

Heteroskedastičnost

1. U tabeli su dati podaci o ukupnim izdacima za obrazovanje i BDP-u za uzorak od 24 zemlje:

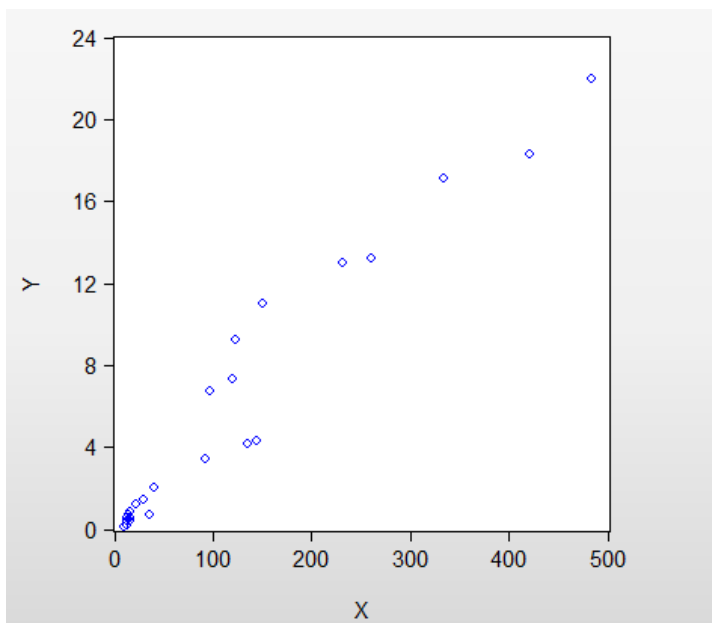
Y	10.97	0.12	0.39	17.10	1.21	0.40	1.45	0.52	1.99	3.43	18.27	6.70
X	151.27	10.35	16.25	334.29	22.38	12.92	30.35	16.61	39.90	92.60	420.79	97.62

7.31	9.27	0.19	0.54	4.17	4.34	0.72	13.02	0.85	13.25	0.70	21.96
120.17	122.93	12.98	13.75	135.96	145.17	14.39	232.01	15.63	261.39	36.67	483.65

- Nacrtati dijagram rasturanja tačaka i na osnovu njega prokomentarisati heteroskedastičnost.
- Ispitati postojanje heteroskedastičnosti Goldfeld-Quandtovim testom.
- Prema dobijenim rezultatima testiranja predložiti transformaciju modela u cilju eliminisanja heteroskedastičnosti ako je $\delta_i^2 = k^2 X_i^2$.

Rešenje:

a)



Pri nižim nivoima BDP izdaci za obrazovanje ne razlikuju se mnogo između posmatranih zemalja, dok sa rastom BDP dolazi i do rasta varijabiliteta izdataka za obrazovanje zemalja sa sličnim nivoom BDP. Očekujemo da je u modelu prisutna heteroskedastičnost.

b)

$$H_0: V(\varepsilon_i) = \delta^2$$

$$H_1: V(\varepsilon_i) = \delta_i^2$$

1.korak: Opservacije iz uzorka se poređaju prema rastućim vrednostima za X:

Y	0.12	0.40	0.19	0.54	0.72	0.85	0.39	0.52	1.21	1.45	0.70	1.99
X	10.35	12.92	12.98	13.75	14.39	15.63	16.25	16.61	22.38	30.35	36.67	39.90

3.43	6.70	7.31	9.27	4.17	4.34	10.97	13.02	13.25	17.10	18.27	21.96
92.60	97.62	120.17	122.93	135.96	145.17	151.27	232.01	261.39	334.29	420.79	483.65

2.korak: Izostavlja se c centralnih opservacija ($n/4$), u primeru je to šest opservacija.

3.korak: Pojedinačno ocenjujemo dve odvojene regresije na osnovu prvih devet ($\frac{n-c}{2}$) i poslednjih devet ($\frac{n-c}{2}$) opservacija.

4.korak: Dobijamo sume kvadrata reziduala iz dve ocenjene regresije.

1.regresija

	Y	X	y	x	yx	x ²	y ²
	0.12	10.35	-0.43	-4.68	2.00	21.88	0.18
	0.40	12.92	-0.15	-2.11	0.32	4.45	0.02
	0.19	12.98	-0.36	-2.04	0.73	4.18	0.13
	0.54	13.75	-0.01	-1.28	0.02	1.64	0.00
	0.72	14.39	0.17	-0.64	-0.11	0.41	0.03
	0.85	15.63	0.30	0.60	0.18	0.36	0.09
	0.39	16.25	-0.16	1.22	-0.19	1.50	0.02
	0.52	16.61	-0.03	1.58	-0.05	2.49	0.00
	1.21	22.38	0.66	7.35	4.88	54.07	0.44
Suma	4.95	135.25	0.00	0.00	7.78	90.97	0.92
Prosek	0.55	15.03					

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{7.78}{90.97} = 0.085$$

$$\Sigma e_{1i} = \Sigma y_i^2 - b \Sigma x_i y_i = 0.92 - 0.085 * 7.78 = \mathbf{0.255}$$

