

Uji Kolmogorov Smirnov

Pengertian

- Uji Kolmogorov Smirnov merupakan pengujian normalitas yang banyak dipakai, terutama setelah adanya banyak program statistik yang beredar.
- Kelebihan dari uji ini adalah sederhana dan tidak menimbulkan perbedaan persepsi di antara satu pengamat dengan pengamat yang lain, yang sering terjadi pada uji normalitas dengan menggunakan grafik.

- Konsep dasar dari uji normalitas Kolmogorov Smirnov adalah dengan membandingkan distribusi data (yang akan diuji normalitasnya) dengan distribusi normal baku.
- Distribusi normal baku adalah data yang telah ditransformasikan ke dalam bentuk Z-Score dan diasumsikan normal.
- Jadi sebenarnya uji Kolmogorov Smirnov adalah uji beda antara data yang diuji normalitasnya dengan data normal baku.

- Seperti pada uji beda biasa, jika signifikansi di bawah 0,05 berarti terdapat perbedaan yang signifikan, dan jika signifikansi di atas 0,05 maka tidak terjadi perbedaan yang signifikan.
- Penerapan pada uji Kolmogorov Smirnov adalah bahwa jika signifikansi di bawah 0,05 berarti data yang akan diuji mempunyai perbedaan yang signifikan dengan data normal baku, berarti data tersebut tidak normal.

$p < 0,05 \rightarrow$ distribusi data tidak normal

$p \geq 0,05 \rightarrow$ distribusi data normal

- kelemahan dari Uji Kolmogorov Smirnov, yaitu bahwa jika kesimpulan kita memberikan hasil yang tidak normal, maka kita tidak bisa menentukan transformasi seperti apa yang harus kita gunakan untuk normalisasi.
- Jadi ya kalau tidak normal, gunakan plot grafik untuk melihat menceng ke kanan atau ke kiri, atau menggunakan Skewness dan Kurtosis sehingga dapat ditentukan transformasi seperti apa yang paling tepat dipergunakan.

Contoh

Hasil dari outopsi mendapatkan data berat otak 15 orang dewasa yang menderita penyakit X sbb:

1348	1140	1086	1039	920
1233	1146	1002	1012	904
1255	1168	1016	1001	973

Berdasarkan data di atas apakah dapat di tarik kesimpulan bahwa populasi asal sampel adalah normal dengan rata-rata 1083 dan simpangan baku 129 ?

Prinsip uji K-S :

Menghitung selisih absolut $F_s(x)$ (distribusi frekuensi kumulatif sampel) dengan $F_t(x)$ (distribusi frekuensi kumulatif teoritis)

$$D = |F_s(x) - F_t(x)| \max$$

Step Cara Manual

Asumsi: sampel adalah acak dari distr populasi kontinu

1. Susun hipotesis

- i. $H_0 : F(x) = F_t(x), \forall x$
- ii. $H_1 : F(x) \neq F_t(x), \text{palingsedikit satu } x$

2. Urutkan data dari yang terkecil ke terbesar

904	920	973	1001	1002	1012	1016	1039
1086	1140	1146		1168	1233	1255	1348

3. Hitung distribusi FS(x_i) dengan rata-rata 1083 dan simpangan baku 129 !

4. Hitung $F_t(x_i)$ dibantu tabel distribusi normal baku z , $z = (x_i - \text{rerata}) / \text{simpangan baku}$

5. Hitung $D!$, tentukan D max

6. Tentukan kuantil statistik uji kolmogorov pada tabel K-S, sebut k

7. Jika $D > k$ maka H_0 tidak ditolak

3. Menyusun FS(xi)

xi	f	f _{kum}	F _S (xi)
904	1	1	0,066667
920	1	2	0,133333
973	1	3	0,2
1001	1	4	0,266667
1002	1	5	0,333333
1012	1	6	0,4
1016	1	7	0,466667
1039	1	8	0,533333
1086	1	9	0,6
1140	1	10	0,666667
1146	1	11	0,733333
1168	1	12	0,8
1233	1	13	0,866667
1255	1	14	0,933333
1348	1	15	1

