



twitter: @kh_notodiputro



IPB University
— Bogor Indonesia —

Department of Statistics
Study Program in Statistics and Data Science

Pengantar Data Kategorik

Prof. Dr. Ir. Khairil Anwar Notodiputro, MS

email: khairil@apps.ipb.ac.id

**Ketua Program Studi
Statistika dan Sains Data**



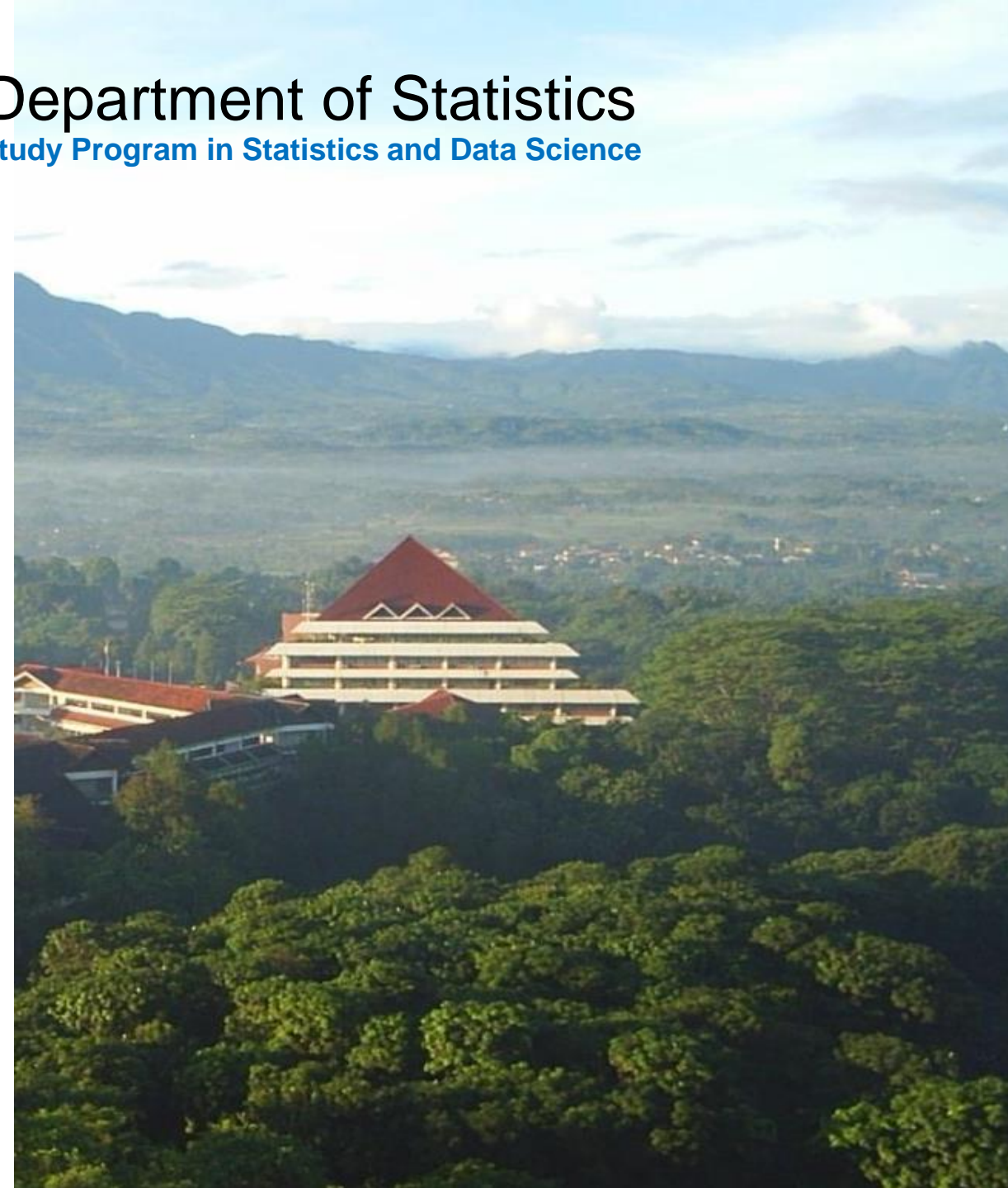


IPB University
— Bogor Indonesia —

Department of Statistics
Study Program in Statistics and Data Science

Outline

- **1. Percobaan Multinom**
- **2. *Goodness of Fit Test***
- **3. *Contingency Tables***
- **4. Ukuran Keeratan**
- **5. Paradoks Simpson**



