**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ**

**ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΠΑΤΡΑΣ**

Εργαστήριο Λήψης Αποφάσεων & Επιχειρησιακού Προγραμματισμού

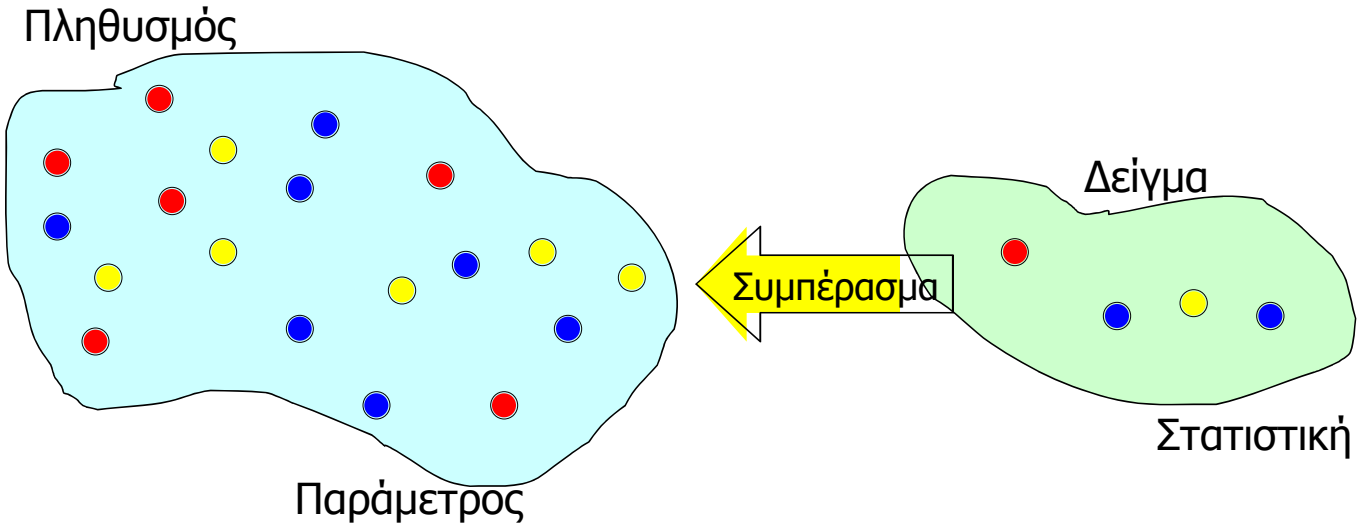
*Καθηγητής Ι. Μητρόπουλος*

# ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΕΙΔΙΚΑ ΘΕΜΑΤΑ

## Κεφάλαιο 12

## Εκτίμηση των παραμέτρων ενός πληθυσμού

# Εκτιμητική σχετικά με έναν πληθυσμό...




Θα αναπτύξουμε τεχνικές προκειμένου να εκτιμηθούν – ελεγχθούν πληθυσμιακοί παράμετροι:

Πληθυσμιακός μέσος αριθμητικός $\mu$

Πληθυσμιακή αναλογία $\boldsymbol{p}$

# Εκτιμητική όταν η διακύμανση σ είναι άγνωστη...

Στα προηγούμενα κεφάλαια είδαμε τρόπους εκτίμησης & ελέγχου του μέσου ενός πληθυσμού όταν η διακύμανσή του & η τυπική απόκλισή του ήταν γνωστή. Συγκεκριμένα τόσο ο εκτιμητής διαστήματος όσο και ο έλεγχος υπόθεσης υπολογίστηκαν από τον τύπο:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

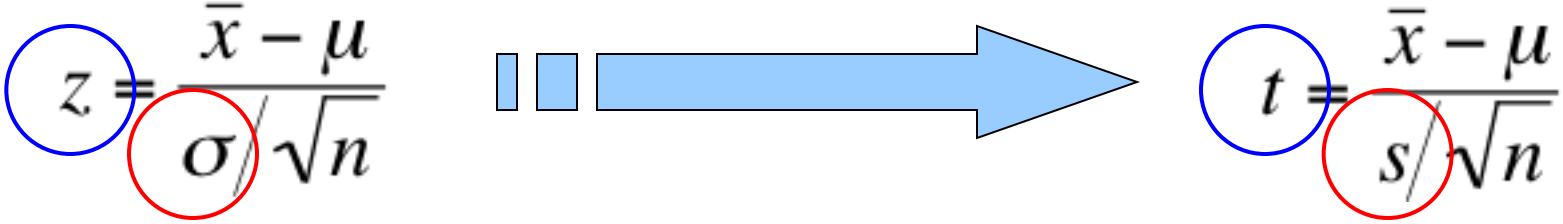
*Αλλά πόσο ρεαλιστικό είναι να γνωρίζουμε την πληθυσμιακή διακύμανση? Η απάντηση είναι ΟΧΙ!*

Για το λόγο αυτό, χρησιμοποιούμε τον στατιστικό έλεγχο ***Student t***, που δίνεται από τον τύπο:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

# Εκτιμητική όταν η διακύμανση σ είναι άγνωστη...

Όταν **σ** άγνωστη τότε χρησιμοποιούμε τον σημειακό εκτιμητή της ***s***




και ο z- έλεγχος αντικαθίσταται από τον t-έλεγχο, που ακολουθεί την κατανομή **student t** με ν = n-1 βαθμούς ελευθερίας.

