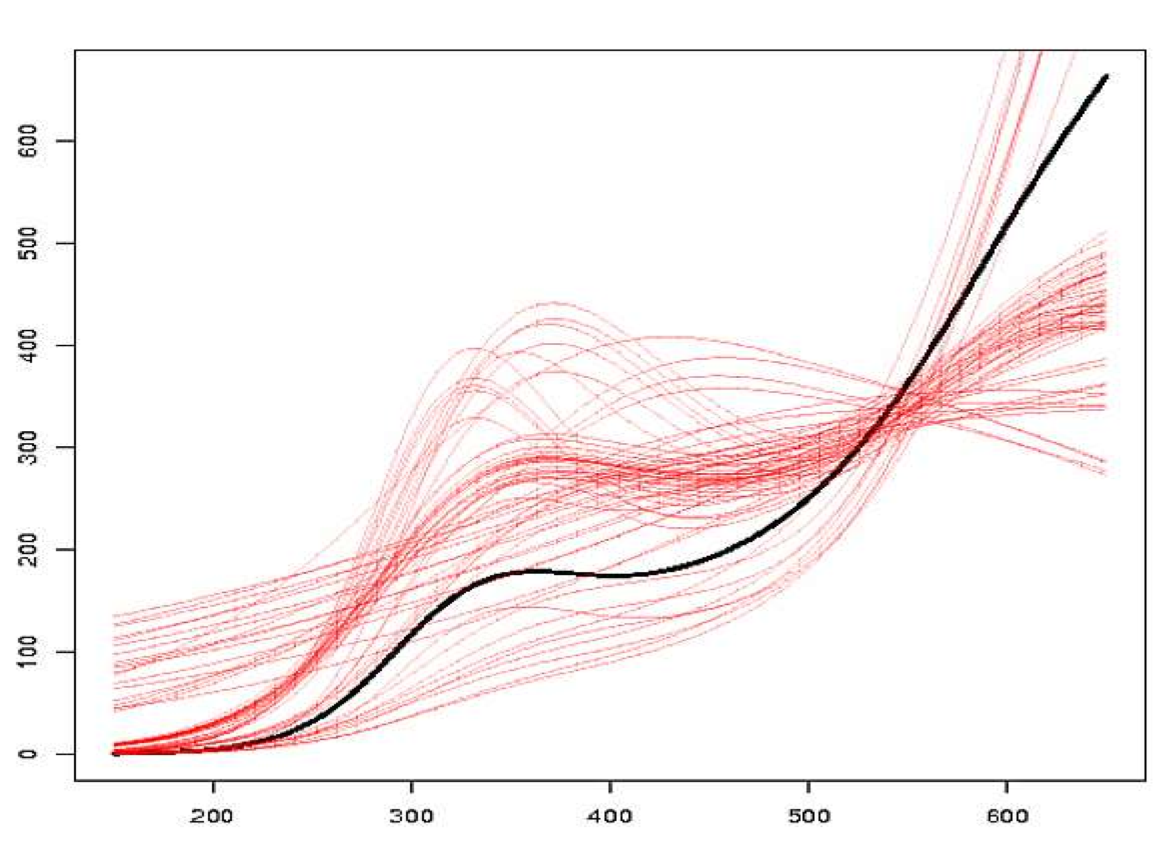
Διάλεξη 4

Εκτίμηση και Διαστήματα Εμπιστοσύνης



ΣΤΟΧΟΙ

1. Ορισμός *εκτίμηση σημείου*.
2. Ορισμός *επιπέδου εμπιστοσύνης*.
3. Κατασκευή διαστήματος εμπιστοσύνης για τη μέση τιμή του πληθυσμού, όταν η τυπική απόκλιση του πληθυσμού είναι γνωστή.
4. Κατασκευή διαστήματος εμπιστοσύνης για την μέση τιμή του πληθυσμού, όταν η τυπική απόκλιση του πληθυσμού είναι άγνωστη.
5. Κατασκευή διαστήματος εμπιστοσύνης για την αναλογία του πληθυσμού.
6. Καθορισμός μεγέθους του δείγματος.

Δειγματοληψία και Εκτιμήσεις

Γιατί διεξάγουμε δειγματοληψία;

1. Το να επικοινωνήσουμε με το σύνολο του πληθυσμού είναι χρονοβόρα *διαδικασία*.
2. Το *κόστος* μελέτης όλων των στοιχείων ενός πληθυσμού είναι *απαγορευτικό*.
3. Υπάρχει *φυσική αδυναμία* ελέγχου όλων των στοιχείων ενός πληθυσμού.
4. Τα *αποτελέσματα βάσει ενος δείγματος* είναι *επαρκή*.