Data Kategorik

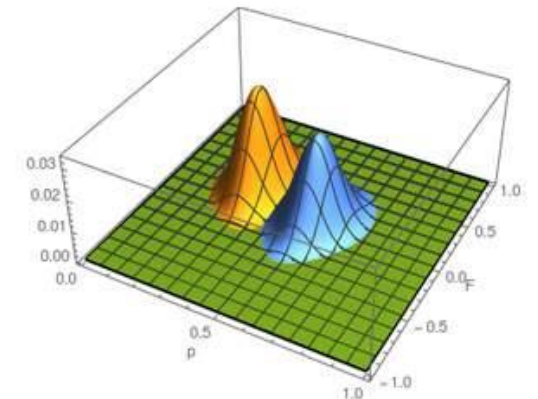


IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

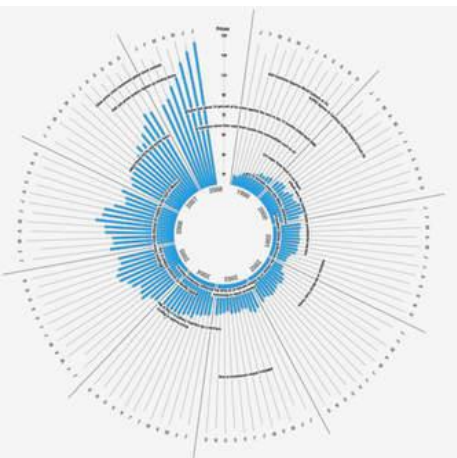
Pengantar

- Sering kali hasil pengamatan yang diperoleh dari survei atau percobaan bersifat kualitatif (klasifikasi), bukan kuantitatif.
- Respon kualitatif biasanya membentuk data yang bersifat cacahan (*count*) seperti jumlah bunga yang berwarna merah, warna putih, atau warna merah muda.
- Data yang diperoleh dari peubah kualitatif ini dikenal sebagai **data kategorik**.
- Kita akan membahas bagaimana analisis statistika sederhana untuk data yang bersifat kategorik.
- Pembahasan mencakup: **percobaan multinom, statistik khi-kuadrat Pearson, uji kecocokan model, tabel kontingensi, perbandingan populasi multinom, dan kesetaraan uji statistika.**



Percobaan multinom (*multinomial experiment*)

- Dalam praktik data kualitatif bisa berasal dari percobaan multinom. Ciri-ciri percobaan multinom adalah:
 - Percobaan itu terdiri atas n buah trial;
 - Setiap trial menghasilkan salah satu dari k buah kategori;
 - Peluang bhw hasilnya adalah kategori ke- i , yaitu $p_{i,}$ konstan;
 - Antar-trial bebas satu sama lain;
 - Cacahan hasil percobaan, yaitu frekuensi setiap kategori $O_i, i = 1, 2, \dots, k$, dengan $\sum O_i = n$.
 - O_i dikenal sebagai frekuensi amatan (*observed frequency*)



Percobaan multinom (*multinomial experiment*)

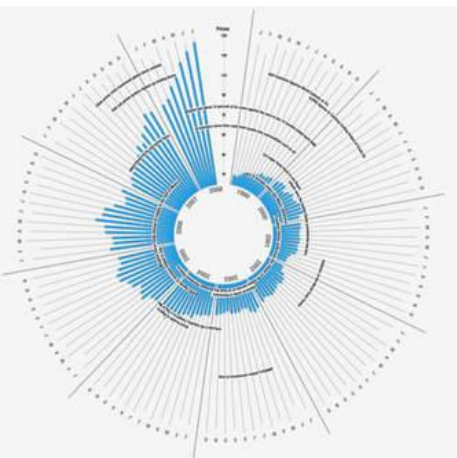
- Dalam kotak ada 4 macam kelereng: putih, kuning, merah dan hitam dengan peluang p_1, p_2, p_3 , dan p_4 . Ambil secara acak (**dengan pemulihan**) satu kelereng dan catat warna kelereng yang terambil. Kalau diulangi sebanyak n kali maka kita memiliki data cacahan dari hasil percobaan multinom.
- Percobaan multinom mirip dg percobaan binom utk contoh besar (ingat statistik Z_{hit} untuk **uji proporsi**) sebelumnya.

Fungsi peluang multinom adalah

$$f(x_1, \dots, x_k; n, p_1, \dots, p_k) = \Pr(X_1 = x_1 \text{ and } \dots \text{ and } X_k = x_k) \\ = \begin{cases} \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \times \cdots \times p_k^{x_k}, & \text{when } \sum_{i=1}^k x_i = n \\ 0 & \text{otherwise,} \end{cases}$$



Ambil 1 kelereng
sebanyak n kali dengan
pemulihan \rightarrow multinom



Data Kategorik

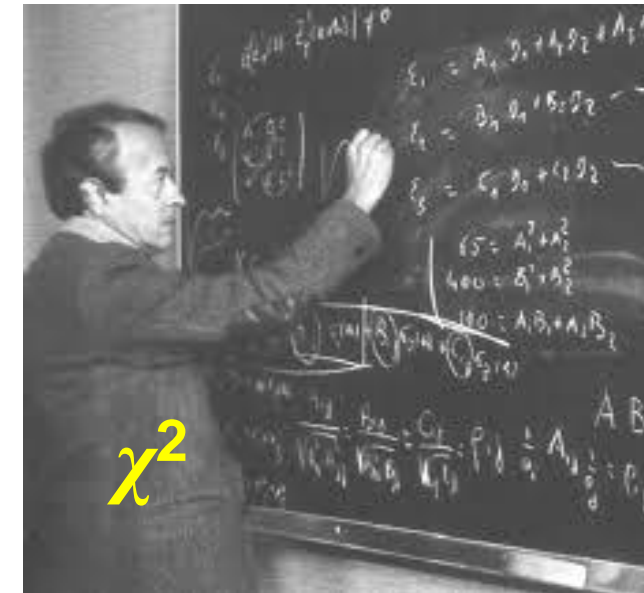


IPB University
— Bogor Indonesia —

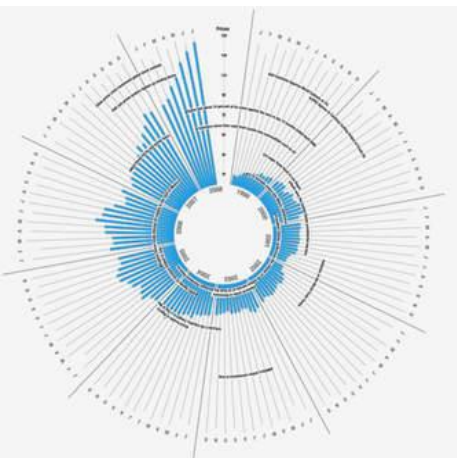
Department of Statistics
Statistics and Data Science Study Program

