



ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
AGRICULTURAL UNIVERSITY OF ATHENS

Πολλαπλή Συμμεταβολή

Κατσιλέρος Αναστάσιος

2017

ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΣΥΜΜΕΤΑΒΟΛΗ

Σε αρκετά πειράματα η εξαρτημένη μεταβλητή Y μπορεί να επηρεάζεται από περισσότερες από μια ανεξάρτητες μεταβλητές X .

Το γραμμικό μοντέλο για k ανεξάρτητες μεταβλητές και n παρατηρήσεις μπορεί να γραφεί ως εξής:

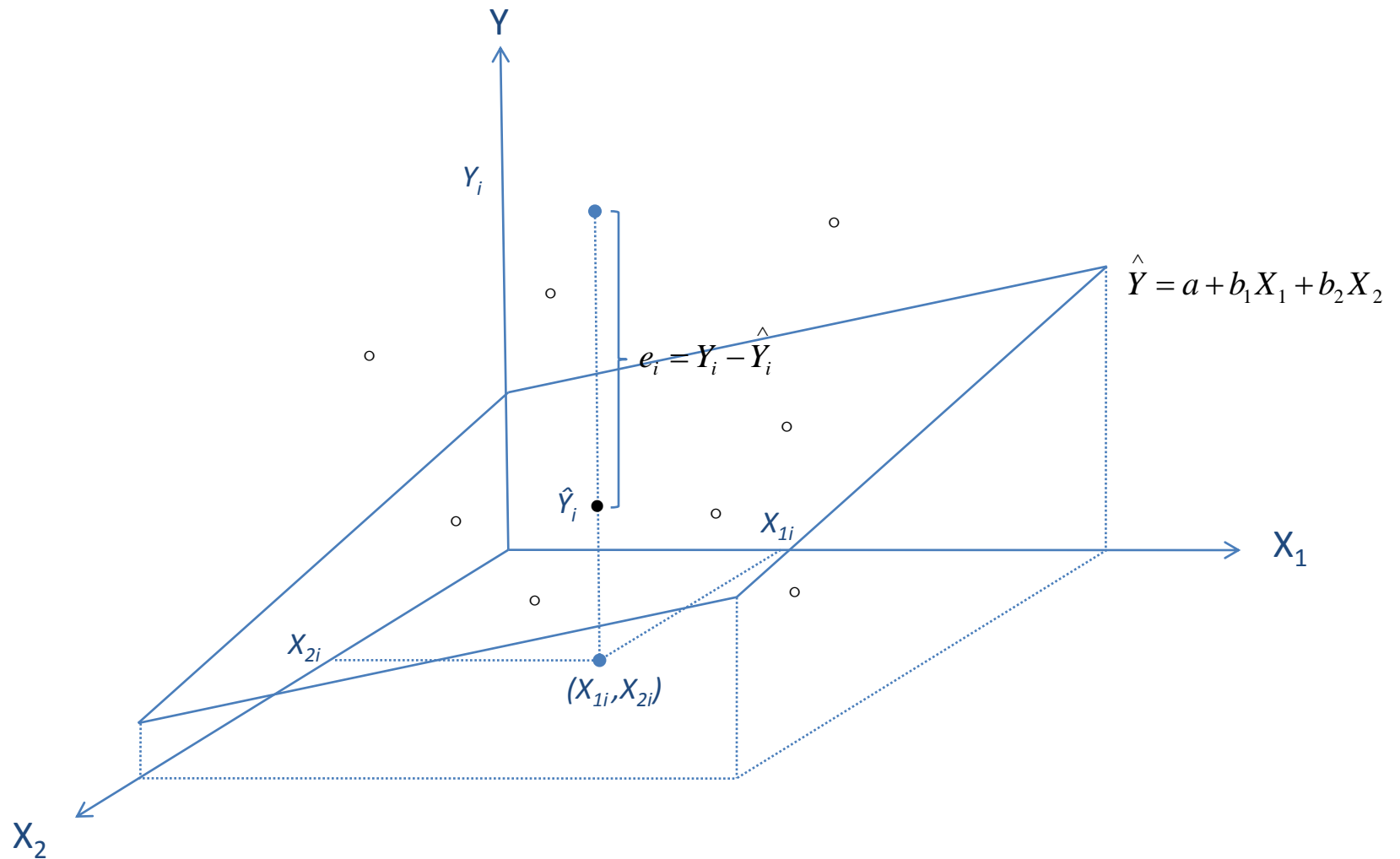
$$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki} + e_i \quad i = 1, 2, \dots, n$$

όπου e_i τα σφάλματα, τα οποία είναι ανεξάρτητα και $N(0, \sigma^2)$.

