ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Εισαγωγή στην Ανάλυση Γλωσσικών Δεδομένων

Ενότητα 2: Περιγραφική Στατιστική

Γεώργιος Κ. Μικρός
Φιλοσοφική Σχολή
Τμήμα Ιταλικής Γλώσσας και Φιλολογίας

Είδη στατιστικών δεδομένων

- Αριθμητικά δεδομένα
  - Μετρήσεις που οι τιμές τους είναι αριθμοί, π.χ. «ορθογραφική ικανότητα», «λεξιλογική πυκνότητα», «διάρκεια άρθρωσης ενός φωνήεντος» κ.ά.
- Ποιοτικά δεδομένα
  - Μετρήσεις που οι τιμές τους είναι ποιοτικές διαβαθμίσεις, π.χ. τιμές όπως «θετική, αρνητική», «μηδαμινή, μέτρια, επαρκή» ως απαντήσεις σε στάση γύρω από τη διδασκαλία των αρχαίων στο γυμνάσιο.
- Κατηγορικά δεδομένα
  - Μετρήσεις που οι τιμές τους είναι ονοματικοί χαρακτηρισμοί που διαφέρουν σε είδος π.χ. τιμές όπως «άνδρας, γυναίκα», «Κείμενα δημοσιογραφικού λόγου, λογοτεχνίας, επιστήμης».

Δείκτες κεντρικής τάσης

- Μέσος όρος (mean)
- Δεσπόζουσα τιμή (mode)
  - Είναι η τιμή που εμφανίζεται συχνότερα σε κάποια δεδομένα
  - Π.χ. 8 12 15 17 **19 19 19** 27 56 56 89
- Διάμεσος (median)
  - Είναι η τιμή που βρίσκεται στο μέσο της κατανομής των δεδομένων. Οι μισές τιμές της κατανομής βρίσκονται κάτω από τη διάμεσο και οι άλλες μισές πάνω από αυτήν
  - Π.χ. 17 23 45 **46** 83 84 96

Δείκτες Διασποράς

- Εύρος (range)
- Διακύμανση (variance)
  - Είναι ο μέσος όρος των τετράγωνων των αποκλίσεων των τιμών μιας κατανομής από τον μέσο όρο της.
- Τυπική απόκλιση (standard deviation)
  - Είναι η τετραγωνική ρίζα των τιμών της διακύμανσης

X	**X-MO**	**$(X-MO)^2$**	
5	-6,4	40,96	
8	-3,4	11,56	
13	1,6	2,56	
14	2,6	6,76	
17	5,6	31,36	
MO		**s^2**	**s**
11,4		**18,64**	**4,32**

Στατιστικός Πληθυσμός

- Οι εμφανίσεις των γλωσσικών φαινομένων γίνονται σε ένα περιβάλλον που στη στατιστική το αποκαλούμε πληθυσμό (population). Η έννοια του πληθυσμού στην ΠΓ μπορεί να συλληφθεί δισυπόστατα:
  - Όλα τα κείμενα (προφορικά ή γραπτά) που έχουν παραχθεί κατά τη διάρκεια ανάπτυξης της γλώσσας
  - Το σύνολο της γλώσσας ως δομή (π.χ. όλες οι λέξεις στο λεξικό μιας γλώσσας, το σύνολο των επιτρεπτών συντακικών δομών μιας γλώσσας κ.ά.)

Μονάδα πληθυσμού (population unit)

- Η βασική μονάδα του πληθυσμού (ΜΠ) διαφέρει ανάλογα με τον κλάδο της ΠΓ.
- Μερικά παραδείγματα:
  - στην λεξικοστατιστική η βασική ΜΠ είναι η λέξη,
  - στην ανάλυση της γλωσσικής ποικιλίας η βασική ΜΠ είναι οποιοδήποτε γλωσσικό χαρακτηριστικό το οποίο εμφανίζει ποικιλία χρήσης (από εναλλαγή φωνημάτων έως και διαφοροποίηση χρήσης συντακτικών δομών).

Δείγμα (sample)

- Επειδή η μελέτη του πληθυσμού τις περισσότερες φορές είναι δύσκολη έως αδύνατη (λόγω του μεγέθους του, του είδους του κ.ά.) συνήθως αποσπούμε για μελέτη ένα υποσύνολο το οποίο αποκαλούμε δείγμα (sample).
- Αν διατηρηθούν ορισμένες προϋποθέσεις (η σημαντικότερη από τις οποίες είναι η τυχαιότητα) τότε τα ποσοτικά χαρακτηριστικά του δείγματος είναι ίδια με του πληθυσμού και τα συμπεράσματα που θα εξάγουμε για το δείγμα θα ισχύουν και για τον πληθυσμό.

Υπολογισμός των χαρακτηριστικών του πληθυσμού

- Ο στόχος του τυχαίου δείγματος είναι ο υπολογισμός κάποιων χαρακτηριστικών του πληθυσμού. Τα χαρακτηριστικά αυτά μπορεί να είναι:
  - Ο μέσος όρος κάποιας μεταβλητής (Π.χ. Ο μέσος όρος χρήσης παθητικής φωνής σε κείμενα δημοσιογραφικού λόγου.).
  - Η αναλογία σε κάποια κατηγορία (π.χ. Η αναλογία επιθέτων σε κάποιο κείμενο).
- Τα χαρακτηριστικά του πληθυσμού είναι άγνωστα, αλλά οι τιμές που μετράμε στο δείγμα μπορούν να χρησιμοποιηθούν για να τις εκτιμήσουμε.
- Όταν τα χαρακτηριστικά του πληθυσμού εκτιμούνται από αυτά του δείγματος τότε αναμένεται η ύπαρξη ενός δειγματοληπτικού λάθους.

Είδη πληθυσμών

- Δείγμα πεπερασμένου πληθυσμού
  - Σε ορισμένες περιπτώσεις το δείγμα προέρχεται από έναν πληθυσμό που είναι συγκεκριμένος και μπορεί να μετρηθεί.
  - Π.χ. η γλωσσομάθεια των ανδρών στην Ελλάδα
- Δείγμα υποθετικού πληθυσμού
  - Είναι πιθανό να περιγράψουμε έναν πληθυσμό ως ένα σύνολο όλων των πιθανών μετρήσεων που θα μπορούσαν να γίνουν αν η συλλογή των δεδομένων επαναλαμβανόταν.
  - Αυτός ο τύπος του πληθυσμού δεν έχει συγκεκριμένη ύπαρξη – είναι υποθετικός και μάλιστα άπειρος.
  - Π.χ. Η μέτρηση της ταχύτητας του φωτός μας δίνει διαφορετικές τιμές κάθε φορά όλες πολύ κοντινές, αλλά με υπαρκτή και συνεχόμενη διαφοροποίηση κάθε φορά.

Επίδραση του μεγέθους του δείγματος στο δειγματοληπτικό λάθος

- Όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο μικρότερο είναι το δειγματοληπτικό λάθος. Ωστόσο, όταν ο πληθυσμός είναι μεγάλος, η δειγματοληψία ενός μικρού ποσοστού του δείγματος μπορεί να δώσει εξίσου αξιόπιστες εκτιμήσεις.
- **Το δειγματοληπτικό λάθος εξαρτάται πολύ περισσότερο από το μέγεθος του δείγματος παρά από το ποσοστό του πληθυσμού από τον οποίο παίρνουμε το δείγμα.**
  - Για παράδειγμα, ένα δείγμα 10 ατόμων από ένα πληθυσμό 10.000 θα εκτιμήσει την αναλογία ανδρών / γυναικών στον πληθυσμό με την ίδια περίπου ακρίβεια, όσο και ένα δείγμα 10 ατόμων για ένα πληθυσμό 100.