- Sekarang kita perluas konsep tersebut dengan melakukan inferensi terhadap parameter multinom p_1, p_2, \dots, p_k menggunakan statistik yang berbeda.
- Statistik yang akan kita gunakan dikenal sebagai statistik khi-kuadrat (***chi-square statistic***), pertama kali diperkenalkan oleh Pearson (1900).

$$\chi^2 = \sum_{\text{all categories}} \frac{\overset{\text{Hasil pengamatan}}{\text{(observed frequency - expected frequency)}^2}}{\underset{\text{Teoritis}}{\text{expected frequency}}} = \sum_{i=1}^c \frac{(O_i - E_i)^2}{E_i}$$

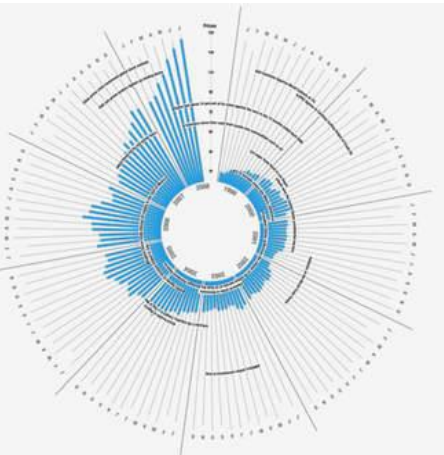


Karl Pearson (1857-1936)



Ilustrasi

- Sebuah dadu yang setimbang dilempar sebanyak $n=300$ kali. Peluang munculnya angka 1-6 sama, yaitu $p_i = \frac{1}{6}$.
- Berapa harapan untuk memperoleh angka 1 dalam pelemparan tersebut? Secara intuitif, frekuensi harapannya (*expected frequency*) adalah $300 * (\frac{1}{6}) = 50$ kali.
- Jika E_i merupakan frekuensi harapan munculnya angka i maka $E_1 = E_2 = \dots = E_6 = n * p_i = 50$.
- Misal ingin diuji $H_0: p_1 = p_{01}; p_2 = p_{02}; \dots, p_i = p_{0i}$.
- Jika H_0 benar, maka frekuensi amatan (yang kita amati), O_i dari hasil percobaan, akan mirip dengan frekuensi harapannya secara teoritis, $E_i = n * p_i$.



Data Kategorik



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

- Statistik uji khi-kuadrat:

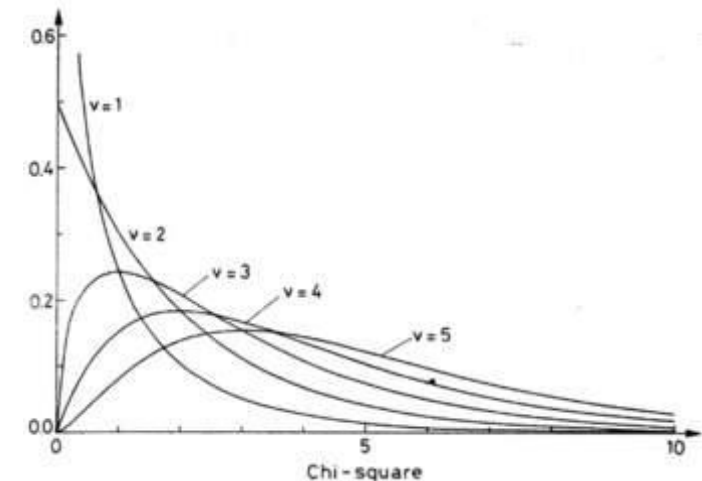
$$\chi_{hit}^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

← Hasil pengamatan
← Teoritis

Untuk n besar maka χ_{hit}^2 mengikuti sebaran χ^2 dengan $db = k-1$, dimana k adalah banyaknya kategori.

$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	df
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6751	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14

$\chi^2_{0.05(4)}$



THE GOODNESS-OF-FIT TEST



IPB University
— Bogor Indonesia —

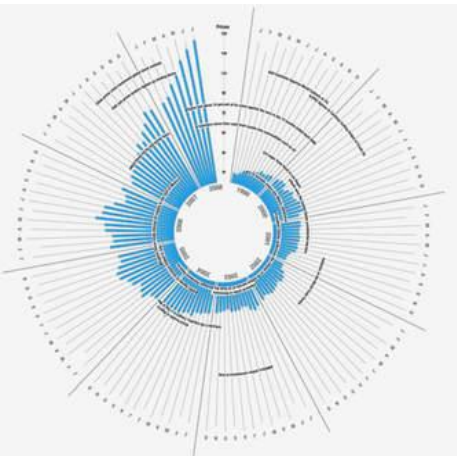
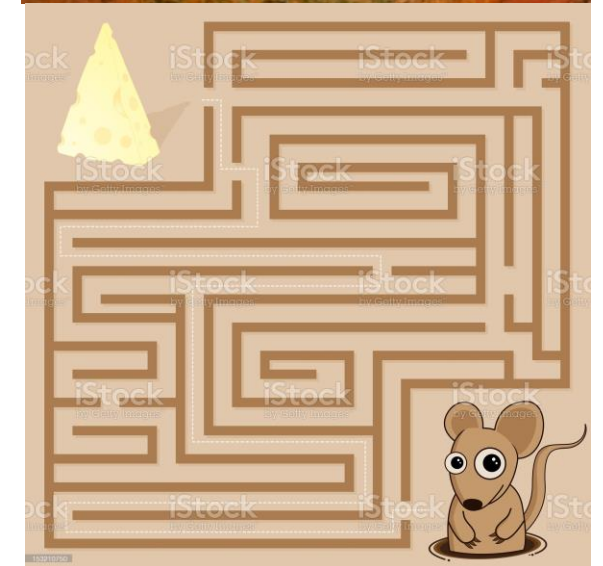
Department of Statistics
Statistics and Data Science Study Program

