Γεωργικός Πειραματισμός

10ο Εργαστήριο

Αναστάσιος Κατσιλέρος

Γεωπονικό Πανεπιστήμιο Αθηνών
Εργαστήριο Βελτίωσης Φυτών και Γεωργικού Πειραματισμού

katsileros@aua.gr

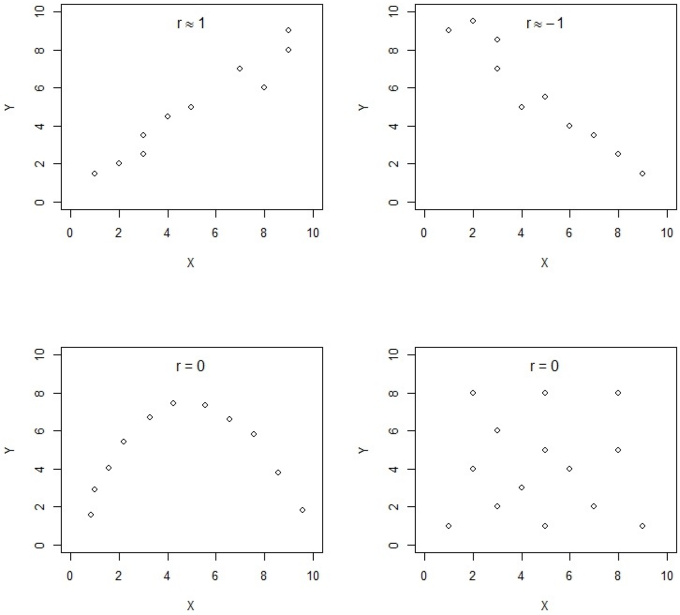
Αθήνα 2020

Ανάλυση Συσχέτισης

Η **ανάλυση συσχέτισης** (correlation analysis) είναι μία στατιστική μέθοδος που χρησιμοποιείται για τον έλεγχο της σύνδεσης μεταξύ δύο μεταβλητών, ανεξάρτητα από την ύπαρξη σχέσης αιτίου - αποτελέσματος μεταξύ τους. Για την εκτίμηση της έντασης αλλά και του τρόπου με τον οποίο δύο μεταβλητές συνδέονται-σχετίζονται, χρησιμοποιείται ο **συντελεστής γραμμικής συσχέτισης Pearson's r** (Pearson correlation coefficient).

Ο συντελεστής γραμμικής συσχέτισης του Pearson κυμαίνεται από -1 έως 1 και είναι ανεξάρτητος από τις μονάδες μέτρησης. Όταν συντελεστής παίρνει θετικές τιμές οι δύο μεταβλητές είναι θετικά γραμμικά συσχετισμένες, ενώ όταν παίρνει αρνητικές τιμές, είναι αρνητικά γραμμικά συσχετισμένες. Όσο το r πλησιάζει την απόλυτη τιμή 1, η ένταση της σύνδεσης είναι ισχυρή, ενώ τιμές κοντά στο μηδέν υποδεικνύει απουσία γραμμικής συσχέτισης μεταξύ των δύο μεταβλητών.

Η ύπαρξη ισχυρής συσχέτισης δεν σημαίνει ότι οι δύο μεταβλητές συνδέονται υποχρεωτικά με σχέση αιτιότητας αλλά μπορεί οι δύο μεταβλητές να επηρεάζονται από μια τρίτη μεταβλητή, όπως επίσης απουσία συσχέτισης μπορεί να υπάρχει μεταξύ δύο μεταβλητών που συνδέονται αλλά με μη γραμμική σχέση. Μία πρώτη εικόνα της σύνδεσης των δύο μεταβλητών δίνεται γραφικά με το **διάγραμμα διασποράς** των παρατηρήσεων (scatter plot).


r ≈ 1
r ≈ −1
r = 0
r = 0
X
Y

Ο συντελεστής r από ένα δείγμα n ζευγαρωτών παρατηρήσεων δύο μεταβλητών, υπολογίζεται ως εξής:

$$r = \frac{s_{xy}}{s_x s_y}$$

όπου:

$$s_{xy} = Cov(X, Y) = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$s_x^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \kappa\alpha\iota \quad s_y^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

οπότε:

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

Για τον προσδιορισμό της σχέσης των δύο μεταβλητών μπορεί να χρησιμοποιηθεί ο ισοδύναμος τύπος:

$$r = \frac{A\Gamma_{xy}}{\sqrt{AT_x AT_y}}$$

όπου:

$$A\Gamma_{xy} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}$$

$$AT_x = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}$$

$$AT_y = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$$

Παράδειγμα 13. (correlation.txt) Έλεγχος συσχέτισης πρωτεΐνης (X) και ελαίων (Y), σε σπόρους ελαιοκράμβης.

