

Μέθοδοι Ποιοτικής & Ποσοτικής Έρευνας

Μάθημα 12^ο - Κεφάλαιο 12^ο

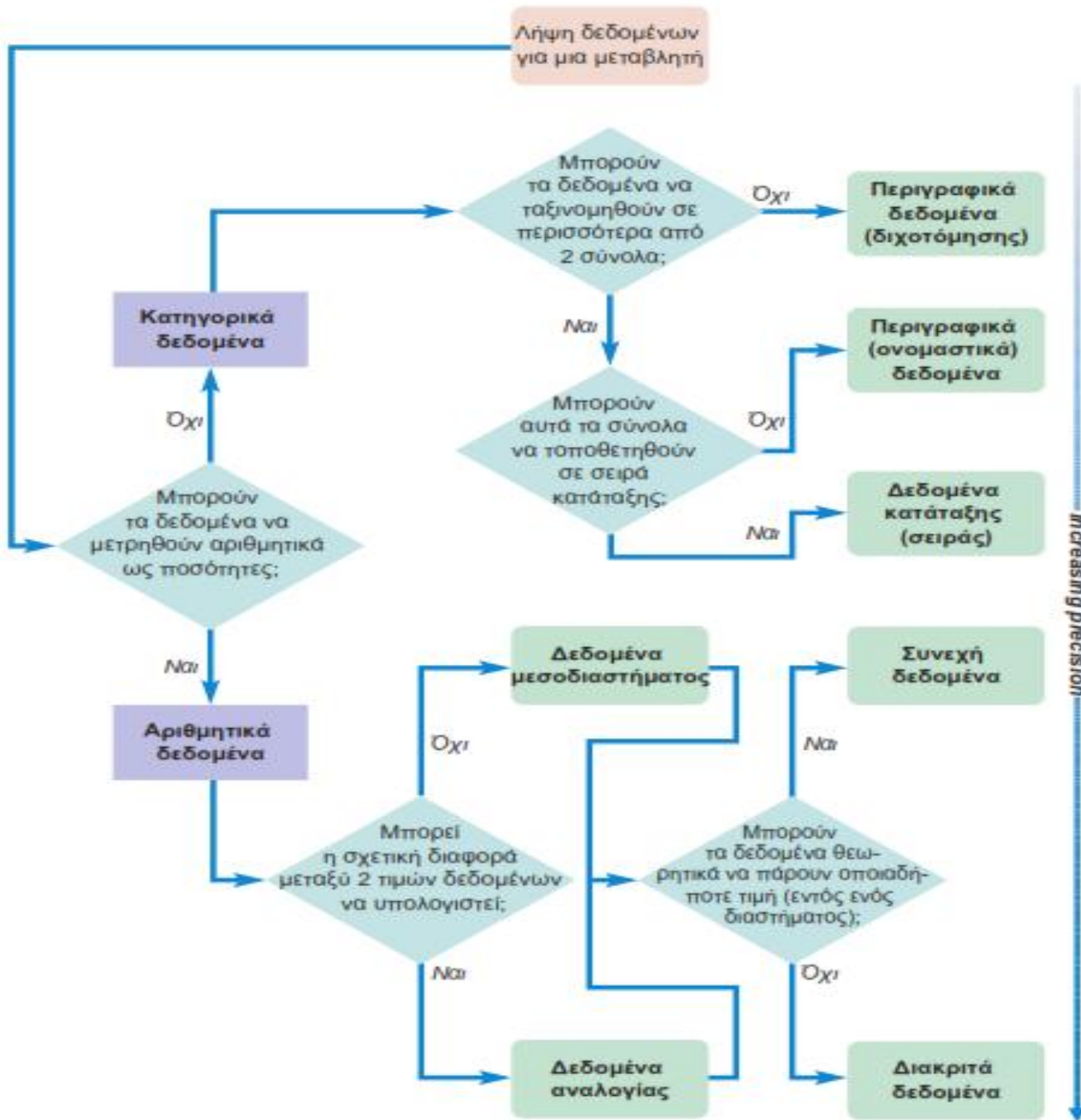
Ανάλυση ποσοτικών δεδομένων

Διδάσκων: Δρ. Α. Κουμπάρλης

Οι σημειώσεις που ακολουθούν προέρχονται από το βιβλίο **Μέθοδοι Έρευνας στις Επιχειρήσεις και την Οικονομία** των **M. Saundres, P. Lewis και A. Thornhill**, εκδόσεις ΔΙΣΙΓΜΑ

Μαθησιακοί στόχοι κεφαλαίου

- Να αναγνωρίζετε τα βασικά ζητήματα προετοιμασίας ποσοτικών δεδομένων για ανάλυση και να αναλύετε αυτά τα δεδομένα με τη βοήθεια λογισμικού,
- Να αναγνωρίζετε διαφορετικούς τύπους δεδομένων και να αντιλαμβάνεστε τις επιπτώσεις του τύπου των δεδομένων όσον αφορά την επιλογή μεθόδων ανάλυσης,
- Να δημιουργείτε μήτρες δεδομένων και να κωδικοποιείτε δεδομένα,
- Να επιλέγετε τους πιο κατάλληλους πίνακες και διαγράμματα για να παρουσιάζετε και να απεικονίζετε γραφικά διαφορετικές πτυχές των δεδομένων σας,
- Να επιλέγετε τις πιο κατάλληλες στατιστικές παραμέτρους για να περιγράφετε μεμονωμένες μεταβλητές και να εξετάζετε τις σχέσεις ανάμεσα σε μεταβλητές και τάσεις,
- Να ερμηνεύετε σωστά τους πίνακες, τα διαγράμματα και τις στατιστικές παραμέτρους που χρησιμοποιείτε.