Εκτίμηση Σημείου VS Εκτίμηση Διαστήματος Εμπιστοσύνης

- Ένα σημείο εκτίμησης είναι μια μεμονωμένη τιμή (σημείο) η οποία υπολογίζεται από ένα δείγμα και χρησιμοποιείται για την εκτίμηση της τιμής του πληθυσμού.
- Μια εκτίμηση διαστήματος εμπιστοσύνης είναι ένα εύρος τιμών που κατασκευάζεται από το δείγμα, έτσι ώστε η παράμετρος του πληθυσμού είναι πιθανό να προκύψει εντός αυτού του διαστήματος σε μια προκαθορισμένη πιθανότητα. Η προκαθορισμένη αυτή πιθανότητα ονομάζεται επίπεδο εμπιστοσύνης.

Ποιοι είναι παράγοντες που καθορίζουν το πλάτος του διαστήματος εμπιστοσύνης;

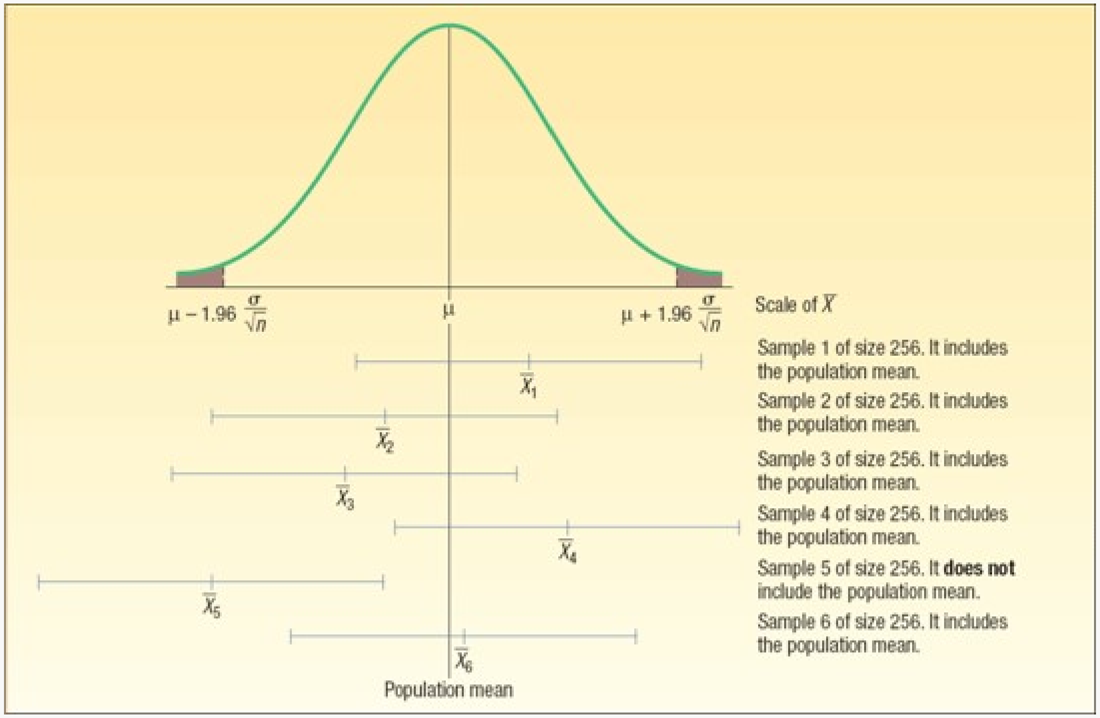
1.Το μέγεθος του δείγματος, ***n***.

2.Η μεταβλητότητα του πληθυσμού, συνήθως το **σ** εκτιμάται από το ***s***.

3.Το επιθυμητό επίπεδο εμπιστοσύνης.