Ανάλογα ο εκτιμητής διαστήματος εμπιστοσύνης του πληθυσμιακού μέσου μ δίνεται από τον τύπο: $\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$

# Έλεγχος μ όταν σ άγνωστο ...

## 1.Έλεγχος υποθέσεων.

Όταν η πληθυσμιακή τυπική απόκλιση είναι άγνωστη και ο πληθυσμός ακολουθεί την κανονική κατανομή, ο έλεγχος υποθέσεων για τον πληθυσμιακό μέσο μ υπολογίζεται από τον τύπο: $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

και ακολουθεί την κατανομή **student t** με ν = n-1 βαθμούς ελευθερίας

## 2.Εκτιμητής διαστήματος εμπιστοσύνης.

Όταν η πληθυσμιακή τυπική απόκλιση είναι άγνωστη και ο πληθυσμός ακολουθεί την κανονική κατανομή, ο εκτιμητής διαστήματος του πληθυσμιακού μέσου μ υπολογίζεται από τον τύπο: $$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

# Παράδειγμα 12.1

Στο κοντινό μέλλον όλες οι χώρες θα χρειαστεί να λάβουν μέτρα για την προστασία του περιβάλλοντος.

Πιθανές δράσεις ο περιορισμός της κατανάλωσης ενέργειας και η ανακύκλωση,

Σήμερα τα περισσότερα προϊόντα που παράγονται από ανακυκλωμένα υλικά είναι σημαντικά ακριβότερα σε σχέση με αυτά που κατασκευάζονται με παραδοσιακές τεχνικές.

# Παράδειγμα 12.1

Οι εφημερίδες αποτελούν εξαίρεση.

Είναι επικερδές η ανακύκλωση εφημερίδων.

Η σημαντικότερη πηγή κόστους στην ανακύκλωση εφημερίδων αποτελεί η συλλογή τους από τα νοικοκυριά και τελευταίως έχουν εμφανιστεί αρκετές εταιρίες που δραστηριοποιούνται στον τομέα αυτό.

Ο οικονομικός αναλυτής μιας τέτοιας εταιρίας υπολόγισε ότι για να είναι κερδοφόρα μια τέτοια δράση θα πρέπει η μέση εβδομαδιαία ποσότητα να είναι μεγαλύτερη από 2 λίβρες (0,9 kg) ανά νοικοκυριό.

# Παράδειγμα 12.1

Έτσι στα πλαίσια της μελέτης σκοπιμότητας για την κατασκευή ενός νέου κέντρου συλλογής η εταιρία επέλεξε ένα τυχαίο δείγμα 148 νοικοκυριών της περιοχής και κατέγραψε το βάρος (σε λίβρες) των εφημερίδων που απέρριψε κάθε νοικοκυριό του δείγματος κατά την διάρκεια μιας εβδομάδας. Xm12-01*

Μπορούμε από τα στοιχεία να συμπεράνουμε ότι το σχεδιαζόμενο νέο κέντρο συλλογής εφημερίδων για ανακύκλωση θα είναι κερδοφόρο.

# Παράδειγμα 12.1

**ΚΑΘΟΡΙΣΜΟΣ**

Το ζητούμενο είναι ο έλεγχος μια υπόθεσης για τον μέσο του βάρους των εφημερίδων που μπορεί να ανακυκλώσουν τα νοικοκυριά της περιοχής.

Έτσι η υπό έλεγχο παράμετρος είναι ο πληθυσμιακός μέσος μ

Θέλουμε να γνωρίσουμε εάν υπάρχει επαρκείς αποδείξεις για να συμπεράνουμε ότι ο μέσος είναι μεγαλύτερος του 2. Συνεπώς,

$$H_1: \mu > 2$$

Η μηδενική υπόθεση ως συνήθως:

$$H_0: \mu = 2$$

# Παράδειγμα 12.1

**ΚΑΘΟΡΙΣΜΟΣ**

Επειδή η τυπική απόκλιση σ του πληθυσμού είναι άγνωστη θα χρησιμοποιηθεί ο έλεγχος t :

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad \nu = n - 1 = 148\text{-}1 = 147$$

Επειδή η εναλλακτική υπόθεση είναι:

$$H_1: \mu > 2$$

Η περιοχή απόρριψης γίνεται για α=0,01:

$$t > t_{\alpha,\nu} = t_{.01,148} \approx t_{.01,150} = 2.351$$

*Example 12.1 Manual Calculations*

# Παράδειγμα 12.1

ΚΑΘΟΡΙΣΜΟΣ

Υπολογισμοί:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2,5 + 0,7 + \ldots + 2,6 + 3,1}{148} = \frac{322,7}{148} = 2,18$$

