



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ»**

ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΚΑΙ ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

ΕΡΓΑΣΙΑ ΔΕΥΤΕΡΗ



ΜΠΟΥΡΑΝΤΑΝΗΣ ΑΝΔΡΕΑΣ

mpouras.math@gmail.com

ΑΜ:ΜΕΣ22003

ΕΙΣΑΓΩΓΗ

Γνωστή πολυεθνική εταιρία επιθυμεί να αξιολογήσει τα αποτελέσματα της δραστηριότητας franchising, την οποία έχει αναπτύξει τα τελευταία έτη. Για το σκοπό αυτό, κατέγραψε για κάθε κατάστημα franchise το οποίο λειτούργησε με δικά της προϊόντα, τα εξής χαρακτηριστικά:

- Ετήσια καθαρά κέρδη σε χιλιάδες Ευρώ (1000 €) του καταστήματος franchise (Net profit [Y])
- Επιφάνεια του καταστήματος franchise σε εκατοντάδες τετραγωνικών μέτρων (sq.ft [X1])
- Απόθεμα προϊόντων (σε κιβώτια) του καταστήματος franchise (Inventory [X2])
- Έξοδα διαφήμισης του καταστήματος franchise (Advertising expenses [X3])
- Πλήθος ανταγωνιστικών καταστημάτων στην περιοχή του καταστήματος franchise (competing stores [X4])

A.) Θέλουμε να εξετάσουμε όλα τα δυνατά γραμμικά μοντέλα που έχουν ως ερμηνευτικές μεταβλητές κάποιες από τις X1,X2,X3,X4 και ως εξαρτημένη μεταβλητή την Y και να βρούμε το καλύτερο μοντέλο βάσει της διαδικασίας Enter.

Θα ξεκινήσουμε με το πλήρες μοντέλο:

Correlations						
		Net profit	Square fit	Inventory	Advertising expenses	Competing stores
Pearson Correlation	Net profit	1.000	.841	.938	.900	-.893
	Square fit	.841	1.000	.790	.691	-.690
	Inventory	.938	.790	1.000	.896	-.775
	Advertising expenses	.900	.691	.896	1.000	-.798
	Competing stores	-.893	-.690	-.775	-.798	1.000
	Square fit	24	24	24	24	24
	Inventory	24	24	24	24	24
	Advertising expenses	24	24	24	24	24
	Competing stores	24	24	24	24	24

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	174.204	79.770		2.184	.042		
	Square fit	19.173	6.842	.203	2.802	.011	.353	2.834
	Inventory	.384	.114	.387	3.369	.003	.141	7.080
	Advertising expenses	.007	.005	.142	1.342	.195	.167	5.990
	Competing stores	-13.440	2.992	-.340	-4.493	.000	.325	3.079

Αρχικά από τον πρώτο πίνακα παρατηρούμε πως υπάρχει μεγάλη συσχέτιση μεταξύ των μεταβλητών μας ενώ στο δεύτερο πίνακα βλέπουμε πως ο παράγοντας διόγκωσης διακύμανσης VIF των επεξηγηματικών μεταβλητών δεν είναι πολύ κοντά στο 1 ωστόσο δεν είναι μεγαλύτερες του 10 επομένως χρειάζεται περεταίρω διερεύνηση.

Θα συνεχίσουμε με τους υπόλοιπους συνδυασμούς των επεξηγηματικών μεταβλητών:

Μοντέλο	Μεταβλητές	R^2	R^2_{adj}	MSE _p
1	X ₁ ,X ₂ ,X ₃ ,X ₄	0.965	0.957	3.368
2	X ₁ ,X ₂ ,X ₃	0.927	0.916	6.600
3	X ₁ ,X ₂ ,X ₄	0.961	0.956	3.503
4	X ₁ ,X ₃ ,X ₄	0.944	0.935	5.112
5	X ₂ ,X ₃ ,X ₄	0.950	0.943	4.523
6	X ₁ ,X ₂	0.906	0.897	8.107
7	X ₁ ,X ₃	0.902	0.893	8.458
8	X ₁ ,X ₄	0.894	0.884	9.138
9	X ₂ ,X ₃	0.897	0.887	8.862
10	X ₂ ,X ₄	0.948	0.944	4.447
11	X ₃ ,X ₄	0.894	0.884	9.158
12	X ₁	0.708	0.695	24.055
13	X ₂	0.879	0.873	9.963
14	X ₃	0.810	0.801	15.675
15	X ₄	0.797	0.788	16.686