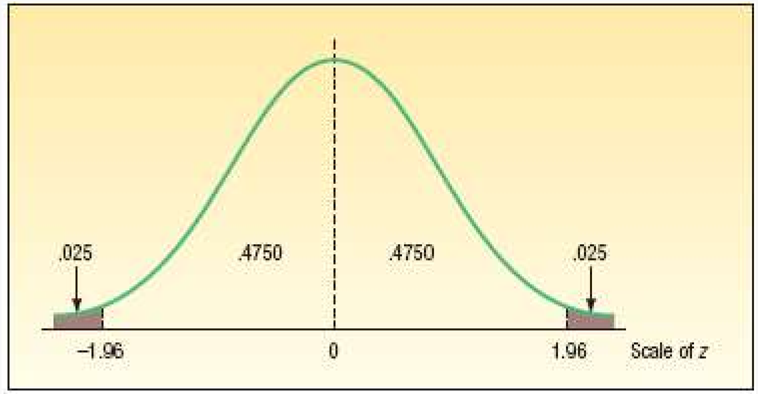
Διάστημα Εμπιστοσύνης - Ερμηνεία

Για ένα διάστημα εμπιστοσύνης 95%, το 95% των όμοια κατασκευασμένων διαστημάτων θα περιέχουν την εκτιμημένη παράμετρο. Επίσης το 95% των δειγματικών μέσων για ένα συγκεκριμένο μέγεθος δείγματος θα βρίσκεται εντός 1.96 τυπικών αποκλίσεων του υποθετικού πληθυσμού.



Πως προκύπτει η τιμή z για ένα δεδομένο Διάστημα Εμπιστοσύνης

Το 95% του διαστήματος εμπιστοσύνης αναφέρεται στο μεσαίο 95% της κατανομής των παρατηρήσεων. Ως εκ τούτου, το υπόλοιπο 5% κατανέμεται ισομερώς μεταξύ των δύο ουρών.



Following is a portion of Appendix B.1.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

Εκτίμηση Σημείου και Εκτίμηση Διαστήματος Εμπιστοσύνης για τον μέσο όρο – το σ είναι γνωστό

CONFIDENCE INTERVAL FOR POPULATION MEAN WITH σ KNOWN $\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$

1. Το πλάτος του διαστήματος καθορίζεται από το επίπεδο εμπιστοσύνης και το μέγεθος του τυπικού σφάλματος του μέσου όρου.
2. Το τυπικό σφάλμα επηρεάζεται από δύο τιμές:
   - Την τυπική απόκλιση
   - Τον αριθμο των παρατηρήσεων του δείγματος

ΠΑΡΑΔΕΙΓΜΑ

Η American Management Association επιθυμεί να έχει πληροφορίες σχετικά με το μέσο εισόδημα των μεσαίων στελεχών στον κλάδο του λιανικού εμπορίου. Ένα τυχαίο δείγμα 256 διευθυντών φανερώνει μια μέση τιμή δείγματος των 45.420 δολαρίων. Η τυπική απόκλιση του πληθυσμού είναι 2.050 δολάρια. Ο σύλλογος θα ήθελε απαντήσεις στα ακόλουθα ερωτήματα:

1. Ποιος είναι ο μέσος όρος του πληθυσμού;

   Σε αυτήν την περίπτωση δεν ξέρουμε. Αλλά ξέρουμε ότι ο μέσος όρος του δείγματος είναι $45,420. Συνεπώς, καλύτερη εκτίμηση μας για την άγνωστη τιμή του πληθυσμού είναι το αντίστοιχο στατιστικό δείγμα.

2. Ποιο θα είναι ένα εύλογο εύρος τιμών για τον μέσο όρο του πληθυσμού? (Χρησιμοποιήστε το 95% διάστημα εμπιστοσύνης)

   $$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = \$45{,}420 \pm 1.96\frac{\$2{,}050}{\sqrt{256}} = \$45{,}420 \pm \$251$$

   Τα όρια του διαστήματος είναι $45,169 και $45,671

   Το ±$251 αναφέρεται ως περιθώριο σφάλματος.

3. Τι σημαίνουν αυτά τα αποτελέσματα;

   Εάν συλλέξουμε πολλά δείγματα μεγέθους 256 στελεχών, και για κάθε δείγμα υπολογίσουμε το μέσο όρο και στη συνέχεια κατασκευάσουμε το 95% διάστημα εμπιστοσύνης, θα μπορούσαμε να περιμένουμε περίπου 95% αυτών των διαστημάτων εμπιστοσύνης να περιέχουν την πληθυσμιακή τιμή.

