ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΠΑΤΡΑΣ
Εργαστήριο Λήψης Αποφάσεων & Επιχειρησιακού Προγραμματισμού
Καθηγητής Ι. Μητρόπουλος

ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΕΙΔΙΚΑ ΘΕΜΑΤΑ

Κεφάλαιο 10

Εισαγωγή στην εκτιμητική

Είδαμε …

Κεφάλαιο 7 και 8: Διωνυμική, Poisson, Κανονική, και η Εκθετική κατανομή μας επιτρέψουν να κάνουμε δηλώσεις πιθανότητα για X (μέσος του πληθυσμού).

Για να γίνει αυτό χρειαζόμαστε τις παραμέτρους του πληθυσμού.

Binomial: p

Poisson: μ

Normal: μ and σ

Exponential: λ or μ

Είδαμε ...

Κεφάλαιο 9:

Κατανομές δειγματοληψίας μας επιτρέψουν να κάνουμε δηλώσεις πιθανότητας σχετικά με στατιστικά στοιχεία.

Χρειαζόμαστε τις παραμέτρους του πληθυσμού.

Μέσο δείγματος (Sample mean): μ and σ

Αναλογία δείγματος (Sample proportion): p

Διαφορά των μέσων δείγματος: μ_1, σ_1, μ_2, and σ_2

Είδαμε …

- **ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ:**

  Η κατανομή του δειγματικού μέσου σε τυχαίο δείγμα κάθε πληθυσμού ακολουθεί κατά προσέγγιση την κανονική κατανομή. Όσο μεγαλύτερο το μέγεθος του δείγματος τόσο περισσότερο η κατανομή του προσεγγίζει την κανονική κατανομή.

Αναλογίες (ποσοστά)

Κατανομή δειγματικού μέσου (χαρακτηριστικά):

$$\mu_{\bar{x}} = \mu$$

$$\sigma^2_{\bar{x}} = \sigma^2 / n \quad and \quad \sigma_{\bar{x}} = \sigma / \sqrt{n}$$

Η κατανομή του δειγματικού μέσου μπορεί να εκφραστεί ως:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

$$E(\hat{P}) = p$$

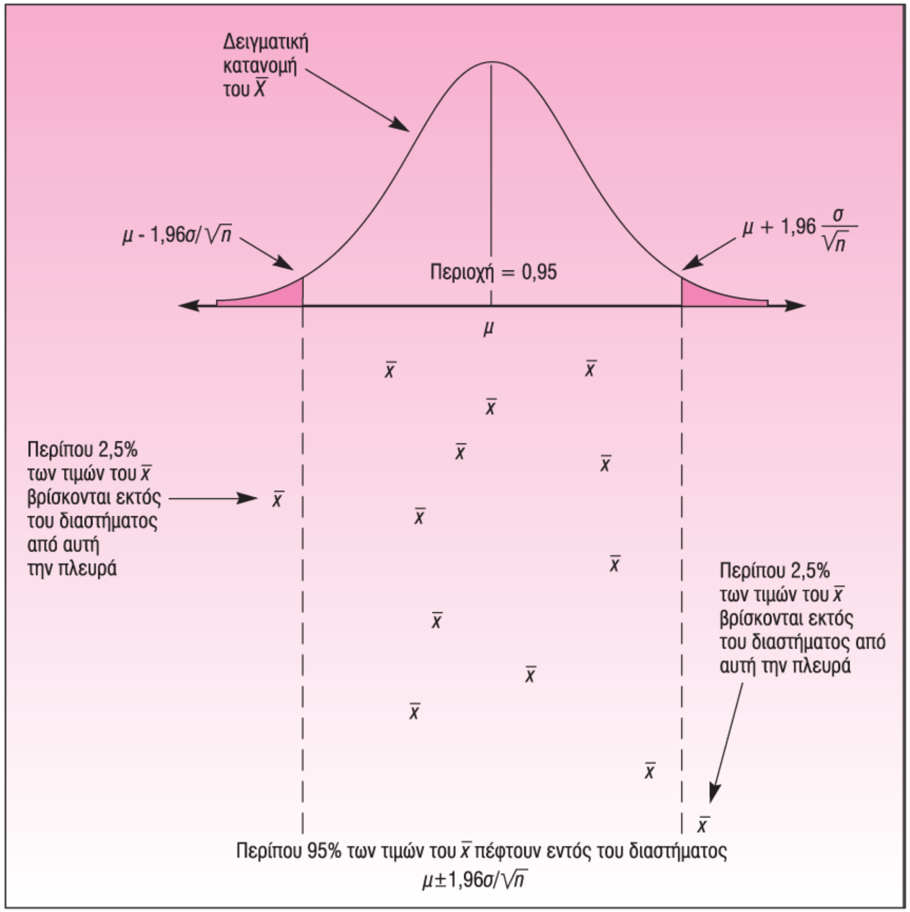
$$V(\hat{P}) = \sigma^2_{\hat{p}} = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$$

