

UJIAN AKHIR SEMESTER STATISTIKA DAN PROBABILITAS

Senin, 11 Desember 2017 | 100 menit
[Boleh membuka buku | Tidak boleh memakai komputer]

SOAL 1 [SO A-3, BOBOT NILAI 50%]

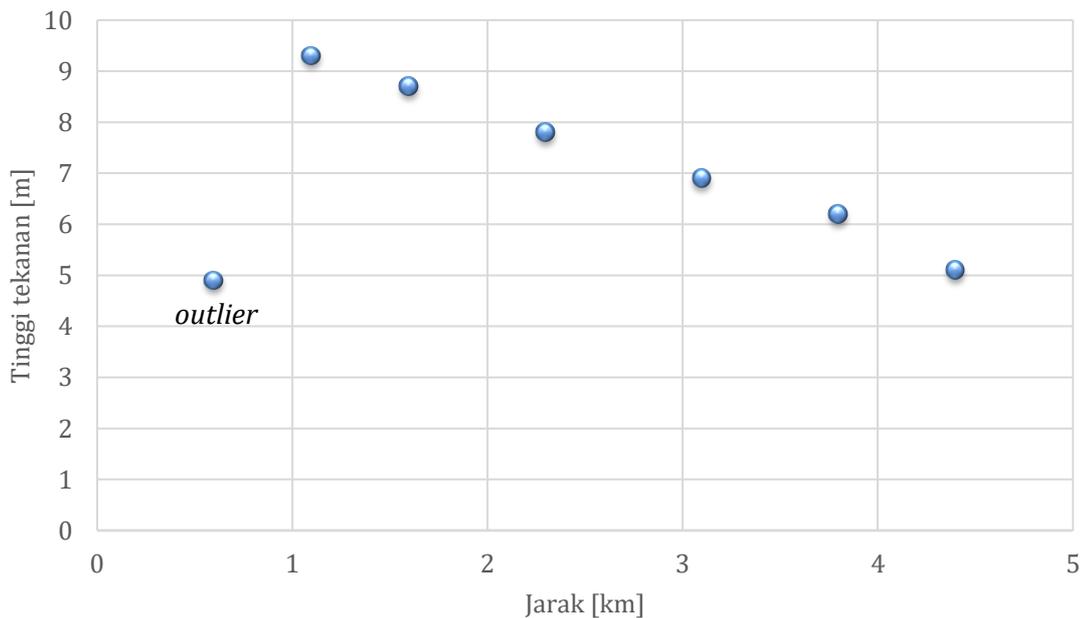
Sebuah PDAM melakukan pengukuran tinggi tekanan air di pipa distribusi utama. Hasil pengukuran disajikan dalam tabel di bawah ini.

Nomor data	Jarak [km]	Tinggi tekanan [m]
1	0.6	4.9
2	1.1	9.3
3	1.6	8.7
4	2.3	7.8
5	3.1	6.9
6	3.8	6.2
7	4.4	5.1

- Buatlah grafik yang menampilkan data tersebut (*scatter plot*). [Bobot 5%]
- Perhatikan salah satu pasang data yang tampak berbeda dari pasang data yang lain. Apa pendapat Saudara terhadap pasang data ini? Apa yang menurut Saudara perlu dilakukan? [Bobot 5%]
- Temukanlah persamaan tinggi tekanan sebagai fungsi jarak dengan teknik regresi linear, metode kuadrat terkecil. [Bobot 20%]
- Berapakah koefisien korelasi hubungan linear kedua variabel tersebut? [Bobot 20%]

PENYELESAIAN

(a) Grafik data [bobot nilai 5%]



(b) Outlier [bobot nilai 5%]

Dari gambar tampak bahwa salah satu pasang data berada pada posisi yang menyimpang dari pola data yang lain. Di titik stasiun ini, yaitu pada jarak 0.6 kilometer, tinggi tekanan adalah 4.9 meter. Pola data secara keseluruhan menunjukkan tekanan tinggi di stasiun yang berjarak dekat. Tekanan turun seiring dengan pertambahan jarak. Pada jarak 0.6 meter, justru tekanan lebih rendah daripada tekanan di tempat lain. Datum seperti ini dikenal sebagai *outlier*. Adanya *outlier* dapat disebabkan oleh sifat keragaman (variabilitas) sampel atau diakibatkan oleh kesalahan pengukuran. 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 pengukuran sudah benar, maka *outlier* tetap diikuti dalam pengolahan data.