Έλεγχος για Πολυσυγγραμμικότητα

Στην πολλαπλή συμμεταβολή είναι πιθανόν να υπάρχουν ισχυρές συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών X_1, \dots, X_p .

Πειραματικοί Σχεδιασμοί



Πίνακας Ανάλυσης Πολλαπλής Συμμεταβολής

Πηγή παρ/τητας	ΒΕ	ΑΤ	ΜΤ	F
Συμμεταβολή	k	$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	$\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{k}$	$\frac{ΜΤ_{συμ}}{ΜΤ_{υπ}}$
Υπόλοιπο	n - k - 1	$\sum_{i=1}^n (Y_i - \hat{Y})^2$	$\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n - k - 1}$	
Σύνολο	n - 1	$\sum_{i=1}^n (Y_i - \bar{Y})^2$		

Πειραματικοί Σχεδιασμοί

Παράδειγμα

Oxy	Weight	Runtime	RunPulse	RstPulse	MaxPulse
59,571	68,15	8,17	166	40	172
60,055	81,87	8,63	170	48	186
54,297	85,84	8,65	156	45	168
54,625	70,87	8,92	146	48	155
49,156	81,42	8,95	180	44	185
49,874	89,02	9,22	178	55	180
48,673	76,32	9,4	186	56	188
45,441	76,32	9,63	164	48	166
50,545	59,08	9,93	148	49	155
46,672	77,91	10	162	48	168
45,313	75,07	10,07	185	62	185
50,388	73,37	10,08	168	67	168
50,541	73,03	10,13	168	45	168
46,774	91,63	10,25	162	48	164
51,855	83,12	10,33	166	50	170
45,79	73,71	10,47	186	59	188
47,467	82,78	10,5	170	53	172
47,273	79,15	10,6	162	47	164
49,091	81,19	10,85	162	64	170
40,836	69,63	10,95	168	57	172
45,118	67,25	11,08	172	48	172
44,754	66,45	11,12	176	51	176
46,08	79,38	11,17	156	62	165
44,609	89,47	11,37	178	62	182
47,92	61,24	11,5	170	52	176
44,811	77,45	11,63	176	58	176
45,681	75,98	11,95	176	70	180
39,407	73,37	12,63	174	58	176
39,203	91,63	12,88	168	44	172
39,442	81,42	13,08	174	63	176
37,388	87,66	14,03	186	56	192

Πειραματικοί Σχεδιασμοί

```
> attach(MultRegression)
> fitmult=lm(Oxy~Weight + Runtime + RunPulse + RstPulse + MaxPulse)
> library(car)
> vif(fitmult)
```

Weight	Runtime	RunPulse	RstPulse	MaxPulse
1.114716	1.373507	8.351864	1.343349	8.139587

Πειραματικοί Σχεδιασμοί

```
> library(agricolae)
> multcor=correlation(MultRegression[,1:6], method="pearson")
> multcor
```

\$correlation

	Oxy	Weight	Runtime	RunPulse	RstPulse	MaxPulse
Oxy	1.00	-0.16	-0.86	-0.40	-0.40	-0.24
Weight	-0.16	1.00	0.14	0.18	0.04	0.25
Runtime	-0.86	0.14	1.00	0.31	0.45	0.23
RunPulse	-0.40	0.18	0.31	1.00	0.35	0.93
RstPulse	-0.40	0.04	0.45	0.35	1.00	0.31
MaxPulse	-0.24	0.25	0.23	0.93	0.31	1.00

\$pvalue

	Oxy	Weight	Runtime	RunPulse	RstPulse	MaxPulse
Oxy	1,00E+06	0.3816911	4,59E-04	2,66E+04	0.02603592	2,00E+05
Weight	3,82E+05	1,00E+06	4,41E+05	3,28E+05	0.81429423	1,76E+05
Runtime	4,59E-04	0.4411926	1,00E+06	8,58E+04	0.01100668	2,21E+05
RunPulse	2,66E+04	0.3284411	8,58E+04	1,00E+06	0.05181166	4,09E-08
RstPulse	2,60E+04	0.8142942	1,10E+04	5,18E+04	1,00E+06	9,51E+04
MaxPulse	2,00E+05	0.1760793	2,21E+05	4,09E-08	0.09509402	1,00E+06