X_i	Y_i	$X_i Y_i$
1	9	9
2	6	12
3	8	24
3	4	12
4	5	20
5	3	15
5	2	10
6	3	18
$\sum X_i = 29$	$\sum Y_i = 40$	$\sum X_i Y_i = 120$
$\sum X_i^2 = 125$	$\sum Y_i^2 = 244$	

$$A\Gamma_{xy} = \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} = 120 - \frac{29 \cdot 40}{8} = -25$$

$$AT_x = \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n} = 125 - \frac{29^2}{8} = 19,9$$

$$AT_y = \sum_{i=1}^{n} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n} = 244 - \frac{40^2}{8} = 44$$

οπότε:

$$r = \frac{A\Gamma_{xy}}{\sqrt{AT_x AT_y}} = \frac{-25}{\sqrt{19,9 \cdot 44}} = -0,845$$

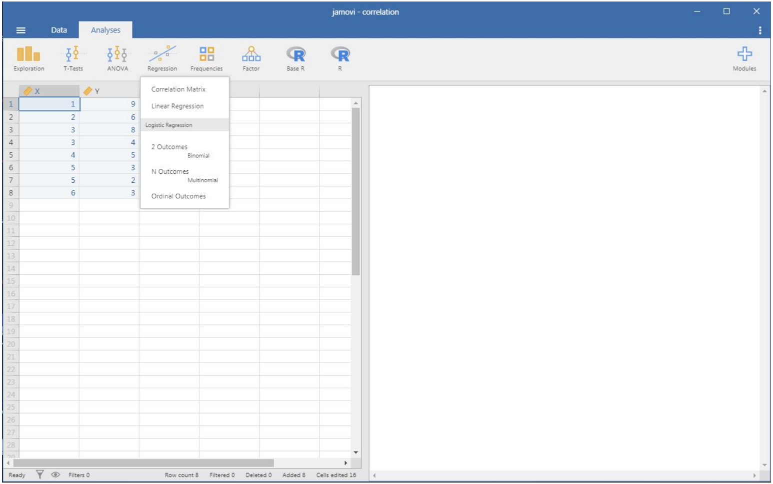
Ο έλεγχος των υποθέσεων ($H_0 : r = 0, H_1 : r \neq 0$) για τον συντελεστή συσχέτισης r μπορεί να γίνει με την χρήση πινάκων κρίσιμων τιμών του Pearson's r ή με τη δοκιμασία t.

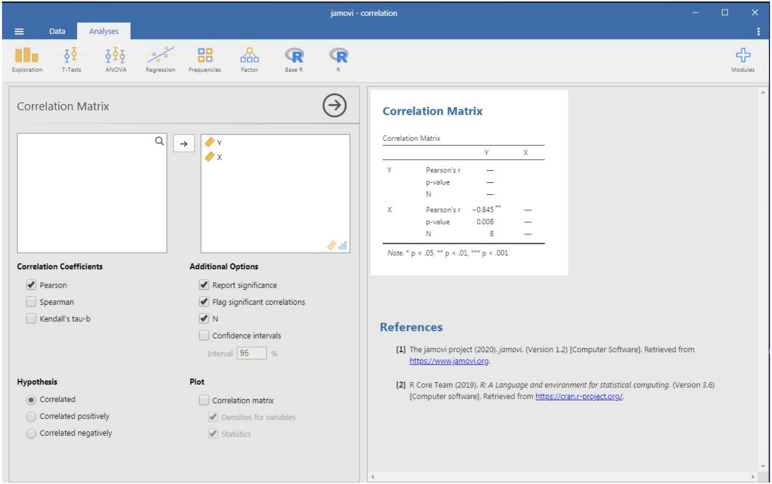
Η κρίσιμη τιμή του σχετικού πίνακα για βαθμούς ελευθερίας $n-2=6$ και για $\alpha = 0,05$, είναι 0,707. Επειδή η απόλυτη τιμή r $(-0,845)$ είναι μεγαλύτερη από την κρίσιμη τιμή του πίνακα, απορρίπτεται η μηδενική υπόθεση, επομένως υπάρχει στατιστικά σημαντική αρνητική συσχέτιση μεταξύ των μεταβλητών της περιεκτικότητας ελαίων και πρωτεΐνης.

Για τον έλεγχο σημαντικότητας με την δοκιμασία του t, χρησιμοποιείται ο τύπος:

$$t_r = \frac{(r-0)}{\sqrt{(1-r^2)/(n-2)}} = \frac{-0,845}{\sqrt{(1-(-0,845)^2)/6}} = 3,87$$

και η τιμή της δοκιμασίας t_r συγκρίνεται με την κρίσιμη τιμή του πίνακα t για $n-2$ βαθμούς ελευθερίας. Επειδή η τιμή $t_r = 3,87$ είναι μεγαλύτερη είναι από την κρίσιμη τιμή του πίνακα (2,44), απορρίπτεται η μηδενική υπόθεση.


