Y:	6,15	7,86	9,05	13,83	13,02	6,40	7,69		
	36,67	38,62	33,12	39,76	40,63	38,77	34,68	38,49	
	192,18	196,03	197,24	201,13	191,24	198,28	190,47	190,79	190,52
	566,53	561,31	566,17	566,83	562,37	557,16	566,08	565,08	
	2350,25	2348,15	2347,14	2344,92	2346,34	2346,58	2346,48		
	3947,96	3948,50	3945,06	3944,19	3949,17	3941,45	3941,86	3947,43	
	9037,24	9040,31	9035,98	9038,56	9040,33	9036,72	9037,75	9038,46	9036,45
	12721,09	12724,81	12725,66	12718,27	12725,52	12727,36	12722,09	12721,23	
	22855,06	22861,06	22853,33	22853,75	22856,98	22857,72	22857,45		

Examining a physical or economic law, i.e. the dependence of a quantity Y on X, we have done several experiments and measurements. For some carefully preselected values of the regressor X, i.e. the first nine prime numbers, we have measured the quantity Y several times, i.e. seven up to nine times for each value of X. Test the null hypothesis that the law under consideration is cubic, i.e. of the form $Y = \gamma_0 + \gamma_1 X + \gamma_2 X^2 + \gamma_3 X^3 + \varepsilon$ for some (unknown) parameters $\gamma_0, \gamma_1, \gamma_2, \gamma_3 \in \mathbb{R}$, where ε is random error; assume that the random error is normally distributed and that its variance is constant (homoskedasticity).

- Form the 71×71 design matrix $\boldsymbol{X}$ and find the estimates $b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9$ of the unknown parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$.
- Calculate the Residual Sum of Squares (RSS) and the Residual Variance s^2=RSS/(71−rank($\boldsymbol{X}$)).
- Form the 71×4 design matrix $\boldsymbol{F}$ and find the estimates g_0, g_1, g_2 g_3 of the unknown parameters $\gamma_0, \gamma_1, \gamma_2$ γ_3.
- Calculate the Residual Sum of Squares (RSS_h) and the Residual Variance s^2_h=RSS_h/(9−rank($\boldsymbol{F}$)).
- Calculate the statistic F = [(RSS_h−RSS)/RSS] / [(9−rank($\boldsymbol{F}$)/(71−rank($\boldsymbol{X}$)].
- At the significance level of α = 5 %, test the null hypothesis that the law under consideration is c

X:	2
	3
	5
	7
	11
	13
	17
	19
	23

ubic.

Y	X1	X2	X3
410,49	33,7	56,0	1,8
110,92	10,8	5,5	3,6
276,48	20,9	39,2	3,1
67,22	2,6	10,6	3,2
323,11	14,5	74,2	0,5
223,21	1,9	65,2	2,1
124,73	5,3	21,7	4,9
473,16	35,0	71,1	2,4
253,89	28,6	13,6	2,2
239,10	29,9	5,7	1,6
372,56	24,9	63,9	1,9
104,52	10,8	1,6	4,8
398,03	32,9	48,8	3,8
285,26	14,9	54,5	3,6
239,77	26,0	10,1	4,6
296,60	13,9	65,7	0,3
538,10	39,2	79,4	4,8
269,62	34,3	3,3	3,9
198,72	8,3	39,8	4,0
239,98	0,2	73,6	3,4
118,78	7,3	17,5	2,8
152,86	11,9	24,3	0,2
408,86	22,5	76,7	3,9
106,78	1,5	25,2	3,5
120,38	14,3	1,5	2,4
269,00	29,3	17,2	2,4
434,29	39,4	45,5	3,6
161,77	7,5	29,8	3,7
357,90	31,8	40,3	3,2
376,68	39,6	26,6	4,1
233,52	1,8	72,0	0,6
435,01	39,1	48,6	3,3
216,14	9,2	47,1	1,7
84,35	4,3	9,9	4,5
383,87	35,5	45,4	0,2
245,17	10,1	50,8	4,8
322,30	27,7	38,9	3,0
220,81	18,6	29,9	1,1
496,53	36,0	73,1	5,0
126,57	13,4	2,6	3,9
383,64	38,7	31,0	3,4
156,69	17,0	10,5	1,0

You are given a dataset of 50 observations of the proc
regressor X_1 = the workforce productivity.
regressor X_2 = the amount of labour.
regressor X_3 = the amount of capital.
regressand Y = the amount of production.
Assume (for simplicity) that the production depends linearly on the workforce productivity, the labour, and the capital, without the intercept term; that is, assume $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ where ε is rando
error; assume that the random error is normally distri
and that its variance is constant (homoskedasticity).

