

Στατιστική II

Έλεγχοι Υποθέσεων — Κλασικοί Έλεγχοι & Η Τριάδα LR/Wald/LM

Τμήμα Οικονομικής Επιστήμης, ΟΠΑ

Εισαγωγή: από την εκτίμηση στον έλεγχο

Οι δύο προηγούμενες ενότητες αφιερώθηκαν στην *εκτίμηση*: πώς να δίνουμε σε μια άγνωστη παράμετρο μια εύλογη αριθμητική τιμή και πώς να ποσοτικοποιούμε την αβεβαιότητα. Σε αυτή την ενότητα προχωράμε στο δεύτερο μεγάλο σκέλος της συμπερασματολογίας: τον **έλεγχο υποθέσεων** (hypothesis testing). Εδώ δεν ψάχνουμε «ποια είναι η τιμή της παραμέτρου», αλλά «ταιριάζουν τα δεδομένα μας με μια συγκεκριμένη οικονομική υπόθεση»;

Τυπικά οικονομικά ερωτήματα ελέγχου:

- Είναι ο μέσος μηνιαίος μισθός σε αυτό τον κλάδο πραγματικά 1.500 €, όπως υποστηρίζει ο εργοδοτικός σύλλογος;
- Η ημερήσια μεταβλητότητα μιας μετοχής έχει αυξηθεί μετά την κρίση;
- Είναι ο πληθωρισμός υψηλότερος από τον στόχο της ΕΚΤ;
- Έχει η μεταρρύθμιση στην αγορά εργασίας μειώσει την ανεργία;

Σε κάθε περίπτωση, διατυπώνουμε δύο ανταγωνιστικούς ισχυρισμούς για την αληθινή τιμή της παραμέτρου, και χρησιμοποιούμε τα δεδομένα μας για να επιλέξουμε μεταξύ τους με ελεγχόμενο κίνδυνο λάθους.

Δομή της ενότητας:

Κλασικοί έλεγχοι: Η δομή του ελέγχου (H_0, H_1 , σφάλματα τύπου I/II, ισχύς, p -values), οι κλασικοί έλεγχοι (Z, t, χ^2 , ένα και δύο δείγματα), και το λήμμα Neyman-Pearson για βέλτιστους ελέγχους.

Η τριάδα LR/Wald/LM: Οι τρεις «ασυμπτωτικοί» έλεγχοι που είναι ο πυρήνας κάθε σύγχρονης οικονομετρικής εφαρμογής. Παρουσιάζονται ξεχωριστά στο πλαίσιο της κανονικής κατανομής (όπου, εκπληκτικά, συμπίπτουν αλγεβρικά) και *ενοποιημένα* σε μια Bernoulli εφαρμογή, όπου διαφέρουν αριθμητικά.

Σύνδεση με την Οικονομετρία

Η τριάδα Likelihood Ratio (LR), Wald, και Lagrange Multiplier (LM, ή Score) είναι το πιο σημαντικό σύνολο εργαλείων που θα μάθετε σε όλο το προπτυχιακό πρόγραμμα Οικονομετρίας. Κάθε στατιστική δοκιμασία που θα δείτε στο Stata, R, EViews ή Python (statsmodels) — από τον έλεγχο σημαντικότητας ενός συντελεστή μέχρι την ετεροσκεδαστικότητα και την αυτοσυσχέτιση — ανήκει σε μία από τις τρεις κατηγορίες. Είναι τόσο θεμελιώδεις που στις εξετάσεις εισόδου σε μεταπτυχιακά οικονομικά (UCL, LSE, Bocconi, NES, κ.λπ.) εμφανίζονται κάθε χρόνο.

Προαπαιτούμενα: ιδιότητες MLE, πληροφορία Fisher, συνάρτηση score, ασυμπτωτική κανονικότητα, μέθοδος Delta, και κανονική/ t/χ^2 κατανομές.

Περιεχόμενα

Εισαγωγή	1
1 Δομή Ελέγχου και Κλασικοί Έλεγχοι	2
1.1 Διατύπωση: μηδενική και εναλλακτική υπόθεση	2
1.2 Σφάλματα τύπου I και II — Ισχύς	3
1.3 Στατιστική έλεγχου, περιοχή απόρριψης, p-value	4
1.4 Κλασικοί έλεγχοι για μέση τιμή	6
1.5 Έλεγχοι δύο δειγμάτων	8
1.6 Σύνδεση μεταξύ ελέγχων και διαστημάτων εμπιστοσύνης	8
1.7 Λήμμα Neyman-Pearson	9
2 Λυμένες Ασκήσεις	9
3 Η Τριάδα LR / Wald / LM	14
3.1 Το γενικό πλαίσιο	14
4 Λυμένες Ασκήσεις	16
4.1 Ομάδα A: Τριάδα στην κανονική κατανομή — όπου όλα συμπίπτουν . .	16
4.2 Ομάδα B: Unified case study — Bernoulli, όπου διαφέρουν	18
5 Σύνοψη Τύπων	20

1 Δομή Ελέγχου και Κλασικοί Έλεγχοι

1.1 Διατύπωση: μηδενική και εναλλακτική υπόθεση

Ορισμός 3.1 (Στατιστικός έλεγχος υποθέσεων)

Ένας **στατιστικός έλεγχος υποθέσεων** είναι μια διαδικασία απόφασης μεταξύ δύο ανταγωνιστικών ισχυρισμών για την άγνωστη παράμετρο θ :

- **Μηδενική υπόθεση** (null hypothesis), συμβολισμός H_0 : ο «αρχικός» ή «status quo» ισχυρισμός. Π.χ. $H_0 : \mu = 1.500$ (ο μέσος μισθός είναι 1.500 €).
- **Εναλλακτική υπόθεση** (alternative hypothesis), συμβολισμός H_1 : ο ισχυρισμός που υιοθετούμε αν η H_0 απορριφθεί. Π.χ. $H_1 : \mu \neq 1.500$, ή $H_1 : \mu > 1.500$.

Τύποι εναλλακτικών:

- **Αμφίπλευρη** (two-sided): $H_1 : \theta \neq \theta_0$.
- **Άνω μονόπλευρη** (right-tailed): $H_1 : \theta > \theta_0$.
- **Κάτω μονόπλευρη** (left-tailed): $H_1 : \theta < \theta_0$.

Απλές vs σύνθετες υποθέσεις:

- **Απλή** (simple): καθορίζει πλήρως την κατανομή, π.χ. $H_0 : \mu = 1.500$ σε γνωστή σ .
- **Σύνθετη** (composite): αφήνει αδιευκρίνιστες παραμέτρους, π.χ. $H_1 : \mu > 1.500$ ή κάθε υπόθεση όταν η σ είναι άγνωστη.

Ασύμμετροι ρόλοι

Η H_0 και η H_1 δεν είναι συμμετρικές. Η H_0 θεωρείται «αληθής εκτός αν υπάρχουν ισχυρές αποδείξεις για το αντίθετο» — όπως ο κατηγορούμενος σε δικαστήριο που τεκμαίρεται αθώος. Έτσι, αν τα δεδομένα δεν είναι αρκετά πειστικά εναντίον της H_0 , **αποτυγχάνουμε να την απορρίψουμε** — δεν λέμε «την αποδεχόμαστε». Η διάκριση είναι κρίσιμη: η μη απόρριψη είναι ασθενέστερη απόφαση από την αποδοχή.