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}}$$

Που θα φτάσουμε…

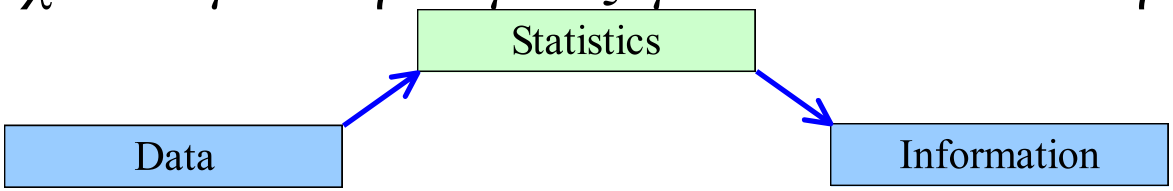
Ωστόσο, σε όλες σχεδόν τις ρεαλιστικές καταστάσεις πληθυσμιακοί παράμετροι είναι άγνωστοι. Θα χρησιμοποιήσουμε τη κατανομή δειγματοληψίας για να εξαγάγουμε συμπεράσματα σχετικά με τις άγνωστες πληθυσμιακές παραμέτρους.

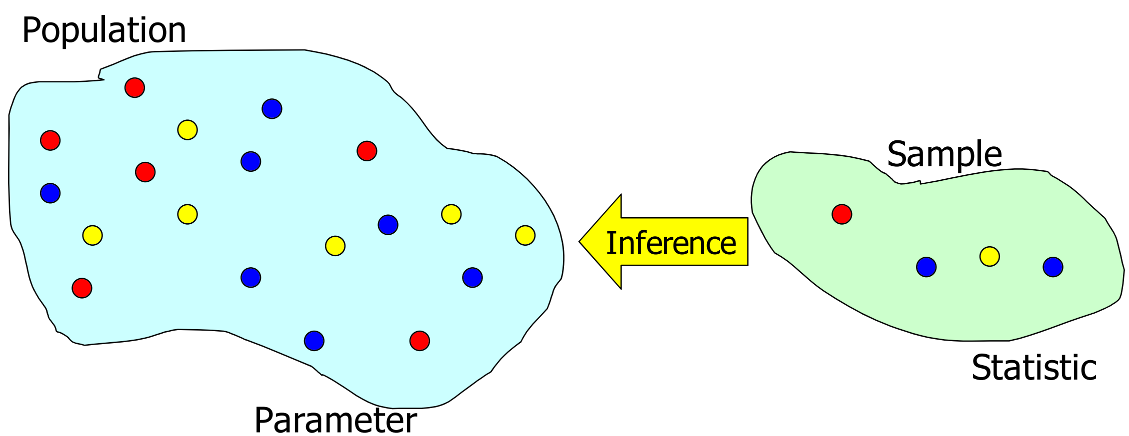


Διάγραμμα 6-1 Κατανομή πιθανοτήτων του $\bar{X}$ και κάποιες τιμές του στατιστικού σε επαναλαμβανόμενες δειγματοληψίες.

Statistical Inference…

Στατιστική Συμπερασματολογία (inference) *είναι η διαδικασία με την οποία συλλέγουμε πληροφορίες και εξάγουμε συμπεράσματα σχετικά με πληθυσμούς μέσα από κατάλληλα δείγματα.*





Προκειμένου να γίνει συμπερασματολογία, χρειαζόμαστε τις δεξιότητες & γνώση της Περιγραφικής στατιστικής, κατανομές πιθανότητας, και κατανομές δειγματοληψίας.

Εκτίμηση...

Υπάρχουν δύο τύποι συμπερασματολογίας: εκτίμηση και έλεγχος υποθέσεων. Η ***εκτίμηση*** εισάγεται πρώτα.