jamovi - correlation
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
Correlation Matrix
Linear Regression
Logistic Regression
2 Outcomes
Binomial
N Outcomes
Multinomial
Ordinal Outcomes
X
Y
Ready
Filters 0
Row count 8
Filtered 0
Deleted 0
Added 8
Cells edited 16


jamovi - correlation
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
Correlation Matrix
Y
X
Correlation Coefficients
Pearson
Spearman
Kendall's tau-b
Additional Options
Report significance
Flag significant correlations
N
Confidence intervals
Interval 95 %
Hypothesis
Correlated
Correlated positively
Correlated negatively
Plot
Correlation matrix
Densities for variables
Statistics
Correlation Matrix
Correlation Matrix
Y X
Y Pearson's r —
p-value —
N —
X Pearson's r −0.845 ** —
p-value 0.008 —
N 8 —
Note. * p < .05, ** p < .01, *** p < .001
References
[1] The jamovi project (2020). jamovi. (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org.
[2] R Core Team (2019). R: A Language and environment for statistical computing. (Version 3.6) [Computer software]. Retrieved from https://cran.r-project.org/.

Η **ανάλυση συμμεταβολής ή παλινδρόμησης** (regression analysis) είναι μια στατιστική μέθοδος προσδιορισμού της σχέσης μεταξύ δύο ή και περισσότερων μεταβλητών, με σκοπό την πρόβλεψη των τιμών της μιας μεταβλητής μέσω των άλλων. Ο όρος παλινδρόμηση (regression) εισήχθηκε από τον ανθρωπολόγο Galton (1885), ο οποίος μελέτησε την σχέση του ύψους των παιδιών, με το ύψος των γονέων τους.

Η πιο απλή περίπτωση συμμεταβολής, είναι η απλή γραμμική (simple linear regression), στην οποία μελετάται η επίδραση μίας **μεταβλητής** (variable), η οποία ονομάζεται **ανεξάρτητη** (independent) ή **επεξηγηματική** (explanatory), σε μία άλλη μεταβλητή, η οποία ονομάζεται **εξαρτημένη** (dependent) ή **προβλέψιμη** (predicted). Μεταξύ των δύο μεταβλητών υπάρχει σχέση αιτίου-αποτελέσματος, που μπορεί να οφείλεται εν μέρει ή εξ ολοκλήρου στην επίδραση της ανεξάρτητης, στην εξαρτημένη μεταβλητή.

Το γραμμικό μοντέλο που περιγράφει την σχέση των δύο μεταβλ-
ητών είναι το εξής:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

όπου:
Y_i η i παρατηρούμενη τιμή της εξαρτημένης μεταβλητής Y
X_i η i τιμή της ανεξάρτητης μεταβλητής X
α, β είναι δύο άγνωστες σταθερές
ε_i το πειραματικό σφάλμα, $N(0, \sigma^2)$

Ο προσδιορισμός της σχέσης των δύο μεταβλητών αποδίδεται από μια ευθεία, η οποία προσαρμόζεται καλύτερα στα δεδομένα. Για την εύρεση της καλύτερης ευθείας, χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων, η οποία προσδιορίζει τις παραμέτρους α και β, έτσι ώστε το άθροισμα των τετραγώνων των αποκλίσεων των παρατηρήσεων από την ευθεία, να είναι ελάχιστο.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

Οι τιμές των παραμέτρων α και β, που ελαχιστοποιούν την σχέση καλούνται εκτιμήτριες ελαχίστων τετραγώνων, συμβολίζονται με $\hat{\alpha}$ και $\hat{\beta}$ και δίνονται από τις σχέσεις:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

