

UJIAN AKHIR SEMESTER STATISTIKA DAN PROBABILITAS

Senin, 19 Desember 2016 | 100 menit
[Boleh membuka buku | Tidak boleh memakai komputer]

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

Hasil pengukuran sampel di beberapa sekolah dan universitas terhadap kondisi fisik siswa/mahasiswa menghasilkan data amatan seperti disajikan pada tabel di bawah ini.

Nomor data	Usia [tahun]	Tinggi badan [cm]
1	12	155
2	19	169
3	18	171
4	20	173
5	17	167
6	16	161
7	15	158
8	14	170
9	13	156

- Temukanlah persamaan hubungan antara kedua variabel 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%]

Jika n adalah ukuran sampel, usia siswa/mahasiswa dinyatakan sebagai variabel X ($x_i = x_1, x_2, \dots, x_n$) dan tinggi badan siswa dinyatakan sebagai variabel Y ($y_i = y_1, y_2, \dots, y_n$), maka hubungan antara kedua variabel, yang diperoleh dari regresi linear, dapat dinyatakan dalam persamaan linear di bawah ini:

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

Variabel \hat{Y} atau sering pula disimbolkan dengan Y_r adalah tinggi badan siswa sebagai fungsi usia siswa. 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 tinggi badan rata-rata dan usia rata-rata. Untuk menghemat penulisan, indeks pada operator penjumlahan tidak dituliskan, sehingga $\sum_{i=1}^n x_i$ dituliskan sebagai $\sum x_i$.

Hitungan regresi linear dengan metode kuadrat terkecil disajikan pada Tabel 1 pada halaman setelah halaman ini. Dari Tabel 1, diperoleh informasi sebagai berikut:

- jumlah siswa, $n = 9$;
- tinggi badan rata-rata, $\bar{Y} = \sum y_i / n = 1480 / 9 = 164.4 \text{ cm}$;
- usia rata-rata, $\bar{X} = \sum x_i / n = 144 / 9 = 16 \text{ tahun}$.

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

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{9 \times 23802 - 144 \times 1480}{9 \times 2364 - 144^2} = 2.03 \text{ cm/tahun.}$$

$$a_0 = \bar{Y} - a_1 \bar{X} = 164.4 - 2.03 \times 16 = 131.92 \text{ cm.}$$

Perhatikan bahwa usia dan tinggi badan siswa memiliki satuan. Koefisien a_0 bersatuan [cm] dan a_1 bersatuan [cm/tahun]. Hubungan antara tinggi badan dan usia siswa yang diperoleh dari regresi linear antara kedua variabel adalah:

$$\hat{Y} = 131.92 + 2.03 X \text{ atau } \hat{y}_i = 131.92 + 2.03 x_i.$$

TABEL 1 HITUNGAN REGRESI LINEAR HUBUNGAN ANTARA TINGGI BADAN DAN USIA SISWA DENGAN METODE KUADRAT TERKECIL

i	x_i [tahun]	y_i [cm]	$x_i y_i$ [tahun.cm]	x_i^2 [tahun ²]
1	12	155	1860	144
2	19	169	3211	361
3	18	171	3078	324
4	20	173	3460	400
5	17	167	2839	289
6	16	161	2576	256
7	15	158	2370	225
8	14	170	2380	196
9	13	156	2028	169
Σ	144	1480	23802	2364

(b) Koefisien korelasi [bobot nilai 10%]

Koefisien korelasi, r , dinyatakan dalam persamaan berikut:

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

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.