- **Ilustrasi:** Suatu penelitian ingin mempelajari apakah tikus memiliki preferensi warna kesukaan. Untuk itu seekor tikus dimasukkan ke dalam kandang yang ujungnya memiliki 3 pintu dengan 3 warna (**hijau**, **merah**, **biru**).
- Percobaan diulangi sebanyak 90 kali dan setiap kali dicatat pintu warna apa yang dilalui oleh tikus untuk keluar. Hasilnya **hijau: 20**, **merah: 39**, dan **biru: 31** kali.
- Jika tikus memilih pintu keluar secara **acak** maka $p_1 = p_2 = p_3 = \frac{1}{3}$.
- Jadi ingin diuji

$$H_0: p_1 = p_2 = p_3 = \frac{1}{3} \text{ lawan}$$

$$H_1: \text{sedikitnya ada satu } p_i \neq \frac{1}{3}$$

(p_i : peluang tikus memilih keluar melalui pintu berwarna ke- i)



THE GOODNESS-OF-FIT TEST



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

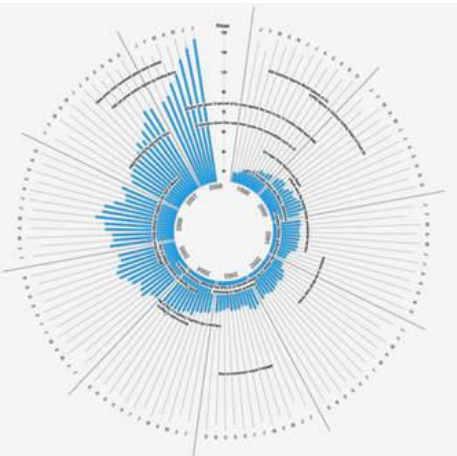
- Frekuensi amatan $O_1=20$, $O_2=39$, dan $O_3=31$.

- Frekuensi harapan E_i (jika H_0 benar) adalah

$$E_1 = 90 \times \left(\frac{1}{3}\right) = 30, \text{ dan } E_2 = E_3 = 30.$$

- Statistik uji $\chi_{hit}^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = \frac{(20-30)^2}{30} + \dots + \frac{(31-30)^2}{30} = 6.067$

- Statistik uji ini dibandingkan dengan sebaran χ^2 dengan derajat bebas $(k-1) = 2 \rightarrow \chi_{0.05}^2(2) = 5.99$.
- Dari tabel χ^2 diperoleh nilai-p ada di antara 0.025 sampai 0.05 sehingga $H_0: p_1 = p_2 = p_3 = \frac{1}{3}$ ditolak, artinya tikus tidak memilih pintu secara acak.
- Tikus cenderung keluar dari **pintu berwarna merah**. Apakah berarti tikus lebih menyukai warna merah ketimbang lainnya? Masih perlu penelitian yg lebih mendalam.



Diskusi Dulu.....



Sesi 1... beres!!!



Sesi 2...



THE GOODNESS-OF-FIT TEST



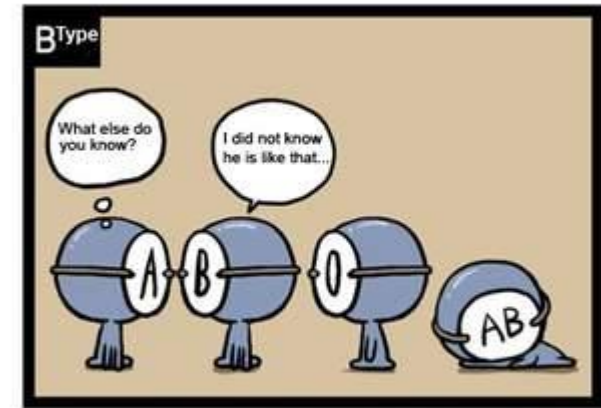
IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

- **Ilustrasi:** Proporsi empat golongan darah dari suatu komunitas selama bertahun-tahun dipercaya sbb: A (0.41), B (0.10), AB (0.04), dan O (0.45).
- Dilakukan suatu survei terhadap 200 orang dari komunitas ini dan hasilnya adalah golongan darah A (89), B (18), AB (12), dan O (81).
- Apakah telah terjadi perubahan proporsi golongan darah di dalam komunitas tersebut?
- Jika tidak ada perubahan proporsi golongan darah maka:
 $p_1 = 0.41; p_2 = 0.10; p_3 = 0.04; p_4 = 0.45$
- Jadi hipotesis yang ingin diuji adalah

$$H_0: p_1 = 0.41; p_2 = 0.10; p_3 = 0.04; p_4 = 0.45$$

$$H_1: \text{ada } p_i \text{ yang tidak sama dengan nilai dalam } H_0.$$



B type doesn't really know well about many things...