1.2 Σφάλματα τύπου I και II — Ισχύς

Σε κάθε έλεγχο μπορούμε να κάνουμε δύο είδη λάθους:

		Πραγματικότητα	
		H_0 αληθής	H_0 ψευδής
Απόφαση	Απορρίπτω H_0	Σφάλμα τύπου I πιθανότητα α	Σωστή απόφαση ισχύς = $1 - \beta$
	Δεν απορρίπτω H_0	Σωστή απόφαση εμπιστοσύνη $1 - \alpha$	Σφάλμα τύπου II πιθανότητα β

Ορισμός 3.2 (Σφάλματα και ισχύς)

- **Σφάλμα τύπου I:** απόρριψη της H_0 ενώ είναι αληθής. Πιθανότητα:

$$\alpha = \mathbb{P}(\text{απόρριψη } H_0 \mid H_0 \text{ αληθής}).$$

Ονομάζεται **επίπεδο σημαντικότητας** (significance level).

- **Σφάλμα τύπου II:** μη απόρριψη της H_0 ενώ είναι ψευδής. Πιθανότητα:

$$\beta = \mathbb{P}(\text{μη απόρριψη } H_0 \mid H_0 \text{ ψευδής}).$$

- **Ισχύς ελέγχου** (power): $1 - \beta = \mathbb{P}(\text{απόρριψη } H_0 \mid H_0 \text{ ψευδής})$. Είναι η πιθανότητα να εντοπίσουμε σωστά μια ψευδή H_0 .

Το κλασικό trade-off

Τα α και β είναι αντίστροφα συνδεδεμένα: για δεδομένο μέγεθος δείγματος, μειώνοντας το α (γινόμαστε αυστηρότεροι στην απόρριψη) αυξάνουμε το β (χάνουμε περισσότερες πραγματικά ψευδείς H_0). Η τυπική πρακτική θέτει $\alpha = 0,05$ ή $0,01$, και μετά μεγαλώνει το δείγμα για να μειωθεί το β .

Παρανόηση που πρέπει να αποφεύγετε: Το $\alpha = 0,05$ δεν σημαίνει «5% πιθανότητα να κάνω λάθος». Σημαίνει «5% πιθανότητα να απορρίψω εσφαλμένα μια αληθινή H_0 » — υπό συνθήκη ότι η H_0 είναι αληθής.

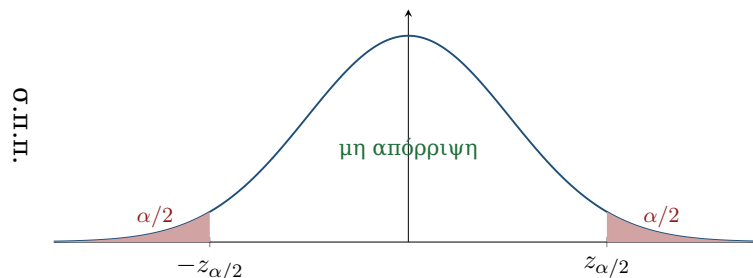
1.3 Στατιστική ελέγχου, περιοχή απόρριψης, p-value

Ορισμός 3.3 (Στατιστική ελέγχου και περιοχή απόρριψης)

Ένας έλεγχος ορίζεται από:

- Μια **στατιστική ελέγχου** (test statistic) $T = T(X_1, \dots, X_n)$ της οποίας η κατανομή υπό την H_0 είναι γνωστή.
- Μια **περιοχή απόρριψης** (rejection region) R : αν η παρατηρούμενη τιμή του T ανήκει στην R , απορρίπτουμε την H_0 .

Η περιοχή R καθορίζεται έτσι ώστε $\mathbb{P}(T \in R \mid H_0) = \alpha$.



Σχήμα 1: Αμφίπλευρος έλεγχος με στατιστική $T \sim N(0, 1)$ υπό H_0 . Περιοχή απόρριψης: ουρές με συνολικό εμβαδόν α .

Ορισμός 3.4 (p-value)

Το **p-value** είναι η πιθανότητα, υπό την H_0 , να παρατηρήσουμε τιμή της στατιστικής **τουλάχιστον** τόσο ακραία όσο η παρατηρούμενη t_{obs} :

$$p\text{-value} = \mathbb{P}(T \text{ τουλάχιστον τόσο ακραίο όσο } t_{\text{obs}} \mid H_0).$$

Κανόνας απόφασης με p-value:

$$p\text{-value} \leq \alpha \implies \text{απορρίπτουμε } H_0.$$

$$p\text{-value} > \alpha \implies \text{δεν απορρίπτουμε } H_0.$$

Σωστή ερμηνεία του p-value

Το **p-value** δεν είναι η πιθανότητα η H_0 να είναι αληθής. Είναι η πιθανότητα, αν υποθέσουμε την H_0 , να βλέπαμε ένα δείγμα τόσο ακραίο όσο το δικό μας. Μικρό **p-value** σημαίνει: «αν η H_0 ίσχυε, αυτό που παρατηρώ θα ήταν εξαιρετικά απίθανο — μάλλον η H_0 δεν ισχύει».

Διαισθητική κατανόηση του p-value — ένα καθημερινό παράδειγμα

Η σκηνή. Φαντάσου ότι ένας φίλος σου ισχυρίζεται: «έχω ένα δίκαιο νόμισμα» (δηλαδή κορώνα/γράμματα με πιθανότητα 50% το καθένα). Εσύ τον υποπτεύεσαι. Λες: «για να δω, ρίξε το 10 φορές». Εκείνος το ρίχνει και βγαίνει **κορώνα 9 φορές, γράμματα μόνο 1 φορά**.

Το ερώτημα. Είναι το νόμισμα όντως δίκαιο, ή πειραγμένο; Σκέψου: αν το νόμισμα ήταν πραγματικά δίκαιο, πόσο παράξενο θα ήταν να δω 9 κορώνες σε 10 ρίψεις; Αυτό ακριβώς υπολογίζει το **p-value**:

«Υποθέτοντας ότι ο φίλος μου λέει αλήθεια (δίκαιο νόμισμα), πόσο πιθανό ήταν να συμβεί κάτι τόσο ακραίο όσο αυτό που είδα;»

Ο υπολογισμός. Αν το νόμισμα ήταν δίκαιο, η πιθανότητα να βγουν 9 ή 10 κορώνες σε 10 ρίψεις είναι περίπου 1%. Άρα $p\text{-value} \approx 0,01$.

Η ερμηνεία. Το 1% σημαίνει: «αν το νόμισμα ήταν δίκαιο, αυτό που είδα θα συνέβαινε μόνο σε 1 στις 100 περιπτώσεις». Έχεις δύο επιλογές:

- Είτε ο φίλος σου είναι ειλικρινής και απλώς συνέβη κάτι πολύ απίθανο,
- Είτε ο φίλος σου ψεύδεται και το νόμισμα δεν είναι δίκαιο.

Επειδή το «απίθανο» είναι όντως πολύ απίθανο (1%), η λογική επιλογή είναι να **απορρίψεις** τον ισχυρισμό του.

Η σύνδεση με το α . Πριν δεις τα δεδομένα, αποφασίζεις πόσο αυστηρός κριτής θες να είσαι. Θέτεις ένα όριο α — π.χ. $\alpha = 5\%$ — που σημαίνει: «θα απορρίψω τον φίλο μου μόνο αν αυτό που βλέπω θα συνέβαινε λιγότερο από 5% των φορών υπό την υπόθεσή του».

