



ΕΛΛΗΝΙΚΗ
ΔΗΜΟΚΡΑΤΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΜΑΚΕΔΟΝΙΑΣ

**Λυμένες Ασκήσεις για το μάθημα:
ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΜΕ ΧΡΗΣΗ Η/Υ**

ΚΩΝΣΤΑΝΤΙΝΟΣ ΖΑΦΕΙΡΟΠΟΥΛΟΣ

Τμήμα: ΔΙΕΘΝΩΝ ΚΑΙ ΕΥΡΩΠΑΪΚΩΝ ΣΠΟΥΔΩΝ

Άδειες Χρήσης

Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons. Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα. Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Μακεδονίας**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.



Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα: Ποιος είναι ο μέσος πληθυσμός και η τυπική του απόκλιση για τις χώρες της Αφρικής

Λύση:

Επιλέγουμε-θέτουμε με χρήση της SELECT, region=4

Εκτελούμε την Analyze->Descriptive Statistics->Descriptives για την μεταβλητή populatn

	N	Minimum	Maximum	Mean	Std. Deviation
populatn πληθυσμός σε χιλιάδες	19	959	98100	19757,11	24357,858
Valid N (listwise)	19				

Τα ζητούμενα είναι το Mean=19757 και Std. Deviation=24357

Φροντίστε τελειώνοντας να διαγράψετε το φίλτρο.

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα: Ποιος είναι ο μέσος πληθυσμός και η τυπική απόκλιση για τις χώρες της Λατινικής Αμερικής

Λύση:

Επιλέγουμε-θέτουμε με χρήση της SELECT, region=6

Εκτελούμε την Analyze->Descriptive Statistics->Descriptives για την μεταβλητή populatn

Όπως κι στο προηγούμενο ερώτημα

	N	Minimum	Maximum	Mean	Std. Deviation
populatn πληθυσμός σε χιλιάδες	21	256	156600	21928,86	36806,778
Valid N (listwise)	21				

Φροντίστε τελειώνοντας να διαγράψετε το φίλτρο.

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα: Πόσες χώρες με μεσογειακό κλίμα είναι Μουσουλμανικές

Λύση:

Επιλέγουμε-θέτουμε με χρήση της SELECT, climate=6

Εκτελούμε την Analyze->Descriptive Statistics->Frequencies για την μεταβλητή religion

religion κύρια θρησκεία

	Frequency	Percent	Valid Percent	Cumulative Percent
Animist	1	10,0	10,0	10,0
Buddhist	1	10,0	10,0	20,0
Catholic	3	30,0	30,0	50,0
Valid Hindu	1	10,0	10,0	60,0
Muslim	3	30,0	30,0	90,0
Orthodox	1	10,0	10,0	100,0
Total	10	100,0	100,0	

3 χώρες (10% των χωρών με μεσογειακό κλίμα) είναι Μουσουλμανικές

Φροντίστε τελειώνοντας να διαγράψετε το φίλτρο.

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα: Υπάρχει στατιστικά σημαντική διαφορά της παιδικής θνησιμότητας ανάμεσα στις χώρες της Αφρικής και τις χώρες της Μέσης Ανατολής;

Λύση:

Η μεταβλητή region παίρνει την τιμή 4 και 5 αντίστοιχα για την Αφρική και την Μέση Ανατολή.

Η ποσοτική μεταβλητή που θα αναλύσουμε είναι η παιδικής θνησιμότητα babydeath.

Πρόκειται για έλεγχο υποθέσεων.

Ξεκινάμε με έλεγχο κανονικότητας.

Εκτελούμε την Analyze->Descriptive Statistics->Explore, εισάγοντας στην dependent list την babydeath και στη factor list την region. Επιλέγουμε στα plots το normality plots with tests.

Tests of Normality							
	region περιοχή	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
babydeath παιδική θνησιμότητα (deaths per 1000 live births)	1 OECD	,130	21	,200 [*]	,969	21	,710
	2 East Europe	,150	14	,200 [*]	,957	14	,672
	3 Pacific/Asia	,163	17	,200 [*]	,901	17	,071
	4 Africa	,183	19	,093	,932	19	,187
	5 Middle East	,069	17	,200 [*]	,984	17	,984
	6 Latin America	,151	21	,200 [*]	,907	21	,047

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Στον παραπάνω πίνακα βλέπουμε τα df για την Αφρική και την Μέση Ανατολή είναι μικρότερα του 50. Επιλέγουμε λοιπόν τα τεστ των Shapiro-Wilk. Επειδή και για τα δύο τα p είναι μεγαλύτερα από 0,05 έχουμε κανονικότητα. Μπορούμε λοιπόν να προχωρήσουμε με t-test ανεξάρτητων δειγμάτων.