Είδη δειγματοληψίας

- Τυχαία
  - Απλή τυχαία
    - Με αντικατάσταση
    - Χωρίς αντικατάσταση
    - Συστηματική
  - Διαστρωματωμένη
  - Συστάδων
- Μη τυχαία
  - Ευκολίας
  - Κρίσης
  - Κρίσιμων περιπτώσεων
  - Προκαθορισμένης ποσόστωσης
  - Εθελοντών
  - «Χιονοστιβάδας»

Απλή τυχαία δειγματοληψία (1)

- Σε ένα τυχαίο δείγμα μεγέθους *n* το οποίο έχει παρθεί από έναν πεπερασμένο πληθυσμό *N* τιμών, κάθε τιμή έχει την ίδια πιθανότητα να περιληφθεί στο δείγμα.
- **Δύο διαφορετικές μέθοδοι** απλής τυχαίας δειγματοληψίας είναι οι πλέον συνηθισμένες: Και οι δύο μπορούν να εφαρμοστούν εν σειρά, ξεκινώντας από μια τυχαία τιμή του πληθυσμού.
- **Δειγματοληψία χωρίς αντικατάσταση (Sampling without replacement - SWOR)**
  - Στην SWOR, η πρώτη τιμή που επιλέγεται αφαιρείται από τον πληθυσμό και η δεύτερη επιλέγεται από τις υπόλοιπες N – 1 τιμές του πληθυσμού.
  - Στην SWOR κάθε πιθανό υποσύνολο n τιμών από τον πληθυσμό έχει την ίδια πιθανότητα να επιλεγεί.

Απλή τυχαία δειγματοληψία (2)

- **Δειγματοληψία με αντικατάσταση (Sampling with replacement - SWR)**
  - Στην SWR, η πρώτη τιμή που επιλέγεται **επιστρέφει στον πληθυσμό** και η δεύτερη τιμή επιλέγεται τυχαία από το σύνολο N των τιμών του πληθυσμού.
  - Σε αντίθεση με την SWOR, ένα δείγμα με αντικατάσταση μπορεί να περιέχει την ίδια τιμή του πληθυσμού περισσότερο από μία φορά.
- Η SWOR προτιμάται όταν υπάρχει η δυνατότητα να χρησιμοποιηθούν και οι δύο. Ωστόσο, περιστασιακά υπάρχει η πιθανότητα οι τιμές ενός πληθυσμού να μην μπορούν να αφαιρεθούν από αυτόν. Π.χ. Ένα παράδειγμα θα ήταν ένας βιολόγος που καταγράφει χαρακτηριστικά ζώων που εντοπίζονται σε μια περιοχή. Σε μια τέτοια περίπτωση δεν θα ξέρει αν κάποιο ζώο το έχει ήδη καταγράψει ή όχι.
- **Όταν ο πληθυσμός είναι μεγάλος (και σημαντικά μεγαλύτερος από το μέγεθος του δείγματος), τότε τα SWR and SWOR είναι σχεδόν ίδια.**

Όταν το τυχαίο δείγμα δεν είναι και τόσο ... τυχαίο

- Σημαντικό ρόλο στην απλή τυχαία δειγματοληψία παίζει η σωστή καταγραφή των μονάδων παρατήρησης του πληθυσμού.
- Προεδρικές εκλογές στις ΗΠΑ (1936): Στις 31 Οκτωβρίου του 1936 δημοσκόπηση του Reader's Digest προέβλεψε:
  - Alf Landon: 55%
  - Ρούσβελτ: 41%
- Τα αποτελέσματα έδωσαν … :
  - Alf Landon: 37%
  - Ρούσβελτ: 61%
- Η σημαντικότατη αυτή απόκλιση είχε πολλές αιτίες (Squire 1988) με σημαντικότερη ίσως την λανθασμένη καταγραφή του πληθυσμού. Το δείγμα επιλέχθηκε τυχαία μέσα από:
  - τους τηλεφωνικούς καταλόγους
  - τα αρχεία των αδειών οδήγησης αυτοκινήτων στις ΗΠΑ

Διαλέγοντας ένα τυχαίο δείγμα Ι

- Μια μέθοδος για να επιλεγεί ένα τυχαίο δείγμα μεγέθους *n* είναι...
  - Γράφουμε τα ονόματα (ή άλλα χαρ/κά) του πληθυσμού σε όμοια τμήματα χαρτιού,
  - Τα αναμειγνύουμε σε κάποιο κουτί
  - Επιλέγουμε *n* χαρτιά.
- Η μέθοδος αυτή ωστόσο, είναι ... μη πρακτική για μεγάλους πληθυσμούς.

Διαλέγοντας ένα τυχαίο δείγμα ΙΙ

- Μια εναλλακτική μέθοδος επιλογής τυχαίου δείγματος περιλαμβάνει την παραγωγή τυχαίων αριθμών (0, 1, ..., 9). Υπάρχουν πολλοί τρόποι με τους οποίους μπορούν να παραχθούν τυχαίοι αριθμοί, έτσι ώστε ο κάθε ένας από αυτούς έχει την ίδια πιθανότητα εμφάνισης. Μερικοί από αυτούς είναι:
  - Να ρίξουμε ένα ζάρι 10 πλευρών αρκετές φορές.
  - Να χρησιμοποιήσουμε πίνακες τυχαίων αριθμών. Μπορείτε να ανοίξετε μια σελίδα στην τύχη και να επιλέξετε από εκείνην και έπειτα ακολουθία ψηφίων.
  - Να χρησιμοποιήσουμε λογισμικό (π.χ. Microsoft Excel). Πολλά προγράμματα περιέχουν αλγόριθμους παραγωγής τυχαίων αριθμών (ψευδοτυχαίων για την ακρίβεια).

Χρησιμοποιώντας τους τυχαίους αριθμούς

- Η χρήση τυχαίων αριθμών για την επιλογή ενός δείγματος (δίχως αντικατάσταση) μπορεί να περιγραφεί στα ακόλουθα 4 βήματα:
  1. Αριθμείστε όλα τα μέλη του πληθυσμού, ξεκινώντας από το 0
  2. Παράγετε μια τυχαία τιμή με τον ίδιο αριθμό ψηφίων όσων έχει ο μεγαλύτερος αριθμός που σχετίζεται με κάποιο μέλος του πληθυσμού.
  3. Αν η τιμή που παραχθεί ανήκει σε κάποιο μέλος του πληθυσμού, και αυτό το μέλος δεν περιλαμβάνεται ήδη στο δείγμα, προσθέστε το στο δείγμα. (Διαφορετικά αγνοήστε τον συγκεκριμένο τυχαίο αριθμό και παράγετε άλλον).
  4. Επαναλαμβάνετε τα βήματα 2 και 3 μέχρι ένα αρκετά μεγάλο δείγμα έχει συλλεχθεί.

Συστηματική δειγματοληψία (1/2)

- Ο ερευνητής θα πρέπει με κάποιο τυχαίο τρόπο να επιλέξει μια μονάδα παρατήρησης από τον πληθυσμό Από την στιγμή που βρεθεί η τυχαία μονάδα εκκίνησης ο ερευνητής παίρνει όλες τις μονάδες που απέχουν κάποιο συγκεκριμένο διάστημα από αυτήν.
- Στα πλεονεκτήματα αυτής της μεθόδου συγκαταλέγεται και η εύκολη δυνατότητα εφαρμογής της σε πληθυσμούς που δεν έχουν αντιστοιχηθεί με νούμερα. Για παράδειγμα θα μπορούσε να χρησιμοποιηθεί πολύ γρήγορα σε έναν τηλεφωνικό κατάλογο και να επιλεγεί ένα δείγμα ξεκινώντας από μια τυχαία σελίδα του καταλόγου και προχωρώντας με ένα σταθερό διάστημα μέχρι να καταρτιστεί το επιθυμητό μέγεθος του δείγματος.

