

UJIAN AKHIR SEMESTER STATISTIKA DAN PROBABILITAS

Senin, 10 Desember 2018 | 100 menit
[Boleh membuka buku | Tidak boleh memakai komputer]

SOAL 1 [SO A-3, BOBOT NILAI 40%]

Hasil *Home Interview (HI) Survey* di kawasan perumahan perkotaan dalam delapan zona menghasilkan data hubungan antara kepemilikan kendaraan bermotor (mobil) dan jumlah perjalanan per hari yang terjadi di kawasan tersebut. Data disajikan pada tabel di bawah ini.

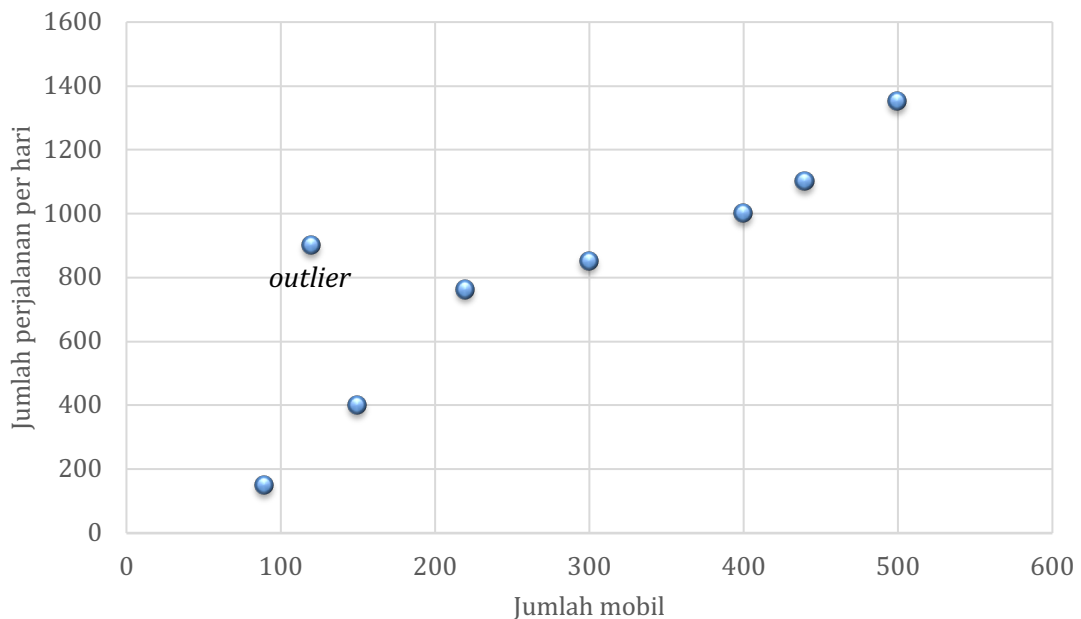
Zona	jumlah mobil	jumlah perjalanan per hari
1	220	760
2	300	850
3	400	1000
4	120	900
5	440	1100
6	500	1350
7	150	400
8	90	150

- Temukanlah persamaan jumlah perjalanan per hari sebagai fungsi jumlah mobil dengan teknik regresi linear, metode kuadrat terkecil. [Bobot 20%]
- Berapakah koefisien korelasi hubungan linear kedua variabel tersebut? [Bobot 10%]
- Apakah yang dapat Saudara simpulkan dari hubungan kedua variabel tersebut? [Bobot 10%]

PENYELESAIAN

(a) Regresi linear [bobot nilai 20%]

Langkah cerdas dalam penyelesaian soal ini adalah mengawalinya dengan memplotkan data.