THE GOODNESS-OF-FIT TEST



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

- Frekuensi amatan $O_1 = 89$, $O_2 = 18$, $O_3 = 12$, dan $O_4 = 81$, sedangkan frekuensi harapan E_i (jika H_0 benar) adalah $E_1 = 200 \times (0.41) = 82$, begitu juga $E_2 = 20$, $E_3 = 8$, dan $E_4 = 90$.
- Sehingga statistik uji: $\chi_{hit}^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = 3.70$
- Statistik uji ini dibandingkan dengan sebaran χ^2 dengan derajat bebas $(k-1) = 3 \rightarrow \chi_{0.05}^2(3) = 7.81$.
- Dari tabel χ^2 diperoleh bhw nilai $\chi_{hit}^2 \leq \chi_{0.05}^2(3)$ (nilai-p > 0.10), sehingga $H_0: p_1 = 0.41; p_2 = 0.10; p_3 = 0.04; p_4 = 0.45$ diterima (**INKONKLUSIF**).
- Jadi, **tidak cukup bukti** untuk mengatakan bahwa proporsi golongan darah dalam komunitas itu telah berubah.

	Group A	Group B	Group AB	Group O
Red blood cell type				
Antibodies in plasma			None	
Antigens in red blood cell	A antigen	B antigen	A and B antigens	None

Tabel Kontingensi

- Seringkali objek/individu dapat diklasifikasikan berdasarkan dua klasifikasi. Misal, seseorang diklasifikasi menurut gender (pria, wanita) dan pekerjaan (PNS, nonPNS).
- Data yang dicatat biasanya adalah data jumlah individu (cacahan) di dalam berbagai kategori yang terbentuk.
- Dalam ilustrasi di atas ada 4 kategori: pria-PNS, pria-nonPNS, wanita-PNS, dan wanita-nonPNS.
- [Ilustrasi lain](#), pabrik pembuat HP mencatat jumlah kerusakan yang terjadi pada hasil produksinya menurut jenis kerusakan (*defects*) dan waktu pengerjaan (*shift*) → lihat tabel di samping ini.

Observed counts

Gender	Pekerjaan	
	PNS	nonPNS
Pria	O_{11}	O_{12}
Wanita	O_{21}	O_{22}

Contingency Table				
Type of Defects	Shift			Total
	1	2	3	
A	15	26	33	74
B	21	31	17	69
C	45	34	49	128
D	13	5	20	38
Total	94	96	119	309

Ilustrasi

- Pertanyaannya apakah jenis kerusakan HP yg diproduksi berhubungan dgn waktu pengerjaan? Misal, apakah jenis kerusakan A sering terjadi pada *shift* 3? Apakah jenis kerusakan B sering terjadi pada *shift* 1? Dan seterusnya.
- Jadi ingin dievaluasi **apakah baris dan lajur** (kolom) itu berhubungan, atau **tidak bebas** satu sama lain.
- Kebebasan antara baris dan lajur dalam tabel kontingensi dapat dievaluasi menggunakan uji khi-kuadrat dengan hipotesis:

H_0 : baris dan lajur bebas satu sama lain

H_1 : baris dan lajur tidak bebas satu sama lain.

Observed counts

Gender	Pekerjaan	
	PNS	nonPNS
Pria	O_{11}	O_{12}
Wanita	O_{21}	O_{22}

Contingency Table

Type of Defects	Shift			Total
	1	2	3	
A	15	26	33	74
B	21	31	17	69
C	45	34	49	128
D	13	5	20	38
Total	94	96	119	309

Frekuensi Harapan

- Perhatikan tabel di samping ini. Jika baris dan lajur bebas (H_0 benar), maka peluang suatu individu masuk kategori (i,j) sama dengan peluang masuk kategori i dikalikan peluang masuk kategori j . Atau: $p_{ij} = p_{i.} p_{.j}$
- Besarnya peluang tersebut diduga sbb:

$\widehat{p}_{ij} = \frac{O_{ij}}{O_{..}}; \widehat{p}_{i.} = \frac{O_{i.}}{O_{..}}; \text{ dan } \widehat{p}_{.j} = \frac{O_{.j}}{O_{..}}; \text{ dengan } O_{..} = \text{total amatan dalam keseluruhan sel, } O_{i.} = \text{total amatan pada baris ke-}i, \text{ sedangkan } O_{.j} = \text{total amatan pada lajur ke-}j.$

- Sehingga $E_{ij} = (p_{i.} p_{.j})(O_{..})$, atau $E_{ij} = \frac{O_{i.} O_{.j}}{O_{..}}$

Baris	Lajur			Total
	1	...	j	
1	p_{11}	...	p_{1j}	$p_{1.}$
:	:	...	:	:
i	p_{i1}	...	p_{ij}	$p_{i.}$
Total	$p_{.1}$...	$p_{.j}$	$p_{..}$

Baris	Lajur			Total
	1	...	j	
1	O_{11}	...	O_{1j}	$O_{1.}$
:	:	...	:	:
i	O_{i1}	...	O_{ij}	$O_{i.}$
Total	$O_{.1}$...	$O_{.j}$	$O_{..}$

Statistik uji

- Seperti sebelumnya maka statistik uji khi-kuadrat adalah

$$\chi_{hit}^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- Jika statistik ini lebih besar dari nilai χ^2 tabel berderajat bebas $(r-1)(c-1)$ maka H_0 ditolak.
- Untuk ilustrasi perhatikan data jenis kerusakan (*defect*) HP dan waktu pengerjaan (*shift*) sebelumnya.

$$\begin{aligned} \chi_{hit}^2 &= \frac{(15 - 22.51)^2}{22.51} + \frac{(26 - 22.99)^2}{22.99} + \dots + \frac{(20 - 14.63)^2}{14.63} \\ &= 19.18 \end{aligned}$$

- Dari tabel ternyata $\chi_{0.05}^2 = 12.59$ (db=6) $\rightarrow H_0$ ditolak.
- Artinya ada bukti yang kuat bahwa jenis kerusakan HP tergantung pada waktu pengerjaan (*shift*). Jenis rusak C banyak terjadi pada *shift* 3, jenis rusak B pada *shift* 2, dan jenis rusak A pada *shift* 3.