Συστηματική δειγματοληψία (2/2)

- Τα συστηματικά τυχαία δείγματα συμπεριφέρονται ακριβώς όπως τα τυχαία δείγματα χωρίς αντικατάσταση με την προϋπόθεση ότι η καταγραφή των μονάδων παρατήρησης του πληθυσμού δεν παρουσιάζει κάποιας μορφής κανονικότητα ή κυκλικότητα. Για τον λόγο αυτό είναι σημαντικό ο κατάλογος καταγραφής να παρουσιάζει τις μονάδες παρατήρησης με τυχαίο τρόπο. Εναλλακτικά, μπορεί να χρησιμοποιεί μια μορφή συστηματικής οργάνωσης που ωστόσο δεν θα επηρεάζει το φαινόμενο το οποίο ερευνούμε.

Διαστρωματωμένη δειγματοληψία

- Αν τα άτομα στον πληθυσμό μπορούν να χωριστούν σε διαφορετικές ομάδες (ονομάζονται στρώματα (**strata)** στην ορολογία της δειγματοληψίας), είναι καλύτερο να παίρνουμε ένα απλό τυχαίο δείγμα μέσα σε κάθε ξεχωριστή ομάδα , από το να δειγματοληπτούμε σε όλο τον πληθυσμό. Αυτό ονομάζεται διαστρωματωμένη τυχαία δειγματοληψία (**stratified random sample**).
- Για παράδειγμα ένα τυχαίο δείγμα 40 φοιτητών από μια τάξη 200 ανδρών και 200 γυναικών θα μπορούσε να περιλαμβάνει 25 άνδρες και 15 γυναικών. Ένα τυχαίο διαστρωματωμένο δείγμα θα μπορούσε να επιλέξει 20 άνδρες και 20 γυναίκες διασφαλίζοντας ότι η αναλογία του φύλου ταιριάζει αυτήν του πληθυσμού.
- Η ακρίβεια της συγκεκριμένης μεθόδου είναι πολύ μεγαλύτερη από την απλή τυχαία δειγματοληψία όταν:
  - Οι μέσοι όροι του φαινομένου που μετράμε διαφοροποιούνται σημαντικά μεταξύ των στρωμάτων
  - Η διακύμανση των τιμών του φαινομένου που μετράμε στο εσωτερικό του κάθε στρώματος είναι μικρή, δηλαδή οι τιμές του είναι κοντά η μία στην άλλη και δεν παρουσιάζουν σημαντικές αποκλίσεις.

Δειγματοληψία συστάδων

Δειγματοληπτικό πλαίσιο (Sampling frame)

- Τόσο τα απλά όσο και τα διαστρωμένα τυχαία δείγματα απαιτούν τη γνώση όλων των μονάδων του πληθυσμού. Αυτό ωστόσο, δεν είναι πάντα εφικτό και τότε θα πρέπει να επιλεγεί μια διαφορετική μεθοδολογία. Για παράδειγμα κάποιος ερευνητής θέλει να πάρει δείγμα της εφηβικής ομιλίας. Χωρίς μια λίστα με τα σπίτια που έχουν εφήβους πώς μπορούμε να προσεγγίσουμε το δείγμα μας;

Δειγματοληψία συστάδων (Cluster sampling)

- Μια λύση είναι να ομαδοποιηθούν τα άτομα σε σχετικά μικρές ομάδες, τις **συστάδες** (**clusters)**, για τις οποίες μια πλήρης λίστα είναι διαθέσιμη. Οι συστάδες είναι παρόμοιες με τα στρώματα, αλλά συνήθως πολύ μικρότερα.
- Στην δειγματοληψία συστάδων, επιλέγεται ένα τυχαίο δείγμα συστάδων με όλα τα μέλη τους.
- Ένα από τα σημαντικότερα πλεονεκτήματά του είναι η οικονομία που επιτυγχάνεται.

Διεπίπεδη δειγματοληψία

- Η διεπίπεδη δειγματοληψία (two-stage sampling) σχετίζεται με τη δειγματοληψία συστάδων αλλά χρησιμοποιείται σε μεγάλους πληθυσμούς.
- Στην διεπίπεδη δειγματοληψία ο πληθυσμός χωρίζεται σε ομάδες «γειτονικών» μονάδων που ονομάζονται **πρωτεύουσες δειγματοληπτικές μονάδες** (primary sampling units). Αυτές οι μονάδες είναι μεγάλες, όπως π.χ. Οι νομοί μιας χώρας. Ένας μικρός αριθμός από αυτές τις μονάδες επιλέγεται με βάση κάποια δειγματοληπτική μονάδα και εν συνεχεία σε κάθε μονάδα επαναλαμβάνεται η δειγματοληψία.
- Ένα σημαντικό πλεονέκτημα είναι και εδώ το μειωμένο κόστος, αν και η ακρίβεια είναι μειωμένη.

Μη τυχαίες μέθοδοι δειγματοληψίας: Δείγμα Ευκολίας

- Η δειγματοληψία ευκολίας στηρίζεται σε άτομα που είναι διαθέσιμα και μπορούν εύκολα να συμμετάσχουν σε μια έρευνα.
- Ωστόσο, μια σειρά από μεταβλητές, όπως η ηλικία, το φύλο, η ψυχολογική και κοινωνική ταυτότητα το γλωσσικό υπόστρωμα, η εκπαίδευση και άλλοι παράγοντες δεν μπορούν να ελεγχθούν αποτελεσματικά ως προς την συσχέτισή τους με τον ερευνητικό πληθυσμό με συνακόλουθο αποτέλεσμα να μην μπορεί να υπολογιστεί η ακρίβεια των μετρήσεων που μας παρέχει το δείγμα.
- Το δείγμα ευκολίας μπορεί να αποδειχθεί ιδιαίτερα αποτελεσματικό στην περίπτωση την οποία ο ερευνητής θέλει να ανιχνεύσει το υπό μελέτη φαινόμενο και να πάρει μια χονδρική εικόνα για την συχνότητά του και τις βασικότερες μεταβλητές με τις οποίες συσχετίζεται. Το δείγμα ευκολίας μπορεί να συλλεχθεί γρήγορα και να δώσει μεγάλο αριθμό δεδομένων δίχως να απαιτεί από τον ερευνητή μεγάλο κόστος ερευνητικού σχεδιασμού.
- Μερικές φορές το δείγμα ευκολίας είναι εξίσου αποτελεσματικό με το τυχαίο δείγμα όταν το υπό μελέτη φαινόμενο παρουσιάζει σημαντική ομοιογένεια στην διασπορά του μέσα στον πληθυσμό.

Δειγματοληψία κρίσης

- Ο ερευνητής πολλές φορές έχοντας μια ικανοποιητική εικόνα της ενός φαινομένου μπορεί να επιλέξει το δείγμα βασιζόμενος στην προσωπική του κρίση. Αξιοποιώντας την γνώση που του παρέχει η βιβλιογραφία θα συμπεριλάβει περιπτώσεις που κατά τη γνώμη του θα αποτελέσουν σημαντικό παράγοντα διαμόρφωσης των ερευνητικών αποτελεσμάτων και θα αποκλείσει ακραίες ή μη τυπικές περιπτώσεις που θα μπορούσαν να διαστρεβλώσουν την ποσοτική απεικόνιση του υπό μελέτη φαινομένου.
- Η επιλογή με βάση την κρίση του ερευνητή όσο και αν βασίζεται σε προηγούμενη εμπεριστατωμένη έρευνα, ενέχει σημαντικά περιθώρια λάθους, αφού η αντικειμενικότητα της κρίσης του μπορεί να επηρεαστεί από ιδεολογικούς, κοινωνικούς, ψυχολογικούς, προσωπικούς κ.ά. παράγοντες που δεν μπορούν να αποκλειστούν και να υπολογιστεί η επίδρασή τους.
- Σε ορισμένες περιπτώσεις, μπορεί να ενταχθεί σε μια ευρύτερη ερευνητική μεθοδολογία προκαταρκτικής εξέτασης ενός μεγάλου αριθμού φαινομένου από την οποία θα επιλεγεί τελικά ένας μικρός αριθμός που παρουσιάζει ιδιαίτερες διακυμάνσεις μέσα στα δείγματα.