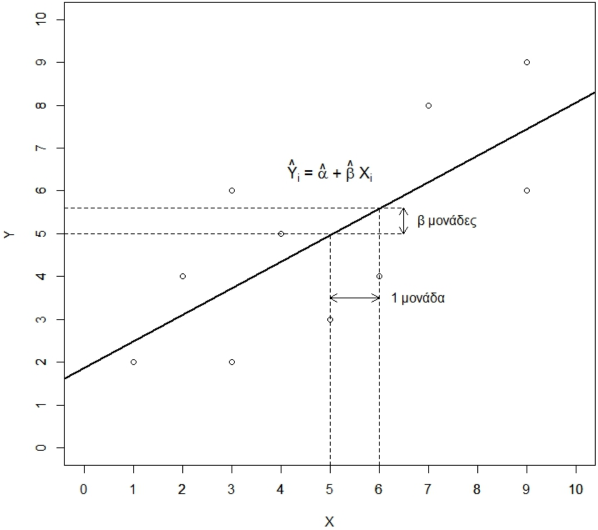
και

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

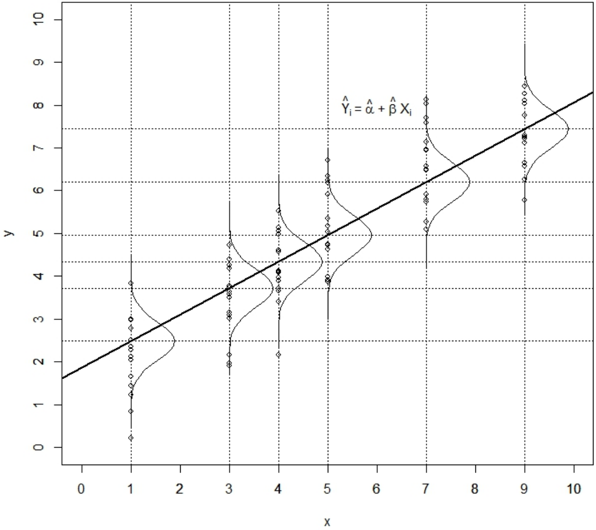
Επομένως η ευθεία η οποία προσαρμόζεται καλύτερα στα δεδομένα ονομάζεται ευθεία συμμεταβολής ή ευθεία ελαχίστων τετραγώνων και δίνεται από τον τύπο:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \tag{1}$$

όπου $\hat{Y}_i$ είναι η αναμενόμενη ή προβλεπόμενη τιμή που αντιστοιχεί σε μία δοθείσα τιμή X_i και η οποία είναι πάνω στην ευθεία συμμεταβολής και είναι συνήθως διαφορετική από την πραγματική ή παρατηρηθείσα τιμή Y_i. Η τιμή της εκτιμήτριας $\hat{\alpha}$ δίνει το σημείο όπου η ευθεία τέμνει τον άξονα Y, ενώ η εκτιμήτρια $\hat{\beta}$ είναι ο συντελεστής διεύθυνσης της ευθείας και δίνει τη μεταβολή της εξαρτημένης μεταβλητής Y όταν το X μεταβληθεί κατά μια μονάδα.


$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$
β μονάδες
1 μονάδα
Y
X
0
1
2
3
4
5
6
7
8
9
10

Οι προϋποθέσεις για τη χρήση της γραμμικής συμμεταβολής στην περιγραφή της σχέσης δύο μεταβλητών είναι ότι, η σχέση είναι πραγματικά γραμμική, οι τιμές της ανεξάρτητης μεταβλητής X είναι ελεγχόμενες ή καθορισμένες από τον ερευνητή, ενώ για κάθε επίπεδο της X, οι παρατηρούμενες τιμές του Y είναι ανεξάρτητες και ακολουθούν την κανονική κατανομή με ίσες διακυμάνσεις.



$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$
y
x
0
1
2
3
4
5
6
7
8
9
10

Για τον έλεγχο σημαντικότητας του συντελεστή συμμεταβολής, χρησιμοποιείται η δοκιμασία του t για BE = n − 2, ως εξής:

$H_0 : \beta = 0$ και $H_1 : \beta \neq 0$

$$t = \frac{\hat{\beta}}{s_b}$$

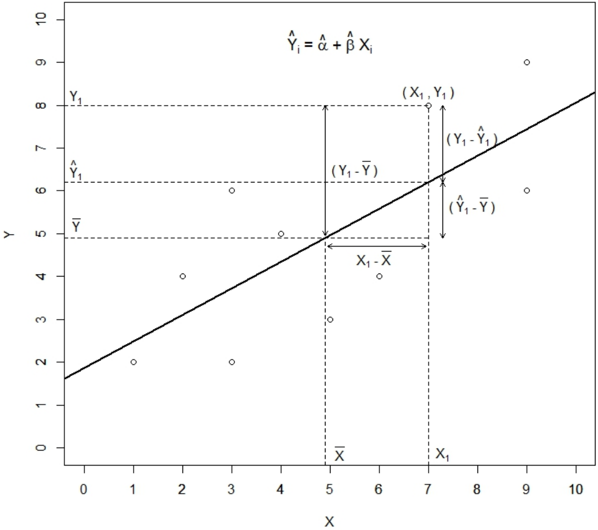
όπου:

$$s_b^2 = \frac{s^2}{(n-1)s_X^2} = \frac{s^2}{AT_X}$$

και

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Ένας άλλος τρόπος ελέγχου της σημαντικότητας του συντελεστή συμμεταβολής είναι με την ανάλυση διακύμανσης και την δοκιμασία του F, η οποία είναι ισοδύναμη με την δοκιμασία του t. Επιπλέον με την ανάλυση διακύμανσης υπολογίζεται το ποσοστό την παραλλακτικότητας που παρατηρείται στην εξαρτημένη μεταβλητή Υ και εξηγείται από την ανεξάρτητη μεταβλητή.



$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$
(X_1, Y_1)
Y_1
$\hat{Y}_1$
$\bar{Y}$
$(Y_1 - \bar{Y})$
$(Y_1 - \hat{Y}_1)$
$(\hat{Y}_1 - \bar{Y})$
$X_1 - \bar{X}$
$\bar{X}$
X_1
X
Y

Από το σχήμα φαίνεται ότι η διαφορά της παρατηρηθείσας τιμής Y_i από το μέσο $\bar{Y}$ δίνεται από το άθροισμα της διαφοράς της παρατηρηθείσας τιμής Y_i από την αναμενόμενη τιμή $\hat{Y}_i$ και της διαφοράς της αναμενόμενης τιμής $\hat{Y}_i$ από το μέσο $\bar{Y}$.