Τύποι ποσοτικών δεδομένων

Εικόνα 12.1 Ορισμός του τύπου δεδομένων

Μήτρα δεδομένων

- κάθε στήλη αναπαριστά μια ξεχωριστή μεταβλητή
- κάθε γραμμή περιέχει τις τιμές των μεταβλητών για μια μεμονωμένη περίπτωση για την οποία έχετε αποκτήσει δεδομένα.

	A	B	C	D	E
1	κωδικός	ηλικία	φύλο	υπηρεσία	εργάζεται
2	1	27	1	2	1
3	2	19	2	1	2
4	3	24	2	3	1
5					

Κωδικοποίηση Δεδομένων

- **Όλοι οι τύποι δεδομένων καλό είναι να καταγράφονται κωδικοποιημένοι αριθμητικά.**
 - καταχωρίζετε τα δεδομένα γρήγορα και με λιγότερα λάθη
 - διευκολύνονται οι αναλύσεις ειδικά αν απαιτείται επανακωδικοποίηση
 - προσοχή στις ακατανόητες αναλύσεις (π.χ. ο υπολογισμός μέσου όρου για τη μεταβλητή «φύλο»)
 - εάν χρησιμοποιείτε υπολογιστικό φύλλο (Excel), θα χρειαστείτε μια χωριστή λίστα με κωδικούς για κάθε μεταβλητή. Τα ειδικά λογισμικά στατιστικής ανάλυσης (SPSS κλπ) συνδέουν αυτόματα κάθε κωδικό με μια προκαθορισμένη ετικέτα.
- **Επανακωδικοποίηση δεδομένων**
 - Ομαδοποίηση ή συνδυασμός δεδομένων σε λιγότερο λεπτομερείς κατηγορικές μεταβλητές.
 - Για παράδειγμα η ηλικία ενός ατόμου, π.χ. 28 ετών θα μπορούσε με επανακωδικοποίηση να τοποθετηθεί στην ηλικιακή κατηγορία, π.χ., από 25 ως 30 ετών (κατηγορικά δεδομένα κατάταξης).

Κωδικοποίηση κατηγορικών δεδομένων

- **Σχεδιάστε ένα σύστημα κωδικοποίησης το οποίο θα διευκολύνει την ανάλυση των δεδομένων.**
 - πολλές πηγές δευτερογενών δεδομένων έχουν ενσωματωμένο το σύστημα κωδικοποίησης (π.χ. ΣΤΑΚΟΔ κλπ)
 - προκαθορισμένοι κωδικοί που περιλαμβάνονται στη φόρμα συλλογής δεδομένων (ερωτηματολόγιο) ή
 - βιβλίο κωδικών εκ των υστέρων που είναι καλύτερο να το αναπτύξετε αφού θα έχουν συγκεντρωθεί δεδομένα από τις πρώτες 50 ως 100 περιπτώσεις.
- **Λάβετε υπόψη:**
 - το επίπεδο ακρίβειας που απαιτείται,
 - τα συστήματα κωδικοποίησης που χρησιμοποιήθηκαν σε έρευνες με τις οποίες θα γίνουν συγκρίσεις

Καταχώρηση δεδομένων

- «αν καταχωρίσεις σκουπίδια, σκουπίδια θα πάρεις»
 - προηγμένο λογισμικό ανάλυσης (π.χ. SPSS) σας επιτρέπει να συνδέετε κάθε μεταβλητή με ετικέτες και τους κωδικούς που σχετίζονται με καθεμία.
- Έλεγχος για σφάλματα
 - Αναζητήστε μη αποδεκτούς κωδικούς. Κοινά λάθη είναι η προσθήκη του γράμματος O (πεζό και κεφαλαίο) αντί για το μηδέν (0), τα λατινικά γράμματα L και I αντί του 1, ο αριθμός 7 αντί του 1 και τα λατινικά A, B, E, H, I, K, M, N, O, P, T, Y, X,Z αντί για τα αντίστοιχα ελληνικά και αντίστροφα.
 - Αναζητήστε μη λογικές σχέσεις. Για παράδειγμα, αν κάποιος κωδικοποιείται στα «ανώτερα διοικητικά επαγγέλματα» στην κατηγορία κοινωνικοοικονομικής κατάταξης και περιγράφει τη δουλειά του ως «χειρωνακτική», τότε μάλλον υπάρχει λάθος.
 - Ελέγξτε ότι τηρούνται οι κανόνες στις ερωτήσεις φίλτρου
 - Ο έλεγχος δεδομένων είναι πολύ χρονοβόρος και συχνά δε γίνεται. Αν όμως δεν γίνει τα αποτελέσματα κινδυνεύουν να είναι εσφαλμένα και τα συμπεράσματα εντελώς ακατάλληλα