Η Πληθυσμιακή τυπική απόκλιση (σ) είναι άγνωστη – Η *t*-κατανομή

Στις περισσότερες περιπτώσεις δειγματοληψίας η τυπική απόκλιση πληθυσμού (**σ**) δεν είναι γνωστή. Παρακάτω είναι μερικά παραδείγματα όπου είναι απίθανο οι τυπικές αποκλίσεις του πληθυσμού να είναι γνωστές.

1. Ο Πρύτανης του Business College θέλει να υπολογίσει για φοιτητές πλήρους φοίτησης το μέσο αριθμό των ωρών που απασχολούνται σε αμειβόμενες θέσεις ανά εβδομάδα. Επιλέγει ένα δείγμα 30 φοιτητών με τους οποίους έρχεται σε επαφή και τους ρωτά πόσες ώρες εργάστηκαν την τελευταία εβδομάδα.

2. Ο Πρύτανης θέλει να εκτιμήσει την απόσταση που διανύει ένας τυπικός φοιτητής για να πάει στην τάξη. Επιλέγοντας ένα δείγμα 40 φοιτητών έρχεται σε επαφή με τον καθένα και καθορίζει την απόσταση της διαδρομής από το σπίτι του κάθε μαθητή στην πανεπιστημιούπολη.

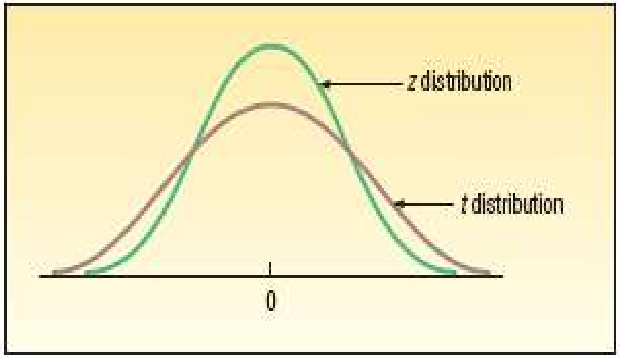
ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΗΣ *t*-Κατανομής

1. Είναι όπως η κατανομή **z,** μια συνεχής κατανομή.

2. Είναι όπως η κατανομή **z,** σε σχήμα καμπάνας και συμμετρική.

3. Δεν υπάρχει μόνο μία κατανομή *t*, αλλά μια οικογένεια κατανομών *t*. Όλες οι *t* κατανομές έχουν μέσο όρο 0 αλλά η τυπική τους απόκλιση διαφέρει ανάλογα με το μέγεθοςτου δείγματος, ***n***.

4. Η *κατανομή t* είναι περισσότερο απλωμένη και επίπεδη στο κέντρο από την τυπική κανονική κατανομή. Παρόλα αυτά όσο το δείγμα μεγαλώνει, η κατανομή *t* προσεγγίζει την τυπική κανονική κατανομή.



Εκτιμήσεις Διαστήματος Εμπιστοσύνης για τον Μέσο Όρο

Χρησιμοποιώντας την Z-Κατανομή

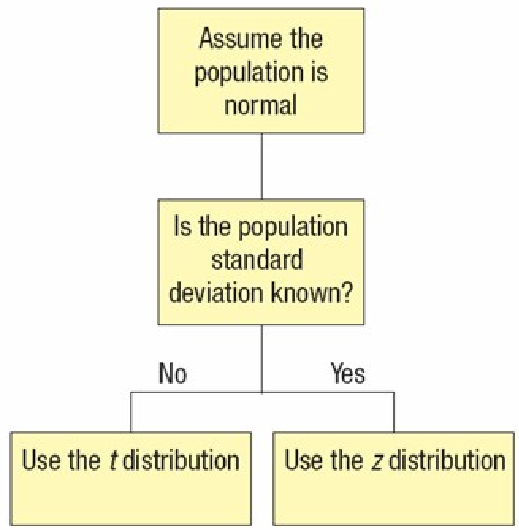
Αν τυπική απόκλιση του πληθυσμού είναι γνωστή ή το δείγμα είναι μεγαλύτερο από 30.

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$$

Χρησιμοποιώντας την *t*-Κατανομή

Αν τυπική απόκλιση του πληθυσμού είναι άγνωστη και το δείγμα είναι μικρότερο από 30.

$$\bar{X} \pm t\frac{s}{\sqrt{n}}$$



Determining When to Use the *z* Distribution or the *t* Distribution

Εκτιμήσεις Διαστήματος Εμπιστοσύνης για τον Μέσο Όρο – Παράδειγμα με την *t*-Κατανομή

ΠΑΡΑΔΕΙΓΜΑ

Ένας κατασκευαστής ελαστικών επιθυμεί να διερευνήσει τη διάρκεια ζωής του πέλματος των ελαστικών του. Ένα δείγμα 10 ελαστικών χρησιμοποιούμενο για 50.000 μίλια έδειξε ότι κατά μέσο όρο απομένει πέλμα 0.32 ιντσών με τυπική απόκλιση 0.09 ιντσών.