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

Υψώνοντας και τα δύο μέλη της σχέσης στο τετράγωνο και επαναλαμβάνοντας για όλες τις παρατηρήσεις, το συνολικό άθροισμα δίνεται από τη σχέση:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{n-1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{n-1}^{n}(Y_i - \hat{Y}_i)^2$$

Το ολικό άθροισμα τετραγώνων εκφράζει τη συνολική παραλλακτικότητα των παρατηρήσεων Y_i από την μέση τιμή $\bar{Y}$.

$$AT o\lambda = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = ATy$$

Το άθροισμα τετραγώνων της συμμεταβολής εκφράζει τη παραλλακτικότητα των αναμενόμενων τιμών $\hat{Y}_i$ από την μέση τιμή $\bar{Y}$.

$$AT\sigma\upsilon\mu = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = \frac{A\Gamma_{xy}^2}{ATx}$$

Το άθροισμα τετραγώνων των αποκλίσεων ή υπολοίπου εκφράζει τη παραλλακτικότητα των παρατηρήσεων Y_i σε σχέση με τις αναμενόμενες τιμές $\hat{Y}_i$.

$$AT\alpha\pi o\kappa\lambda = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = ATy - \frac{A\Gamma_{xy}^2}{ATx}$$

Πηγή Παρ/τας	**Βαθμοί Ελευθερίας**	**Άθροισμα Τετραγώνων**	**Μέσο Τετράγωνο**	**Δοκιμασία F**	**Κρίσιμη Τιμή F**
Συμμεταβολή	1	$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	$AT\sigma\upsilon\mu/1$	$MT\sigma\upsilon\mu/MT\upsilon$	$F_{(1,n-2)}$
Αποκλίσεις ή Υπόλοιπο	$n-2$	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$AT\upsilon\pi/(n-2)$		
Σύνολο	$n-1$	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$			

Το ποσοστό την παραλλακτικότητας της εξαρτημένης μεταβλητή Y που εξηγείται από την ανεξάρτητη μεταβλητή, δίνεται από τον συντελεστή προσδιορισμού R^2 (coefficient of determination) ως εξής:

$$R^2 = \frac{\text{ΑΤσυμμεταβολής}}{\text{ΑΤσυνόλου}} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Παράδειγμα 14. (regression.txt) Μελετήθηκε σε ένα πείραμα η επίδραση της ποσότητας του λιπάσματος X, στην απόδοση Y μίας ποικιλίας.

X_i	Y_i	$X_i Y_i$
1	2	2
2	4	8
3	2	6
3	6	18
4	5	20
5	3	15
6	4	24
7	8	56
9	7	63
9	9	81
$\sum X_i = 49$	$\sum Y_i = 50$	$\sum X_i Y_i = 293$
$\sum X_i^2 = 311$	$\sum Y_i^2 = 304$	

$$A\Gamma_{xy} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} =$$

$$= 293 - \frac{49 \cdot 50}{10} = 48$$

$$AT_x = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n} = 311 - \frac{49^2}{10} = 70,9$$

$$AT_y = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n} = 304 - \frac{50^2}{10} = 154$$

Ο συντελεστής της ευθείας συμμεταβολής $\hat{\beta}$ υπολογίζεται ως εξής:

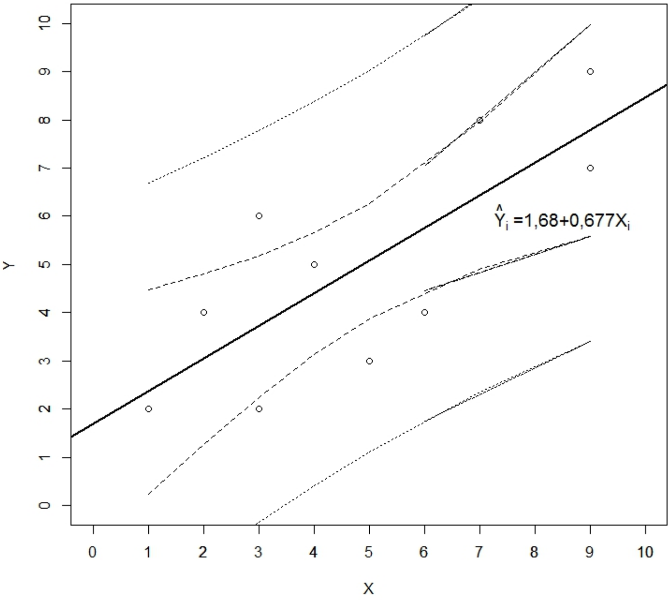
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})} = \frac{A\Gamma xy}{ATx} = \frac{48}{70,9} = 0,677$$