Στάθμιση δεδομένων

- Εφαρμόζεται συχνά σε διαστρωματική τυχαία δειγματοληψία με διαφορετικό κλάσμα δειγματοληψίας ή διαφορετικό ποσοστό απόκρισης για κάθε στρώμα.
- Για να αποκτήσετε μια ακριβή συνολική εικόνα μπορείτε να εφαρμόσετε συντελεστές στάθμισης ως εξής:
 - Υπολογίζετε το κλάσμα δειγματοληψίας ή το ποσοστό απόκρισης κάθε στρώματος.
 - Εντοπίζετε το στρώμα με το υψηλότερο κλάσμα ή ποσοστό απόκρισης.
 - Υπολογίζετε το βάρος για κάθε στρώμα, χρησιμοποιώντας τον τύπο:
 - Εφαρμόζετε τον κατάλληλο συντελεστή στάθμισης σε κάθε περίπτωση

Ανάλυση και παρουσίαση δεδομένων (1)

- Διερευνητική ανάλυση δεδομένων (EDA)
 - χρησιμοποιεί διαγράμματα για την εξερεύνηση και κατανόηση των δεδομένων
 - δίνει την ευελιξία να χρησιμοποιείτε νέες τεχνικές ανάλυσης στην πορεία
 - ανοίγει το δρόμο για άλλες γόνιμες μεθόδους ανάλυσης
 - οι υπολογιστές διευκολύνουν και επιταχύνουν τις διεργασίες.



Πλαίσιο 12.8

Λίστα ελέγχου

Σχεδίαση διαγραμμάτων και πινάκων

Για διαγράμματα και πίνακες

- ✓ Έχει σύντομο, αλλά σαφή και περιγραφικό τίτλο;
- ✓ Οι μονάδες μέτρησης που χρησιμοποιούνται αναφέρονται σαφώς;
- ✓ Οι πηγές δεδομένων που χρησιμοποιούνται αναφέρονται σαφώς;
- ✓ Υπάρχουν σημειώσεις που εξηγούν τις συντομογραφίες και την τυχόν ασυνήθιστη ορολογία;
- ✓ Δηλώνεται το μέγεθος του δείγματος στο οποίο βασίζονται οι τιμές του πίνακα;

Για διαγράμματα

- ✓ Έχει ξεκάθαρες ετικέτες αξόνων;
- ✓ Οι ράβδοι και τα στοιχεία τους έχουν την ίδια λογική σειρά;
- ✓ Για τις μικρότερες περιοχές, χρησιμοποιείται πιο πυκνή σκίαση;
- ✓ Έχετε αποφύγει τη λάθος παρουσίαση ή την αλλοίωση των δεδομένων;
- ✓ Συμπεριλαμβάνεται ένα υπόμνημα (όπου είναι απαραίτητο);

Για πίνακες

- ✓ Έχει ξεκάθαρες επικεφαλίδες στηλών και γραμμών;
- ✓ Η σειρά των στηλών και των γραμμών είναι λογική;

Ανάλυση και παρουσίαση δεδομένων (2)

Πίνακας 12.2 Παρουσίαση δεδομένων κατά τύπο δεδομένων: μια περίληψη

	Κατηγορικά		Αριθμητικά	
	Περιγραφικά	Κατάταξης	Συνεχή	Διακριτά
Για παρουσίαση μίας μεταβλητής, ώστε οποιαδήποτε συγκεκριμένη τιμή να μπορεί να διαβαστεί εύκολα.	Πίνακας/κατανομή συχνοτήτων (τα δεδομένα συχνά ομαδοποιούνται)			
Για παρουσίαση της συχνότητας εμφάνισης κατηγοριών ή τιμών μίας μεταβλητής, ώστε να είναι σαφείς η υψηλότερη και η χαμηλότερη τιμή.	Ραβδόγραμμα ή εικονογράφημα (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν).		Ιστόγραμμα ή πολύγωνο συχνοτήτων (τα δεδομένα πρέπει να ομαδοποιηθούν).	Ραβδόγραμμα ή πικτογράφημα (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν).
Για παρουσίαση της τάσης μίας μεταβλητής.		Γραμμικό διάγραμμα ή ραβδόγραμμα	Γραμμικό διάγραμμα ή ιστόγραμμα	Γραμμικό διάγραμμα ή ραβδόγραμμα
Για παρουσίαση της αναλογίας εμφάνισης κατηγοριών ή τιμών μίας μεταβλητής.	Διάγραμμα πίτας ή ραβδόγραμμα (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν).		Ιστόγραμμα ή διάγραμμα πίτας (τα δεδομένα πρέπει να ομαδοποιηθούν).	Διάγραμμα πίτας ή ραβδόγραμμα (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν).
Για παρουσίαση της κατανομής τιμών μίας μεταβλητής.			Πολύγωνο συχνοτήτων, ιστόγραμμα (τα δεδομένα πρέπει να ομαδοποιηθούν) ή διάγραμμα πλαισίου απολήξεων.	Πολύγωνο συχνοτήτων, ραβδόγραμμα (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν) ή διάγραμμα πλαισίου απολήξεων.

