

Inferensi Nonparameterik

Chi-square test & Kolmogorov-Smirnov test

Week 4 – Metode Statistika



Outline

1

Chi-square GoF test

2

Normality test (K-S)





1

Chi-square GoF test

Penggunaan Uji Kesesuaian Chi-Square

- Chi Square adalah salah satu alat analisis yang paling sering digunakan pada statistik, dengan tujuan untuk Uji Homogenitas, Uji Independensi dan Uji Goodness of Fit Test.
- Uji ini merupakan uji non-parametrik yang dirancang khusus untuk distribusi diskrit. Uji ini sebagai alternatif untuk uji Kolmogorov-Smirnov atau uji Shapiro-Wilk ketika distribusi diskrit, seperti binomial dan Poisson, digunakan.

Penggunaan Uji Kesesuaian Chi-Square

- Uji Chi-square menguji hipotesis bahwa sampel data untuk satu variabel kategori dengan k kategori berasal dari populasi dengan distribusi tertentu.
- Frekuensi yang diamati untuk setiap kategori dibandingkan dengan frekuensi yang diharapkan yang dihitung berdasarkan distribusi yang diuji.
- Jika hanya ada dua kategori, pertimbangkan untuk menggunakan uji binomial.
- Pengujian dapat diterapkan pada distribusi kontinu hanya dengan menggabungkannya (*binning*), yaitu mengubahnya menjadi distribusi diskrit.

Uji Kesesuaian Chi-Square

- Perumusan hipotesis:
H0 : data sampel berasal dari suatu populasi berdistribusi Chi-square
H1 : data sampel tidak berasal dari suatu populasi berdistribusi Chi-square

- Statistik uji:

$$\chi^2 = \sum_{i=1}^r \frac{(\theta_i - e_i)^2}{e_i}$$

Dengan θ_i adalah nilai pengamatan dan e_i adalah nilai harapan.

Uji Kesesuaian Chi-Square

- Daerah penolakan:
Tolak H_0 apabila $\chi^2 \geq \chi^2_{\alpha, (r-k-1)}$
 r : banyak karakteristik yang diamati
 k : banyaknya parameter yang ditaksir
- Struktur data:

Kategori X	1	2	...	r	$\Sigma = \text{total}$
θ_i	θ_1	θ_2	...	θ_r	n
$e_i = n P_i$	e_1	e_2	...	e_r	n

Pada Uji Chi Square ini akan dibahas tentang uji-uji untuk keselarasan dari distribusi : Uniform dan Normal



2

Normality test

Studi Kasus Metode Nonparametrik

- Seorang peneliti ingin mengetahui apakah jam tidur dapat mempengaruhi frekuensi seseorang terserang penyakit.
- Setelah mendapatkan data, ternyata frekuensi seseorang terserang penyakit **tidak berdistribusi normal** karena sebagian besar responden sangat jarang sakit sedangkan beberapa orang lainnya sering merasa sakit, sehingga terdapat outlier atau terdapat skewness.
- Oleh karena itu, peneliti tidak dapat menggunakan analisis regresi linear yang mensyaratkan asumsi distribusi normal pada variabelnya.
- Sehingga peneliti dapat menggunakan metode non-parametrik untuk mengetahui hubungan tersebut seperti analisis regresi non-parametrik

Bagaimana Menguji Distribusi Normalitas Data?

