

**ΣΗΜΕΙΩΣΕΙΣ ΓΙΑ ΤΟ ΜΑΘΗΜΑ
ΣΤΑΤΙΣΤΙΚΗ ΓΙΑ ΠΟΛΙΤΙΚΟΥΣ ΜΗΧΑΝΙΚΟΥΣ**

ΜΕΡΟΣ Β

Δημήτρης Κουγιουμτζής

<http://users.auth.gr/dkugiu/Teach/CivilEngineer>

E-mail: dkugiu@gen.auth.gr

1/11/2009

Περιεχόμενα

1	ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ	7
1.1	Περιγραφή Στατιστικών Δεδομένων	7
1.2	Περιγραφικά Μέτρα Στατιστικών Δεδομένων	13
1.2.1	Μέτρα θέσης	13
1.2.2	Μέτρα μεταβλητότητας	14
2	ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ	23
2.1	Σημειακή Εκτίμηση	23
2.1.1	Κριτήρια καλών εκτιμητριών	24
2.1.2	Μέθοδος υπολογισμού της σημειακής εκτίμησης	26
2.2	Εκτίμηση Διαστήματος Εμπιστοσύνης	29
2.2.1	Διάστημα εμπιστοσύνης της μέσης τιμής μ	29
2.2.2	Διάστημα εμπιστοσύνης της διασποράς σ^2	36
2.2.3	Διάστημα εμπιστοσύνης της αναλογίας p	38
2.2.4	Διάστημα εμπιστοσύνης της διαφοράς δύο μέσων τιμών $\mu_1 - \mu_2$	40
2.2.5	Διάστημα της διαφοράς δύο αναλογιών $p_1 - p_2$	44
3	ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ	47
3.1	Διαδικασία Ελέγχου Στατιστικής Υπόθεσης	47
3.1.1	Στατιστική υπόθεση	47
3.1.2	Στατιστική ελέγχου και περιοχή απόρριψης	48
3.1.3	Απόφαση ελέγχου	49
3.2	Έλεγχος Παραμέτρων Πληθυσμού	50
3.2.1	Έλεγχος για τη μέση τιμή μ	50
3.2.2	Έλεγχος για τη διασπορά σ^2	52
3.2.3	Έλεγχος για την αναλογία p	54
3.2.4	Έλεγχος για τη διαφορά μέσων τιμών $\mu_1 - \mu_2$	55
3.2.5	Έλεγχος για τη διαφορά δύο αναλογιών $p_1 - p_2$	58
3.3	Χαρακτηριστικά Ελέγχου	59
3.3.1	Αποφάσεις ελέγχου και πραγματική κατάσταση	60
3.3.2	Δύναμη ελέγχου	61
3.3.3	Δίπλευρος έλεγχος και διάστημα εμπιστοσύνης	64
4	ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ	67
4.1	Συσχέτιση δύο τ.μ.	67
4.1.1	Ο συντελεστής συσχέτισης ρ	68

4.1.2	Σημειακή εκτίμηση του συντελεστή συσχέτισης	68
4.2	Απλή Γραμμική Παλινδρόμηση	71
4.2.1	Το πρόβλημα της γραμμικής παλινδρόμησης	72
4.2.2	Σημειακή εκτίμηση των παραμέτρων της γραμμικής παλινδρόμησης . . .	74
4.2.3	Σχέση του συντελεστή συσχέτισης και παλινδρόμησης	77