- Στο παράδειγμά μας: $p\text{-value} = 1\% < 5\% \rightarrow$ **απορρίπτω** την υπόθεση «δίκαιο νόμισμα».
- Αν είχαν βγει 6 κορώνες (αντί 9), το $p\text{-value}$ θα ήταν περίπου 38%. Πολύ πιθανό υπό δίκαιο νόμισμα $\rightarrow$ **δεν απορρίπτω**.

Τι ΔΕΝ είναι το $p\text{-value}$. Τρεις συνηθισμένες παρανοήσεις:

1. **Δεν είναι** «η πιθανότητα να είναι δίκαιο το νόμισμα». Το $p\text{-value}$ υπολογίζεται υπό την υπόθεση ότι είναι δίκαιο. Δεν μάς λέει την πιθανότητα της υπόθεσης — μάς λέει πόσο εύλογα είναι τα δεδομένα δοθείσας της υπόθεσης.
2. **Δεν είναι** «η πιθανότητα να κάνω λάθος αν απορρίψω». Είναι κάτι πιο συγκεκριμένο: η πιθανότητα να δω αυτά τα δεδομένα (ή πιο ακραία) αν η μηδενική υπόθεση είναι όντως αληθής.
3. **Δεν μετράει** το «πόσο μεγάλη» είναι η διαφορά. Μετράει μόνο πόσο απίθανη είναι, δεδομένου του μεγέθους δείγματος. Με τεράστιο δείγμα, ακόμη και πολύ μικρές διαφορές μπορεί να δώσουν πολύ μικρό $p\text{-value}$ — αυτό δεν σημαίνει ότι είναι «σημαντικές» πρακτικά, μόνο στατιστικά.

Η αναλογία του δικαστή. Είσαι δικαστής. Ο κατηγορούμενος τεκμαίρεται **αθώος** (αυτή είναι η H_0). Σου παρουσιάζουν στοιχεία. Το $p\text{-value}$ απαντάει: «αν ο κατηγορούμενος ήταν όντως αθώος, πόσο πιθανό θα ήταν να εμφανίζονταν αυτά τα στοιχεία εις βάρος του;»

- Αν η απάντηση είναι «πολύ απίθανο» (μικρό $p\text{-value}$), τα στοιχεία είναι ενοχοποιητικά $\rightarrow$ **καταδικάζεις** (απορρίπτεις την αθωότητα).
- Αν η απάντηση είναι «αρκετά εύλογο» (μεγάλο $p\text{-value}$), τα στοιχεία δεν είναι πειστικά $\rightarrow$ **αθωνώνεις** (δεν απορρίπτεις).

Το $\alpha = 5\%$ είναι ο **βαθμός βεβαιότητας** που απαιτείς για να καταδικάσεις. Όσο πιο σοβαρή η απόφαση, τόσο μικρότερο α επιλέγεις — π.χ. στη φαρμακευτική έρευνα $\alpha = 0,01$ ή και $0,001$.

Σύνοψη σε μία γραμμή:

Το $p\text{-value}$ είναι μια **μέτρηση έκπληξης** — πόσο «ακραία» είναι τα δεδομένα μας αν η μηδενική υπόθεση είναι αληθινή.

Μικρό $p\text{-value}$ = ακραία δεδομένα = «μάλλον η υπόθεση δεν στέκει».

1.4 Κλασικοί έλεγχοι για μέση τιμή

Θεώρημα 3.1 (Έλεγχος Z — γνωστή διασπορά)

Έστω $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ με σ^2 γνωστή. Για τον έλεγχο $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, η στατιστική

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

ακολουθεί $N(0, 1)$ υπό την H_0 . Απορρίπτουμε σε επίπεδο α αν $|Z| > z_{\alpha/2}$.

Θεώρημα 3.2 (Έλεγχος t — άγνωστη διασπορά)

Έστω $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ με σ^2 άγνωστη, και $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$. Για τον έλεγχο $H_0 : \mu = \mu_0$, η στατιστική

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

ακολουθεί κατανομή Student- t με $n-1$ βαθμούς ελευθερίας υπό την H_0 . Απορρίπτουμε αν $|T| > t_{n-1, \alpha/2}$.

Θεώρημα 3.3 (Έλεγχος χ^2 για διασπορά)

Έστω $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$. Για τον έλεγχο $H_0 : \sigma^2 = \sigma_0^2$, η στατιστική

$$W = \frac{(n-1)S^2}{\sigma_0^2}$$

ακολουθεί χ_{n-1}^2 υπό την H_0 . Για αμφίπλευρο έλεγχο, απορρίπτουμε αν $W < \chi_{n-1, 1-\alpha/2}^2$ ή $W > \chi_{n-1, \alpha/2}^2$.

Η κατανομή Student- t

Η κατανομή **Student- t** με ν βαθμούς ελευθερίας (degrees of freedom) εμφανίζεται φυσικά όταν χρησιμοποιούμε τη δειγματική τυπική απόκλιση S αντί για τη γνωστή σ . Συγκεκριμένα, αν $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ με άγνωστη σ , τότε

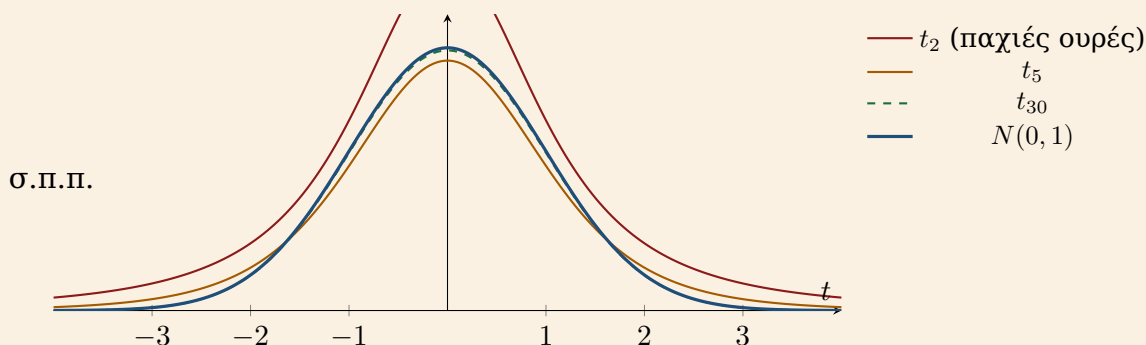
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Δηλαδή ο t_{n-1} ενσωματώνει την πρόσθετη αβεβαιότητα από την εκτίμηση του σ .

Βασικές ιδιότητες:

- Συμμετρική γύρω από το 0 (όπως η κανονική).
- **Παχύτερες ουρές** από την κανονική — δίνει μεγαλύτερη πιθανότητα σε ακραίες τιμές.
- Όσο μεγαλώνει το ν , η t_ν συγκλίνει στην τυπική κανονική $N(0, 1)$. Για $\nu \geq 30$ είναι σχεδόν ταυτόσημες.
- Μέση τιμή: 0 (για $\nu > 1$). Διασπορά: $\nu/(\nu-2)$ (για $\nu > 2$) — μεγαλύτερη από 1, και τείνει στο 1 καθώς $\nu \rightarrow \infty$.

Πού εμφανίζεται: σε όλους τους ελέγχους και διαστήματα εμπιστοσύνης για τη μέση τιμή όταν η σ είναι άγνωστη — που είναι ουσιαστικά πάντα στην πράξη.