Πειραματικοί Σχεδιασμοί

> summary(fitmult)

Call:

lm(formula = Oxy ~ Weight + Runtime + RunPulse + RstPulse + MaxPulse)

Residuals:

Min	1Q	Median	3Q	Max
-5.8780	-1.3001	0.0914	1.3451	4.0598

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	82.39361	9.17856	8.977	2.70e-09 ***
Weight	-0.05091	0.05792	-0.879	0.38784
Runtime	-2.95182	0.38597	-7.648	5.29e-08 ***
RunPulse	-0.39704	0.12880	-3.083	0.00495 **
RstPulse	0.01239	0.06950	0.178	0.85995
MaxPulse	0.38479	0.14225	2.705	0.01212 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.503 on 25 degrees of freedom

Multiple R-squared: 0.8161, Adjusted R-squared: 0.7793

F-statistic: 22.19 on 5 and 25 DF, p-value: 1.861e-08

Πειραματικοί Σχεδιασμοί

```
> anova(fitmult)
```

Analysis of Variance Table

Response: Oxy

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Weight	1	22.55	22.55	3.6006	0.06937 .
Runtime	1	611.67	611.67	97.6589	4.07e-10 ***
RunPulse	1	14.38	14.38	2.2964	0.14221
RstPulse	1	0.36	0.36	0.0575	0.81238
MaxPulse	1	5.83	45.83	7.3172	0.01212 *
Residuals	25	56.58	6.26		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ΒΗΜΑΤΙΚΗ ΣΥΜΜΕΤΑΒΟΛΗ

(Stepwise regression)

Η βηματική συμμεταβολή είναι μια τεχνική επιλογής των σημαντικότερων μεταβλητών από ένα μεγάλο αριθμό ανεξάρτητων μεταβλητών. Για την επιλογή των μεταβλητών χρησιμοποιούνται διάφορα κριτήρια όπως R^2_{adj} , Akaike information criterion, Bayesian information criterion, Mallows's C_p .

Η επιλογή προς τα εμπρός (Forward Selection). Το μοντέλο ξεκινάει χωρίς μεταβλητές και εξετάζει αν η προσθήκη κάθε μεταβλητής δίνει στατιστικά σημαντική βελτίωση της προσαρμογής. Η διαδικασία επαναλαμβάνεται μέχρι η προσθήκη μεταβλητών στο μοντέλο δεν το βελτιώνει στατιστικά σημαντικά.

Η διαγραφή προς τα πίσω (Backward Elimination). Το μοντέλο περιλαμβάνει όλες τις υποψήφιες μεταβλητές και εξετάζει αν η διαγραφή κάθε μεταβλητής προκαλεί την στατιστικώς μη σημαντική αλλοίωση της προσαρμογής του μοντέλου. Η διαδικασία επαναλαμβάνεται μέχρι καμία μεταβλητή δεν μπορεί να διαγραφεί χωρίς στατιστικά σημαντική απώλεια προσαρμογής.

Αμφίδρομη απομάκρυνση (Both or Mixed). Ένας συνδυασμός των παραπάνω.

Πειραματικοί Σχεδιασμοί

```
> step(fitmult, direction='forward')
```

Start: AIC=62.21

Oxy ~ Weight + Runtime + RunPulse + RstPulse + MaxPulse

Call:

lm(formula = Oxy ~ Weight + Runtime + RunPulse + RstPulse + MaxPulse)

Coefficients:

(Intercept)	Weight	Runtime	RunPulse	RstPulse	MaxPulse
82.39361	-0.05091	-2.95182	-0.39704	0.01239	0.38479

Πειραματικοί Σχεδιασμοί

> step(fitmult, direction='backward')

Start: AIC=62.21

Oxy ~ Weight + Runtime + RunPulse + RstPulse + MaxPulse

	Df	Sum of Sq	RSS	AIC
- RstPulse	1	0.20	156.78	60.247
- Weight	1	4.84	161.42	61.151
<none>			156.58	62.208
- MaxPulse	1	45.83	202.41	68.166
- RunPulse	1	59.52	216.10	70.194
- Runtime	1	366.34	522.92	97.589

