



Τρίτη Εργαστηριακή Άσκηση

Περιγραφική Στατιστική, Συσχέτιση και Πολλαπλή Γραμμική Παλινδρόμηση με το SPSS

Στον παρακάτω πίνακα παρουσιάζονται δεδομένα από 36 πελάτες ενός φυσικού καταστήματος. Για κάθε πελάτη καταγράφονται το φύλο του (1 = Άνδρας, 0 = Γυναίκα), το συνολικό ποσό αγορών που έχει πραγματοποιήσει στο κατάστημα (Total Purchase, σε €), το ετήσιο εισόδημά του (Yearly Revenue, σε €) και η ετήσια συχνότητα επισκέψεων στο φυσικό κατάστημα (Frequency of Visits to the Physical Store per Year).

A/A	Total Purchase	Yearly Revenue	Gender	Frequency of Visits to the Physical Store per year	A/A	Total Purchase	Yearly Revenue	Gender	Frequency of Visits to the Physical Store per year
1	5.000	100.000	1	15	19	0	100.000	0	37
2	1.700	25.000	0	2	20	10.000	25.000	0	20
3	6.000	85.000	0	11	21	7.500	105.000	0	10
4	2.000	20.000	1	25	22	18.000	95.000	1	10
5	7.000	105.000	0	16	23	16.000	110.000	0	21
6	3.000	30.000	1	12	24	12.000	100.000	1	30
7	8.000	75.000	1	45	25	11.000	105.000	1	15
8	4.000	20.000	0	12	26	20.000	95.000	1	20
9	9.000	110.000	1	9	27	21.000	110.000	1	36
10	5.000	43.000	0	10	28	22.000	100.000	1	25
11	7.500	100.000	1	60	29	23.000	105.000	1	50
12	6.000	25.000	0	5	30	24.000	95.000	0	35
13	6.500	105.000	0	22	31	25.000	110.000	0	30
14	7.000	30.000	0	15	32	26.000	100.000	0	45
15	6.000	95.000	1	35	33	27.000	105.000	0	50
16	8.000	20.000	0	20	34	28.000	95.000	0	52
17	6.500	110.000	1	15	35	29.000	110.000	1	121
18	9.000	60.000	0	30	36	30.000	100.000	0	100

Ερωτήματα

1. Να καταχωρήσετε τα δεδομένα στο **SPSS**.
2. Να ονομάσετε τις μεταβλητές, να ορίσετε τις κατάλληλες ετικέτες τιμών (όπου απαιτείται) και να επιλέξετε την κατάλληλη κλίμακα μέτρησης για καθεμία.
3. Να υπολογίσετε τα κυριότερα μέτρα θέσης της μεταβλητής **Total Purchase**.
4. Να υπολογίσετε τα κυριότερα μέτρα διασποράς της μεταβλητής **Total Purchase**.
5. Να υπολογίσετε τα κυριότερα μέτρα μορφής (ασυμμετρία και κύρτωση) της μεταβλητής **Total Purchase**.
6. Να κατασκευάσετε τον πίνακα συχνοτήτων της μεταβλητής **Gender** και τα αντίστοιχα διαγράμματα.
7. Να υπολογίσετε τα κυριότερα μέτρα θέσης της μεταβλητής **Total Purchase** για κάθε φύλο ξεχωριστά και να κατασκευάσετε κατάλληλο διάγραμμα.
8. Να κατασκευάσετε τα διαγράμματα διασποράς (Scatter Plots) μεταξύ των μεταβλητών **Yearly Revenue** και **Total Purchase**, καθώς και μεταξύ των **μεταβλητών Frequency of Visits to the Physical Store per Year** και **Total Purchase**.
9. Να κατασκευάσετε τον πίνακα συσχέτισης **Pearson** για όλες τις ποσοτικές μεταβλητές και να σχολιάσετε την κατεύθυνση (θετική ή αρνητική) και την ισχύ των συσχετίσεων. Με βάση τον πίνακα συσχέτισης **Pearson**, να εντοπίσετε το ζεύγος μεταβλητών που παρουσιάζει την ισχυρότερη γραμμική συσχέτιση και να αναφέρετε αν αυτή είναι θετική ή αρνητική.
10. Να κατασκευάσετε μοντέλο πολλαπλής γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή τη **Total Purchase** και ανεξάρτητες μεταβλητές τις **Yearly Revenue** και **Frequency of Visits to the Physical Store per Year**.
11. Να γράψετε την εκτιμημένη εξίσωση της πολλαπλής γραμμικής παλινδρόμησης.
12. Με βάση τη στήλη Sig. του πίνακα Coefficients, να αναφέρετε ποιες ανεξάρτητες μεταβλητές επηρεάζουν στατιστικά σημαντικά τη μεταβλητή **Total Purchase**, χρησιμοποιώντας ως επίπεδο σημαντικότητας το 5%.
13. Να χρησιμοποιήσετε το μοντέλο για να εκτιμήσετε το Total Purchase ενός πελάτη με συγκεκριμένα χαρακτηριστικά που θα σας δοθούν.