Δειγματοληψία κρίσιμων περιπτώσεων

- Ο ερευνητής επιλέγει περιπτώσεις που έχουν συγκεκριμένα χαρακτηριστικά και που βάσει της κρίσης του ερευνητή είναι ουσιώδη για την έρευνα και αντιπροσωπεύουν επαρκώς το φαινόμενο που μελετά.
- Μια τέτοια μεθοδολογία εκτός των προβλημάτων της μη τυχαιότητας παρουσιάζει τις περισσότερες φορές έλλειψη θεωρητικής τεκμηρίωσης. Αν κάποια άτομα παρουσιάζουν μια συμπεριφορά που εμπειρικά συγκλίνει με τον ερευνητικό πληθυσμό δεν σημαίνει ότι αυτή η σύγκλιση θα συνεχίζει να επαναλαμβάνεται για πάντα. Ειδικά όταν η σύγκλιση δεν ερμηνεύεται στο πλαίσιο κάποιου συγκεκριμένου θεωρητικού πλαισίου, τότε καθίσταται συμπτωματική και δεν μπορεί να χρησιμοποιηθεί με ασφάλεια για μελλοντικές έρευνες.

Δειγματοληψία προκαθορισμένης ποσόστωσης

- Αποτελεί την μη τυχαία εφαρμογή της διαστρωματωμένης δειγματοληψίας. Πρώτα καθορίζεται ένας συγκεκριμένος αριθμός για κάθε υπο-ομάδα του δείγματος και στη συνέχεια επιλέγονται μέλη βασιζόμενα στην κρίση του ερευνητή.
- Η συγκεκριμένη μέθοδος:
  - Επιλέγει με μη τυχαίο τρόπο τα μέλη του δείγματος με ό,τι αυτό συνεπάγεται
  - Αποκρύβει τα ποσοστά των ατόμων που δεν θέλουν να απαντήσουν σε μια συγκεκριμένη ερώτηση ή να συμμετάσχουν στην έρευνα
- Το δείγμα προκαθορισμένης ποσόστωσης θεωρείται ποιοτικότερο από τα άλλα δείγματα που προκύπτουν από μη τυχαίες μεθόδους. Είναι ευκολότερο και αρκετά φθηνότερο από την ομόλογη του τυχαία μέθοδο της διαστρωματωμένης δειγματοληψίας. Είναι ιδιαίτερα κατάλληλο για έρευνες που προσεγγίζουν το δείγμα τηλεφωνικά και γενικότερα πληθυσμούς που δεν μπορούν εύκολα να καταγραφούν τα μέλη τους.

Δειγματοληψία εθελοντών

- Τα συγκεκριμένα δείγματα προκύπτουν μέσα από την εθελοντική συμμετοχή των ατόμων σε μια έρευνα.
- Η συγκεκριμένη μέθοδος επιτρέπει τον γρήγορο σχηματισμό μη τυχαίων δειγμάτων με σοβαρά όμως προβλήματα αντιπροσωπευτικότητας (π.χ. Τηλεψηφοφορίες, σφυγμομετρήσεις των ΜΜΕ κ.ά.)
- Σε κάποιες άλλες περιπτώσεις ο ερευνητής για να ολοκληρώσει μια πολύπλοκη έρευνα θα πρέπει να έχει συχνή επαφή με το δείγμα του και να απασχολεί αρκετές ώρες τα μέλη του. Σε οποιοδήποτε τυχαίο δείγμα η πιθανότητα κάποιος να δεχτεί να συμμετάσχει σε μια έρευνα που απαιτεί πολλές ώρες προσωπικής δέσμευσης είναι πολύ μικρή. Στην περίπτωση αυτή η μόνη εναλλακτική λύση είναι η δημιουργία δείγματος εθελοντών με την χρήση διαφόρων κινήτρων, όπως π.χ. κάποια μικρό χρηματικό αντίτιμο ως αποζημίωση για την συμμετοχή στην έρευνα. Δίχως μια τέτοια εθελοντική βάση συνεργασίας (έστω και μερικώς αποζημιούμενη) πολύπλοκες έρευνες με απαιτήσεις από τα μέλη του δείγματος δεν θα μπορούσαν να ολοκληρωθούν σε λογικά πλαίσια κόστους και χρόνου.

Δειγματοληψία «χιονοστιβάδας» (1)

- Όταν η μελέτη επικεντρώνεται σε υπο-πληθυσμούς που είναι δύσκολο να προσεγγιστούν λόγω της κλειστής οργάνωσής τους ή του περιθωριακού τους χαρακτήρα τότε μια τυχαία δειγματοληψία πολύ δύσκολα θα συγκεντρώσει ικανοποιητικό μέγεθος δείγματος.
- Σε τέτοιες περιπτώσεις συχνά ο ερευνητής χρησιμοποιεί την δειγματοληψία «χιονοστιβάδας».
  - Με αυτή την μέθοδο ο ερευνητής χρησιμοποιεί ένα δίκτυο προσωπικών του επαφών για να έρθει σε επικοινωνία με έναν μικρό αριθμό ατόμων που σχετίζονται με τον υπο-πληθυσμό που ερευνά.
  - Εν συνεχεία και αφού έχει κερδίσει την εμπιστοσύνη του αρχικού αυτού μικρού δείγματος ζητά από τα μέλη του να τον συστήσουν σε παρόμοια άτομα για να συνεχίσει την έρευνα. Με τον τρόπο αυτό αυξάνεται γεωμετρικά το μέγεθος του δείγματος του, ακριβώς όπως μια χιονοστιβάδα μεγαλώνει συνεχώς καθώς κατηφορίζει μια πλαγιά.

Δειγματοληψία «χιονοστιβάδας» (2)

- Η έλλειψη τυχαιότητας προκαλεί και εδώ ένα σημαντικό πρόβλημα αξιοπιστίας που σχετίζεται με την ακρίβεια των εκτιμήσεων που παίρνουμε από ένα τέτοιο δείγμα. Τα δείγματα χιονοστιβάδας παρουσιάζουν σημαντικότερη απόκλιση από άλλα μη τυχαία δείγματα όπως αυτό της κρίσης ή της προκαθορισμένης ποσόστωσης γιατί σε κάθε επέκταση του αρχικού δείγματος πολλοί και διαφορετικοί άνθρωποι με καθαρά προσωπικά κριτήρια επιλέγουν τα επόμενα μέλη του δείγματος. Το κάθε άτομο που μετέχει σε ένα τέτοιο δείγμα γίνεται δυνάμει ερευνητής και προχωρά στην άτυπη κατάρτιση ενός δικού του δείγματος με καθαρά προσωπικά κριτήρια, δίχως να έχει τα θεωρητικά εφόδια να το κάνει. Έτσι το δείγμα που προκύπτει από μια τέτοια αλυσιδωτή διαδικασία είναι μοναδικό και εξαιρετικά ιδιόμορφο.
- Η δειγματοληψία χιονοστιβάδας είναι ιδιαίτερα χρήσιμη όταν το επίκεντρο της είναι κλειστές κοινωνικές ομάδες όπως χρήστες ναρκωτικών, παράνομοι μετανάστες, θρησκευτικές ή πολιτικές μειονότητες καθώς και άτομα που ανήκουν στα άκρα του κοινωνικού συστήματος δηλ. άτομα πολύ πλούσια και με υψηλή κοινωνική θέση ή άτομα που ανήκουν στην ανέχεια.