Οι στατιστικές έννοιες και μέθοδοι είναι εργαλεία όχι απλά χρήσιμα αλλά κι απαραίτητα για να καταλάβουμε τον κόσμο γύρω μας. Αν για ένα φαινόμενο που μελετάμε δεν υπάρχει καθόλου αβεβαιότητα ή τυχαιότητα δε χρειάζεται η στατιστική προσέγγιση αφού **καθοριστικά μοντέλα** (deterministic models) μπορούν να περιγράψουν το φαινόμενο επακριβώς. Στα πραγματικά φαινόμενα όμως σχεδόν πάντα υπάρχει ο παράγοντας της αβεβαιότητας ή τυχαιότητας κι αυτό κάνει σημαντική τη γνώση των **στατιστικών μεθόδων και μοντέλων** για την επίλυση πραγματικών προβλημάτων. Για παράδειγμα αν θέλουμε να μελετήσουμε την απόσταση από το σπίτι του φοιτητή στο Πανεπιστήμιο, τότε δε μπορούμε να μιλήσουμε με βεβαιότητα γι αυτό το μέγεθος όσους φοιτητές κι αν συμπεριλάβουμε για να συλλέξουμε στοιχεία για την απόσταση. Μπορούμε όμως να υπολογίσουμε τη μέση απόσταση από ένα δείγμα φοιτητών. Ένα άλλο παράδειγμα είναι η σχέση της κατανάλωσης ηλεκτρικής ενέργειας ή νερού με τη διάσταση του χώρου ενός κτιρίου. Η εκτίμηση μιας τέτοιας σχέσης μπορεί να γίνει μόνο με στατιστικές μεθόδους αφού οι μετρήσεις ηλεκτρικής ενέργειας και διάστασης χώρου για κάθε κτίριο επηρεάζονται από πολλούς άλλους παράγοντες κι εμπεριέχουν έτσι τυχαιότητα. Συγκεκριμένα για κλάδους της μηχανικής όπως οι πολιτικοί μηχανικοί, η στατιστική μπορεί να βοηθήσει στην κατανόηση κάποιων φαινομένων και διαδικασιών σχετικών με τη μηχανική που περιέχουν αβεβαιότητα ή μεταβλητότητα, όπως η μελέτη της αντοχής θραύσης σκυροδέματος, ή η επιμήκυνση αλουμινίου για διάφορες τάσεις.

Για να μάθουμε για ένα φαινόμενο συνήθως πρώτα συλλέγουμε δεδομένα. **Στατιστική** (statistics) είναι η επιστήμη ή 'τέχνη' του να μαθαίνουμε από τα δεδομένα. Η στατιστική συνίσταται στη συλλογή δεδομένων, στην περιγραφή τους και κυρίως στην ανάλυση τους που οδηγεί και στην απόκτηση συμπερασμάτων. Η συλλογή των δεδομένων αναφέρεται κι ως **δειγματοληψία** (sampling) κι αποτελεί ξεχωριστό πεδίο της στατιστικής που δεν θα μας απασχολήσει εδώ. Η περιγραφή των δεδομένων και η παρουσίαση συνοπτικών στοιχείων και γραφημάτων αναφέρεται ως **περιγραφική στατιστική** (descriptive statistics). Η ανάλυση στατιστικών δεδομένων και η ερμηνεία των αποτελεσμάτων αναφέρεται ως **στατιστική συμπερασματολογία** (statistical inference) και αποτελεί την βάση της στατιστικής και γι αυτό συχνά αναφέρεται απλά και ως **στατιστική**.

Παραθέτουμε κάποιους βασικούς ορισμούς που χρησιμοποιούνται στην στατιστική ανάλυση :

τυχαία μεταβλητή (τ.μ.) (random variable): οποιοδήποτε χαρακτηριστικό του οποίου η τιμή αλλάζει στα διάφορα στοιχεία του πληθυσμού, π.χ. η αντοχή θραύσης σκυροδέματος.

δεδομένα (data): ένα σύνολο τιμών μιας τ.μ. που έχουμε στη διάθεση μας, π.χ. μετρήσεις αντοχής θραύσης σκυροδέματος.

πληθυσμός (population): μια ομάδα ή μια κατηγορία στην οποία αναφέρεται η τ.μ., π.χ. σκυρόδεμα γενικά, ή ένας τύπος σκυροδέματος.

δείγμα (sample): ένα υποσύνολο του πληθυσμού που μελετάμε, π.χ. 25 δοκίμια σκυροδέματος που χρησιμοποιήσαμε για να κάνουμε μετρήσεις αντοχής θραύσης.

παράμετρος (parameter): ένα μέγεθος που συνοψίζει με κάποιο τρόπο τις τιμές της τ.μ. στον πληθυσμό, π.χ. η μέση τιμή αντοχής θραύσης σκυροδέματος.

στατιστική (statistic): ένα μέγεθος που συνοψίζει με κάποιο τρόπο τις τιμές της τ.μ. στο δείγμα, π.χ. ο μέσος όρος αντοχής θραύσης από τα 25 δοκίμια σκυροδέματος που μετρήσαμε.

Στο Μέρος Α ασχοληθήκαμε με την τυχαία μεταβλητή (τ.μ.) X , μελετήσαμε τη συνάρτηση πυκνότητας ή μάζας πιθανότητας $f_X(x)$ και τη συνάρτηση κατανομής $F_X(x)$ καθώς και κύριες παραμέτρους της κατανομής, όπως τη μέση (ή προσδοκώμενη) τιμή $E(X) \equiv \mu$ και τη διασπορά (ή διακύμανση) $\text{Var}(X) \equiv \sigma^2$.

