ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Ανάλυση Δεδομένων στη Λογιστική και Χρηματοοικονομική

Ενότητα 3η

Συμπεράσματα από Το Απλό Γραμμικό Υπόδειγμα Παλινδρόμησης,

Έλεγχος Υποθέσεων και Διαστήματα Εμπιστοσύνης

Inference on the simple regression model,

Hypothesis testing and confidence intervals

Περιγραφή Ενότητας

- Οι υποθέσεις του απλού γραμμικού μοντέλου παλινδρόμησης
- Έλεγχος υποθέσεων
- Διαστήματα εμπιστοσύνης

Οι Υποθέσεις του Απλού Γραμμικού Μοντέλου Παλινδρόμησης

Οι Υποθέσεις του Κλασσικού Γραμμικού Μοντέλου Παλινδρόμησης
Classical Linear Regression Model (CLRM)

- Το κλασσικό γραμμικό μοντέλο παλινδρόμησης στηρίζεται πάνω σε αρκετές υποθέσεις οι οποίες πρέπει να εκπληρώνονται από τις εκτιμήσεις μας, ώστε το μοντέλο να θεωρείται αξιόπιστο και να μπορεί να χρησιμοποιηθεί, π.χ. για την πρόβλεψη της εξαρτημένης μεταβλητής. Ιδιαίτερη έμφαση δίνεται στο διαταρακτικό όρο, καθώς τα δεδομένα για τις μεταβλητές Y και X είναι δεδομένα (δηλαδή είναι παρατηρήσεις). Ο διαταρακτικός όρος θα πρέπει να πληροί τις ακόλουθες υποθέσεις:

Μαθηματικός Τύπος	Ερμηνεία
1. $E(u_t) = 0$	Ο διαταρακτικός όρος έχει μέσο όρο μηδέν, *The errors have zero mean*
2. $Var(u_t) = \sigma^2$	Η διακύμανση του διαταρακτικού όρου είναι σταθερή και συγκεκριμένη για όλες τις τιμές της ανεξάρτητης μεταβλητής X, *The variance of the errors is constant and finite over all values of* x_t
3. $Cov(u_i,u_j)=0$	Οι διαταρακτικοί όροι είναι στατιστικά ανεξάρτητοι μεταξύ τους (έχουν συνδιακύμανση μηδέν, *The errors are statistically independent of one another*
4. $Cov(u_t,x_t)=0$	Η συνδιακύμανση μεταξύ διαταρακτικών όρων και ανεξάρτητης μεταβλητής είναι μηδέν, *No relationship between the error and corresponding x variate*
5. u_t is normally distributed	Ο διαταρακτικός όρος κατανέμεται σύμφωνα με την κανονική κατανομή.

Υποθέσεις και ιδιότητες των βέλτιστων γραμμικών αμερόληπτων εκτιμητών (Best Linear Unbiased Estimators, BLUE)

- Η πέμπτη υπόθεση (δηλαδή ότι τα κατάλοιπα κατανέμονται σύμφωνα με την κανονική κατανομή) αποτελεί προϋπόθεση εάν θέλουμε να εξάγουμε συμπεράσματα (inferences) για τις πληθυσμιακές τιμές των συντελεστών α και β, βασισμένοι στις δειγματικές τιμές των συντελεστών $\hat{\alpha}$ και $\hat{\beta}$.
- Ενώ εάν οι υποθέσεις 1 έως 4 ισχύουν τότε οι εκτιμητές που χρησιμοποιήσαμε (OLS) θα είναι BLUE (Best Linear Unbiased Estimators). Τι σημαίνει ότι είναι BLUE ;
- "Estimator", δηλαδή εκτιμητής της πραγματικής τιμής του β.
- "Linear", δηλαδή ένας γραμμικός εκτιμητής.
- "Unbiased", δηλαδή, κατά μέσο όρο οι δειγματικές τιμές των α και β θα είναι ίσες με τις πραγματικές (πληθυσμιακές) τιμές τους.
- "Best", δηλαδή ο εκτιμητής κατά OLS έχει την μικρότερη διακύμανση ανάμεσα σε όλους τους πιθανούς γραμμικούς αμερόληπτους εκτιμητές. Το θεώρημα Gauss – Markov αναφέρει ότι ο εκτιμητής ελάχιστων τετραγώνων (OLS) έχει τη χαμηλότερη διακύμανση δειγματοληψίας εντός της κατηγορίας γραμμικών αμερόληπτων εκτιμητών, εάν τα σφάλματα στη γραμμική παλινδρόμηση είναι ασυσχέτιστα, έχουν σταθερές διακυμάνσεις και αναμενόμενη τιμή (μέσο όρο) μηδέν. Τα σφάλματα δεν χρειάζεται να κατανέμονται κανονικά, ούτε χρειάζεται να είναι ανεξάρτητα και πανομοιότυπα κατανεμημένα (i.i.d.), (αλλά χρειάζεται να είναι μη συσχετισμένα μεταξύ τους, με μέσο μηδέν και ομοσκεδαστικά, με πεπερασμένη διακύμανση).

Συνέπεια/Αμεροληψία/Αποτελεσματικότητα
Consistency/Unbiasedness/Efficiency

- Συνέπεια: Οι εκτιμητές των $\hat{\alpha}$ και $\hat{\beta}$ κατά OLS είναι συνεπείς. Αυτό σημαίνει ότι οι εκτιμήσεις τους θα συγκλίνουν στις πραγματικές τους τιμές όσο το δείγμα μεγαλώνει προς το άπειρο. Χρειαζόμαστε αυτές τις υποθέσεις για να το αποδείξουμε αυτό: $E(x_t u_t)=0$ και $Var(u_t)=\sigma^2 < \infty$. Η συνέπεια των εκτιμητών συνεπάγεται ότι:

$$\lim_{T\to\infty} \Pr\left[\left|\hat{\beta}-\beta\right| > \delta\right] = 0 \quad \forall\delta > 0$$

- Αμεροληψία: Οι εκτιμητές των $\hat{\alpha}$ και $\hat{\beta}$ κατά OLS είναι αμερόληπτοι. Αυτό συνεπάγεται ότι $E(\hat{\alpha})=\alpha$ και $E(\hat{\beta})=\beta$. Άρα κατά μέσο όρο οι εκτιμήσεις μας θα είναι ίσες με τις πραγματικές τιμές. Για να το αποδείξουμε αυτό χρειάζεται να ισχύει αυτή η υπόθεση $E(u_t)=0$. Η αμεροληψία των εκτιμητών είναι πιο ισχυρή συνθήκη από τη συνέπεια τους.

- Αποτελεσματικότητα: Ένας εκτιμητής $\hat{\beta}$ της παραμέτρου β μπορεί να οριστεί ως αποτελεσματικός εάν είναι αμερόληπτος και κανένας άλλος αμερόληπτος εκτιμητής δεν έχει μικρότερη διακύμανση. Εάν ο εκτιμητής είναι αποτελεσματικός, τότε ελαχιστοποιούμε την πιθανότητα να είναι πολύ μακριά από την πραγματική τιμή για την παράμετρο β.

Ακρίβεια και Τυπικά Σφάλματα - Precision and Standard Errors

- Κάθε σύνολο εκτιμήσεων των συντελεστών της παλινδρόμησης $\hat{\alpha}$ and $\hat{\beta}$ είναι δεμένο με το δείγμα που χρησιμοποιήθηκε για την εκτίμηση τους.
- Ανακαλέστε ότι οι εκτιμήσεις για τα α και β δίνονται από τους ακόλουθους τύπους (δηλαδή από τις εκτιμήσεις για τα $\hat{\alpha}$ και $\hat{\beta}$ με βάση το δείγμα που έχουμε:

$$\hat{\beta} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2} \quad \text{και } \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

- Χρειαζόμαστε όμως και ένα μέτρο της αξιοπιστίας της εκτίμησης που έχουμε κάνει για τα $\hat{\alpha}$ και $\hat{\beta}$, δηλαδή για την ακρίβεια της εκτίμησης, αυτό είναι το τυπικό σφάλμα. Με δεδομένες τις υποθέσεις των OLS εκτιμητών που αναφέρθηκαν, τα τυπικά σφάλματα είναι:

$$SE(\hat{\alpha}) = s\sqrt{\frac{\sum x_t^2}{T\sum(x_t - \bar{x})^2}} = s\sqrt{\frac{\sum x_t^2}{T\sum x_t^2 - T^2\bar{x}^2}},$$

$$SE(\hat{\beta}) = s\sqrt{\frac{1}{\sum(x_t - \bar{x})^2}} = s\sqrt{\frac{1}{\sum x_t^2 - T\bar{x}^2}}$$

όπου s η εκτιμημένη (δειγματική) τυπική απόκλιση των καταλοίπων.

Μερικές Παρατηρήσεις στους Εκτιμητές των Τυπικών Σφαλμάτων
Some Comments on the Standard Error Estimators

1. Τα τυπικά σφάλματα SE($\hat{\alpha}$) και SE($\hat{\beta}$) εξαρτώνται από τη δειγματική διακύμανση s^2 (ή τυπική απόκλιση s). Όσο πιο μεγάλη η s^2, τόσο πιο πολύ τα σφάλματα κατανέμονται γύρω από τη μέση τιμή τους και τόσο πιο πολύ η μεταβλητή y θα απέχει από τη μέση τιμή της.
2. Το άθροισμα των τετραγώνων της μεταβλητής x από το μέσο της $\sum(x_t - \bar{x})^2$ εμφανίζεται στον παρονομαστή των τύπων για τα τυπικά σφάλματα, δηλαδή έχει αντίστροφη σχέση με τα τυπικά σφάλματα.
3. Όσο μεγαλύτερο μέγεθος του δείγματος, T, τόσο μικρότερη θα είναι η διακύμανση των συντελεστών.