Ανάλυση και παρουσίαση δεδομένων (3)

Για παρουσίαση της αλληλεξάρτησης ανάμεσα σε δύο ή περισσότερες μεταβλητές, ώστε να μπορεί να διαβαστεί εύκολα οποιαδήποτε συγκεκριμένη τιμή.	Πίνακας συνάφειας/πίνακας διασταύρωσης μεταβλητών (τα δεδομένα συχνά ομαδοποιούνται).	
Για σύγκριση της συχνότητας εμφάνισης κατηγοριών ή τιμών δύο ή περισσότερων μεταβλητών, ώστε να είναι σαφείς η υψηλότερη και η χαμηλότερη τιμή.	Διάγραμμα πολλαπλών ράβδων (τα συνεχή δεδομένα πρέπει να ομαδοποιούνται, άλλα δεδομένα ίσως να πρέπει να ομαδοποιηθούν).	
Για σύγκριση των τάσεων δύο ή περισσότερων μεταβλητών, ώστε οι συνδέσεις να είναι σαφείς.		Διάγραμμα πολλαπλών γραμμών ή διάγραμμα πολλαπλών ράβδων.
Για σύγκριση των αναλογιών εμφάνισης κατηγοριών ή τιμών δύο ή περισσότερων μεταβλητών.	Συγκριτικά διαγράμματα πίτας ή ραβδόγραμμα στοιχείων ποσοστών (τα συνεχή δεδομένα πρέπει να ομαδοποιούνται, τα άλλα δεδομένα ίσως να πρέπει να ομαδοποιηθούν).	
Για σύγκριση της κατανομής τιμών δύο ή περισσότερων μεταβλητών.		Πολλαπλό διάγραμμα πλαισίου απολήξεων
Για σύγκριση της συχνότητας εμφάνισης κατηγοριών ή τιμών δύο ή περισσότερων μεταβλητών, ώστε τα σύνολα να είναι σαφή.	Διάγραμμα ράβδων σε στοίβα (τα συνεχή δεδομένα πρέπει να ομαδοποιούνται, τα άλλα δεδομένα ίσως πρέπει να ομαδοποιηθούν).	
Για σύγκριση των αναλογιών και των συνόλων εμφανίσεων κατηγοριών ή τιμών δύο ή περισσότερων μεταβλητών.	Συγκριτικά αναλογικά διαγράμματα πίτας (τα συνεχή δεδομένα πρέπει να ομαδοποιούνται, τα άλλα δεδομένα ίσως πρέπει να ομαδοποιηθούν).	
Για παρουσίαση της σχέσης ανάμεσα σε περιπτώσεις δύο μεταβλητών.		Διάγραμμα διασποράς

Πίνακας 12.4 Περιγραφικά στατιστικά στοιχεία κατά τύπο δεδομένων: μια περίληψη

Για να μετρήσουμε:		Κατηγορικά		Αριθμητικά	
		Περιγραφικά	Κατάταξης	Συνεχή	Διακριτά
Κεντρική τάση που...	... αναπαριστά την τιμή που παρατηρείται πιο συχνά.	Επικρατούσα τιμή			
	... αναπαριστά την ενδιάμεση τιμή.			Διάμεσος	
	... υπολογίζεται από όλες τις τιμές δεδομένων (μέσος όρος).			Μέσος	
Διασπορά που...	... δηλώνει τη διαφορά ανάμεσα στην υψηλότερη και τη χαμηλότερη τιμή.			Διάστημα (τα δεδομένα δε χρειάζεται να κατανέμονται κανονικά, αλλά πρέπει να τοποθετηθούν σε σειρά κατάταξης).	
	... δηλώνει τη διαφορά μέσα στο μεσαίο 50% των τιμών.			Ενδοτεταρτημοριακό εύρος (τα δεδομένα δε χρειάζεται να κατανέμονται κανονικά, αλλά πρέπει να τοποθετηθούν σε σειρά κατάταξης).	
	... δηλώνει τη διαφορά μέσα σε κάποιο άλλο διάστημα τιμών.			Δεκατημόρια ή εκατοστημόρια (τα δεδομένα δε χρειάζεται να κατανέμονται κανονικά, αλλά πρέπει να τοποθετηθούν σε σειρά κατάταξης).	
	... περιγράφει το βαθμό στον οποίο οι τιμές δεδομένων διαφέρουν από το μέσο.			Διακύμανση, ή τυπική απόκλιση (τα δεδομένα δε χρειάζεται να κατανέμονται κανονικά).	
	... συγκρίνει το βαθμό στον οποίο οι τιμές δεδομένων διαφέρουν από το μέσο ανάμεσα σε μεταβλητές			Συντελεστής μεταβλητότητας (τα δεδομένα δε χρειάζεται να κατανέμονται κανονικά).	
	... επιτρέπει τη σύγκριση του σχετικού βαθμού διαφοράς των διαφορετικών τιμών δεδομένων			Αριθμοδείκτες	

