

ΜΑΘΗΜΑ : ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΕΡΕΥΝΑΣ

Διδάσκων : ΦΛΩΡΟΥ ΓΙΑΝΝΟΥΛΑ

ΕΞΕΤΑΣΤΙΚΗ ΠΕΡΙΟΔΟΣ Α	ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ 7/2/2020	ΔΙΑΡΚΕΙΑ ΕΞΕΤΑΣΗΣ 3 ώρες
<u>ΟΔΗΓΙΕΣ</u>		
<ul style="list-style-type: none"> • Η εξέταση γίνεται με κλειστά βιβλία-σημειώσεις. • Τα θέματα να επιστραφούν στο τέλος της εξέτασης. • Ο/Η εξεταζόμενος/η μπορεί να αποχωρήσει το νωρίτερο 20 λεπτά από την παράδοση των θεμάτων. • Απαγορεύεται η χρήση κινητών τηλεφώνων. 		

**ΟΛΑ ΤΑ ΘΕΜΑΤΑ ΕΙΝΑΙ ΙΣΟΒΑΘΜΑ ΜΕ 2 ΜΟΝΑΔΕΣ ΤΟ ΚΑΘΕΝΑ.
ΑΠΑΝΤΗΣΤΕ 5 ΑΠΟ ΤΑ 6 ΘΕΜΑΤΑ**

ΘΕΜΑΤΑ ΕΞΕΤΑΣΕΩΝ

ΘΕΜΑ 1.

α) Ποιες οι προσεγγίσεις μιας πρωτογενούς έρευνας και ποια σημαντικά χαρακτηριστικά πρέπει να έχει η ακολουθία των ερωτήσεων σε ένα ερωτηματολόγιο;

β) Ποια τα πλεονεκτήματα της ταχυδρομικής έρευνας (ταχυδρομούμενα ερωτηματολόγια) και πως μπορούμε να περιορίσουμε την μη απόκριση;

γ) Ένας οργανισμός σας έχει αναθέσει να σχεδιάζετε ένα ερωτηματολόγιο για μία δημοσκόπηση σχετικά με την ικανοποίηση των πελατών του από τις υπηρεσίες του. Διατυπώστε δύο ερωτήματα χρησιμοποιώντας την κλίμακα Likert. Στη συνέχεια, αναδιατυπώστε τα ερωτήματα αυτά ώστε να χρησιμοποιούν κλίμακα σημαντικού διαφορισμού.

ΘΕΜΑ 2.

α) Σας δίνονται οι παράμετροι θέσης και διασποράς για τη μεταβλητή "Εισόδημα" σε τέσσερα δείγματα διαφορετικών ηλικιών όπως προέκυψαν από το πρόγραμμα SPSS. Υπάρχει διαφορά στη θέση των δειγμάτων; Ποιο από τα δείγματα έχει μεγαλύτερη διασπορά; Ποιο μεγαλύτερη ομοιογένεια;

Case Summaries

Εισόδημα (χιλιάδες)

ηλικία	N	Mean	Median	Std. Deviation	Minimum	Maximum
Κάτω από 25	589	9,4630	9,7000	1,96623	4,20	12,60
25 - 49	1130	17,6634	17,1000	3,58871	11,90	25,70
50 - 74	567	30,2125	29,8000	3,52136	23,60	37,90
75 και άνω	824	60,9275	59,3500	16,16260	36,70	99,80
Σύνολο	3110	29,8611	22,3000	21,57632	4,20	99,80

β) Αν υποθέσουμε ότι και τα δύο δείγματα (δεύτερο και τρίτο) προέρχονται από πληθυσμούς που ακολουθούν κανονική κατανομή, με άγνωστες αλλά ίσες διακυμάνσεις, μπορούμε να ισχυριστούμε ότι οι μέσες τιμές των αντίστοιχων πληθυσμών δεν διαφέρουν; Δικαιολογήστε την απάντησή σας με στάθμη σημαντικότητας 95%.

Μπορείτε να χρησιμοποιήσετε και τον παρακάτω πίνακα του SPSS για να δικαιολογήσετε την απάντησή σας.

Independent Samples Test

	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
Εισόδημα Equal variances assumed	,275	,600	-68,372	1695	,000	-12,54916	,18354	-12,90915	-12,18917
Equal variances not assumed			-68,803	1152,711	,000	-12,54916	,18239	-12,90702	-12,19130

ΘΕΜΑ 3.

