

# ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ

Γιαννούλα Φλώρου

Καθηγήτρια

Τμήμα Λογιστικής & Χρηματοοικονομικής

Δημοκρίτειο Πανεπιστήμιο Θράκης

Μεταπτυχιακό Πρόγραμμα

Λογιστική Ελεγκτική

# Παρουσίαση ύλης

- Σκοπός μαθήματος
- Στόχοι μαθήματος
- Οργάνωση μαθήματος
  
- Διδασκαλία (ανά εβδομάδα)
- Διδακτικά βοηθήματα
- Αξιολόγηση

# ΣΚΟΠΟΣ ΜΑΘΗΜΑΤΟΣ

- Ανάπτυξη και συγγραφή ερευνητικής επιστημονικής εργασίας
- Εξαγωγή και γραπτή παρουσίαση συμπερασμάτων που προκύπτουν από πραγματικά δεδομένα
- Μέθοδοι επεξεργασίας ποιοτικών και ποσοτικών δεδομένων με χρήση στατιστικού λογισμικού σε ηλεκτρονικό υπολογιστή

# ΣΤΟΧΟΙ ΜΑΘΗΜΑΤΟΣ

## μαθησιακά αποτελέσματα

- Κατανόηση δομής και τρόπου συγγραφής επιστημονικής εργασίας (Βιβλιογραφική ανασκόπηση, εύρεση βιβλιογραφικών πηγών, παράθεση παραπομπών)
- Σχεδιασμός και πραγματοποίηση επιστημονικής έρευνας με χρήση πραγματικών δεδομένων (συλλογή δεδομένων, καταγραφή τους, επεξεργασία και εξαγωγή συμπερασμάτων)
- Επιλογή θέματος διπλωματικής εργασίας (προετοιμασία δομής, σχεδιασμός ερωτηματολογίου, διατύπωση ερευνητικών υποθέσεων)
- Εκμάθηση στατιστικών μεθόδων μέσω του στατιστικού λογισμικού SPSS (καταχώρηση δεδομένων, μέθοδοι παρουσίασης και επεξεργασίας, τεχνικές ανάλυσης πολυμεταβλητών δεδομένων)

# ΟΡΓΑΝΩΣΗ ΜΑΘΗΜΑΤΟΣ

- Επιστημονική έρευνα
  - Ορισμός, διακρίσεις, στάδια
- Δεοντολογία επιστημονικής έρευνας
  - Ορισμός, ηθική, μέτρηση εγκυρότητας, αξιοπιστίας
- Συγγραφή επιστημονικής εργασίας
  - Δομή, περιεχόμενο, αναφορές, παρουσίαση
- Ανάλυση ποιοτικών και ποσοτικών δεδομένων
  - Συλλογή, επεξεργασία, παρουσίαση
- Οικονομετρία
  - Ανάλυση οικονομικών δεδομένων, προβλέψεις
- Προετοιμασία δομής διπλωματικής εργασίας

# Εβδομαδιαία κατανομή της ύλης

- 13 τρίωρες ενότητες
  - 6 ενότητες με θεωρία κι ανάλυση στατιστικών δεδομένων
  - 3 ενότητες με θεωρία κι ανάλυση οικονομετρικών δεδομένων
  - 2 εργαστηριακές ενότητες με εκμάθηση στατιστικού λογισμικού SPSS
  - 2 θεωρητικές ενότητες προετοιμασίας διπλωματικής εργασίας

# Διδακτικά βοηθήματα

- Κύριο σύγγραμμα σχετικό με τη στατιστική επεξεργασία
  - Keller, G. (2010). Στατιστική για Οικονομικά & Διοίκηση Επιχειρήσεων. Αθήνα: Επίκεντρο.

## ΗΛΕΚΤΡΟΝΙΚΟ ΒΙΒΛΙΟ

- Μεθοδολογία έρευνας και εισαγωγή στη Στατιστική Ανάλυση Δεδομένων με το IBM SPSS STATISTICS-  
<https://repository.kallipos.gr/handle/11419/5075>
- Βοηθητικά συγγράμματα για στατιστική επεξεργασία
  - Δημητριάδης, Ε. (2017). Στατιστική επιχειρήσεων με εφαρμογές σε SPSS και LISREL, Αθήνα: Κριτική.
  - Berenson L. Mark, Levine M. David, Szabat A. Kathryn (2018). Βασικές Αρχές Στατιστικής για Επιχειρήσεις-Έννοιες και Εφαρμογές, Broken Hill.
- Βοηθήματα για επιστημονική εργασία
  - Saunders, M., Lewis, P. Thornhill, A. (2014). *Μέθοδοι έρευνας στις επιχειρήσεις & την οικονομία*. Αθήνα: Δισίγμα.
  - Ρόντος, Κ. & Παπάνης, Ε. (2006). *Στατιστική έρευνα*. Αθήνα: Ι Σιδέρης.
  - Howard, K., Sharp, J., (1996). *Η επιστημονική μελέτη*. Αθήνα: Gutenberg.

# Βοηθητικό υλικό του μαθήματος

- Σημειώσεις, διαφάνειες, παραδείγματα, εργασίες, ενδεικτική βιβλιογραφία, θα βρίσκονται ηλεκτρονικά στην ιστοσελίδα την ηλεκτρονικής εκπαίδευσης του μεταπτυχιακού προγράμματος.
- Απορίες – παρατηρήσεις – προτάσεις – γενικότερη επικοινωνία στο email

[gflorou@af.duth.gr](mailto:gflorou@af.duth.gr)



# Αξιολόγηση φοιτητών

Συμμετοχή των φοιτητών σε εργασίες με χρήση του στατιστικού προγράμματος SPSS.

Εργασία (coursework) 20% και γραπτές εξετάσεις (exams) 80%.

# περιγραφικά μέτρα – κατανομές – εκτίμηση

Μεταπτυχιακό Πρόγραμμα  
Λογιστική Ελεγκτική

# Περιεχόμενα

- Είδη μεταβλητών
- Περιγραφικά χαρακτηριστικά
- Κατανομές
- Εκτίμηση με Διάστημα εμπιστοσύνης

# Είδη μεταβλητών

- Ποσοτικές (αντιστοιχούν σε ποσότητα)
  - Συνεχείς (Δεκαδικές τιμές)
  - Διακριτές (ακέραιες τιμές)
- Ποιοτικές (δεν αντιστοιχούν σε ποσότητα)
  - Κατηγορικές (nominal)
  - Διάταξης (Ordinal)
  - Διχοτομικές (0-1)

# Κλίμακες μέτρησης

- Κλίμακα λόγου ή αναλογίας (ratio scale)
  - Ποσοτικές μετρήσεις που μπορούν να διαιρεθούν μεταξύ τους. Αντιστοιχούν σε πραγματική μονάδα μέτρησης
- Κλίμακα ιεράρχησης (ordinal scale)
  - Βαθμολογίες, μετρήσεις που μπορούν να ταξινομηθούν από τη μικρότερη στη μεγαλύτερη αλλά δεν έχει νόημα να διαιρεθούν ή να αφαιρεθούν μεταξύ τους. Αντιστοιχούν σε «αυθαίρετες» μονάδα μέτρησης
- Κλίμακα διαστήματος (interval scale)
  - Βαθμολογίες, μετρήσεις που μπορούν να ταξινομηθούν από τη μικρότερη στη μεγαλύτερη, έχει νόημα να αφαιρεθούν μεταξύ τους αλλά δεν έχει νόημα να διαιρεθούν ή μεταξύ τους. Πχ. Η θερμοκρασία.