Ο στόχος της εκτίμησης είναι να καθοριστεί ***η κατά προσέγγιση τιμή*** μιας παραμέτρου του πληθυσμού με βάση ένα στατιστικό στοιχείο του δείγματος.

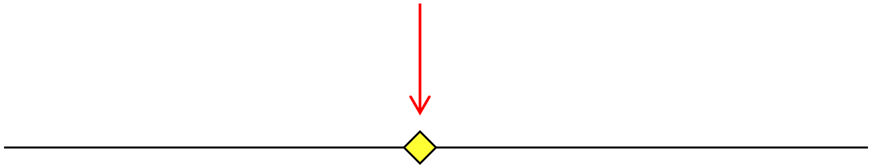
Π.χ. ο μέσος δείγματος($\bar{x}$) χρησιμοποιείται για την εκτίμηση (μ) του πληθυσμιακού μέσου.

Εκτίμηση ...

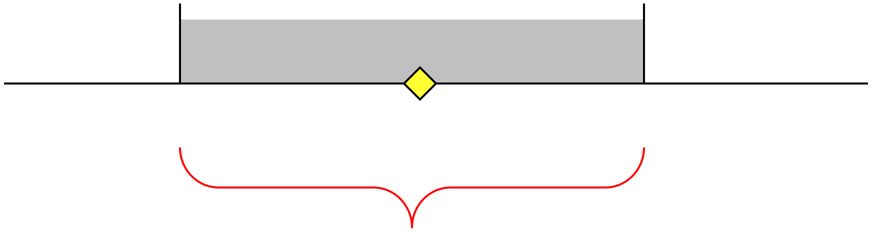
Ο στόχος της εκτίμησης είναι να καθοριστεί η κατά ***προσέγγιση τιμή*** μιας παραμέτρου του πληθυσμού με βάση ένα στατιστικό στοιχείο του δείγματος.

<u>Υπάρχουν δύο τύποι εκτιμητών:</u>

Point Estimator

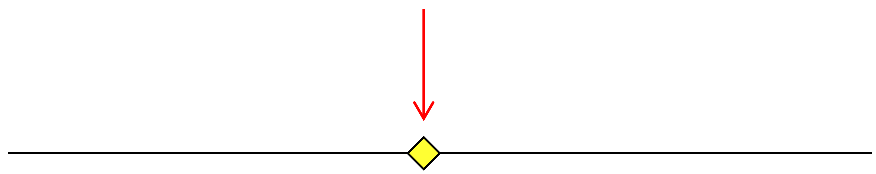


Interval Estimator



Σημειακός Εκτιμητής…

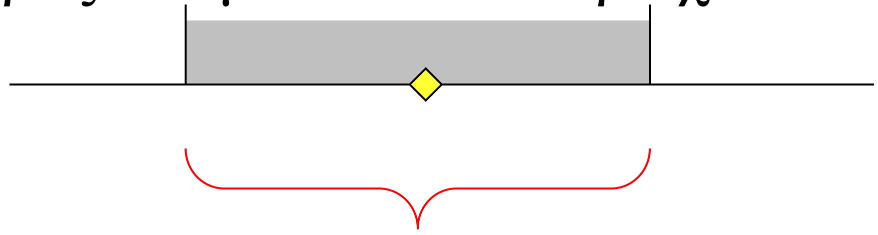
Ο εκτιμητής σημείου (σημειακός εκτιμητής) είναι η τιμή ενός στατιστικού μεγέθους ενός δείγματος που χρησιμοποιείται απευθείας ως προσέγγιση της αντίστοιχης παραμέτρου του πληθυσμού.



Είδαμε νωρίτερα ότι η πιθανότητα εμφάνισης ενός συγκεκριμένου σημείου (τιμής) σε συνεχείς κατανομές ήταν πρακτικά μηδενικό. Επίσης θα περίμενε κανείς ο εκτιμητής σημείου να πλησιάζει περισσότερο την τιμή της πληθυσμιακής παραμέτρου όσο το δείγμα αυξάνει, όμως αυτό δεν συμβαίνει. Ως εκ τούτου χρησιμοποιείται ο **εκτιμητής διαστήματος (*interval estimator*)** για τη εκτίμηση των πληθυσμιακών παραμέτρων…

Εκτιμητής διαστήματος…

Ένας εκτιμητής διαστήματος (***interval estimator)*** δίνει την τιμή μιας παραμέτρου με την μορφή ενός διαστήματος τιμών, στο οποίο η παράμετρος αναμένεται να περιέχεται.