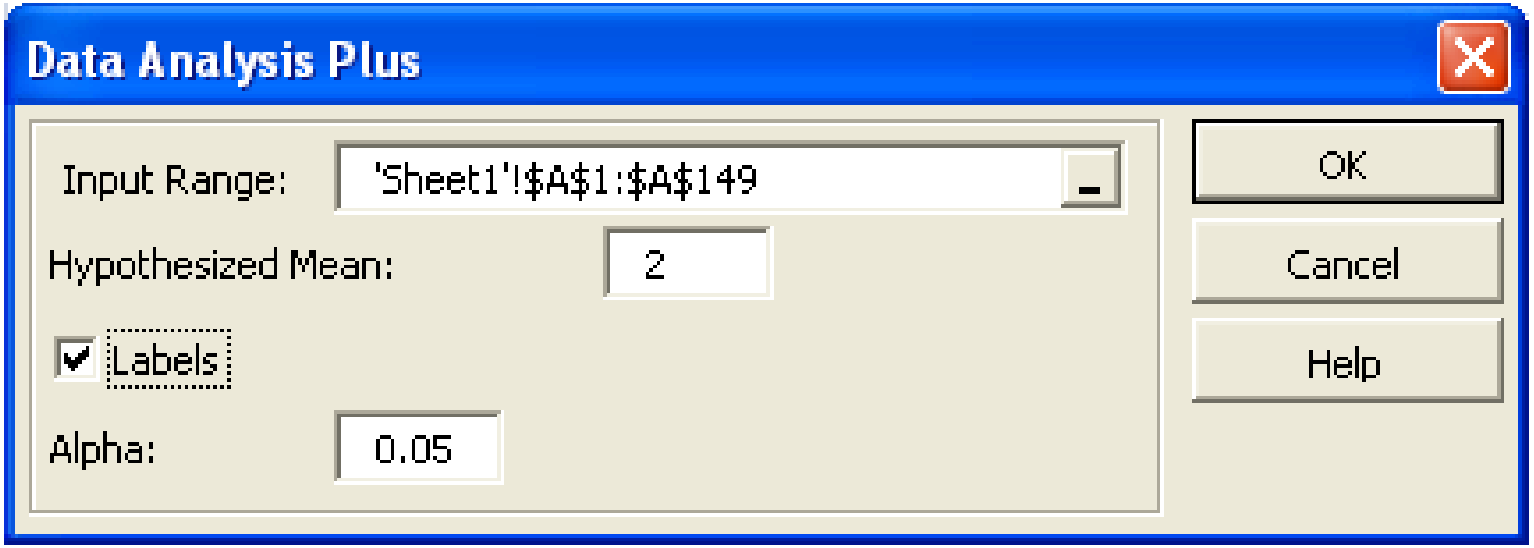
$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right] = 0,962 \;\&\; s = \sqrt{s^2} = 0,981$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{2,18 - 2}{0,981/\sqrt{148}} = 2,24$$

# Παράδειγμα 12.1

**ΥΠΟΛΟΓΙΣΜΟΣ**

Click Add-Ins, Data Analysis Plus, t-test: Mean



# Παράδειγμα 12.1

**ΥΠΟΛΟΓΙΣΜΟΣ**

| | A | B | C | D |
|---|---|---|---|---|
| 1 | **t-Test: Mean** | | | |
| 2 | | | | |
| 3 | | | | *Newspaper* |
| 4 | Mean | | | 2.18 |
| 5 | Standard Deviation | | | 0.98 |
| 6 | Hypothesized Mean | | | 2 |
| 7 | df | | | 147 |
| 8 | t Stat | | | 2.24 |
| 9 | P(T<=t) one-tail | | | 0.0134 |
| 10 | t Critical one-tail | | | 1.6553 |
| 11 | P(T<=t) two-tail | | | 0.0268 |
| 12 | t Critical two-tail | | | 1.9762 |

:

# Παράδειγμα 12.1

**ΕΡΜΗΝΕΙΑ**

Η τιμή του στατιστικού ελέγχου t = 2.24 και η αντίστοιχη p-τιμή είναι .0134.

Συνεπώς δεν υπάρχουν επαρκείς αποδείξεις ώστε να συμπεράνουμε ότι το μέσο βάρος των εφημερίδων προς απόρριψη είναι μεγαλύτερο των 2.0 λιβρών.

*Παρατηρήστε ότι υπάρχει κάποια ένδειξη, η p-τιμή είναι 0.0134. Ωστόσο, επειδή θέλαμε να είμαστε πολύ σίγουροι για το συμπέρασμα ορίστηκε ως α = 0,01*

**Έτσι, δεν καταλήγουμε στο συμπέρασμα ότι το εργοστάσιο ανακύκλωσης θα ήταν κερδοφόρο.**

# Παράδειγμα 12.2

In 2004 (the latest year reported) 130,134,000 tax returns were files in the United States.

The Internal Revenue Service (IRS) examined 0.77% or 1,008,000 of them to determine if they were correctly done.

To determine how well the auditors are performing a random sample of these returns was drawn and the additional tax was reported. Xm12-02

Estimate with 95% confidence the mean additional income tax collected from the 1,008,000 files audited.

# Example 12.2

**IDENTIFY**

The objective is to describe the population of additional tax collected.

The data are interval.

The parameter to be estimated is the mean additional tax.