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

(c) Regresi linear [bobot nilai 20%]

Hubungan antara tinggi tekanan $H (h_i = h_1, h_2, \dots, h_n)$ dan jarak $X (x_i = x_1, x_2, \dots, x_n)$ dapat diperoleh dengan teknik regresi metode kuadrat terkecil yang dinyatakan dalam persamaan linear berikut:

$$\hat{H} = a_0 + a_1X$$

Variabel \hat{H} atau sering pula disimbolkan dengan H_r adalah tinggi tekanan dalam satuan meter sebagai fungsi jarak X dalam satuan kilometer. Nilai a_0 dan a_1 dalam persamaan regresi dicari dengan persamaan berikut:

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

Dalam persamaan di atas, n adalah jumlah data, \bar{H} dan \bar{X} adalah tinggi tekanan rata-rata dan jarak rata-rata. Untuk menghemat penulisan, indeks pada operator penjumlahan tidak dituliskan, sehingga $\sum_{i=1}^n x_i$ dituliskan sebagai $\sum x_i$ dan $\sum_{i=1}^n h_i$ dituliskan sebagai $\sum h_i$.

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

TABEL 1 HITUNGAN REGRESI LINEAR HUBUNGAN ANTARA TINGGI TEKanan DAN JARAK STASIUN DENGAN METODE KUADRAT TERKECIL

i	x_i [km]	h_i [m]	$x_i h_i$ [km.m]	x_i^2 [km ²]
1	1.1	9.3	10.23	1.21
2	1.6	8.7	13.92	2.56
3	2.3	7.8	17.94	5.29
4	3.1	6.9	21.39	9.61
5	3.8	6.2	23.56	14.44
6	4.4	5.1	22.44	19.36
Σ	16.3	44	109.48	52.47

Dari Tabel 1, diperoleh informasi sebagai berikut:

- jumlah pasang data, $n = 6$;
- tinggi tekanan rata-rata, $\bar{H} = \sum h_i / n = 44 / 6 = 7.33$ m;
- jarak rata-rata, $\bar{X} = \sum x_i / n = 16.3 / 6 = 2.72$ km.

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

$$a_1 = \frac{n \sum x_i h_i - \sum x_i \sum h_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{6 \times 109.48 - 16.3 \times 44}{6 \times 52.47 - 16.3^2} = -1.23 \text{ m/km.}$$

$$a_0 = \bar{H} - a_1 \bar{X} = 7.33 + 1.23 \times 2.72 = 10.67 \text{ m.}$$

Perhatikan bahwa jarak stasiun pengukuran dan tinggi tekanan memiliki satuan. Koefisien a_0 bersatuan [m] dan a_1 bersatuan [m/km]. Hubungan antara tinggi tekanan dan jarak stasiun yang diperoleh dari regresi linear antara kedua variabel adalah:

$$\hat{H} = 10.67 - 1.23 X \text{ atau } \hat{h}_i = 10.67 - 1.23 x_i.$$

Tinggi tekanan H dalam satuan meter dan jarak stasiun X dalam satuan kilometer.

(d) Koefisien korelasi [bobot nilai 20%]

Koefisien korelasi, r , dinyatakan dalam persamaan berikut:

$$r = \frac{S_t - S_r}{S_t} = \frac{\sum(h_i - \bar{H})^2 - \sum(h_i - \hat{h}_i)^2}{\sum(h_i - \bar{H})^2} \text{ atau } r = \frac{n \sum x_i h_i - (\sum x_i)(\sum h_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum h_i^2 - (\sum h_i)^2}}$$