Η μελέτη μιας τ.μ. με τη βοήθεια της πιθανοθεωρίας προϋποθέτει ότι γνωρίζουμε (ή υποθέτουμε) την κατανομή της τ.μ., κι άρα και τον πληθυσμό στον οποίο παίρνει τιμές, καθώς και τις παραμέτρους της. Στην πράξη, οι παράμετροι πληθυσμού είναι *σταθερές* αλλά *άγνωστες* ενώ οι στατιστικές δείγματος είναι *γνωστές* αλλά *μεταβλητές*. Θα ασχοληθούμε μόνο με τις πιο γνωστές από την πιθανοθεωρία παραμέτρους κατανομής μιας τ.μ., δηλαδή τη μέση τιμή μ , τη διασπορά σ^2 και την αναλογία p .

Γενικά δε γνωρίζουμε τον πληθυσμό αλλά συλλέγουμε δεδομένα για την τ.μ. που μας ενδιαφέρει από ένα αντιπροσωπευτικό δείγμα του πληθυσμού. Πρώτα παρουσιάζουμε τα δεδομένα μέσα από πίνακες και διαγράμματα και συνοψίζουμε τα δεδομένα υπολογίζοντας συνοπτικά μέτρα (αυτό είναι το θέμα της *περιγραφικής στατιστικής*). Στη συνέχεια προσπαθούμε να βγάλουμε συμπεράσματα για την τ.μ. και τις παραμέτρους του πληθυσμού που μελετάμε από τα δεδομένα του δείγματος και τις στατιστικές που υπολογίζουμε από αυτά τα δεδομένα (αυτό είναι το θέμα της *στατιστικής συμπερασματολογίας*).

Για παράδειγμα, για τη μελέτη της αντοχής θραύσης κάποιου τύπου σκυροδέματος και συγκεκριμένα για τη μέση τιμή της δεν είναι δυνατόν να γνωρίζουμε την αντοχή θραύσης για κάθε δοκίμιο που παράγουμε αλλά μπορούμε να μετρήσουμε την αντοχή θραύσης σε κάποια δοκίμια που επιλέγουμε τυχαία. Πρώτα παρουσιάζουμε τις μετρήσεις αντοχής θραύσης που συλλέξαμε χρησιμοποιώντας τις τεχνικές της περιγραφικής στατιστικής για να σχηματίσουμε μια πρώτη εντύπωση για την κατανομή της αντοχής θραύσης. Μετά χρησιμοποιώντας μεθόδους της στατιστικής συμπερασματολογίας προχωράμε στην εκτίμηση της μέσης τιμής της και στον έλεγχο κάποιας υπόθεσης σχετικά με τιμές που μπορεί να πάρει η μέση τιμή.

Ένα άλλο παράδειγμα είναι η μελέτη της μεταβολής της επιμήκυνσης αλουμινίου με την τάση. Το δείγμα περιλαμβάνει τις μετρήσεις της επιμήκυνσης δοκιμίων αλουμινίου σε διάφορες τάσεις. Με την περιγραφική στατιστική μπορούμε να απεικονίσουμε και να περιγράψουμε την εξάρτηση της επιμήκυνσης του αλουμινίου από την τάση και με μεθόδους της στατιστικής (συμπερασματολογίας) μπορούμε να εκτιμήσουμε ποσοτικά τη σχέση αυτής της εξάρτησης.

Στο Μέρος Β θα ασχοληθούμε με κάποια κύρια θέματα στατιστικής που αναφέραμε παραπάνω. Στο Κεφάλαιο 1 θα παρουσιάσουμε συνοπτικά κύριες τεχνικές της περιγραφικής στατιστικής. Στο Κεφάλαιο 2 θα μελετήσουμε την εκτίμηση κυρίων παραμέτρων κατανομής μιας τ.μ. ενός πληθυσμού και στο Κεφάλαιο 3 θα περιγράψουμε στατιστικούς ελέγχους για συγκεκριμένες τιμές αυτών των παραμέτρων. Στο Κεφάλαιο 4 θα μελετήσουμε τη συσχέτιση δύο τ.μ. καθώς και την εξάρτηση (παλινδρόμηση) μιας τ.μ. από μια άλλη.