Σε διαφορετικές ομάδες ατόμων, μετρήθηκε η "Εισόδημα". Στη συνέχεια εφαρμόστηκε η μέθοδος της Ανάλυση Διακύμανσης ενός παράγοντα (One-way Anova), σύμφωνα με το εντελώς τυχαιοποιημένο σχέδιο ανάλυσης.

α) Γράψτε την αρχική υπόθεση ελέγχου της Ανάλυσης Διακύμανσης. Πόσα άτομα μετρήθηκαν και πόσες ομάδες;

β) Να αξιολογήσετε τα αποτελέσματα της ανάλυσης σε επίπεδο σημαντικότητας $\alpha=5\%$, και να αποφανθείτε αν οι ομάδες διαφέρουν σε ότι αφορά την Τιμή οχήματος.

γ) Τι προκύπτει από τον πίνακα αποτελεσμάτων πολλαπλών συγκρίσεων;

ANOVA

Price of primary vehicle

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26076,593	4	6519,148	14,242	,000
Within Groups	1421279,297	3105	457,739		
Total	1447355,890	3109			

Εισόδημα

Duncan^{a,b}

Level of education	N	Subset for alpha = 0.05			
		1	2	3	4
Did not complete high school	682	25,8012			
High school degree	973		29,3621		
Some college	643		30,3261	30,3261	
College degree	636			32,3785	
Post-undergraduate degree	176				37,5568
Sig.		1,000	,503	,154	1,000

Means for groups in homogeneous subsets are displayed.

ΘΕΜΑ 4.

Δίνεται ο παρακάτω πίνακας αποτελεσμάτων της εκτίμησης και των διαγνωστικών ελέγχων ενός υποδείγματος για την σχέση του δείκτη τιμών μετοχών και κάποιων βασικών μακρο-οικονομικών δεικτών όπου,

FDI είναι οι εισροές ξένων επενδύσεων, Y είναι το ΑΕΠ, TAX είναι η φορολογία και EBD δείκτης που εκφράζει την «φιλικότητα» του θεσμικού συστήματος για την δημιουργία νέων επενδύσεων.

Ordinary Least Squares Estimation

Dependent variable is FDI

70 observations used for estimation from 1998Q1 to 2016Q4

```
*****
Regressor                Coefficient      Standard Error      TRatio[Prob]
C                        2.48          1.2                 2.2 [.025]
Y                        .42           .21                -2.16 [.02]
TAX                      -.12          .04                -3.12 [.005]
EBD                      .41           .18                -2.473 [.01]
*****
R-Squared                .887          R-Bar-Squared      .865
*****
```

Diagnostic Tests

```
*****
* Test Statistics *      LM Version *
*****
* A:Serial Correlation*CHSQ( 4)= .21 [.560]*
* B:Functional Form *CHSQ( 1)= 1.13 [.183]*
* C:Normality *CHSQ( 2)= .73 [.698]*
* D:Heteroscedasticity*CHSQ( 1)= 1.88 [.128]*
*****
```

α) Να διατυπώσετε την εκτιμημένη μορφή του υποδείγματος

β) Να δώσετε την ερμηνεία των συντελεστών

γ) Να σχολιάσετε τον συντελεστή προσδιορισμού R²

δ) Να σχολιάσετε την σημαντικότητα των ανεξαρτήτων μεταβλητών του υποδείγματος.

ε) Να σχολιάσετε την αξιοπιστία του υποδείγματος με βάση τους βασικούς διαγνωστικούς ελέγχους.

ΘΕΜΑ 5.

α) Σε μια έρευνα για την σχέση του φύλου με την επιτυχία στις εξετάσεις απόκτησης διπλώματος οδήγησης από ένα δείγμα 200 προσφάτως επιτυχόντων οδηγών είχαμε τα εξής:

	ΑΠΟΤΥΧΙΑ	ΕΠΙΤΥΧΙΑ	ΣΥΝΟΛΑ
ΑΝΔΡΑΣ	75	50	125
ΓΥΝΑΙΚΑ	25	50	75
ΣΥΝΟΛΑ	100	100	200

Ο έλεγχος με το SPSS έδωσε τα παρακάτω αποτελέσματα:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13,333 ^a	1	,000		
Continuity Correction ^b	12,288	1	,000		
Likelihood Ratio	13,529	1	,000		
Fisher's Exact Test				,000	,000
N of Valid Cases	200				

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13,333 ^a	1	,000		
Continuity Correction ^b	12,288	1	,000		
Likelihood Ratio	13,529	1	,000		
Fisher's Exact Test				,000	,000
N of Valid Cases	200				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 37,50.

b. Computed only for a 2x2 table

Να ελέγξετε αν φύλο και επιτυχία συσχετίζονται. Δίνεται η κριτική τιμή του ελέγχου $\chi^2 = 3,841$.