The confidence interval estimator is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

# Example 12.2

COMPUTE

For manual calculations click

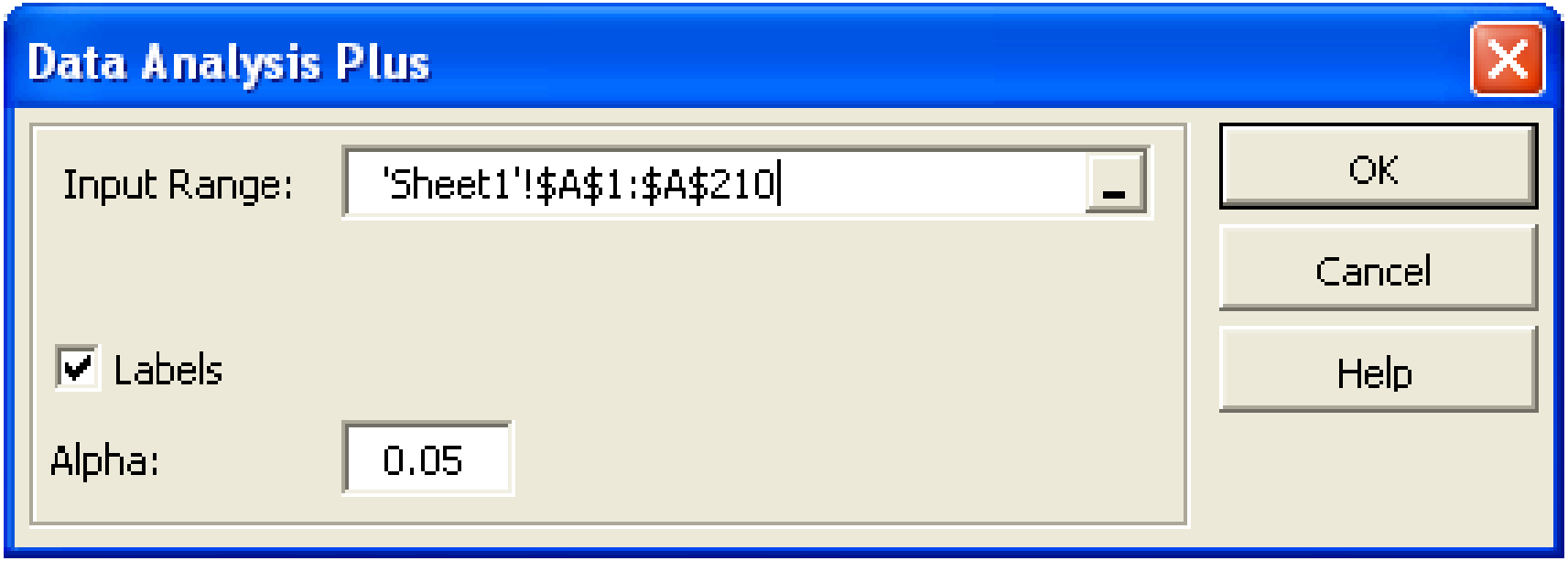
Example 12.2 manual calculations

For Excel skip to next slide.

# Example 12.2

**COMPUTE**

Click Add-Ins, Data Analysis Plus, t-Estimate: Mean



# Example 12.2

**COMPUTE**

| | A | B | C | D |
|---|---|---|---|---|
| 1 | **t-Estimate: Mean** | | | |
| 2 | | | | |
| 3 | | | | *Taxes* |
| 4 | Mean | | | 6001 |
| 5 | Standard Deviation | | | 2864 |
| 6 | LCL | | | 5611 |
| 7 | UCL | | | 6392 |

# Example 12.2…

**INTERPRET**

We estimate that the mean additional tax collected lies between \$5,611 and \$6,392 .

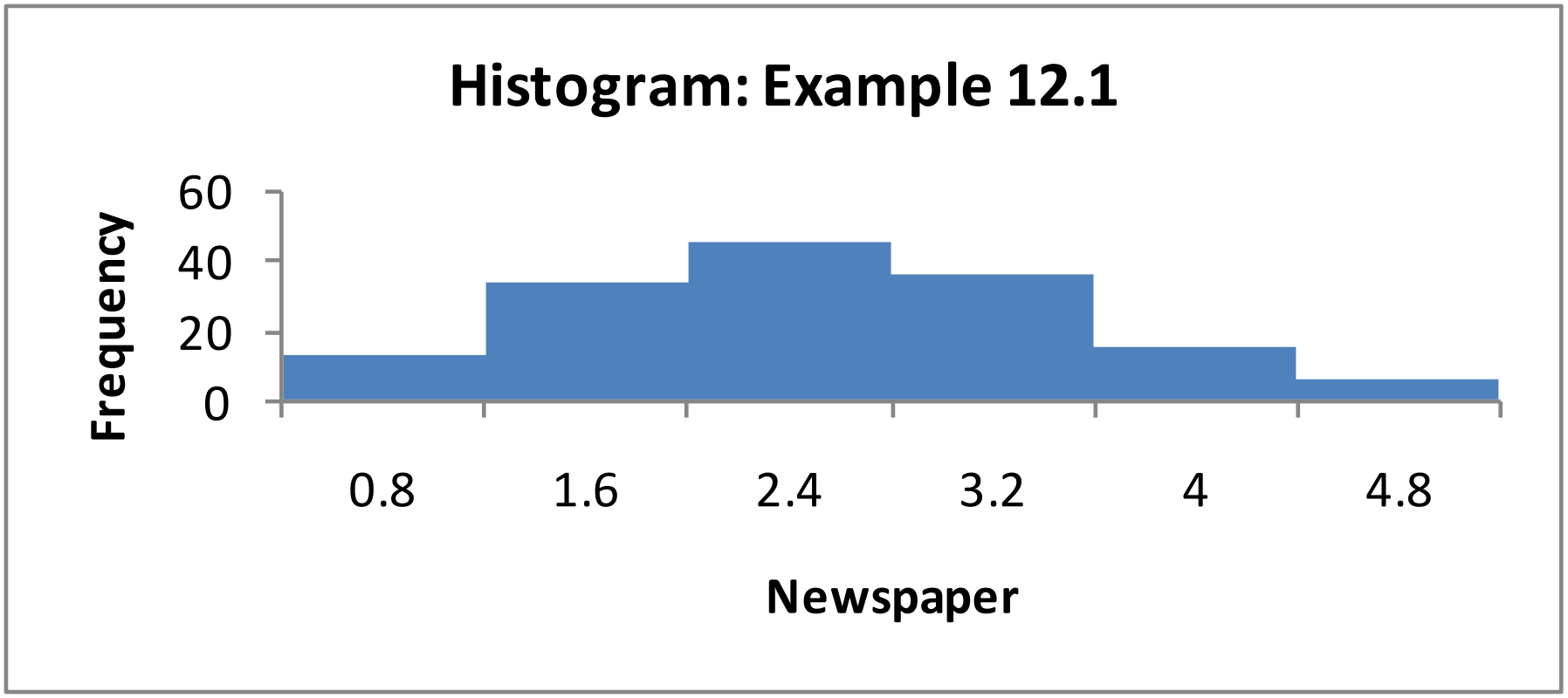
We can use this estimate to help decide whether the IRS is auditing the individuals who should be audited.

