HUBUNGAN ANTAR PEUBAH

DALAM ANALISIS INGIN DIKETAHUI ATAU DIEVALUASI HUBUNGAN ATAU KETERKAITAN ANTAR PEUBAH

Hubungan Antar Peubah

Besarnya gaji

Lama bekerja

Hubungan Antar Peubah (lanjutan)

Pendapatan yang diterima

Kepuasan Kerja

Hubungan Antar Peubah (lanjutan)

Hubungan antara keputusan pembelian suatu produk tertentu dikaitkan dengan jenis kelamin

Hubungan Antar Peubah (lanjutan)

Hubungan antara keputusan pembelian suatu produk tertentu dikaitkan dengan tingkat pendapatan konsumen

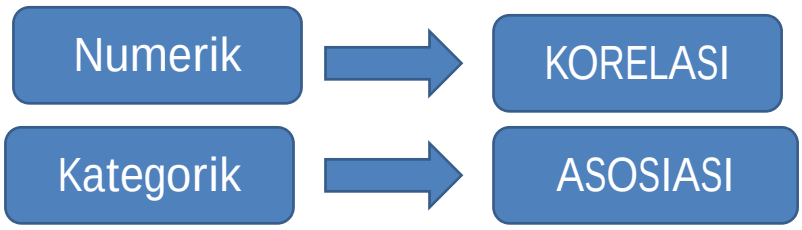
Hubungan Antar Peubah (lanjutan)

Hubungan antara status kredit nasabah (lancar atau macet)

dengan

status rumah (sendiri atau kontrak) dan lokasi tinggal (desa atau kota)

Alat Analisis yang Digunakan



Relationship	Numerik	Kategorik
Numerik	Korelasi Pearson, Spearman	Tabel Ringkasan
Kategorik	Tabel Ringkasan	Spearman (ordinal), **Chi Square**

No.	Pokok Bahasan	Sub Pokok Bahasan	Perkiraan Waktu (menit)	Daftar Kepustakaan
3.	Tabel Kontingensi 2x2	• Peluang Bersama, Peluang Marjinal, dan Peluang Bersyarat • Kepekaan dan Kekhususan dalam Uji Diagnostik • Kebebasan	1 x (2 x 50')	1: Bab 2.1
4.	Tabel Kontingensi 2x2	• Percontohan Binomial dan Multinomial • Beda Proporsi • Risiko Relatif	1 x (2 x 50')	1: Bab 2.1.5 Bab 2.2
5.	Tabel Kontingensi 2x2	• Rasio Odd • Uji Kebebasan Khi-kuadrat	1 x (2 x 50')	1: Bab 2.3, 2.4
6.	Tabel Kontingensi 2x2	• Uji Kebebasan untuk Data Ordinal • Uji Eksak untuk Ukuran Contoh Kecil	1 x (2 x 50')	1: Bab 2.5, 2.6

Tabulasi Silang

Eksplorasi asosiasi antar peubah biasa diawali dengan tabulasi silang antar kedua peubah

Peubah X	Peubah Y				Total
	Kategori 1	Kategori 2	...	Kategori q	
Kategori 1	O_{11}	O_{12}	...	O_{1q}	B_1
Kategori 2	O_{21}	O_{22}	...	O_{2q}	B_2
...	...	...	...	...	...
Kategori p	O_{p1}	O_{p2}	...	O_{pq}	B_p
Total	K_1	K_2	...	K_q	N

Tabel Kontingensi 2x2

Bagian (1)

Tabulasi Silang 2x2

Peubah X	Peubah Y		Total
	Kategori 1	Kategori 2	
Kategori 1	O_{11}	O_{12}	B_1
Kategori 2	O_{21}	O_{22}	B_2
Total	K_1	K_2	N

Struktur Peluang Tabel Kontingensi

- Peluang **Bersama** $\pi_{ij} = P(X = i, Y = j)$

  dengan $\sum_{i,j} \pi_{ij} = 1.$
- Peluang **Marginal** → total (baris/kolom) peluang bersama

$$\pi_{1+} = \pi_{11} + \pi_{12} \qquad \pi_{+1} = \pi_{11} + \pi_{21}$$

- Peluang **Bersyarat** → peluang Y pada setiap level X

$$\pi_{Y|X}$$

Teladan Tabel 2.2

General Social Survey by their gender and by their belief in an afterlife. Table 2.2 illustrates the cell count notation for these data.