Βλέπουμε οπότε πως βάσει του κριτηρίου R^2adj και του MSE_p καταλήγουμε πως το καλύτερο γραμμικό μοντέλο επιτυγχάνεται με τη προσθήκη όλων των επεξηγηματικών μεταβλητών με $R^2adj=0.957$ και $MSE_p=3.368$.

Όμως από τον πίνακα με τις εκτιμήσεις των παραμέτρων διαπιστώνουμε πως το p-value της X_3 είναι 0.195. Οπότε για $\alpha < 19.5\%$ η β_4 δεν είναι στατιστικά σημαντική συνεπώς αφαιρούμε τη X_3 από το πολλαπλό μοντέλο παλινδρόμησης και καταφεύγουμε στο μοντέλο με επεξηγηματικές τις X_1, X_2, X_4 και $R^2adj = 0.956$.

Συνεχίζοντας θέλουμε να εκτιμήσουμε τις παραμέτρους του πολυωνυμικού μοντέλου:

$$E(Y) = c_0 + c_1X_2 + c_2X_2^2 + c_3X_2^3$$

	Mean	Std.Dev	95% LCL	95% UCL	VIF
C ₀	-69.774	185.145	-455.979	316.431	
C ₁	0.456	1.080	-1.797	2.709	223.68
C ₂	0.001	0.002	-0.002	0.005	914.97
C ₃	$-7.9 * 10^{-7}$	0.000	0.000	0.000	267.12

Παρατηρούμε στον παραπάνω πίνακα πως τα VIF των επεξηγηματικών μεταβλητών έχουν τεράστιες τιμές ($>>10$) συνεπώς οι εκτιμήσεις των παραμέτρων δεν είναι αξιόπιστες.

Β.) Μας έχει ζητηθεί να επιλέξουμε το βέλτιστο πλήθος επεξηγηματικών μεταβλητών για το εξής μοντέλο χρησιμοποιώντας τη μέθοδο Stepwise:

$$E(Y) = \beta_0 + \sum_{i=1}^4 \beta_i X_i + \sum_{i=1}^4 \beta_{4+i} X_i^2 + X_1 \sum_{i=2}^4 \beta_{7+i} X_i + X_2 \sum_{i=3}^4 \beta_{9+i} X_i + \beta_{14} X_3 X_4$$

Μοντέλο	Μεταβλητές	R^2adj	Sig.F Change
1	X_2	0.873	0.001
2	X_2, X_4^2	0.949	0.001
3	X_2, X_4^2, X_2^2	0.960	0.015
4	$X_2, X_4^2, X_2^2, X_1 X_3$	0.969	0.016

Διαπιστώνουμε πως με τη μέθοδο Stepwise το καλύτερο μοντέλο που προκύπτει αποτελείται από τέσσερις μεταβλητές τις X_2 , X_4^2 , X_2^2 και X_1X_3 . Επιπρόσθετα βλέπουμε στην τέταρτη στήλη ότι για επίπεδο σημαντικότητας μεγαλύτερο του 1.6% όλες οι προαναφερθείσες μεταβλητές βελτιώνουν το μοντέλο μας.