- Is multicollinearity present?

269,88	17,7	42,8	4,5
296,62	14,2	63,4	2,1
101,17	2,4	26,5	2,1
94,12	5,2	20,0	0,6
475,22	35,4	73,1	2,0
387,67	28,4	61,0	0,1
344,24	36,9	27,3	1,6
181,40	1,3	52,1	2,6

duction of a company:

om
ibuted

Y	X1	X2	X3
2561,95	56,5	48,6	425,9
1486,93	25,4	33,7	251,3
1358,99	29,7	25,8	227,2
1042,72	22,5	21,5	173,3
1460,85	55,6	11,0	230,7
1255,20	37,9	14,8	204,3
1949,24	33,4	42,7	330,3
2146,69	39,0	46,0	362,4
2149,62	51,2	40,1	354,2
1792,02	19,8	49,0	308,7
2000,50	30,8	49,0	340,4
2566,39	55,2	51,8	426,5
1811,29	32,7	40,1	304,5
1512,46	32,9	30,5	251,4
2394,85	57,7	42,8	395,5
1972,46	40,9	38,8	330,7
2005,57	27,1	51,7	342,3
1468,36	10,9	43,6	258,0
1012,06	12,7	24,1	174,9
2110,19	56,3	34,5	344,8
2337,13	43,9	51,2	392,4
752,78	19,3	12,6	125,0
2087,89	36,2	48,7	351,9
1843,97	58,7	21,7	295,8
973,44	27,6	13,8	160,2
1814,47	21,8	49,2	311,3
1985,90	23,2	53,9	341,6
2408,91	58,6	42,3	397,3
1409,80	50,3	14,2	223,0
1253,04	32,6	19,5	207,3
1694,28	50,8	24,0	272,9
2395,14	42,7	53,3	404,2
2059,95	41,4	42,0	344,8
1826,62	50,9	28,6	297,4
1719,13	52,9	24,1	277,5
941,79	17,6	19,9	159,1
1994,72	46,5	37,8	329,3
1920,22	30,9	45,2	326,6
2550,87	50,5	52,9	429,4
2329,50	54,4	43,6	385,2
2173,32	48,5	40,0	362,0
1578,46	47,5	19,9	257,2
2347,79	33,2	58,8	401,7
1381,47	29,0	27,1	230,6
1407,70	41,0	18,9	228,0

You are given a dataset of 60 observations of the proc
regressor X_1 = the workforce productivity.
regressor X_2 = the amount of labour.
regressor X_3 = the amount of capital.
regressand Y = the amount of production.
Assume (for simplicity) that the production depends linearly on the workforce productivity, the labour, and the capital, without the intercept term; that is, assume $Y = ?_1 X_1 + ?_2 X_2 + ?_3 X_3 + ?$ where ? is rando
error; assume that the random error is normally distri
and that its variance is constant (homoskedasticity).

- Is multicollinearity present?

1032,98	28,9	16,5	167,4
1903,33	18,1	52,4	330,6
2047,29	48,1	38,6	337,7
1898,89	59,3	23,8	306,8
1595,62	34,2	32,4	264,4
1128,69	32,3	16,6	183,4
2310,37	41,6	51,9	389,1
2532,82	59,8	45,2	418,6
1892,30	17,7	52,1	330,2
1602,57	20,7	39,8	275,0
2280,90	28,7	58,4	391,7
1336,26	18,4	33,1	229,2
2261,19	58,2	37,8	371,5
2343,06	48,0	47,9	391,9
1699,93	23,4	42,2	291,3

duction of a company:

om
ibuted

Solution to 4.A:
The underlying law here is
$Y = 7X_1 + 3X_2 + 5X_3 + \varepsilon$
with $\varepsilon \sim \mathcal{N}(0,10)$.
The columns of the design matrix $\boldsymbol{X}$ are pretty random.
No multicollinearity is present.

Solution to 4.B:
The underlying law here is
$Y = 5X_1 + 3X_2 + 5X_3 + \varepsilon$
with $\varepsilon \sim \mathcal{N}(0,10)$.
Here the columns of
the design matrix $\boldsymbol{X}$
fulfil the relation
$\boldsymbol{X}_3 = 7 + 3\boldsymbol{X}_1 + 5\boldsymbol{X}_2 + \delta$
with $\delta \sim \mathcal{U}(-10,+10)$.
Multicollinearity is present.