# Ποιοτικές μεταβλητές

Καταγραφή απαντήσεων ποιοτικών μεταβλητών

- Καταγράφονται συνήθως με λέξεις ή εκφράσεις αλλά είναι δύσκολο να ομαδοποιήσουμε παρόμοιες απαντήσεις και να παρουσιάσουμε αποτελέσματα

Κωδικογράφηση

- Για ευκολότερη καταγραφή σε ΗΥ δίνουμε σε κάθε λέξη (απάντηση) έναν κωδικό (συνήθως ακέραιο αριθμό).

# Διευκόλυνση απαντήσεων

- Σε ερωτηματολόγιο που αναφέρεται σε ποιοτική μεταβλητή συνήθως καταγράφουμε όλες τις πιθανές απαντήσεις και ζητάμε από τον ερωτώμενο να επιλέξει μία ή περισσότερες απαντήσεις.

Π.χ.

Οικογενειακή κατάσταση

ο άγαμος      ο έγγαμος      ο διαζευγμένος      ο χήρος

Γνώμη για τα νέα μέτρα

οδιαφωνώ απόλυτα    οδιαφωνώ    οδεν έχω γνώμη    οσυμφωνώ    οσυμφωνώ απόλυτα

# Προσοχή στην ερμηνεία κωδικών-α

- Κατηγορική ποιοτική μεταβλητή
  - Τμήμα φοίτησης
    - Λογιστική 1
    - Διοίκηση Επιχειρήσεων 2
    - Πληροφορική 3
    - Μηχανολογία 4
    - Ηλεκτρολογία 5
    - Τεχνολογία πετρελαίου 6
    - Κλπ 7
- Η διαφορά 5-1 είναι ίδια με τη διαφορά 5-4



# Προσοχή στην ερμηνεία κωδικών-β

- Ποιοτική μεταβλητή διάταξης
  - Γνώμη για το μνημόνιο
    - Διαφωνώ απόλυτα 1
    - Διαφωνώ 2
    - Ούτε συμφωνώ ούτε διαφωνώ 3
    - Συμφωνώ 4
    - Συμφωνώ απόλυτα 5
- Η διαφορά 5-1 είναι πολύ διαφορετική από τη διαφορά 5-4

# Μεταβλητές με πολλαπλές απαντήσεις (multiple response)

– Ξένη γλώσσα που γνωρίζετε

- Αγγλικά 1
- Γαλλικά 2
- Γερμανικά 3
- Ρώσικα 4
- Άλλη ..... 5

- Το ίδιο άτομο μπορεί να απαντήσει από 0 έως και 5 επιλογές απάντησης.
- Προσοχή στην επεξεργασία τους!!

# Περιγραφική παρουσίαση δεδομένων

Πίνακες

Διαγράμματα

Στατιστικές εκθέσεις

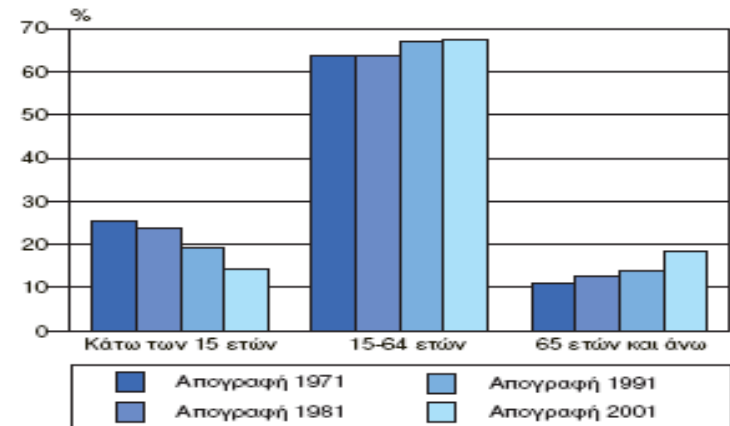
Υπολογισμός παραμέτρων

**Ι. ΠΛΗΘΥΣΜΟΣ**  
 Ια. Πραγματικός πληθυσμός της Ελλάδας κατά φύλο και ομάδες ηλικιών

	Απογραφές			
	1971(1)	1981(1)	1991(1)	2001(2)
<b>Σύνολο Ελλάδος</b> .....	<b>8.768.372</b>	<b>9.739.589</b>	<b>10.259.900</b>	<b>10.964.020</b>
Άρρενες .....	4.286.748	4.779.571	5.055.408	5.431.816
Θήλειες .....	4.481.624	4.960.018	5.204.492	5.532.204
<b>0 - 14 ετών</b> .....	<b>2.223.904</b>	<b>2.307.297</b>	<b>1.974.867</b>	<b>1.666.888</b>
<b>15 - 64 ετών</b> .....	<b>5.567.352</b>	<b>6.192.751</b>	<b>6.680.681</b>	<b>7.423.889</b>
<b>65 ετών και άνω</b> .....	<b>967.116</b>	<b>1.239.541</b>	<b>1.404.352</b>	<b>1.673.243</b>
<b>Ποσοστιαία κατανομή %</b>				
Άρρενες .....	48,89	49,07	49,27	49,54
Θήλειες .....	51,11	50,93	50,73	50,45
<b>0 - 14 ετών</b> .....	<b>25,36</b>	<b>23,69</b>	<b>19,25</b>	<b>15,20</b>
<b>15 - 64 ετών</b> .....	<b>63,72</b>	<b>63,58</b>	<b>67,06</b>	<b>67,71</b>
<b>65 ετών και άνω</b> .....	<b>10,92</b>	<b>12,73</b>	<b>13,69</b>	<b>17,08</b>

(1) Δειγματοληπτική επεξεργασία 25% και 10% των ερωτηματολογίων των απογραφών 1971 και 1981, αντίστοιχα, και καθολική επεξεργασία των ερωτηματολογίων της απογραφής 1991.  
 (2) Οριστικά στοιχεία για το σύνολο της Ελλάδας και προσωρινά κατά φύλο και ομάδες ηλικιών.