Αποτελέσματα – Σχολιασμός

Συντελεστές Γραμμικής Συσχέτισης Pearson

Correlations

		Total Purchase	Yearly Revenue	Frequency of Visits to the Physical Store per year
Total Purchase	Pearson Correlation	1	.512**	.640***
	Sig. (2-tailed)		.001	<.001
	N	36	36	36
Yearly Revenue	Pearson Correlation	.512**	1	.372*
	Sig. (2-tailed)	.001		.026
	N	36	36	36
Frequency of Visits to the Physical Store per year	Pearson Correlation	.640***	.372*	1
	Sig. (2-tailed)	<.001	.026	
	N	36	36	36

** . Correlation is significant at the 0.01 level (2-tailed).

*** . Correlation at 0.001(2-tailed)

* . Correlation is significant at the 0.05 level (2-tailed).

- Από τον πίνακα συσχέτισης Pearson προκύπτει ότι το **Total Purchase** παρουσιάζει μέτρια θετική γραμμική συσχέτιση με το **Yearly Revenue** ($r = 0.512$) και ισχυρότερη θετική γραμμική συσχέτιση με τη **Frequency of Visits to the Physical Store per year** ($r = 0.640$). Επιπλέον, το **Yearly Revenue** και η **Frequency of Visits** εμφανίζουν επίσης θετική συσχέτιση ($r = 0.372$).
- Όλες οι συσχετίσεις είναι στατιστικά σημαντικές, καθώς οι αντίστοιχες τιμές *Sig.* είναι μικρότερες από 0,05.

Πολλαπλή Γραμμική Παλινδρόμηση

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.704 ^a	.496	.465	6667.941

a. Predictors: (Constant), Frequency of Visits to the Physical Store per year, Yearly Revenue

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1443915784.4	2	721957892.22	16.238	<.001 ^b
	Residual	1467227271.1	33	44461432.458		
	Total	2911143055.6	35			

a. Dependent Variable: Total Purchase

b. Predictors: (Constant), Frequency of Visits to the Physical Store per year, Yearly Revenue

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	32.384	2957.065		.011	.991
	Yearly Revenue	.086	.036	.318	2.387	.023
	Frequency of Visits to the Physical Store per year	191.433	48.878	.521	3.917	<.001

a. Dependent Variable: Total Purchase

- Το μοντέλο πολλαπλής γραμμικής παλινδρόμησης παρουσιάζει συντελεστή προσδιορισμού $R^2 = 0,496$, γεγονός που σημαίνει ότι περίπου το 49,6% της μεταβλητότητας του συνολικού ποσού αγορών (Total Purchase) ερμηνεύεται από τις μεταβλητές Yearly Revenue και Frequency of Visits to the Physical Store per year.
- Το *Std. Error of the Estimate* = 6667,941 εκφράζει το τυπικό σφάλμα εκτίμησης του μοντέλου και δείχνει ότι οι προβλεπόμενες τιμές του Total Purchase αποκλίνουν από τις πραγματικές τιμές κατά περίπου 6.668 € κατά μέσο όρο.
- Η εκτιμημένη εξίσωση της παλινδρόμησης είναι:

$$\text{Total Purchase} = 32.384 + 0.086(\text{Yearly Revenue}) + 191.433(\text{Frequency of Visits}).$$

- Και οι δύο ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές, αφού οι τιμές Sig. είναι μικρότερες από 0,05 (0,023 και <0,001, αντίστοιχα). Επομένως, τόσο το ετήσιο εισόδημα όσο και η συχνότητα επισκέψεων επηρεάζουν σημαντικά το συνολικό ποσό αγορών.