# Check Required Conditions

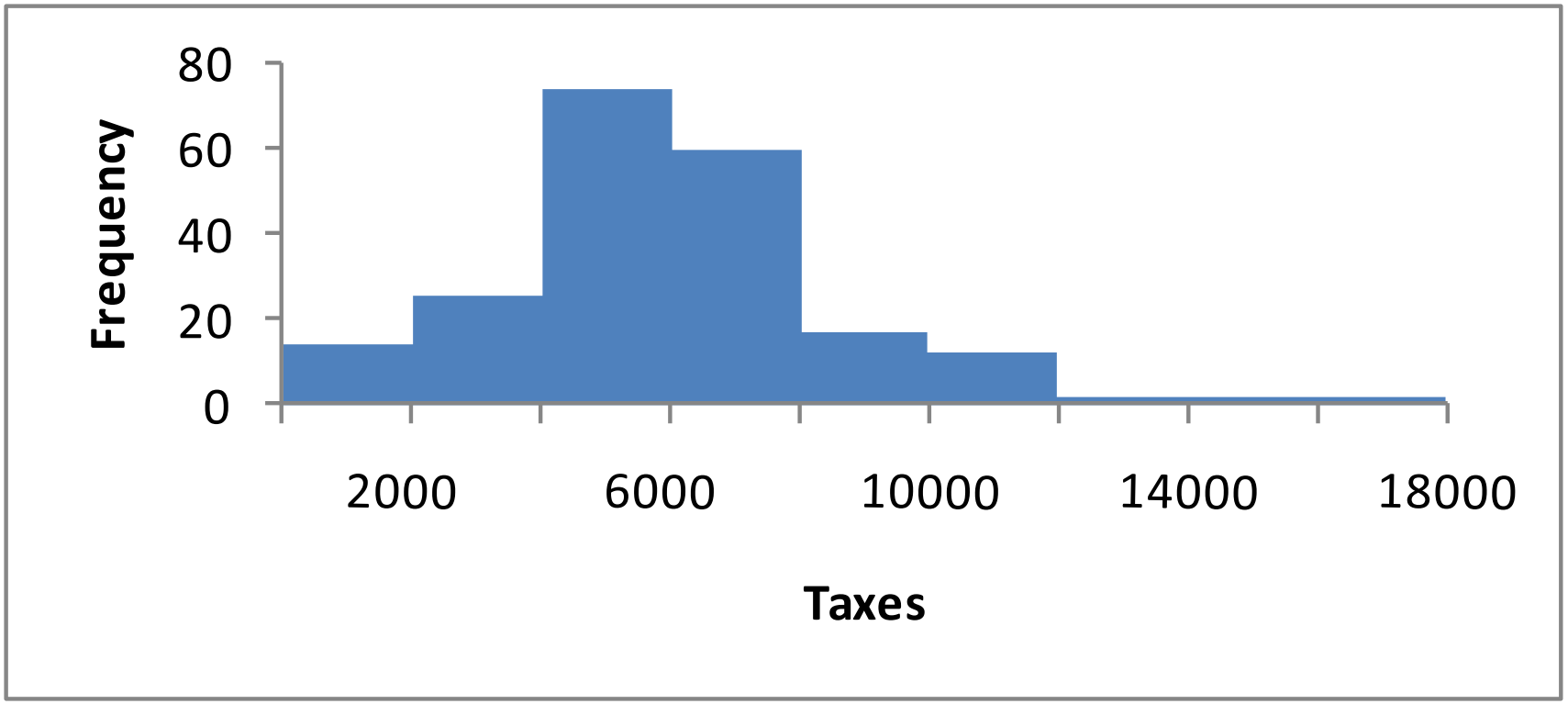
The Student t distribution is ***robust***, which means that if the population is nonnormal, the results of the t-test and confidence interval estimate are still valid provided that the population is "not ***extremely*** nonnormal".