**Ποσοστιαία κατανομή του πληθυσμού κατά ομάδες ηλικιών**



# Περιγραφική παρουσίαση ποιοτικών (κατηγορικών) μεταβλητών

- Μία μεταβλητή μόνο
  - Πίνακας συχνοτήτων
  - Ραβδόγραμμα ή κυκλικό διάγραμμα
- Δύο ποιοτικές μεταβλητές
  - Πίνακας συχνοτήτων διπλής εισόδου
  - Ραβδόγραμμα σωρευτικό ή πολλαπλό
- Πολλές ποιοτικές μεταβλητές μαζί  
Παραγοντική ανάλυση, ιεραρχική ταξινόμηση

# Περιγραφική παρουσίαση ποσοτικών (συνεχών) μεταβλητών

- Μία μεταβλητή μόνο
  - Παράμετροι θέσης διασποράς
  - Ιστόγραμμα
- Δύο ποσοτικές μεταβλητές
  - Διάγραμμα διασποράς
  - Συσχέτιση, απλή παλινδρόμηση
- Πολλές ποσοτικές μεταβλητές μαζί
  - Παλινδρόμηση πολλαπλή, μερική συσχέτιση

# Περιγραφική παρουσίαση ποσοτικής και ποιοτικής μεταβλητής

- Μία ποσοτική να παρουσιαστεί σε σχέση με μία ποιοτική-κατηγορική μεταβλητή
  - Υπολογισμός μέτρων θέσεως για κάθε κατηγορία
  - Θηκόγραμμα
- Παράδειγμα εισόδημα και φύλο
  - Υπολογισμός ξεχωριστά για άνδρες, ξεχωριστά για γυναίκες
  - Παρουσίαση θηκογράμματος για άνδρες και γυναίκες
- Σύγκριση - έλεγχος ισότητας των μέτρων θέσεως μεταξύ των κατηγοριών της ποιοτικής  
σχολιασμός διαφορών, στατιστικοί έλεγχοι

# «Παράμετροι» ποσοτικών δεδομένων

- Μας επιτρέπουν να εκφράζουμε μεγάλο πλήθος δεδομένων με **μια** αντιπροσωπευτική τιμή, ώστε στη συνέχεια να μπορούμε ευκολότερα να κάνουμε συγκρίσεις.
- Ανάλογα με το τι θέλουμε να αντιπροσωπεύει η τιμή αυτή, μπορούμε να υπολογίσουμε παραμέτρους που αντιπροσωπεύουν :
  - την θέση (σε μια κλίμακα μετρήσεων) των ποσοτικών δεδομένων και οι οποίες ονομάζονται μέτρα θέσεως, ή
  - το εύρος διασποράς των τιμών των ποσοτικών δεδομένων οι οποίες ονομάζονται μέτρα διασποράς ή
  - ακόμη το σχήμα του διαγράμματος που περιγράφει όλα τα ποσοτικά δεδομένα κι οι οποίες ονομάζονται μέτρα ασυμμετρίας και κύρτωσης.

# Παράμετροι κεντρικής θέσης

- Αριθμητικός μέσος
  - Υπολογίζεται αθροίζοντας όλα τα δεδομένα και διαιρώντας με το πλήθος τους.
- Διάμεσος
  - Είναι δηλαδή η τιμή κάτω από την οποία βρίσκεται το 50% των δεδομένων και πάνω από την οποία βρίσκεται το υπόλοιπο 50%.
- Τεταρτημόρια
  - Το πρώτο τεταρτημόριο αντιστοιχεί στην τιμή κάτω από την οποία βρίσκεται το 25% των δεδομένων και πάνω από την οποία βρίσκεται το υπόλοιπο 75%.
  - Το τρίτο τεταρτημόριο αντιστοιχεί στην τιμή κάτω από την οποία βρίσκεται το 75% των δεδομένων και πάνω από την οποία βρίσκεται το υπόλοιπο 25%.



# Παράμετροι διασποράς

- Διακύμανση
  - Είναι η τιμή που μετράει πόσο απέχουν τα δεδομένα από τον αριθμητικό μέσο.
  - Υπολογίζεται αθροίζοντας τα τετράγωνα των διαφορών όλων των τιμών από τον αριθμητικό μέσο
- Τυπική απόκλιση
  - Είναι η τιμή που υπολογίζεται από τη διακύμανση αν πάρουμε την τετραγωνική ρίζα της διακύμανσης.

# παράδειγμα

A κατηγο	B κατηγο
724	588
774	673
794	668
750	685
727	790
785	653
719	633
630	568
774	691
718	620
761	723
763	519
767	705
809	549
697	565
847	639
787	584
753	436

- Παρουσιάστε σύντομα τα δεδομένα μοριοδότησης δανειζόμενων
- Υπάρχει διαφορά στις δύο κατηγορίες;
- Ποια έχει μεγαλύτερη ανομοιογένεια;

# Παράμετροι παραδείγματος

	Στήλη A	Στήλη B
Μέσος	754,3888	627,16666
Τυπικό σφάλμα	11,32010	19,716495
Διάμεσος	762	636
Επικρατούσα τιμή	774	#Δ/Υ
Μέση απόκλιση τετραγώνου	48,02712	83,650006
Διακύμανση	2306,604	6997,3235
Κύρτωση	1,779535	0,5001884
Ασυμμετρία	-0,68943	-0,340468
Εύρος	217	354
Ελάχιστο	630	436
Μέγιστο	847	790
Άθροισμα	13579	11289
Πλήθος	18	18

Υπάρχει διαφορά στη θέση των δύο κατηγοριών;

Υπάρχει διαφορά στη μεταβλητότητα;

*Από την εμπειρία στη θεωρία ...*

*Και από τη θεωρία στην εφαρμογή .....*

# Κατανομές πιθανοτήτων

- Εμπειρικές  
Προκύπτουν από την παρατήρηση-καταγραφή πραγματικών δεδομένων  
παρουσιάζονται με πίνακα ή ιστόγραμμα
- Θεωρητικές
  - Υπολογίζονται θεωρητικά και «προσομοιάζουν» ή «αντικαθιστούν» τις άγνωστες εμπειρικές κατανομές
  - Παρουσιάζονται με συνάρτηση και γραφική της παράσταση

# Θεωρητικές κατανομές

Συνεχής μεταβλητή (ποσοτικά δεδομένα)

- Κανονική κατανομή
- Κατανομή  $t$
- Κατανομή  $\chi^2$
- Κατανομή  $F$
- Εκθετική κατανομή

Ασυνεχής μεταβλητή (ποιοτικά – ακέραια δεδομένα)

