

**ΕΡΓΑΣΤΗΡΙΑΚΕΣ ΣΗΜΕΙΩΣΕΙΣ ΓΙΑ ΤΟ ΜΑΘΗΜΑ
ΣΤΑΤΙΣΤΙΚΗ ΓΙΑ ΧΗΜΙΚΟΥΣ ΜΗΧΑΝΙΚΟΥΣ**

ΜΕΡΟΣ Β

Δημήτρης Κουγιουμτζής

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

E-mail: dkugiu@gen.auth.gr

5 Μαΐου 2011

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 δε μπορούμε να διατιμήσουμε με βεβαιότητα ένα μίγμα βενζίνης ως προς τον αριθμό οκτανίων αλλά μπορούμε να εκτιμήσουμε το μέσο αριθμό οκτανίων από ένα δείγμα μιγμάτων βενζίνης ίδιου τύπου. Ένα άλλο παράδειγμα είναι η σχέση της εκπομπής νιτρικού οξέος της μηχανής οχήματος με την παλαιότητα της μηχανής. Η μέτρηση της εκπομπής νιτρικού οξέος επηρεάζεται από πολλούς άλλους παράγοντες εκτός της παλαιότητας της μηχανής που είναι δύσκολο να ερευνηθούν και γι αυτό θεωρούμε ότι οι μετρήσεις αυτές εμπεριέχουν τυχαιότητα και χρησιμοποιούμε στατιστικές μεθόδους για να εκτιμήσουμε τη σχέση εκπομπής ρύπων και παλαιότητας. Άλλα ενδεικτικά παραδείγματα είναι η μελέτη της περιεκτικότητας μιας χημικής ουσίας και ο χρόνος αντίδρασης για διαφορετικές θερμοκρασίες σε μια χημική διαδικασία. Γενικά για κλάδους της μηχανικής όπως οι χημικοί μηχανικοί, η στατιστική μπορεί να βοηθήσει στην κατανόηση κάποιων φαινομένων και διαδικασιών σχετικών με τη μηχανική που περιέχουν αβεβαιότητα ή μεταβλητότητα.

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. ο αριθμός οκτανίων κάποιου τύπου μίγματος βενζίνης.

δεδομένα (data): ένα σύνολο τιμών μιας τ.μ. που έχουμε στη διάθεση μας, π.χ. μετρήσεις του αριθμού οκτανίων ενός μίγματος βενζίνης.

πληθυσμός (population): μια ομάδα ή μια κατηγορία στην οποία αναφέρεται η τ.μ., π.χ. ένας συγκεκριμένος τύπος μίγματος βενζίνης.

δείγμα (sample): ένα υποσύνολο του πληθυσμού που μελετάμε, π.χ. 25 μίγματα βενζίνης ενός συγκεκριμένου τύπου.

παράμετρος (parameter): ένα μέγεθος που συνοψίζει με κάποιο τρόπο τις τιμές της τ.μ. στον πληθυσμό, π.χ. ο μέσος αριθμός οκτανίων για ένα τύπο μίγματος βενζίνης.

στατιστική (statistic): ένα μέγεθος που συνοψίζει με κάποιο τρόπο τις τιμές της τ.μ. στο δείγμα, π.χ. ο μέσος όρος των αριθμών οκτανίων από 25 μίγματα βενζίνης κάποιου τύπου.

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 του αριθμού οκτανίων κάποιου τύπου βενζίνης που παράγει ένα διύλιστήριο πετρελαίου και συγκεκριμένα για τον υπολογισμό της μέση τιμής του αριθμού οκτανίων δεν είναι δυνατόν να καταγράψουμε τον αριθμό οκτανίων της βενζίνης σε κάθε βυτίο αλλά μπορούμε να μετρήσουμε τον αριθμό οκτανίων σε κάποια βυτία που επιλέγουμε τυχαία. Πρώτα παρουσιάζουμε τις μετρήσεις του αριθμού οκτανίων που συλλέξαμε χρησιμοποιώντας τις τεχνικές της περιγραφικής στατιστικής για να σχηματίσουμε μια πρώτη εντύπωση για την κατανομή του αριθμού οκτανίων του μίγματος βενζίνης. Μετά χρησιμοποιώντας μεθόδους της στατιστικής συμπερασματολογίας προχωράμε στην εκτίμηση του μέσου αριθμού οκτανίων και στον έλεγχο κάποιας υπόθεσης για τη μέση τιμή, π.χ. αν η μέση τιμή του αριθμού οκτανίων μπορεί να πέφτει κάτω από κάποιο όριο.

Ένα άλλο παράδειγμα είναι η μελέτη της μεταβολής του χρόνου αντίδρασης με τη θερμοκρασία του θαλάμου όπου γίνεται μια χημική διαδικασία. Σε αυτήν την περίπτωση το δείγμα περιλαμβάνει τις μετρήσεις του χρόνου αντίδρασης για διάφορες θερμοκρασίες. Με την περιγραφική στατιστική μπορούμε να απεικονίσουμε και να περιγράψουμε την εξάρτηση του χρόνου αντίδρασης από τη θερμοκρασία 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.