Step: AIC=60.25

Oxy ~ Weight + Runtime + RunPulse + MaxPulse

	Df	Sum of Sq	RSS	AIC
- Weight	1	4.99	161.77	59.218
<none>			156.78	60.247
- MaxPulse	1	45.99	202.77	66.221
- RunPulse	1	59.34	216.12	68.197
- Runtime	1	421.49	578.28	98.708

Πειραματικοί Σχεδιασμοί

Step: AIC=59.22

Oxy ~ Runtime + RunPulse + MaxPulse

	Df	Sum of Sq	RSS	AIC
<none>			161.77	59.218
- MaxPulse	1	41.35	203.12	64.274
- RunPulse	1	55.14	216.91	66.311
- Runtime	1	443.20	604.98	98.107

Call:

lm(formula = Oxy ~ Runtime + RunPulse + MaxPulse)

Coefficients:

(Intercept)	Runtime	RunPulse	MaxPulse
80.9008	-2.9702	-0.3751	0.3542

Πειραματικοί Σχεδιασμοί

> step(fitmult, direction='both')

Start: AIC=62.21

Oxy ~ Weight + Runtime + RunPulse + RstPulse + MaxPulse

	Df	Sum of Sq	RSS	AIC
- RstPulse	1	0.20	156.78	60.247
- Weight	1	4.84	161.42	61.151
<none>			156.58	62.208
- MaxPulse	1	45.83	202.41	68.166
- RunPulse	1	59.52	216.10	70.194
- Runtime	1	366.34	522.92	97.589

Step: AIC=60.25

Oxy ~ Weight + Runtime + RunPulse + MaxPulse

	Df	Sum of Sq	RSS	AIC
- Weight	1	4.99	161.77	59.218
<none>			156.78	60.247
+ RstPulse	1	0.20	156.58	62.208
- MaxPulse	1	45.99	202.77	66.221
- RunPulse	1	59.34	216.12	68.197
- Runtime	1	421.49	578.28	98.708

Πειραματικοί Σχεδιασμοί

Step: AIC=59.22

Oxy ~ Runtime + RunPulse + MaxPulse

	Df	Sum of Sq	RSS	AIC
<none>			161.77	59.218
+ Weight	1	4.99	156.78	60.247
+ RstPulse	1	0.35	161.42	61.151
- MaxPulse	1	41.35	203.12	64.274
- RunPulse	1	55.14	216.91	66.311
- Runtime	1	443.20	604.98	98.107

Call:

lm(formula = Oxy ~ Runtime + RunPulse + MaxPulse)

Coefficients:

(Intercept)	Runtime	RunPulse	MaxPulse
80.9008	-2.9702	-0.3751	0.3542

ΑΝΑΛΥΣΗ ΔΙΑΔΡΟΜΩΝ (Path analysis)

Η ανάλυση διαδρομών (Path analysis) χρησιμοποιείται για την περιγραφή των σχέσεων εξάρτησης μεταξύ ενός συνόλου μεταβλητών.