- Κατανομή διωνυμική
- Κατανομή Poisson

# Κανονική κατανομή

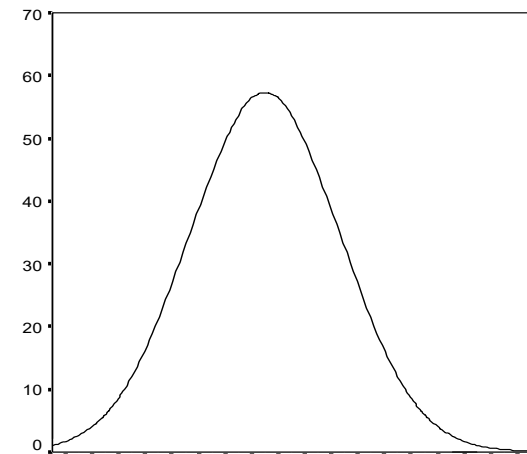
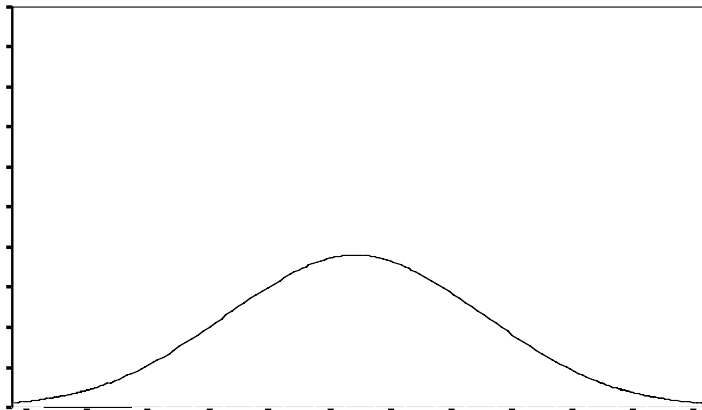
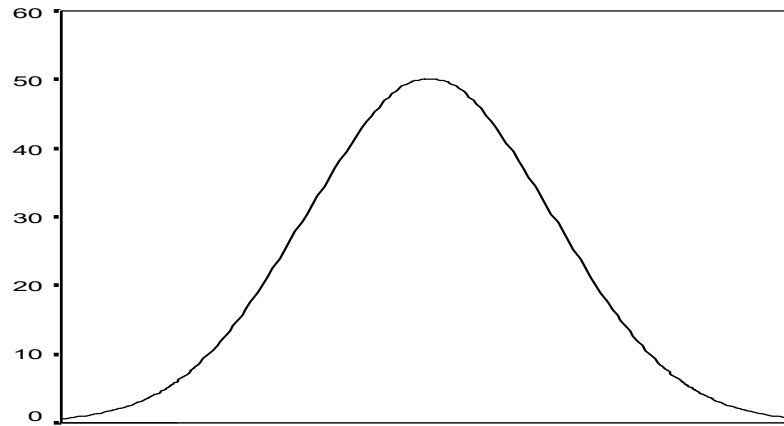
Ονομάζεται και κατανομή Gauss ή «φυσική» κατανομή ή κατανομή της «καμπάνας»

Εφαρμόζεται σε πάρα πολλές περιπτώσεις που αφορούν μετρήσεις της φύσης (ύψος, βάρος, μήκος, πλάτος, κλπ.)

Όταν δεν ξέρουμε το είδος της κατανομής, συνήθως υποθέτουμε κανονική κατανομή.

Είναι συμμετρική γύρω από την αναμενόμενη τιμή της (μέσος όρος) και η καμπύλη της έχει το σχήμα της καμπάνας.

# Σχήμα κανονικής κατανομής

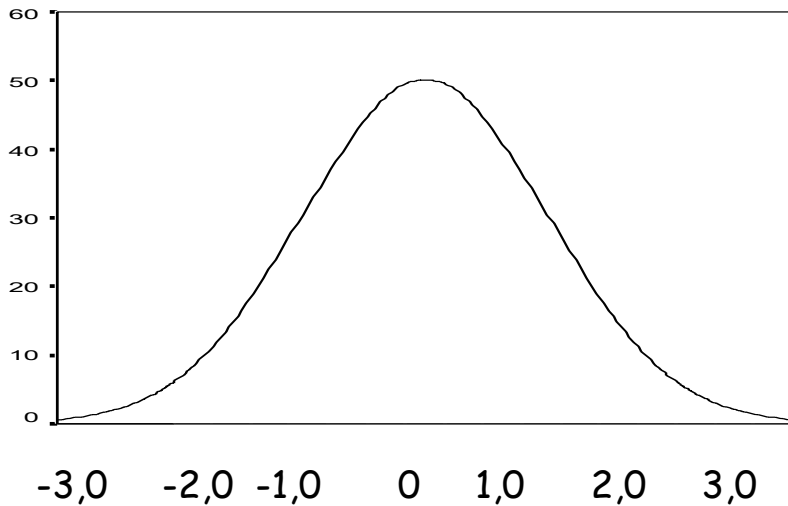




# Χαρακτηριστικά κανονικής κατανομής

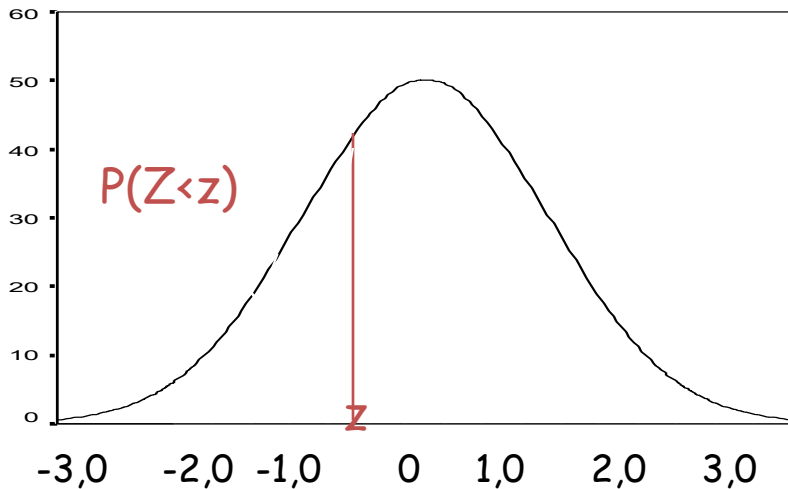
- Μέση τιμή  $\mu$ .
- Διακύμανση  $\sigma^2$ .
- Συμβολίζεται με  $N(\mu, \sigma)$
- $X \sim N(\mu, \sigma)$ ,  
όταν  $E(X) = \mu$  και  $\text{Var}(X) = \sigma^2$