Μία εισαγωγή στην εξαγωγή στατιστικών συμπερασμάτων
An Introduction to Statistical Inference

- Θέλουμε να εξάγουμε ένα συμπέρασμα σχετικά με τις πραγματικές τιμές (του πληθυσμού) βασισμένοι πάνω στα παρακάτω εκτιμημένα αποτελέσματα (παραμέτρους) μίας παλινδρόμησης:

$$\hat{y}_t = 20.3 + 0.5091 x_t$$
$$(14.38)\,(0.2561)$$

- Το $\hat{\beta} = 0.5091$ είναι μία εκτίμηση της άγνωστης πληθυσμιακής παραμέτρου β. Πόσο αξιόπιστη είναι αυτή η εκτίμηση ;
- Η αξιοπιστία της μπορεί να μετρηθεί με το τυπικό σφάλμα του συντελεστή $\hat{\beta}$.

Έλεγχος Υποθέσεων (Hypothesis Testing)

- Ο έλεγχος υποθέσεων μπορεί να μας βοηθήσει να εξάγουμε συμπεράσματα για την αξιοπιστία (ακρίβεια) των εκτιμημένων συντελεστών για ένα επίπεδο εμπιστοσύνης.
- Στον έλεγχο υποθέσεων έχουμε πάντα δύο υποθέσεις, τη μηδενική υπόθεση H_0 (the null hypothesis) και την εναλλακτική υπόθεση H_1 (alternative hypothesis).
- Η μηδενική υπόθεση αναφέρεται στην υπόθεση που θέλουμε να εξετάσουμε, ενώ η εναλλακτική υπόθεση αναπαριστά όλα τα υπόλοιπα ενδεχόμενα.
- Για παράδειγμα υποθέστε ότι με δεδομένα τα αποτελέσματα παλινδρόμησης στην προηγούμενη διαφάνεια, ενδιαφερόμαστε να εξετάσουμε στατιστικά την υπόθεση εάν η πραγματική (πληθυσμιακή) τιμή για την παράμετρο β είναι πράγματι 0.5 όπως την εκτιμήσαμε με το δείγμα μας. Ο έλεγχος υποθέσεων θα ήταν: $H_0 : \beta = 0.5$ *και* $H_1 : \beta \neq 0.5$

  Το οποίο είναι επίσης γνωστό ως ένα δίπλευρο τεστ.

Μονόπλευρος Έλεγχος Υποθέσεων
One-Sided Hypothesis Tests

- Μερικές φορές μπορεί να περιμένουμε ότι για παράδειγμα $\beta > 0.5$ αντί για $\beta < 0.5$. Σε αυτή την περίπτωση θα πρέπει να κάνουμε ένα μονόπλευρο έλεγχο υποθέσεων:

$$\mathrm{H}_0 : \beta = 0.5$$
$$\mathrm{H}_1 : \beta > 0.5$$

ή εναλλακτικά:

$$\mathrm{H}_0 : \beta = 0.5$$
$$\mathrm{H}_1 : \beta < 0.5$$

- Υπάρχουν δύο τρόποι να γίνει ο έλεγχος υποθέσεων, είτε με ένα τεστ σημαντικότητας (ο πιο κοινός τρόπος) ή με την εκτίμηση ενός διαστήματος εμπιστοσύνης (confidence interval, c.i.).

Η Κατανομή Πιθανότητας των Εκτιμητών Ελαχίστων Τετραγώνων

- Υποθέτουμε ότι $u_t \sim \mathrm{N}(0,\sigma^2)$

- Δεδομένου ότι οι εκτιμητές των ελάχιστων τετραγώνων είναι γραμμικοί συνδυασμοί τυχαίων μεταβλητών
$$\hat{\beta} = \sum w_t y_t$$

- Το σταθμισμένο άθροισμα των κανονικών τυχαίων μεταβλητών κατανέμεται επίσης κανονικά, έτσι, άρα
$$\hat{\alpha} \sim \mathrm{N}(\alpha, \mathrm{Var}(\hat{\alpha}))$$
$$\hat{\beta} \sim \mathrm{N}(\beta, \mathrm{Var}(\hat{\beta}))$$

- **Τι γίνεται εάν τα σφάλματα δεν κατανέμονται κανονικά; Οι εκτιμήσεις παραμέτρων θα εξακολουθούν να κατανέμονται κανονικά;***Υ*

Ναι, εάν ισχύουν οι άλλες υποθέσεις του CLRM και το μέγεθος του δείγματος είναι αρκετά μεγάλο.

- Μπορούμε επίσης να γράψουμε

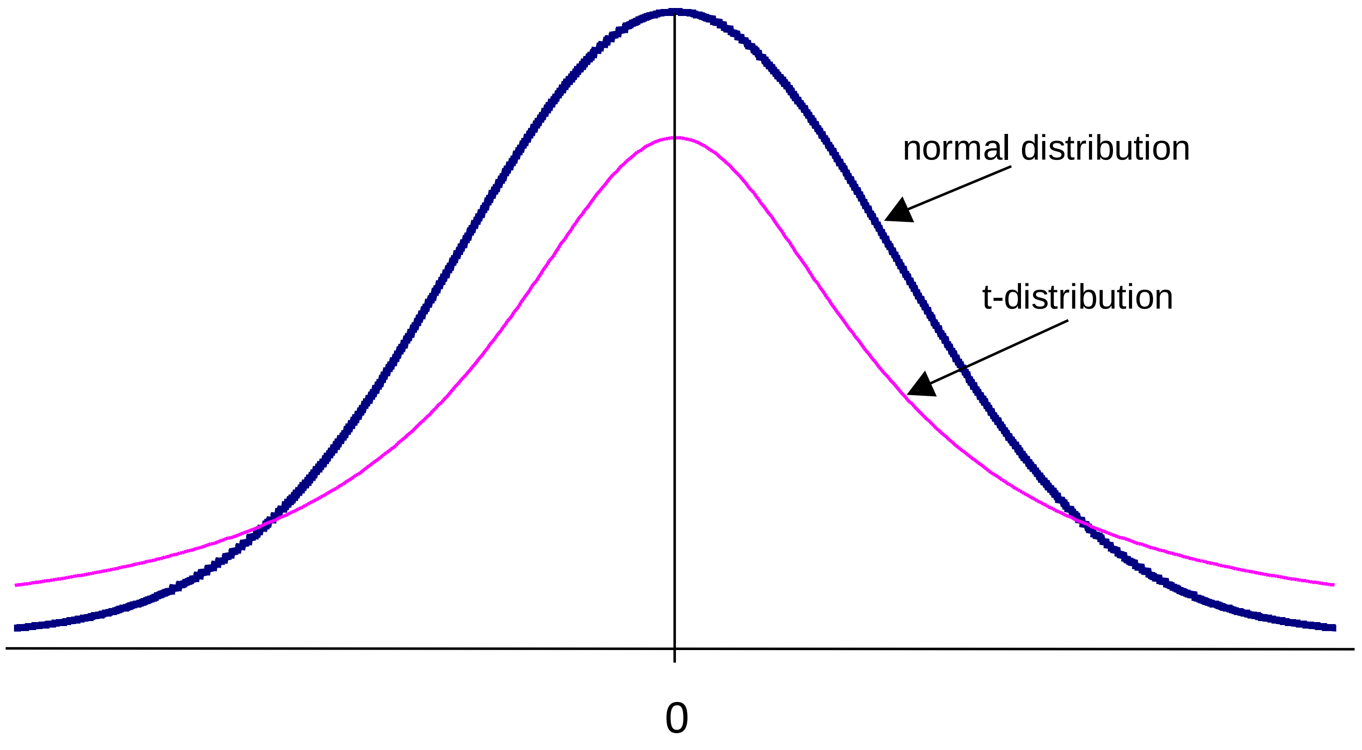
$$\frac{\hat{\alpha}-\alpha}{\sqrt{\operatorname{var}(\alpha)}} \sim N(0,1) \qquad \text{και} \qquad \frac{\hat{\beta}-\beta}{\sqrt{\operatorname{var}(\beta)}} \sim N(0,1)$$

- Αλλά var(α) and var(β) είναι άγνωστες, άρα

$$\frac{\hat{\alpha}-\alpha}{SE(\hat{\alpha})} \sim t_{T-2} \qquad \text{και} \qquad \frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \sim t_{T-2}$$

Η (τυπική) κανονική κατανομή και η κατανομή t

- Σε πολλές περιπτώσεις η πληθυσμιακή τυπική απόκλιση, σ, είναι άγνωστη. Ως αποτέλεσμα τα τυπικά σφάλματα των πληθυσμιακών παραμέτρων α και β είναι άγνωστα. Για αυτό το λόγο, η κατανομή t χρησιμοποιείται σε αυτές τις περιπτώσεις αντί της κανονικής κατανομής.
- Η κατανομή t είναι επίσης συμμετρική, καμπανοειδής και έχει μέσο μηδέν όπως η κανονική κατανομή και η τυπική κανονική κατανομή. Ωστόσο, η κατανομή t είναι πιο πλατιά και έχει μεγαλύτερες ουρές.
- Η κατανομή *t* έχει επίσης βαθμούς ελευθερίας ως T-2.



Έλεγχος Υποθέσεων με Στατιστικά Τεστ (1)
Testing Hypotheses: The Test of Significance Approach

- Ας υποθέσουμε την ακόλουθη γραμμική εξίσωση παλινδρόμησης (όπου t=1,2,...,T):

$$y_t = \alpha + \beta x_t + u_t$$

- Τα βήματα για τον έλεγχο υποθέσεων είναι:

1. Εκτίμηση των $\hat{\alpha}$ and $\hat{\beta}$ και SE($\hat{\alpha}$) και SE($\hat{\beta}$), αντίστοιχα.

