

TUGAS ANALISIS REGRESI LINIER SEDERHANA K1/P1

Abie Septian Wibowo

G1401231069

1. Data

Data yang digunakan adalah data Tingkat Pengangguran Terbuka (TPT) dan Rata-rata Lama Sekolah (RLS) di beberapa kabupaten/kota di provinsi Jawa Barat tahun 2023. Terdapat 15 amatan dengan peubah Y adalah Tingkat Pengangguran Terbuka (TPT) dan peubah X adalah Rata-rata Lama Sekolah (RLS).

Kabupaten/Kota	TPT	RLS
Kabupaten Bandung	8.32	9.10
Kabupaten Tasikmalaya	6.16	7.96
Kabupaten Ciamis	5.06	8.09
Kabupaten Majalengka	5.71	7.52
Kabupaten Sumedang	9.18	8.73
Kabupaten Bekasi	10.09	9.57
Kota Bogor	11.79	10.64
Kota Sukabumi	10.78	10.37
Kota Bandung	11.46	11.06
Kota Cirebon	10.53	10.37
Kota Bekasi	10.88	11.66
Kota Depok	9.76	11.58
Kota Cimahi	13.07	11.39
Kota Tasikmalaya	7.66	9.54
Kota Banjar	6.09	8.79
Rataan	9.10	9.76
Simpangan baku	2.48	1.37

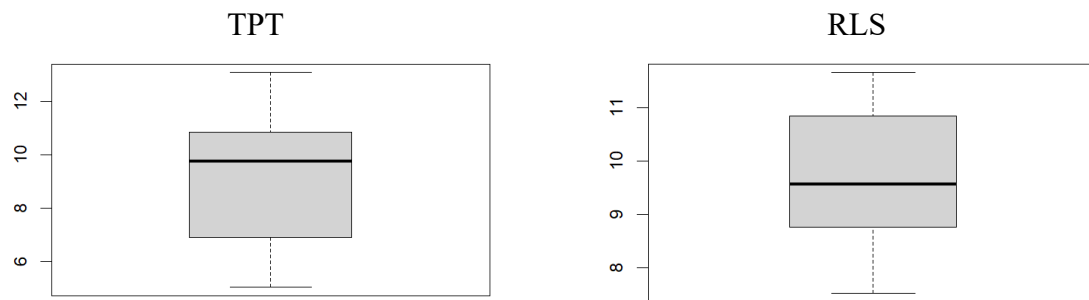
2. Eksplorasi Data

Beberapa eksplorasi yang akan dilakukan di antaranya:

- Statistik 5 serangkai

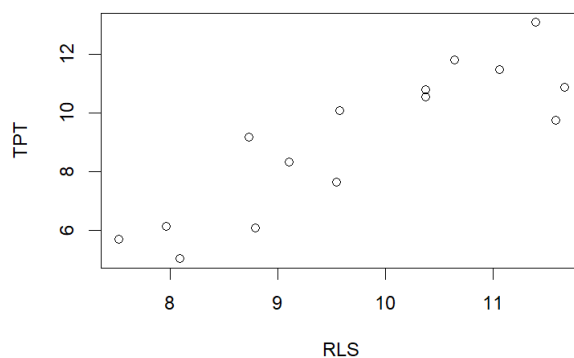
	Min	25%	Median	75%	Max
TPT	5.06	6.91	9.76	10.83	13.07
RLS	7.52	8.76	9.57	10.85	11.66

- *Boxplot*



Tidak ada pencilan baik dari peubah TPT (Y) maupun peubah RLS (X)

- *Scatter plot*



Dari grafik diatas tidak terlihat adanya pencilan multivariat. Selain itu, hubungan antar peubah tampak positif.

3. Persamaan Dugaan Regresi Linier

Untuk mendapatkan persamaan dugaan regresi diperlukan nilai koefisien regresi yaitu $\hat{\beta}_0$ dan $\hat{\beta}_1$ yang didapat dari rumus berikut:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})(y_i - \bar{y})} = \frac{s_{xy}}{s_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

y	x	$(y_i - \bar{y})$	$(x_i - \bar{x})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
8.32	9.1	-0.783	-0.658	0.613	0.433	0.515
6.16	7.96	-2.943	-1.798	8.659	3.233	5.291
5.06	8.09	-4.043	-1.668	16.343	2.782	6.743
5.71	7.52	-3.393	-2.238	11.510	5.009	7.593
9.18	8.73	0.077	-1.028	0.006	1.057	-0.079
10.09	9.57	0.987	-0.188	0.975	0.035	-0.186