Σχήμα: Η t_ν έχει παχύτερες ουρές για μικρά ν και συγκλίνει στην $N(0,1)$ όσο $\nu \rightarrow \infty$.

Πρακτική συνέπεια: Οι κρίσιμες τιμές της t είναι μεγαλύτερες από εκείνες της κανονικής. Π.χ. για αμφίπλευρο 95%:

ν	5	10	20	30	100	∞ (κανονική)
$t_{\nu,0,025}$	2,571	2,228	2,086	2,042	1,984	1,960

Με μικρό δείγμα ζητάμε «μεγαλύτερη απόδειξη» πριν απορρίψουμε — αυτό αντικατοπτρίζει τη μεγαλύτερη αβεβαιότητα από την εκτίμηση του σ .

Η κατανομή χ^2 (Χι-τετράγωνο)

Η κατανομή χ^2 με k βαθμούς ελευθερίας ορίζεται ως το άθροισμα τετραγώνων ανεξάρτητων τυπικών κανονικών:

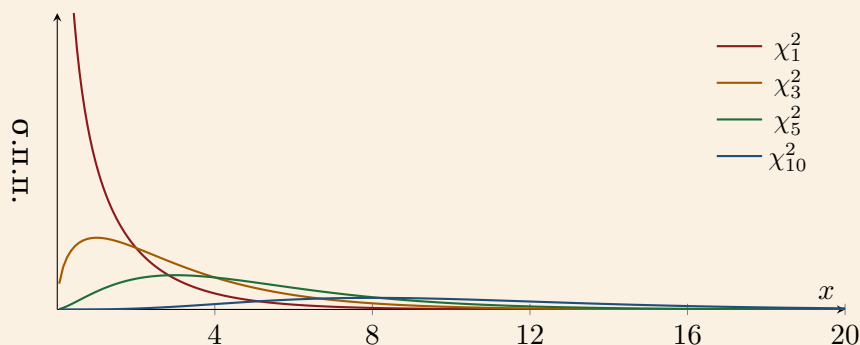
$$\text{Αν } Z_1, \dots, Z_k \stackrel{\text{iid}}{\sim} N(0,1), \text{ τότε } Z_1^2 + Z_2^2 + \dots + Z_k^2 \sim \chi_k^2.$$

Βασικές ιδιότητες:

- Παίρνει μόνο μη αρνητικές τιμές ($X \geq 0$) — αφού είναι άθροισμα τετραγώνων.
- **Ασύμμετρο** (δεξιά λοξή / right-skewed), ειδικά για μικρά k . Όσο μεγαλώνει το k , η κατανομή γίνεται πιο συμμετρική και πλησιάζει την κανονική.
- Μέση τιμή: $\mathbb{E}(X) = k$. Διασπορά: $\text{Var}(X) = 2k$.
- **Προσθετική ιδιότητα:** αν $X \sim \chi_{k_1}^2$ και $Y \sim \chi_{k_2}^2$ ανεξάρτητες, τότε $X + Y \sim \chi_{k_1+k_2}^2$.

Πού εμφανίζεται στη Στατιστική II:

- Δειγματική διασπορά: αν $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, τότε $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.
- Έλεγχος διασποράς (Θεώρημα 3.3).
- Ασυμπτωτική κατανομή της τριάδας LR/Wald/LM: όλες $\xrightarrow{d} \chi_1^2$ (Θεώρημα 3.6).
- Έλεγχοι καλής προσαρμογής (chi-square goodness-of-fit).



Σχήμα: Η χ_k^2 ξεκινά εξαιρετικά λοξή για $k = 1$ και γίνεται όλο και πιο συμμετρική όσο μεγαλώνει το k .

Σχέση με την κατανομή t : Η t -στατιστική γεννιέται ως πηλίκο κανονικής προς ρίζα χ -τετραγώνου διηρημένου με τους β.ε.:

$$T = \frac{Z}{\sqrt{V/\nu}}, \quad Z \sim N(0, 1), \quad V \sim \chi_\nu^2, \quad \text{ανεξάρτητες} \implies T \sim t_\nu.$$

Δηλαδή τα t και χ^2 είναι «αδέρφια» — μαζί με την κανονική, αποτελούν τη βασική τριάδα κατανομών δειγματοληψίας.

Κρίσιμες τιμές για άνω μονόπλευρο σε $\alpha = 0,05$:

k	1	5	10	20	30	50
$\chi_{k,0,05}^2$	3,841	11,07	18,31	31,41	43,77	67,50

1.5 Έλεγχοι δύο δειγμάτων

Ανεξάρτητα δείγματα (independent samples) — Welch t -test. Για δύο ανεξάρτητα δείγματα από $N(\mu_X, \sigma_X^2)$ και $N(\mu_Y, \sigma_Y^2)$ μεγέθους n_X και n_Y , ο έλεγχος $H_0 : \mu_X = \mu_Y$ χρησιμοποιεί:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_X^2/n_X + S_Y^2/n_Y}}$$

η οποία υπό H_0 ακολουθεί προσεγγιστικά Student- t με βαθμούς ελευθερίας που υπολογίζονται από τον τύπο Welch-Satterthwaite.

Ζεύγος δειγμάτων (paired samples). Όταν έχουμε ζευγαρωμένες παρατηρήσεις (π.χ. πριν-μετά για κάθε υποκείμενο), ορίζουμε $D_i = X_i - Y_i$ και εκτελούμε έναν συνηθισμένο t -test μιας μεταβλητής για $H_0 : \mu_D = 0$ στο διάνυσμα $\{D_i\}$.

1.6 Σύνδεση μεταξύ ελέγχων και διαστημάτων εμπιστοσύνης

Θεώρημα 3.4 (Δυσκότητα Ελέγχου-Διαστήματος)

Ένα $(1 - \alpha) \cdot 100\%$ διάστημα εμπιστοσύνης για το θ αποτελείται από όλες τις τιμές θ_0 που δεν απορρίπτονται από έναν αμφίπλευρο έλεγχο επιπέδου α στην $H_0 : \theta = \theta_0$.

Παρατήρηση

Αυτή η δυϊκότητα είναι θεμελιώδης: ένα διάστημα εμπιστοσύνης είναι «η συλλογή των μη απορριπτόμενων υποθέσεων». Σας λέει αμέσως ποιες τιμές της παραμέτρου είναι συμβατές με τα δεδομένα.

1.7 Λήμμα Neyman-Pearson**Θεώρημα 3.5 (Λήμμα Neyman-Pearson)**

Για έλεγχο μεταξύ δύο απλών υποθέσεων $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, ο **πιο ισχυρός** (Most Powerful — MP) έλεγχος επιπέδου α απορρίπτει την H_0 όταν ο λόγος πιθανοφάνειας

$$\Lambda(\mathbf{x}) = \frac{L(\theta_1; \mathbf{x})}{L(\theta_0; \mathbf{x})} > k,$$

όπου το k επιλέγεται έτσι ώστε $\mathbb{P}(\Lambda > k \mid H_0) = \alpha$.

Σημασία

Το Λήμμα μάς λέει ότι μεταξύ όλων των πιθανών στατιστικών ελέγχου επιπέδου α , ο λόγος πιθανοφάνειας μεγιστοποιεί την ισχύ. Αυτό είναι ο πρόγονος του γενικευμένου λόγου πιθανοφάνειας (Generalized LR).