β) Α. Ποιά από τις παρακάτω τεχνικές θα παράξει ένα απλό τυχαίο δείγμα των νοσοκομείων;

1. Τυχαία επιλογή περιφερειών, και στη συνέχεια, δειγματοληψία όλων των νοσοκομείων μέσα στην περιφέρεια
2. Απαρίθμηση όλων των νοσοκομείων του δειγματοληπτικού πλαισίου και στη συνέχεια επιλογή με χρήση γεννήτριας τυχαίων αριθμών.
3. Κατηγοριοποίηση των νοσοκομείων ανά περιφέρεια και τυχαία επιλογή ενός συγκεκριμένου ποσοστού από κάθε περιφέρεια.
4. Επιλογή όσων νοσοκομείων προσφερθούν να συμμετέχουν

Β. Ποιό από τα παρακάτω ΔΕΝ ισχύει για τη δειγματοληψία σκοπιμότητας;

1. Η επιλογή των αντικειμένων του δείγματος βασίζεται στην κρίση του ερευνητή
2. Το δείγμα είναι αντιπροσωπευτικό
3. Η επιλογή για το δείγμα των ακραίων περιπτώσεων του υπό μελέτη φαινομένου αποτελεί περίπτωση δειγματοληψίας σκοπιμότητας
4. Η επιλογή για το δείγμα των τυπικών περιπτώσεων είναι αποδεκτή όταν η σκοπιμότητα είναι ο σχηματισμός ενός ενδεικτικού προφίλ

Γ. Ποια από τις παρακάτω δηλώσεις αληθεύει;

1. Όσο μεγαλύτερο είναι το δείγμα, τόσο μεγαλύτερο είναι το διάστημα εμπιστοσύνης
2. Όσο πιο μικρό είναι το δείγμα, τόσο πιο μεγάλο είναι το σφάλμα δειγματοληψίας
3. Όσο πιο πολλές κατηγορίες μετρούνται, τόσο πιο μικρό είναι το μέγεθος του δείγματος
4. Ένα επίπεδο εμπιστοσύνης της τάξεως του 95 τοις εκατό είναι πάντα αρκετό

Δ. Όταν το δείγμα σχηματίζεται από ανθρώπους που είτε εθελοντικά προσφέρθηκαν να συμμετέχουν, ή έτυχε να είναι διαθέσιμοι, πρόθυμοι, ή βολικοί εκείνη την περίοδο, η δειγματοληψία καλείται:

1. Δειγματοληψία χιονοστιβάδας
2. Δειγματοληψία ευκολίας
3. Στρωματοποιημένη δειγματοληψία
4. Τυχαία δειγματοληψία

ΘΕΜΑ 6.

Οι παρακάτω πίνακες παρουσιάζουν τα αποτελέσματα από την παραγοντική ανάλυση η οποία πραγματοποιήθηκε σε 18 μεταβλητές.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0,844
Bartlett's Test of Sphericity	Approx. Chi-Square
	2778,703
	df
	153
	Sig.
	0,000