**Περιγραφή
δεδομένων με
χρήση
στατιστικών
παραμέτρων**

Εξέταση σχέσεων, διαφορών και τάσεων με χρήση στατιστικών παραμέτρων (1)

- **μη παραμετρικοί έλεγχοι:**
 - χρησιμοποιούνται όταν τα δεδομένα δεν κατανέμονται κανονικά.
 - συνήθως χρησιμοποιούνται με κατηγορικά δεδομένα
- **παραμετρικοί έλεγχοι :**
 - θεωρούνται πιο ισχυροί επειδή χρησιμοποιούν αριθμητικά δεδομένα,
 - πρέπει να ικανοποιούνται οι παρακάτω προϋποθέσεις

Εξέταση σχέσεων, διαφορών και τάσεων με χρήση στατιστικών παραμέτρων (2)

1. οι περιπτώσεις δεδομένων που επιλέγονται για το δείγμα είναι ανεξάρτητες
2. τα δείγματα εξάγονται από κανονικά κατανεμημένους πληθυσμούς
3. οι πληθυσμοί από τους οποίους εξάγονται τα δείγματα έχουν ίσες διακυμάνσεις
4. χρησιμοποιούνται αριθμητικά δεδομένα.
5. το μέγεθος του δείγματος είναι αρκετά μεγάλο και ικανοποιεί τις απαιτήσεις των στατιστικών αναλύσεων που χρησιμοποιούνται

Εξέταση σχέσεων, διαφορών και τάσεων με χρήση στατιστικών παραμέτρων (3)

Πίνακας 12.5 Στατιστικοί έλεγχοι για την εξέταση σχέσεων, διαφορών και τάσεων κατά τύπο δεδομένων: μια περίληψη

	Κατηγορικά		Αριθμητικά	
	Περιγραφικά	Κατάταξης	Συνεχή	Διακριτά
Έλεγχος κανονικότητας κατανομής			Έλεγχος Kolmogorov-Smirnov, έλεγχος Shapiro-Wilk	
Έλεγχος συσχέτισης μεταβλητών	χ τετράγωνο (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν).		χ τετράγωνο αν η μεταβλητή ομαδοποιείται σε διακριτές κλάσεις.	
	V και Phi του Cramer (για διχοτομικές μεταβλητές).			
Έλεγχος για το αν δύο ομάδες (κατηγορίες) διαφέρουν.	Kolmogorov-Smirnov (τα δεδομένα ίσως πρέπει να ομαδοποιηθούν) ή έλεγχος Mann-Whitney U.		Έλεγχος t ανεξάρτητων δειγμάτων ή έλεγχος t σε ζεύγη (συνικά χρησιμοποιείται για τον έλεγχο των αλλαγών ως προς το χρόνο) ή έλεγχος Mann-Whitney U (αν τα δεδομένα παρουσιάζουν ασυμμετρία ή το δείγμα είναι μικρό).	
Έλεγχος για το αν τρεις ή περισσότερες ομάδες (κατηγορίες) διαφέρουν.			Ανάλυση διακύμανσης (ANOVA).	
Αξιολόγηση της δύναμης της σχέσης μεταξύ δύο μεταβλητών.	Συντελεστής Spearman (συντελεστής rho του Spearman) ή συντελεστής Kendall (συντελεστής tau του Kendall).		Συντελεστής γραμμικής συσχέτισης του Pearson (PMCC).	

Εξέταση σχέσεων, διαφορών και τάσεων με χρήση στατιστικών παραμέτρων (4)

Αξιολόγηση της δύναμης μιας σχέσης ανάμεσα σε μία εξαρτημένη και μία ανεξάρτητη μεταβλητή.		Συντελεστής προσδιορισμού (συντελεστής παλινδρόμησης).
Αξιολόγηση της δύναμης μιας σχέσης ανάμεσα σε μία εξαρτημένη και δύο ή περισσότερες ανεξάρτητες μεταβλητές.		Συντελεστής πολλαπλού προσδιορισμού (συντελεστής πολλαπλής παλινδρόμησης).
Πρόβλεψη της τιμής μιας εξαρτημένης μεταβλητής από μία ή περισσότερες ανεξάρτητες μεταβλητές.		Εξίσωση παλινδρόμησης (ανάλυση παλινδρόμησης).
Εξερεύνηση της σχετικής αλλαγής (τάσης) ως προς το χρόνο.		Αριθμοδείκτες
Σύγκριση σχετικών αλλαγών (τάσεων) ως προς το χρόνο.		Αριθμοδείκτες
Προσδιορισμός της τάσης ως προς το χρόνο σε μια σειρά δεδομένων.		Χρονολογική σειρά: κινητοί μέσοι ή εξίσωση παλινδρόμησης (ανάλυση παλινδρόμησης).