Table 2.2. Notation for Cell Counts in Table 2.1

	Belief in Afterlife		
Gender	Yes	No or Undecided	Total
Females	$n_{11} = 509$	$n_{12} = 116$	$n_{1+} = 625$
Males	$n_{21} = 398$	$n_{22} = 104$	$n_{2+} = 502$
Total	$n_{+1} = 907$	$n_{+2} = 220$	$n = 1127$

Beri contoh manakah peluang bersama, marginal, dan bersyarat!!

Sensitivity and Specificity in Diagnostic Tests

- Diagnostic testing is used to detect many medical conditions.
- For example, the mammogram can detect breast cancer in women, and the prostate-specific antigen (PSA) test can detect prostate cancer in men.
- The result of a diagnostic test is said to be ***positive*** *if it states that the disease is* ***present*** *and* ***negative*** *if it states that the disease* is **absent**.

The accuracy of diagnostic tests

- Given that a subject has the disease, the probability the diagnostic test is positive is called the ***sensitivity***
- Given that the subject does not have the disease, the probability the test is negative is called the *specificity*.

- misalkan *X adalah kondisi sebenarnya,* dengan kategori 1 = sakit, 2 = tidak sakit, and let *Y* = *hasil uji diagnostik*, dengan kategori 1 = positif, 2 = negatif.
- Sensitivity = $P(Y = 1 \mid X = 1)$, *specificity* = $P(Y = 2 \mid X = 2)$
- Semakin tinggi nilai sensitivity danspecificity, semakin baik uji diagnostik tersebut.

the most common form of cancer in women

Mammogram

1% terkena kanker payudara

Typical values reported for mammograms are sensitivity = *0.86 and specificity = 0.88.*

- ***If these are* true**, then given that a mammogram has a positive result, the probability that the woman truly has breast cancer is only 0.07.

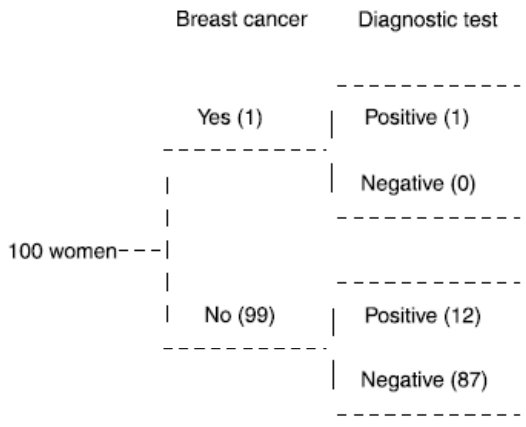


Figure 2.1. Tree diagram showing results of 100 mammograms, when sensitivity = 0.86 and specificity = 0.88.

Independence

- Two variables are said to be *statistically independent if the population conditional* distributions of *Y are identical at each level of X.*
- When two variables are independent, the probability of any particular column outcome *j is the same in each row.*
- *Belief in* an afterlife is independent of gender, for instance, if the actual probability of believing in an afterlife equals 0.80 both for females and for males.

Binomial and Multinomial Sampling

Peubah X	Peubah Y		Total
	Kategori 1	Kategori 2	
Kategori 1	O_{11}	O_{12}	B_1
Kategori 2	O_{21}	O_{22}	B_2
Total	K_1	K_2	N

Random sampling or randomized experiments

Sebaran Binomial atau multinomial

- Tiap baris → kelompok/grup berbeda, **sampel pada baris fix** → sample tersebar pada 2 perlakuan berbeda → sebaran binomial
- Ukuran **sampel total fix** → dibagi ke 4 kategori → sebaran multinomial

Latihan

2.1 An article in the *New York Times* (February 17, 1999) about the PSA blood test for detecting prostate cancer stated that, of men who had this disease, the test fails to detect prostate cancer in 1 in 4 (so called false-negative results), and of men who did not have it, as many as two-thirds receive false-positive results. Let C ($\bar{C}$) denote the event of having (not having) prostate cancer and let $+$ ($-$) denote a positive (negative) test result.

a. Which is true: $P(-|C) = 1/4$ or $P(C|-) = 1/4$? $P(\bar{C}|+) = 2/3$ or $P(+|\bar{C}) = 2/3$?

b. What is the sensitivity of this test?

c. Of men who take the PSA test, suppose $P(C) = 0.01$. Find the cell probabilities in the 2×2 table for the joint distribution that cross classifies Y = diagnosis $(+, -)$ with X = true disease status $(C, \bar{C})$.

d. Using (c), find the marginal distribution for the diagnosis.

e. Using (c) and (d), find $P(C|+)$, and interpret.

