

Korelacije

Osmo predavanje

Sadržaj

1. Smisao i princip korelacije
2. Formule
3. Linearna bivarijantna regresija

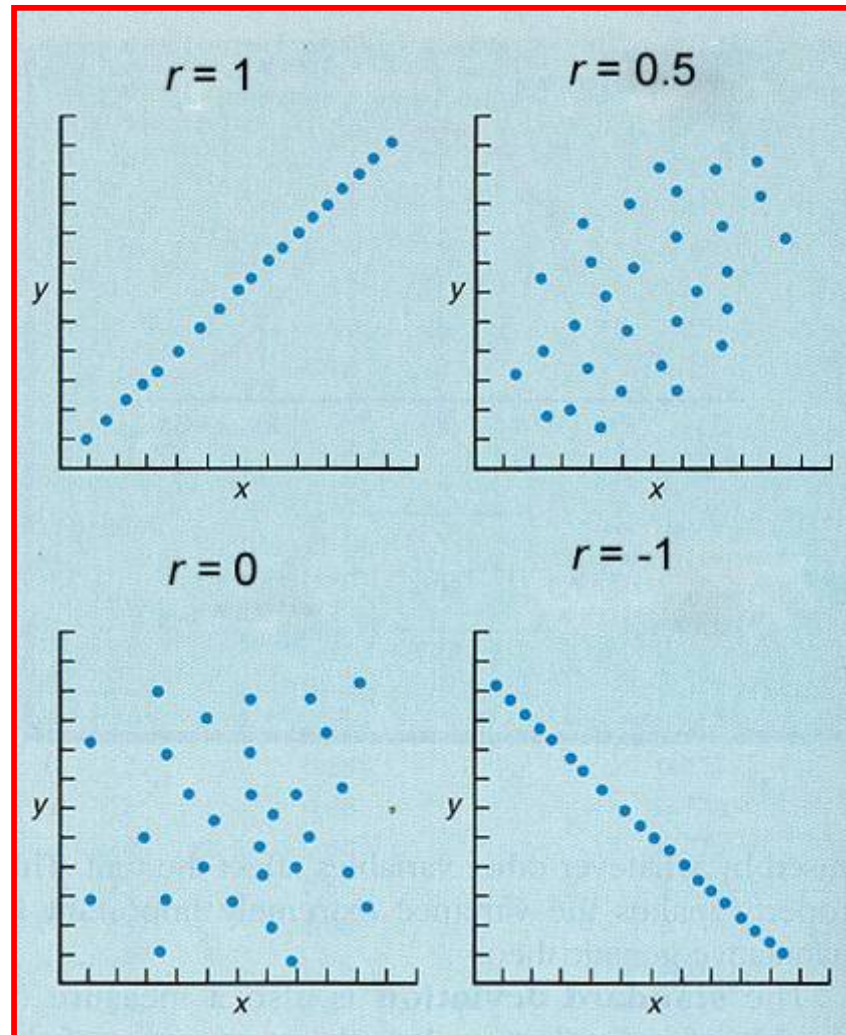
Smisao i princip korelacije

Brojna vrednost (koeficijent) koji predstavlja u kojoj meri su dve varijable međusobno povezane...

Koeficijent korelacije

Vrednosti od -1 do 1

Smisao i princip korelacije



Izračunavanje korelacionog koeficijenta

U slučaju većeg broja podataka u praksi se najčešće koristi sledeća formula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2} \sqrt{\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2}}$$

Izračunavanje korelacionog koeficijenta

...Odnosno računar... (vidi uputstvo za sedmu vežbu)

Značajnost korelacije

- Određen je brojem parova (N)
- Što je N manje to r mora biti veće da bi korelacija bila značajna (i obrnuto)
- Značajnost korelacija se određuje uz pomoć tablica ili upotrebom računarskih programa:

Značajnost korelacije

Pearson Product Moment Correlation Distribution [?] [X]

N: [▲] [▼] Two-tailed

r: [▲] [▼] Compute p from r

p: [▲] [▼] Compute p from p

Fisher z: [▲] [▼] Compute r from z

Značajnost korelacije

Korisna Web adresa. Dobar “udžbenik” statistike sa lepim ilustracijama (JAVA applets).