2. Υπολογισμός του αντίστοιχου τεστ:

$$test\ statistic = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$$

όπου β^* είναι η τιμή της παραμέτρου β σύμφωνα με την μηδενική υπόθεση (π.χ. συνήθως μηδέν).

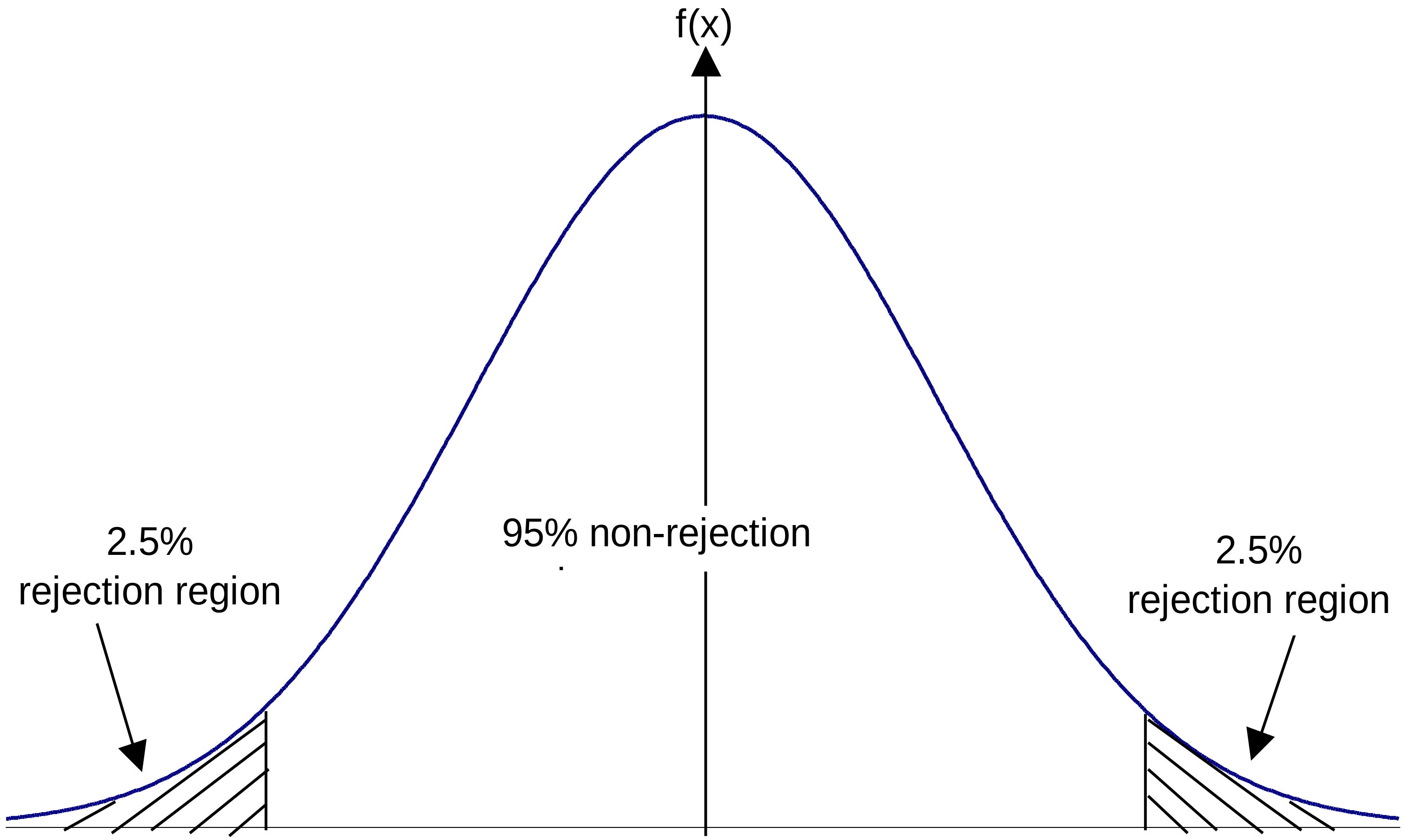
Έλεγχος Υποθέσεων με Στατιστικά Τεστ (2)
Testing Hypotheses: The Test of Significance Approach

3. Χρειαζόμαστε κάποια κατανομή για να τη συγκρίνουμε σε σχέση με την τιμή του στατιστικού τεστ που μόλις υπολογίσαμε. Αυτού του τύπου τα τεστ ακολουθούν την κατανομή t με T-2 βαθμούς ελευθερίας, όπου T ο αριθμός των παρατηρήσεων του δείγματος. Όσο μεγαλύτεροι οι βαθμοί ελευθερίας (το δείγμα), τόσο λιγότερο προσεκτικοί χρειάζεται να είμαστε σε ότι αφορά τη σταθερότητα της αξιοπιστίας των αποτελεσμάτων μας (robustness).

4. Χρειάζεται να διαλέξουμε ένα επίπεδο σημαντικότητας (α), γνωστό ως significance level. Συνήθως χρησιμοποιείται το 5% αλλά συχνές εναλλακτικές είναι το 10% και 1%. Το επίπεδο σημαντικότητας επίσης καλείται μέγεθος του τεστ καθώς αντιπροσωπεύει πόσο μεγάλη είναι η περιοχή (της κατανομής) όπου θα απορρίψουμε (ή δεν θα απορρίψουμε) τη μηδενική υπόθεση. Με άλλα λόγια, το επίπεδο σημαντικότητας δηλώνει ότι θα περιμέναμε ένα αντίθετο αποτέλεσμα (τυχαίο) από το συμπέρασμα που θα καταλήξουμε μόνο με 5% πιθανότητα.

Έλεγχος Υποθέσεων με Στατιστικά Τεστ (3)
Testing Hypotheses: The Test of Significance Approach

5. Δοθέντος του επιπέδου σημαντικότητας (α), μπορούμε να προσδιορίσουμε την περιοχή απόρριψης και μη-απόρριψης της μηδενικής υπόθεσης. Για ένα δίπλευρο τεστ με επίπεδο σημαντικότητας α=5% αυτές περιοχές είναι αντίστοιχα:

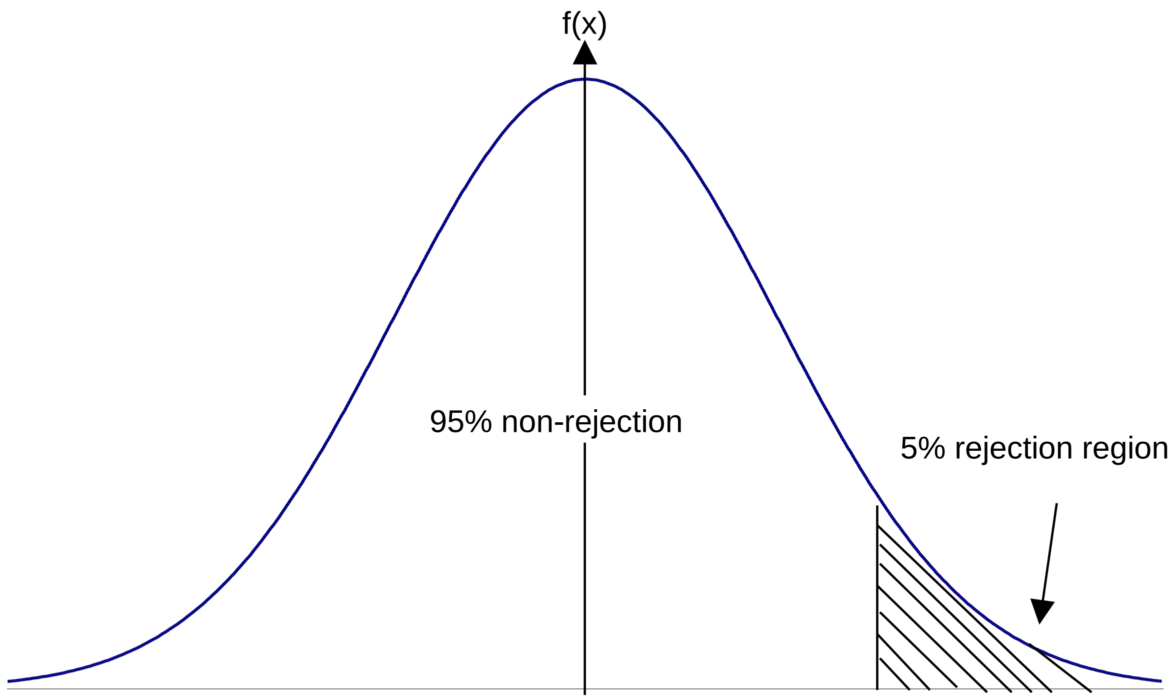


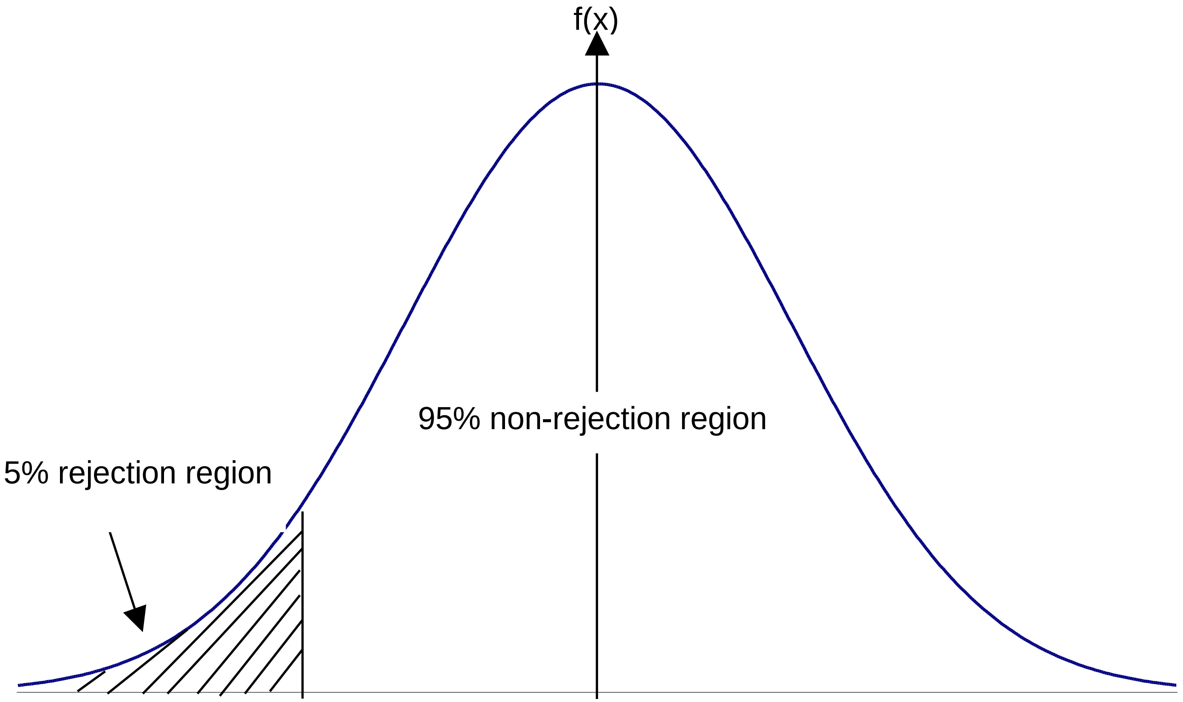
Έλεγχος Υποθέσεων με Στατιστικά Τεστ: Καταλήγοντας σε συμπεράσματα (4)
The Test of Significance Approach: Drawing Conclusions

6. Βρίσκουμε από τους πίνακες της κατανομής t την κριτική τιμή που αντιστοιχεί στο επίπεδο σημαντικότητας (α=5%) ή ισοδύναμα στο επίπεδο εμπιστοσύνης (confidence level = 95%) και στους βαθμούς ελευθερίας (T-2) που αντιστοιχούν στη δική μας περίπτωση. Στη συνέχεια συγκρίνουμε την κριτική τιμή των πινάκων με την τιμή του τεστ που υπολογίσαμε για τον έλεγχο υποθέσεων.

7. Τέλος, εάν η τιμή του τεστ που υπολογίσαμε είναι εντός της περιοχής (-χων) απόρριψης όπως την (τις) είδαμε στην προηγούμενη διαφάνεια, τότε απορρίπτουμε τη μηδενική υπόθεση. Εάν είναι εκτός περιοχής απόρριψης τότε δεν απορρίπτουμε την μηδενική υπόθεση. Παρατηρήστε ότι, για να είμαστε στατιστικά συνεπείς με την ορολογία, ποτέ δεν δεχόμαστε την εναλλακτική υπόθεση (H_1), αλλά μόνο απορρίπτουμε ή δεν απορρίπτουμε την μηδενική υπόθεση.

The Rejection Region for a 1-Sided Test (Upper Tail-Lower Tail)





Έλεγχος Υποθέσεων με Στατιστικά Τεστ: Διάστημα Εμπιστοσύνης

1. Υπολογίστε $\hat{\alpha}$, $\hat{\beta}$ και $SE(\hat{\alpha})$, $SE(\hat{\beta})$

2. Επιλέξτε το επίπεδο σημαντικότητας, α, (π.χ. 5%). Αυτό ισοδυναμεί με επιλογή διαστήματος εμπιστοσύνης (1-α)×100% δηλ. 5% επίπεδο σημασίας = 95% διάστημα εμπιστοσύνης

3. Χρησιμοποιείστε τους t-πίνακες για να βρείτε την κατάλληλη κρίσιμη τιμή, η οποία θα έχει και πάλι βαθμούς ελευθερίας T-2.

4. Το διάστημα εμπιστοσύνης δίνεται από $(\hat{\beta} - t_{crit} \times SE(\hat{\beta}), \hat{\beta} + t_{crit} \times SE(\hat{\beta}))$

5. Εκτελέστε τη δοκιμή: Εάν η υποθετική τιμή β (β*) βρίσκεται εκτός του διαστήματος εμπιστοσύνης, τότε απορρίψτε την μηδενική υπόθεση ότι β = β*, διαφορετικά μην απορρίψετε την μηδενική.

Έλεγχος Υποθέσεων: Ένα Παράδειγμα
Constructing Tests of Significance: An Example

- Τα αποτελέσματα για την προηγούμενη παλινδρόμηση ήταν τα ακόλουθα (ο αριθμός των παρατηρήσεων ήταν T=22):

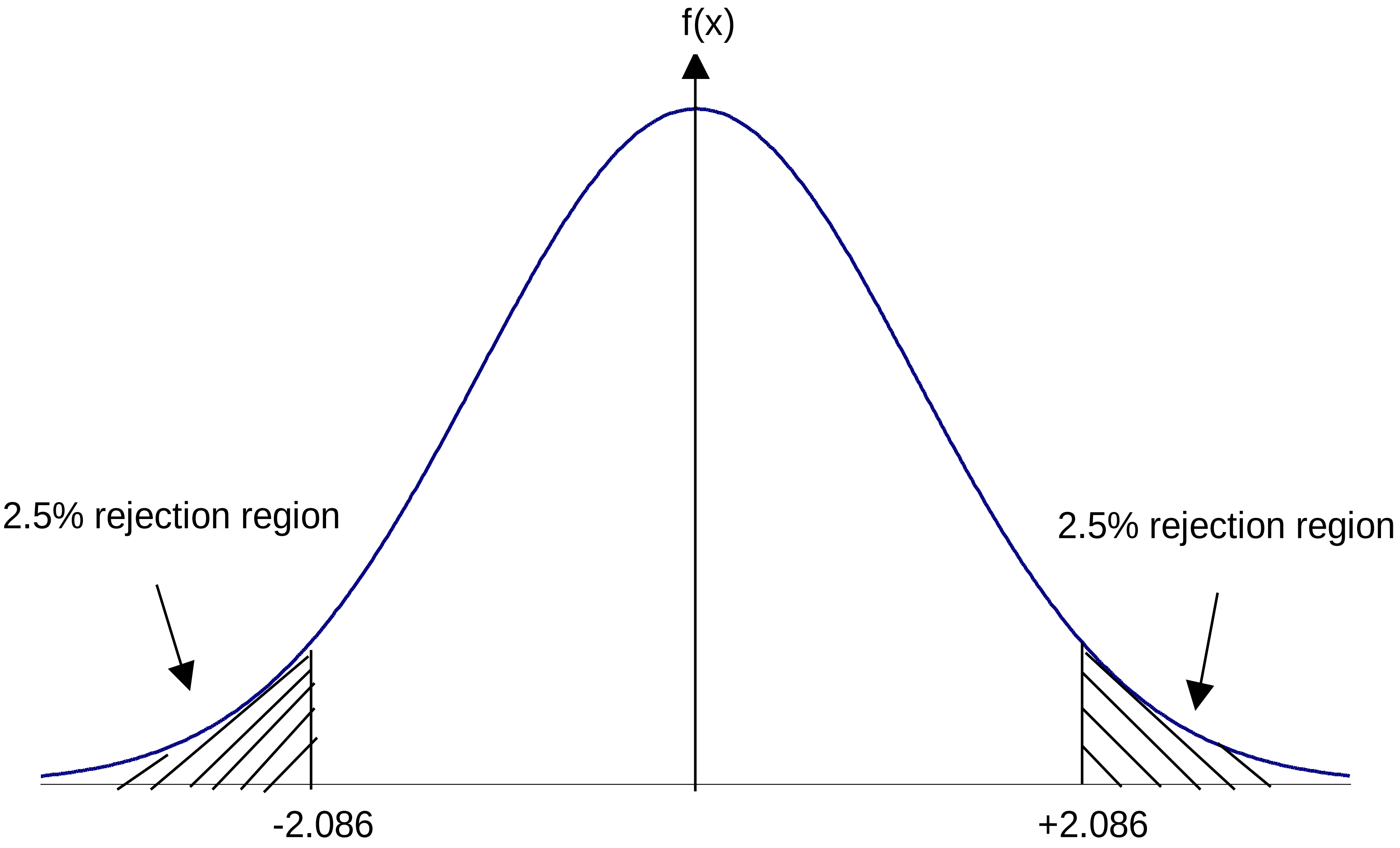
$$\hat{y}_t = 20.3 + 0.5091x_t$$
$$(14.38) \quad (0.2561)$$

- Χρησιμοποιώντας το τεστ σημαντικότητας για τον έλεγχο υποθέσεων, θέλουμε να ελέγξουμε τη μηδενική υπόθεση ότι β=1 έναντι της εναλλακτικής ότι το β≠1 (δίπλευρο τεστ).

- Το πρώτο βήμα είναι να βρούμε την κριτική τιμή (critical value) από τους πίνακες για την κατανομή t, για επίπεδο σημαντικότητας α=5% και για T-2=22-2=20 βαθμού ελευθερίας:

$$t_{critical\ value} = t_{20,2.5\%} = -2.086$$

Οι περιοχές απόρριψης της H_0



Έλεγχος Υποθέσεων με (i) ένα στατιστικό τεστ, (ii) ένα διάστημα εμπιστοσύνης

- Οι Υποθέσεις:

$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

$$test\ stat = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$$
$$= \frac{0.5091 - 1}{0.2561} = -1.917$$

Εφόσον το -1.917 είναι μεγαλύτερο από το -2.086, δεν απορρίπτουμε την H_0.