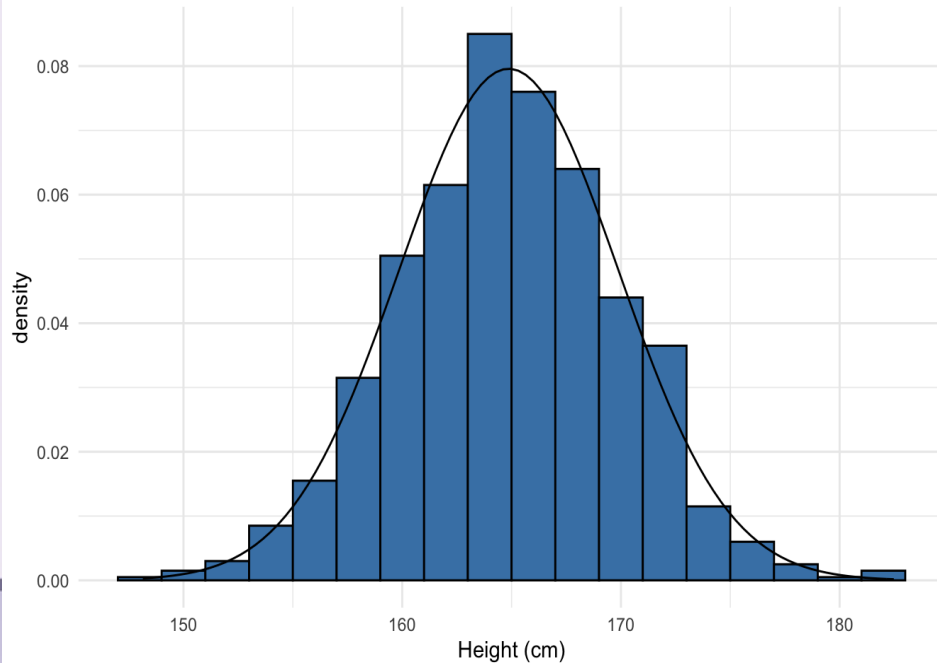
1. ANALISIS GRAFIK

HISTOGRAM

- Metode yang lebih umum dan mudah tetapi kurang sesuai jika digunakan pada ukuran sampel kecil

Histogram of adult height and normal curve

N = 1000, mean = 164.87, variance = 25.13



NORMAL PROBABILITY PLOT

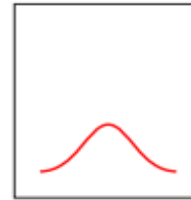
- Dilihat dengan membandingkan distribusi kumulatif data dengan distribusi kumulatif distribusi normal.
- Bentuk distribusi normal membentuk garis diagonal yang lurus dan data dibandingkan dengan diagonal tersebut.
- Jika distribusi datanya normal, maka garis yang merepresentasikan data tersebut akan mendekati diagonal

Normal Probability Plots and Corresponding Univariate Distributions

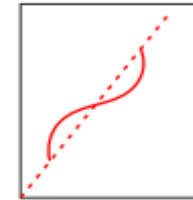


Normal probability plot

(a) Normal distribution

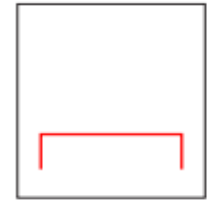


Univariate distribution

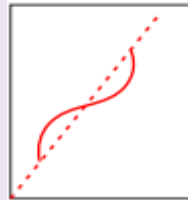


Normal probability plot

(b) Uniform distribution

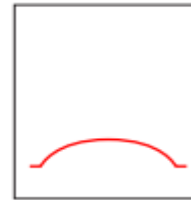


Univariate distribution

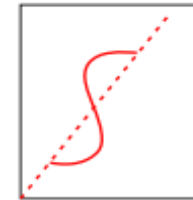


Normal probability plot

(c) Non-peaked distribution

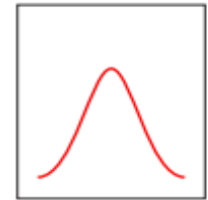


Univariate distribution

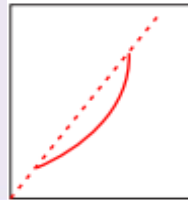


Normal probability plot

(d) Peaked distribution

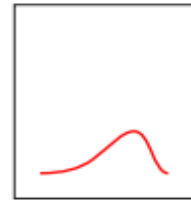


Univariate distribution

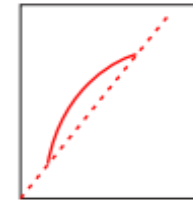


Normal probability plot

(e) Negative distribution

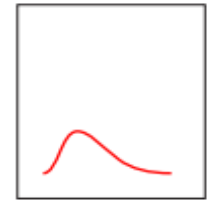


Univariate distribution



Normal probability plot

(f) Positive distribution



Univariate distribution

— Plot of univariate distribution

- - - - Cumulative normal distribution

Bagaimana Menguji Distribusi Normalitas Data?

2. UJI STATISTIK

UJI KOLMOGOROV –SMIRNOV & SHAPIRO WILK

- Menghitung tingkat signifikansi dari distribusi normal.
- Uji signifikansi tidak terlalu terlihat untuk sampel kecil (kurang dari 30) tetapi cukup sensitif untuk sampel besar (lebih dari 1000)
- Dapat menggunakan Minitab, R, atau Python
- Kolmogorov-Smirnov untuk sample ≥ 50 (data normal bila $\text{sig.} > 0.05$)
- Shapiro-Wilk untuk sample < 50 (data normal bila $\text{sig.} > 0.05$)
- Selain uji KS dan SW, uji normalitas juga dapat dilakukan menggunakan uji Liliefors.

