

Εισαγωγή στην οικονομετρία (1404)

Διάλεξη 12

Αγγελος Αλεξόπουλος
Οικονομικό Πανεπιστήμιο Αθηνών -
Τμήμα Οικονομικής Επιστήμης
angelos@aub.gr

Περιγραμματα Διάλεξης

➤ Στατιστική Επαγωγή

✓ Διαστήματα Εμπιστοσύνης

- Γνωστή διακύμανση διαταρακτικού όρου σ^2
- Άγνωστη διακύμανση διαταρακτικού όρου σ^2

✓ Έλεγχοι στατιστικών υποθέσεων

- Μικρά vs. Μεγάλα δείγματα
- Μονόπλευρος vs. Δίπλευρος στατιστικός έλεγχος

Σιοπός Οικονομετρίας

- **Εκτίμηση οικονομικών σχέσεων (ποσοτικοποίηση)**
- Έλεγχος υποθέσεων (οικονομικών σχέσεων) – Στατιστική Επαγωγή
- Πρόβλεψη οικονομικών μεταβλητών

Επιτίμηση οικονομικών σχέσεων

Κλασικές υποθέσεις

Αν X_i είναι σταθερή σε επαναλαμβανόμενα δείγματα (προκαθορισμένη - fixed)

1. Γραμμικότητα ως προς τις παραμέτρους
2. $E(u_i) = 0$ or $E(u_i|x_i) = 0 \forall i$
3. $var(u_i) = E(u_i^2) = \sigma^2$ or $var(u_i|x_i) = E(u_i^2|x_i) = \sigma^2, \forall i$
4. $cov(u_i, u_j) = 0, \forall i \neq j$
5. $u_i \sim N(0, \sigma^2), \forall i$

Εκτίμηση οικονομικών σχέσεων

Απλό γραμμικό υπόδειγμα: $Y_i = \beta_0 + \beta_1 X_i + u_i, i = 1, \dots, n$

Εκτιμημένο υπόδειγμα: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad \forall i$

Εκτιμητές ΕΤ: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ και $\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$ ή $\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$

Κατάλοιπα: $\hat{u}_i = Y_i - \hat{Y}_i$

Ιδιότητες: $\sum_{i=1}^n \hat{u}_i = 0, \sum_{i=1}^n X_i \hat{u}_i = 0, \sum_{i=1}^n \hat{Y}_i \hat{u}_i = 0$

Εκτίμηση οικονομικών σχέσεων

- ✓ Θεώρημα Gauss-Markov: οι εκτιμητές ΕΤ $\hat{\beta}_0$ και $\hat{\beta}_1$ αποτελούν τους **καλύτερους γραμμικούς αμερόληπτους εκτιμητές** και αναφέρονται ως **BLUE** (best linear unbiased estimator)
- ✓ Όταν $u_i \sim N(0, \sigma^2)$, $\forall i$ και X_i προκαθορισμένες τότε

$$\hat{\beta}_0 \sim N[E(\hat{\beta}_0), \text{var}(\hat{\beta}_0)] \sim N\left[\beta_0, \sigma^2 \frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}\right]$$

και

$$\hat{\beta}_1 \sim N[E(\hat{\beta}_1), \text{var}(\hat{\beta}_1)] \sim N\left[\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right]$$

- ✓ Αμερόληπτος εκτιμητής της διακύμανση διαταρακτικού όρου $\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-k}$
- ✓ Συντελεστής προσδιορισμού (coefficient of determination)

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

$$0 < R^2 < 1$$

Σιοπός Οικονομετρίας

- Εκτίμηση οικονομικών σχέσεων (ποσοτικοποίηση)
- Έλεγχος υποθέσεων (οικονομικών σχέσεων) – Στατιστική Επαγωγή
- Πρόβλεψη οικονομικών μεταβλητών

Στατιστική επαγωγή

- **Διάστημα εμπιστοσύνης:** διάστημα τιμών για τον συντελεστή του γραμμικού υποδείγματος εντός του οποίου είναι πιθανό να βρίσκεται η αληθινή τιμή του συντελεστή στον πληθυσμό με βάση τις εκτιμήσεις που παίρνουμε από τον εκτιμητή στο δείγμα με κάποια πιθανότητα
 - Απαιτείται η γνώση της κατανομής των εκτιμητών του υποδείγματος
- **Έλεγχοι στατιστικών υποθέσεων:** έλεγχος αν οι συντελεστές ενός υποδείγματος παίρνουν συγκεκριμένες τιμές
 - Έλεγχος στατιστικής σημαντικότητας: έλεγχος αν ο συντελεστής κλίσης είναι ίσος με το μηδέν