2 Λυμένες Ασκήσεις**Άσκηση 1: Z-test για μέσο μισθό + υπολογισμός ισχύος**

Ο εργοδοτικός σύλλογος ενός κλάδου ισχυρίζεται ότι ο μέσος μηνιαίος μισθός είναι $\mu_0 = 1.500$ €. Ένας ερευνητής υποπτεύεται ότι είναι χαμηλότερος και θέλει να τον ελέγξει. Παίρνει τυχαίο δείγμα $n = 36$ εργαζομένων και βρίσκει $\bar{x} = 1.430$ €. Δεχόμαστε ότι η τυπική απόκλιση των μισθών είναι γνωστή: $\sigma = 240$ €.

(α') Διατυπώστε H_0 και H_1 (μονόπλευρα).

(β') Εκτελέστε τον έλεγχο σε επίπεδο σημαντικότητας $\alpha = 0,05$.

(γ') Υπολογίστε το p -value.

(δ') Υπολογίστε την ισχύ του ελέγχου αν η αληθινή μέση τιμή είναι $\mu = 1.400$ €.

Λύση

(α) Αφού ο ερευνητής υποπτεύεται χαμηλότερο μισθό:

$$H_0 : \mu = 1.500 \quad \text{vs} \quad H_1 : \mu < 1.500.$$

(β) Στατιστική έλεγχου:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.430 - 1.500}{240/\sqrt{36}} = \frac{-70}{40} = -1,75.$$

Για μονόπλευρο έλεγχο σε $\alpha = 0,05$, η κρίσιμη τιμή είναι $-z_{0,05} = -1,645$. Επειδή

$$Z_{\text{obs}} = -1,75 < -1,645,$$

απορρίπτουμε την H_0 : υπάρχει στατιστική απόδειξη ότι ο μέσος μισθός είναι χαμηλότερος από $1.500 \in$.

(γ) p -value για αριστερό μονόπλευρο έλεγχο:

$$p = \mathbb{P}(Z \leq -1,75) = \Phi(-1,75) \approx 0,0401.$$

Αφού $p \approx 0,0401 < 0,05$, απορρίπτουμε. ✓ (Συμφωνεί με το προηγούμενο βήμα.)

(δ) Ισχύς υπό $\mu = 1.400$. Απορρίπτουμε όταν $Z < -1,645$, δηλαδή όταν

$$\bar{X} < \mu_0 - 1,645 \cdot \frac{\sigma}{\sqrt{n}} = 1.500 - 1,645 \cdot 40 = 1.500 - 65,8 = 1.434,2.$$

Υπό $\mu = 1.400$, το $\bar{X} \sim N(1.400, (40)^2)$. Άρα

$$\text{Ισχύς} = \mathbb{P}(\bar{X} < 1.434,2 \mid \mu = 1.400) = \mathbb{P}\left(Z < \frac{1.434,2 - 1.400}{40}\right) = \mathbb{P}(Z < 0,855) \approx 0,8037.$$

Δηλαδή ο έλεγχος έχει $\approx 80\%$ ισχύ να εντοπίσει διαφορά $100 \in$ στον μέσο μισθό. Πρακτικά αποδεκτό (κανόνας του 80%).

Άσκηση 2: One-sample t -test για απόδοση μετοχής

Ένας διαχειριστής χαρτοφυλακίου ισχυρίζεται ότι η αναμενόμενη ημερήσια απόδοση μιας μετοχής είναι $\mu_0 = 0,05\%$. Σε δείγμα $n = 20$ τυχαίων ημερών παρατηρούνται $\bar{x} = 0,03\%$ και $s = 0,08\%$. (Υποθέτουμε κανονικότητα.)

(α') Ελέγξτε αμφίπλευρα την υπόθεση σε $\alpha = 0,05$.

(β') Δώστε το p -value (κατά προσέγγιση).

(γ') Δώστε ένα 95% διάστημα εμπιστοσύνης για το μ και χρησιμοποιήστε τη δυϊκότητα ελέγχου-διαστήματος για να επιβεβαιώσετε την απόφαση.

Λύση

(α) $H_0 : \mu = 0,05$ vs $H_1 : \mu \neq 0,05$. Στατιστική:

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0,03 - 0,05}{0,08/\sqrt{20}} = \frac{-0,02}{0,01789} \approx -1,118.$$

Κρίσιμη τιμή σε $\alpha = 0,05$ αμφίπλευρα με $n - 1 = 19$ βαθμούς ελευθερίας: $t_{19,0,025} \approx 2,093$. Επειδή $|T_{\text{obs}}| = 1,118 < 2,093$, **δεν απορρίπτουμε την H_0** .

(β) Από πίνακες t_{19} , $|T| = 1,118$ αντιστοιχεί περίπου σε αμφίπλευρο p -value $\approx 0,28$. Πολύ μεγαλύτερο του 0,05 — δεν απορρίπτουμε.

(γ) Διάστημα εμπιστοσύνης 95%:

$$\bar{x} \pm t_{19,0,025} \cdot \frac{s}{\sqrt{n}} = 0,03 \pm 2,093 \cdot 0,01789 = 0,03 \pm 0,0374.$$

Δηλαδή $[-0,0074\%, 0,0674\%]$. Το $\mu_0 = 0,05$ *περιέχεται* στο διάστημα $\rightarrow$ δεν απορρίπτουμε (δυϊκότητα). ✓

Άσκηση 3: Paired t -test — αξιολόγηση πολιτικής

Ένας δήμος εφαρμόζει νέο πρόγραμμα επιδότησης ενεργειακής αναβάθμισης σε νοικοκυριά. Για 10 νοικοκυριά καταγράφεται η ετήσια κατανάλωση ηλεκτρισμού (kWh) πριν (X) και μετά (Y):

Νοικ.	1	2	3	4	5	6	7	8	9	10
X	4500	5200	3800	6000	4700	5500	4100	5800	4900	5300
Y	4200	4900	3700	5400	4500	5100	4000	5400	4600	5000

Ελέγξτε σε $\alpha = 0,05$ αν το πρόγραμμα μειώνει σημαντικά την κατανάλωση.

Λύση

Ορίζουμε $D_i = X_i - Y_i$ (μείωση κατανάλωσης για κάθε νοικοκυριό). Υπολογίζουμε:

$$D = (300, 300, 100, 600, 200, 400, 100, 400, 300, 300).$$

$$\bar{D} = \frac{300+300+100+600+200+400+100+400+300+300}{10} = \frac{3000}{10} = 300.$$

Για το s_D , υπολογίζουμε αποκλίσεις από τη μέση: $(0, 0, -200, 300, -100, 100, -200, 100, 0, 0)$.

Τετράγωνα: $(0, 0, 40000, 90000, 10000, 10000, 40000, 10000, 0, 0)$. Άθροισμα: 200.000. Άρα

$$s_D^2 = \frac{200.000}{n-1} = \frac{200.000}{9} \approx 22.222,2, \quad s_D \approx 149,07.$$

Έλεγχος: $H_0 : \mu_D = 0$ vs $H_1 : \mu_D > 0$ (μονόπλευρα, αναμένουμε μείωση $\rightarrow D > 0$).

$$T = \frac{\bar{D} - 0}{s_D/\sqrt{n}} = \frac{300}{149,07/\sqrt{10}} = \frac{300}{47,14} \approx 6,364.$$

Κρίσιμη τιμή: $t_{9,0,05} \approx 1,833$. Αφού $T = 6,364 \gg 1,833$, **απορρίπτουμε ισχυρά την H_0** . Το πρόγραμμα μειώνει σημαντικά την κατανάλωση. (p -value $\ll 0,001$.)