		Unstandardized Coefficients		Standardized Coefficients	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta		Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-109.167	46.809		.029	-206.243	-12.091		
	Inventory	.932	.074	.938	.000	.779	1.084	1.000	1.000
2	(Constant)	189.753	59.670		.005	65.663	313.842		
	Inventory	.639	.069	.643	.000	.495	.782	.461	2.169
	X4_square	-.687	.119	-.401	.000	-.934	-.440	.461	2.169
3	(Constant)	77.866	67.321		.261	-62.563	218.295		
	Inventory	1.120	.191	1.128	.000	.722	1.519	.047	21.432
	X4_square	-.693	.105	-.405	.000	-.911	-.474	.461	2.170
	X2_square	.000	.000	-.500	.015	-.001	.000	.049	20.451
4	(Constant)	50.038	60.077		.415	-75.704	175.781		
	Inventory	1.016	.172	1.023	.000	.656	1.377	.044	22.619
	X4_square	-.583	.101	-.341	.000	-.795	-.372	.383	2.612
	X2_square	-.001	.000	-.622	.002	-.001	.000	.045	22.087
	X1_X3	.001	.001	.292	.016	.000	.002	.108	9.274

Στον από πάνω πίνακα αναγράφονται οι εκτιμήσεις των καλύτερων μοντέλων που προέκυψαν με τη μέθοδο Stepwise καθώς επίσης τα 95% διαστήματα εμπιστοσύνης τους όπως και οι παράγοντες διόγκωσης διακύμανσης.

Συγκρίνοντας τώρα τα αποτελέσματα του ερωτήματος α. με αυτά του β. έχουμε πως το R^2_{adj} του τελευταίου είναι μεγαλύτερο του πρώτου ωστόσο οι τιμές των VIF του α. κυμαίνονται μεταξύ διαστήματος (2.61 ,3.63) ενώ στο β. παρατηρούμε ότι μία μεταβλητή έχει τιμή κοντά στο 10 ενώ δύο άλλες έχουνε πάρει ανησυχητικούς τιμές κοντά στο 22. Επομένως το μοντέλο α. αποτελεί καλύτερο καθώς έχει ίσο σχεδόν τροποποιημένο συντελεστή προσδιορισμού με τον β. ενώ παράλληλα διατηρεί μικρές τιμές VIF.

i.) Προτού ξεκινήσουμε να εκτιμήσουμε τις παραμέτρους του μοντέλου που προτιμήσαμε θα το ελέγξουμε ως προς τις συνθήκες κανονικότητας, ανεξαρτησίας, ομοσκεδαστικότητας και πολυσυγγραμικότητας.

Έλεγχος	p-value	X_i	VIF
Κανονικότητας	0.200	X_1	2.77
Ανεξαρτησίας	0.835	X_2	3.63
Ομοσκεδαστικότητας	0.717	X_4	2.61

Επομένως για οποιοδήποτε επίπεδο σημαντικότητας μικρότερο του 20% τηρούνται οι συνθήκες κανονικότητας, ανεξαρτησίας και ομοσκεδαστικότητας. Επιπλέον οι VIF παίρνουνε πολύ μικρότερες τιμές από το 10 οπότε δεν έχουμε πρόβλημα πολυσυγγραμικότητας

Model		Unstandardized Coefficients		95.0% Confidence Interval for B	
		B	Std. Error	Lower Bound	Upper Bound
1	(Constant)	225.106	71.567	75.820	374.392
	Square fit	17.807	6.900	3.414	32.199
	Inventory	.491	.083	.317	.665
	Competing stores	-15.012	2.807	-20.868	-9.156

ii.) Από το στατιστικό πακέτο υπολογίσαμε τις εκτιμήσεις και τα διαστήματα εμπιστοσύνης των παραμέτρων του πλήρες μοντέλο

Στο 95% των δειγμάτων:

- το β_0 θα παίρνει τιμές εντός του διαστήματος [75.82 , 374.392]
- το β_1 θα παίρνει τιμές εντός του διαστήματος [3.414, 32.199]
- το β_2 θα παίρνει τιμές εντός του διαστήματος [0.317 , 0.665]
- το β_4 θα παίρνει τιμές εντός του διαστήματος [-20.868 , -9.156]

Σε επίπεδο σημαντικότητας 5%:

Η επιφάνεια του καταστήματος είναι θετικά συσχετισμένη με τα κέρδη της εταιρίας και όταν αυξηθεί η επιφάνεια του καταστήματος κατά 100τ.μ. διατηρώντας το απόθεμα και τα ανταγωνιστικά καταστήματα σταθερά, τα μέσα ετήσια κέρδη θα αυξηθούν από 3.41 μέχρι 32.2 χιλιάδες..