- Seperti sebelumnya maka statistik uji khi-kuadrat adalah

$$\chi_{hit}^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- Jika statistik ini lebih besar dari nilai χ^2 tabel berderajat bebas $(r-1)(c-1)$ maka H_0 ditolak.
- Untuk ilustrasi perhatikan data jenis kerusakan (*defect*) HP dan waktu pengerjaan (*shift*) sebelumnya.



	Shift			
Type of Defects	1	2	3	Total
A	15 (22.51)	26 (22.99)	33 (28.50)	74
B	21 (20.99)	31 (21.44)	17 (26.57)	69
C	45 (38.94)	34 (39.77)	49 (49.29)	128
D	13 (11.56)	5 (11.81)	20 (14.63)	38
Total	94	96	119	309

$$E_{32} = \frac{128 \times 96}{309} = 39.77$$

Ilustrasi

- Survei dilakukan untuk mengevaluasi efektivitas obat flu di suatu daerah. Vaksin diberikan kepada masyarakat 2 kali dalam dua minggu. Ada yang dapat dua kali vaksin tapi ada yang cuma mendapat satu kali.
- Data dari 1000 orang seperti tabel di samping ini.
- Vaksin efektif jika orang yang mendapat vaksin lebih baik kondisinya (flu/tdkFlu) drpd yang tidak mendapat vaksin.

H_0 : Kondisi (flu/tdkFlu) tidak tergantung pada pemberian vaksin

H_1 : Kondisi (flu/tdkFlu) tergantung pada pemberian vaksin

Kondisi	Vaksin			Total
	0 kali	1 kali	2 kali	
Flu	24	9	13	46
TdkFlu	289	100	565	954
Total	313	109	578	1000



www.shutterstock.com · 1911303223

- Vaksin efektif jika orang yang mendapat vaksin lebih baik kondisinya (flu/tdkFlu) drpd yang tidak mendapat vaksin.

H_0 : Kondisi (flu/tdkFlu) tidak tergantung pada pemberian vaksin

H_1 : Kondisi (flu/tdkFlu) tergantung pada pemberian vaksin

- Hasil output komputer $\chi_{hit}^2 = 17.313$ (db=2) sehingga H_0 ditolak pada $\alpha = 5\%$, artinya ada bukti yang kuat bahwa vaksin tersebut efektif untuk menghindari FLU.

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	No Vaccine	One Shot	Two Shots	Total
1	24 14.40 6.404	9 5.01 3.169	13 26.59 6.944	46
2	289 298.60 0.309	100 103.99 0.153	565 551.41 0.335	954
Total	313	109	578	1000

Chi-Sq = 17.313, DF = 2, P-Value = 0.000

Incidence of Flu for Three Treatments

No Vaccine	One Shot	Two Shots
$\frac{24}{313} = .08$	$\frac{9}{109} = .08$	$\frac{13}{578} = .02$

Terkecil

$$\begin{aligned} O_{23} &= 565 \\ E_{23} &= 551.41 \\ \frac{(O_{23} - E_{23})^2}{E_{23}} &= 0.335 \end{aligned}$$

Diskusi Dulu.....



Sesi 2... beres!!!



Sesi 3...



Jumlah Baris/Lajur Tetap



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Fixed margins

- Bisa terjadi dalam pengumpulan data peneliti terlebih dahulu menetapkan total baris atau total lajur karena ingin dibandingkan antar-baris atau antar-lajur.
- Uji χ^2 yang sebelumnya tetap dapat digunakan hanya interpretasinya yang agak berbeda karena hipotesisnya berbeda.
- Persoalannya lalu berubah seperti kita membandingkan proporsi dari dua populasi. Hipotesisnya pun berubah menjadi hipotesis tentang perbedaan proporsi dua populasi.

Pilihan	Kota			
	1	2	3	4
A	76	53	59	48
A ^c	124	147	141	152
Total	200	200	200	200



Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	Ward 1	Ward 2	Ward 3	Ward 4	Total
1	76 59.00 4.898	53 59.00 0.610	59 59.00 0.000	48 59.00 2.051	236
2	124 141.00 2.050	147 141.00 0.255	141 141.00 0.000	152 141.00 0.858	564
Total	200	200	200	200	800

Chi-Sq = 10.722 DF = 3, P-Value = 0.013

Jumlah Baris/Lajur Tetap



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Ilustrasi

- Suatu survei dilakukan untuk membandingkan keterpilihan calon A di 4 daerah. Peneliti menetapkan jumlah responden di setiap kota adalah 200 orang.

H_0 : Proporsi pemilih calon A di empat kota sama saja;

H_1 : Sedikitnya ada dua kota dg proporsi pemilih A berbeda.