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 TINGGI BADAN DAN USIA SISWA

i	x_i [km]	h_i [m]	$(h_i - \bar{H})^2$ [m ²]	\hat{h}_i [m]	$(h_i - \hat{h}_i)^2$ [m ²]
1	1.1	9.3	3.87	9.3	0.00
2	1.6	8.7	1.87	8.7	0.00
3	2.3	7.8	0.22	7.8	0.00
4	3.1	6.9	0.19	6.9	0.00
5	3.8	6.2	1.28	6.0	0.04
6	4.4	5.1	4.99	5.3	0.04
$S_t =$			12.41	$S_r =$	0.08

$$r = \frac{S_t - S_r}{S_t} = \frac{12.41 - 0.08}{12.41} = -0.997.$$

Akar kuadrat dapat bernilai positif atau negatif. Koefisien korelasi dapat bernilai positif atau negatif. Karena gradien kurva regresi, a_1 , bernilai negatif, atau dengan kata lain tinggi tekanan berbanding terbalik dengan jarak stasiun, maka koefisien korelasi bernilai negatif, $r = -0.997$.

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

Tabel di bawah ini adalah data jumlah penumpang bus suatu PO dari Yogyakarta ke Cilacap untuk keberangkatan tengah hari selama bulan November 2017. Data disajikan sebagai rasio jumlah penumpang terhadap kapasitas tempat duduk (dinyatakan dalam persen, %).

Jumlah penumpang (%)	Frekuensi
50 - 60	2
60 - 70	7
70 - 80	10
80 - 90	8
90 - 100	3

- (a) Berapakah nilai rata-rata, median, dan modus jumlah penumpang? [Bobot 10%]
- (b) Berapakah nilai simpangan baku jumlah penumpang? [Bobot 10%]

- (c) Tunjukkan rentang keyakinan jumlah penumpang rata-rata populasi dengan tingkat keyakinan 80%. [Bobot 20%]
- (d) Ujilah hipotesis yang menyatakan bahwa jumlah penumpang rata-rata populasi adalah 80% dengan tingkat keyakinan 90%. [Bobot 10%]

PENYELESAIAN

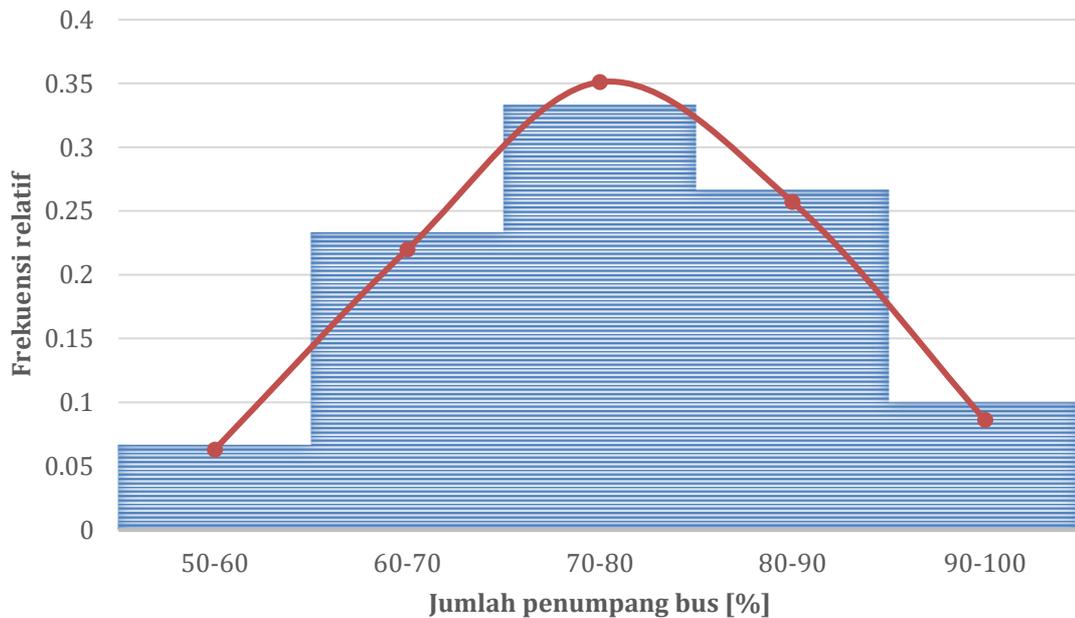
(a) Nilai rata-rata, median, dan modus [bobot nilai 10%]

Data jumlah penumpang bus tersebut adalah data sampel, bukan data populasi. Tabel frekuensi disajikan pada Tabel 3 di bawah ini. Jumlah penumpang bus disimbolkan dengan notasi X.