- Κατασκευάστε το 95% διάστημα εμπιστοσύνης για το μέσο όρο του πληθυσμού.
- Θα ήταν λογικό για τον κατασκευαστή να συμπεράνει ότι μετά από 50.000 μίλια το μέσο πέλμα που απομένει είναι 0.30 ίντσες;

	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.100	0.050	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	0.20	0.10	0.05	0.02	0.01
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	$\boxed{2.262}$	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

Compute the C.I.
using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,10-1}\frac{s}{\sqrt{n}}$$

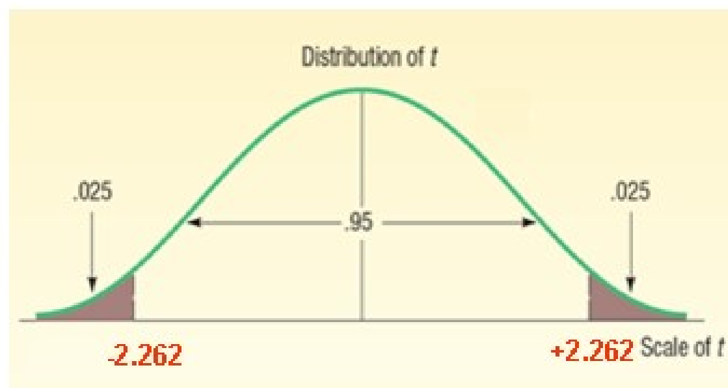
$$= 0.32 \pm t_{.025,9}\frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 2.262\frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

Conclude : the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches.



Διάστημα Εμπιστοσύνης για αναλογία (π)

Ο διευθυντής των υπηρεσιών σταδιοδρομίας στο Southern Technical Institute αναφέρει ότι το 80% των αποφοίτων του, εισέρχεται στην αγορά εργασίας σε θέσεις που σχετίζεται με τον τομέα των σπουδών τους.

1. Ένας εκπρόσωπος της εταιρείας ισχυρίζεται ότι το 45% των πωλήσεων της Burger King γίνονται στο drive-through window.
2. Μια έρευνα για τις κατοικίες στην περιοχή του Σικάγου έδειξε ότι το 85% των νέων κατασκευών έχει κεντρικό κλιματισμό.
3. Μια πρόσφατη έρευνα παντρεμένων ανδρών μεταξύ 35 και 50 έδειξε ότι το 63% πιστεύουν ότι και οι δύο σύντροφοι θα πρέπει να συνεισφέρουν οικονομικά στο νοικοκυριό.

Χρησιμοποιώντας Κανονική Κατανομή για την προσέγγιση της Διωνυμικής Κατανομής

Για να αναπτύξουμε ένα διάστημα εμπιστοσύνης για μία αναλογία, θα πρέπει να πληρούνται

1. Οι ακόλουθες διωνυμικές συνθήκες:
   α. Τα δεδομένα είναι αποτέλεσμα των μετρήσεων.
   β. Υπάρχουν μόνο δύο πιθανά αποτελέσματα.
   γ. Η πιθανότητα επιτυχίας παραμένει η ίδια από δοκιμή σε δοκιμή.
   δ. Οι δοκιμές είναι ανεξάρτητες. Αυτό σημαίνει ότι το αποτέλεσμα σε μία δοκιμή δεν επηρεάζει το αποτέλεσμα σε μία άλλη.

2. Οι τιμές n*π και n*(1-π) θα πρέπει και οι δύο να είναι μεγαλύτερες ή ίσες από με 5. Αυτή η συνθήκη μας επιτρέπει να επικαλεστούμε το κεντρικό οριακό θεώρημα και την τυπική κανονική κατανομή, δηλαδή την Z, για να υπολογισθεί ένα διάστημα εμπιστοσύνης.