<http://www.stat.berkeley.edu/users/stark/Java/Html/Correlation.htm>

Bivarijantna regresija

Bivarijantna regresija

Kada je korelacija između dve varijable dovoljno “velika” (visoka), na osnovu rezultata varijable x (nezavisno promenljive) možemo predvideti rezultat varijable y (zavisno promenljive)...

Bivarijantna regresija

- To je od posebnog značaja u slučajevima kada nam je merenje varijable x znatno lakše nego merenje varijable y (*Npr. procena indirektno potrošnje kiseonika...*).

Bivarijantna regresija

$$y = mx + b$$

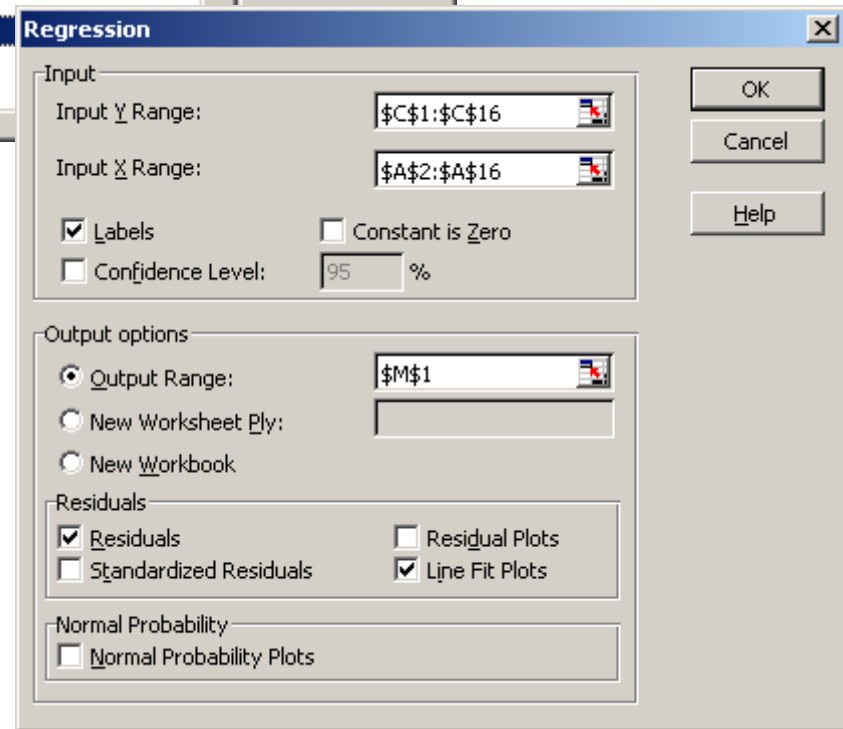
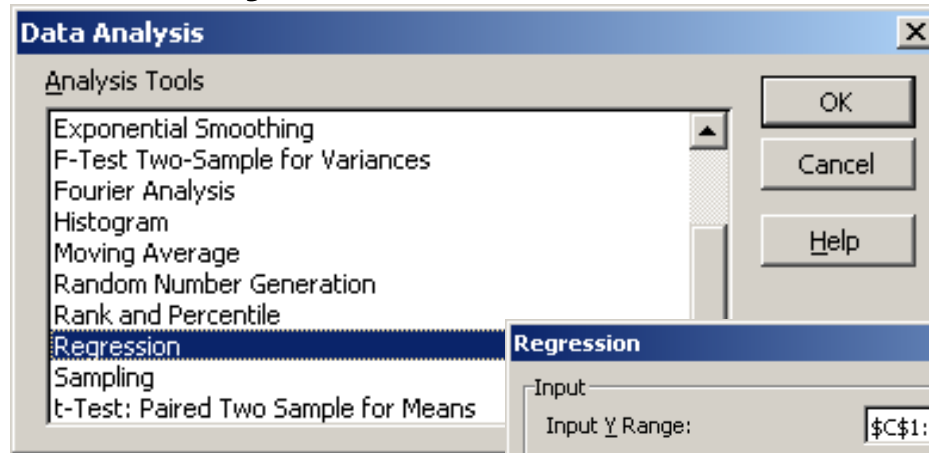
$$m = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}$$