# Τυπική κανονική κατανομή (κατανομή Z)



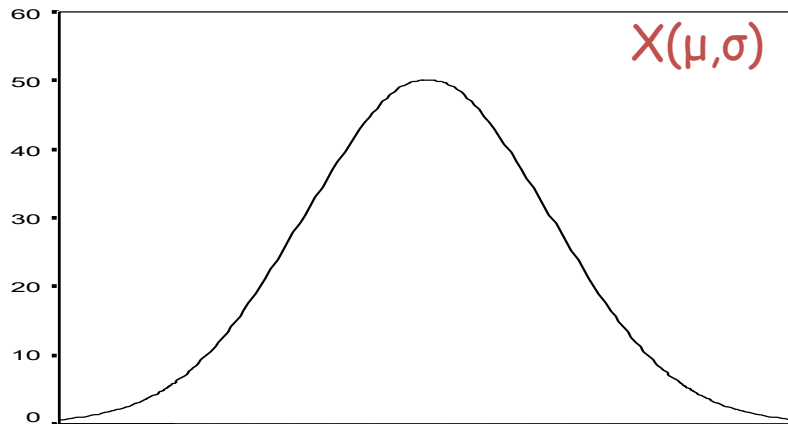
- Μέση τιμή  $\mu=0$ .
- Διακύμανση  $\sigma^2=1$ .
- Συμβολίζεται με  $N(0,1)$
- $Z \sim N(0,1)$ , όταν  $E(Z)=0$   
και  $\text{Var}(Z)=1$
- Παίρνει τιμές από -3 έως +3

# Υπολογισμός πιθανοτήτων τυπικής κανονικής κατανομής

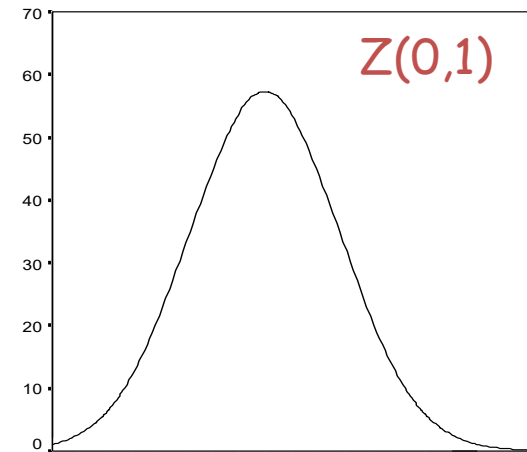


- $P(Z < 0) = 0,5 = 50\%$   
 $P(Z > 0) = 0,5 = 50\%$ .
- $P(-1 < Z < +1) = 0,675$
- $P(-2 < Z < +2) = 0,95$
- $P(-3 < Z < +3) = 0,995$
  
- Πίνακας με υπολογισμένες όλες τις πιθανότητες  $P(Z < z)$

# Μετατροπή κανονικής κατανομής $X(\mu, \sigma)$ σε τυπική κανονική κατανομή $Z(0,1)$



$$Z = \frac{X - \mu}{\sigma}$$



# Υπολογισμός τιμής όταν γνωρίζουμε την πιθανότητα

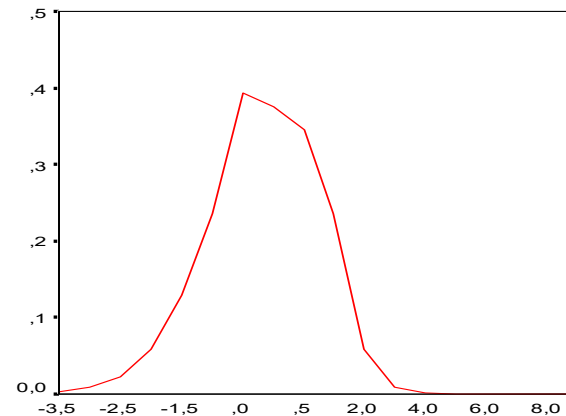
- $P(X < \delta) = 0,95$  Να βρεθεί το  $\delta$
- $P(Z < (\delta - \mu) / \sigma) = 0,95$
- Από τον πίνακα με υπολογισμένες πιθανότητες της  $Z$  κατανομής:
- $P(Z < 1,96) = 0,95$
- Άρα  $(\delta - \mu) / \sigma = 1,96$  και  $\delta = 1,96 * \sigma + \mu$

# t κατανομή

Εφαρμόζεται σε πάρα πολλές περιπτώσεις στην πράξη όταν δεν γνωρίζουμε τη διακύμανση της κανονικής κατανομής

Είναι συμμετρική γύρω από το μηδέν, η καμπύλη της έχει το σχήμα της καμπάνας αλλά το άπλωμα της αλλάζει ανάλογα με τους «βαθμούς ελευθερίας».

Δηλαδή έχει διαφορετικό άπλωμα για  $\beta.ε=2$ . διαφορετικό για  $\beta.ε=10$  και ταυτίζεται με την κανονική κατανομή για  $\beta.ε.\geq 30$ .

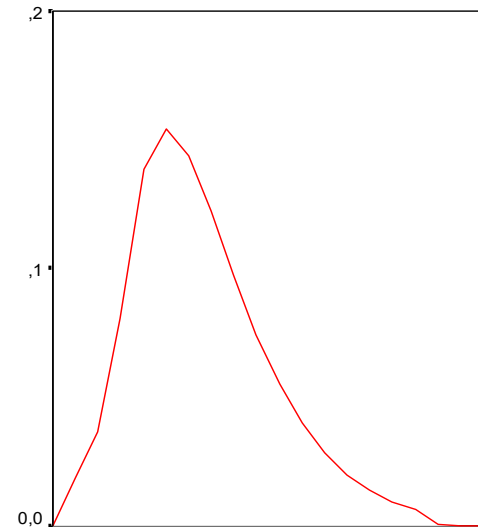


# $\chi^2$ κατανομή

Εφαρμόζεται στις περιπτώσεις ελέγχου  $c$  και καλής προσαρμογής

Προκύπτει από την ύψωση στο «τετράγυο» περισσότερων κανονικών κατανομών (περισσότερους «βαθμούς ελευθερίας», β.ε.= $n$ ).

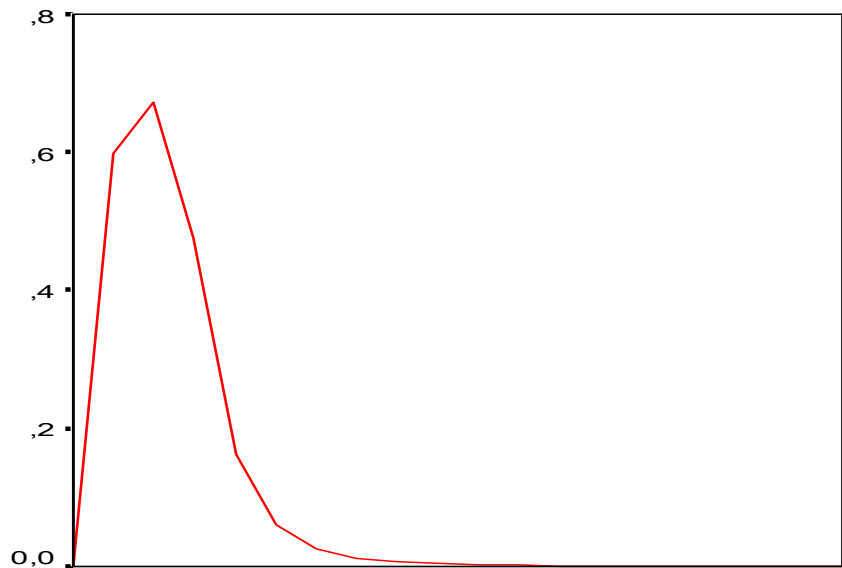
Έχει πάντα θετικές τιμές συνήθως μικρές τιμές της είναι  $E(X)=n$  (β.ε.) και η διακύμανσή της είναι  $Var(X)=2n$ .