- Οι Υποθέσεις:

$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

$$\hat{\beta} \pm t_{crit} \times SE(\hat{\beta})$$
$$= 0.5091 \pm 2.086 \times 0.2561$$
$$= (-0.0251, 1.0433)$$

Εφόσον το β=1 (δηλαδή η H_0) ανήκει στο διάστημα εμπιστοσύνης που εκτιμήσαμε (-0.0251 έως 1.0433) δεν απορρίπτουμε την H_0.

Τα λάθη που μπορούμε να κάνουμε στον έλεγχο υποθέσεων
The Errors That We Can Make Using Hypothesis Tests

- Υπάρχουν δύο λάθη που μπορούν να γίνουν:

1. Να απορρίψουμε την H_0 ενώ ήταν αληθής (Type I error).
2. Να μην απορρίψουμε την H_0 ενώ ήταν ψευδής (θα έπρεπε να απορριφθεί), (Type II error).

		Reality	
		H_0 is true	H_0 is false
Result of Test	Significant (reject H_0)	Type I error $= \alpha$	√
	Insignificant (do not reject H_0)	√	Type II error $= \beta$

Η σχέση ανταλλαγής μεταξύ των σφαλμάτων Type I and Type II
The Trade-off Between Type I and Type II Errors

- Η πιθανότητα ενός σφάλματος τύπου Ι είναι ίση με το επίπεδο σημαντικότητας α. Όπως είδαμε αν διαλέξουμε α=5% τότε αυτό σημαίνει ότι υπάρχει μόλις 5% πιθανότητα ένα αποτέλεσμα που εκτιμήσαμε να μην ισχύει και να έχει δημιουργηθεί από τύχη.
- Όμως τι συμβαίνει εάν μειώσουμε το α, ας πούμε στο α=1%. Τότε μειώνουμε τις πιθανότητες του σφάλματος Τύπου Ι και ταυτόχρονα μειώνουμε την πιθανότητα να απορρίψουμε την H_0. Με αυτόν τον τρόπο (μειώνοντας την πιθανότητα να απορρίψουμε την H_0) αυξάνουμε την πιθανότητα να κάνουμε ένα σφάλμα Τύπου ΙΙ:

Reduce size of test → more strict criterion for rejection → reject null hypothesis less often

↗ less likely to falsely reject

↘ more likely to incorrectly not reject

- Με αυτόν τον τρόπο υπάρχει πάντα μία σχέση ανταλλαγής μεταξύ των σφαλμάτων Τύπου Ι και ΙΙ. Ο μόνος τρόπος να τα μειώσουμε και τα δύο σφάλματα ταυτόχρονα είναι να αυξήσουμε το δείγμα μας.

Μία ειδική (και πολύ δημοφιλής) περίπτωση ελέγχου υποθέσεων: το t-ratio
A Special Type of Hypothesis Test: The *t*-ratio

- Ανακαλέστε ότι ο γενικός τύπος για έναν έλεγχο υποθέσεων χρησιμοποιώντας το t-test (κατανομή t) ήταν:

$$test\ statistic = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

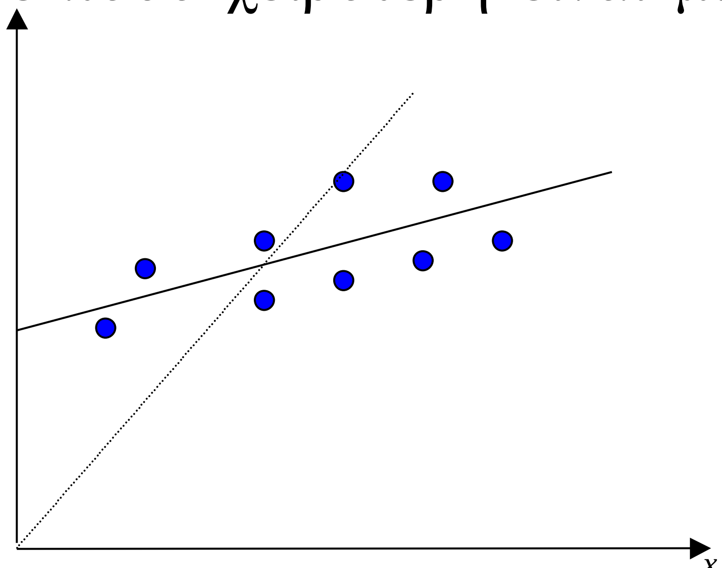
- Εάν ο έλεγχος υποθέσεων είναι: $H_0 : \beta_i = 0$, $H_1 : \beta_i \neq 0$, τότε αυτή η ειδική περίπτωση είναι το *t*-ratio test (δηλαδή απαντά εάν ένας συντελεστής είναι στατιστικά σημαντικός ή όχι). Ισοδύναμα το t-ratio test υποθέτει ότι $\beta_i^* = 0$. Άρα το στατιστικό τεστ γίνεται:

$$test\ stat = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

- Έτσι, μπορούμε να διακρίνουμε ότι το t-ratio test είναι απλά η διαίρεση του συντελεστή που εκτιμήσαμε από μία παλινδρόμηση, $\widehat{\beta}_i$, με το τυπικό του σφάλμα $SE(\widehat{\beta}_i)$.

Τι μας λέει το t-ratio;

- Εάν απορρίψουμε την H_0, τότε λέμε ότι ο συντελεστής που εκτιμήσαμε είναι στατιστικά σημαντικός σε επίπεδο σημαντικότητας (α=5%) ή ισοδύναμα επίπεδο εμπιστοσύνης (95%).
- Εάν ο συντελεστής που εκτιμήσαμε δεν είναι στατιστικά σημαντικός, δηλαδή εάν δεν απορρίψουμε την H_0 (όπως για παράδειγμα συμβαίνει για το σταθερό όρο της παλινδρόμησης που εκτιμήσαμε, δηλαδή για το $\hat{\alpha}$) τότε ο συντελεστής που εκτιμήσαμε είναι στατιστικά μιλώντας ίσος με το 0 (δηλαδή ισχύει αυτό που λέει η H_0: α=0).
- Οι μεταβλητές που δεν είναι στατιστικά σημαντικές συνήθως αφαιρούνται από ένα μοντέλο παλινδρόμησης. Επίσης σημαντικό είναι να υπάρχει σταθερός όρος στην γραμμική παλινδρόμηση. Δείτε πόσο χειρότερη είναι μία γραμμή στο να περιγράψει τα δεδομένα χωρίς σταθερό όρο:



Η επιλογή του επιπέδου σημαντικότητας (α) και η πιθανότητά p-value
The Exact Significance Level or *p*-value

- Η επιλογή του επιπέδου σημαντικότητας επηρεάζει άμεσα τις κριτικές τιμές από τους πίνακες των κατανομών και το τελικό συμπέρασμα περί σημαντικότητας ενός εκτιμημένου συντελεστή μίας παλινδρόμησης.
- Στους πίνακες κατανομών, εκτός από την κριτική τιμή του τεστ, μας δίνεται και η πιθανότητα (p-value), η οποία αντιπροσωπεύει την αληθοφάνεια της μηδενικής υπόθεσης.
- Για παράδειγμα, έστω ότι το t-ratio έχει μία τιμή ίση με $t_{62} = 1.47$, η οποία αντιστοιχεί σε p-value ίσο με 0.12 (την κριτική τιμή και την πιθανότητα p-value μας τη δίνει ο πίνακας της κατανομής t για 62 (T-2) βαθμούς ελευθερίας.
  - Σε αυτή την περίπτωση δεν απορρίπτουμε την H_0 για α=5% (γιατί το p-value=12%).
  - Επίσης δεν απορρίπτουμε την H_0 για α=10% (γιατί το p-value=12%)
  - Όμως απορρίπτουμε την H_0 για α=20% (γιατί το p-value=12%).
  - Παρατηρήστε ότι γενικώς όταν **p-value < α, τότε απορρίπτουμε την H_0**

Διαστήματα Εμπιστοσύνης

Πίνακες κατανομών

Διαστήματα Εμπιστοσύνης

- Σύμφωνα με το Κεντρικό Οριακό Θεώρημα η κατανομή του δειγματικού μέσου $\bar{X}$ είναι γνωστή αν γνωρίζουμε τον πληθυσμιακό μέσο (μ) και την τυπική απόκλιση (σ).
- Είδαμε ότι ακόμη κι αν δε γνωρίζουμε την πληθυσμιακή τυπική απόκλιση (σ) μπορούμε να την υποκαταστήσουμε με τη δειγματική τυπική απόκλιση (s).
- Τι συμβαίνει όμως αν δε γνωρίζουμε τον πληθυσμιακό μέσο (μ);
- Άρα δεν μπορούμε να τον μαντέψουμε ακριβώς αλλά γνωρίζουμε ότι βρίσκεται μέσα σε ένα εύρος με συγκεκριμένη πιθανότητα.
- Υποθέστε ότι θέλουμε να επιλέξουμε ένα συμμετρικό διάστημα γύρω από το μ και να βρούμε πόσο συχνά μια μεταβλητή που ακολουθεί κανονική κατανομή θα βρίσκεται σε αυτό το διάστημα.

Τι είναι το διάστημα εμπιστοσύνης

- Το διάστημα εμπιστοσύνης είναι μία εκτίμηση ενός εύρους τιμών για μία μεταβλητή που είναι άγνωστη σε εμάς (π.χ. ο πληθυσμιακός μέσος μ), βασισμένη σε στοιχεία από ένα δείγμα, με μία συγκεκριμένη πιθανότητα. Η πιθανότητα αυτή ονομάζεται και επίπεδο εμπιστοσύνης (90%, 95%, 99%).