Σημείο εκτίμησης της μείωσης: $\bar{D} = 300$ kWh ετησίως, με τυπικό σφάλμα $s_D/\sqrt{n} \approx 47,14$.

Παρατήρηση

Ο paired t -test είναι πιο ισχυρός από έναν two-sample test εδώ, γιατί ελέγχει τη μεταβλητότητα μεταξύ νοικοκυριών (κάθε νοικοκυριό συγκρίνεται με τον εαυτό του). Στην οικονομετρία, αυτό είναι το «fixed effects» στα panel data — αφαιρούμε τη σταθερή ατομική επίδραση κάθε υποκειμένου.

Άσκηση 4: Two-sample t -test (Welch) — σύγκριση δύο κλάδων

Συγκρίνουμε τα μηνιαία έσοδα (σε χιλ. €) ανά εργαζόμενο σε δύο κλάδους:

- Τεχνολογία (X): $n_X = 25$, $\bar{x} = 7,2$, $s_X = 1,8$.
- Παραδοσιακή βιομηχανία (Y): $n_Y = 30$, $\bar{y} = 5,8$, $s_Y = 2,5$.

Ελέγξτε σε $\alpha = 0,05$ αν τα μέσα έσοδα διαφέρουν.

Λύση

$H_0 : \mu_X = \mu_Y$ vs $H_1 : \mu_X \neq \mu_Y$. Στατιστική Welch:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}} = \frac{7,2 - 5,8}{\sqrt{1,8^2/25 + 2,5^2/30}} = \frac{1,4}{\sqrt{0,1296 + 0,2083}} = \frac{1,4}{\sqrt{0,3379}} \approx 2,409$$

Βαθμοί ελευθερίας Welch-Satterthwaite:

$$\nu \approx \frac{(s_X^2/n_X + s_Y^2/n_Y)^2}{(s_X^2/n_X)^2/(n_X - 1) + (s_Y^2/n_Y)^2/(n_Y - 1)}.$$

Αριθμητής: $0,3379^2 \approx 0,1142$. Παρονομαστής: $0,1296^2/24 + 0,2083^2/29 \approx 0,0007 + 0,00149 \approx 0,00219$.

$$\nu \approx \frac{0,1142}{0,00219} \approx 52,1 \implies \text{στρογγύλεμα προς τα κάτω: } \nu = 52.$$

Κρίσιμη τιμή: $t_{52,0,025} \approx 2,007$. Αφού $|T| = 2,409 > 2,007$, **απορρίπτουμε την H_0** . Υπάρχει στατιστικά σημαντική διαφορά στα μέσα έσοδα.

p-value: κατά προσέγγιση 0,02 αμφίπλευρα.

Άσκηση 5: χ^2 -test για διασπορά — risk management

Ένα risk management team απαιτεί η ημερήσια μεταβλητότητα ενός portfolio να μην υπερβαίνει $\sigma_0 = 1\%$ (δηλαδή $\sigma_0^2 = 1$). Σε δείγμα $n = 25$ ημερών, η δειγματική τυπική απόκλιση είναι $s = 1,30\%$ (δηλαδή $s^2 = 1,69$). Υποθέτουμε κανονικότητα.

Ελέγξτε σε $\alpha = 0,05$ αν η μεταβλητότητα έχει αυξηθεί πάνω από το όριο.

Λύση

$H_0 : \sigma^2 = 1$ vs $H_1 : \sigma^2 > 1$. Στατιστική:

$$W = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \cdot 1,69}{1} = 40,56.$$

Κρίσιμη τιμή για άνω μονόπλευρο: $\chi_{24,0,05}^2 \approx 36,415$. Αφού $W = 40,56 > 36,415$, **απορρίπτουμε την H_0** : η μεταβλητότητα έχει αυξηθεί στατιστικά σημαντικά. Το risk team χρειάζεται να αναθεωρήσει τα όρια.

p-value: $\mathbb{P}(\chi_{24}^2 > 40,56) \approx 0,019$.

Σύνδεση με την Οικονομετρία

Στα μοντέλα GARCH (χρηματοοικονομική οικονομετρία), ο έλεγχος για «ARCH effects» (μη σταθερή μεταβλητότητα) βασίζεται σε ένα χ^2 test πολύ παρόμοιο με αυτό — απλώς εφαρμοσμένο στα τετράγωνα των καταλοίπων αντί στις αρχικές παρατηρήσεις.

Άσκηση 6: Λήμμα Neyman-Pearson — MP test για Bernoulli

Ένα κατάστημα ελέγχει την ποιότητα μιας παρτίδας προϊόντων. Έστω $X_i \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$, όπου $X_i = 1$ αν το προϊόν είναι ελαττωματικό. Θέλουμε να επιλέξουμε μεταξύ:

$$H_0 : p = 0,05 \quad \text{vs} \quad H_1 : p = 0,10.$$

Δίνεται $n = 100$ προϊόντα. Βρείτε τη μορφή του πιο ισχυρού (MP) ελέγχου σε επίπεδο $\alpha = 0,05$.

Λύση

Από το Λήμμα Neyman–Pearson, ο MP έλεγχος βασίζεται στον λόγο πιθανοφάνειας:

$$\begin{aligned}\Lambda(\mathbf{x}) &= \frac{L(0,10)}{L(0,05)} = \frac{(0,10)^{\sum x_i} (0,90)^{n-\sum x_i}}{(0,05)^{\sum x_i} (0,95)^{n-\sum x_i}} = \left(\frac{0,10}{0,05}\right)^{\sum x_i} \left(\frac{0,90}{0,95}\right)^{n-\sum x_i} \\ &= 2^{\sum x_i} \left(\frac{18}{19}\right)^{n-\sum x_i}.\end{aligned}$$

Παίρνουμε λογάριθμο:

$$\ln \Lambda = (\ln 2) \sum x_i + (n - \sum x_i) \ln(18/19) = \sum x_i [\ln 2 - \ln(18/19)] + n \ln(18/19).$$

Ο πρώτος συντελεστής $\ln 2 - \ln(18/19) \approx 0,693 + 0,054 = 0,747 > 0$. Άρα η $\ln \Lambda$ είναι γνησίως αύξουσα στο $\sum x_i$.

Συνεπώς ο κανόνας «απορρίπτω αν $\Lambda > k$ » ισοδυναμεί με «απορρίπτω αν $\sum x_i > c$ » για κάποιο σταθερό c .

Επιλογή του c ώστε $\alpha = 0,05$: Υπό H_0 , $\sum X_i \sim \text{Binomial}(100, 0,05)$, με μέση 5 και διασπορά 4,75. Από κανονική προσέγγιση, $\sum X_i \approx N(5, 4,75)$. Θέλουμε

$$\mathbb{P}\left(\sum X_i > c \mid p = 0,05\right) = 0,05.$$

$$c \approx 5 + 1,645\sqrt{4,75} \approx 5 + 3,58 \approx 8,58 \implies \text{απορρίπτω αν } \sum X_i \geq 9.$$

(Στην πράξη χρησιμοποιούμε την ακριβή binomial CDF: $\mathbb{P}(\sum X_i \geq 9 \mid p = 0,05) \approx 0,0628$, και $\mathbb{P}(\sum X_i \geq 10 \mid p = 0,05) \approx 0,0282$.)