και

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = \frac{50}{10} - 0,677\frac{49}{10} = 1,68$$

Επομένως η εξίσωση συμμεταβολής είναι η εξής:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i = 1,68 + 0,677X_i$$



$\hat{Y}_i = 1{,}68 + 0{,}677X_i$
Y
X
0 1 2 3 4 5 6 7 8 9 10

Υπολογισμός υπολοίπων

$$\hat{Y}_i = 1,68 + 0,677X_i$$

X_i	Y_i	$\hat{Y}_i$	$\varepsilon_i = Y_i - \hat{Y}_i$
1	2	2,357	-0,357
2	4	3,034	0,966
3	2	3,711	-1,711
3	6	3,711	2,289
4	5	4,388	0,612
5	3	5,065	-2,065
6	4	5,742	-1,742
7	8	6,419	1,581
9	7	7,773	-0,773
9	9	7,773	1,227

Έλεγχος Ανεξαρτησίας Durbin Watson

Η στατιστική Durbin Watson είναι ένας έλεγχος για την ανεξαρτησία των υπολοίπων. Η στατιστική d έχει τιμή μεταξύ 0 και 4. Η τιμή 2 σημαίνει ότι δεν εντοπίζεται αυτοσυσχέτιση, ενώ τιμές μεταξύ 0 έως 2 υποδεικνύουν θετική αυτοσυσχέτιση και από 2 έως 4, αρνητική αυτοσυσχέτιση.

$$d = \frac{\sum_{i=2}^{n} (\epsilon_i - \epsilon_{i-1})^2}{\sum_{i=1}^{n} \epsilon_i^2}$$

όπου $\varepsilon_i = Y_i - \hat{Y}_i$

$H_0 : \rho = 0$ και $H_1 : \rho \neq 0$

$$d = \frac{(0,966 - (-0,357))^2 + ... + (1,227 - (-0,773))^2}{21,503} =$$

$$= \frac{55,583}{21,503} = 2,584$$

Ο έλεγχος της σημαντικότητας της τιμής d γίνεται χρησιμοποιώντας τις κρίσιμες τιμές από τον Πίνακα Durbin-Watson.

Δοκιμασία του t

$$t = \frac{\hat{\beta}}{s_b}$$

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \frac{1}{n-2}(ATy - \frac{A\Gamma_{xy}^2}{ATx}) = \frac{21,5}{8} = 2,69$$

$$s_b^2 = \frac{s^2}{(n-1)s_X^2} = \frac{2,69}{70,9} = 0,0379$$

$$t = \frac{\hat{\beta}}{s_b} = \frac{0,677}{\sqrt{0,0379}} = 3,476$$

Η τιμή της δοκιμασίας t συγκρίνεται με την κρίσιμη τιμή του πίνακα t για $n-2$ βαθμούς ελευθερίας. Επειδή η τιμή $t = 3,476$ είναι μεγαλύτερη είναι από την κρίσιμη τιμή του πίνακα (2,44), απορρίπτεται η μηδενική υπόθεση.

Ανάλυση διακύμανσης

$$AT o\lambda = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = ATy = 54$$

$$AT\sigma\upsilon\mu = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \frac{A\Gamma_{xy}^2}{ATx} = \frac{48^2}{70,9} = 32,5$$

$$AT\alpha\pi o\kappa\lambda = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = ATy - \frac{A\Gamma_{xy}^2}{ATx}$$

ή

$$AT\alpha\pi o\kappa\lambda = AT o\lambda - AT\sigma\upsilon\mu = 54 - 32,5 = 21,5$$

Πηγή Παρ/τας	Βαθμοί Ελευθερίας	Άθροισμα Τετραγώνων	Μέσο Τετράγωνο	Δοκιμασία F	Κρίσιμη Τιμή F
Συμμεταβολή	1	32,5	32,5	12,09**	5,31
Αποκλίσεις ή Υπόλοιπο	8	21,5	2,688		
Σύνολο	9	54			

Επειδή το F της συμμεταβολής είναι μεγαλύτερη από την κρίσιμη τιμή, απορρίπτεται η μηδενική υπόθεση και επομένως υπάρχει γραμμική σχέση μεταξύ X και Y.