Pilihan	Kota			
	1	2	3	4
A	76	53	59	48
A ^c	124	147	141	152
Total	200	200	200	200



$$E_{11} = \frac{236 \times 200}{800}$$

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	Ward 1	Ward 2	Ward 3	Ward 4	Total
1	76 59.00 4.898	53 59.00 0.610	59 59.00 0.000	48 59.00 2.051	236
2	124 141.00 2.050	147 141.00 0.255	141 141.00 0.000	152 141.00 0.858	564
Total	200	200	200	200	800

Chi-Sq = 10.722 DF = 3, P-Value = 0.013

- Dari output komputer nampak bahwa nilai-p = 0.013 sehingga kita **menolak** H_0 .
- Dari data dapat dihitung bahwa calon A paling banyak dipilih di kota 1 ($\widehat{p}_1 = \frac{76}{200} = 0.38$), dan paling sedikit dipilih di kota 4 ($\widehat{p}_4 = \frac{48}{200} = 0.24$).

Ukuran Keeratan Baris vs Lajur



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Table 11.11 GSS Data Showing Race and Gender as Explanatory Variables for Opinion About the Death Penalty

Race	Opinion		n	Gender	Opinion		n
	Favor	Oppose			Favor	Oppose	
White	71%	29%	1473	Male	71%	29%	885
Black	46%	54%	259	Female	62%	38%	1017
Chi-squared = 65.55				Chi-squared = 17.78			
df = 1, P-value = 0.00				df = 1, P-value = 0.00			

Source: Data from CSM, UC Berkeley.

Ilustrasi

Hubungan antara ras dengan pendapat terhadap hukuman mati menghasilkan:

$$\chi_{hit}^2 = \frac{(1046 - 990.79)^2}{990.79} + \dots + \frac{(140 - 84.79)^2}{84.79} = 65.55$$

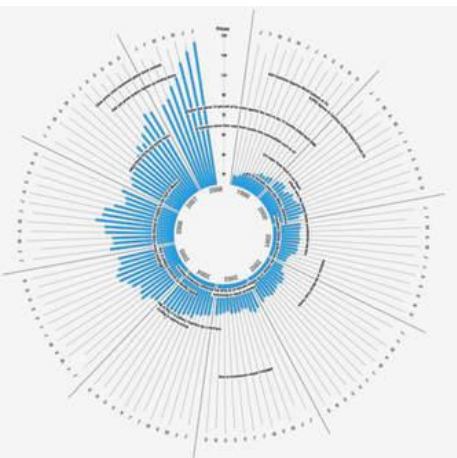
- Sedangkan hubungan antara gender dengan pendapat terhadap hukuman mati menghasilkan $\chi_{hit}^2 = 17.78$.
- Apakah berarti bahwa hubungan ras dengan pendapat thd hukuman mati lebih erat ketimbang gender thd hukuman mati ?
- Disini kita perlu **ukuran keeratan hubungan** antara baris dan lajur.

Ras	Pendapat H.Mati		Total
	Setuju	T.Setuju	
Putih	1046	427	1473
Hitam	119	140	259

O_{ij}

Ras	Pendapat H.Mati		Total
	Setuju	T.Setuju	
Putih	990.79	482.21	1473
Hitam	174.21	84.79	259

E_{ij}



Ukuran Keeratan Baris vs Lajur



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Ukuran keeratan

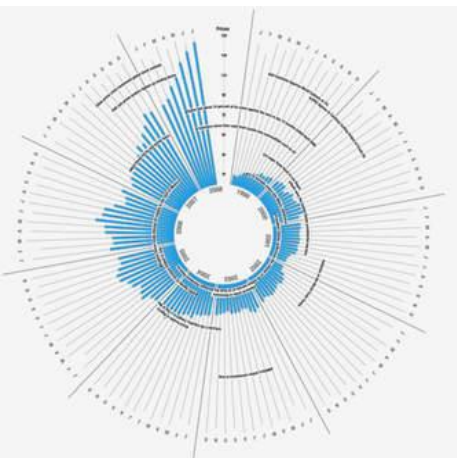
- **1. Beda proporsi:** Ukuran keeratan yang paling mudah adalah beda proporsi yang menyusun respon tertentu.
- Beda proporsi itu sama dengan nol jika sebaran bersyaratnya identik, baris dan lajur bebas satu sama lain.
- Beda proporsi sama dengan -1 atau +1 jika baris dan lajur berhubungan sempurna. Dalam praktek biasanya beda proporsi ini antara -1 sampai +1.

Data hasil survei

Race	Opinion	
	Favor	Oppose
White	71%	29%
Black	46%	54%
Gender		
Male	71%	29%
Female	62%	38%

Data hipotetik

Income	Case A Accept Credit Card			Case B Accept Credit Card		
	No	Yes	Total	No	Yes	Total
High	240 (60%)	160 (40%)	400 (100%)	0 (0%)	400 (100%)	400 (100%)
Low	360 (60%)	240 (40%)	600 (100%)	600 (100%)	0 (0%)	600 (100%)



Ukuran Keeratan Baris vs Lajur



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Ukuran keeratan

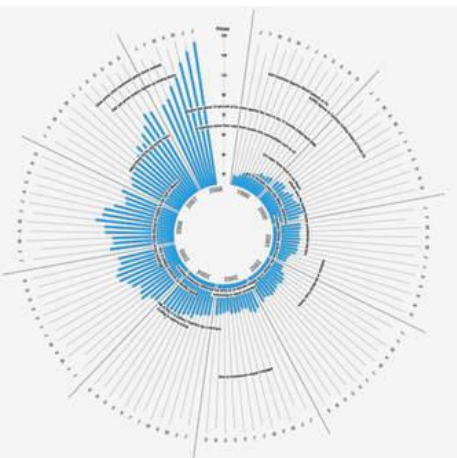
- Untuk kasus A (tabel bawah) nyaris tdk ada hubungan antara pendapatan dengan persetujuan kartu kredit: beda proporsi $(0.6-0.6) = 0$. Tetapi untuk kasus B hubungan baris dan lajur sangat erat yaitu $(0-1) = -1$.
- Sedangkan untuk kasus ras dan gender thd hukuman mati (tabel atas) beda proporsi adalah 0.25 (ras) dan 0.09 (gender). Jadi **hubungan antara ras dgn hukuman mati lebih kuat untuk ras ketimbang untuk gender**.