Έλεγχος κανονικότητας

- οι παραμετρικοί έλεγχοι θεωρούν ότι τα δείγματα προέρχονται από κανονικά κατανομημένους πληθυσμούς που σημαίνει ότι οι τιμές δεδομένων για καθεμία από τις ποσοτικές μεταβλητές θα πρέπει επίσης να είναι κανονικά κατανομημένες.
- **έλεγχος με χρήση γραφημάτων**
 - (ιστογράμματα, διαγράμματα πλαισίου απολήξεων, πολύγωνα συχνοτήτων κλπ)
 - για κανονικά κατανομημένα δεδομένα, οι τιμές του αριθμητικού μέσου, της διάμεσου και της επικρατούσας τιμής είναι ίδιες.
- **έλεγχος με χρήση στατιστικών ελέγχων**
 - έλεγχος Kolmogorov–Smirnov
 - έλεγχος Shapiro–Wilks
 - Με πολύ μεγάλα δείγματα, είναι εύκολο να πάρετε σημαντικές διαφορές ανάμεσα σε μια μεταβλητή δείγματος και σε μια συγκρίσιμη κανονική κατανομή ακόμα και όταν οι πραγματικές διαφορές είναι σχετικά μικρές. Γι' αυτό το λόγο, είναι χρήσιμο να χρησιμοποιήσετε επίσης ένα διάγραμμα προκειμένου η απόφασή σας να είναι εμπειριστατωμένη.

Έλεγχος σημαντικότητας σχέσεων και διαφορών

- ο έλεγχος για την πιθανότητα μια παρατηρούμενη σχέση ή διαφορά ανάμεσα σε μεταβλητές να είναι συμπτωματική και μόνο
- **μηδενική υπόθεση (H_0):** «δεν υπάρχει σημαντική σχέση ή διαφορά ανάμεσα στις μεταβλητές ...»
- **Εναλλακτική υπόθεση (H_1):** «υπάρχει σημαντική σχέση ή διαφορά ανάμεσα στις μεταβλητές ...»

- η στατιστική σημαντικότητα μιας σχέσης που υποδεικνύεται από ένα στατιστικό έλεγχο καθορίζεται εν μέρει από το μέγεθος του δείγματος
- είναι πολύ δύσκολο να προκύψει στατιστικά σημαντικός έλεγχος από ένα πολύ μικρό δείγμα.
- όσο αυξάνεται το μέγεθος του δείγματος, λιγότερο προφανείς σχέσεις και διαφορές διαπιστώνεται ότι είναι στατιστικά σημαντικές
- μικρά δείγματα μπορούν να μειώσουν στο ελάχιστο την ευαισθησία των στατιστικών ελέγχων, ενώ τα πολύ μεγάλα δείγματα μπορούν να κάνουν το αντίθετο

Σφάλματα τύπου I και τύπου II

- η εσφαλμένη απόρριψη μιας μηδενικής υπόθεσης και επομένως, η αποδοχή της εναλλακτικής υπόθεσης είναι σφάλμα τύπου I
- η εσφαλμένη αποδοχή μιας μηδενικής υπόθεσης και επομένως, η απόρριψη της εναλλακτικής υπόθεσης είναι σφάλμα τύπου II

		Πιθανότητα για	
		Σφάλμα τύπου I	Σφάλμα τύπου II
Επίπεδο σημαντικότητας στο	0.05	Αυξημένη	Μειωμένη
	0.01	Μειωμένη	Αυξημένη

Έλεγχος συσχέτισης δύο μεταβλητών (1)

- **έλεγχος χ τετράγωνο (χ^2)**

- δεδομένα σε πίνακα συνάφειας διπλής εισόδου
- υπολογίζει την πιθανότητα τα δεδομένα στον πίνακά σας, ή δεδομένα ακόμα πιο ακραία, να είναι συμπτωματικά και μόνο
- βασίζεται σε σύγκριση των τιμών που παρατηρούνται στον πίνακα μ' αυτές που θα αναμένονταν αν οι δύο μεταβλητές ήταν εντελώς ανεξάρτητες
- βασίζεται στις ακόλουθες προϋποθέσεις:
 - οι κατηγορίες που χρησιμοποιούνται στον πίνακα συνάφειας είναι αμοιβαία αποκλειόμενες, ώστε κάθε παρατήρηση να ανήκει σε μία μόνο κατηγορία ή κλάση,
 - αναμενόμενες τιμές μικρότερες του 5 να μην υπάρχουν για περισσότερο του 25% των κελιών του πίνακα. Για πίνακες συνάφειας δύο γραμμών και δύο στηλών, καλό είναι να μην υπάρχουν αναμενόμενες τιμές μικρότερες του 10

Έλεγχος συσχέτισης δύο μεταβλητών (2)

- **έλεγχος V του Cramer**

- δεδομένα σε πίνακα συνάφειας διπλής εισόδου
- Ενώ η χ^2 υπολογίζει την πιθανότητα τα δεδομένα στον πίνακά σας, ή δεδομένα ακόμα πιο ακραία, να είναι συμπτωματικά και μόνο η V του Cramer μετρά τη συσχέτιση μεταξύ των μεταβλητών του πίνακα στο κλειστό διάστημα $[0,1]$
- 0 σημαίνει απουσία συσχέτισης και το 1 απόλυτη συσχέτιση.