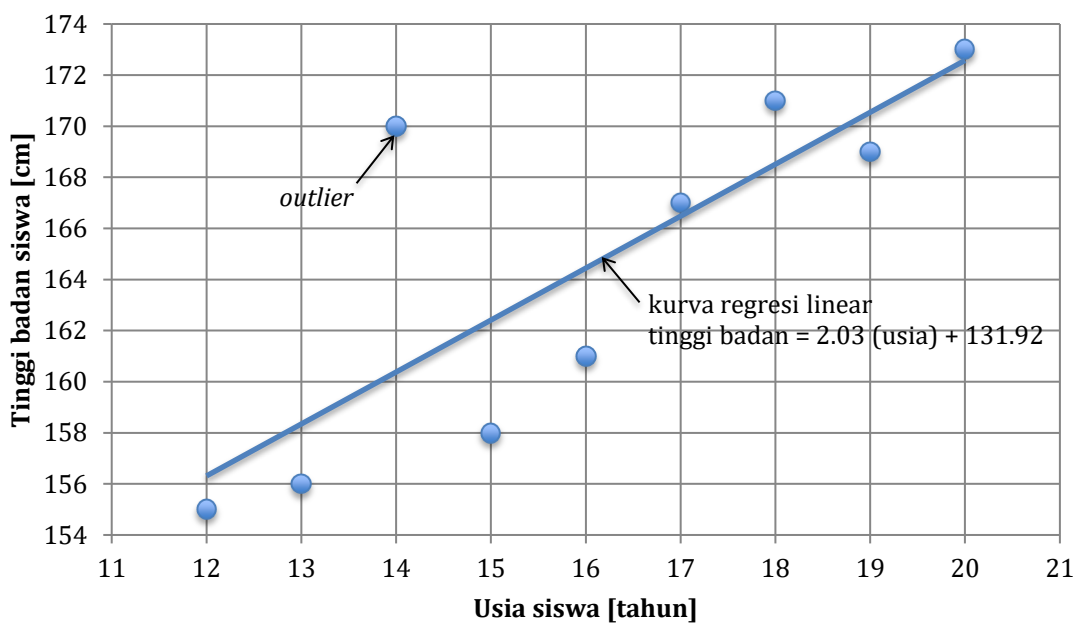
TABEL 2 HITUNGAN KOEFISIEN KORELASI ANTARA TINGGI BADAN DAN USIA SISWA

i	x_i [tahun]	y_i [cm]	$(y_i - \bar{Y})^2$ [cm ²]	\hat{y}_i [cm]	$(y_i - \hat{y}_i)^2$ [cm ²]
1	12	155	88.36	156.3	1.69
2	19	169	21.16	170.5	2.25
3	18	171	43.56	168.5	6.25
4	20	173	73.96	172.5	0.25
5	17	167	6.76	166.4	0.36
6	16	161	11.56	164.4	11.56
7	15	158	40.96	162.4	19.36
8	14	170	31.36	160.3	94.09
9	13	156	70.56	158.3	5.29
		$S_t =$	388.24	$S_r =$	141.10

$$r = \frac{S_t - S_r}{S_t} = \frac{388.24 - 141.10}{388.24} = 0.80.$$

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 tinggi badan siswa berbanding lurus dengan usia siswa, maka koefisien korelasi bernilai positif, $r = 0.80$.

Gambar 1 menyajikan hubungan antara usia dan tinggi badan siswa secara grafis. Gambar ini tidak wajib dibuat karena soal tidak memintanya. Salah satu data tampak berada jauh dari kurva regresi linear. Si siswa ini berusia 14 tahun dan memiliki tinggi badan 170 centimeter. 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 diputuskan. 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.



GAMBAR 1 HUBUNGAN LINEAR ANTARA USIA SISWA DALAM SATUAN TAHUN DAN TINGGI BADAN SISWA DALAM SATUAN CENTIMETER

(c) Hubungan antara usia dan tinggi badan [bobot nilai 10%]

Tinggi badan dan usia siswa menunjukkan hubungan linear yang erat, walau linearitas hubungan antara kedua variabel tidak sepenuhnya sempurna. Nilai koefisien korelasi 0.80 cukup mendukung simpulan yang menyatakan bahwa tinggi badan siswa berbanding lurus dengan usia siswa. Ini tampak jelas pada tampilan grafis titik-titik data (Gambar 1). Keberadaan sebuah *outlier* (14,170) mengurangi keeratan hubungan linear antara kedua variabel.

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

Angka-angka di bawah ini adalah sampel kelembaban udara relatif, dalam satuan persen, yang diperoleh dari sebuah stasiun cuaca.

88	92	79	86	81	77	82	83
87	94	81	90	85	70	84	78
74	85	75	85	75	78	80	82
78	76	90	85	71	83	90	73
87	89	77	81	94	78	84	81

- (a) Buatlah tabel frekuensi dengan rentang kelas 4%, batas bawah rentang kelas pertama adalah 68% (rentang kelas pertama 68-72). [Bobot 10%]
- (b) Hitunglah nilai rata-rata, median, dan modus berdasarkan sampel kelembaban udara dengan memakai tabel frekuensi. [Bobot 10%]
- (c) Hitunglah nilai simpangan baku sampel kelembaban udara dengan memakai tabel frekuensi. [Bobot 10%]
- (d) Hitunglah rentang keyakinan kelembaban udara rata-rata populasi dengan tingkat keyakinan 95%. [Bobot 15%]
- (e) Ujilah hipotesis yang menyatakan bahwa kelembaban udara rata-rata populasi adalah 80% dengan tingkat keyakinan 90%. [Bobot 15%]