Έτσι λέγεται ότι (με κάποια ___% βεβαιότητα) ότι η παράμετρος του πληθυσμού ενδιαφέροντος είναι μεταξύ κάποιου κατώτερου και ανώτερου όριου.

Σημειακή & Διαστήματος Εκτίμηση...

Για παράδειγμα, ας υποθέσουμε ότι θέλουμε να υπολογιστεί το μέσο καλοκαιρινό εισόδημα $\overline{x}$ μιας κατηγορίας σπουδαστών.

Για n = 25 μαθητές, υπολογίζεται σε 400 €/ εβδομάδα.

Εκτίμηση σημείου (point estimate)

Εκτίμηση διαστήματος (interval estimate)

Είναι μια εναλλακτική δήλωση:

Το μέσο εισόδημα είναι μεταξύ $\underbrace{\text{380 and 420}}$ €/ εβδομάδα.

Ιδιότητες εκτιμητών...

Επιθυμητές ιδιότητες σε εκτιμητές σημείου είναι:
Αμεροληψία, συνέπεια, και σχετική αποδοτικότητα:

Ένας ***αμερόληπτος εκτιμητής*** (***unbiased estimator***) μιας παραμέτρου πληθυσμού είναι ένας εκτιμητής αναμενόμενη τιμή του οποίου είναι ισούται με την τιμή της παραμέτρου του πληθυσμού.

Ένας αμερόληπτος εκτιμητής λέγεται ότι είναι ***συνεπής (consistent)***, αν η διαφορά μεταξύ της τιμής του εκτιμητή και της τιμής της παραμέτρου (πληθυσμιακής) μειώνεται όσο μεγαλώνει το μέγεθος του δείγματος.

Εάν υπάρχουν δύο αμερόληπτοι εκτιμητές μιας παραμέτρου, αυτός που έχει την μικρότερη διακύμανση λέγεται ότι είναι ***σχετικά αποδοτικός (relative efficient)***.

Αμερόληπτος εκτιμητής…

Ένας ***αμερόληπτος εκτιμητής*** (***unbiased estimator***) μιας παραμέτρου πληθυσμού είναι ένας εκτιμητής που η αναμενόμενη τιμή του οποίου είναι ίση με την εν λόγω παράμετρο

Π.χ. ο αριθμητικός μέσος δείγματος $\bar{X}$ είναι ένας αμερόληπτος εκτιμητής της μέση της τιμής μ πληθυσμού, αφού:

$$E(\bar{X}) = \mu$$

Αμερόληπτος εκτιμητής …

Ένας ***αμερόληπτος εκτιμητής*** (***unbiased estimator***) μιας παραμέτρου πληθυσμού είναι ένας εκτιμητής που η αναμενόμενη τιμή του οποίου είναι ίση με την εν λόγω παράμετρο.

Π.χ. ο μέσος δείγματος είναι ένας αμερόληπτος εκτιμητής του πληθυσμιακού μέσου μ αφού:

Ε(μέσος δείγματος) = μ

Συνέπεια…

Ένας αμερόληπτος εκτιμητής ονομάζεται ***συνεπής*** εάν η διαφορά της τιμής του εκτιμητή και της τιμής της παραμέτρου μειώνεται καθώς αυξάνεται το μέγεθος του δείγματος.

Π.χ. $\overline{X}$ είναι συνεπής εκτιμητής του μ γιατί:

$V(\overline{X}) = \sigma^2/n$

Δηλαδή, καθώς το n αυξάνει, η διακύμανση του $\overline{X}$ μειώνεται.

Συνέπεια …

Ένας αμερόληπτος εκτιμητής ονομάζεται ***συνεπής*** εάν η διαφορά της τιμής του εκτιμητή και της τιμής της παραμέτρου μειώνεται καθώς αυξάνεται το μέγεθος του δείγματος.

Π.χ. ο μέσος δείγματος είναι ένας ***συνεπής*** εκτιμητής της τιμής της παραμέτρου μ επειδή:

V(μέσος δείγματος) = $1.57\sigma^2/n$

Δηλαδή, καθώς το n αυξάνεται, η διακύμανση του μέσου δείγματος μειώνεται.

Σχετική Αποδοτικότητα...