Data hasil survei

Race	Opinion	
	Favor	Oppose
White	71%	29%
Black	46%	54%
Gender		
Male	71%	29%
Female	62%	38%

Data hipotetik

Income	Case A Accept Credit Card			Case B Accept Credit Card		
	No	Yes	Total	No	Yes	Total
High	240 (60%)	160 (40%)	400 (100%)	0 (0%)	400 (100%)	400 (100%)
Low	360 (60%)	240 (40%)	600 (100%)	600 (100%)	0 (0%)	600 (100%)



Ukuran Keeratan Baris vs Lajur



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Table 11.14 Outcome of Auto Accident by Whether or Not Subject Wore Seat Belt

Wore Seat Belt	Outcome		Total
	Survived	Died	
Yes	412,368	510	412,878
No	162,527	1601	164,128

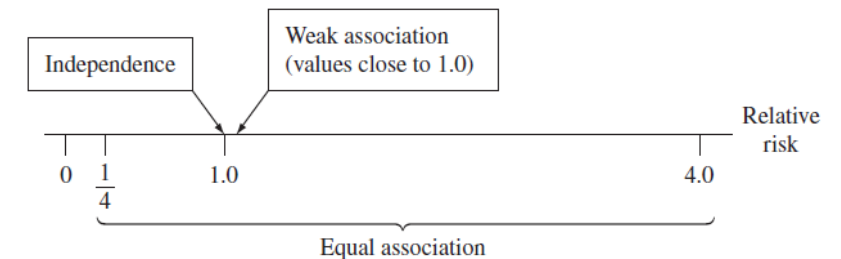
Source: Department of Highway Safety and Motor Vehicles, Florida.

Ukuran keeratan

- **2. Risiko relatif (*relative risk*):** Jika p_1 proporsi kejadian untuk kelompok 1 dan p_2 adalah proporsi kejadian untuk kelompok 2 maka **risiko relatif** kejadian itu adalah $\frac{p_1}{p_2}$.
- Tabel 11.14 menunjukkan proporsi meninggal dari pengguna sabuk pengaman: $510/412,878 = 0.00124$, untuk yg tidak menggunakan sabuk pengaman: $1601/164,128 = 0.127$. Risiko relatif: $0.127/0.00124 = 7.9$.
- Jika terjadi kecelakaan maka risiko meninggal bagi yang tidak menggunakan sabuk adalah 7.9 kali lebih besar dibanding jika menggunakan sabuk pengaman.

Risiko relatif:

- The relative risk can equal any nonnegative number.
- When $p_1 = p_2$, the variables are independent and relative risk = 1.0.
- Values farther from 1.0 (in either direction) represent stronger associations. Two values for the relative risk represent the same strength of association, but in opposite directions, when one value is the reciprocal of the other.



RR = 4 sama saja dengan RR = $\frac{1}{4}$.

Paradoks Simpson



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

Ilustrasi

- Apakah benar telah terjadi diskriminasi dalam penjatuhan hukuman mati berdasar warna kulit (putih/hitam) di AS? (Lihat Tabel 10.16)
- Di Florida dari 326 kasus, ada 19/160 (11.9%) tertuduh dihukum mati jika jaksanya berkulit putih.
- Ada 17/166 atau 10.2% tertuduh yg dihukum mati jika jaksanya berkulit hitam.

Table 10.16 Defendant's Race and Death Penalty Verdict for Homicide Cases in Florida

Defendant's Race	Death Penalty		Total	Percentage Yes
	Yes	No		
White	19	141	160	11.9
Black	17	149	166	10.2

Defendant =
pelaku kriminal

- Risiko relatifnya kecil, yaitu $11.9/10.2 = 1.17$ (**dekat dgn satu**).
- **Tidak cukup bukti** adanya hub antara warna kulit dg hukuman mati. **Benarkah?**

Table 10.17 Defendant's Race and Death Penalty Verdict, Controlling for Victim's Race

Victim's Race	Defendant's Race	Death Penalty		Total	Percentage Yes
		Yes	No		
White	White	19	132	151	12.6
	Black	11	52	63	17.5
Black	White	0	9	9	0.0
	Black	6	97	103	5.8

Paradoks Simpson



IPB University
— Bogor Indonesia —

Department of Statistics
Statistics and Data Science Study Program

- Tetapi jika diungkap **siapa yang menjadi korban** pembunuhan itu maka lihat Tabel 10.17.

Table 10.17 Defendant's Race and Death Penalty Verdict, Controlling for Victim's Race

Victim's Race	Defendant's Race	Death Penalty		Total	Percentage Yes
		Yes	No		
White	White	19	132	151	12.6
	Black	11	52	63	17.5
Black	White	0	9	9	0.0
	Black	6	97	103	5.8

- Jika korban berkulit hitam dan pembunuhnya berkulit putih maka hukuman mati sama dengan nol. Jika korban berkulit putih dan pembunuhnya berkulit hitam maka hukuman mati sebesar 17.5%.
- Terjadi paradoks, ini dikenal sebagai **paradoks Simpson**.

😊 **THANK YOU** 😊



IPB University
— Bogor Indonesia —

Department of Statistics
Study Programs in Statistics and Data Science