TABEL 3 JUMLAH PENUMPANG BUS SEBUAH PO DARI YOGYAKARTA KE CILACAP UNTUK JADWAL KEBERANGKATAN TENGAH HARI SELAMA NOVEMBER 2017

<i>i</i>	Jumlah penumpang, X [%] Kelas	x_i	Frekuensi f_i	Frekuensi relatif	$f_i x_i$ [%]	$f_i x_i^2$ [% ²]
1	50 – 60	55	2	0.067	110	6050
2	60 – 70	65	7	0.233	455	29575
3	70 – 80	75	10	0.333	750	56250
4	80 – 90	85	8	0.267	680	57800
5	90 – 100	95	3	0.100	285	27075
$\Sigma =$			30	1	2280	176750

Jumlah data dalam sampel adalah $\Sigma f_i = 30$. Operator penjumlahan Σf_i dibaca $\Sigma_{i=1}^5 f_i$. Data jumlah penumpang bus dalam tabel frekuensi di atas dapat pula disajikan dalam bentuk grafik batang atau histogram seperti disajikan pada Gambar 1.



GAMBAR 1 DISTRIBUSI PENUMPANG BUS SEBUAH PO DARI YOGYAKARTA KE CILACAP UNTUK JADWAL KEBERANGKATAN TENGAH HARI SELAMA NOVEMBER 2017

Grafik ini tidak wajib dibuat karena soal tidak memintanya. Perhatikan bentuk kurva pada gambar tersebut. Tampak jelas bahwa bentuk kurva mirip dengan kurva pdf distribusi normal. Dengan demikian, sampel jumlah penumpang bus tersebut berdistribusi normal.

Nilai rata-rata dihitung dengan bantuan tabel frekuensi, yaitu dengan menambahkan satu kolom yang berisi nilai frekuensi dikalikan dengan nilai data, $f_i x_i$.

Kelembaban udara rata-rata adalah:

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2280\%}{30} = 76\%$$

Nilai rata-rata dapat pula dibaca pada histogram (Gambar 1). Karena histogram data jumlah penumpang bus mirip dengan kurva pdf distribusi normal, maka jumlah penumpang bus rata-rata berada di tengah, yaitu dalam kelas 70-80 [%].

Nilai median adalah nilai data yang berada di tengah dalam deret data yang diurutkan dari kecil ke besar atau dari besar ke kecil. Tabel 3 telah mengatur data dalam deret dari kecil ke besar. Karena jumlah data adalah 30, maka nilai median adalah nilai yang berada di tengah antara data ke-15 dan ke-16. Dari histogram data (Gambar 1) dan kolom frekuensi pada tabel frekuensi data (Tabel 3), tampak bahwa data berdistribusi secara simetris dengan sumbu simetri kelas 70-80 [%]. Nilai median jumlah penumpang bus, dengan demikian, adalah di antara 70% s.d. 80%. Nilai median dapat pula dihitung dengan persamaan berikut:

$$X_{median} = x_l + \left(\frac{\frac{n}{2} - \sum_{i=1}^{m-1} f_i}{f_m} \right) (x_u - x_l)$$

Dalam persamaan di atas, x_l adalah batas bawah kelas yang mengandung nilai median, x_u adalah batas atas kelas yang mengandung nilai median, m adalah nomor urut kelas yang mengandung nilai median, dan f adalah frekuensi data.

$$X_{median} = 70 + \left(\frac{15 - 9}{10} \right) (80 - 70) = 70 + \frac{6}{10} 10 = 76\%.$$

Nilai modus adalah nilai data yang memiliki frekuensi tertinggi, yaitu kelas data 70-80 [%]. Nilai modus dapat pula dihitung dengan persamaan berikut:

$$X_{modus} = x_l + \left\{ \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \right\} (x_u - x_l)$$

Dalam persamaan di atas, x_l adalah batas bawah kelas yang mengandung nilai modus, x_u adalah batas atas kelas yang mengandung nilai modus, m adalah nomor urut kelas yang mengandung nilai modus, dan f adalah frekuensi data.