Uji Kolmogorov-Smirnov

Uji K-S satu sampel

- Uji Kolmogorov Smirnov pada prinsipnya, digunakan sebagai uji *goodness of fit* (kesesuaian) antara frekuensi pengamatan dengan frekuensi yang diharapkan, yang tidak memerlukan asumsi tertentu tentang bentuk distribusi data populasi dimana sampel tersebut diambil.
- Hipotesis
 - H_0 : data sampel berasal dari distribusi normal
 - H_1 : data sampel tidak berasal dari distribusi normal

K-S satu sampel Test

Langkah Analisis:

1. Mengurutkan data hasil pengamatan dari nilai yang terkecil sampai dengan yang terbesar.
2. Menyusun distribusi kumulatif relatif data hasil pengamatan dan diberi simbol $F_a(X)$.
3. Menghitung nilai Z dengan rumus:

$$Z = \frac{X - \mu}{\sigma}$$

Dimana, μ adalah nilai rata-rata, sedangkan σ adalah nilai standar deviasi.

K-S satu sampel Test

Langkah Analisis (lanjutan):

4. Menghitung distribusi kumulatif teoritis (berdasarkan area kurva normal), dan notasikan dengan $F_e(X)$
5. Menghitung selisih antara $F_a(X)$ dengan $F_e(X)$
6. Mengambil selisih mutlak maksimum antara $F_a(X)$ dengan $F_e(X)$ dan notasikan dengan D .

$$D = \text{Max}|F_a(X) - F_e(X)|$$

4. Membandingkan nilai D yang diperoleh dengan nilai D_α dari tabel nilai D untuk uji Kolmogorov-Smirnov sampel tunggal. Dengan kriteria pengambilan keputusan adalah:
 - H_0 diterima apabila $D \leq D_\alpha$
 - H_0 ditolak apabila $D > D_\alpha$

Contoh K-S satu sampel Test

Seorang dosen olah raga melakukan penelitian untuk menguji apakah tinggi mahasiswa mengikuti distribusi normal atau tidak, untuk keperluan tersebut diambil sampel secara acak sebanyak 10 mahasiswa. ($\alpha = 5\%$)

Data Penelitian

Nama Mahasiswa	Tinggi Badan
Tjang Kung	175
Bung Kuk	160
Ce Ball	150
Tjing Grink	170
Land Dunk	174
Thing Gie	170
Thiang Bonk Soer	176
Bhon Cell	160
Bho Gell	165
Methek Shell	155

K-S satu sampel Test

Penyelesaian:

1. Judul Penelitian

Uji Distribusi Kenormalan Tinggi Mahasiswa

2. Variabel Penelitian

Tinggi Mahasiswa

3. Pertanyaan Penelitian

Apakah tinggi badan mahasiswa berdistribusi normal ?

4. Hipotesis:

H_0 : Tinggi badan mahasiswa berdistribusi normal.

H_1 : Tinggi badan mahasiswa berdistribusi tidak normal.

5. Kriteria Pengujian

H_0 tidak dapat ditolak, jika $\text{Sig.} > \alpha$ (α).

H_0 ditolak, jika $\text{Sig.} \leq \alpha$ (α).

K-S satu sampel Test

Analisis Data

No	X	X Urut	Kum	Fa (X)	Nilai Z	Fe (X)	Fa (X)- Fe (X)
1	175	150	150	0.09	-1.728	0.042	0.049
2	160	155	305	0.18	-1.170	0.121	0.063
3	150	160	465	0.28	-0.613	0.270	0.011
4	170	160	625	0.38	-0.613	0.270	0.108
5	174	165	790	0.48	-0.056	0.478	0.000
6	170	170	960	0.58	0.502	0.692	-0.112
7	176	170	1130	0.68	0.502	0.692	-0.009
8	160	174	1304	0.79	0.947	0.828	-0.040
9	165	175	1479	0.89	1.059	0.855	0.038
10	155	176	1655	1.00	1.170	0.879	0.121
Rata		165.5				Max	0.121
St Dev		8.9722					

$$D_{max}(0.121) \leq D_{0.05,9}(0.430)$$

Keputusan: gagal tolak H_0

K-S satu sampel Test

Output SPSS

One-Sample Kolmogorov-Smirnov Test		
		Tinggi Badan
N		10
Normal Parameters ^a	Mean	165.5000
	Std. Deviation	8.97218
Most Extreme Differences	Absolute	.192
	Positive	.130
	Negative	-.192
Kolmogorov-Smirnov Z		.607
Asymp. Sig. (2-tailed)		.855
a. Test distribution is Normal.		

K-S satu sampel Test

Kesimpulan:

Berdasarkan hasil analisis diperoleh Asymp Sig.(2-tailed) sebesar 0.855, karena nilai Asymp Sig.(2-tailed) sebesar 0,855 lebih besar dari alpha (0.05), maka hipotesis nol tidak dapat ditolak, sehingga hipotesis yang menyatakan "*Rata-rata tinggi mahasiswa berdistribusi normal*", diterima.

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