Είδη δειγματοληπτικών λαθών

- Ελλιπής κάλυψη δείγματος (coverage error)
  - Πολλές φορές ο σχεδιασμός της έρευνας δεν υπολογίζει το σύνολο του πληθυσμού και περιορίζεται μόνο σε ένα μέρος του.
- Ελλιπής ανταπόκριση στην έρευνα (non-response error). Πιθανές αιτίες:
  - Αποτυχία να επικοινωνήσουμε με τα άτομα. Π.χ. σε μια τηλεφωνική έρευνα δεν θα υπάρχει απάντηση σε κάποιους αριθμούς.
  - Άρνηση να συμμετάσχουν τα άτομα στην έρευνα. Π.χ. Μερικά άτομα είναι πολύ απασχολημένα ή θα αρνηθούν την «εισβολή» στο ιδιωτικό τους χώρο.
  - Άρνηση να απαντήσουν σε συγκεκριμένες ερωτήσεις. Π.χ. Οι άνθρωποι είναι συχνά απρόθυμοι να αποκαλύψουν πληροφορίες που έχουν κάποια προσωπική ή εμπορική αξία για αυτούς.
- Λάθος ερευνητικού εργαλείου (instrument error). Προκύπτει από ελλιπή σχεδιασμό του ερωτηματολογίου.
- Λάθος που προέρχεται από τον ερευνητή (interviewer error). Ιδιαίτερα από κάποια χαρακτηριστικά του ή ακόμα και από την ίδια την παρουσία του (παράδοξο του παρατηρητή).

Μέθοδοι προσέγγισης του δείγματος Ι

- **Τηλέφωνο**
  - Οικονομική
  - Η έλλειψη απαντήσεων μπορεί να είναι υψηλή.
  - Το λάθος κάλυψης είναι μικρό.
  - Ο αριθμός των ερωτήσεων θα πρέπει να είναι μικρός.
- **Ταχυδρομημένο ερωτηματολόγιο**
  - Η έλλειψη απαντήσεων είναι πολύ υψηλή.
  - Ακριβότερη από τις τηλεφωνικές έρευνες.
  - Ο αριθμός των ερωτήσεων μπορεί να είναι μεγαλύτερος.

Μέθοδοι προσέγγισης του δείγματος ΙΙ

- **Συνέντευξη**
  - Η ερευνητές που προσεγγίζουν τα άτομα στο σπίτι τους έχουν μεγαλύτερη πιθανότητα να πάρουν απαντήσεις σε μεγάλα ερωτηματόλογια.
  - Χαμηλά λάθη ελλιπών απαντήσεων.
  - Οι ερευνητής θα πρέπει να είναι καλά εκπαιδευμένοι για να μειώσουν το λάθος που προκύπτει από την παρουσία τους.
  - Ακριβή έρευνα.
  - Τα ερωτηματολόγια μπορούν να είναι σχετικά μεγάλα.
  - Τα σπίτια σπάνια επιλέγονται τυχαία. Συνήθως επιλέγονται τυχαία δρόμοι και μετά κάθε 5ο ή 10ο σπίτι επιλέγεται ένα. Αυτό ονομάζεται και **συστηματική δειγματοληψία**.

Μέθοδοι προσέγγισης του δείγματος ΙΙΙ

- **Δημόσια μέρη: κάποιες έρευνες γίνονται προσεγγίζοντας ανθρώπους σε εμπορικά κέντρα ή άλλα πολυσύχναστα δημόσια μέρη.**
  - Μεγάλα λάθη κάλυψης
  - Οικονομική και γι' αυτό ευρέως χρησιμοποιούμενη
  - Τα ερωτηματολόγια πρέπει να είναι μικρά
  - Για να μειωθούν τα λάθη κάλυψης συχνά χρησιμοποιείται **δειγματοληψία ποσόστωσης** (quota sample). Ο κάθε ερευνητής καλείται να επιλέξει συγκεκριμένο αριθμό ανδρών, γυναικών, νέων, ηλικιωμένων κ.ά.. Η αναλογία για κάθε χαρακτηριστικό είναι τέτοια που αντανακλά την αναλογία στον ερευνούμενο πληθυσμό.
- **Αυτοεπιλογή (Self-selected)**
  - Η αυτόκλητη συμμετοχή με τηλεφωνήματα ή γράμματα σε έρευνες γίνεται συνήθως από τηλεοπτικούς, ραδιοφωνικού σταθμούς και περιοδικά. Το δείγμα αυτό δεν έχει καμία αντιπροσωπευτικότητα και τα αποτελέσματα δεν έχουν καμία σημασία. Τέτοιου τύπου έρευνες θα πρέπει γενικά να αποφεύγονται.

Ιδιότητες των πιθανοτήτων Ι

- Οι πιθανότητες είναι πάντα μεταξύ 0 και 1.
  - Για κάθε γεγονός, *A*, **0 ≤ P(*A*) ≤ 1**
- Για κάθε γεγονός, *A*:
  - **P(*A*) = 0 σημαίνει ότι το γεγονός Α δεν μπορεί να συμβεί**
  - **P(*A*) = 1 σημαίνει ότι το γεγονός Α είναι σίγουρο ότι θα συμβεί**
- **Η πιθανότητα να μην συμβεί κάποιο γεγονός Α είναι:**
  - **P(Α δεν συμβαίνει)= 1 – P(Α)**

Ιδιότητες των πιθανοτήτων ΙΙ

- **Αθροιστικός νόμος:** όταν δύο γεγονότα δεν μπορούν να συμβούν μαζί , λέγεται ότι είναι αμοιβαία αποκλειόμενα. Για κάθε δύο αμοιβαία αποκλειόμενα γεγονότα, Α και Β:
  - **P(*A* ή *B*) = P(*A*) + P(*B*) P(9 ή 10)= P(9) + P(10) = (0,1) + (0,1)= 0,2**
- Αν τα γεγονότα Α και Β δεν είναι αμοιβαία αποκλειόμενα τότε:
  - **P(*A* ή *B*) = P(*A*) + P(*B*) – P (Α και Β)**
  - **Η πιθανότητα επιλογής ενός αριθμού μεγαλύτερου από 8 (P(A)) ή η επιλογή μονού αριθμού (P(B)).**
  - **P(9, 10) + P(1,3,5,7,9) – P(9) = P(0,2) + P(0,5) – P(0,1)=0,6**

Ιδιότητες των πιθανοτήτων ΙΙΙ

- Η εξέταση της πιθανότητας δύο ή περισσότερων γεγονότων όταν συμβαίνουν ταυτόχρονα ή διαδοχικά απαιτεί τον πολλαπλασιασμό των πιθανοτήτων του κάθε γεγονότος ξεχωριστά.
- ***Στατιστικά ανεξάρτητα γεγονότα*:** Τέτοια είναι η περίπτωση που θέλουμε να υπολογίσουμε την πιθανότητα να επιλέξουμε το 4 από μια κληρωτίδα με 10 αριθμούς και αφού ξαναβάλουμε το νούμερο πίσω να επιλέξουμε το 2. Η επιλογή του ενός αριθμού είναι ανεξάρτητη από τον άλλο αφού όταν επιλέγουμε τον πρώτο δεν μένουν λιγότεροι αριθμοί στην κληρωτίδα μειώνοντας την πιθανότητα επιλογής του δεύτερου. Η πιθανότητα δίνεται ακολούθως:
  - **P(X *και* Y) = P(X) . P(Y)** Π.χ. P(2 και 4) = P(2) . P(4) = 0,1 . 0,1 = 0,01

Ιδιότητες των πιθανοτήτων IV

- ***Στατιστικά εξαρτημένα γεγονότα*:** Είναι αυτά στα οποία η πιθανότητα εμφάνισης του ενός εξαρτάται από την πιθανότητα εμφάνισης του άλλου. Στο προηγούμενο παράδειγμα της επιλογής του 4 και ύστερα του 2 αν δεν ξαναβάλουμε το 4 στην κληρωτίδα η πιθανότητα να επιλεχθεί το 2 θα είναι μεγαλύτερη (αφού θα υπάρχουν συνολικά 9 αριθμοί αντί 10 που θα ήταν σε αντίστροφη περίπτωση) και δίνεται παρακάτω:
  - **P(X και Y) = P(X) . P(Y|X) ή P(Y) . P(X|Y)** Π.χ. P(2 και 4) = P(2) . P(4|2) = 0,1 . 0,111 = 0,011
  - P(X|Y) αντιπροσωπεύει τη λεγόμενη εξαρτημένη πιθανότητα (conditional propability) δηλ. την πιθανότητα να επιλεγεί το δεύτερο γεγονός αφού το πρώτο έχει ήδη συμβεί.