Συμπέρασμα: ο MP έλεγχος απορρίπτει την H_0 αν παρατηρηθούν *πέρα πολλά* ελαττωματικά προϊόντα — διαισθητικά σωστό. Το λήμμα Neyman–Pearson επιβεβαιώνει ότι αυτή είναι η *βέλτιστη* στρατηγική: κανένας άλλος έλεγχος επιπέδου 0,05 δεν μπορεί να έχει μεγαλύτερη ισχύ στο $p = 0,10$.

Παρατήρηση

Αυτό το παράδειγμα δείχνει ότι ο λόγος πιθανοφάνειας είναι «φυσικά» η βέλτιστη στατιστική ελέγχου. Στο Δίωρο 2 θα γενικεύσουμε αυτή την ιδέα σε σύνθετες υποθέσεις (Generalized LR), και θα δούμε ότι ισοδύναμα μπορούμε να εργαστούμε με τη score (LM) ή τον MLE (Wald).

3 Η Τριάδα LR / Wald / LM

3.1 Το γενικό πλαίσιο

Στα παραπάνω είδαμε κλασικούς ελέγχους που βασίζονται σε γνωστές κατανομές (κανονική, t , χ^2). Σε σύγχρονες οικονομετρικές εφαρμογές όμως, η ακριβής δειγματοληπτική κατανομή δεν είναι γνωστή — διαθέτουμε μόνο τις *ασυμπτωτικές* ιδιότητες του MLE της προηγούμενης ενότητας. Από αυτές προκύπτουν τρεις γενικοί ασυμπτωτικοί έλεγχοι, η περίφημη **τριάδα**:

Πλαίσιο: έστω $X_1, \dots, X_n$ τυχαίο δείγμα με κοινή σ.π.π. $f(x; \theta)$, και θέλουμε να ελέγξουμε

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0.$$

Ορίζουμε:

- $\hat{\theta}_{\text{ML}}$ = MLE από όλο τον παραμετρικό χώρο (unrestricted).
- $\ell(\theta)$ = λογαριθμική πιθανοφάνεια.
- $S(\theta) = \ell'(\theta)$ = συνάρτηση score.
- $I_n(\theta) = nI_1(\theta)$ = πληροφορία Fisher.

Ορισμός 3.5 (Τριάδα ελέγχων)

Likelihood Ratio (LR) test:

$$LR = 2[\ell(\hat{\theta}_{\text{ML}}) - \ell(\theta_0)].$$

Συγκρίνει τις τιμές της log-likelihood στον unrestricted MLE και στην υπόθεση.

Wald test:

$$W = (\hat{\theta}_{\text{ML}} - \theta_0)^2 \cdot I_n(\hat{\theta}_{\text{ML}}).$$

Βασίζεται στην απόκλιση του MLE από την θ_0 , σταθμισμένη με την (εκτιμώμενη) πληροφορία.

Lagrange Multiplier (LM) / Score test:

$$LM = \frac{[S(\theta_0)]^2}{I_n(\theta_0)}.$$

Βασίζεται στην κλίση της log-likelihood στην θ_0 — αν είναι μηδέν, η θ_0 είναι όντως η μεγιστοποιούσα.

Θεώρημα 3.6 (Ασυμπτωτική κατανομή)

Κάτω από την H_0 και τις συνηθισμένες συνθήκες ομαλότητας:

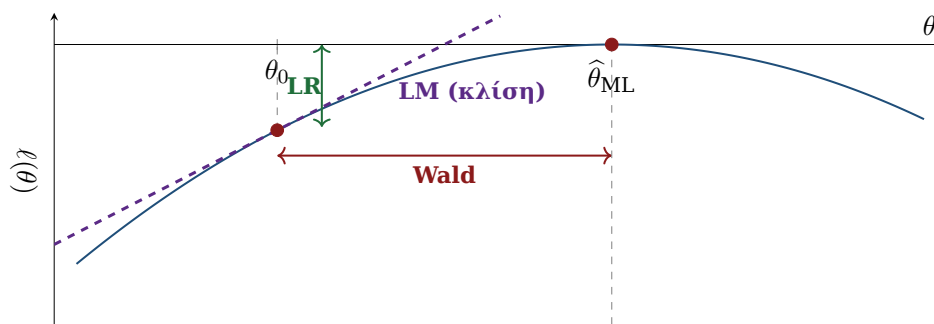
$$LR \xrightarrow{d} \chi_1^2, \quad W \xrightarrow{d} \chi_1^2, \quad LM \xrightarrow{d} \chi_1^2.$$

Απορρίπτουμε την H_0 σε επίπεδο α αν η αντίστοιχη στατιστική υπερβαίνει το $\chi_{1,\alpha}^2$.

Γεωμετρική διαίσθηση των τριών ελέγχων

Φαντάσου τη log-likelihood $\ell(\theta)$ ως καμπύλη με μέγιστο στο $\hat{\theta}_{ML}$. Καθένας έλεγχος μετράει με διαφορετικό τρόπο «πόσο μακριά» από τη θ_0 είμαστε:

- **LR**: μετράει *κάθετα* — πόσο χαμηλά πέφτει η ℓ στο θ_0 σε σχέση με το $\hat{\theta}_{ML}$.
- **Wald**: μετράει *οριζόντια* — πόσο μακριά είναι το $\hat{\theta}_{ML}$ από τη θ_0 , σταθμισμένο με την καμπυλότητα.
- **LM (Score)**: μετράει την *κλίση* της ℓ στη θ_0 — αν η θ_0 είναι κορυφή, η κλίση εκεί πρέπει να είναι 0.



Σχήμα 2: Οι τρεις έλεγχοι μετρούν διαφορετικά τη «διαφορά» μεταξύ θ_0 και $\hat{\theta}_{ML}$: LR = κάθετο χάσμα στη ℓ , Wald = οριζόντια απόσταση, LM = κλίση στο θ_0 .

Πότε διαφέρουν

Ασυμπτωτικά ($n \rightarrow \infty$): οι τρεις στατιστικές είναι *ασυμπτωτικά ισοδύναμες* — έχουν την ίδια οριακή κατανομή χ_1^2 και δίνουν την ίδια απόφαση με πιθανότητα 1.

Σε πεπερασμένα δείγματα: γενικά διαφέρουν αριθμητικά. Υπάρχει η περίφημη **ανισότητα**

$$W \geq LR \geq LM$$

που ισχύει συχνά (αλλά όχι πάντα) στην πράξη.

Ειδική περίπτωση: όταν η log-likelihood είναι ακριβώς τετραγωνική στο θ (όπως στην κανονική με γνωστή σ), οι τρεις στατιστικές συμπίπτουν αλγεβρικά. Αυτό θα το δούμε στις Ασκήσεις 7–9.

Υπολογιστικό κριτήριο επιλογής

- **LR**: χρειάζεται και τον unrestricted MLE και την τιμή της log-likelihood στο θ_0 .
- **Wald**: χρειάζεται μόνο τον unrestricted MLE και την Fisher στο $\hat{\theta}$.
- **LM**: χρειάζεται μόνο την υπόθεση θ_0 (όχι τον MLE!) και την Fisher στο θ_0 .

Έτσι, αν ο MLE είναι δύσκολο να υπολογιστεί (π.χ. απαιτεί αριθμητική βελτιστοποίηση), προτιμάμε το LM. Αν η log-likelihood στο θ_0 είναι δύσκολη, προτιμάμε το Wald.