Κατανομή εκτιμητή ΕΤ

- Έστω το γραμμικό υπόδειγμα $Y_i = \beta_0 + \beta_1 X_i + u_i, i = 1, \dots, n$
- Αν υποθέσουμε ότι ισχύουν οι κλασιές υποθέσεις για τον διαταρακτικό όρο u_i και ότι κατανέμεται κανονικά $u_i \sim N(0, \sigma^2), \forall i$ τότε και οι εκτιμητές ΕΤ θα κατανέμονται κανονικά γύρω από τις αληθινές τους τιμές στον πληθυσμό

$$\hat{\beta}_0 \sim N[E(\hat{\beta}_0), var(\hat{\beta}_0)] \sim N \left[\beta_0, \sigma^2 \frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} \right]$$

και

$$\hat{\beta}_1 \sim N[E(\hat{\beta}_1), var(\hat{\beta}_1)] \sim N \left[\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]$$

Τυπικό σφάλμα εκτιμητή $\hat{\beta}_1$

- Η διαφορά του εκτιμητή $\hat{\beta}_1$ από την θεωρητική του τιμή στον πληθυσμό β_1 αποτελεί μια τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή

$$\hat{\beta}_1 - \beta_1 \sim N[0, \text{var}(\hat{\beta}_1)]$$

- Επειδή η μεταβλητή $\hat{\beta}_1 - \beta_1$ έχει διακύμανση που εξαρτάται από τις τιμές της διακύμανσης του διαταρακτικού όρου σ^2 και της ανεξάρτητης μεταβλητής X_i θα την τυποποιήσουμε διαιρώντας την με το **τυπικό σφάλμα** του εκτιμητή $\hat{\beta}_1$ (τυπική απόκλιση)

$$se(\hat{\beta}_1) \equiv \sqrt{\text{var}(\hat{\beta}_1)} = \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Στατιστικό κριτήριο - γνωστή σ^2

- Αυτή η τυποποιημένη μεταβλητή αναφέρεται στην κατασκευή διαστημάτων εμπιστοσύνης (και στους ελέγχους υποθέσεων) ως **στατιστικό κριτήριο** ή **στατιστική** και η κατανομή του ακολουθεί την τυποποιημένη κανονική κατανομή

$$z \equiv \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}} \sim N(0,1)$$

- Η διακύμανση του διαταρακτικού όρου σ^2 θεωρείται **γνωστή!**

Στατιστικό κριτήριο – άγνωστη σ^2

- Ωστόσο η διακύμανση του διαταρακτικού όρου σ^2 αποτελεί άγνωστη παράμετρο του γραμμικού υποδείγματος και επιτιμάται με βάση τα κατάλοιπα

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n - k}$$

- Αν η διακύμανση αντικατασταθεί με μια επίτιμηση της τότε το κριτήριο της τυποποιημένης μεταβλητής $\hat{\beta}_1 - \beta_1$, z , ακολουθεί την t-student κατανομή με $n - k$ βαθμούς ελευθερίας και δίνεται ως εξής

$$t \equiv \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}} \sim t_{(n-k)}$$

Διάστημα εμπιστοσύνης

- Έχοντας ορίσει τα κριτήρια z και t και προσδιορίσει τις κατανομές τους μπορούμε να προχωρήσουμε στον υπολογισμό του διαστήματος εμπιστοσύνης εντός του οποίου θα βρίσκεται η αληθινή τιμή του συντελεστή β_1 στον πληθυσμό με κάποια πιθανότητα $1 - \alpha$
- Ο εκτιμητής ΕΤ $\hat{\beta}_1$ είναι αμερόληπτος εκτιμητής του β_1 και επομένως το διάστημα εμπιστοσύνης θα είναι ένα κλειστό και συμμετρικό διάστημα γύρω από την αληθινή τιμή του β_1 με ανώτερη $\hat{\beta}_1^U$ τιμή και κατώτερη $\hat{\beta}_1^L$ (U: Upper, L: Lower)

$$\hat{\beta}_1^L \leq \beta_1 \leq \hat{\beta}_1^U$$

- Δεξιά και αριστερά των ορίων του διαστήματος εμπιστοσύνης $\hat{\beta}_1^L$ και $\hat{\beta}_1^U$ θα βρίσκεται το $\alpha/2$ των τιμών της κατανομής του $\hat{\beta}_1$
- α επίπεδο πιθανότητας (επίπεδο σημαντικότητας ή σφάλμα τύπου I): σφάλμα η αληθινή τιμή του β_1 να μη συμπεριλαμβάνεται στο διάστημα εμπιστοσύνης