Dari gambar tampak bahwa salah satu pasangan data berada pada posisi yang menyimpang dari pola data yang lain. Pasangan data ini, yaitu di Zona 4, jumlah mobil adalah 120 dan jumlah perjalanan per hari adalah 900. Pola data secara keseluruhan menunjukkan keteraturan hubungan antara jumlah mobil dan jumlah perjalanan per hari. Semakin banyak jumlah mobil, jumlah perjalanan per hari bertambah. Data di Zona 4 mengubah pola ini. Datum seperti ini dikenal sebagai *outlier*. Adanya *outlier* dapat disebabkan oleh sifat keragaman (variabilitas) sampel atau diakibatkan oleh kesalahan survei. Apabila penyebab *outlier* diketahui, maka perlakuan terhadapnya dapat ditentukan dengan mudah. Jika *outlier* disebabkan oleh kesalahan pengukuran, maka *outlier* dikeluarkan dari data dan tidak diikuti dalam pengolahan data. Sebaliknya, jika survei sudah benar, maka *outlier* tetap diikuti dalam pengolahan data.

Dengan asumsi bahwa *outlier* tersebut disebabkan oleh kesalahan survei, maka *outlier* dikeluarkan dari daftar data yang diolah.

Hubungan antara jumlah perjalanan per hari $Y (y_1, y_2, \dots, y_n)$ dan jumlah mobil $X (x_1, x_2, \dots, x_n)$ dapat diperoleh dengan teknik regresi metode kuadrat terkecil yang dinyatakan dalam persamaan linear berikut:

$$\hat{Y} = a_0 + a_1 X$$

Variabel \hat{Y} atau sering pula disimbolkan dengan Y_r adalah jumlah perjalanan per hari sebagai fungsi jumlah mobil X . Nilai a_0 dan a_1 dalam persamaan regresi dicari dengan persamaan berikut:

$$a_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad \text{dan} \quad a_0 = \bar{Y} - a_1 \bar{X}$$

Dalam persamaan di atas, n adalah jumlah data, \bar{Y} dan \bar{X} adalah jumlah perjalanan per hari rata-rata dan jumlah mobil rata-rata. Untuk menghemat penulisan, indeks pada operator penjumlahan tidak dituliskan, sehingga $\sum_{i=1}^n x_i$ dituliskan sebagai $\sum x_i$ dan $\sum_{i=1}^n y_i$ dituliskan sebagai $\sum y_i$.

Hitungan regresi linear dengan metode kuadrat terkecil disajikan pada Tabel 1 di bawah ini.

TABEL 1 HITUNGAN REGRESI LINEAR HUBUNGAN ANTARA JUMLAH PERJALANAN PER HARI DAN JUMLAH MOBIL DENGAN METODE KUADRAT TERKECIL

x_i	y_i	$x_i y_i$	x_i^2
220	760	167200	48400
300	850	255000	90000
400	1000	400000	160000
440	1100	484000	193600
500	1350	675000	250000
150	400	60000	22500
90	150	13500	8100
2100	5610	2054700	772600

Dari Tabel 1, diperoleh informasi sebagai berikut:

- jumlah pasang data, $n = 7$;
- jumlah perjalanan per hari rata-rata, $\bar{Y} = \sum y_i / n = 5610 / 7 = 801$;
- jumlah mobil rata-rata, $\bar{X} = \sum x_i / n = 2100 / 7 = 300$.

Koefisien a_1 dan a_0 pada persamaan kurva regresi dihitung sebagai berikut:

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{7 \times 2054700 - 2100 \times 5610}{7 \times 772600 - 2100^2} = 2.6066.$$

$$a_0 = \bar{Y} - a_1 \bar{X} = 801 - 2.6066 \times 300 = 19.0224.$$

Perhatikan bahwa jumlah mobil dan jumlah perjalanan per hari adalah variabel diskrit sehingga nilai rata-rata jumlah mobil dan jumlah perjalanan per hari adalah bilangan bulat. Nilai koefisien a_0

dan a_1 tidak harus bilangan bulat, tetapi nilai \hat{Y} dan Y haruslah bilangan bulat. Hubungan antara jumlah perjalanan per hari dan jumlah mobil yang diperoleh dari regresi linear antara kedua variabel adalah:

$$\hat{Y} = 19.0224 + 2.6066 X \text{ atau } \hat{y}_i = 19.0224 + 2.6066 x_i.$$

(b) Koefisien korelasi [bobot nilai 10%]