Έλεγχος διαφορετικότητας δύο ομάδων (3)

- **Δεδομένα κατάταξης**

- ο έλεγχος **Kolmogorov–Smirnov** δίνει τη δυνατότητα να διαπιστωθεί αν η κατανομή των τιμών ενός δείγματος διαφέρει από μια καθορισμένη κατανομή, εκτός της κανονικής
 - ελέγχεται η πιθανότητα η κατανομή των υπό παρατήρηση δεδομένων να διαφέρει από εκείνη του καθορισμένου πληθυσμού μόνο συμπτωματικά
 - πιθανότητα 0,05 ή μικρότερη σημαίνει ότι μπορείτε να είστε τουλάχιστον 95% βέβαιοι ότι η διαφορά ανάμεσα στις δύο κατανομές δεν προέρχεται μόνο από σύμπτωση

- **Αριθμητικά δεδομένα (2 διακριτές ομάδες)**

- έλεγχος t ανεξαρτήτων δειγμάτων
- έλεγχος t συζευγμένων δειγμάτων

- **Αριθμητικά δεδομένα (3 ή περισσότερες ομάδες)**

- μονόδρομη ανάλυση διακύμανσης ή μονόδρομη ANOVA
- έλεγχος t συζευγμένων δειγμάτων

Έλεγχος διαφορετικότητας δύο ομάδων (4)

- **Αριθμητικά δεδομένα (3 ή περισσότερες ομάδες)**
 - μονόδρομη ανάλυση διακύμανσης ή μονόδρομη ANOVA
 - **προϋποθέσεις χρήσης**
 - οι τιμές των δεδομένων είναι ανεξάρτητες και δε σχετίζονται μεταξύ τους
 - τα δεδομένα για κάθε ομάδα είναι κανονικά κατανομημένα (η προϋπόθεση δεν είναι ιδιαίτερα σημαντική, εφόσον ο αριθμός των περιπτώσεων σε κάθε ομάδα είναι > 30).
 - οι διακυμάνσεις των ομάδων δεδομένων δε διαφέρουν μεταξύ τους, στατιστικά (όταν ο αριθμός των περιπτώσεων στη μεγαλύτερη ομάδα δεν υπερβαίνει περισσότερο του 50% τον αριθμό των περιπτώσεων στη μικρότερη ομάδα, η επίδραση στα αποτελέσματα του ελέγχου, από την παραβίαση της συγκεκριμένης προϋπόθεσης, είναι αμελητέα).

Αποτίμηση της έντασης της σχέσης

- αλλαγή σε μία μεταβλητή συνοδεύεται από μια αλλαγή σε άλλη μεταβλητή, αλλά δεν είναι σαφές ποια μεταβλητή προκάλεσε την αλλαγή στην άλλη, **συσχέτιση μεταβλητών,**
- αλλαγή σε μία ή περισσότερες (ανεξάρτητες) μεταβλητές προκαλεί αλλαγή σε μια άλλη (εξαρτημένη) μεταβλητή, **σχέση αιτίας και αποτελέσματος.**

Αξιολόγηση της έντασης της σχέσης μεταξύ δύο μεταβλητών (1)

- **συντελεστής συσχέτισης:**
 - δίνει τη δυνατότητα να ποσοτικοποίησης της έντασης της γραμμικής σχέσης ανάμεσα σε δύο μεταβλητές κατάταξης ή αριθμητικές.
 - συμβολίζεται με το γράμμα r
 - μπορεί να πάρει οποιαδήποτε τιμή ανάμεσα στο +1 που αναπαριστά μια πλήρως θετική συσχέτιση και -1 που αναπαριστά μια πλήρως αρνητική συσχέτιση
 - συντελεστές συσχέτισης ανάμεσα στο +1 και το -1 αναπαριστούν ασθενέστερες θετικές και αρνητικές συσχετίσεις, ενώ η τιμή 0 σημαίνει ότι οι μεταβλητές είναι απόλυτα ανεξάρτητες
 - στην έρευνα στους τομείς των επιχειρήσεων και της διοίκησης, οι τέλει ή μηδενικές συσχετίσεις είναι εξαιρετικά ασυνήθιστες.

Αξιολόγηση της έντασης της σχέσης μεταξύ δύο μεταβλητών (2)

- **συντελεστής γραμμικής συσχέτισης του Pearson (PMCC)**
 - και οι δύο μεταβλητές είναι αριθμητικές
 - τα δείγματα είναι τυχαία
- **αν η μία ή και οι δύο μεταβλητές περιέχουν δεδομένα κατάταξης**
 - συντελεστής συσχέτισης κατάταξης του Spearman (ρ του Spearman)
 - συντελεστής συσχέτισης κατάταξης του Kendall (τ του Kendall)
 - ο συντελεστής του Kendall θεωρείται πιο κατάλληλος αν τα δεδομένα μιας εκ των μεταβλητών περιέχουν συνδεμένες βαθμίδες (tied ranks).