Διάστημα εμπιστοσύνης - γνωστή σ^2

- Ας υποθέσουμε ότι η διακύμανση σ^2 είναι γνωστή τότε μπορούμε να υπολογίσουμε το διάστημα εμπιστοσύνης $\hat{\beta}_1^L \leq \beta_1 \leq \hat{\beta}_1^U$ με βάση το κλειστό διάστημα τιμών του στατιστικού κριτηρίου Z (τυποποιημένη κανονική κατανομή)
- Το διάστημα εντός του οποίου θα βρίσκεται η αληθινή τιμή του Z με κάποια πιθανότητα ορίζεται ως εξής:

$$P(-z_c \leq z \leq z_c) = 1 - \alpha$$

όπου z_c (ή $z_{\alpha/2}$) και $-z_c$ (ή $-z_{\alpha/2}$) οι κριτικές τιμές της τυποποιημένης κανονικής κατανομής δεξιά και αριστερά των οποίων θα βρίσκεται το $\frac{\alpha}{2}$ των τιμών της κατανομής του κριτηρίου Z

Κανονική κατανομή: συμμετρική επομένως

- $P(z < -z_c) = \alpha/2$
- $P(z > z_c) = \alpha/2$

Διάστημα εμπιστοσύνης - γνωστή σ^2

$$P(-z_c \leq z \leq z_c) = 1 - \alpha$$

- Αντικαθιστώντας $z = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$ στην παραπάνω σχέση, όπου $se(\hat{\beta}_1) = \sqrt{var(\hat{\beta}_1)}$, έχουμε:

$$\Pr\left(-z_c \leq \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \leq z_c\right) = 1 - \alpha \Rightarrow$$

$$\Pr\left(z_c \geq -\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \geq -z_c\right) = 1 - \alpha \Rightarrow$$

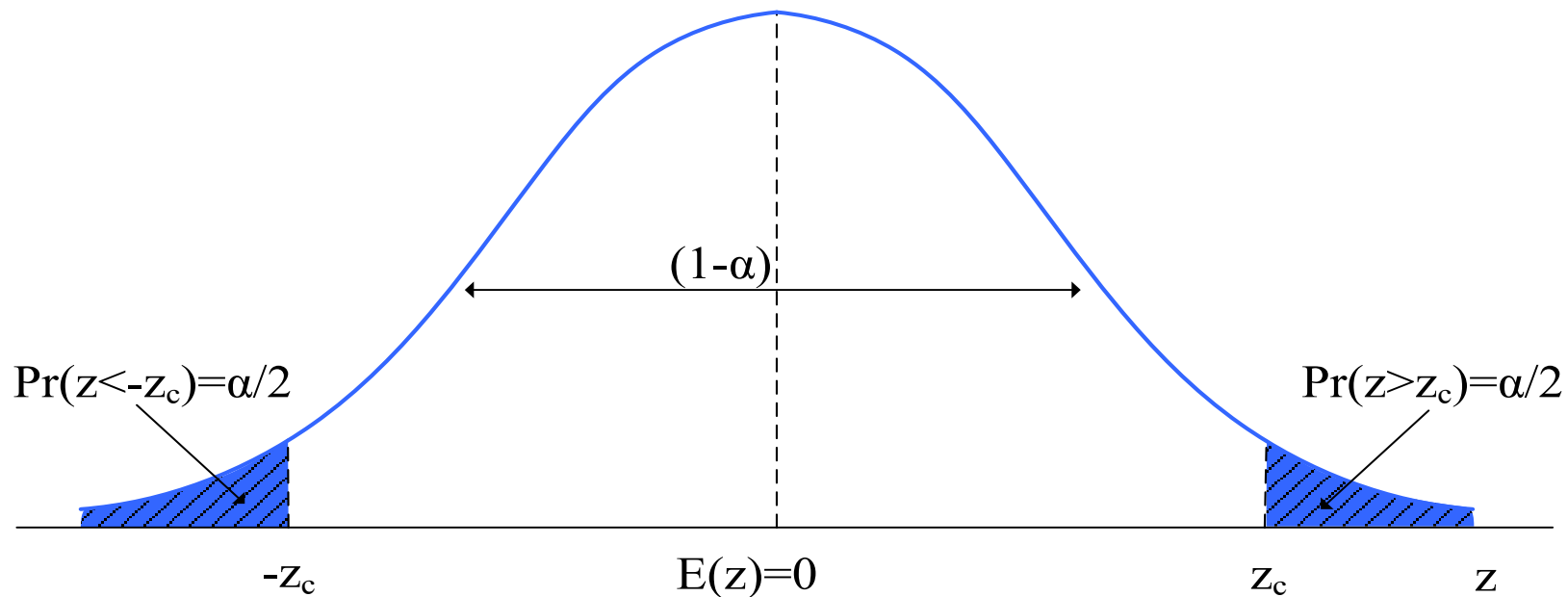
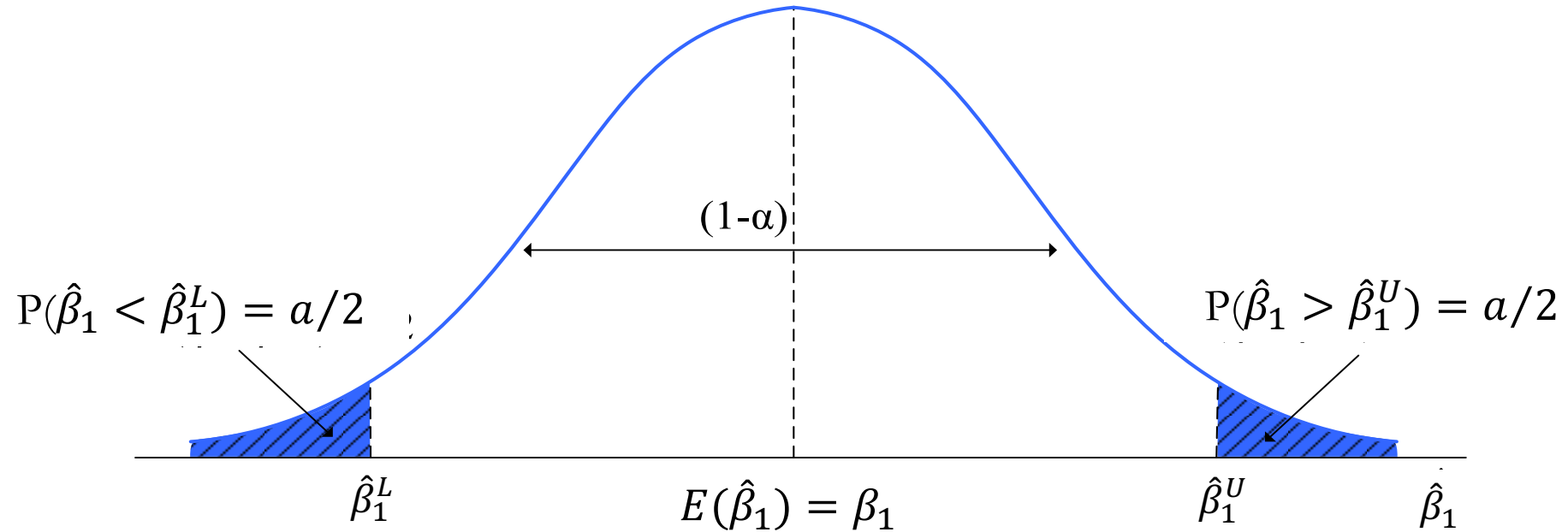
$$\Pr\left(\hat{\beta}_1 + z_c \cdot se(\hat{\beta}_1) \geq \beta_1 \geq \hat{\beta}_1 - z_c \cdot se(\hat{\beta}_1)\right) = 1 - \alpha \Rightarrow$$