To check this requirement, ***draw a histogram*** of the data and see how "bell shaped" the resulting figure is. If a histogram is extremely skewed (say in the case of an exponential distribution), that could be considered "extremely nonnormal" and hence t-statistics would be not be valid in this case.

# Histogram for Example 12.1

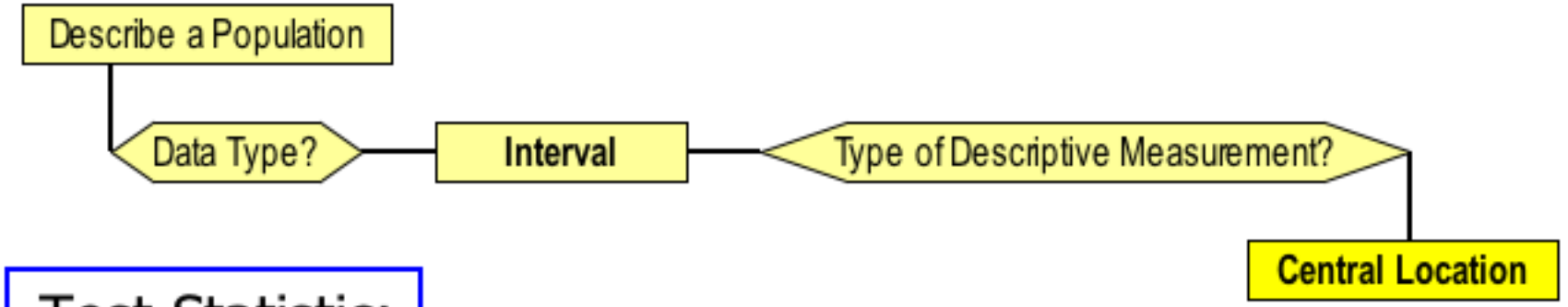


# Histogram for Example 12.2



# Εντοπισμός παραγόντων

Παράγοντες που προσδιορίζουν τον t-έλεγχο και τον εκτιμητή του μ :




Parameter of interest: $\mu$

Test Statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

*ν=n-1 degrees of freedom*

Interval Estimator:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Required Condition: Population is normal (or at least "not extremely nonnormal")

# Εκτιμητική: Πληθυσμιακή Αναλογία...

Όταν τα δεδομένα είναι ονομαστικά, καταμετρούμε τον αριθμό των εμφανίσεων της κάθε τιμής και υπολογίζουμε τις αναλογίες. Έτσι, η παράμετρος που μας ενδιαφέρει για να περιγράψουμε ένα πληθυσμό ονομαστικών δεδομένων είναι η πληθυσμιακή αναλογία ***p***.

Η παράμετρος αυτή βασίζεται στο δυωνυμικό πείραμα. Ο στατιστικός δείκτης που χρησιμοποιείται για την εκτίμηση και τον έλεγχο μιας αναλογίας ορίζεται ως: $\hat{p} = \frac{x}{n}$

Όπου p-καπελάκι (hat) $\hat{p}$ ) είναι η αναλογία του δείγματος με **x** αριθμό επιτυχιών σε ένα δείγμα μεγέθους **n**.

# Εκτιμητική: Πληθυσμιακή Αναλογία…

Όταν ***np*** και ***n(1–p)*** είναι ταυτόχρονα μεγαλύτερα του 5, η κατανομή του δείγματος του $\hat{p}$ είναι κατά προσέγγιση κανονική με μέσο: $\mu = p$

Τυπική απόκλιση: $\sigma = \sqrt{\dfrac{p(1-p)}{n}}$

Και τυποποιημένο έλεγχο: $z = \dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}}$

# Εκτιμητική: Πληθυσμιακή Αναλογία…

**1. Έλεγχος για μια αναλογία πληθυσμού**

Ο έλεγχος υπόθεσης που χρησιμοποιούμε για μια αναλογία p ενός πληθυσμού είναι:

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

**2. Εκτιμητής διαστήματος εμπιστοσύνης για μια αναλογία πληθυσμού**

Ο εκτιμητής διαστήματος εμπιστοσύνης α για μια αναλογία πληθυσμού υπολογίζεται ως:

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

(και για τα δύο απαιτείται ότι $np>5$ και $n(1-p)>5$)

# Παράδειγμα 12.5

Στις Αμερικάνικες προεδρικές εκλογές υπάρχουν δύο μόνο υποψήφιοι και αυτός που κερδίζει την πλειοψηφία σε μια πολιτεία κερδίζει το σύνολο των εκλεκτόρων της πολιτείας αυτής. Στην πράξη, αυτό σημαίνει ότι είτε ο Δημοκρατικός ή ο Ρεπουμπλικάνος υποψήφιος θα κερδίσει.

Υποθέστε οτι τα αποτελέσματα της δημοσκόπησης εξόδου (exit poll) καταγράφηκαν και κωδικοποιήθηκαν ως 1 = Δημοκρατικός υποψήφιος 2 = Ρεπουμπλικάνος υποψήφιος.

Οι κάλπες κλείνουν στις 8:00 μ.μ.. Μπορεί ο τηλεοπτικός σταθμός που παρήγγειλε την δημοσκόπηση εξόδου με βάση τα δεδομένα να αναγγείλει στις 8,01 ότι ο Ρεπουμπλικάνος υποψήφιος θα κερδίσει?