# F κατανομή

Εφαρμόζεται στις περιπτώσεις ελέγχου  
ανάλυσης διακύμανσης.

Προκύπτει από την διαίρεση δύο κατανομών  $\chi^2$   
(ανάλογα με «βαθμούς ελευθερίας», αριθμητή= $n_1$  και  
παρανομαστή= $n_2$  ).

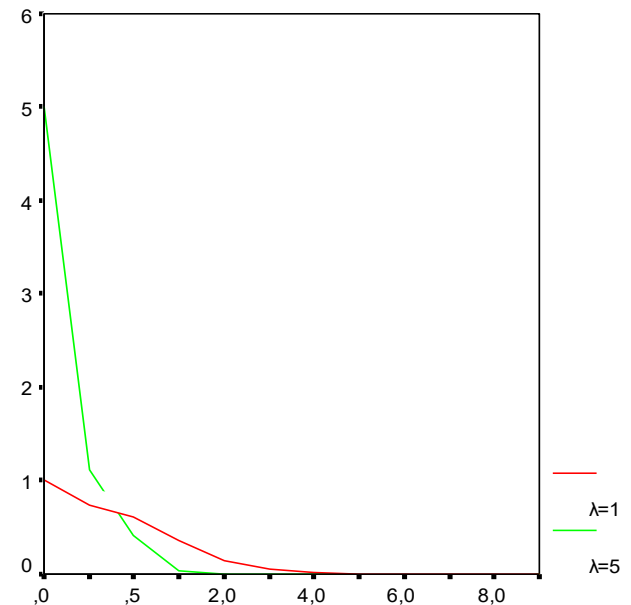
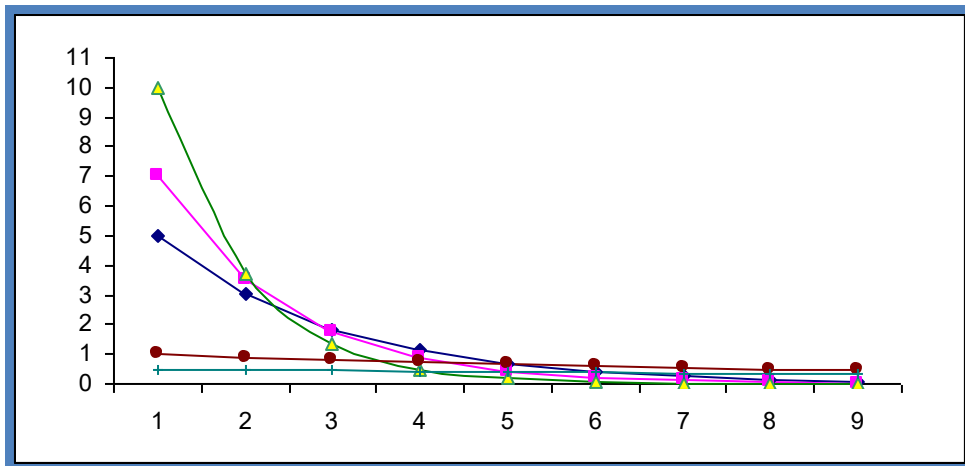




# εκθετική κατανομή

Εφαρμόζεται σε περιπτώσεις στην πράξη όταν η τιμή μιας τυχαίας μεταβλητής, ξεκινάει από μια αρχική τιμή  $\lambda > 0$  και στη συνέχεια μειώνεται μέχρι να μηδενιστεί.

Έχει πάντα θετικές τιμές, η μέση τιμή της είναι  $E(X) = 1/\lambda$ , η τυπική απόκλιση  $\sigma = 1/\lambda$  και η διακύμανσή της είναι  $\text{Var}(X) = 1/\lambda^2$ .



# Θεωρία $\rightarrow \rightarrow$ εφαρμογή

- Κάθε θεωρητική κατανομή θεωρείται γνωστή αν γνωρίζουμε τις «παραμέτρους» της δηλαδή τα βασικά στοιχεία της, που συνήθως αντιστοιχούν στη μέση τιμή ή και στην τυπική απόκλιση
- Αφού είναι «γνωστή» προχωρούμε στους υπολογισμούς βασιζόμενοι σε κάποιον τύπο ή σε κάποια ήδη υπολογισμένη τιμή και κάνουμε εφαρμογή στα δεδομένα μας.

# Πως καταλαβαίνουμε ποια κατανομή ακολουθούν τα δεδομένα;

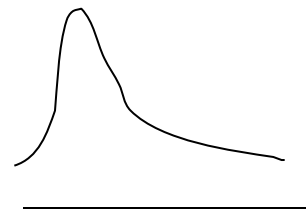
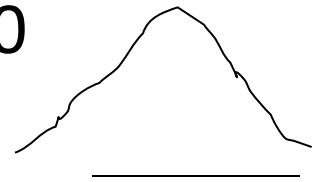
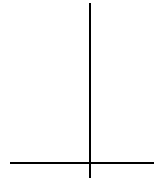
- Από το τι μετράνε (τι καταγράφουν).
- Από την αντίστοιχη θεωρία.
- Από το σχήμα του ιστογράμματος αν είναι πολλά.
- Από υπόθεση που κάνουμε.
- Από παλαιότερα δεδομένα για την ίδια μεταβλητή.

# Σε τι χρησιμεύουν οι κατανομές;

- Γνωρίζοντας μόνο τις παραμέτρους μόνο θέσης και διασποράς έχουμε μια «γενική ιδέα» για μια μεταβλητή, δηλαδή μόνο τη φανταζόμαστε.
- Γνωρίζοντας την κατανομή της μεταβλητής, έχουμε μια πλήρη εικόνα της, δηλαδή μπορούμε να βρούμε την πιθανότητα κάθε τιμή της.

# Παράδειγμα γνώσης

- Εισόδημα μηνιαίο φοιτητών με μέση τιμή 300 ευρώ και τυπική απόκλιση 50 ευρώ.
- Εισόδημα μηνιαίο φοιτητών με μέση τιμή 300 ευρώ και τυπική απόκλιση 50 ευρώ, που ακολουθεί κανονική κατανομή.
- Εισόδημα μηνιαίο φοιτητών με μέση τιμή 300 ευρώ και τυπική απόκλιση 50 ευρώ, που ακολουθεί  $\chi^2$  κατανομή.