11.79	10.64	2.687	0.882	7.222	0.778	2.370
10.78	10.37	1.677	0.612	2.813	0.375	1.027
11.46	11.06	2.357	1.302	5.557	1.695	3.069
10.53	10.37	1.427	0.612	2.037	0.375	0.874
10.88	11.66	1.777	1.902	3.159	3.618	3.380
9.76	11.58	0.657	1.822	0.432	3.320	1.198
13.07	11.39	3.967	1.632	15.740	2.663	6.475
7.66	9.54	-1.443	-0.218	2.081	0.048	0.315
6.09	8.79	-3.013	-0.968	9.076	0.937	2.916

Didapat

$$s_{xx} = \sum (x_i - \bar{x})^2 = 26.356$$

$$s_{yy} = \sum (y_i - \bar{y})^2 = 86.224$$

$$s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 41.500$$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{41.500}{26.356} = 1.574$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 9.10 - (1.574)(9.76) = -6.262$$

Didapat persamaan duga regresi $\hat{y}_i = -6.262 + 1.574x_i$. Artinya Ketika RLS berubah 1 satuan akan menyebabkan dugaan rata-rata TPT naik sebesar 1.574%, sementara dugaan rata-rata TPT ketika RLS-nya 0 tahun adalah sebesar -6.262%. Namun tidak terdapat RLS kabupaten/kota 0 tahun dalam selang data dan juga besar TPT tidak mungkin negatif, sehingga tidak dapat diinterpretasikan demikian.

4. Penguraian Sumber Keragaman

Terdapat dua sumber keragaman dalam model regresi, yaitu sumber keragaman yang dapat dijelaskan oleh model atau jumlah kuadrat regresi (JKR) dan sumber keragaman yang tidak dapat dijelaskan oleh model atau jumlah kuadrat galat (JKG), sehingga total sumber keragamannya atau jumlah kuadrat totalnya merupakan penjumlahan dari kedua sumber keragaman tersebut.

$$\mathbf{JKT} = \mathbf{JKR} - \mathbf{JKG}$$

Dengan

$$\mathbf{JKR} = \sum (\hat{y}_i - \bar{y})^2$$

$$\mathbf{JKG} = \sum (y_i - \hat{y}_i)^2$$

$$\mathbf{JKT} = \sum (y_i - \bar{y})^2$$

y	x	\hat{y}_i	$\hat{y}_i - \bar{y}$	$y_i - \hat{y}_i$	$y_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$
8.32	9.1	8.067	-1.036	0.253	-0.783	1.073	0.064	0.613
6.16	7.96	6.272	-2.831	-0.112	-2.943	8.015	0.012	8.659
5.06	8.09	6.476	-2.626	-1.416	-4.043	6.898	2.006	16.343
5.71	7.52	5.579	-3.524	0.131	-3.393	12.418	0.017	11.510
9.18	8.73	7.484	-1.619	1.696	0.077	2.620	2.876	0.006
10.09	9.57	8.807	-0.296	1.283	0.987	0.088	1.647	0.975
11.79	10.64	10.491	1.389	1.299	2.687	1.929	1.686	7.222
10.78	10.37	10.066	0.964	0.714	1.677	0.929	0.509	2.813
11.46	11.06	11.153	2.050	0.307	2.357	4.203	0.094	5.557
10.53	10.37	10.066	0.964	0.464	1.427	0.929	0.215	2.037
10.88	11.66	12.098	2.995	-1.218	1.777	8.969	1.482	3.159
9.76	11.58	11.972	2.869	-2.212	0.657	8.230	4.891	0.432
13.07	11.39	11.672	2.570	1.398	3.967	6.603	1.953	15.740
7.66	9.54	8.759	-0.343	-1.099	-1.443	0.118	1.209	2.081
6.09	8.79	7.578	-1.524	-1.488	-3.013	2.323	2.216	9.076

Didapat

$$\mathbf{JKR} = \sum(\hat{y}_i - \bar{y})^2 = 65.345$$

$$\mathbf{JKG} = \sum(y_i - \hat{y}_i)^2 = 20.879$$

$$\mathbf{JKT} = \sum(y_i - \bar{y})^2 = 86.224$$

Selain jumlah kuadrat, terdapat juga derajat bebas, kuadrat tengah, dan nilai F hitung. Derajat bebas adalah nilai yang digunakan untuk mengukur jumlah informasi independen yang tersedia untuk mengestimasi parameter model. Kuadrat tengah merupakan hasil bagi dari jumlah kuadrat dengan derajat bebasnya, sementara F hitung didapat dari pembagian antara kuadrat tengah regresi dengan kuadrat tengah galat. Pada regresi linier sederhana, rumus-rumus tersebut dapat dituliskan sebagai berikut