Υποθέσεις:

H_0 : οι μέσες τιμές παιδικής θνησιμότητας είναι ίσες ανάμεσα σε χώρες της Αφρικής και Μέσης Ανατολής

H_1 : οι μέσες τιμές δεν είναι ίσες

Επιλέγουμε Analyze-compare means-independent samples t-test. Θέτουμε στο test variable το `babymort` και στο grouping variable το `region(4,5)`

	region περιοχή	N	Mean	Std. Deviation	Std. Error Mean
babymort παιδική θνησιμότητα (deaths per 1000 live births)	4 Africa	19	94,179	28,6532	6,5735
	5 Middle East	17	41,394	19,1789	4,6516

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
babymort παιδική θνησιμότητα (deaths per 1000 live births)	Equal variances assumed	4,258	,047	6,414	34	,000	52,7848	8,2302	36,0590	69,5106
	Equal variances not assumed			6,555	31,620	,000	52,7848	8,0528	36,3741	69,1956

Επειδή το sig του τεστ του Levene είναι $0,047 < 0,05$ συμπεραίνουμε ότι έχουμε άνισες διασπορές. Συνεχίζουμε την ανάγνωση του πίνακα στην δεύτερή του γραμμή

$t=6,555$

$df=31$

$p=0,000 < 0,05$ άρα απορρίπτουμε την μηδενική υπόθεση.

Η απάντηση στην άσκηση είναι ότι υπάρχει στατιστικά σημαντική διαφορά της παιδικής θνησιμότητας ανάμεσα σε χώρες της Αφρικής και Μέσης Ανατολής

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα: Υπάρχει στατιστικά σημαντική διαφορά του ΑΕΠ ανάμεσα στις χώρες της Αφρικής, της Λατινικής Αμερικής και της Ασίας/Ειρηνικού;

Λύση:

Η μεταβλητή region παίρνει την τιμή 4, 6 και 3 αντίστοιχα για την Αφρική, τη Λατινική Αμερική και την Ασία/Ειρηνικό.

Η ποσοτική μεταβλητή που θα αναλύσουμε είναι η gdp_cap.

Πρόκειται για έλεγχο υποθέσεων. Ξεκινάμε με έλεγχο κανονικότητας.

Εκτελούμε την Analyze->Descriptive Statistics->Explore, εισάγοντας στην dependent list την gdp_cap και στη factor list την region. Επιλέγουμε στα plots το normality plots with tests.

Tests of Normality							
	region περιοχή	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
gdp_cap ΑΕΠ	1 OECD	,192	21	,042	,926	21	,114
	2 East Europe	,141	14	,200 [*]	,936	14	,367
	3 Pacific/Asia	,299	17	,000	,695	17	,000
	4 Africa	,273	19	,001	,709	19	,000
	5 Middle East	,156	17	,200 [*]	,865	17	,018
	6 Latn America	,193	21	,039	,820	21	,001

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Επειδή οι df είναι λιγότεροι από 50, αναφερόμαστε το τεστ των Shapiro-Wilk. Τα αντίστοιχα p είναι μικρότερα από 0,05 άρα δεν έχουμε κανονικότητα.

Θα εκτελέσουμε μη παραμετρικό έλεγχο Kruskal-Wallis

Υποθέσεις:

H_0 : υπάρχει ομοιογένεια ανάμεσα στις τρεις περιοχές ως προς το ΑΕΠ

H_1 : δεν υπάρχει ομοιογένεια

Για να εργαστούμε θα πρέπει να επιλέξουμε με την εντολή select τις τρεις περιοχές

Επιλέγουμε-θέτουμε με χρήση της SELECT, region=4 or region=6 or region=3

και στη συνέχεια Analyze->Nonparametric tests->Independent samples. Στο Fields θέτουμε το gdp_cap στο test fields και το region στο groups.