Koefisien korelasi, r , dinyatakan dalam persamaan berikut:

$$r = \frac{S_t - S_r}{S_t} = \frac{\sqrt{\sum(y_i - \bar{Y})^2 - \sum(y_i - \hat{y}_i)^2}}{\sqrt{\sum(y_i - \bar{Y})^2}} \text{ atau } r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Dalam persamaan di atas, operator penjumlahan $\sum x_i$ dibaca $\sum_{i=1}^n x_i$ dan indeks $i = 1, 2, \dots, n$.

Hitungan untuk mendapatkan nilai S_t dan nilai S_r dilakukan secara tabulasi dalam Tabel 2 di bawah ini. Hitungan mengacu kepada persamaan r di atas yang di sebelah kiri.

TABEL 2 HITUNGAN KOEFISIEN KORELASI ANTARA JUMLAH PERJALANAN PER HARI DAN JUMLAH MOBIL

x_i	y_i	$(y_i - \bar{Y})^2$	\hat{y}_i	$(y_i - \hat{y}_i)^2$
220	760	1681	592	28224
300	850	2401	801	2401
400	1000	39601	1062	3844
440	1100	89401	1166	4356
500	1350	301401	1322	784
150	400	160801	410	100
90	150	423801	254	10816
$S_t =$		1019087	$S_r =$	
			50525	

$$r = \frac{S_t - S_r}{S_t} = \frac{1019087 - 50525}{1019087} = 0.975.$$

Akar kuadrat dapat bernilai positif atau negatif. Koefisien korelasi dapat bernilai positif atau negatif. Karena gradien kurva regresi, a_1 , bernilai positif, atau dengan kata lain jumlah perjalanan per hari berbanding lurus dengan jumlah mobil, maka koefisien korelasi bernilai positif, $r = 0.975$.

(c) Hubungan linear antara jumlah mobil dan jumlah perjalanan per hari di delapan zona [bobot nilai 10%]

Jumlah perjalanan per hari dan jumlah mobil menunjukkan hubungan linear yang erat, walau linearitas hubungan antara keduanya tidak sepenuhnya sempurna. Nilai koefisien korelasi 0.975 sangat mendukung simpulan bahwa jumlah perjalanan per hari berbanding lurus dengan jumlah mobil.

SOAL 2 [SO B-4, BOBOT NILAI 60%]

Dari hasil pengujian suatu campuran beton, diperoleh data yang ditunjukkan pada tabel berikut (MPa = megapascal):

Nomor data	1	2	3	4	5	6	7	8	9	10
Mutu [MPa]	26	30	28	30	32	30	32	28	33	30
Nomor data	11	12	13	14	15	16	17	18	19	20
Mutu [MPa]	30	28	29	32	31	28	30	32	35	26

- (a) Berapakah nilai rata-rata, median, dan modus mutu beton tersebut? [Bobot 10%]
 (b) Berapakah nilai simpangan baku mutu beton tersebut? [Bobot 10%]
 (c) Tunjukkan rentang keyakinan mutu beton dengan tingkat keyakinan 90% dan 95%. [Bobot 20%]
 (d) Jika dinyatakan bahwa campuran tersebut menghasilkan mutu beton 27.5 MPa pada tingkat keyakinan 80%, apa pendapat Anda? [Bobot 20%]

PENYELESAIAN

(a) Nilai rata-rata, median, dan modus [bobot nilai 10%]

Data mutu beton tersebut adalah data sampel, bukan data populasi. Tabel di bawah ini menyajikan 20 nilai mutu beton yang diurutkan dari nilai mutu terendah (terkecil) sampai tertinggi (terbesar).