Xm12-05*

# Παράδειγμα 12.5

**ΚΑΘΟΡΙΣΜΟΣ**

Το πρόβλημα είναι η περιγραφή του πληθυσμού των ψήφων σε μια πολιτεία. Τα δεδομένα είναι ονομαστικά. Η παράμετρος που ελέγχεται είναι η αναλογία ενός εκ των υποψηφίων έστω του Ρεπουμπλικάνου υποψηφίου.

Για να μπορεί ο τηλεοπτικός σταθμός να αναγγείλει τον Ρεπουμπλικάνο υποψήφιο ως νικητή θα πρέπει να ελέγξει την υπόθεση

$$H_1: p > .50$$

Και ως εκ τούτου η μηδενική υπόθεση γίνεται:

$$H_0: p = .50$$

# Παράδειγμα 12.5

**ΚΑΘΟΡΙΣΜΟΣ**

Η κατανομή είναι κατά προσέγγιση η τυποποιημένη κανονικής κατανομή με επίπεδο σημαντικότητας 5%.
Επομένως $z > z_\alpha = z_{0.05} = 1{,}645$

και $$\hat{p} = \frac{x}{n} = \frac{407}{765} = 0{,}532$$

Ο στατιστικός έλεγχος είναι:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0{,}532 - 0{,}5}{\sqrt{0{,}5(1-0{,}5)/765}} = 1{,}77$$

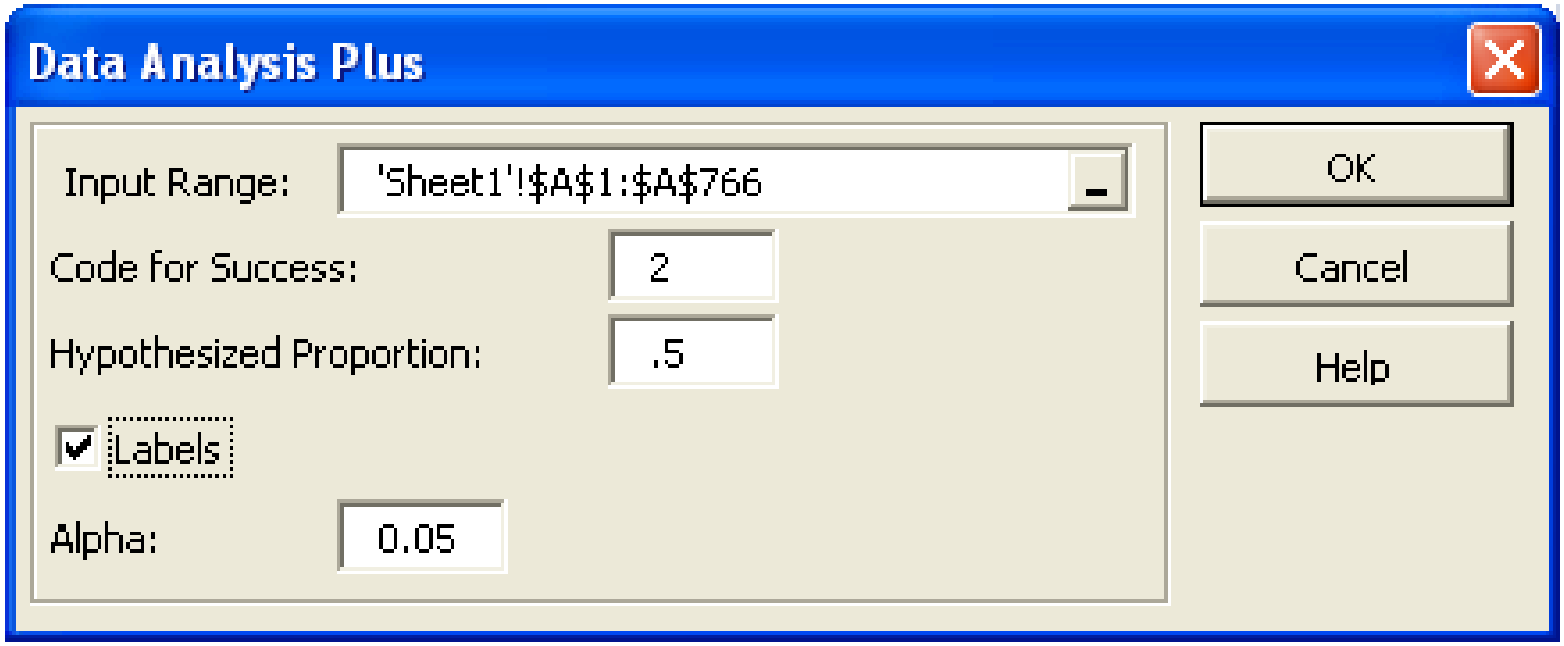
Example 12.5 Manual calculations

Για υπολογισμούς μέσω excel δείτε την συνέχεια:

# Παράδειγμα 12.5

**ΥΠΟΛΟΓΙΣΜΟΣ**

Click Add-Ins, Data Analysis Plus, z-test: Proportion



# Παράδειγμα 12.5

**ΥΠΟΛΟΓΙΣΜΟΣ**

| | A | B | C | D |
|---|---|---|---|---|
| 1 | **z-Test: Proportion** | | | |
| 2 | | | | |
| 3 | | | | *Votes* |
| 4 | Sample Proportion | | | 0.532 |
| 5 | Observations | | | 765 |
| 6 | Hypothesized Proportion | | | 0.5 |
| 7 | z Stat | | | 1.77 |
| 8 | P(Z<=z) one-tail | | | 0.0382 |
| 9 | z Critical one-tail | | | 1.6449 |
| 10 | P(Z<=z) two-tail | | | 0.0764 |
| 11 | z Critical two-tail | | | 1.96 |

# Παράδειγμα 12.5

**ΕΡΜΗΝΕΙΑ**

Σε επίπεδο σημαντικότητας 5% μπορούμε να απορρίψουμε την μηδενική υπόθεση και να συμπεράνουμε ότι υπάρχουν επαρκείς αποδείξεις που δείχνουν ότι ο Ρεπουμπλικάνος υποψήφιος θα κερδίσει την πολιτεία.

Ωστόσο, αυτό είναι η σωστή απόφαση;

Όσοι γνωρίζουν Αγγλικά μπορούν να δουν στην συνέχεια τι πραγματικά συνέβη.

# Παράδειγμα 12.5

**INTERPRET**

One of the key issues to consider here is the cost of Type I and Type II errors.

A Type I error occurs if we conclude that the Republican will win when in fact he has lost.

# Παράδειγμα 12.5

INTERPRET

Such an error would mean that a network would announce at 8:01 P.M. that the Republican has won and then later in the evening would have to admit to a mistake.

If a particular network were the only one that made this error it would cast doubt on their integrity and possibly affect the number of viewers.

# Παράδειγμα 12.5

**INTERPRET**

This is exactly what happened on the evening of the U. S. presidential elections in November 2000.

Shortly after the polls closed at 8:00 P.M. all the networks declared that the Democratic candidate Albert Gore would win in the state of Florida.

A couple of hours later, the networks admitted that a mistake had been made and the Republican candidate George W. Bush had won.

# Παράδειγμα 12.5

INTERPRET

Several hours later they again admitted a mistake and finally declared the race too close to call.

Fortunately for each network all the networks made the same mistake.

However, if one network had not done this it would have developed a better track record, which could have been used in future advertisements for news shows and likely drawn more viewers.

Considering the costs of Type I and II errors it would have been better to use a 1%significance level.

# Selecting the Sample Size

When we introduced the sample size selection method to estimate a mean in Section 10.3, we pointed out that the sample size depends on the confidence level and the bound on the error of estimation that the statistics practitioner is willing to tolerate.

When the parameter to be estimated is a proportion the bound on the error of estimation is

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

# Selecting the Sample Size

Solving for n we produce the required sample size to estimate p and where B is the bound on the error of Estimation

$$n = \left( \frac{z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})}}{B} \right)^2$$

Unfortunately we do not know the value of $\hat{p}$

# Selecting the Sample Size

Two methods – in each case we choose a value for $\hat{p}$ then solve the equation for n.

Method 1 : no knowledge of even a rough value of $\hat{p}$ This is a 'worst case scenario' so we substitute $\hat{p} = .50$

Method 2 : we have some idea about the value of $\hat{p}$ This is a better scenario and we substitute in our estimated $\hat{p}$ value.

# Selecting the Sample Size

Method 1 :: no knowledge of value of $\hat{p}$, use 50%:

$$n = \left( \frac{1.96\sqrt{.50(1-.50)}}{.03} \right)^2 = 1{,}068$$

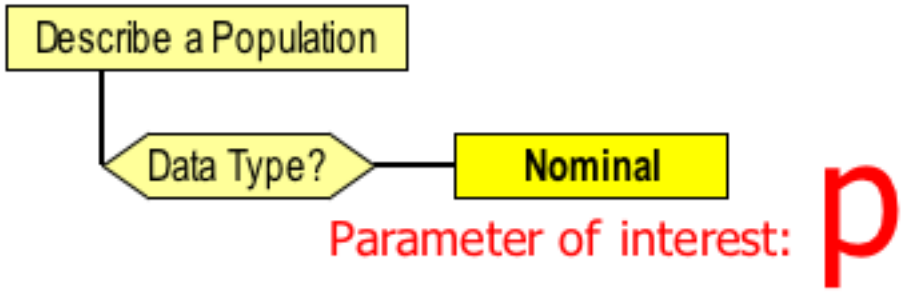
Method 2 :: some idea about a possible $\hat{p}$ value, say 20%:

$$n = \left( \frac{1.96\sqrt{.20(1-.20)}}{.03} \right)^2 = 683$$

Thus, we can sample fewer people if we already have a reasonable estimate of the population proportion before starting.

# Εντοπισμός παραγόντων

Παράγοντες που προσδιορίζουν τον z-έλεγχο και τον εκτιμητή διαστήματος του ***p***:




Test Statistic:

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

Interval Estimator:

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

Required Conditions: $np \geq 5$ *and* $n(1-p) \geq 5$ *(for test)*

$n\hat{p} \geq 5$ *and* $n(1-\hat{p}) \geq 5$ *(for estimate)*

# Διάγραμμα ροής μεθοδολογιών κεφαλαίου