PENYELESAIAN

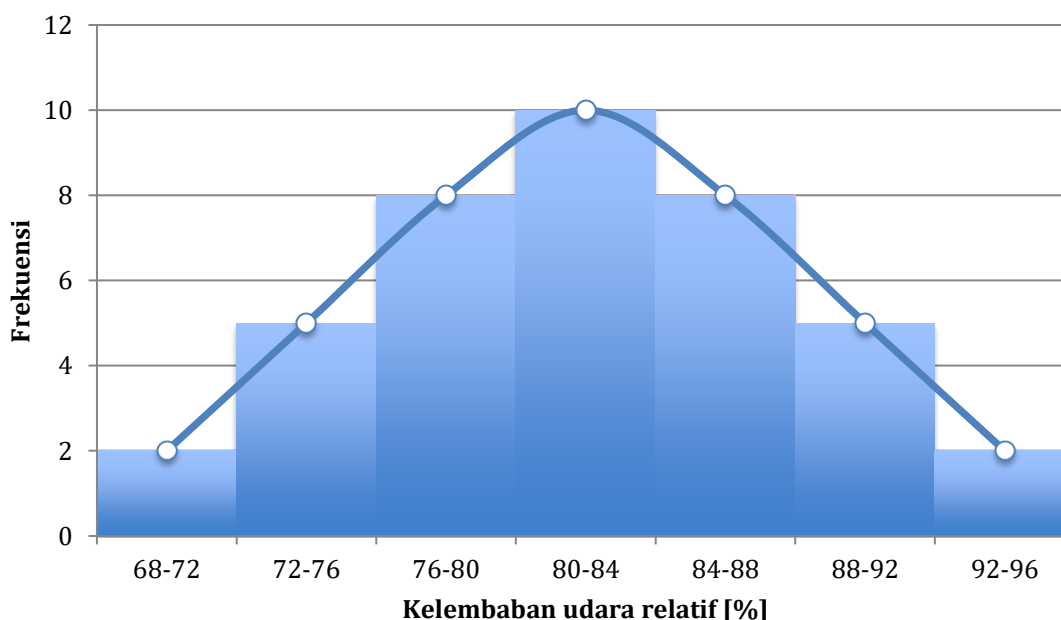
(a) Tabel frekuensi [bobot nilai 10%]

Kelembaban udara adalah variabel random kontinu. Data kelembaban udara tersebut adalah data sampel, bukan data populasi. Tabel frekuensi disajikan pada Tabel 3 di bawah ini. Kelembaban udara disimbolkan dengan notasi X .

TABEL 3 KELEMBABAN UDARA RELATIF DI SEBUAH STASIUN CUACA

i	Kelembaban udara, X [%] Kelas	x_i	Frekuensi f_i	Frekuensi kumulatif	$f_i x_i$ [%]	$f_i x_i^2$ [% ²]
1	68 - 72	70	2	2	140	9800
2	72 - 76	74	5	7	370	27380
3	76 - 80	78	8	15	624	48672
4	80 - 84	82	10	25	820	67240
5	84 - 88	86	8	33	688	59168
6	88 - 92	90	5	38	450	40500
7	92 - 96	94	2	40	188	17672
$\Sigma =$			40		3280	270432

Jumlah data dalam sampel kelembaban udara adalah $\Sigma f_i = 40$. Operator penjumlahan Σf_i dibaca $\Sigma_{i=1}^7 f_i$. Data kelembaban udara dalam tabel frekuensi di atas dapat pula disajikan dalam bentuk grafik batang atau histogram seperti disajikan pada Gambar 2.



GAMBAR 2 KELEMBABAN UDARA RELATIF DI SUATU STASIUN CUACA

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 kelembaban udara tersebut berdistribusi normal.

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

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:

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{3280}{40} = 82\%$$

Nilai rata-rata dapat pula dibaca pada histogram (Gambar 2). Karena histogram data kelembaban udara mirip dengan kurva pdf distribusi normal, maka kelembaban udara rata-rata berada di tengah, yaitu dalam kelas 80-84 [%]. Jadi, kelembaban udara rata-rata adalah 82%.

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 40, maka nilai median adalah nilai yang berada di tengah antara data ke-20 dan ke-21. Dari histogram data (Gambar 2) dan kolom frekuensi pada tabel frekuensi data (Tabel 3), tampak bahwa data berdistribusi secara simetris dengan sumbu simetri kelas 80-84 [%]. Nilai median kelembaban udara, dengan demikian, adalah 82%. 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.

$$X_{median} = 80 + \left(\frac{20 - 15}{10} \right) (84 - 80) = 80 + \frac{5}{10} 4 = 82\%.$$

Nilai modus adalah nilai data yang memiliki frekuensi tertinggi, yaitu kelas data 80-84 [%]. Mengingat distribusi data adalah simetri, maka modus kelembaban udara adalah 82%. 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.

$$X_{modus} = 80 + \left\{ \frac{10 - 8}{(10 - 8) + (10 - 8)} \right\} (84 - 80) = 80 + \frac{2}{2 + 2} 4 = 82\%.$$

Tampak bahwa nilai rata-rata, median, dan modus kelembaban udara adalah sama, yaitu 82%. Kesamaan ketiga nilai ini merupakan salah satu sifat data yang berdistribusi normal.

