

STK 211
Kuliah ke-14 &15

ANALISIS DATA KATEGORI

1. Uji Proporsi
2. Uji Khi-kuadrat



Data Kategori

Data berupa hasil pengukuran karakteristik/atribut individu yang bukan berupa numerik

Misal: - preferensi konsumen terhadap produk
- tingkat pendidikan
- suku bangsa

Analisis: Persentase (proporsi) individu dg karakteristik tertentu.



UJI PROPORSI

- Satu Contoh $\implies H_0: p = p_0$ vs (i) $H_1: p < p_0$,
(ii) $H_1: p > p_0$
atau
(iii) $H_1: p \neq p_0$

Membandingkan proporsi penduduk yang berpendidikan tinggi dengan sebuah nilai tertentu.

Misal: Apakah proporsi orang yg bergelar paling rendah sarjana adalah 25%?

$H_0: p = 0,25$ vs $H_1: p \neq 0,25$



Prosedur Pengujian

1. **Sebaran Binom:** untuk $n \leq 20$

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \text{ untuk } x = 0, 1, 2, \dots, n$$

2. **Sebaran Normal:** untuk $n > 20$

- Statistik Z (Normal baku)



Pengujian dengan Sebaran Binom

Prosedur:

1. Nyatakan hipotesis statistik
 $H_0: p = p_0$ vs $H_1: p < p_0, p > p_0$, atau $H_1: p \neq p_0$
2. Tentukan taraf nyata, α
3. Wilayah kritik:
 - $x \leq k_\alpha$ bila $H_1: p < p_0$
 - $x \geq k_\alpha$ bila $H_1: p > p_0$
 - $x \leq k_{\alpha/2}$ atau $x \geq k_{\alpha/2}$ bila $H_1: p \neq p_0$
4. Perhitungan: Hitunglah x , yaitu banyaknya keberhasilan
5. Keputusan: Tolak H_0 bila x jatuh dlm wilayah kritik; bila tidak, terima H_0 .



Penentuan nilai k_α

k_α adalah bilangan bulat terbesar yg bersifat:

1. $P(X \leq k_\alpha \text{ bila } p = p_0) = \sum_{x=0}^{k_\alpha} b(x; n, p_0) \leq \alpha$

Untuk $H_1: p < p_0$

2. $P(X \geq k_\alpha \text{ bila } p = p_0) = \sum_{x=k_\alpha}^n b(x; n, p_0) \leq \alpha$

Untuk $H_1: p > p_0$

3. $P(X \leq k_{\alpha/2} \text{ bila } p = p_0) = \sum_{x=0}^{k_{\alpha/2}} b(x; n, p_0) \leq \alpha/2$

atau

$$P(X \geq k_{\alpha/2} \text{ bila } p = p_0) = \sum_{x=k_{\alpha/2}}^n b(x; n, p_0) \leq \alpha/2$$

Untuk $H_1: p \neq p_0$



Teladan 1:

Untuk menguji pernyataan produsen TV bhw paling sedikit 90% produk terbarunya akan dipilih oleh konsumen dilakukan survei dg 10 responden secara acak. Dari 10 responden, ternyata hanya 7 orang yang membeli TV model terbaru tersebut.

• $H_0: p = 0.9$ vs $H_1: p < 0.9$

• Taraf nyata $\alpha = 5\%$

• Wilayah kritik:

$$P(X \leq k_{0,05} \text{ bila } p = 0,9) = \sum_{x=0}^{k_{0,05}} b(x;10, 0,9) \leq 0,05$$



Dari Tabel A.2 (Walpole):

$n = 10, p = 0,9$

r	$\Sigma b(x;n,p)$
0	0.0000
1	0.0000
2	0.0000
3	0.0000
4	0.0002
5	0.0016
6	0.0128
7	0.0702

maka $k_{0,05} = 6$

• Perhitungan: $x = 7$

• Keputusan: Tidak menolak H_0 (menerima H_0)



Teladan 2:

a. Sebuah perusahaan rokok mengatakan bahwa 20% di antara para perokok lebih menyukai merk X. Untuk menguji pendapat ini, diambil 20 perokok secara acak dan ditanyakan rokok merk apa yg mereka sukai. Bila 6 di antara 20 perokok itu menyukai merk X, kesimpulan apa yg dapat ditarik? Gunakan taraf nyata 0,05.

a. Misalkan di masa lalu 40% orang dewasa setuju dg hukuman mati. Apakah kita mempunyai alasan utk percaya bhw pada saat ini proporsi org yg setuju hukuman mati telah meningkat bila di antara 15 org dewasa, 8 org setuju hukuman tsb. Gunakan taraf nyata 1%.



2. SEBARAN NORMAL:

• Untuk $n > 20$

• Menggunakan statistik Z:
$$z_{hi} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

• Daerah penolakan H_0 :

$z_h < -z_\alpha$ untuk $H_1: p < p_0$

$z_h > z_\alpha$ untuk $H_1: p > p_0$

$|z_h| > z_{\alpha/2}$ untuk $H_1: p \neq p_0$



Teladan 3:

Sebuah sampel 100 benih kedelai diambil secara acak dari sebuah kantong benih. Sampel tsb kemudian diberi perlakuan utk mendeteksi ada/tidaknya penyakit pd benih tsb. Ternyata 12 di antaranya positif mengandung penyakit.

Apakah pernyataan produsen yg menyatakan bhw 90% benihnya sehat dpt dipercaya? Gunakan taraf nyata 0.05.