Παράμετροι και στατιστική

- Οι τιμές που περιγράφουν την ποικιλία του πληθυσμού ονομάζονται παράμετροι πληθυσμού και είναι σταθερές.
- Οι τιμές που περιγράφουν το δείγμα ποικίλλουν από δείγμα σε δείγμα.

	Στατιστικές Δείγματος	**Παράμετροι Πληθυσμού**
Μέσος όρος	x	μ
Τυπική απόκλιση	s	σ

Κατανομή του δειγματικού μέσου όρου

- Όλες οι τιμές που περιγράφουν ένα δείγμα ποικίλλουν από δείγμα σε δείγμα. Η ευρύτερα χρησιμοποιούμενη τιμή είναι ο μέσος όρος.
- Όταν μια συγκεκριμένη τιμή αποσπάται από έναν πληθυσμό παρουσιάζει μια κατανομή που περιγράφεται από την κατανομή του πληθυσμού. Όταν ένα τυχαίο δείγμα n τιμών δειγματοληπτείται, ο μέσος όρος του δείγματος είναι επίσης τυχαίος, αλλά παρουσιάζει μια κατανομή που έχει πολύ μικρότερη ποικιλία από την συνολική ποικιλία του πληθυσμού.
- Οι μέσοι όροι των δειγμάτων «κανονικοποιούν» τις ακραίες τιμές στο δείγμα και έτσι οι μέσοι όροι των δειγμάτων τείνουν να «κλείνουν» εγγύτερα στο κέντρο της κατανομής του πληθυσμού.

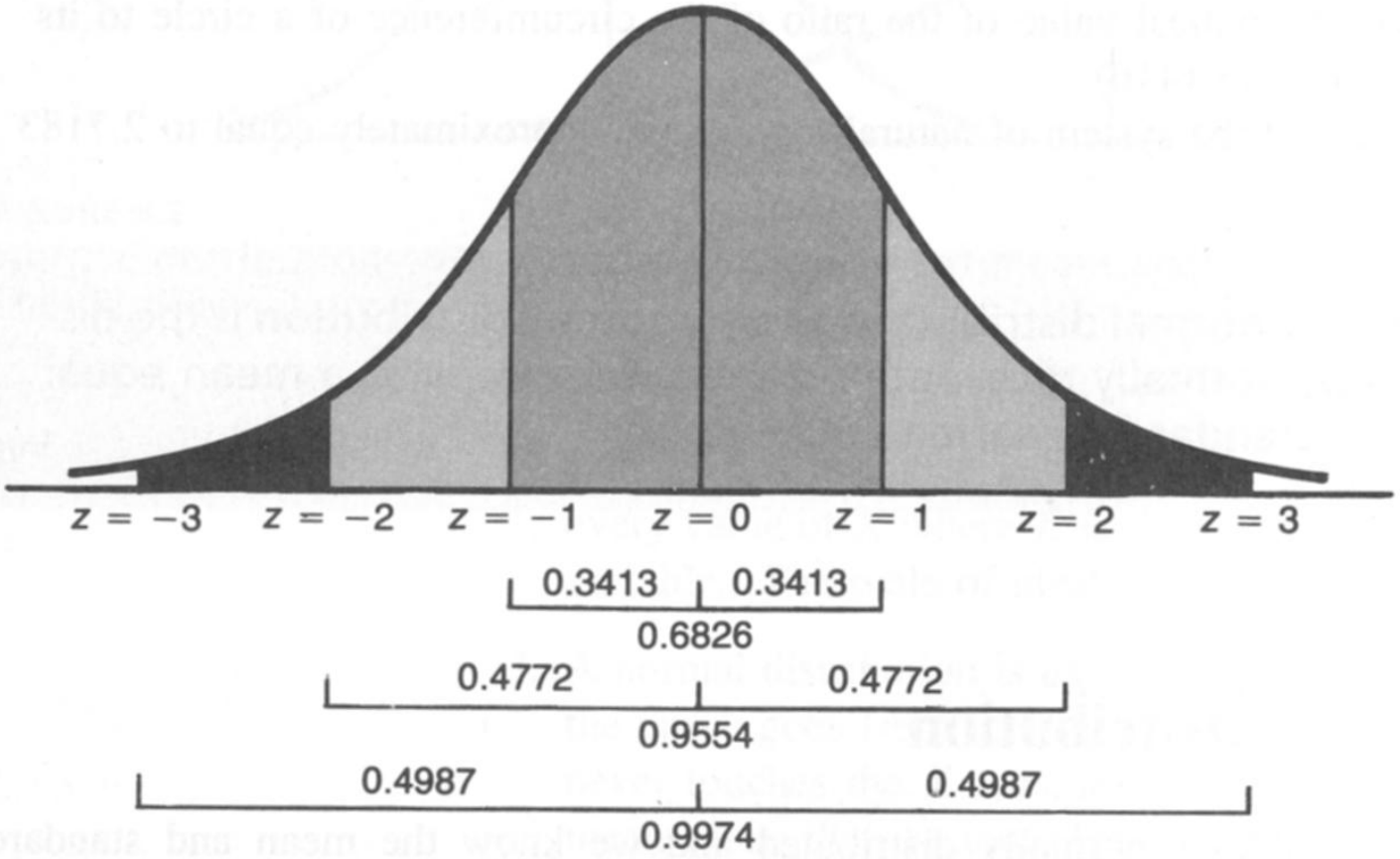
Ιδιότητες του δειγματοληπτικού μέσου όρου

- Ο μέσος όρος του δείγματος έχει μια κατανομή με τις ακόλουθες ιδιότητες:
  - **Έχει κατανομή που επικεντρώνεται στον μέσο όρο του πληθυσμού**
  - **Η ποικιλία του μειώνεται καθώς το μέγεθος του δείγματος μεγαλώνει**
- **Όταν η κατανομή του πληθυσμού είναι κανονική, ο μέσος όρος του δείγματος έχει επίσης κανονική κατανομή**
- Όταν η κατανομή του πληθυσμού δεν είναι κανονική τότε και η κατανομή των μέσων όρων δεν είναι κανονική, αλλά...

  Το θεώρημα του κεντρικού ορίου μας λέει ότι:

  Για τις περισσότερες άλλες κατανομές, η κατανομή του μέσου όρου ενός δείγματος **τείνει στην κανονική** όσο το μέγεθος του δείγματος μεγαλώνει.