Αν έχουμε δύο αμερόληπτους εκτιμητές μιας πληθυσμιακής παραμέτρου, αυτός που έχει την μικρότερη διακύμανση ονομάζεται σχετικά αποδοτικός (***relatively efficient)***.

Π.χ. Και ο μέσος δείγματος και ενδιάμεσος δείγματος είναι αμερόληπτοι εκτιμητές όμως ο ενδιάμεσος δείγματος έχει μεγαλύτερη διακύμανση σε σχέση με τον μέσο δείγματος. Έτσι επιλέγουμε το μέσο δείγματος $\bar{x}$ ως σχετικά αποδοτικότερο εκτιμητή.

Συνεπώς ο μέσος δείγματος $\bar{x}$ είναι ο καλύτερος εκτιμητής του πληθυσμιακού μέσου μ.

Εκτίμηση του μ όταν σ γνωστό ...

Στο κεφάλαιο 8 είδαμε την ακόλουθη γενική διατύπωση της πιθανότητας του $\overline{X}$

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

Από το κεφάλαιο 9 η κατανομή δείγματος του $\overline{X}$ είδαμε ότι κατά προσέγγιση ακολουθεί την κανονική κατανομή με τυπική απόκλιση: $\sigma / \sqrt{n}$

Συνεπώς το $$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

ακολουθεί (κατά προσέγγιση) την τυπική κανονική κατανομή.

Εκτίμηση του μ όταν σ γνωστό …

Έτσι, αντικαθιστώντας Z παράγουμε

$$P\left(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

Στο κεφάλαιο 9 (με ένα μικρό κομμάτι της άλγεβρας) εκφράσαμε τα εξής:

$$P\left(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Με ένα μικρό βήμα της άλγεβρας διαφορετικά έχουμε

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Εκτίμηση του μ όταν σ γνωστό …

Συνεπώς

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Είναι μια άλλη έκφραση της πιθανότητας του $\bar{x}$.

Ωστόσο, ισοδύναμη έκφραση είναι επίσης ότι αυτό αποτελεί τον ***εκτιμητή διαστήματος εμπιστοσύνης του μ***

Εκτίμηση του μ όταν σ γνωστό ...

Το διάστημα μπορεί επίσης να εκφραστεί ως

Κατώτερο όριο εμπιστοσύνης
Lower confidence limit (LCI) = $$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Ανώτερο όριο εμπιστοσύνης
Upper confidence limit (UCI) = $$\left(\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Η πιθανότητα ***1 – α*** ονομάζεται ***επίπεδο ή στάθμη εμπιστοσύνης***, η οποία εκφράζει το πόσο συχνά (πιθανά) το εν λόγω διάστημα περιέχει πραγματικά το μ.

Παράδειγμα 10.1…

Η εταιρία Doll Computer Company κατασκευάζει υπολογιστές και τους παραδίδει κατευθείαν στους πελάτες οι οποίοι τους παραγγέλνουν στο διαδίκτυο.

Προκειμένου να ανταποκριθεί άμεσα, η Doll κατασκευάζει εκ των πρότερων τον καθένα από τα πέντε δημοφιλέστερους υπολογιστές και τους μεταφέρει αποθήκες, από τις οποίες παίρνει γενικά 1 μέρα προκειμένου να παραδώσει έναν υπολογιστή στον πελάτη.

Η στρατηγική αυτή απαιτεί υψηλά επίπεδα αποθέματος που επιβαρύνει σημαντικά το κόστος.

Παράδειγμα 10.1…

Προκειμένου να μειωθούν τα σχετικά κόστη ο operations manager θέλει να χρησιμοποιήσει ένα υπόδειγμα αποθεμάτων (inventory model). Αυτός σημειώνει ως τυχαίες μεταβλητές την ημερήσια ζήτηση και τον χρόνο παράδοσης που ακολουθούν την κανονική κατανομή (κατά προσέγγιση) και χρειάζεται να μάθει το μέσο προκειμένου να υπολογίσει το βέλτιστο επίπεδο αποθεμάτων.

Καταγράφονται 25 περιπτώσεις παράδοσης και καταγράφει την ζήτηση κάθε περιόδου. Xm10-01

Ο manager θέλει να κατασκευάσει ένα 95% διάστημα εμπιστοσύνης εκτίμησης της μέσης ζήτησης. Υποθέστε ότι ο manager γνωρίζει ότι η τυπική απόκλιση είναι 75 υπολογιστές.