2.regresija	Y	X	y	x	yx	x2	y2
	9.27	122.93	-3.21	-131.23	421.90	17222.43	10.34
	4.17	135.96	-8.31	-118.20	982.15	13971.03	69.04
	4.34	145.17	-8.14	-108.99	887.74	11879.21	66.34
	10.97	151.27	-1.51	-102.89	155.51	10587.25	2.28
	13.02	232.01	0.54	-22.15	-11.91	490.79	0.29
	13.25	261.39	0.76	7.23	5.52	52.24	0.58
	17.10	334.29	4.61	80.13	369.59	6420.14	21.28
	18.27	420.79	5.79	166.63	964.89	27764.87	33.53
	21.96	483.65	9.48	229.49	2174.82	52666.43	89.81
Suma	112.34	2287.44	0	0	5950.21	141054.40	293.49
Prosek	12.48	254.16					

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{5950.21}{141054.40} = 0.04$$

$$\sum e_{2i} = \sum y_i^2 - b \sum x_i y_i = 293.49 - 0.04 * 5950.21 = \mathbf{42.49}$$

$$F = \frac{\sum e_{2i}^2}{\sum e_{1i}^2} = \frac{42.49}{0.255} = \mathbf{166.43}$$

$$F_7^7(\alpha = 0.05) = 3.87$$

$F > F_7^7(\alpha = 0.05)$ -> Na nivou značajnosti od 5% odbacujemo H_0 i zaključujemo da u modelu postoji heteroskedastičnost.

c)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$V(\varepsilon_i) = \delta_i^2 = k^2 X_i^2$$

U cilju eliminisanja heteroskedastičnosti delimo model sa X_i :

$$\frac{Y_i}{X_i} = \beta_0 \cdot \frac{1}{X_i} + \beta_1 + \frac{\varepsilon_i}{X_i}$$

$$\text{Novi model: } Y_i^* = \beta_0^* + \beta_1^* X_i^* + \varepsilon_i^*$$

$$\left(Y_i^* = \frac{Y_i}{X_i}, X_i^* = \frac{1}{X_i}, \beta_0^* = \beta_1, \beta_1^* = \beta_0 \right)$$

U novom modelu varijansa slučajne greške je stabilna:

$$V(\varepsilon_i^*) = V\left(\frac{\varepsilon_i}{X_i}\right) = \frac{V(\varepsilon_i)}{X_i^2} = \frac{k^2 X_i^2}{X_i^2} = k^2 = \text{const}$$

2. Na osnovu 500 podataka o zaradama (Y) i broju godina školovanja (X₁) ocenjen je sledeći model:

$$\hat{Y}_i = 0.76 + 1.26X_{1i} \quad , \quad R^2 = 0.1$$

(2.8) (0.18)

- a) Testirati postojanje heteroskedastičnosti u modelu ako je poznato $r_{|e|x} = 0.157$.
 b) Model je proširen sa novom objašnjavajućom promenljivom (X₂- godine radnog iskustva) i ocenjen:

$$\hat{Y}_i = -14.67 + 1.87X_{1i} + 0.98X_{2i} \quad , \quad R^2 = 0.15$$

(4.29) (0.22) (0.2)

Definisati pomoćnu regresiju za Whiteov test prilikom testiranja heteroskedastičnosti.

- c) Testirati postojanje heteroskedastičnosti u model pod b) ako se zna da je koeficijent determinacije pomoćne regresije Whiteov-og testa 0.05.

Rešenje:

a) $H_0: V(\varepsilon_i) = \delta^2$

$H_1: V(\varepsilon_i) = \delta_i^2$

Glejserov test

$|\hat{e}_i| = \hat{\delta}_0 + \hat{\delta}X_{1i}$

$r_{|e|x} = 0.157$

Podsećanje: U jednostavnoj regresiji važi da je kvadrat koeficijenta korelacije jednak koeficijentu determinacije.

$R^2 = r_{|e|x}^2 = 0.025$

$$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{0.025}{1}}{\frac{1-0.025}{500-2}} = 12.77$$

$F_{498}^1(\alpha = 0.05) = 3.84$

$F > F_{498}^1(\alpha = 0.05) \rightarrow$ Na nivou značajnosti od 5% odbacujemo H_0 i zaključujemo da u modelu postoji heteroskedastičnost.

b)

Pomoćna regresija:

$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_1^* X_{1i}^2 + \alpha_2^* X_{2i}^2 + \alpha_3^* X_{1i} X_{2i} + \text{greška}$

c)

$$H_0: V(\varepsilon_i) = \delta^2$$

$$H_1: V(\varepsilon_i) = \delta_i^2$$

$$WH = n \cdot R^2 = 500 \cdot 0.05 = 25$$

$$\chi_5^2(\alpha = 0.05) = 11.07$$

$WH > \chi_5^2(\alpha = 0.05)$ -> Na nivou značajnosti od 5% odbacujemo H_0 i zaključujemo da u modelu postoji heteroskedastičnost.