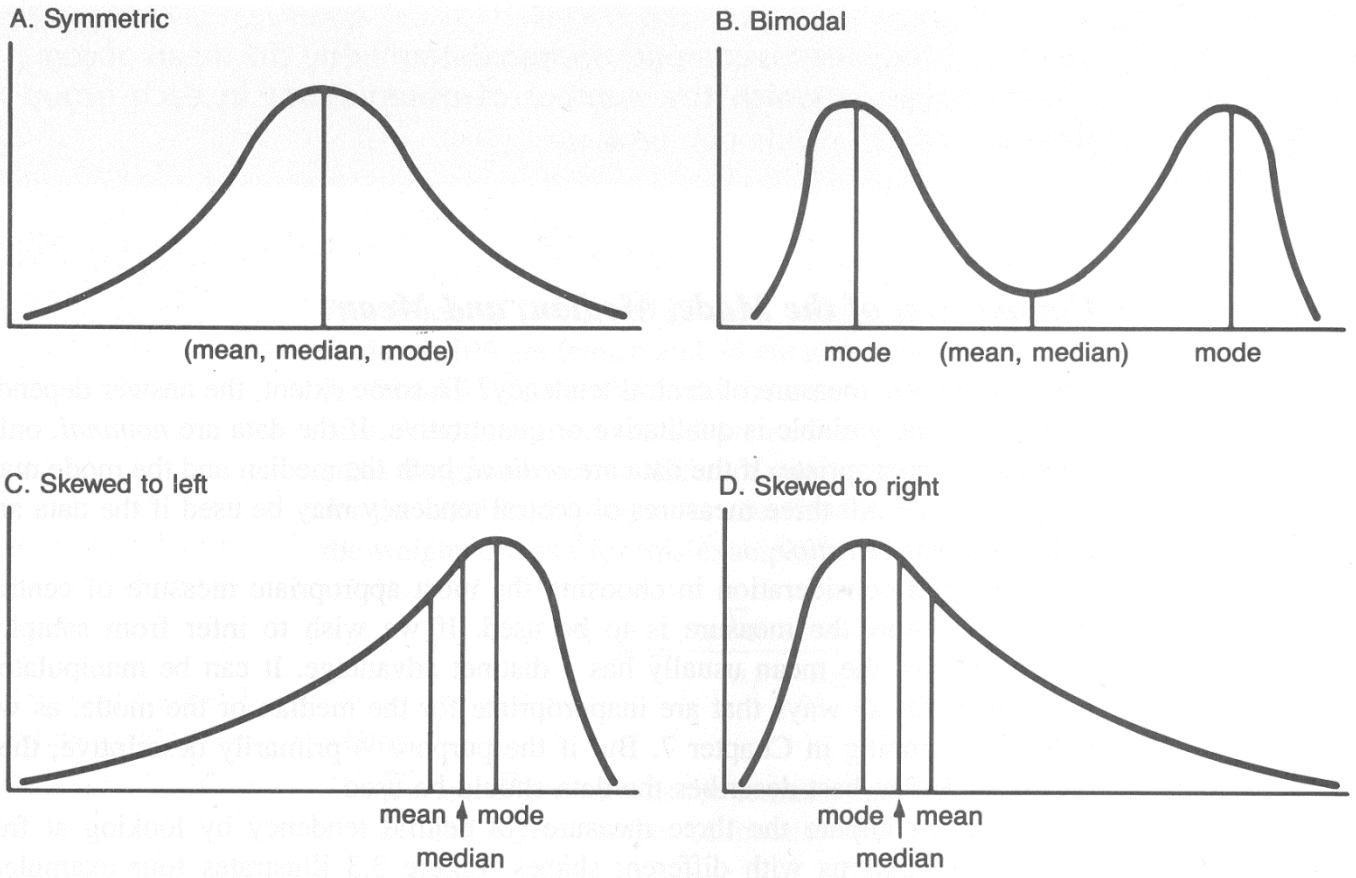
Το σχήμα της κανονικής κατανομής



Κανονική Κατανομή

- Η οικογένεια των κανονικών κατανομών αποτελείται από συμμετρικές, κωδωνοειδείς κατανομές που καθορίζεται από δύο παραμέτρους, τον μέσο όρο (μ) και την τυπική απόκλιση (σ).
- Η κανονική κατανομή χρησιμοποιείται ως πληθυσμός μοντέλο για να εξηγήσει την ποικιλία σε δεδομένα. Ωστόσο, πολλά δεδομένα δεν μπορούν να μοντελοποιηθούν με τη κανονική κατανομή. Μια κανονική κατανομή **δεν** είναι κατάλληλο μοντέλο για ...
  - Δεδομένα που είναι διακριτά
  - Δεδομένα που έχουν μια στρεβλή κατανομή (με μακριά «ουρά» στα αριστερά ή τα δεξιά)
  - Δεδομένα που έχουν πολύ μακριές «ουρές» (με τις περισσότερες τιμές κοντά στο κέντρο, αλλά μικρό ποσοστό των τιμών πολύ μακριά από τον μέσο όρο)
  - Δεδομένα που περιέχουν δύο ή περισσότερες συσπειρώσεις τιμών
- Δεδομένα με στρεβλές κατανομές μπορούν να μετασχηματιστούν σε συμμετρική μορφή και να πλησιάσουν την κανονική κατανομή.

Οικογένειες κατανομών



Τυπική Κανονική Κατανομή

- Όλες οι κανονικές κατανομές έχουν το ίδιο σχήμα ανεξαρτήτως της κλίμακας των δεδομένων. Πώς μπορούμε να καταλήξουμε σε μια κανονική κατανομή που να έχει έναν κοινό οριζόντιο άξονα; Η λύση δίνεται με την τυποποίηση (**standardising**) των τιμών:

$$z = \frac{x - \mu}{\sigma} \qquad z = \frac{X - \overline{X}}{s}$$

- Η τυπική τιμή z έχει την τυπική κανονική κατανομή (**standard normal distribution**) με μέσο όρο = 0 και τυπική απόκλιση = 1.

Ιδιότητες της Τυπικής Κανονικής Κατανομής

- Η P(τιμή μέσα σε διάστημα **1 τ.α.** του μ.ό.) είναι περίπου **0,68**
- Η P(τιμή μέσα σε διάστημα **2 τ.α.** του μ.ό.) είναι περίπου **0,95**
- Η P(τιμή μέσα σε διάστημα **3 τ.α.** του μ.ό.) είναι περίπου **0,997**
- Είναι σημαντικό να θυμάστε ότι περίπου το 95% των τιμών σε έναν πληθυσμό που ακολουθεί την κανονική κατανομή βρίσκεται μέσα σε διάστημα 2 τ.α. της κατανομής του μέσου όρου.
- Για να είμαστε ακριβέστεροι, το **95%** των τιμών σε έναν πληθυσμό κανονικής κατανομής είναι μέσα σε ένα διάστημα **1,96** τυπικών αποκλίσεων από το μέσο όρου.