Παράδειγμα 10.1…

235	374	309	499	253
421	361	514	462	369
394	439	348	344	330
261	374	302	466	535
386	316	296	332	334

Παράδειγμα 10.1…

“Θέλουμε να υπολογίσετε τη μέση ζήτηση διαχρονικά ώστε να οδηγήσει με βεβαιότητα 95% στον καθορισμό των επίπεδων αποθεμάτων…”

ΠΡΟΣΔΙΟΡΙΣΜΟΣ

Συνεπώς, η παράμετρος που πρέπει να εκτιμηθεί είναι ο πληθυσμιακός μέσος:μ

Και επομένως ο εκτιμητής του διαστήματος εμπιστοσύνης είναι:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Παράδειγμα 10.1... ΥΠΟΛΟΓΙΣΜΟΙ

Προκειμένου να υπολογιστεί ο εκτιμητής του διαστήματος εμπιστοσύνης, προκύπτει:

$\bar{x}$	370.16
$z_{\alpha/2}$	1.96
σ	75
n	25

Υπολογισμός από στοιχεία…

$1-\alpha = .95, \therefore \alpha/2 = .025$

$so \;\; z_{\alpha/2} = z_{.025} = 1.96$

Δεδομένα προβλήματος (γνωστά)

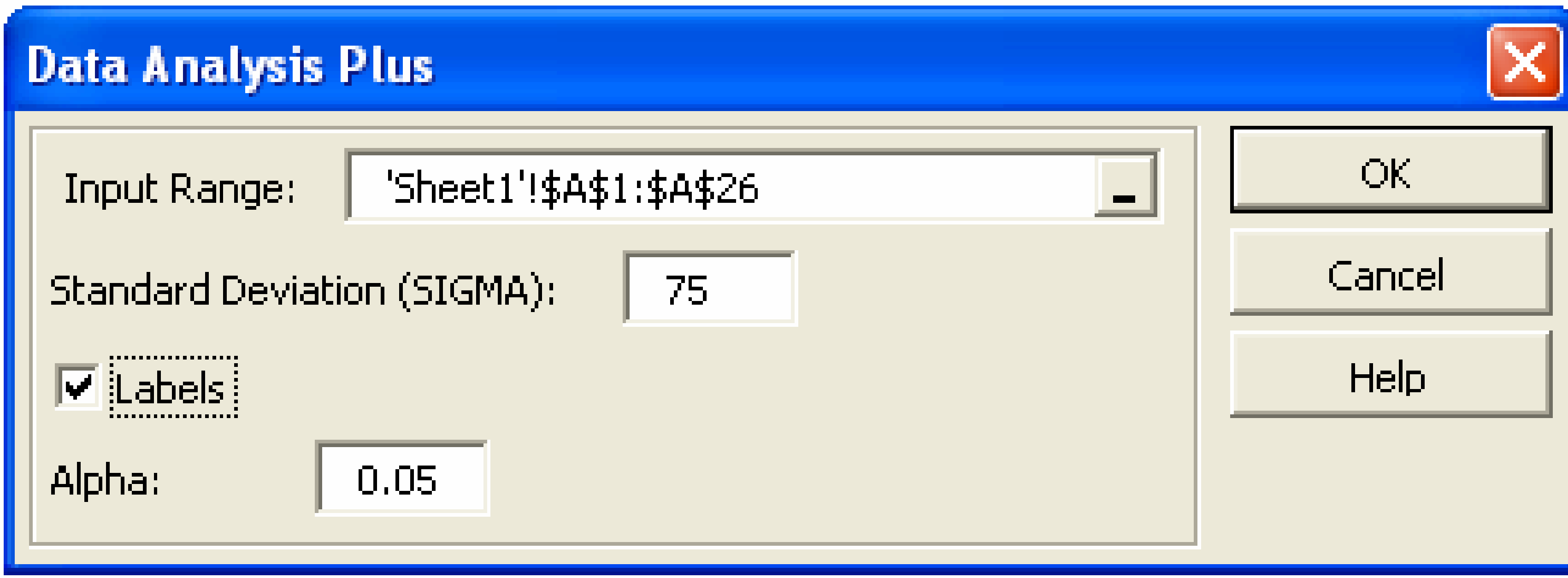
συνεπώς: $$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm z_{.025}\frac{75}{\sqrt{25}} = 370.16 \pm 1.96\frac{75}{\sqrt{25}} = 370.16 \pm 29.40$$

Το **κατώτερο** και **ανώτερο** επίπεδο εμπιστοσύνης είναι 340.76 & 399.56.