$$\Pr\left(\hat{\beta}_1 - z_c \cdot se(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + z_c \cdot se(\hat{\beta}_1)\right) = 1 - \alpha \Rightarrow$$

$$\Pr(\hat{\beta}_1^L \leq \beta_1 \leq \hat{\beta}_1^U) = 1 - \alpha$$

Όπου $\hat{\beta}_1^L = \hat{\beta}_1 - z_c \cdot se(\hat{\beta}_1)$ και $\hat{\beta}_1^U = \hat{\beta}_1 + z_c \cdot se(\hat{\beta}_1)$ αποτελούν αντίστοιχα το κάτω και το άνω όριο του διαστήματος εμπιστοσύνης β_1

Διάστημα εμπιστοσύνης για τον συντελεστή $\hat{\beta}_1$



Διάστημα εμπιστοσύνης- άγνωστη σ^2

- Ανάλογα μπορεί να υπολογιστεί το διάστημα εμπιστοσύνης του συντελεστή β_1 όταν η διακύμανση του διαταρακτικού όρου σ^2 είναι άγνωστη χρησιμοποιώντας το κριτήριο t που ακολουθεί την t-student κατανομή

$$\Pr\left(\hat{\beta}_1 - t_c \cdot se(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_c \cdot se(\hat{\beta}_1)\right) = 1 - \alpha$$

όπου $-t_c$ (ή $-t_{\alpha/2, n-k}$) και t_c (ή $t_{\alpha/2, n-k}$) οι κριτικές τιμές της t-student με $n - k$ βαθμούς ελευθερίας

- t-student κατανομή: συμμετρική επομένως

- $P(t < -t_c) = \alpha/2$ και $P(t > t_c) = \alpha/2$

και $\hat{\beta}_1^L = \hat{\beta}_1 - t_c \cdot se(\hat{\beta}_1)$ και $\hat{\beta}_1^U = \hat{\beta}_1 + t_c \cdot se(\hat{\beta}_1)$ αποτελούν αντίστοιχα το κάτω και το άνω όριο του διαστήματος εμπιστοσύνης β_1

Έλεγχος στατιστικών υποθέσεων

Στάδια στατιστικών ελέγχων:

- i. Ορισμός δύο αντίθετων υποθέσεων:
 - μηδενική υπόθεση (null hypothesis), H_0 και εναλλακτική υπόθεση (alternative hypothesis), H_1
- ii. Υιοθέτηση ενός στατιστικού κριτηρίου (test statistic) για την διεξαγωγή του ελέγχου και ο προσδιορισμός της στατιστικής της κατανομής κάτω από τη μηδενική υπόθεση
- iii. Υιοθέτηση ενός κανόνα απόφασης για την απόρριψη ή μη της μηδενικής υπόθεσης

Μηδενική & εναλλακτική υπόθεση

➤ Έστω ότι ενδιαφερόμαστε να ελέγξουμε κάποια στατιστική υπόθεση για το συντελεστή β_1 του απλού γραμμικού υποδείγματος

✓ Δίπλευρος ή διακτάληκτος έλεγχος (two tailed test)

$$H_0: \beta_1 = \beta_1^0$$

$$H_1: \beta_1 \neq \beta_1^0$$

✓ Μονόπλευρος ή μονοκτάληκτος έλεγχος (one tailed test)

$$H_0: \beta_1 \leq \beta_1^0 \text{ ή } \beta_1 = \beta_1^0$$

$$H_1: \beta_1 > \beta_1^0$$

και

$$H_0: \beta_1 \geq \beta_1^0 \text{ ή } \beta_1 = \beta_1^0$$

$$H_1: \beta_1 < \beta_1^0$$

Στατιστικά κριτήρια

- Η απόρριψη ή η αποδοχή της H_0 για κάποιο συντελεστή του υποδείγματος γίνεται με βάση κάποιο στατιστικό κριτήριο
 - z (γνωστή σ^2 / μεγάλα δείγματα) και t (άγνωστη σ^2 / μικρά δείγματα)
- Απαραίτητη η γνώση της θεωρητικής κατανομής του στατιστικού κριτηρίου κάτω από την H_0
 - Βάση της κατανομής μπορούν να υπολογιστούν οι πιθανότητες απόρριψής ή μη της H_0

Κανόνας απόφασης

	Αληθής H_0	Ψευδής H_0
Απόρριψη H_0	Σφάλμα τύπου I	Σωστή απόφαση
Αποδοχή H_0	Σωστή απόφαση	Σφάλμα τύπου II

- Σωστή απόφαση: να γίνει αποδεκτή η H_0 όταν είναι σωστή ή να απορριφθεί όταν είναι λανθασμένη
- $\alpha = P(\text{Σφάλμα τύπου I}) = P(\text{Απόρριψη } H_0 | \text{Αληθής } H_0)$
 - α η μέγιστη πιθανότητα του σφάλματος τύπου I όταν η H_0 είναι αληθής ονομάζεται **επίπεδο σημαντικότητας & μέγεθος κριτηρίου** (significance level & size of the test)
- $\beta = P(\text{Σφάλμα τύπου II}) = P(\text{Αποδοχή } H_0 | \text{Αληθής } H_1)$
 $= P(\text{Αποδοχή } H_0 | \text{Ψευδής } H_0)$

Κανόνας απόφασης

	Αληθής H_0	Ψευδής H_0
Απόρριψη H_0	Σφάλμα τύπου I	Σωστή απόφαση
Αποδοχή H_0	Σωστή απόφαση	Σφάλμα τύπου II

- Η πιθανότητα απόρριψης της H_0 ενώ η H_1 είναι αληθής μπορεί να υπολογιστεί με βάση την πιθανότητα του σφάλματος τύπου II και ονομάζεται **ισχύς ή δύναμη κριτηρίου** (power of the test):

$$\begin{aligned} P(\text{Απόρριψη } H_0 | \text{Αληθής } H_1) &= 1 - P(\text{Αποδοχή } H_0 | \text{Αληθής } H_1) \\ &= 1 - P(\text{Σφάλμα τύπου II}) = 1 - \beta \end{aligned}$$

- Η ελαχιστοποίηση του σφάλματος τύπου II για κάποιο επίπεδο σημαντικότητας α σημαίνει μεγιστοποίηση της ισχύς του κριτηρίου