$$\mathbf{db_r} = 1$$

$$\mathbf{db_g} = n - 2 = 13$$

$$\mathbf{db_t} = n - 1 = 14$$

$$\mathbf{KTR} = \frac{JKR}{db_r} = 65.345$$

$$\mathbf{KTG} = \frac{JKG}{db_g} = \frac{20.879}{13} = 1.606$$

$$\mathbf{F_{hit}} = \frac{KTR}{KTG} = \frac{65.345}{1.606} = 40.686$$

Tabel sidik ragam

	db	JK	KT	Fhit
Regresi	1	65.345	65.345	40.686
Galat	13	20.879	1.606	

Total	14	86.224
-------	----	--------

5. Koefisien Determinasi

Koefisien determinasi (R^2) adalah ukuran statistik yang menunjukkan seberapa baik model regresi menjelaskan variabilitas data dependen (Y). Semakin tinggi R square, semakin bagus model regresinya karena lebih banyak variabilitas data dependen yang dapat dijelaskan oleh peubah X

$$R^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT} = \frac{65.345}{86.224} = \mathbf{0.7578}$$

Keragaman peubah TPT yang dapat dijelaskan oleh peubah RLS sebesar 75.78%, sementara sebesar 24.22% keragaman peubah TPT dijelaskan oleh peubah lain yang tidak disertakan dalam model.

6. Uji Signifikansi

- Uji Signifikansi Simultan (F)

Hipotesis uji:

H_0 : $\beta_1 = 0$ (tidak ada pengaruh yang signifikan antara RLS (x) terhadap TPT (y)).

H_1 : $\beta_1 \neq 0$ (ada pengaruh yang signifikan antara RLS (x) terhadap TPT (y)).

Statistik uji:

$$F_{hit} = \frac{KTR}{KTG} = \frac{65.345}{1.606} = \mathbf{40.686}$$

$$F_{tabel} = F_{(\alpha, db_r, db_g)} = F_{(0.05, 1, 13)} = \mathbf{4.667}$$

Keputusan uji:

Tolak H_0 karena $F_{hit}(40.686) > F_{tabel}(4.667)$. Peubah RLS memengaruhi TPT secara signifikan pada taraf nyata 5%

- Uji Signifikansi Parameter (t)

Parameter β_1

Hipotesis uji:

H_0 : $\beta_1 = 0$ (tidak ada pengaruh yang signifikan antara TPT terhadap RLS).

H_1 : $\beta_1 \neq 0$ (ada pengaruh yang signifikan antara TPT terhadap RLS).

Statistik uji:

$$S_{\hat{\beta}_1} = \sqrt{\frac{KTG}{s_{xx}}} = \sqrt{\frac{1.606}{26.356}} = 0.2468$$

$$t_{hit} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} = \frac{1.574}{0.2468} = 6.378$$

$$t_{tabel} = t_{(\frac{\alpha}{2}, db_g)} = 2.1603$$

Keputusan uji:

Tolak H_0 karena $|t_{hit}|(6.378) > t_{tabel}(2.1603)$. Peubah RLS memengaruhi TPT secara signifikan pada taraf nyata 5%

Parameter β_0

Hipotesis uji:

H_0 : $\beta_0 = 0$ (semua TPT dapat dijelaskan oleh RLS).

H_1 : $\beta_0 \neq 0$ (ada TPT yang tidak dapat dijelaskan oleh RLS).

Statistik uji:

$$S_{\hat{\beta}_0} = \sqrt{KTG\left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)} = \sqrt{1.606\left(\frac{1}{15} + \frac{(9.76)^2}{26.356}\right)} = \sqrt{1.606(3.681)} = 2.431$$

$$t_{hit} = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} = \frac{-6.262}{2.431} = -2.576$$

$$t_{tabel} = t_{(\frac{\alpha}{2}, db_g)} = 2.1603$$

Keputusan uji:

Tolak H_0 karena $|t_{hit}|(-2.576) > t_{tabel}(2.1603)$. Ada nilai TPT yang tidak dapat dijelaskan oleh RLS pada taraf nyata 5%.

7. Dugaan Selang Kepercayaan Paramater

selang kepercayaan untuk $\hat{\beta}_0$ dan $\hat{\beta}_1$ terdiri atas batas bawah dan batas atas yang dapat dituliskan sebagai berikut.

$$\hat{\beta}_0 - t_{(\frac{\alpha}{2}, db_g)} S_{\hat{\beta}_0} < \beta_0 < \hat{\beta}_0 + t_{(\frac{\alpha}{2}, db_g)} S_{\hat{\beta}_0}$$

$$\hat{\beta}_1 - t_{(\frac{\alpha}{2}, db_g)} S_{\hat{\beta}_1} < \beta_1 < \hat{\beta}_1 + t_{(\frac{\alpha}{2}, db_g)} S_{\hat{\beta}_1}$$

Selang kepercayaan 95% bagi β_0 :

$$-6.262 - (2.1603)(2.431) < \beta_0 < -6.262 + (2.1603)(2.431)$$

$$-6.262 - 5.252 < \beta_0 < -6.262 + 5.252$$

$$-11.514 < \beta_0 < -1.010$$

Pada tingkat kepercayaan 95%, selang kepercayaan bagi koefisien kemiringan garis berada di antara -11.514 hingga -1.010.