# Εφαρμογή κανονικής κατανομής

- Ο αριθμός των πελατών που εισέρχονται σε ένα πολυκατάστημα ανά ώρα, στις ώρες αιχμής έχει κανονική κατανομή με μέσο 200 πελάτες και τυπική απόκλιση 50 πελάτες.
- Να υπολογισθεί η πιθανότητα σε μια συγκεκριμένη ώρα αιχμής να εισέλθουν λιγότεροι από 250 πελάτες

*Από το ειδικό στο γενικό ...*

*επαγωγή .....*

*Από το δείγμα στον πληθυσμό  
εκτίμηση ....*

# Δύο τρόποι εκτίμησης (επίτευξη στόχου)

- 1ος «Μπαμ και κάτω» «μια και έξω»
  - Πλεονέκτημα ταχύτητα
  - Μειονέκτημα κίνδυνος αποτυχίας
- 2ος «πλάγια – περικυκλωτική προσέγγιση»
  - Πλεονέκτημα σίγουρα αργά βήματα
  - Μειονέκτημα μεγαλύτερη προσπάθεια



# Δύο τρόποι εκτίμησης (επίτευξη στόχου)

- 1ος «Μπαμ και κάτω»
- **σημειακή εκτίμηση**
  - Πλεονέκτημα ταχύτητα
  - Μειονέκτημα κίνδυνος αποτυχίας
- 2ος «πλάγια – περικυκλωτική προσέγγιση»
- **Εκτίμηση διαστήματος**
  - Πλεονέκτημα σίγουρα αργά βήματα
  - Μειονέκτημα μεγαλύτερη προσπάθεια

# Εκτίμηση διαστήματος

- Είναι ένα διάστημα τιμών, στο οποίο αναμένουμε να βρίσκεται η ζητούμενη παράμετρος.
- Πότε δεν είμαστε **βέβαιοι** ότι η παράμετρος θα βρίσκεται στα σίγουρα στο διάστημα εμπιστοσύνης, εκτός και αν αυτό είναι το διάστημα  $[-\infty, +\infty]$
- Οπότε έχουμε στάθμη εμπιστοσύνης του διαστήματος η οποία μετρείται με την πιθανότητα να εκτιμούμε με το διάστημα τη ζητούμενη παράμετρο. (ή με την αντίστοιχη πιθανότητα να κάνουμε λάθος εκτίμηση)

# Βασικός τύπος διαστήματος εμπιστοσύνης

- Αν ξέρουμε τη εκτίμηση της παραμέτρου  $m$  και την τυπική της απόκλιση  $s$ , υποθέτοντας ότι ακολουθεί κανονική κατανομή,
- η τιμή της θα βρίσκεται μέσα στα όρια της κανονικής κατανομής,
- με πιθανότητα 90% ή 95% ή 99%, ανάλογα με τις τιμές των ορίων  $\times z_{\alpha/2}=1,645$  ή  $z_{\alpha/2}= 1,96$  ή  $z_{\alpha/2}= 2,58$
- Επομένως το εκτιμημένο διάστημα τιμών θα είναι  
$$[m - z_{\alpha/2} s, m + z_{\alpha/2} s]$$

# Αναλογία κατηγορίας πληθυσμού

- Στην περίπτωση που τα δεδομένα μας είναι ποιοτικά, δεν μπορούμε να εκτιμήσουμε μέτρα θέσεως ή διασποράς, παρά μόνο συχνότητα και αναλογία μιας κατηγορίας.
- Στον πληθυσμό από τον οποίο προέρχονται τα δεδομένα, συμβολίζουμε την αναλογία που μας ενδιαφέρει με το σύμβολο  $p$ .
- Η τιμή  $p$  είναι  $<1$  και η αναλογία να μην συμβεί αυτό που μας ενδιαφέρει συμβολίζεται  $1-p$ .

# Εκτίμηση αναλογίας

- Έστω  $x$  ο αριθμός των «επιτυχιών» για την κατηγορία που μας ενδιαφέρει στο δείγμα μεγέθους  $n$ .

$$\hat{p} = \frac{x}{n}$$

- Η εκτίμηση αυτή της αναλογίας του πληθυσμού έχει μέση τιμή  $p$  και τυπική απόκλιση  $\sqrt{p(1-p)/n}$  και ακολουθεί κανονική κατανομή.
- (με την προϋπόθεση  $np > 5$  και  $n(1-p) > 5$ )

# Διάστημα εμπιστοσύνης αναλογίας

- Το διάστημα εμπιστοσύνης για την αναλογία του πληθυσμού με στάθμη σημαντικότητας  $\alpha\%$  είναι:

$$\hat{p} \pm z_{\alpha/2} \sqrt{p(1-p)/n}$$

- Επειδή ο τύπος περιέχει την άγνωστη τιμή  $p$ , αντικαθιστάμε με την εκτιμημένη τιμή του.

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$$

# Διάστημα εμπιστοσύνης αναλογίας μικρού πληθυσμού

- Το διάστημα εμπιστοσύνης για την αναλογία του πληθυσμού  $N$  (όταν αυτός δεν είναι πολύ μεγάλος), αλλάζει αφού χρησιμοποιούμε ένα διορθωτικό παράγοντα για την εκτίμηση της τυπικής απόκλισης και γίνεται:

$$\hat{p} \pm z_{\alpha/2} \sqrt{p(1-p)/n} \sqrt{\frac{N-n}{N-1}}$$

# Διάστημα εμπιστοσύνης συνόλου επιτυχιών

- Το διάστημα εμπιστοσύνης για το σύνολο των «επιτυχιών» στον πληθυσμό  $N$  με στάθμη σημαντικότητας  $\alpha\%$  είναι:

$$N \left[ \hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p}) / n} \right]$$

- Όπου  $N$  αντιπροσωπεύει το μέγεθος του πληθυσμού. Δηλαδή πολλαπλασιάζουμε το άνω και κάτω όριο του διαστήματος με  $N$ .



# Εφαρμογή

(διάστημα εμπιστοσύνης αναλογίας)

- Ένας οργανισμός αστικών συγκοινωνιών, κατήργησε τα μηχανήματα έκδοσης εισιτηρίων μέσα στα οχήματα και οι επιβάτες πρέπει να έχουν το εισιτήριο τους πριν ανέβουν στο όχημα. Σε μια διαδρομή ελέγχθησαν 300 επιβάτες και βρέθηκαν 63 χωρίς εισιτήριο. Να εκτιμήσετε ένα 95% διάστημα εμπιστοσύνης για την αναλογία των επιβατών που επιβιβάζονται χωρίς εισιτήριο.
- Αν υποθέσουμε ότι το εισιτήριο στοιχίζει 1,20 ευρώ και κάθε μήνα μετακινούνται 10.000 επιβάτες, να εκτιμήσετε διάστημα εμπιστοσύνης 95% για τη ζημιά του οργανισμού.