SAMPLE PROPORTION $$p = \frac{X}{n}$$

CONFIDENCE INTERVAL FOR A POPULATION PROPORTION $$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

Διάστημα Εμπιστοσύνης για αναλογία πληθυσμού – Παράδειγμα

ΠΑΡΑΔΕΙΓΜΑ

Η ένωση που εκπροσωπεί τους Bottle Blowers of America (BBA) εξετάζει μια πρόταση για τη συγχώνευση με την Teamsters Union. Σύμφωνα με το καταστατικό της BBA, τουλάχιστον τα τρία-τέταρτα των μελών της ένωσης θα πρέπει να εγκρίνει οποιαδήποτε συγχώνευση. Ένα τυχαίο δείγμα 2.000 σημερινών μελών της BBA δείχνει ότι 1.600 πρόκειται να υπερψηφίσουν την πρόταση συγχώνευσης. Ποια είναι η εκτίμηση της αναλογίας για τον πληθυσμό;

Κατασκευάστε το 95% διάστημα εμπιστοσύνης για την αναλογία του πληθυσμού. Βάσει του δείγματος, μπορείτε να συμπεράνετε ότι το υφίσταται το απαραίτητο ποσοστό των μελών BBA για να γίνει η συγχώνευση; Γιατί;

First, compute the sample proportion :

$$p = \frac{x}{n} = \frac{1{,}600}{2000} = 0.80$$

Compute the 95% C.I.

$$\text{C.I.} = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2{,}000}} = .80 \pm .018$$

$$= (0.782,\ 0.818)$$

Conclude : The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership .

Συντελεστής Διορθωσης Πεπερασμένου Πληθυσμού

- Ένας πληθυσμός που έχει σταθερό άνω όριο λέγεται πεπερασμένο.
- Για έναν πεπερασμένο πληθυσμό, όπου συνολικός αριθμός των αντικειμένων είναι N και το μέγεθος του δείγματος είναι n, η ακόλουθη προσαρμογή γίνεται στα τυπικά σφάλματα των μέσων του δείγματος και της αναλογίας:

Τυπικό σφάλμα του Μέσου Όρου

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Τυπικό σφάλμα μιας Αναλογίας

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

- Ωστόσο, αν n/N <.05, ο συντελεστής διόρθωσης πεπερασμένου πληθυσμού μπορεί να αγνοηθεί. Γιατί; Δείτε τι συμβαίνει με την τιμή του συντελεστή διόρθωσης στον πίνακα που ακολουθεί, όταν το κλάσμα n/N μικραίνει.

Όταν το n/N μικραίνει ο ΣΔΠΠ προσεγγίζει το 1.

Finite-Population Correction Factor for Selected Samples When the Population Is 1,000

Sample Size	Fraction of Population	Correction Factor
10	.010	.9955
25	.025	.9879
50	.050	.9752
100	.100	.9492
200	.200	.8949
500	.500	.7075

Δ.Ε. για το Μέσο Όρο με ΣΔΠΠ - Παράδειγμα

ΠΑΡΑΔΕΙΓΜΑ

Υπάρχουν μόνο 250 οικογένειες στην Πενσυλβάνια. Ένα τυχαίο δείγμα των 40 οικογενειών έδειξε ότι η μέση ετήσια συνεισφορά τους στην εκκλησία ήταν $ 450 και η τυπική απόκλιση ήταν $ 75.

Θα μπορούσε ο μέσος όρος του πληθυσμού να είναι $445 ή $425?

Ποιος είναι ο μέσος όρος του πληθυσμού; Ποια είναι η καλυτερη εκτίμηση για το μέσου του πληθυσμού;

<u>Δεδομένα του προβλήματος:</u>

$N = 250$

$n = 40$

$s = 75$

Αφού $n/N = 40/250 = 0.16$, πρέπει να χρησιμοποιηθεί ο συντελεστής διόρθωσης πεπερασμένου πληθυσμού.

Η τυπική απόκλιση του πληθυσμού δεν είναι γνωστή, επομένως χρησιμοποιούμε την t-κατανομή (μπορεί να χρησιμοποιηθεί και η z αφού n> 30)

$$\bar{X} \pm t \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \$450 \pm t_{.10/2,40-1} \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm 1.685 \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm \$19.98 \sqrt{.8434}$$

$$= \$450 \pm \$18.35$$

$$= (\$431.65, \quad \$468.35)$$

Είναι πιθανό ο μέσος όρος του πληθυσμού να είναι μεγαλύτερος από $431.65 αλλά μικρότερος από $468.35.