2.2 For diagnostic testing, let X = true status (1 = disease, 2 = no disease) and Y = diagnosis (1 = positive, 2 = negative). Let $\pi_i = P(Y = 1|X = i)$, $i = 1, 2$.

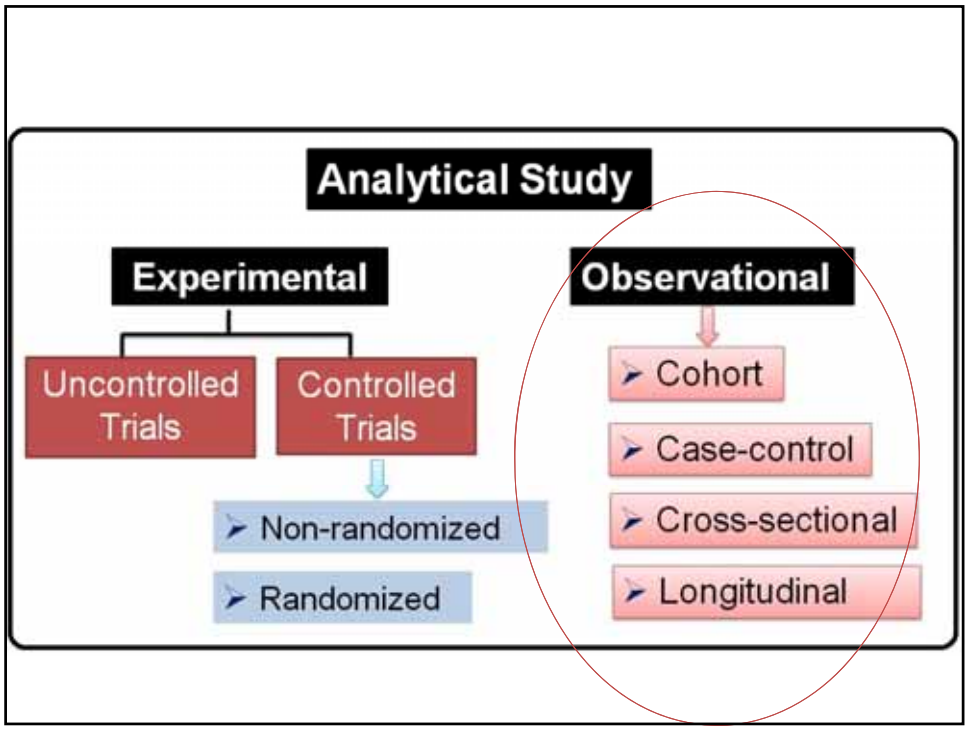
a. Explain why sensitivity $= \pi_1$ and specificity $= 1 - \pi_2$.

b. Let γ denote the probability that a subject has the disease. Given that the diagnosis is positive, use Bayes's theorem to show that the probability a subject truly has the disease is

$$\pi_1\gamma/[\pi_1\gamma + \pi_2(1 - \gamma)]$$

c. For mammograms for detecting breast cancer, suppose $\gamma = 0.01$, sensitivity $= 0.86$, and specificity $= 0.88$. Given a positive test result, find the probability that the woman truly has breast cancer.

d. To better understand the answer in (c), find the joint probabilities for the 2×2 cross classification of X and Y. Discuss their relative sizes in the two cells that refer to a positive test result.



Tugas 1 (Kelompok)

- Kelompok 1&5 →observational study
- Kelompok 2 & 6 →
  - Percontohan Binomial dan Multinomial
  - Beda Proporsi
  - Risiko Relatif
- Kelompok 3 & 7
  - Rasio Odd
  - Uji Kebebasan Khi-kuadrat
- Kelompok 4 & 8
  - Uji Kebebasan untuk Data Ordinal
  - Uji Eksak untuk Ukuran Contoh Kecil