Διάστημα Εμπιστοσύνης = Εκτίμηση ± Περιθώριο Λάθους

Παράδειγμα: Διάστημα εμπιστοσύνης για τον πληθυσμιακό μέσο (μ) με την τυπική απόκλιση του πληθυσμού γνωστή (σ)

CONFIDENCE INTERVAL FOR POPULATION MEAN WITH σ KNOWN

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \qquad [9–1]$$

$\bar{x}$ – sample mean

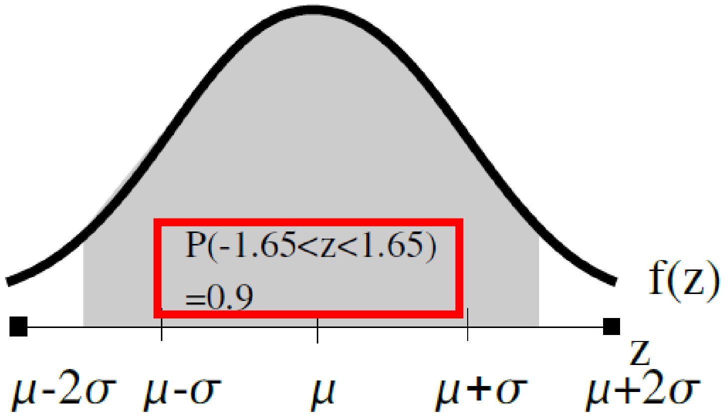
z – z - value for a particular confidence level

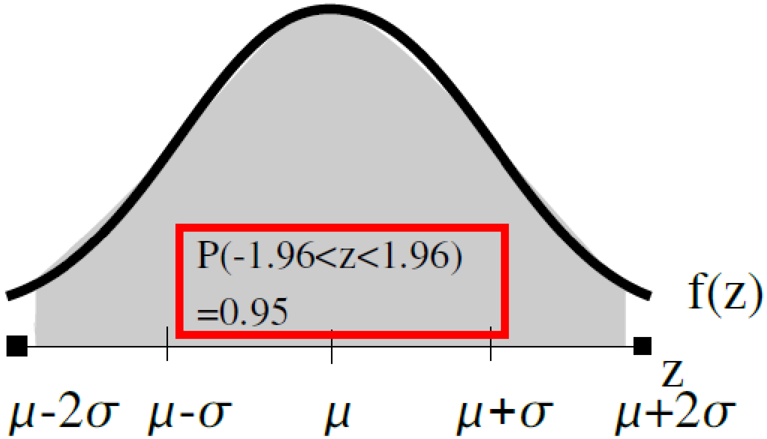
σ – the population standard deviation

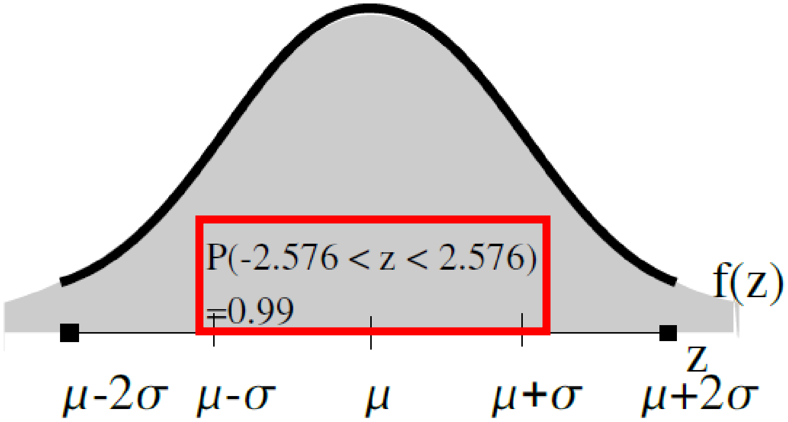
n – the number of observations in the sample

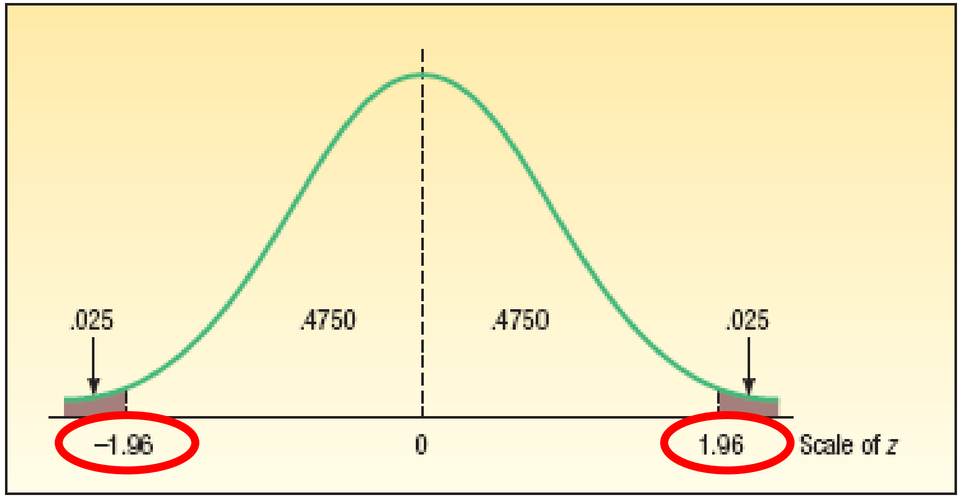
Διάστημα εμπιστοσύνης

- Μια μεταβλητή που κατανέμεται κανονικά με μέσο μ και τυπική απόκλιση σ θα είναι μεταξύ $\mu - 1.645\sigma$ και $\mu + 1.645\sigma$ με πιθανότητα 90%.
  - Αυτό ονομάζεται 90% διάστημα εμπιστοσύνης.
- Μια μεταβλητή που ακολουθεί κανονική κατανομή με μέσο μ και τυπική απόκλιση σ θα είναι μεταξύ $\mu - 1.96\sigma$ και $\mu + 1.96\sigma$ με πιθανότητα 95%.
  - Αυτό αποκαλείται διάστημα εμπιστοσύνης 95%.
- Μια μεταβλητή που ακολουθεί κανονική κατανομή με μέσο μ και τυπική απόκλιση σ θα είναι μεταξύ $\mu - 2.576\sigma$ και $\mu + 2.576\sigma$ με πιθανότητα 99%.
  - Αυτό αποκαλείται διάστημα εμπιστοσύνης 99%.







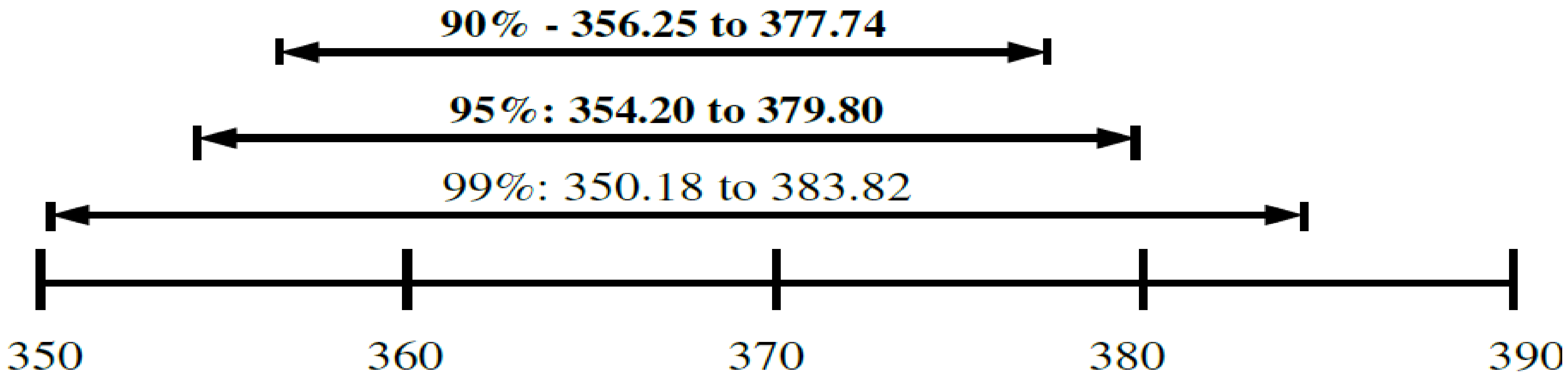


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	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

Παράδειγμα ενός διαστήματος εμπιστοσύνης 1

- Μια έρευνα σε 2,938 πελάτες της υπηρεσίας για τη φροντίδα των άστεγων βρήκε ότι ο μέσος πελάτης κερδίζει \$367 το μήνα με μια τυπική απόκλιση \$354.
- Τυπικό σφάλμα: $\frac{\sigma}{\sqrt{n}} = \frac{354}{2938^{0.5}} = \frac{354}{54.2} = 6.53$