Total Variance Explained

Compon ent	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6,624	36,799	36,799	6,624	36,799	36,799	3,141	17,451	17,451
2	2,126	11,812	48,611	2,126	11,812	48,611	2,748	15,267	32,718
3	1,627	9,036	57,648	1,627	9,036	57,648	2,537	14,095	46,813
4	1,454	8,080	65,727	1,454	8,080	65,727	2,465	13,696	60,510
5	1,229	6,826	72,554	1,229	6,826	72,554	2,168	12,044	72,554
6	,872	4,846	77,399						
7	,629	3,493	80,892						
8	,557	3,096	83,988						
9	,480	2,664	86,652						
10	,461	2,558	89,211						
11	,402	2,235	91,446						
12	,323	1,795	93,241						
13	,300	1,665	94,906						
14	,251	1,393	96,299						
15	,237	1,316	97,615						
16	,171	,953	98,568						
17	,170	,943	99,511						
18	,088	,489	100,000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix ^a					
Component					
	1	2	3	4	5
1. Για πόσο καιρό χρησιμοποιείτε τις υπηρεσίες Διαδικτύου?	-,219	,170	-,123	,077	,679
2. Πόσο συχνά χρησιμοποιείτε τις υπηρεσίες Διαδικτύου;	-,010	,067	,123	,101	,879
3. Πόσο χρόνο ξοδεύετε συνδεδεμένος με το Διαδίκτυο;	,004	,053	,285	,096	,818
4. Αν ένας ιστοχώρος λαμβάνει τα κατάλληλα μέτρα ασφάλειας δεν διστάζω να εισάγω τα στοιχεία της πιστωτικής μου κάρτας	-,236	,373	,611	,011	,138
5. Εμπιστεύομαι το διαδίκτυο	-,193	,171	,787	,163	,039
6. Όταν κάνω μια συνδιαλλαγή στο διαδίκτυο ξέρω ακριβώς τι θα επακολουθήσει	-,264	,043	,626	,274	,314
7. Εμπιστεύομαι τις διαδικτυακές συναλλαγές	-,150	,177	,844	,155	,007
8. Φοβάμαι ότι μπορεί άλλοι να έχουν πρόσβαση στις διαδικτυακές τραπεζικές μου συναλλαγές.	,778	-,050	-,255	,035	-,024

9.	Πιστεύω ότι αρκετά εύκολα μπορούν να κλαπούν χρήματα με τη χρησιμοποίηση της διαδικτυακής τραπεζικής	,813	-,270	-,167	-,030	-,106
10.	Φοβάμαι πως η εχεμύθεια των οικονομικών μου συναλλαγών μπορεί να χαθεί με την χρησιμοποίηση της διαδικτυακής τραπεζικής	,842	-,278	-,122	-,066	-,082
11.	Πιστεύω πως το απόρητο δεν εγγυάται στην διαδικτυακή τραπεζική	,827	-,157	-,154	-,176	-,076
12.	Θα χρησιμοποιούσα τη Διαδικτυακή Τραπεζική για τις τραπεζικές μου ανάγκες.	-,260	,874	,214	,118	,057
13.	Θα χρησιμοποιήσω την διαδικτυακή τραπεζική για να διαχειριστώ τις τραπεζικές μου συναλλαγές.	-,235	,838	,241	,208	,110
14.	Θα μπορούσα να φανταστώ τον εαυτό μου να χρησιμοποιεί τη διαδικτυακή τραπεζική για τις τραπεζικές μου συναλλαγές.	-,195	,817	,161	,213	,168
15.	Η χρησιμοποίηση του διαδικτύου θα μου δώσει την ευκαιρία να φέρω εις πέρας τα καθήκοντα μου γρηγορότερα.	-,188	,321	,027	,773	,154
16.	Η χρησιμοποίηση του διαδικτύου θα με διευκολύνει στην πραγμάτωση των καθηκόντων μου.	-,198	,248	,074	,788	,168
17.	Θα έβρισκα το διαδίκτυο ενδιαφέρον	-,022	,159	,161	,754	,143
18.	Πάνω από όλα θα έβρισκα την χρήση του διαδικτύου περιπετειώδη.	,113	-,100	,187	,622	-,078

- Είναι τα δεδομένα κατάλληλα για παραγοντική ανάλυση;
- Πόσοι παράγοντες προέκυψαν;
- Με ποιο κριτήριο δημιουργήθηκαν οι παράγοντες;
- Ποιο είναι το ποσοστό της διακύμανσης το οποίο ερμηνεύεται από τους παράγοντες;
- Μπορείτε να προσδιορίσετε σε ποιο παράγοντα ανήκει η κάθε μεταβλητή;

ΤΥΠΟΛΟΓΙΟ

Δίνονται οι τύποι:

$$\text{έλεγχος μέσης τιμής ενός δείγματος } t = \frac{(\bar{X}) - \mu_0}{s_p / \sqrt{n}}$$

$$\text{έλεγχος μέσων τιμών } t = \frac{(\bar{X}_1) - (\bar{X}_2) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{όπου } s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Το διάστημα εμπιστοσύνης $\alpha\%$ είναι:

$$\left[\hat{p} - z_{\alpha/2} \sqrt{p(1-p)/n}, \hat{p} + z_{\alpha/2} \sqrt{p(1-p)/n} \right] \quad \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad \alpha=5\% \quad z_{\alpha/2}=1,96$$

ΚΑΛΗ ΕΠΙΤΥΧΙΑ