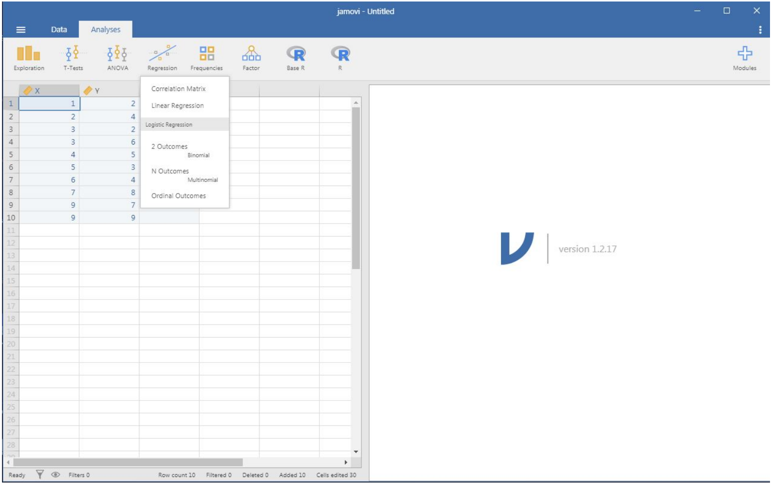
Ο συντελεστής προσδιορισμού R^2 υπολογίζεται ως εξής:

$$R^2 = \frac{\text{ΑΤσυμμεταβολής}}{\text{ΑΤσυνόλου}} = \frac{32,5}{54} = 0,602$$

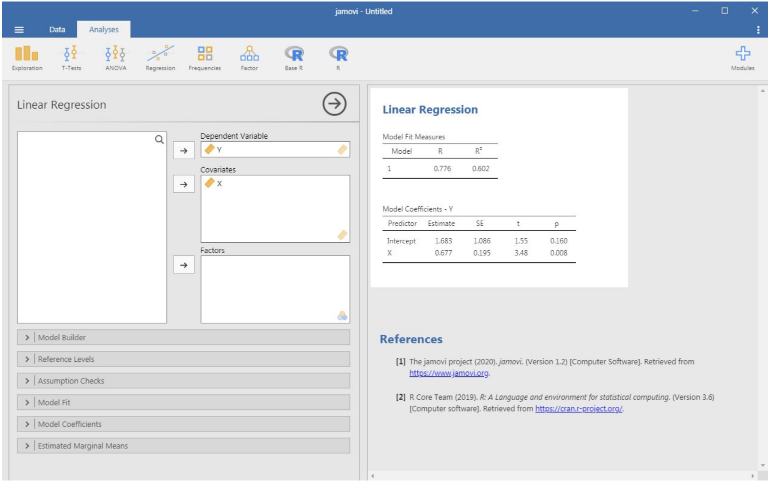
Επομένως χρησιμοποιώντας το συγκεκριμένο γραμμικό μοντέλο, το 60,2% της παραλλακτικότητας της εξαρτημένης μεταβλητή Υ εξηγείται από την ανεξάρτητη μεταβλητή, ενώ το υπόλοιπο ποσοστό οφείλεται σε τυχαία γεγονότα.

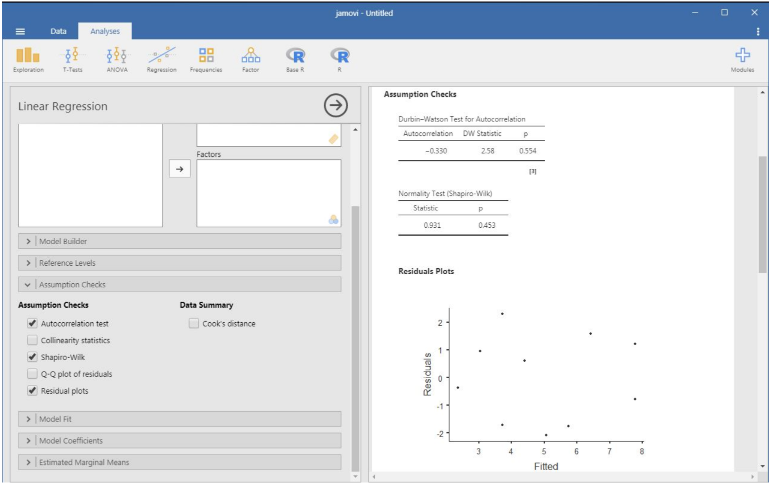
Ο προσαρμοσμένος συντελεστής προσδιορισμού R^2_{adj} χρησιμοποιείται για τη σύγκριση μοντέλων με διαφορετικούς αριθμούς παραμέτρων και υπολογίζεται ως εξής:

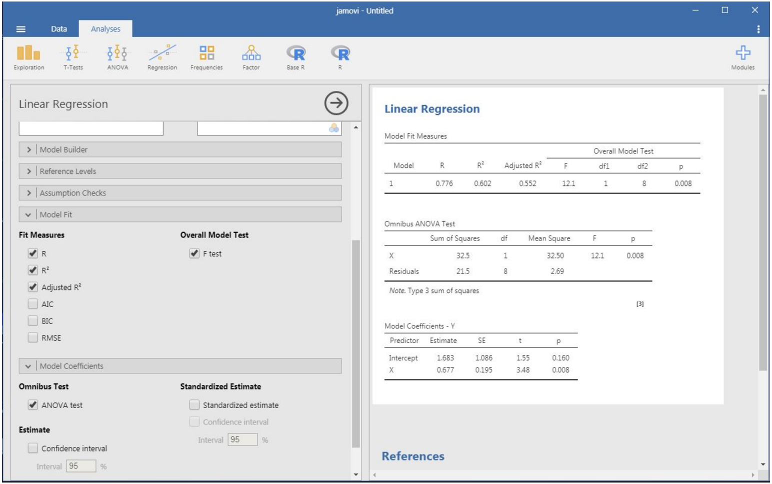
$$R^2_{adj} = 1 - \frac{\text{ΜΤυπολοίπου}}{\text{ΜΤσυνόλου}} = 1 - \frac{2,688}{6} = 0,552$$