Στα settings επιλέγουμε customize tests και kruskal-Wallis:

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of ΑΕΠ is the same across categories of περιοχή.	Independent-Samples Kruskal-Wallis Test	,015	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Επειδή $p=0,015 < 0,05$ απορρίπτουμε τη μηδενική υπόθεση,

άρα υπάρχει διαφορά του ΑΕΠ ανάμεσα στις τρεις περιοχές.

Αφαιρέστε το φίλτρο.

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα:

Υπολογίστε και σχολιάστε τους συντελεστές συσχέτισης ανάμεσα στα: παιδική θνησιμότητα, το ποσοστό γεννήσεων ανά 1000 κατοίκους και το ποσοστό θανάτων ανά 1000 κατοίκους.

Λύση:

Χρησιμοποιούμε τις μεταβλητές bebymort, birth_rt, death_rt

Analyze->Correlate->Bivariate

		Correlations		
		bebymort παιδική θνησιμότητα (deaths per 1000 live births)	birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	death_rt ποσοστό θανάτων ανά 1000 κατοίκους
bebymort παιδική θνησιμότητα (deaths per 1000 live births)	Pearson Correlation Sig. (2-tailed) N	1	,865** ,000 109	,630** ,000 108
birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	Pearson Correlation Sig. (2-tailed) N	,865** ,000 109	1 ,367** 109	,367** ,000 108
death_rt ποσοστό θανάτων ανά 1000 κατοίκους	Pearson Correlation Sig. (2-tailed) N	,630** ,000 108	,367** ,000 108	1 ,000 108

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

Ο συντελεστής συσχέτισης ανάμεσα στις bebymort και birth_rt είναι $r=0,865^{**}$ υψηλή θετική συσχέτιση και στατιστικά σημαντική σε $p=0,01$

Ο συντελεστής συσχέτισης ανάμεσα στις bebymort και death_rt είναι $r=0,63^{**}$ μέτρια θετική συσχέτιση και στατιστικά σημαντική σε $p=0,01$

Ο συντελεστής συσχέτισης ανάμεσα στις birth_rt και death_rt είναι $r=0,367^{**}$ χαμηλή θετική συσχέτιση και στατιστικά σημαντική σε $p=0,01$

Το αρχείο world95.sav περιλαμβάνει κοινωνικούς και οικονομικούς δείκτες για διάφορες χώρες για το 1995. Αφού δείτε τις μεταβλητές και τα δεδομένα του αρχείου απαντήστε στο ερώτημα:

Να υπολογιστεί το μοντέλο παλινδρόμησης με τη μέθοδο stepwise, με εξαρτημένη μεταβλητή την παιδική θνησιμότητα και ανεξάρτητες το ποσοστό γεννήσεων ανά 1000 κατοίκους, το ποσοστό θανάτων ανά 1000 κατοίκους, το ποσοστό αλφαριθμητισμού και το ποσοστό ανθρώπων που ζουν σε πόλεις. Να γίνει σχολιασμός όλων των δεικτών του τελικού μοντέλου.

Λύση:

Εξαρτημένη μεταβλητή είναι η babymort και ανεξάρτητες οι birth_rt, death_rt, literacy και urban.

Επιλέγουμε Analyze->regression->linear

Θέτουμε την babymort στη λίστα dependents και τις υπόλοιπες μεταβλητές στη λίστα independents.

Επιλέγουμε Method >Stepwise.

Σχολιάζουμε το τελικό μοντέλο (4^ο μοντέλο).

Το R Square ισούται με 0,912 που σημαίνει ότι όλες οι ανεξάρτητες μεταβλητές (και οι τέσσερις) που έχουν εισαχθεί στο μοντέλο ερμηνεύουν το 91,2% της διασποράς της εξαρτημένης μεταβλητής. Άρα έχουμε μια πολύ καλή ερμηνευτική ικανότητα του μοντέλου.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,901 ^a	,811	,810	16,7144
2	,929 ^b	,864	,861	14,2656
3	,950 ^c	,902	,899	12,1434
4	,955 ^d	,912	,908	11,5936

a. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού

b. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους

c. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους, birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους

d. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους, birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους, urban % ανθρώπων που ζουν σε πόλεις