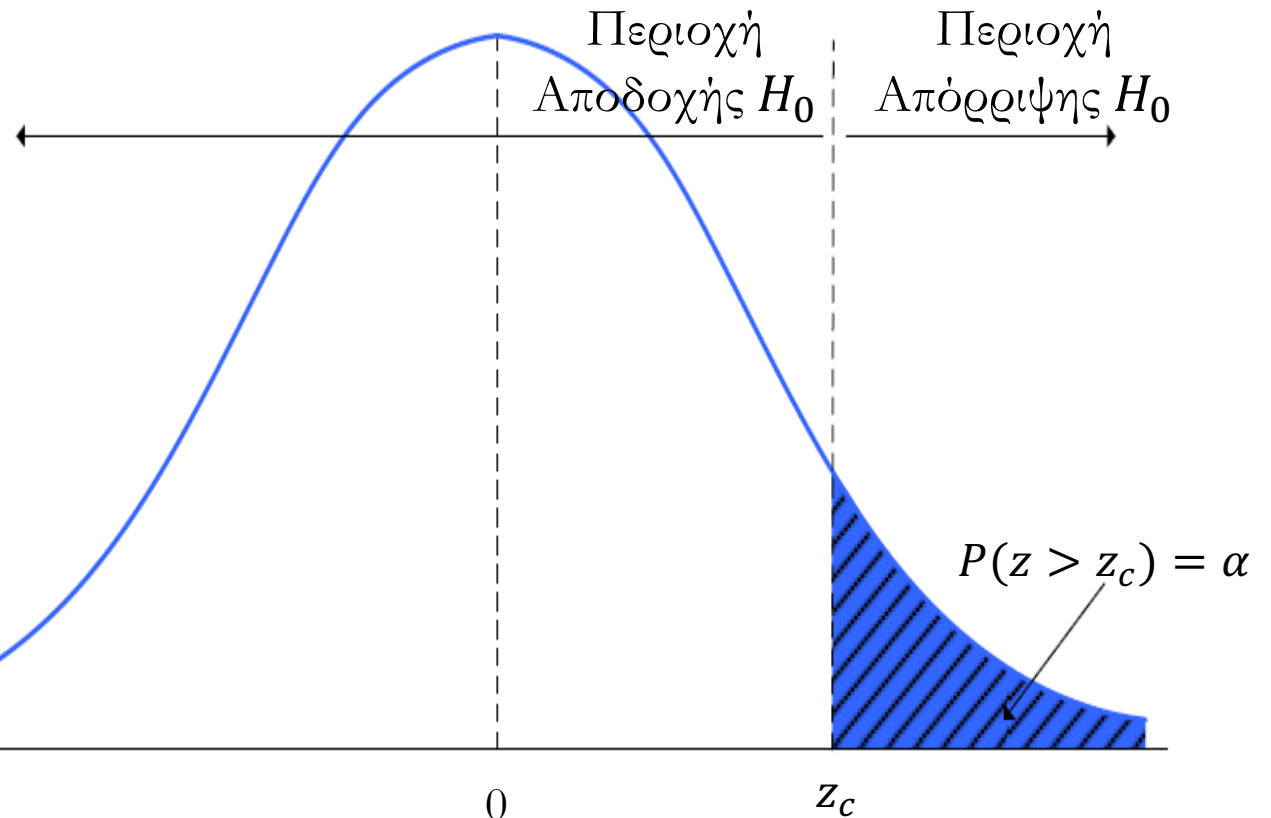
Σφάλμα τύπου I vs. Σφάλμα τύπου II

- Ιδανικά θα θέλαμε να διατηρούμε αυτά τα δύο σφάλματα όσο το δυνατό μικρότερα
 - όμως αν \downarrow Σφάλμα τύπου I, το Σφάλμα τύπου II \uparrow ή αντίστροφα
- Κλασικό έλεγχο υποθέσεων: θέτουμε τη μέγιστη πιθανότητα του σφάλματος τύπου I σε κάποιο επίπεδο σημαντικότητάς που είναι γενικά αποδεκτό και προσπαθούμε να επιλέξουμε το στατιστικό κριτήριο που ελαχιστοποιεί την πιθανότητα σφάλματος τύπου II
 - $\alpha = 1\%$, $\alpha = 5\%$ ή $\alpha = 10\%$
- Γενική αρχή του δικαίου: είναι προτιμότερο να μην καταδικάσουμε έναν αθώο παρά να αθώσουμε έναν ένοχο

Μονόπλευρος έλεγχος προς τα δεξιά

$$H_0: \beta_1 \leq \beta_1^0 \text{ ή } \beta_1 = \beta_1^0$$
$$H_1: \beta_1 > \beta_1^0$$