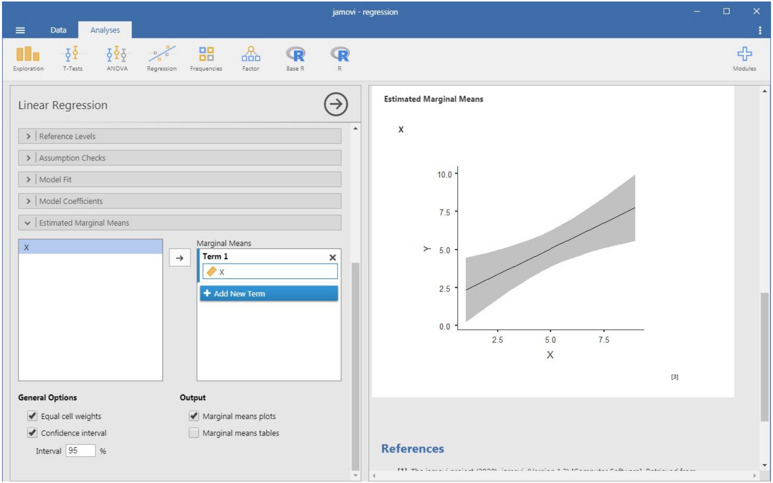


jamovi - Untitled
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
X
Y
Correlation Matrix
Linear Regression
Logistic Regression
2 Outcomes
Binomial
N Outcomes
Multinomial
Ordinal Outcomes
version 1.2.17
Ready
Filters 0
Row count 10
Filtered 0
Deleted 0
Added 10
Cells edited 30


jamovi - Untitled
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
Linear Regression
Dependent Variable
Y
Covariates
X
Factors
Model Builder
Reference Levels
Assumption Checks
Model Fit
Model Coefficients
Estimated Marginal Means
Linear Regression
Model Fit Measures
Model R R²
1 0.776 0.602
Model Coefficients - Y
Predictor Estimate SE t p
Intercept 1.683 1.086 1.55 0.160
X 0.677 0.195 3.48 0.008
References
[1] The jamovi project (2020). jamovi. (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org.
[2] R Core Team (2019). R: A Language and environment for statistical computing. (Version 3.6) [Computer software]. Retrieved from https://cran.r-project.org/.


jamovi - Untitled
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
Linear Regression
Factors
Model Builder
Reference Levels
Assumption Checks
Assumption Checks
Autocorrelation test
Collinearity statistics
Shapiro-Wilk
Q-Q plot of residuals
Residual plots
Data Summary
Cook's distance
Model Fit
Model Coefficients
Estimated Marginal Means
Assumption Checks
Durbin–Watson Test for Autocorrelation
Autocorrelation DW Statistic p
−0.330 2.58 0.554
[3]
Normality Test (Shapiro-Wilk)
Statistic p
0.931 0.453
Residuals Plots
Residuals
Fitted


jamovi - Untitled
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
Linear Regression
Model Builder
Reference Levels
Assumption Checks
Model Fit
Fit Measures
R
R²
Adjusted R²
AIC
BIC
RMSE
Overall Model Test
F test
Model Coefficients
Omnibus Test
ANOVA test
Standardized Estimate
Standardized estimate
Confidence interval
Interval 95 %
Estimate
Confidence interval
Interval 95 %
Linear Regression
Model Fit Measures
Overall Model Test
Model R R² Adjusted R² F df1 df2 p
1 0.776 0.602 0.552 12.1 1 8 0.008
Omnibus ANOVA Test
Sum of Squares df Mean Square F p
X 32.5 1 32.50 12.1 0.008
Residuals 21.5 8 2.69
Note. Type 3 sum of squares
[3]
Model Coefficients - Y
Predictor Estimate SE t p
Intercept 1.683 1.086 1.55 0.160
X 0.677 0.195 3.48 0.008
References


jamovi - regression
Data
Analyses
Exploration
T-Tests
ANOVA
Regression
Frequencies
Factor
Base R
R
Modules
Linear Regression
Reference Levels
Assumption Checks
Model Fit
Model Coefficients
Estimated Marginal Means
X
Marginal Means
Term 1
X
Add New Term
General Options
Equal cell weights
Confidence interval
Interval 95 %
Output
Marginal means plots
Marginal means tables
Estimated Marginal Means
X
10.0
7.5
5.0
2.5
0.0
Y
2.5
5.0
7.5
X
[3]
References