$$X_{modus} = 70 + \left\{ \frac{10 - 7}{(10 - 7) + (10 - 8)} \right\} (80 - 70) = 70 + \frac{3}{3 + 2} 10 = 76\%.$$

Tampak bahwa nilai rata-rata, median, dan modus kelembaban udara adalah sama, yaitu 76%. 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:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{X})^2}{(\sum f_i) - 1}} = \sqrt{\frac{\sum (f_i x_i^2) - (\sum f_i)(\bar{X})^2}{(\sum f_i) - 1}} = \sqrt{\frac{176750 - 30 \times 76^2}{30 - 1}} = 10.94\%.$$

(c) Rentang keyakinan jumlah penumpang bus rata-rata [bobot nilai 20%]

Rentang keyakinan jumlah penumpang bus rata-rata, dengan asumsi bahwa jumlah penumpang bus tersebut berdistribusi normal, dinyatakan dengan persamaan berikut:

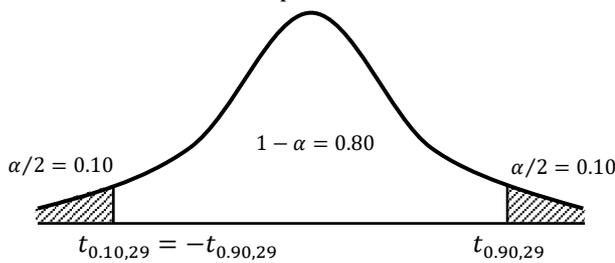
$$\text{prob}(l \leq \mu_x \leq u) = 1 - \alpha$$

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 jumlah penumpang bus 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). Karena tingkat keyakinan telah ditetapkan, yaitu $1 - \alpha = 80\%$, maka $1 - \alpha/2 = 90\%$. Nilai $t_{1-\alpha/2, n-1} = t_{0.90, 29}$ dibaca pada tabel distribusi t . Bacaan tabel menjadi mudah dilakukan dengan cara membuat sketsa pdf distribusi t .

Dari tabel distribusi t , diperoleh:



$$t_{0.90, 29} = 1.3114$$

Dengan demikian, batas bawah dan batas atas rentang adalah:

$$l = 76 - 1.3114 \frac{10.94}{\sqrt{30}} = 73.40\%$$

$$u = 76 + 1.3114 \frac{10.94}{\sqrt{30}} = 78.60\%$$

Jadi, rentang keyakinan 80% jumlah penumpang bus rata-rata adalah:

$$\text{prob}(73.40\% \leq \mu_x \leq 78.60\%) = 0.80.$$

(d) Uji hipotesis jumlah penumpang bus rata-rata [bobot nilai 10%]

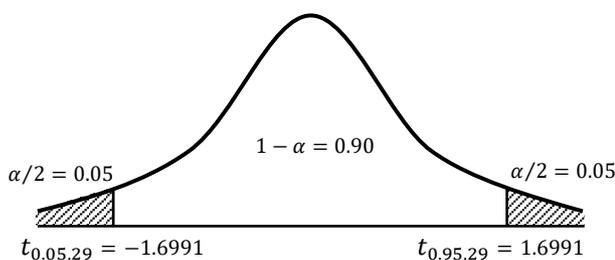
$$H_0: \mu_0 = 80\%$$

$$H_1: \mu_0 \neq 80\%$$

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

$$T = \frac{\sqrt{n}}{S_X} (\bar{X} - \mu_0) = \frac{\sqrt{30}}{10.94} (76 - 80) = -2.0029.$$

Batas-batas penerimaan atau penolakan statistika uji dengan tingkat keyakinan $1 - \alpha = 90\%$ dan jumlah sampel $n = 30$ adalah: $t_{\alpha/2, n-1} = t_{0.05, 29}$ dan $t_{1-\alpha/2, n-1} = t_{0.95, 29}$.



Dari tabel distribusi t , diperoleh:

$$t_{0.95, 29} = 1.6991 \text{ dan}$$

$$t_{0.05, 29} = -1.6991$$

Dengan demikian, statistika uji $T = -2.0029$ berada di luar rentang penerimaan hipotesis H_0 ($|T| > t_{0.95, 29}$), sehingga hipotesis yang menyatakan

bahwa jumlah penumpang bus rata-rata adalah 80% tidak diterima atau ditolak.