# Λύση εφαρμογής

- Η εκτίμηση της αναλογίας από το δείγμα είναι  $63/300=0,21$ . Το διάστημα εμπιστοσύνης για την άγνωστη αναλογία του πληθυσμού όλων των επιβατών θα είναι:

$$\begin{aligned} & \left[ \hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right] = \\ & \left[ 0,21 \pm 1,96 \sqrt{(0,21(1-0,21)/300)} \right] = \\ & \left[ 0,21 \pm 0.0472 \right] = \left[ 0.1628, 0.2572 \right] \end{aligned}$$

- Δηλαδή θα είναι από  $16,28\%$  έως  $25,72\%$ .
- Συνολικά οι επιβάτες που μετακινούνται το μήνα χωρίς εισιτήριο θα είναι  $10.000 * [0,1628, 0,2572] = [1.628, 2.572]$ .
- Το διάστημα εμπιστοσύνης για τη ζημιά θα είναι
- $1,20 * [1.628, 2.572] = [1.953,6 \text{ έως } 3.086,4]$  ευρώ.

# Υπολογισμός διαστήματος εμπιστοσύνης για τη μέση τιμή πληθυσμού

- Εξαρτάται
  - Από το πλήθος των παρατηρήσεων που μετρήσαμε (μέγεθος δείγματος  $n$ )
  - Από την τυπική απόκλιση  $\sigma$
  - Από τη στάθμη εμπιστοσύνης  $\alpha$  που επιθυμούμε
- Τύπος

$$\left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

- Όπου  $z_{\alpha/2}$  είναι σταθερή τιμή που δίνεται από την κανονική κατανομή

# Υπολογισμός διαστήματος εμπιστοσύνης για τη μέση τιμή πληθυσμού

- Εξαρτάται
  - Από το πλήθος των παρατηρήσεων που μετρήσαμε (μέγεθος δείγματος  $n$ )
  - Από την τυπική απόκλιση  $s$
  - Από τη στάθμη εμπιστοσύνης  $\alpha$  που επιθυμούμε
- Τύπος

$$\left[ \bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

- Όπου  $t_{\alpha/2}$  είναι σταθερή τιμή που δίνεται από την  $t$  κατανομή

# Οικονομικός έλεγχος με χρήση διαστήματος εμπιστοσύνης

- Οικονομικός έλεγχος είναι ο έλεγχος που γίνεται για να εξασφαλισθεί η ορθότητα των αποτελεσμάτων της οικονομικής λογιστικής.
- Κάποιες εκτιμήσεις του οικονομικού ελέγχου αφορούν στο σύνολο (άθροισμα) μιας συνεχούς μεταβλητής. Π.χ. ετήσιο σύνολο δαπανών γευμάτων, σύνολο επισκευών αυτοκινήτων κ.α.
- Χρησιμοποιείται ο εκτιμητής του διαστήματος εμπιστοσύνης πολλαπλασιασμένος με  $N$  (πλήθος πληθυσμού). Αν το  $N$  σχετικά μικρό, χρησιμοποιούμε και τον διορθωτικό παράγοντα της τυπικής απόκλισης  $(N-n)/(N-1)$ .

$$N \left[ \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

# Εφαρμογή

(διάστημα εμπιστοσύνης μέσης τιμής)

- Ένας ελεγκτής πραγματοποίησε έλεγχο για να διαπιστώσει αν καταχωρούνται σωστά τα παραστατικά. Ελέχθησαν 96 παραστατικά από τα 866 και βρέθηκαν διαφορές με μέσο όρο 6,46 ευρώ και τυπική απόκλιση 17,58 ευρώ. Το 95% διάστημα εμπιστοσύνης για το σύνολο των διαφορών στις καταχωρήσεις είναι:

$$N \left[ \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right] = 866 \left[ 6.46 \pm 1.984 \frac{17.58}{\sqrt{96}} \sqrt{\frac{866-96}{866-1}} \right] =$$
$$[5.594.36 \pm 2.909.76] = [2.684,6 \acute{\epsilon}\omega\varsigma 8.504,12]$$

# Τυπολόγιο διάστημα εμπιστοσύνης

- Διάστημα εμπιστοσύνης αναλογίας

- Μεγάλο δείγμα

$$\left[ \tilde{p} - z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}, \tilde{p} + z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} \right]$$

- Διάστημα εμπιστοσύνης μέσης τιμής

- Μεγάλο δείγμα

$$\left[ \bar{X} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

- Μικρό δείγμα

$$\left[ \bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

- Διάστημα εμπιστοσύνης διαφοράς δύο μέσων τιμών

- Μεγάλα δείγματα

$$\left[ (\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} s, (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} s \right]$$

$$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- Μικρά δείγματα

$$\left[ (\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} s_t, (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} s_t \right]$$

$$s_t = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$
$$s = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

# Τιμές t κατανομής

B.ε.	$\alpha=10\%$	$\alpha=5\%$	$\alpha=2,5\%$	$\alpha=1,0\%$	$\alpha=0,5\%$
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
>=30	1,282	1,645	1,960	2,326	2,576
<b>za</b>	<b>1,282</b>	<b>1,645</b>	<b>1,960</b>	<b>2,326</b>	<b>2,576</b>



# ασκήσεις

1. Οι ελεγκτές μιας υπηρεσίας, διαπίστωσαν ότι 1450 συναλλαγές αφορούσαν αγορές με κατάτμηση διαγωνισμού. Επέλεξαν τυχαίο δείγμα 100 τέτοιων συναλλαγών και υπολόγισαν την αξία τους. Στο δείγμα η μέση τιμή 313,47 ευρώ ήταν και η τυπική απόκλιση 55,53 ευρώ.

Να εκτιμήσετε διάστημα εμπιστοσύνης 95% για την συνολική αξία των αγορών με κατάτμηση διαγωνισμού.

2. Όταν μια τράπεζα απαίτησε την πληρωμή δανείου της, η εταιρεία ανταλλακτικών αυτοκινήτων κήρυξε πτώχευση. Ο εκκαθαριστής βρήκε 73.544 ανταλλακτικά και σε έλεγχο που έκανε σε τυχαίο δείγμα 500 ανταλλακτικών βρήκε μέση αξία 229,17 ευρώ με τυπική απόκλιση 67,36 ευρώ.

Να εκτιμήσετε διάστημα εμπιστοσύνης 95% για την συνολική αξία των αποθεμάτων ανταλλακτικών.

3. Οικονομικός έλεγχος ενός προγράμματος αποζημιώσεων έγινε σε 118.653 εγκεκριμένες αιτήσεις. Επιλέχθηκε τυχαίο δείγμα 559 αιτήσεων και βρέθηκαν 29 που δεν έπρεπε να είχαν εγκριθεί. Να εκτιμήσετε διάστημα εμπιστοσύνης 99% για το σύνολο των αποζημιώσεων που δόθηκαν ενώ δεν θα έπρεπε να είχαν εγκριθεί.