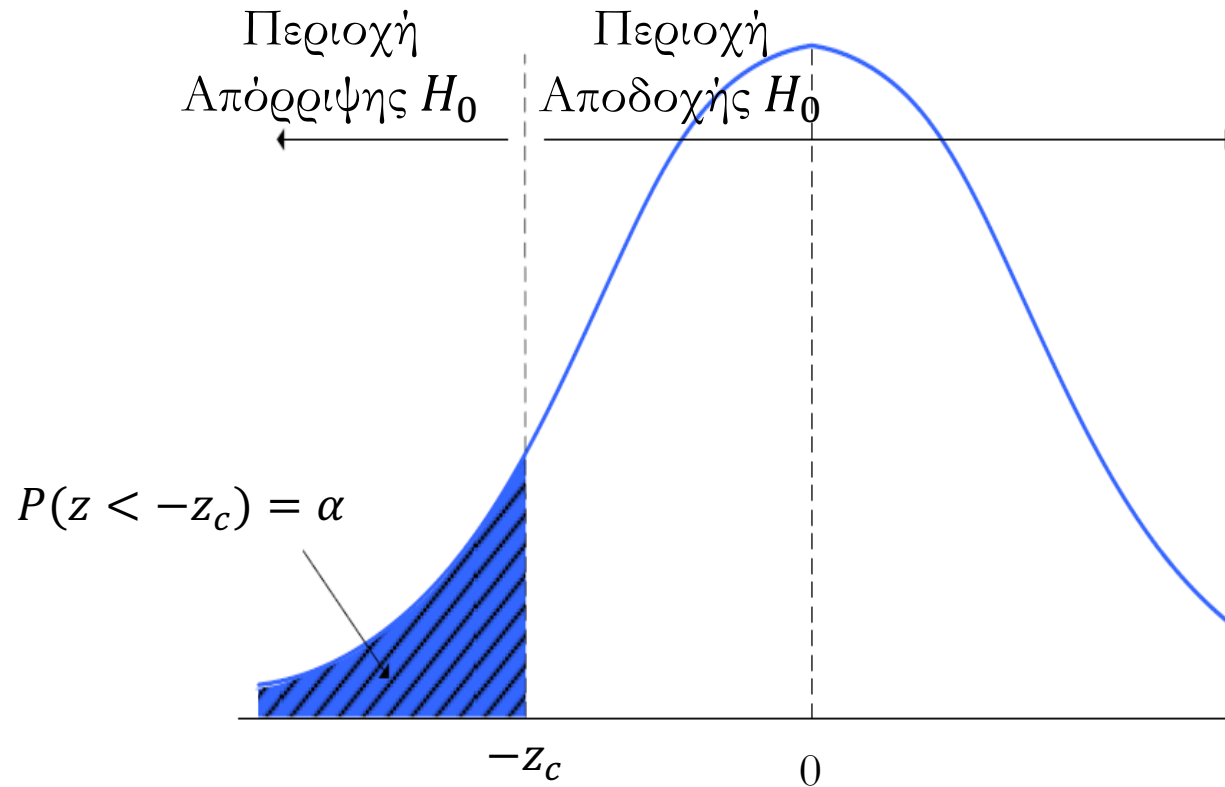
- Για κάποιο επίπεδο σημαντικότητας α η H_0 απορρίπτεται όταν η τιμή του στατιστικού κριτηρίου είναι μεγαλύτερη της κριτικής τιμής της κατανομής του, $z > z_c$ ή $t > t_c$
- Τιμή πιθανότητας (probability value, p-value) στατιστικού κριτηρίου, p-τιμή: η πιθανότητα το κριτήριο να παίρνει μια τιμή μεγαλύτερη από αυτή που εκτιμάται από τα δεδομένα του δείγματος
- Η H_0 απορρίπτεται αν p – value $< \alpha$
- Αν ζ η τιμή του στατιστικού κριτηρίου τότε



$$\text{p-τιμή} = P(z \geq \zeta) = 1 - P(z \leq \zeta) = 1 - \Phi(\zeta) = 1 - \int_{-\infty}^{\zeta} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Μονόπλευρος έλεγχος προς τα αριστερά

$$H_0: \beta_1 \geq \beta_1^0 \text{ ή } \beta_1 = \beta_1^0$$
$$H_1: \beta_1 < \beta_1^0$$



- Για κάποιο επίπεδο σημαντικότητας α η H_0 απορρίπτεται όταν η τιμή του στατιστικού κριτηρίου είναι μικρότερη της κριτικής τιμής της κατανομής του, $z < -z_c$ ή $t < -t_c$
- Η H_0 απορρίπτεται αν p - value $< \alpha$