TABLE C.1
Areas under Standard Normal Curve for Values of *z*

z	*Area between $\bar{x}$ and z*	*Area beyond z*	*Ordinate*	*z*	*Area between $\bar{x}$ and z*	*Area beyond z*	*Ordinate*	*z*	*Area between $\bar{x}$ and z*	*Area beyond z*	*Ordinate*
0.00	.0000	.5000	.3989	0.40	.1554	.3446	.3683	0.80	.2881	.2119	.2897
0.01	.0040	.4960	.3989	0.41	.1591	.3409	.3668	0.81	.2910	.2090	.2874
0.02	.0080	.4920	.3989	0.42	.1628	.3372	.3653	0.82	.2939	.2061	.2850
0.03	.0120	.4880	.3988	0.43	.1664	.3336	.3637	0.83	.2967	.2033	.2827
0.04	.0160	.4840	.3986	0.44	.1700	.3300	.3621	0.84	.2995	.2005	.2803
0.05	.0199	.4801	.3984	0.45	.1736	.3264	.3605	0.85	.3023	.1977	.2780
0.06	.0239	.4761	.3982	0.46	.1772	.3228	.3589	0.86	.3051	.1949	.2756
0.07	.0279	.4721	.3980	0.47	.1808	.3192	.3572	0.87	.3078	.1922	.2732
0.08	.0319	.4681	.3977	0.48	.1844	.3156	.3555	0.88	.3106	.1894	.2709
0.09	.0359	.4641	.3973	0.49	.1879	.3121	.3538	0.89	.3133	.1867	.2685
0.10	.0398	.4602	.3970	0.50	.1915	.3085	.3521	0.90	.3159	.1841	.2661
0.11	.0438	.4562	.3965	0.51	.1950	.3050	.3503	0.91	.3186	.1814	.2637
0.12	.0478	.4522	.3961	0.52	.1985	.3015	.3485	0.92	.3212	.1788	.2613
0.13	.0517	.4483	.3956	0.53	.2019	.2981	.3467	0.93	.3238	.1762	.2589
0.14	.0557	.4443	.3951	0.54	.2054	.2946	.3448	0.94	.3264	.1736	.2565
0.15	.0596	.4404	.3945	0.55	.2088	.2912	.3429	0.95	.3289	.1711	.2541
0.16	.0636	.4364	.3939	0.56	.2123	.2877	.3410	0.96	.3315	.1685	.2516
0.17	.0675	.4325	.3932	0.57	.2157	.2843	.3391	0.97	.3340	.1660	.2492
0.18	.0714	.4286	.3925	0.58	.2190	.2810	.3372	0.98	.3365	.1635	.2468
0.19	.0753	.4247	.3918	0.59	.2224	.2776	.3352	0.99	.3389	.1611	.2444
0.20	.0793	.4207	.3910	0.60	.2257	.2743	.3332	1.00	.3413	.1587	.2420
0.21	.0832	.4168	.3902	0.61	.2291	.2709	.3312	1.01	.3438	.1562	.2396
0.22	.0871	.4129	.3894	0.62	.2324	.2676	.3292	1.02	.3461	.1539	.2371
0.23	.0910	.4090	.3885	0.63	.2357	.2643	.3271	1.03	.3485	.1515	.2347
0.24	.0948	.4052	.3876	0.64	.2389	.2611	.3251	1.04	.3508	.1492	.2323
0.25	.0987	.4013	.3867	0.65	.2422	.2578	.3230	1.05	.3531	.1469	.2299
0.26	.1026	.3974	.3857	0.66	.2454	.2546	.3209	1.06	.3554	.1446	.2275
0.27	.1064	.3936	.3847	0.67	.2486	.2514	.3187	1.07	.3577	.1423	.2251
0.28	.1103	.3897	.3836	0.68	.2517	.2483	.3166	1.08	.3599	.1401	.2227
0.29	.1141	.3859	.3825	0.69	.2549	.2451	.3144	1.09	.3627	.1379	.2203
0.30	.1179	.3821	.3814	0.70	.2580	.2420	.3123	1.10	.3643	.1357	.2179
0.31	.1217	.3783	.3802	0.71	.2611	.2389	.3101	1.11	.3665	.1335	.2155
0.32	.1255	.3745	.3790	0.72	.2642	.2358	.3079	1.12	.3686	.1314	.2131
0.33	.1293	.3707	.3778	0.73	.2673	.2327	.3056	1.13	.3708	.1292	.2107
0.34	.1331	.3669	.3765	0.74	.2704	.2296	.3034	1.14	.3729	.1271	.2083
0.35	.1368	.3632	.3752	0.75	.2734	.2266	.3011	1.15	.3749	.1251	.2059
0.36	.1406	.3594	.3739	0.76	.2764	.2236	.2989	1.16	.3770	.1230	.2036
0.37	.1443	.3557	.3725	0.77	.2794	.2206	.2966	1.17	.3790	.1210	.2012
0.38	.1480	.3520	.3712	0.78	.2823	.2177	.2943	1.18	.3810	.1190	.1989
0.39	.1517	.3483	.3697	0.79	.2852	.2148	.2920	1.19	.3830	.1170	.1965

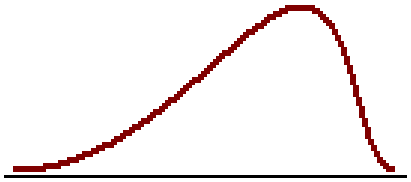
Source: Taken from Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed., 1974. Published by Longman Group Ltd., London (previously published by Oliver and Boyd, Edinburgh), and by permission of the authors and publishers.

Μετατροπή z τιμών

- Οι z τιμές μπορούν εύκολα να μετατραπούν με τη σειρά τους σε κατανομή με συγκεκριμένο μέσο όρο και τυπική απόκλιση πολλαπλασιάζοντάς τες με συγκεκριμένη τυπική απόκλιση και προσθέτοντας τον επιθυμητό μέσο όρο.
- X'= (s')(z) + X
- Αυτό είναι επιθυμητό γιατί πολλές φορές οι z τιμές είναι δύσκολο να εξηγηθούν και να παρουσιαστούν. Παρουσιάζουν αρνητικό πρόσημο και χρησιμοποιούν δεκαδικά ψηφία γεγονός που δυσκολεύει την παρουσίασή τους. Για το λόγο αυτό μπορούμε να τους μετασχηματίσουμε σε διαφορετικές κατανομές με διαφορετικό μέσο όρο και διαφορετική τυπική απόκλιση. Δύο γνωστοί τέτοιοι μετασχηματισμοί είναι οι ακόλουθοι:
- T τιμές: Μέσος όρος 50 και τυπική απόκλιση 10. Έτσι αποκλείεται η ύπαρξη αρνητικών τιμών και δεκαδικών ψηφίων.
- IQ τιμές: Μέσος όρος 100 και τυπική απόκλιση 15. Τα διάφορα τεστ ευφυΐας μετατρέπουν τις τιμές τους σε κλίμακα IQ. Έτσι το σύνολο του πληθυσμού κατατάσσεται στην κλίμακα IQ μεταξύ των τιμών 55 – 145 (± 3 τυπικές αποκλίσεις).

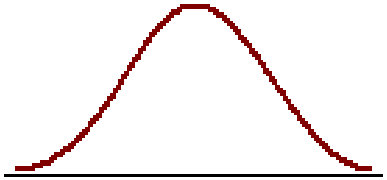
Περιγραφικές τιμές της κατανομής Κλίση

- Κλίση (skewness): Μας λέει πόσο «στραβή» είναι η κατανομή.



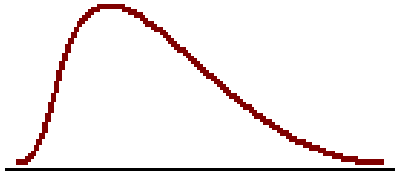
Αρνητικά κεκλιμένη κατανομή

Κλίση < 0



Κανονική κατανομή

Κλίση = 0

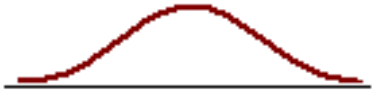


Θετικά κεκλιμένη κατανομή

Κλίση > 0

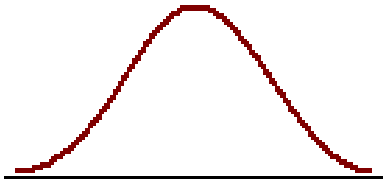
Περιγραφικές τιμές της κατανομής
Κύρτωση

- Κύρτωση (kyrtosis): Μας λέει πόσο «οξύ» είναι το σχήμα της κατανομής.



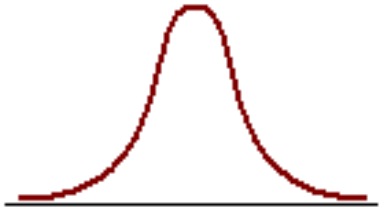
Κατανομή πλατύκυρτη

Κύρτωση < 0



Κανονική κατανομή

Κύρτωση = 0



Κατανομή λεπτόκυρτη

Κύρτωση > 0

Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την 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.

Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων

"Η δομή και οργάνωση της παρουσίασης, καθώς και το υπόλοιπο περιεχόμενο, αποτελούν πνευματική ιδιοκτησία του συγγραφέα και του Πανεπιστημίου Αθηνών και διατίθενται με άδεια Creative Commons Αναφορά Μη Εμπορική Χρήση Παρόμοια Διανομή Έκδοση 4.0 ή μεταγενέστερη.

Οι εικόνες/σχήματα/διαγράμματα/φωτογραφίες που περιέχονται στην παρουσίαση αποτελούν πνευματική ιδιοκτησία τρίτων. Απαγορεύεται η αναπαραγωγή, αναδημοσίευση και διάθεσή τους στο κοινό με οποιονδήποτε τρόπο χωρίς τη λήψη άδειας από τους δικαιούχους. "