Primer (dva merenja iste veličine u dva dana)

<i>X (day 1)</i>	<i>Y (day 2)</i>
72	68
70	64
69	58
67	61
61	62
58	51
51	55
49	44
47	43
43	60
42	46
41	38
40	36
40	37
35	40



<i>r</i>	<i>k</i>	<i>n</i>
0.845	0.726	12.86

$$Y = 0.726X + 12.86$$

Primer (dva merenja iste veličine u dva dana)

Microsoft Excel - Table 7-3 Page 110.xls

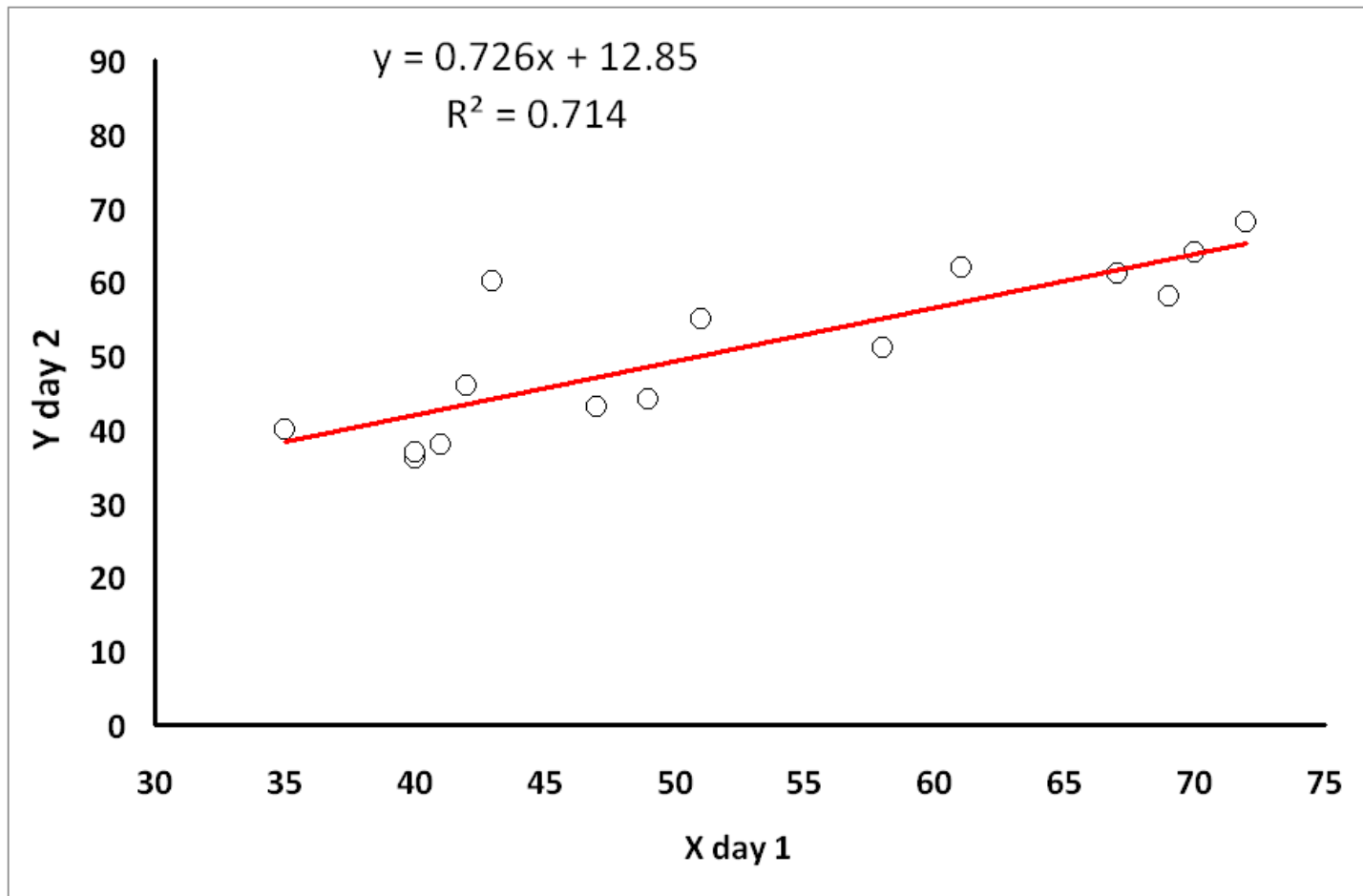
File Edit View Insert Format Tools Data Window Help Adobe PDF