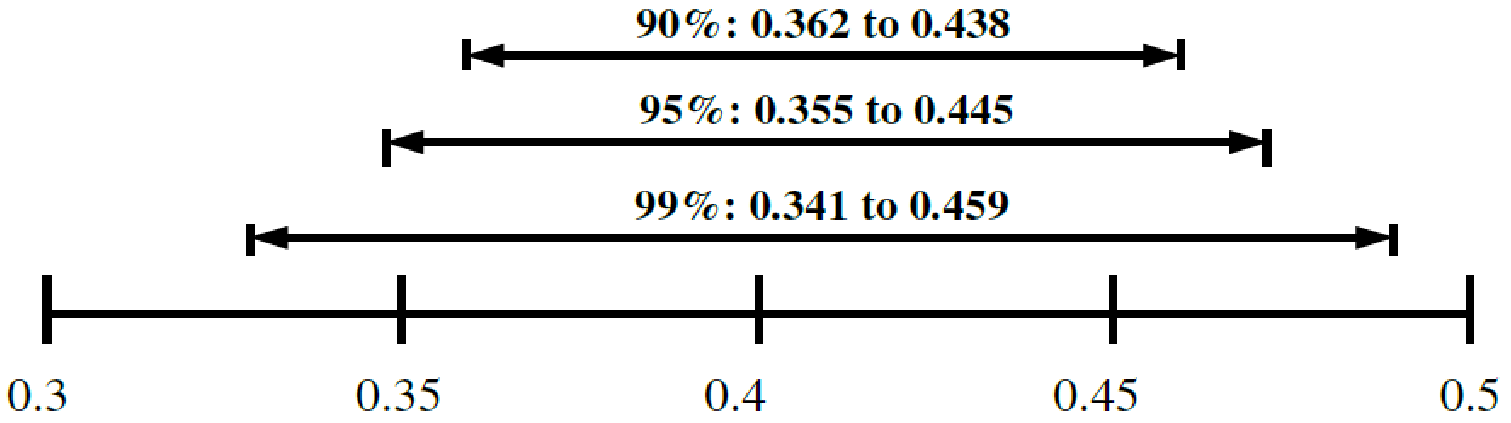
Κατώτατο όριο για το διάστημα εμπιστοσύνης	**Ανώτατο όριο για το διάστημα εμπιστοσύνης**
90% 367 – 1.645(6.53) = 356.26	367 + 1.645(6.53) = 377.74
95% 367 – 1.960(6.53) = 354.20	367 + 1.960(6.53) = 379.80
99% 367 – 2.576(6.53) = 350.18	367 + 2.576(6.53) = 383.82



Παράδειγμα ενός διαστήματος εμπιστοσύνης 2

- Δείγμα 300 σπιτιών. Μέση κατανάλωση κρέατος είναι 0.4 γραμμάρια μέρα με τυπική απόκλιση 0.2.
- Το τυπικό σφάλμα είναι: $\frac{\sigma}{\sqrt{n}} = \frac{0.2}{300^{0.5}} = \frac{0.2}{17.32} = 0.023$

Κατώτατο όριο για το διάστημα εμπιστοσύνης	**Ανώτατο όριο για το διάστημα εμπιστοσύνης**
90% 0.4 – 1.645(0.023) = 0.362	0.4 + 1.645(0.023) = 0.438
95% 0.4 – 1.960(0.023) = 0.355	0.4 + 1.960(0.023) = 0.445
99% 0.4 – 2.576(0.023) = 0.341x	0.4 + 2.576(0.023) = 0.459



Πως διαχειριζόμαστε μικρά δείγματα;

- Για το κεντρικό οριακό θεώρημα το μέγεθος του δείγματος χρειάζεται να είναι πάνω από 30.
- Αν η μεταβλητή από την οποία προέρχεται ακολουθεί κανονική κατανομή τότε ο μέσος υποθέτουμε ότι ακολουθεί κατανομή t.
- Για μεγάλα δείγματα πολλαπλασιάζουμε το τυπικό σφάλμα με τον ίδιο αριθμό για όλα τα μεγέθη δείγματος.
- Για την κατανομή t χρησιμοποιούμε έναν διαφορετικό αριθμό αναλόγως πόσες παρατηρήσεις έχουμε.
- Αν το δείγμα μας έχει n παρατηρήσεις, τότε χρησιμοποιούμε την κατανομή t με $n-1$ βαθμούς ελευθερίας.

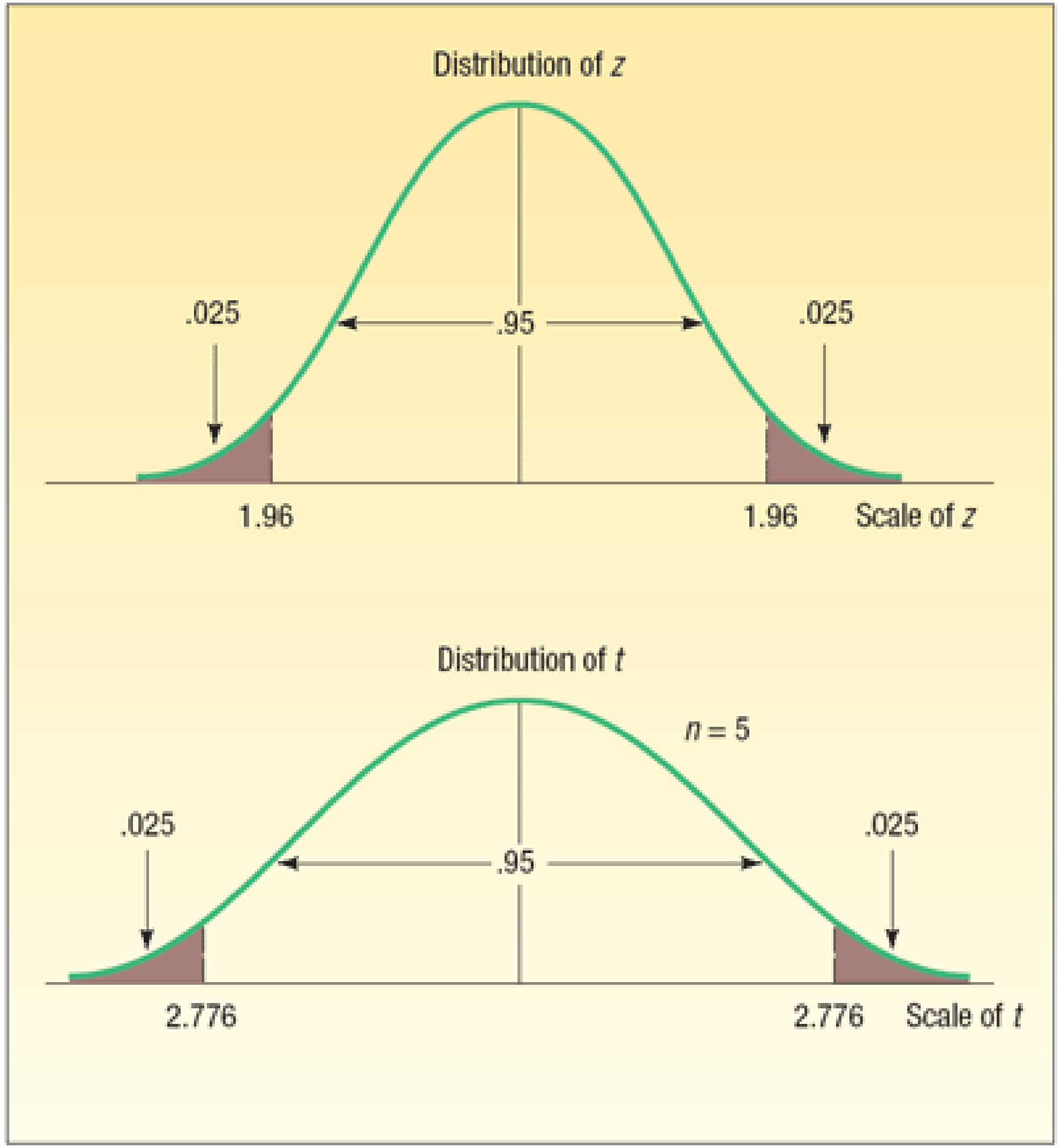


CHART 9–2 Values of z and t for the 95 Percent Level of Confidence

Πως διαχειριζόμαστε μικρά δείγματα; Παράδειγμα

- Ένα δείγμα μαθητών έχει τις παρακάτω βαθμολογίες:
  84, 76 , 98, 43, 65, 76, 90, 92, 64, 87.
- Ποιο είναι το 90% διάστημα εμπιστοσύνης για τον πληθυσμιακό μέσο;

- Ξεκινάμε υπολογίζοντας τον δειγματικό μέσο, τη δειγματική απόκλιση και το τυπικό σφάλμα με τον ίδιο τρόπο.
- Ο δειγματικός μέσος είναι $\bar{X} = 76.6$
- Η δειγματική τυπική απόκλιση είναι s = 18.7
- Άρα το τυπικό σφάλμα είναι $\frac{s}{\sqrt{n}} = \frac{18.7}{\sqrt{10}} = 5.91$
- Υπάρχουν 10 παρατηρήσεις άρα χρησιμοποιούμε 10 – 1 = 9 βαθμοί ελευθερίας.

Πως διαχειριζόμαστε μικρά δείγματα; Παράδειγμα

- $\bar{X} = 76.6, \mathrm{SE} = 5.91$
- 9 βαθμοί ελευθερίας
- 90% διάστημα εμπιστοσύνης
- $\bar{X}-1.83\times SE$ και
- $\bar{X}+1.83\times SE$=76.6 – 1.83(5.91) έως 76.6 + 1.83(5.91) = 65.78 έως 87.42

Όλα τα διαστήματα εμπιστοσύνης είναι two-tailed!