Με άλλα λόγια, θα μπορούσε ο μέσος του πληθυσμού να είναι $445;

Ναι, αλλά δεν είναι πιθανό να είναι $425 γιατί

η τιμή $445 είναι εντός του διαστήματος εμπιστοσύνης και

η τιμή $425 είναι εκτός του διαστήματος εμπιστοσύνης.

Επιλέγοντας το κατάλληλο μέγεθος δείγματος

Υπάρχουν 3 παράγοντες που καθορίζουν το μέγεθος του δείγματος, εκ των οποίων κανένας δεν έχει άμεση σχέση με το μέγεθος του πληθυσμού.

- Το επιθυμητό επίπεδο εμπιστοσύνης.
- Το περιθώριο λάθους που αποδέχεται ο ερευνητής.
- Η διακύμανση του πληθυσμό που μελετάται.

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

where:

n is the size of the sample.

z is the standard normal value corresponding to the desired level of confidence.

σ is the population standard deviation.

E is the maximum allowable error.

ΠΑΡΑΔΕΙΓΜΑ

Ένας φοιτητής στη δημόσια διοίκηση θέλει να προσδιορίσει το μέσο χρηματικό ποσό των μελών των δημοτικών συμβουλίων το οποίο κερδίζουν μηνιαίως (ως αμοιβή για να είναι μέλη του συμβουλίου).

Το σφάλμα στην εκτίμηση του μέσου θα πρέπει να είναι λιγότερο από \$ 100, για το 95% επίπεδο εμπιστοσύνης. εμπιστοσύνης. Ο φοιτητής βρήκε μία μελέτη του Υπουργείου εργασίας στην οποία σημειώνεται ότι η πληθυσμιακή τυπική απόκλιση είναι \$ 1.000. Ποιο είναι το απαιτούμενο μέγεθος τυο δείγματος;

Δoμένα του προβλήματος:

- E, το μέγιστο επιτρεπόμενο σφάλμα είναι \$100
- Η τιμή του **z** για το 95 % επίπεδο εμπιστοσύνης είναι1.96,
- Η τυπική απόκλιση είναι \$1,000.

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$
$$= \left(\frac{(1.96)(\$1{,}000)}{\$100}\right)^2$$
$$= (19.6)^2$$
$$= 384.16$$
$$= 385$$

Μέγεθος του δείγματος για την εκτίμηση της αναλογίας του Πληθυσμού

$$n = p(1-p)\left(\frac{Z}{E}\right)^2$$

όπου:

n είναι το μέγεθος τυο δείγματος

z είναι η τυπική τιμή της κανονικής κατανομής που αντιστοιχεί στο επιθυμητό επίπεδο εμπιστοσύνης

E είναι το μέγιστο επιθυμητό σφάλμα

ΣΗΜΕΙΩΣΗ:

Χρησιμοποιήστε το p = 0.5 αν δεν ειναι διαθέσιμη η πληροφορία σχετικά με την πιθανότητα επιτυχίας.

ΠΑΡΑΔΕΙΓΜΑ 1

Ο σύλλογος American Kennel θέλει να εκτιμηθεί το ποσοστό των παιδιών που έχουν σκύλο ως κατοικίδιο. Αν ο σύλλογος επιθυμεί η ακρίβεια της εκτίμησης να είναι εντός του 3% της πληθυσμιακής αναλογίας, πόσα παιδιά θα πρέπει να επιλέξει ως δείγμα; Υποθέστε επίπεδο εμπιστοσύνης 95% και ότι ο σύλλογος εκτιμά ότι περίπου το 30% των παιδιών έχουν σκύλο ως κατοικίδιο ζώο.

$$n = (.30)(.70)\left(\frac{1.96}{.03}\right)^2 = 897$$

ΠΑΡΑΔΕΙΓΜΑ 2

Σε μια μελέτη πρέπει να εκτιμηθεί η αναλογία των πόλεων που έχουν ιδιώτες συλλέκτες απορριμμάτων. Ο ερευνητής επιθυμεί το περιθώριο σφάλματος να είναι εντός του 10% της αναλογία του πληθυσμού και επίπεδο εμπιστοσύνης 90%. Δεν υπάρχει διαθέσιμη εκτίμηση για το ποσοστό του πληθυσμού. Ποιο είναι το απαιτούμενο μέγεθος του δείγματος;

$$n = (.5)(1-.5)\left(\frac{1.65}{.10}\right)^2 = 68.0625$$

$n = 69$ cities