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	124987,799	1	124987,799	447,388	,000 ^b
	Residual	29054,737	104	279,372		
	Total	154042,537	105			
2	Regression	133081,409	2	66540,705	326,972	,000 ^c
	Residual	20961,127	103	203,506		
	Total	154042,537	105			
3	Regression	139001,438	3	46333,813	314,209	,000 ^d
	Residual	15041,098	102	147,462		
	Total	154042,537	105			
4	Regression	140466,958	4	35116,739	261,263	,000 ^e
	Residual	13575,579	101	134,412		
	Total	154042,537	105			

a. Dependent Variable: babymort παιδική θνησιμότητα (deaths per 1000 live births)

b. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού

c. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους

d. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους, birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους

e. Predictors: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους, birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους, urban % ανθρώπων που ζουν σε πόλεις

Η τελική ANOVA (4^ο μοντέλο) έχει $p=0,000 < 0,05$ και είναι στατιστικά σημαντική. Αυτό σε συνδυασμό με τον υψηλό συντελεστή προσδιορισμού που είδαμε, τεκμηριώνουν ότι έχουμε καλή προσαρμογή του μοντέλου στα δεδομένα.

Περεταίρω, ελέγχουμε τον πίνακα Coefficients με τους συντελεστές του μοντέλου και τα στατιστικά τους.

Το μοντέλο είναι (από τη στήλη B)

$$\text{Babymort} = 55,463 - 0,638 * \text{literacy} + 2,177 * \text{death_rt} + 1,114 * \text{birth_rt} - 0,218 * \text{urban}$$

Επίσης όλοι οι συντελεστές είναι στατιστικά σημαντικοί αφού $p < 0,05$ (Sig)

t	Sig.
3,898	,000
-5,965	,000
6,793	,000
5,904	,000
-3,302	,001

Παρατηρώντας τα Beta ταξινομούμε της ανεξάρτητες μεταβλητές (κατ' απόλυτη τιμή των BETA) ως προς την ερμηνευτική τους ικανότητα.

literacy ποσοστό αλφαριθμητισμού	-,382
birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	,359
death_rt ποσοστό θανάτων ανά 1000 κατοίκους	,243
urban % ανθρώπων που ζουν σε πόλεις	-,137

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	160,572	5,789		27,735	,000
	literacy ποσοστό αλφαριθμητισμού	-1,503	,071	-,901	-21,152	,000
2	(Constant)	121,385	7,939		15,290	,000
	literacy ποσοστό αλφαριθμητισμού	-1,290	,069	-,773	-18,597	,000
	death_rt ποσοστό θανάτων ανά 1000 κατοίκους	2,354	,373	,262	6,306	,000
3	(Constant)	41,812	14,261		2,932	,004

	literacy ποσοστό αλφαριθμητισμού	-,701	,110	-,420	-6,365	,000
	death_rt ποσοστό θανάτων ανά 1000 κατοίκους	2,510	,319	,280	7,876	,000
	birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	1,230	,194	,397	6,336	,000
	(Constant)	55,463	14,230		3,898	,000
	literacy ποσοστό αλφαριθμητισμού	-,638	,107	-,382	-5,965	,000
4	death_rt ποσοστό θανάτων ανά 1000 κατοίκους	2,177	,320	,243	6,793	,000
	birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	1,114	,189	,359	5,904	,000
	urban % ανθρώπων που ζουν σε πόλεις	-,218	,066	-,137	-3,302	,001

a. Dependent Variable: babymort παιδική θνησιμότητα (deaths per 1000 live births)

Excluded Variables^a

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics	
					Tolerance	
1	birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	,359 ^b	4,553	,000	,409	,246
	death_rt ποσοστό θανάτων ανά 1000 κατοίκους	,262 ^b	6,306	,000	,528	,764
	urban % ανθρώπων που ζουν σε πόλεις	-,256 ^b	-5,088	,000	-,448	,579
2	birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους	,397 ^c	6,336	,000	,531	,244
	urban % ανθρώπων που ζουν σε πόλεις	-,182 ^c	-3,884	,000	-,359	,528
3	urban % ανθρώπων που ζουν σε πόλεις	-,137 ^d	-3,302	,001	-,312	,510

a. Dependent Variable: babymort παιδική θνησιμότητα (deaths per 1000 live births)

b. Predictors in the Model: (Constant), literacy ποσοστό αλφαριθμητισμού

c. Predictors in the Model: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους

d. Predictors in the Model: (Constant), literacy ποσοστό αλφαριθμητισμού, death_rt ποσοστό θανάτων ανά 1000 κατοίκους, birth_rt ποσοστό γεννήσεων ανά 1000 κατοίκους