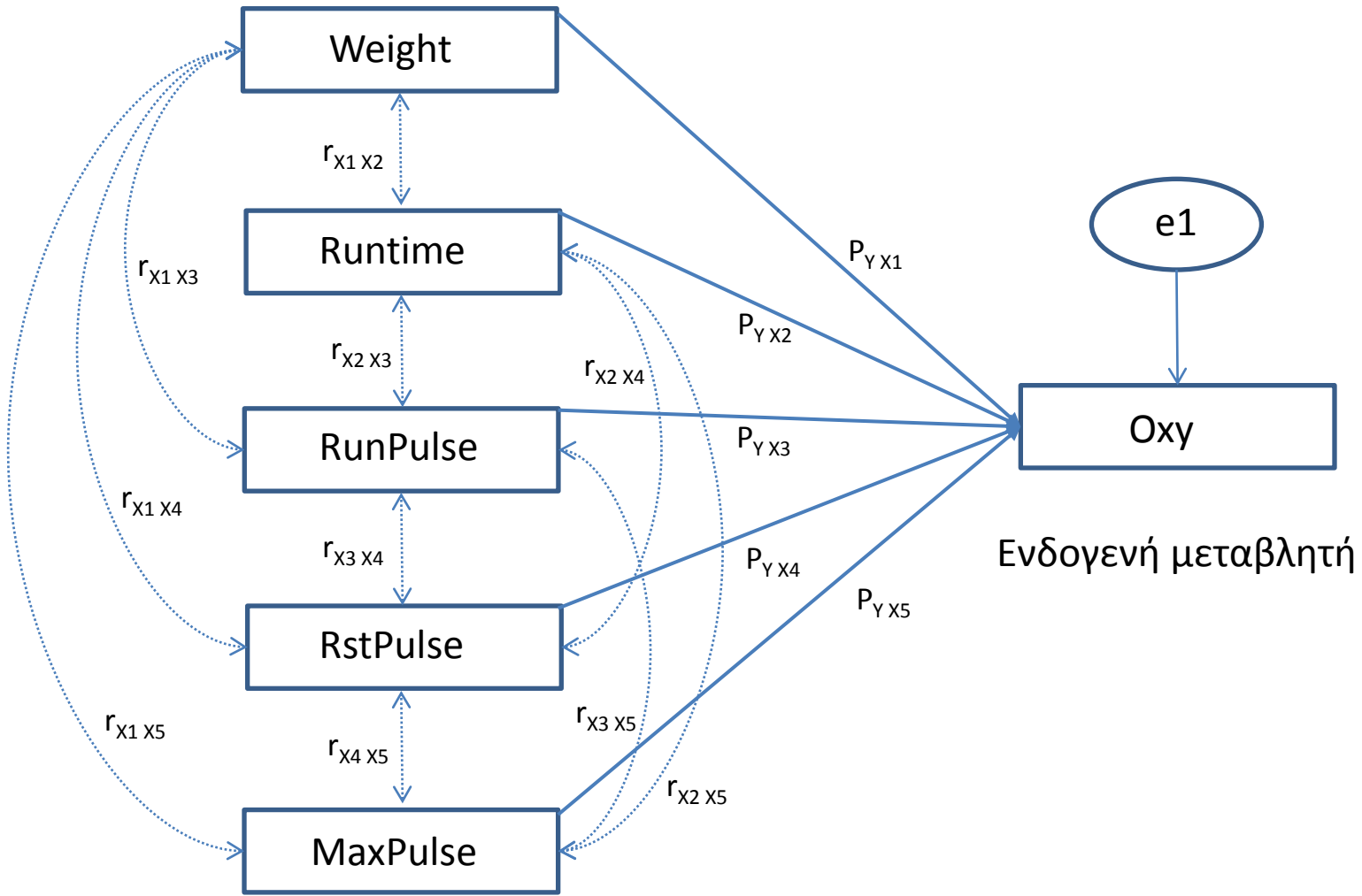
Άμεση επίδραση:
$$P_{YX_i} = \hat{b}_i \frac{S_X}{S_Y}$$
 (standardized regression coefficient)

Έμμεση επίδραση:
$$r_{X_i X_j} P_{YX_i}$$

Όπου: $r_{X_i X_j}$ ο συντελεστής συσχέτισης Pearson μεταξύ των μεταβλητών i και j .

Πειραματικοί Σχεδιασμοί

Εξωγενείς μεταβλητές



Πειραματικοί Σχεδιασμοί

```
> library(agricolae)
> y=MultRegression[,1]
> x=MultRegression[,c(2:6)]
> cor.y=correlation(y, x)$correlation
> cor.x=correlation(x)$correlation
> path.analysis(cor.x,cor.y)
```

Direct(Diagonal) and indirect effect path coefficients

```
=====
```

	Weight	Runtime	RunPulse	RstPulse	MaxPulse
Weight	-0.083697775	-0.1072208	-0.1443327	0.0004764935	0.1747748
Runtime	-0.011717688	-0.7658626	-0.2485730	0.0053605519	0.1607928
RunPulse	-0.015065599	-0.2374174	-0.8018484	0.0041693181	0.6501621
RstPulse	-0.003347911	-0.3446382	-0.2806470	0.0119123376	0.2167207
MaxPulse	-0.020924444	-0.1761484	-0.7457191	0.0036928246	0.6990991

Residual Effect^2 = 0.1797758