Το απόθεμα προϊόντος είναι θετικά συσχετισμένο με τα ετήσια κέρδη, επιπλέον όταν αυξάνουμε το απόθεμα του προϊόντος κατά 1 κουτί διατηρώντας την επιφάνεια του καταστήματος καθώς και τα ανταγωνιστικά καταστήματα στην περιοχή σταθερά, η αναμενόμενη αύξηση των ετήσιων κερδών κυμαίνεται μεταξύ 0.32 και 0.67 χιλιάδες.

Το πλήθος των ανταγωνιστικών καταστημάτων στην περιοχή συσχετίζεται αρνητικά με τα ετήσια κέρδη της πολυεθνικής, πιο συγκεκριμένα όταν κατασκευαστεί ένα ανταγωνιστικό κατάστημα στην περιοχή ενώ διατηρούμε την επιφάνεια και το απόθεμα σταθερό, αναμένουμε μείωση των ετήσιων κερδών από 9.16 μέχρι 20.87 χιλιάδες ευρώ.

Συνεπώς η εξίσωση πολλαπλής παλινδρόμησης είναι:

$$\text{Net.Profit} = 225.106 + 7.807 * \text{Square.Fit} + 0.491 * \text{Inventory} - 15.012 * \text{Comp.Stores}$$

iii .) Θέλουμε να προβλέψουμε το καθαρό ετήσιο κέρδος ενός νέου καταστήματος franchise για ένα κατάστημα με επιφάνεια 400τ.μ. ,απόθεμα 200 κιβώτια και 5 ανταγωνιστικά καταστήματα στην γύρω περιοχή.

Ετήσια Κέρδη(€)	Επιφάνεια(sq.ft)	Απόθεμα	Αντ. Καταστήματα
319.47	4(εκατοντάδες)	200	5

Με το πολλαπλό μοντέλο παλινδρόμησης εκτιμάμε πως τα κέρδη της εταιρίας θα είναι 319.47 .

	Επιφάνεια(sq.ft)	Απόθεμα	Αντ. Καταστήματα
Std. Deviation	2.98	282	7.1

Δυστυχώς λόγω μικρού δείγματος δε μπορούμε να βρούμε παρατηρήσεις όπου οι επεξηγηματικές τους να παίρνουν τιμές ‘κοντά’ με αυτής που θέλουμε να προβλέψουμε για να έχουμε ένα μέτρο αξιολόγησης.

iv .) Παρακάτω δίνονται τα από κοινού (τουλάχιστον 95%) Δ.Ε. για τις παραμέτρους β_1 και β_4 με $\alpha_i=2.5$ για την κάθε μία:

β	LCL	UCL
β_1	1.088	34.525
β_4	-21.814	-.8.21

Επομένως το από κοινού διάστημα εμπιστοσύνης θα είναι το καρτεσιανό γινόμενο των δύο διαστημάτων.

Γ .) Προχωρώντας θα υπολογίσουμε τους συντελεστές προσδιορισμού $R^2_{Y,X_1/X_4}$ και $R^2_{Y,X/X_2}$:

Μοντέλο	X_2	X_4	X_1, X_2	X_1, X_4
SSR	1.592.015	1.444.112	1.640.945	1.619.290
SSE	219.191	367.094	-	-
$R^2_{Y,X_i/X_j}$	-	-	0.223	0.477

Οπότε συμπεραίνουμε πως :

Από το ποσοστό μεταβλητότητας που έχει μείνει ανερμηνευτο στο μοντέλο με επεξηγηματική μόνο την X_2 (και εξαρτημένη την Y),η προσθήκη της X_1 εξηγεί το 22.3% αυτής.

Αντίστοιχα προσθέτοντας την X_1 στο μοντέλο με επεξηγηματική μόνο την X_4 , η X_1 επεξηγεί το 47.7% της ανερμηνευτης μεταβλητότητας στο μοντέλο.