4. Membuat Ft(xi)

$$\bar{x} = 1082,87, \quad s = 128,792$$

xi	$x_i - \bar{x}$	$\frac{x_i - \bar{x}}{s}$	F _t (xi)
904	-178,867	-1,38881	0,0823
920	-162,867	-1,26457	0,1038
973	-109,867	-0,85306	0,1977
1001	-81,867	-0,63565	0,2611
1002	-80,867	-0,62789	0,2643
1012	-70,867	-0,55024	0,2912
1016	-66,867	-0,51919	0,3015
1039	-43,867	-0,3406	0,3669
1086	3,133	0,024326	0,508
1140	57,133	0,443607	0,67
1146	63,133	0,490193	0,6879
1168	85,133	0,661012	0,7454
1233	150,133	1,165701	0,877
1255	172,133	1,336519	0,9082
1348	265,133	2,058614	0,9798

5. Menentukan D

$F_s(x_i)$	$F_t(x_i)$	$ F_s(x_i) - F_t(x_i) $
0,066667	0,0823	0,015597
0,133333	0,1038	0,029498
0,2	0,1977	0,002338
0,266667	0,2611	0,005567
0,333333	0,2643	0,069033
0,4	0,2912	0,1088
0,466667	0,3015	0,165167
0,533333	0,3669	D=Maks 0,166433
0,6	0,508	0,092
0,666667	0,67	0,003333
0,733333	0,6879	0,045433
0,8	0,7454	0,0546
0,866667	0,877	0,010333
0,933333	0,9082	0,025133
1	0,9798	0,0202

6. Lihat Tabel KS, dengan $\alpha=0,05$, $n=15$ diperoleh $k=0,338$

7. Keputusan Uji:
Karena $D=0,1664 < k$
Maka H_0 tidak ditolak
d.k.l
Sampel berat otak berasal
dari distribusi Normal

1. Langkah-langkah prinsip uji Kolmogorov-Smirnov ialah sebagai berikut:

Susun frekuensi-frekuensi dari tiap nilai teramati, berurutan dari nilai terkecil sampai nilai terbesar. Kemudian susun frekuensi kumulatif dari nilai-nilai teramati itu.

2. Konversikan frekuensi kumulatif itu ke dalam probabilitas, yaitu ke dalam fungsi distribusi frekuensi kumulatif $[S(x)]$. Sekali lagi ingat bahwa, distribusi frekuensi teramati harus merupakan hasil pengukuran variabel paling sedikit dalam skala ordinal (tidak bisa dalam skala nominal).

3. Hitung nilai z untuk masing-masing nilai teramati di atas dengan rumus $z = (x_i - \bar{x}) / s$. dengan mengacu kepada tabel distribusi normal baku (tabel B), carilah probabilitas (luas area) kumulatif untuk setiap nilai teramati. Hasilnya ialah sebagai $F_0(x_i)$.

4. Susun $F_s(x)$ berdampingan dengan $F_0(x)$. hitung selisih absolut antara $S(x)$ dan $F_0(x)$ pada masing-masing nilai teramati.

5. Statistik uji Kolmogorov-Smirnov ialah selisih absolut terbesar $F_s(x_i)$ dan $F_t(x_i)$ yang juga disebut deviasi maksimum D

6. Dengan mengacu pada distribusi pencuplikan kita bisa mengetahui apakah perbedaan sebesar itu (yaitu nilai D maksimum teramati) terjadi hanya karena kebetulan. Dengan mengacu pada tabel D, kita lihat berapa probabilitas (dua sisi) kejadian untuk menemukan nilai-nilai teramati sebesar D , bila H_0 benar. Jika probabilitas itu sama atau lebih kecil dari α , maka H_0 ditolak

Berikut ini usia mulai haid pada sejumlah wanita diambil sampel sebanyak 18 orang dengan distribusi sebagaimana tersaji pada tabel berikut :

Usia mulai haid (tahun)	frekuensi
10	3
11	4
12	6
13	2
14	1
15	1
16	1

Ujilah hipotesis nol yang menyatakan bahwa data ini berasal dari suatu populasi berdistribusi normal; diketahui bahwa pada populasi, rata-rata usia mulai haid =12; dengan SD=50.