Selang kepercayaan 95% bagi β_1 :

$$1.574 - (2.1603)(0.2468) < \beta_1 < -6.262 + (2.1603)(0.2468)$$

$$1.574 - 0.533 < \beta_1 < 1.574 + 0.533$$

$$1.041 < \beta_1 < 2.107$$

Pada tingkat kepercayaan 95%, selang kepercayaan bagi intersep berada di antara 1.041 hingga 2.107.

8. Selang Kepercayaan Bagi Nilai Harapan

Akan diduga nilai rata-rata (harapan) amatan ketika nilai $x=10.69$ (4 angka terakhir NIM).

$$\hat{y}(x_0) = -6.262 + 1.574(10.69) = \mathbf{10.570}$$

$$\begin{aligned} \hat{y}(x_0) \pm t_{\left(\frac{\alpha}{2}, db_g\right)} \sqrt{KTG \left[\left(\frac{1}{n}\right) + \left(\frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right) \right]} \\ = 10.570 \pm 2.1603 \sqrt{(1.606) \left[\left(\frac{1}{15}\right) + \left(\frac{(10.69 - 9.76)^2}{26.356}\right) \right]} \\ = 10.570 \pm 2.1603(0.3997) \\ = 10.570 \pm 0.863 \\ = \mathbf{(9.706, 11.433)} \end{aligned}$$

Dugaan rata-rata nilai TPT ketika nilai RLS = 10.69 tahun adalah 10.57%. Selain itu, dalam taraf kepercayaan 95% diyakini bahwa nilai dugaan rata-rata nilai TPT Ketika nilai RLS = 10.69 berada dalam selang 9.706 hingga 11.433.

9. Selang Kepercayaan Bagi Nilai Individu

Akan diduga nilai individu amatan ketika nilai $x=10.69$ (4 angka terakhir NIM).

$$\hat{y}(x_0) = -6.262 + 1.574(10.69) = \mathbf{10.570}$$

$$\begin{aligned}
\hat{y}(x_0) \pm t_{\left(\frac{\alpha}{2}, db_g\right)} \sqrt{KTG \left[1 + \left(\frac{1}{n}\right) + \left(\frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right) \right]} \\
= 10.570 \pm 2.1603 \sqrt{(1.606) \left[1 + \left(\frac{1}{15}\right) + \left(\frac{(10.69 - 9.76)^2}{26.356}\right) \right]} \\
= 10.570 \pm 2.1603(1.3288) \\
= 10.570 \pm 2.871 \\
= (7.699, 13.441)
\end{aligned}$$

Dugaan nilai individu TPT ketika nilai RLS = 10.69 tahun adalah 10.57%. Selain itu, dalam taraf kepercayaan 95% diyakini bahwa nilai dugaan individu TPT Ketika nilai RLS = 10.69 berada dalam selang 7.699 hingga 13.441.

---oOo---

Lampiran Hasil R (fungsi lm)

1. Model (menggunakan `summary()`)

```
Call:
lm(formula = y ~ x, data = abcd)

Residuals:
    Min       1Q   Median       3Q      Max
-2.2115 -1.1585  0.2534  0.9985  1.6960

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.2620     2.4309  -2.576   0.023 *
x              1.5746     0.2469   6.379 2.42e-05 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.267 on 13 degrees of freedom
Multiple R-squared:  0.7579,    Adjusted R-squared:  0.7392
F-statistic: 40.69 on 1 and 13 DF, p-value: 2.424e-05
```

2. Tabel Sidik Ragam (menggunakan `anova()`)

```
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)    
x       1  65.345    65.345   40.686 2.424e-05 ***
Residuals 13  20.879     1.606
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


3. Dugaan Selang Kepercayaan Parameter (menggunakan `confint()`)

```
              2.5 %      97.5 %  
(Intercept) -11.513748 -1.010327  
x             1.041278  2.107873
```

4. Selang Kepercayaan Nilai Harapan (menggunakan `predict(,interval="confidence")`)

```
      fit      lwr      upr  
10.57017  9.706014 11.43433
```

5. Selang Kepercayaan Nilai Individu (menggunakan `predict(,interval="prediction")`)

```
      fit      lwr      upr  
10.57017  7.69917 13.44117
```