Παράδειγμα 10.1 Χρήση Excel ΥΠΟΛΟΓΙΣΜΟΙ

Χρησιμοποιώντας το Data Analysis Plus™ στο παράδειγμα Xm10-01_file, λαμβάνουμε τις ίδιες απαντήσεις ευκολότερα

Click Add-In > Data Analysis Plus > Z-Estimate: Mean



Παράδειγμα 10.1

	A	B	C
1	**z-Estimate: Mean**		
2			
3			*Demand*
4	Mean		370.16
5	Standard Deviation		80.78
6	Observations		25
7	SIGMA		75
8	LCL		340.76
9	UCL		399.56

Παράδειγμα 10.1… **ΕΡΜΗΝΕΙΑ**

Η εκτίμηση για τη μέση ζήτηση κατά το χρόνο παράδοσης βρίσκεται μεταξύ 340.76 και 399.56 — μπορούμε να το χρησιμοποιήσουμε αυτό ως συμβολή στην ανάπτυξη μιας πολιτικής αποθεμάτων.

Δηλαδή, εκτιμήσαμε ότι η μέση ζήτηση πέφτει μεταξύ 340.76 και 399.56, και αυτός ο τύπος εκτίμησης είναι σωστός κατά 95%

Παρεμπιπτόντως, τα μέσα ενημέρωσης συχνά αναφέρεται το ποσοστό 95% ως "19 φορές από 20," το οποίο δίνει έμφαση στη μακροπρόθεσμη πτυχή του επιπέδου εμπιστοσύνης.

Εύρος διαστήματος …

***Ένα ευρύ διάστημα παρέχει λίγες πληροφορίες*.**

Για παράδειγμα, ας υποθέσουμε ότι υπολογίζουμε με βεβαιότητα 95% ότι ο μέσος αρχικός μισθός του λογιστή είναι μεταξύ \$15.000 και \$100.000.

Αντίθεση με αυτό: το 95% διάστημα εμπιστοσύνης των αρχικών μισθών είναι από \$42,000 έως \$45,000.

Η δεύτερη εκτίμηση είναι πολύ στενότερο, παρέχοντας ακριβέστερες πληροφορίες σχετικά με την έναρξη των μισθών.

Εύρος διαστήματος …

Το ***εύρος του διαστήματος εμπιστοσύνης*** αποτελεί συνάρτηση του επίπεδου εμπιστοσύνης, της πληθυσμιακής τυπικής απόκλισης, και του μεγέθους του δείγματος…

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

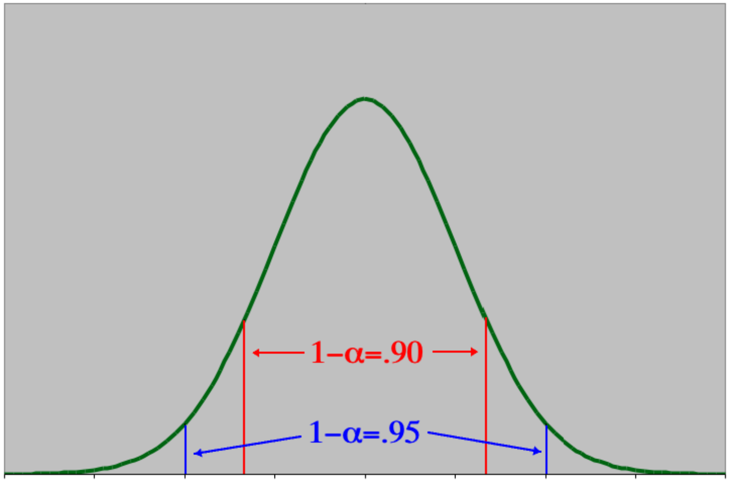
Εύρος διαστήματος ...

Το ***εύρος του διαστήματος εμπιστοσύνης*** αποτελεί συνάρτηση του επίπεδου εμπιστοσύνης, της πληθυσμιακής τυπικής απόκλισης, και του μεγέθους του δείγματος ...