(c) Simpangan baku [bobot nilai 10%]

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

$$s_x = \sqrt{\frac{\sum(x_i - \bar{X})^2}{(\sum f_i) - 1}} = \sqrt{\frac{\sum(f_i x_i^2) - (\sum f_i)(\bar{X})^2}{(\sum f_i) - 1}} = \sqrt{\frac{270432 - 40 \times 82^2}{40 - 1}} = 6.1\%$$

(d) Rentang keyakinan kelembaban udara rata-rata [bobot nilai 15%]

Rentang keyakinan kelembaban udara rata-rata, dengan asumsi bahwa kelembababn udara tersebut berdistribusi normal, dinyatakan dengan persamaan berikut:

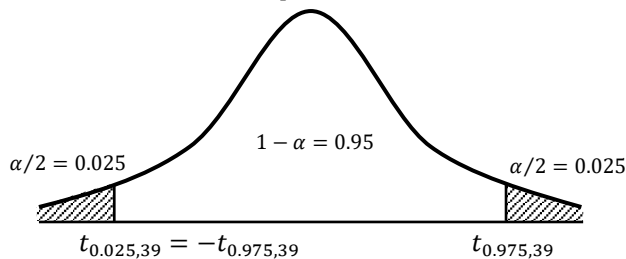
$$\text{prob}(l \leq \mu_x \leq u) = 1 - \alpha$$

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 kelembaban udara 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 $\nu = n - 1$, n ukuran sampel (jumlah data). Karena tingkat keyakinan telah ditetapkan, yaitu $1 - \alpha = 95\%$, maka $1 - \alpha/2 = 97.5\%$. Nilai $t_{1-\alpha/2, n-1} = t_{0.975, 39}$ dibaca pada tabel distribusi t . Bacaan tabel menjadi mudah dilakukan dengan cara membuat sketsa pdf distribusi t .

Dari tabel distribusi t , diperoleh:



$$t_{0.975, 39} = 2.0227$$

Dengan demikian, batas bawah dan batas atas rentang adalah:

$$l = 82 - 2.0227 \frac{6.1}{\sqrt{40}} = 80\%$$

$$u = 82 + 2.0227 \frac{6.1}{\sqrt{40}} = 84\%$$

Jadi, rentang keyakinan 95% kelembaban udara rata-rata adalah:

$$\text{prob}(80\% \leq \mu_x \leq 84\%) = 0.95.$$

(e) Uji hipotesis kelembaban udara rata-rata [bobot nilai 15%]

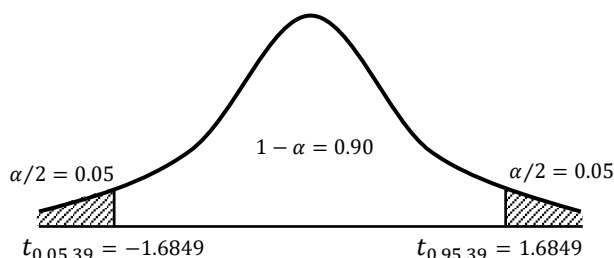
$$H_0: \mu_0 = 80\%$$

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

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

$$T = \frac{\sqrt{n}}{s_x} (\bar{X} - \mu_0) = \frac{\sqrt{40}}{6.1} (82 - 80) = 2.0736.$$

Batas-batas penerimaan atau penolakan statistika uji dengan tingkat keyakinan $1 - \alpha = 90\%$ dan jumlah sampel $n = 40$ adalah: $t_{\alpha/2, n-1} = t_{0.05, 39}$ dan $t_{1-\alpha/2, n-1} = t_{0.95, 39}$.



Dari tabel distribusi t , diperoleh:

$$t_{0.95, 39} = 1.6849 \text{ dan}$$

$$t_{0.05, 39} = -1.6849$$

Dengan demikian, statistika uji $T = 2.0736$ berada di luar rentang penerimaan hipotesis H_0 ($|T| > t_{0.95, 39}$),

sehingga hipotesis yang menyatakan bahwa kelembaban udara rata-rata adalah 80% tidak diterima atau ditolak.

-o0o-