Odds

P(A|Pria) = 35/220 = 0.16

P(B|Pria) = 185/220 = 0.84

Odds pria = 0.16/0.84 = 0.19

P(A|Wanita) = 50/180 = 0.28

P(B|Wanita) = 130/180 = 0.72

Odds wanita = 0.28/0.72 = 0.38

Rasio odds = 0.19/0.38 = 0.49