Δ .) Γνωρίζουμε ότι για τα καταστήματα που φιλοξενούνται σε χώρους επιφάνειας άνω των 400 τετραγωνικών μέτρων, η πολυεθνική εταιρία προσφέρει δωρεάν service στον εξοπλισμό που λειτουργεί στο κατάστημα.

i.) Για να συνεχίσουμε την ανάλυση με τα νέα δεδομένα πρέπει να προσαρμόσουμε το πολλαπλό μοντέλο παλινδρόμησης με την εισαγωγή μιας ψευδομεταβλητής X_5 . Επομένως το μοντέλο θα έχει τη μορφή:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_4 * X_4 + \beta_5 * X_5 + \varepsilon, \text{ όπου } X_5 = 1 \text{ αν } X_1 \geq 4$$

και $X_5 = 0$ διαφορετικά

ii.) Θα μελετήσουμε αν η παροχή δωρεάν service σε ένα κατάστημα επηρεάζει τα ετήσια κέρδη του καταστήματος σε επίπεδο σημαντικότητας 5%.

Model		Unstandardized Coefficients		Sig.	95.0% Confidence Interval for B		VIF
		B	Std. Error		Lower Bound	Upper Bound	
1	(Constant)	243.793	56.457	.000	125.626	361.959	
	Square fit	8.325	6.008	.182	-4.249	20.899	3.404
	Inventory	.401	.070	.000	.255	.547	4.141
	Competing stores	-15.081	2.206	.000	-19.698	-10.465	2.607
	Free Service	127.941	34.948	.002	54.795	201.087	3.012

Παρατηρούμε στον πίνακα πως το p-value που ελέγχει αν ο συντελεστής του X_5 είναι μηδέν έχει τιμή 0.002. Επομένως απορρίπτουμε τη μηδενική υπόθεση και συμπεραίνουμε πως η παροχή δωρεάν service σε ένα κατάστημα επηρεάζει(θετικά) τα ετήσια κέρδη του.

iii.) Θέλουμε να εξετάσουμε σε επίπεδο σημαντικότητας 5% αν το απόθεμα προϊόντων που το κατάστημα διαθέτει επηρεάζει σε διαφορετικό βαθμό τα κέρδη ενός καταστήματος ανάλογα με το αν στο κατάστημα παρέχεται δωρεάν service ή όχι. Για την επίτευξη αυτού του στόχου πρέπει να αναπροσαρμόσουμε το μοντέλο παλινδρόμησης:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_4 * X_4 + \beta_5 * X_5 + \beta_6 * X_2 X_5 + \varepsilon$$

, όπου $X_5 = 1$ αν $X_1 \geq 4$

και $X_5 = 0$ διαφορετικά

και μένει να ελέγξουμε αν η β_6 είναι στατιστικά σημαντική.

Model		Unstandardized Coefficients		Sig.	95.0% Confidence Interval for B	
		B	Std. Error		Lower Bound	Upper Bound
1	(Constant)	157.839	56.122	.012	39.931	275.746
	Square fit	8.882	5.088	.098	-1.808	19.572
	Inventory	.921	.188	.000	.527	1.315
	Competing stores	-18.197	2.150	.000	-22.714	-13.680
	Free Service	309.457	68.843	.000	164.823	454.091
	X2_X5	-.619	.212	.009	-1.065	-.174

Με το p-value στη τελευταία γραμμή να ισούται με 0.009 απορρίπτουμε τη μηδενική υπόθεση και συνάγουμε πως το απόθεμα των προϊόντων επηρεάζει σε διαφορετικό βαθμό τα έσοδα του καταστήματος ανάλογα με το αν στο κατάστημα παρέχεται δωρεάν service ή όχι. Συγκεκριμένα, αν σε ένα κατάστημα παρέχεται δωρεάν service (+309 χιλιάδες), για κάθε κιβώτιο προϊόντος τα ετήσια κέρδη αναμένονται να πέφτουν κατά 620 ευρώ (διατηρώντας τις υπόλοιπες επεξηγηματικές σταθερές).