Εικόνα 12.15 Τιμές συντελεστή συσχέτισης

Πηγές Αναπτύχθηκε από προηγούμενες εκδόσεις, Haig και άλλοι (2006)

Αξιολόγηση της έντασης σχέσης αιτιώδους συνάφειας

- **συντελεστής προσδιορισμού ή συντελεστής παλινδρόμησης (r^2)**
 - αξιολογεί την ένταση της σχέσης ανάμεσα σε μια αριθμητική εξαρτημένη μεταβλητή και μία ή περισσότερες αριθμητικές ανεξάρτητες μεταβλητές
 - τα δεδομένα πρέπει να έχουν επιλεγεί τυχαία
 - μπορεί να πάρει οποιαδήποτε τιμή ανάμεσα στο 0 και το +1
 - μετρά το ποσοστό διακύμανσης μιας εξαρτημένης μεταβλητής που μπορεί να εξηγηθεί στατιστικά από τη διακύμανση μίας ανεξάρτητης μεταβλητής ή πολλών ανεξάρτητων μεταβλητών
 - αν η μεταβολή μπορεί να εξηγηθεί, στο σύνολό της, ο συντελεστής προσδιορισμού θα είναι 1.
 - αν κανένα ποσοστό της μεταβολής δεν μπορεί να εξηγηθεί, ο συντελεστής θα είναι 0
 - στις έρευνές μας, σπάνια παίρνουμε τιμές συντελεστή πάνω από 0,8
 - η διαδικασία υπολογισμού του συντελεστή προσδιορισμού και της εξίσωσης παλινδρόμησης με μία ανεξάρτητη μεταβλητή ορίζεται ως ανάλυση παλινδρόμησης.
 - ο υπολογισμός συντελεστή πολλαπλού προσδιορισμού (ή συντελεστής πολλαπλής παλινδρόμησης R^2) και η εξίσωση παλινδρόμησης με δύο ή περισσότερες ανεξάρτητες μεταβλητές ορίζεται ως ανάλυση πολλαπλής παλινδρόμησης.

Πρόβλεψη της τιμής μιας μεταβλητής από μία ή περισσότερες άλλες μεταβλητές

- Ανάλυση παλινδρόμησης (μέσω της εξίσωσης παλινδρόμησης)

$$AoS_i = a + b_1ME_i + b_2NSS_i$$

όπου:

AoS_t είναι ο όγκος των πωλήσεων στο χρόνο t

ME_t είναι οι δαπάνες για το μάρκετινγκ στο χρόνο t

NSS_t είναι ο αριθμός των υπαλλήλων του τμήματος πωλήσεων στο χρόνο t

a είναι η σταθερά παλινδρόμησης

b_1 και b_2 είναι οι συντελεστές κλίσης βήτα

Αυτή η εξίσωση μπορεί να μεταφραστεί ως εξής:

$$\text{Όγκος πωλήσεων στο χρόνο t} = \text{τιμή} + (b_1 * \text{Δαπάνες μάρκετινγκ στο χρόνο t}) + (b_2 * \text{Αριθμός υπαλλήλων τμήματος πωλήσεων στο χρόνο t})$$

Προϋποθέσεις εφαρμογής ανάλυσης παλινδρόμησης

- η σχέση ανάμεσα στις εξαρτημένες και ανεξάρτητες μεταβλητές είναι γραμμική.
- οι τιμές δεδομένων των εξαρτημένων και ανεξάρτητων μεταβλητών έχουν ίσες διακυμάνσεις – ομοσκεδαστικότητα (έλεγχος Levene για την ομογένεια της διακύμανσης)
- εάν υπάρχει ετεροσκεδαστικότητα (δηλαδή άνισες διακυμάνσεις), πιθανόν να μπορείτε να εκτελέσετε την ανάλυσή σας κάτω από προϋποθέσεις
- απουσία συσχέτισης ανάμεσα σε δύο ή περισσότερες ανεξάρτητες μεταβλητές (συγγραμμικότητα ή πολυσυγγραμμικότητα), καθώς έτσι δυσχεραίνεται ο προσδιορισμός των επιμέρους επιδράσεων των μεμονωμένων μεταβλητών (παράγοντας διόγκωσης διακύμανσης - VIF)
- τα δεδομένα για τις ανεξάρτητες μεταβλητές και την εξαρτημένη μεταβλητή κατανέμονται κανονικά

Εξέταση τάσεων

- σχεδιάστε ένα διάγραμμα γραμμής, ώστε να αποκτήσετε μια οπτική αναπαράσταση της τάσης
- τρεις από τις πιο συνηθισμένες χρήσεις αναλύσεις είναι:
 - εξερεύνηση της τάσης ή της σχετικής αλλαγής για μία μόνο μεταβλητή ως προς το χρόνο,
 - σύγκριση των τάσεων ή των σχετικών αλλαγών μεταβλητών που μετριοούνται σε διαφορετικές μονάδες ή έχουν διαφορετικά μεγέθη,
 - προσδιορισμός της μακροπρόθεσμης τάσης και η πρόβλεψη μελλοντικών τιμών μιας μεταβλητής.