	Confidence Intervals					
	80%	90%	95%	98%	99%	99.9%
of Significance for	One-Tailed	Test				
df	0.100	0.050	0.025	0.010	0.005	0.0005
of Significance for	Two-Tailed	Test				
Level of	0.20	0.10	0.05	0.02	0.01	0.001
1	3.08	6.31	12.71	31.82	63.657	636.619
2	1.89	2.920	4.3	6.97	9.925	31.599
3	1.64	2.35	3.18	4.54	5.841	12.924
4	1.53	2.13	2.78	3.75	4.604	8.610
5	1.48	2.02	2.57	3.37	4.032	6.869
6	1.440	1.943	2.45	3.14	3.707	5.959
7	1.42	1.895	2.37	3	3.499	5.408
8	1.4	1.860	2.31	2.87	3.355	5.041
9	1.38	1.833	2.26	2.82	3.250	4.781
10	1.37	1.812	2.23	2.76	3.169	4.587
11	1.36	1.796	2.2	2.72	3.106	4.437
12	1.36	1.782	2.18	2.68	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.35	1.761	2.15	2.62	2.977	4.140
15	1.34	1.753	2.13	2.6	2.947	4.073

Αυτός ο πίνακας είναι one-tailed, οπότε κοιτάμε το 5% για δ.ε. 90%

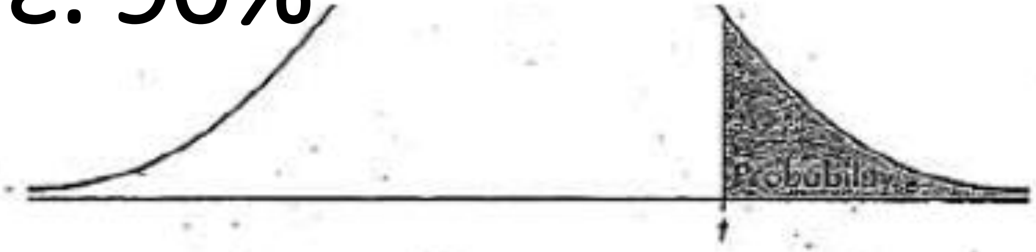


TABLE B: t-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073

Διαχείριση μικρού δείγματος παράδειγμα (1)

- Ο χρόνος ζωής 7 ασθενών με σπάνια μορφή καρκίνου μετά τη μετάσταση είναι: 29, 67, 65, 42, 33, 97 εβδομάδες
- Ποιο είναι το 95% διάστημα εμπιστοσύνης για τον μέσο χρόνο ζωής του πληθυσμού;

	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
29 weeks	-26.57	706.04
67 weeks	11.42	130.61
65 weeks	9.42	88.90
42 weeks	-13.57	184.18
33 weeks	-22.57	509.47
97 weeks	41.42	1716.33
56 weeks	0.42	0.18
Άθροισμα: 389 εβδομάδες Μέσος: 55.57		Άθροισμα: 3335.71 Διακύμανση: 555.95

Διαχείριση μικρού δείγματος παράδειγμα (1)

- Τυπική απόκλιση: 23.58, Τυπικό σφάλμα: $\frac{23.58}{\sqrt{7}} = 8.91$
- Ποιο είναι το 95% διάστημα εμπιστοσύνης για τον μέσο χρόνο ζωής του πληθυσμού;
- Δειγματικός μέσος: 55.57
- Τυπικό σφάλμα: 8.91
- Βαθμοί ελευθερίας: 6
- 95% διάστημα εμπιστοσύνης:$55.7 \pm 2.45(8.91)$
- $= 33.74$ ως 77.40

Διάστημα εμπιστοσύνης για υποσύνολο πληθυσμού (1)

- Ένα υποσύνολο είναι απλώς το υποσύνολο των απαντήσεων σε ένα σετ δεδομένων που ισούνται με έναν συγκεκριμένο αριθμό.
  - Για μια ψευδομεταβλητή (0/1, ναι/όχι) το υποσύνολο των «ναι» ή «1»
  - Για μια κατηγορική μεταβλητή π.χ. η μάρκα του αυτοκινήτου που οδηγεί ένας ερωτώμενος, τι ποσοστό οδηγούν Buick;
    - 1 για «οδηγεί Buick»
    - 0 για «δεν οδηγεί Buick»
  - Για μια πιο γενική διακριτή μεταβλητή π.χ. τι ποσοστό των ανθρώπων έχουν ακριβώς 2 παιδιά;
    - 1 για «έχει ακριβώς δυο παιδία»
    - 0 για «δεν έχει δυο παιδιά»

Διάστημα εμπιστοσύνης για υποσύνολο του πληθυσμού (2)

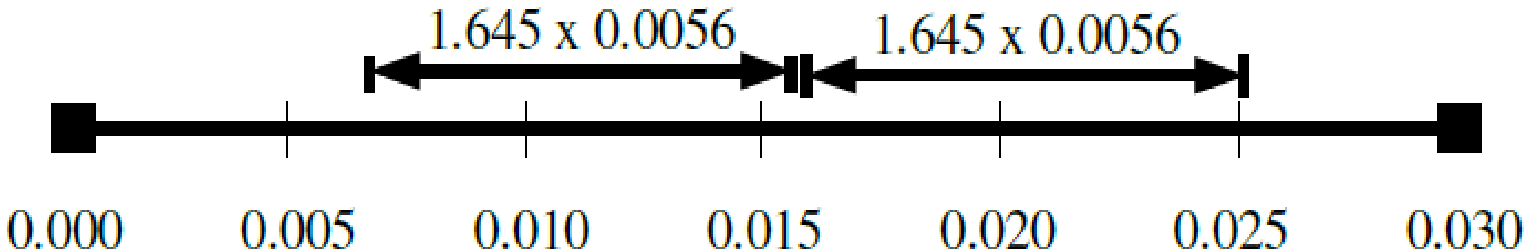
- Σκεφτείτε την ερώτηση: «σε ένα δείγμα μια ψευδομεταβλητής μεγέθους n, ποια είναι η πιθανότητα ότι παρατηρούμε την τιμή «1» k φορές;
- Αυτή είναι μια διωνυμική πιθανότητα
- Άρα ένα υποσύνολο ενός δείγματος είναι στην ουσία μια διωνυμική μεταβλητή που διαιρείται με n.

Διάστημα εμπιστοσύνης για ένα υποσύνολο του πληθυσμού (3)

- Ένα δειγματικό υποσύνολο είναι βασικά μια διωνυμική μεταβλητή που διαιρείται με n.
- Θυμηθείτε ότι καθώς το n γίνεται μεγαλύτερο, μια διωνυμική μεταβλητή πλησιάζει την κανονική με μέσο np και τυπική απόκλιση $\sqrt{np(1-p)}$.
- Άρα αν διαιρέσουμε τη μεταβλητή με n, παίρνουμε μια κανονική μεταβλητή με μέσο p και τυπική απόκλιση $\frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$
- Αν διαιρέσουμε την μεταβλητή με n, παίρνουμε μια κανονική μεταβλητή με μέσο p και τυπική απόκλιση $\frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$
- Άρα η αναμενόμενη τιμή ενός υποσύνολού του δείγματος είναι το πληθυσμιακό υποσύνολο και το τυπικό σφάλμα είναι $\sqrt{\frac{p(1-p)}{n}}$
- Χρησιμοποιούμε αυτή την αναμενόμενη τιμή και τυπικό σφάλμα για δημιουργήσουμε διαστήματα εμπιστοσύνης
- Ζητούμενο: $np \geq 5$ και $np(1-p) \geq 5$

Διάστημα εμπιστοσύνης για ένα υποσύνολο του πληθυσμού – παράδειγμα (1)

- Παράδειγμα: Υποθέτουμε ότι αποφασίζουμε ότι μόνο το 1% των τηλεοράσεων που παράγουμε θα σπάσουν εντός ενός έτους
- Κάνουμε μια έρευνα σε 500 καταναλωτές και βρίσκουμε ότι 8 έχουν σπάσει μέσα σε ένα χρόνο
- Ποιο είναι το 90% διάστημα εμπιστοσύνης για το δείγμα που σπάνε μέσα σε ένα χρόνο;
  - Το υποσύνολο είναι $p = \frac{8}{500} = 0.016$
  - Το τυπικό σφάλμα είναι $\sqrt{\frac{p(1-p)}{n}} = \sqrt{0.016 \times 0.948/500} = \sqrt{0.000031488} = 0.0056$
  - Το διάστημα εμπιστοσύνης; Ίδια μέθοδος όπως πριν
    - Κατώτατο όριο: 0.016 – 1.645(0.0056) = 0.00676
    - Ανώτατο όριο: 0.016 + 1.645(0.0056) = 0.0252



Διάστημα εμπιστοσύνης για ένα υποσύνολο του πληθυσμού – παράδειγμα (2)

- Μια έρευνα που πραγματοποιήθηκε σε ενήλικες βρήκε ότι 83% πιστεύουν ότι η αποστολή γραπτών μηνυμάτων την ώρα της οδήγησης πρέπει να θεωρηθεί μη νόμιμη.
- Ποιο είναι το 90% διάστημα εμπιστοσύνης για το υποσύνολο των ενηλίκων που θεωρούν ότι η αποστολή μηνυμάτων την ώρα της οδήγησης είναι μη νόμιμη;

Διάστημα εμπιστοσύνης για ένα υποσύνολο του πληθυσμού – παράδειγμα (2)

- Ένα δείγμα 1000 πιθανόν ψηφοφόρων βρίσκουν ότι 560 υποστηρίζουν μια καμπανιά δημοψηφίσματος χρηματοοικονομικής αναμόρφωσης
- Ποιο είναι το 99% διάστημα εμπιστοσύνης για το ποσοστό των ψηφοφόρων που υποστηρίζουν το δημοψήφισμα;
- Υποσύνολο πληθυσμού: 560/1000=0.56
- Τυπικό σφάλμα:
  - $\sqrt{p(1-p)/n} = \sqrt{0.56 \times (1-0.56)/1000} = \sqrt{0.0002464} = 0.0157$
  - 99% διάστημα εμπιστοσύνης : $0.56 \pm 0.5195\ to\ 0.6005$