Ε.) Για το παρακάτω μοντέλο θέλουμε να κάνουμε μερικές αναλύσεις.

$$E(Y) = \beta_0 + \beta_2 X_2 + \beta_5 X_5 + \beta_6 X_2 X_5, \text{ όπου } X_5 = 1 \text{ αν } X_1 \geq 4$$

και $X_5 = 0$ διαφορετικά

ι.) Θέλουμε να υπολογίσουμε την διαφορά των αναμενόμενων ετήσιων κερδών ενός καταστήματος στο οποίο παρέχεται δωρεάν service από τα αναμενόμενα ετήσια κέρδη ενός καταστήματος στο οποίο δεν παρέχεται δωρεάν service. Επιπλέον γνωρίζουμε πως και τα δύο καταστήματα έχουν απόθεμα **300 κιβωτίων**.

Αναμενόμενα ετήσια κέρδη σε κατάστημα που ΔΕΝ παρέχεται δωρεάν service	Αναμενόμενα ετήσια κέρδη σε κατάστημα που παρέχεται δωρεάν service
109.319 ευρώ	276.046 ευρώ

Οπότε η διαφορά (δ_1) θα ισούται με **166.727** ευρώ.

ii .) Θέλουμε να υπολογίσουμε την διαφορά των αναμενόμενων ετήσιων κερδών ενός καταστήματος στο οποίο παρέχεται δωρεάν service από τα αναμενόμενα ετήσια κέρδη ενός καταστήματος στο οποίο δεν παρέχεται δωρεάν service. Επιπλέον γνωρίζουμε πως και τα δύο καταστήματα έχουν αποθέμα **500 κιβωτίων**.

Αναμενόμενα ετήσια κέρδη σε κατάστημα που ΔΕΝ παρέχεται δωρεάν service	Αναμενόμενα ετήσια κέρδη σε κατάστημα που παρέχεται δωρεάν service
188.219 ευρώ	423.084 ευρώ

Επομένως η διαφορά(δ_2) θα ισούται με **234.865** .

iii .) Κάποιος ισχυρίζεται ότι η διαφορά δ_1 είναι μισή από τη διαφορά δ_2 . Θα ελέγξουμε σε επίπεδο σημαντικότητας 5% εάν ο ισχυρισμός είναι αληθής.

$$\delta_1 = 2 * \delta_2 \Rightarrow 2 * \beta_5 + 600 * \beta_6 = \beta_5 + 500 * \beta_6 \Rightarrow \beta_5 + 100 * \beta_6 = 0 \Rightarrow$$

$$\beta_5 = -100 * \beta_6$$

Οπότε στο προσαρμοσμένο μοντέλο θα έχουμε:

$$Y = \beta_0 + \beta_2 X_2 - 100\beta_6 X_5 + \beta_6 X_2 X_5 + \varepsilon \Rightarrow$$

$$Y = \beta_0 + \beta_2 X_2 + \beta_6 (-100X_5 + X_2 X_5)$$

Για $X_6 = -100X_5 + X_2 X_5$ έχουμε τον εξής πίνακα ANOVA:

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1651285.005	2	825642.503	108.419	.000 ^b
	Residual	159920.953	21	7615.283		
	Total	1811205.958	23			

Ενώ για το πλήρες μοντέλο έχουμε:

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1658592.894	3	552864.298	72.453	.000 ^b
	Residual	152613.064	20	7630.653		
	Total	1811205.958	23			

Συνεπώς μπορούμε να υπολογίσουμε την τιμή της στατιστικής συνάρτησης F:

$$F = \frac{\frac{SSE_{(H_0)} - SSE}{d_0 - d}}{\frac{SSE}{d}} \Rightarrow F = \frac{\frac{159.920,953 - 152.613,064}{21 - 20}}{\frac{152.613,064}{21}} \Rightarrow$$

$$F = 0.957$$

Το $F_{(1,20)(0.95)} = 4.35$ επομένως σε επίπεδο σημαντικότητας 5% ΔΕΝ απορρίπτουμε τη μηδενική υπόθεση. Δεχόμαστε δηλαδή ότι ο ισχυρισμός του ατόμου είναι σωστός.