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Ένα μεγαλύτερο επίπεδο εμπιστοσύνης παράγει ένα μεγαλύτερο διάστημα εμπιστοσύνης:

Estimators.xls



Εύρος διαστήματος …

Το ***εύρος του διαστήματος εμπιστοσύνης*** αποτελεί συνάρτηση του επίπεδου εμπιστοσύνης, της πληθυσμιακής τυπικής απόκλισης, και του μεγέθους του δείγματος …

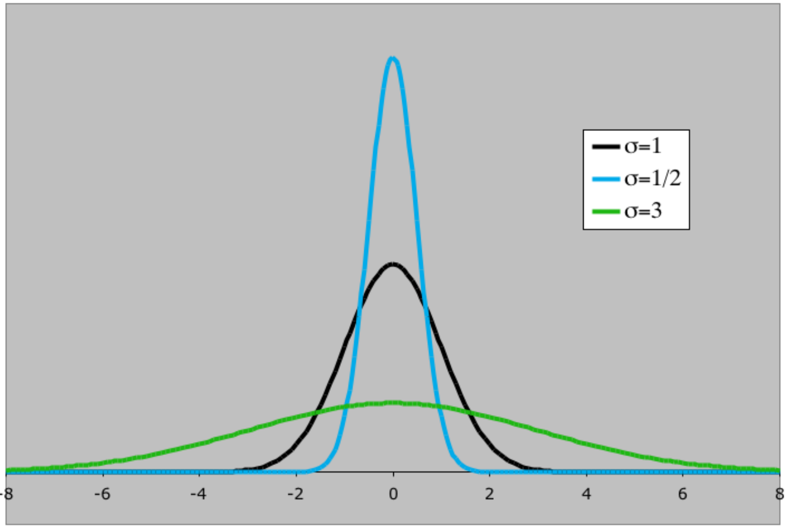
$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Μεγαλύτερες τιμές του σ

Παράγουν ευρύτερα

διαστήματα εμπιστοσύνης

[Estimators.xls]



Εύρος διαστήματος …

Το ***εύρος του διαστήματος εμπιστοσύνης*** αποτελεί συνάρτηση του επίπεδου εμπιστοσύνης, της πληθυσμιακής τυπικής απόκλισης, και του μεγέθους του δείγματος …

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Αύξηση του μεγέθους δείγματος μειώνει το πλάτος του διαστήματος εμπιστοσύνης ενώ το επίπεδο εμπιστοσύνης παραμένει αμετάβλητο. Estimators.xls

Σημείωση: αυτό αυξάνει επίσης το κόστος της απόκτησης πρόσθετων δεδομένων

Επιλογή μεγέθους δείγματος…

Στο κεφάλαιο 5 επισημάναμε ότι η λάθος δειγματοληψία είναι η διαφορά μεταξύ της τιμής του εκτιμητής και της τιμής της παραμέτρου.

Μπορούμε να ορίσουμε αυτή τη διαφορά ως ***σφάλμα εκτίμησης.***

Επιλογή μεγέθους δείγματος …

Το όριο για το σφάλμα της εκτίμησης είναι:

$$B = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Με άλγεβρα βρίσκουμε το μέγεθος του δείγματος για την εκτίμηση του μέσου.

$$n = \left(\frac{z_{\alpha/2}\sigma}{B} \right)^2$$

Επιλογή μεγέθους δείγματος …

Στο παράδειγμα 10.1 πριν την συλλογή στοιχείων ο manager αποφάσισε ότι επιβαλλόταν το σφάλμα εκτίμησης της μέσης ζήτησης να μην ξεπερνά τις16 μονάδες (υπολογιστές).

Επίσης έχουμε $1 - \alpha = .95$ και $\sigma = 75$. Συνεπώς

$$n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2 = \left(\frac{(1.96)(75)}{16}\right)^2 = 84.41$$

Επιλογή μεγέθους δείγματος …

Επειδή το n πρέπει να είναι ακέραιος και επειδή θέλουμε το όριο σχετικά με το σφάλμα της εκτίμησης να είναι όχι περισσότερο από 16 οποιαδήποτε μη ακέραια τιμή πρέπει να στρογγυλοποιηθεί.

Έτσι, η τιμή του n στρογγυλοποιείται 85, πράγμα που σημαίνει ότι για να είναι 95% σίγουρος ότι το σφάλμα της εκτίμησης της θα είναι όχι μεγαλύτερο από 16, θα πρέπει να δοκιμάσετε 85 τυχαία χρονικά διαστήματα.