Mutu [MPa]	26	26	28	28	28	28	29	30	30	30
	30	30	30	31	32	32	32	32	33	35

Mutu beton rata-rata adalah:

$$\bar{X} = \frac{\sum x_i}{n} = \frac{600}{20} = 30 \text{ MPa}$$

Nilai median adalah nilai data yang berada di tengah dalam deret data yang diurutkan dari kecil ke besar atau dari besar ke kecil. Karena jumlah data adalah 20, maka nilai median adalah nilai yang berada di tengah antara data ke-10 dan ke-11. Dari tabel data yang telah diurutkan, tampak bahwa data ke-10 dan ke-11 adalah 30 Mpa sehingga nilai median mutu beton adalah 30 MPa.

$$X_{median} = \frac{30 + 30}{2} = 30 \text{ MPa}$$

Nilai modus adalah nilai data yang memiliki frekuensi tertinggi, yaitu mutu beton 30 Mpa yang berjumlah 6 buah sampel,

$$X_{modus} = 30 \text{ MPa}$$

Tampak bahwa nilai rata-rata, median, dan modus mutu beton adalah sama, yaitu 30 MPa. Kesamaan ketiga nilai ini merupakan salah satu sifat data yang berdistribusi normal.

(b) Simpangan baku [bobot nilai 10%]

Simpangan baku jumlah penumpang bus dihitung dengan bantuan tabel frekuensi, yaitu dengan menambahkan kolom yang berisi $f_i x_i^2$. Nilai simpangan baku adalah:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{100}{20 - 1}} = 2.2942 \text{ MPa}$$

(c) Rentang keyakinan mutu beton rata-rata [bobot nilai 20%]

Rentang keyakinan mutu rata-rata, dengan asumsi bahwa mutu beton tersebut berdistribusi normal, dinyatakan dengan persamaan berikut:

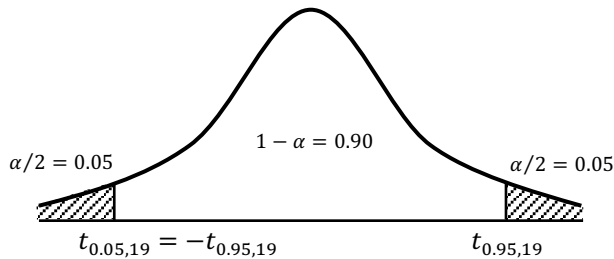
$$\text{prob}(l \leq \mu_x \leq u) = 1 - \alpha$$

Dalam persamaan di atas, l adalah batas bawah rentang keyakinan, u adalah batas atas rentang keyakinan, dan $1 - \alpha$ adalah tingkat keyakinan. Batas bawah dan batas atas rentang keyakinan mutu beton rata-rata dinyatakan dengan persamaan berikut:

$$l = \bar{X} - \frac{s_x}{\sqrt{n}} t_{1-\alpha/2, n-1} \quad \text{dan} \quad u = \bar{X} + \frac{s_x}{\sqrt{n}} t_{1-\alpha/2, n-1}$$

Nilai $t_{1-\alpha/2, n-1}$ adalah nilai t pada pdf distribusi t sedemikian hingga $\text{prob}(T < t) = 1 - \alpha/2$ pada nilai derajat kebebasan $\nu = n - 1$, dan n ukuran sampel (jumlah data). Tingkat keyakinan telah ditetapkan, yaitu $1 - \alpha = 90\%$ dan $1 - \alpha = 95\%$.