$H_0: p = 0.1$ vs $H_1: p \neq 0.1$

$p =$ proporsi benih yg tdk sehat

• Statistik Z:

$$\begin{aligned} z_{hi} &= \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.12 - 0.1}{\sqrt{(0.1)(0.9)/100}} \\ &= \frac{0.02}{0.03} \\ &= 0.667 \end{aligned}$$

$z_{0.025} = 1.96$

⇒ Tidak dapat menolak H_0 , pernyataan produsen yg menyatakan bhw 90% benihnya sehat adalah benar.



- Dua Contoh $\implies H_0: p_1 = p_2$ vs (i) $H_1: p_1 < p_2$,
(ii) $H_1: p_1 > p_2$,
atau
(iii) $H_1: p_1 \neq p_2$

Membandingkan proporsi benih yang bergejala positif dari 2 produsen berbeda.

Apakah produsen pertama menghasilkan proporsi benih sehat lebih banyak daripada produsen kedua?

$$H_0: p_1 = p_2 \text{ vs } H_1: p_1 > p_2$$



Prosedur Pengujian

- Menggunakan statistik Z, syarat: n_1 dan n_2 lebih besar dari 20

$$z_h = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}[(1/n_1) + (1/n_2)]}}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$



Teladan 4:

Sebuah survei dilakukan utk membandingkan viabilitas benih kedelai dari 2 produsen berbeda. Sebanyak 50 benih dari masing-masing produsen diambil secara acak untuk diperiksa daya kecambahnya. Setelah semua benih dibiarkan berkecambah, 80% benih dari produsen A dan 66% benih dari produsen B mampu berkecambah.

Apakah kedua produsen benih menghasilkan viabilitas yang sama? Gunakan taraf nyata 5%.



UJI KHI-KUADRAT

Asosiasi antara dua peubah kategori.

Misal: apakah ada asosiasi (hubungan) antara warna benih dengan warna daun bunga tanaman kacang

Tabel Kontingensi (Uji kebebasan dua peubah):

- Tabel 2 x 2
- Tabel 2 x c
- Tabel r x c



Uji Kebebasan Dua Peubah (Tabel Kontingensi)

Utk melihat apakah warna benih dan warna daun bunga berasosiasi dilakukan pengambilan sampel acak thd 500 tanaman kacang dari sebuah populasi. Setiap tan dikasifikasikan sesuai dg warna benih dan warna daun bunganya.

Warna daun bunga	Warna benih			Total	
	Cokelat gelap	Cokelat	Cokelat terang		
Ungu	21	32	13	4	70
Keunguan	55	102	70	8	235
Putih	14	59	112	10	195
Total	90	193	195	22	500



Tabel Kontingensi r x c

Lajur \ Baris		Peubah II				Total (n _j)
		1	2	...	c	
1	n ₁₁	n ₁₂	...	n _{1c}	n _{1.}	
2	n ₂₁	n ₂₂	...	n _{2c}	n _{2.}	
...	
r	n _{r1}	n _{r2}	...	n _{rc}	n _{r.}	
Total (n _j)	n _{.1}	n _{.2}	...	n _{.c}	n _{..}	

Peubah I



Hipotesis:

H₀: Kedua peubah tidak berasosiasi
H₁: Kedua peubah berasosiasi

Statistik uji:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

dimana $O_{ij} = n_{ij}$ dan

$$E_{ij} = \frac{n_{i.} \times n_{.j}}{n_{..}}$$

dengan derajat bebas (V) = (r-1) x (c-1)



Rule of Five (Aturan Lima):

- Nilai frekuensi harapan tidak kurang dari 5 ($E_{ij} \geq 5$)
- Untuk sel yang mengandung $E_{ij} < 5$, maka gabungkan baris atau lajur yang mengandung $E_{ij} < 5$ tsb.



Warna daun bunga	Warna benih				Total
	Cokelat gelap	Cokelat	Cokelat terang	Putih	
Ungu	21 (12.6)	32 (27.0)	13 (27.3)	4 (3.1)	70
Agak ungu	55 (42.3)	102 (90.7)	70 (91.7)	8 (10.3)	235
Putih	14 (35.1)	59 (75.3)	112 (76.1)	10 (8.6)	195
Total	90	193	195	22	500

$$E_{ij} = \frac{(\text{Total Baris ke } i) \times (\text{Total lajur ke } j)}{\text{Total umum}}$$

$$E_{11} = \frac{70 \times 90}{500} = 12.6, \dots, E_{34} = \frac{195 \times 22}{500} = 8.58$$



Hipotesis:

H_0 : Warna benih tidak berasosiasi dgn warna daun

H_1 : Warna benih berasosiasi dgn warna daun

Statistik uji:

$$\chi^2 = \frac{(21-12.6)^2}{12.6} + \frac{(32-27.02)^2}{27.02} + \dots + \frac{(10-8.58)^2}{8.58}$$
$$= 5.6 + 0.918 + 7.49 + 0.275 + \dots + 0.235 = 58.575$$

$$db = 6, \chi^2_{0.05(6)} = 12.59$$

\Rightarrow Tolak H_0 . Ada asosiasi antara warna benih dengan warna daun bunga tanaman kacang.



Teladan 5 (Tabel 2 x 5):

Pengaruh lima metode penyimpanan thd viabilitas benih diuji dgn menggunakan 124 benih diberi perlakuan dg metode A, 90 benih dg metode B, 120 benih dg metode C, 50 benih dg metode D, dan 100 benih dg metode E. Hasil uji tsb tertera pada tabel di bawah:

	Metode penyimpanan				
	A	B	C	D	E
Jm. bnh berkecambah	112	76	88	43	92
Jm. bnh tdk berkecambah	12	14	32	7	8

Apakah ada asosiasi antara metode penyimpanan dengan viabilitas benih? Gunakan taraf nyata 5%.