100% Arial

P1

	P	Q	R	S	T	U	V	W	X	Y	Z
1			SUMMARY OUTPUT								
2											
3			<hr/> <i>Regression Statistics</i> <hr/>								
4			Multiple R	0.845							
5			R Square	0.715							
6			Adjusted R Square	0.693							
7			Standard Error	6.063							
8			Observations	15							
9			<hr/>								
10			ANOVA								
11				<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
12			Regression	1	1195.9	1195.9	32.5	0.0001			
13			Residual	13	477.8	36.8					
14			Total	14	1673.7						
15			<hr/>								
16				<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
17			<i>Intercept</i>	12.86	6.84	1.88	0.08	-1.93	27.65	-1.93	27.65
18			<i>X (day 1)</i>	0.73	0.13	5.70	0.00	0.45	1.00	0.45	1.00
19			<hr/>								
20			RESIDUAL OUTPUT								
21			<hr/>								
22			<i>Observation</i>	<i>Predicted Y (day 2)</i>		<i>Residuals</i>					
23			1	65.1		2.9					
24			2	63.7		0.3					
25			3	63.0		-5.0					
26			4	61.5		-0.5					
27			5	57.2		4.8					

Primer: (dva merenja iste veličine u dva dana)



Multipla regresija

- Ispituje uticaj dve ili više nezavisnih (prediktorskih) promenljivih na zavisno promenljivu...

$$Y = a + b_1 * X_1 + b_2 * X_2 + \dots + b_p * X_p$$

Multipla regresija

PRIMER*:

Posmatran je uticaj eksplozivne snage (x_1) i brzine trčanja (x_2) na skok u dalj iz zaleta y (10 studenata Fakulteta za fizičku kulturu...)

4	5	6
SDM (m) Y	TSM (m) X1	20m (s) X2
5.24	7.14	2.16
4.82	6.9	2.15
5.17	7.25	2.24
4.98	7.2	2.25
6.08	8.62	2.12
4.54	7.18	2.11
4.94	7.29	2.08
5.12	7.8	2.18
4.68	6.42	1.96
4.43	6.2	1.95

* Primer preuzet iz: *Peri D, Operacionalizacija 2, FINE Graf, 1996*

Multipla regresija

Effect	SDM (m) Y Param.	SDM (m) Y Std.Err	SDM (m) Y t	SDM (m) Y p	-95.00% Cnf.Lmt	+95.00% Cnf.Lmt
Intercept	0.883271	1.740026	0.507619	0.627319	-3.23124	4.997778
TSM (m) X1	0.618137	0.145266	4.255211	0.003769	0.27464	0.961636
20m (s) X2	-0.157479	0.960186	-0.164009	0.874359	-2.42796	2.113000

$$Y = 0.883271 + 0.6181 * X_1 - 0.1574 * X_2$$