Untuk tingkat keyakinan $1 - \alpha = 90\%$, maka $1 - \alpha/2 = 95\%$. Nilai $t_{1-\alpha/2, n-1} = t_{0.95, 19}$ dibaca pada tabel distribusi t . Bacaan tabel menjadi mudah dilakukan dengan cara membuat sketsa pdf distribusi t . Dari tabel distribusi t , diperoleh:



$$t_{0.95, 19} = 1.7293$$

Dengan demikian, batas bawah dan batas atas rentang adalah:

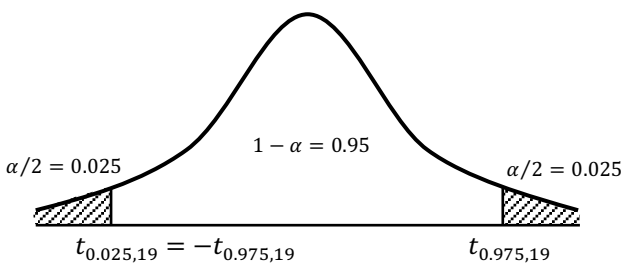
$$l = 30 - 1.7293 \frac{2.2942}{\sqrt{20}} = 28.90 \text{ MPa}$$

$$u = 30 + 1.7293 \frac{2.2942}{\sqrt{20}} = 30.70 \text{ MPa}$$

Jadi, rentang keyakinan 90% mutu beton rata-rata adalah:

$$\text{prob}(28.90 \text{ MPa} \leq \mu_x \leq 30.70 \text{ MPa}) = 0.90.$$

Untuk tingkat keyakinan $1 - \alpha = 95\%$, maka $1 - \alpha/2 = 97.5\%$.



$$t_{0.975, 19} = 2.0930$$

Dengan demikian, batas bawah dan batas atas rentang adalah:

$$l = 30 - 2.0930 \frac{2.2942}{\sqrt{20}} = 28.70 \text{ MPa}$$

$$u = 30 + 2.0930 \frac{2.2942}{\sqrt{20}} = 30.90 \text{ MPa}$$

Jadi, rentang keyakinan 90% mutu beton rata-rata adalah:

$$\text{prob}(28.70 \text{ MPa} \leq \mu_x \leq 30.90 \text{ MPa}) = 0.95.$$

(d) Uji hipotesis jumlah penumpang bus rata-rata [bobot nilai 20%]

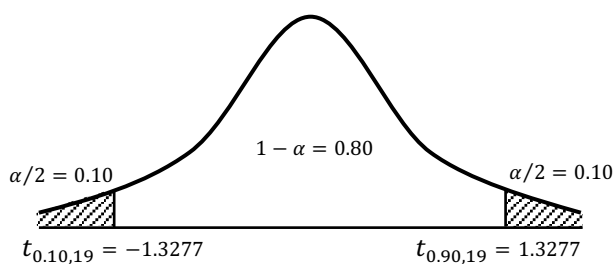
$$H_0: \mu_0 = 27.5 \text{ MPa}$$

$$H_1: \mu_0 \neq 27.5 \text{ MPa}$$

Karena varians populasi tidak diketahui (σ_x^2 tidak diketahui), maka statistika uji adalah:

$$T = \frac{\sqrt{n}}{s_x} (\bar{X} - \mu_0) = \frac{\sqrt{20}}{2.2942} (30 - 27.5) = 4.8013$$

Batas-batas penerimaan atau penolakan statistika uji dengan tingkat keyakinan $1 - \alpha = 80\%$ dan jumlah sampel $n = 20$ adalah: $t_{\alpha/2, n-1} = t_{0.10, 19}$ dan $t_{1-\alpha/2, n-1} = t_{0.90, 19}$.



Dari tabel distribusi t , diperoleh:

$$t_{0.90, 19} = 1.3277 \text{ dan}$$

$$t_{0.10, 19} = -1.3277$$

Dengan demikian, statistika uji $T = 4.8013$ berada di luar rentang penerimaan hipotesis H_0 ($|T| > t_{0.90, 19}$),

sehingga hipotesis yang menyatakan bahwa mutu beton rata-rata adalah 27.5 MPa tidak diterima atau ditolak.

-o0o-