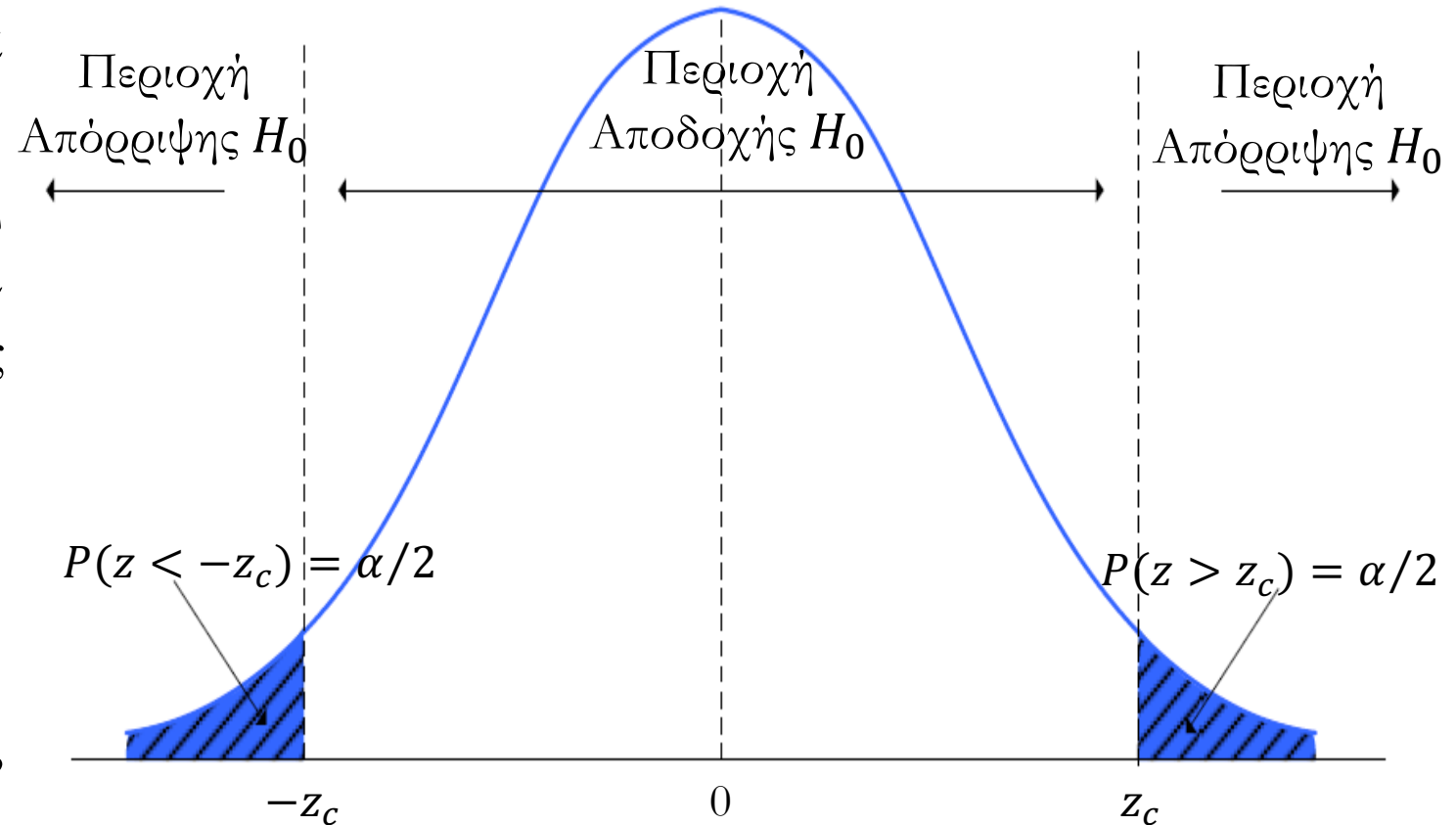
• Αν ζ η τιμή του στατιστικού κριτηρίου τότε

$$p\text{-τιμή} = P(z \leq \zeta) = \Phi(\zeta) = \int_{-\infty}^{\zeta} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Δίπλευρος έλεγχος

$$H_0: \beta_1 = \beta_1^0$$
$$H_1: \beta_1 \neq \beta_1^0$$

- Η απόρριψη της H_0 μπορεί να γίνει είτε προς τα αριστερά είτε προς τα δεξιά της κατανομής του κριτηρίου
- Η H_0 απορρίπτεται όταν η απόλυτη τιμή του στατιστικού κριτηρίου είναι μεγαλύτερη ή ίση της κριτικής τιμής της κατανομής του, $|z| > z_c$ ή $|t| > t_c$
- Η H_0 απορρίπτεται αν $p - \text{value} < \alpha$
- Έλεγχος στατιστικής σημαντικότητας, $\beta_1^0 = 0$



$$p\text{-τιμή} = P(z \leq -\zeta) + P(z \geq \zeta) = 2P(z \leq -\zeta) = \int_{-\infty}^{-\zeta} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Περίληψη

- ✓ Τυπικό σφάλμα του εκτιμητή
- ✓ Διάστημα εμπιστοσύνης
 - Γνωστή vs. Άγνωστη διακύμανση διαταρακτικού όρου
- ✓ Έλεγχος στατιστικών υποθέσεων
 - z (μεγάλα δείγματα) και t (μικρά δείγματα)
 - Μονόπλευρος προς τα δεξιά/αριστερά έλεγχος
 - Δίπλευρος έλεγχος

Βιβλιογραφία

- ✓ Τζαβαλής, κεφ. 4 & 5
- ✓ Stock & Watson, κεφ. 5.2
- ✓ Gujarati & Porter, κεφ. 4 & 5

Ευχαριστώ!

Ερωτήσεις?