HIPOTESIS

Hipotesis yang diuji dinyatakan sebagai berikut (dua sisi):

Ho : Kedua sampel berasal dari populasi dengan distribusi yang sama

Ha : kedua sampel bukan berasal dari populasi dengan distribusi yang sama

PERHITUNGAN

Langkah-langkah menghitung nilai-nilai $S(x_i)$ dan $F_o(x_i)$:

Umur (x)	Frek	Frek Kumulatif	$S(x)$ $\left(\frac{\text{Frek. kom}}{\sum \text{Frek}}\right)$	$Z = \frac{x - \bar{x}}{SD}$	z tabel	$F_o(x)$ (0,5 - Ztabel)	d [$S(x) - F_o(x)$]
10	3	3	0.1667	-0.04	0.0160	0.4840	0.3173
11	4	7	0.3889	-0.02	0.0080	0.4920	0.1031
12	6	13	0.7222	0	0.0000	0.0000	-0.7222
13	2	15	0.8333	0.02	0.0080	0.4920	-0.3413
14	1	16	0.8889	0.04	0.0160	0.4840	-0.4049
15	1	17	0.9444	0.06	0.0239	0.4761	-0.4683
16	1	18	1.0000	0.08	0.0319	0.4681	-0.5319

Untuk memperoleh nilai-nilai $F_o(x)$, pertama-tama yang dilakukan adalah mengkonversikan setiap nilai x teramati menjadi nilai unit variabel normal yang disebut z . Sedang $z = (x_i - \bar{x}) / s$. dari tabel distribusi kumulatif normal baku (Tabel B), kita temukan luas area dari minus tak terhingga sampai z . luas area tersebut memuat nilai-nilai $F_o(x)$. Selanjutnya kita hitung statistik uji D , dari sekian banyak nilai D ternyata statistik uji D maksimum adalah $= 0,7222$. Selanjutnya nilai tersebut dibandingkan dengan nilai D tabel (Tabel Kolmogorv-Smirnov).

SAMPLE SIZE	LEVEL OF SIGNIFICANCE FOR D = MAXIMUM [$F_0(X) - S_n(X)$] SIGNIFIKANSI UNTUK TINGKAT D = MAKSIMUM [$F_0(X) - S_n(X)$]				
(N)	0.20	0.15	0.1	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.510	0.565	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.368	0.410	0.490
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.450
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363

KEPUTUSAN

Dari tabel D diatas, dengan $n=18$ dan α (dua sisi) = 0,05 kita dapatkan nilai tabel 0,309. Karena $0,722 > 0,309$, maka H_0 ditolak, maka kita simpulkan bahwa sampel yang berasal dari populasi tidak dengan distribusi normal.

TES DUA SAMPEL KOLMOGOROV SMIRNOV (*KOLMOGOROV SMIRNOV TWO SAMPLE TEST*)

Fungsi :

Tes dua sampel Kolmogorov Smirnov adalah tes yang digunakan untuk mengetahui apakah dua sampel bebas (independent) berasal dari populasi yang sama. Artinya tes ini diterapkan dalam kaitan pembuktian apakah sampel yang diambil berasal dari satu populasi yang sama atau populasi yang berbeda.

Apakah terdapat kesamaan sebaran data tingkat absensi mahasiswa dan mahasiswi perguruan tinggi berikut apabila digunakan alfa = 1%

Tingkat Absen (Bolos)	Jumlah Mahasiswa	Jumlah Mahasiswi
0	120	240
1-4	300	500
5-8	180	400
9-12	300	600
>12	100	260
Jumlah	1000	2000

Teknik Perhitungan D max :

Bolos	f.1	FK1	FKR1	f.2	FK2	FKR2	Selisish mutlak FKR1- FKR2
0	120	120	0.12	240	240	0.12	0
1-4	300	420	0.42	500	740	0.37	0.05
5-8	180	600	0.60	400	1140	0.57	0.03
9-12	300	900	0.90	600	1740	0.87	0.03
>12	100	1000	1	260	2000	1	-

Sehingga Dmax sebesar = 0,05