Σύνδεση με την Οικονομετρία**Παραδείγματα από την Οικονομετρία:**

- **LM test:** ο Breusch-Pagan test για ετεροσκεδαστικότητα, ο Breusch-Godfrey για αυτοσυσχέτιση, και ο White test είναι όλοι LM tests — επειδή μπορούν να υπολογιστούν χωρίς να εκτιμηθεί το πλήρες μοντέλο.
- **Wald test:** το κλασικό t -statistic σε OLS παλινδρόμηση είναι ένας Wald test (t^2 είναι μια Wald στατιστική για συντελεστή = 0).
- **LR test:** η σύγκριση «restricted» vs «unrestricted» μοντέλου, π.χ. έλεγχος συντελεστών σε logit/probit.

4 Λυμένες Ασκήσεις**4.1 Ομάδα Α: Τριάδα στην κανονική κατανομή — όπου όλα συμπύπτουν****Άσκηση 7: LR test για μέση τιμή κανονικής (γνωστή σ^2)**

Έστω $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, με σ^2 γνωστή. Έλεγχος $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.

(α') Γράψτε τη log-likelihood $\ell(\mu)$ και τη διαφορά $\ell(\bar{X}) - \ell(\mu_0)$.

(β') Παράγετε ρητά τη στατιστική $LR = 2[\ell(\bar{X}) - \ell(\mu_0)]$ και δείξτε ότι

$$LR = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2}.$$

(γ') Δείξτε ότι $LR = Z^2$, όπου $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ είναι η κλασική στατιστική του Z -test.

Λύση**(α)**

$$\ell(\mu) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Από την Εβδομάδα 2 (Άσκηση 7), $\hat{\mu}_{ML} = \bar{X}$. Άρα:

$$\ell(\bar{X}) - \ell(\mu_0) = -\frac{1}{2\sigma^2} \left[\sum (x_i - \bar{X})^2 - \sum (x_i - \mu_0)^2 \right].$$

Χρησιμοποιούμε την ταυτότητα της Εβδομάδας 1: $\sum (x_i - \mu_0)^2 = \sum (x_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2$. Άρα

$$\sum (x_i - \bar{X})^2 - \sum (x_i - \mu_0)^2 = -n(\bar{X} - \mu_0)^2.$$

Συνεπώς

$$\ell(\bar{X}) - \ell(\mu_0) = \frac{n(\bar{X} - \mu_0)^2}{2\sigma^2}.$$

(β)

$$LR = 2 \cdot \frac{n(\bar{X} - \mu_0)^2}{2\sigma^2} = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2}.$$

(γ) Παρατηρώ ότι

$$Z^2 = \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right)^2 = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2} = LR. \quad \checkmark$$

Αφού υπό H_0 , $Z \sim N(0,1)$, ισχύει $Z^2 \sim \chi_1^2$. Άρα και $LR \sim \chi_1^2$, σύμφωνα με το Θεώρημα 3.6.

Άσκηση 8: Wald test για το ίδιο πρόβλημα

Στις ίδιες συνθήκες με την Άσκηση 7, παράγετε τη στατιστική Wald

$$W = (\hat{\mu}_{ML} - \mu_0)^2 \cdot I_n(\hat{\mu}_{ML}).$$

Λύση

Από την Εβδομάδα 2 (Άσκηση 3), για κανονική με γνωστή σ^2 :

$$I_1(\mu) = \frac{1}{\sigma^2}, \quad I_n(\mu) = \frac{n}{\sigma^2}.$$

Παρατηρήστε ότι η Fisher δεν εξαρτάται από μ στην κανονική με γνωστή σ^2 — άρα $I_n(\hat{\mu}) = I_n(\mu_0) = n/\sigma^2$. Συνεπώς:

$$W = (\bar{X} - \mu_0)^2 \cdot \frac{n}{\sigma^2} = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2}.$$

Παρατηρώ ότι $W = LR$! Δηλαδή σε αυτή την περίπτωση οι δύο στατιστικές είναι αλγεβρικά ταυτόσημες.

Άσκηση 9: LM (Score) test για το ίδιο πρόβλημα

Στις ίδιες συνθήκες, παράγετε τη στατιστική LM

$$LM = \frac{[S(\mu_0)]^2}{I_n(\mu_0)}.$$

Λύση

Score (Εβδομάδα 2, Άσκηση 3):

$$S(\mu) = \frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) = \frac{n(\bar{X} - \mu)}{\sigma^2}.$$

Στο $\mu = \mu_0$:

$$S(\mu_0) = \frac{n(\bar{X} - \mu_0)}{\sigma^2}.$$

Πληροφορία: $I_n(\mu_0) = n/\sigma^2$. Άρα

$$LM = \frac{[n(\bar{X} - \mu_0)/\sigma^2]^2}{n/\sigma^2} = \frac{n^2(\bar{X} - \mu_0)^2/\sigma^4}{n/\sigma^2} = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2}.$$

$$LM = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2} = LR = W.$$

Έλεγχος: αλγεβρική ταυτότητα των τριών στατιστικών

$$LR = W = LM = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2} = Z^2.$$

Και οι τρεις στατιστικές είναι **ταυτόσημες** σε αυτό το πρόβλημα, και είναι ίσες με το τετράγωνο της κλασικής Z -στατιστικής. Άρα ο κλασικός Z -test είναι ένας ειδικός LR / Wald / LM test.

Γιατί συμπίπτουν εδώ

Η log-likelihood της κανονικής (ως προς μ , με γνωστή σ^2) είναι ακριβώς τετραγωνική:

$$\ell(\mu) = \text{const} - \frac{n}{2\sigma^2}(\mu - \bar{X})^2 + \text{const}'.$$

Σε τετραγωνική ℓ :

- Το «κάθετο χάσμα» (LR) ισούται με το τετράγωνο της οριζόντιας απόστασης (Wald), σταθμισμένο με την καμπυλότητα.
- Η κλίση στη θ_0 (LM) είναι ευθεία γραμμή της απόστασης από τον MLE — οπότε το τετράγωνό της επίσης συμπίπτει.

Σε μη-τετραγωνική log-likelihood (π.χ. Bernoulli, Poisson, λογιστική παλινδρόμηση κ.λπ.), οι τρεις στατιστικές διαφέρουν αριθμητικά σε πεπερασμένα δείγματα. Αυτό θα το δούμε στην Άσκηση 10.

4.2 Ομάδα B: Unified case study — Bernoulli, όπου διαφέρουν**Άσκηση 10: Unified LR/Wald/LM σε Bernoulli — πρόγραμμα marketing**

Μια εταιρεία υποστηρίζει ότι η πιθανότητα ένας τυχαίος πελάτης να αγοράσει μετά από διαφημιστική εκστρατεία είναι $p_0 = 0,20$. Σε τυχαίο δείγμα $n = 100$ πελατών, $\sum x_i = 27$ αγόρασαν (δηλαδή $\bar{X} = 0,27$).

Ελέγξτε $H_0 : p = 0,20$ vs $H_1 : p \neq 0,20$ και με τις **τρεις** στατιστικές της τριάδας, σε $\alpha = 0,05$.

(α') Υπολογίστε τη **LR**.

(β') Υπολογίστε το **Wald**.

(γ') Υπολογίστε το **LM**.

(δ') Συγκρίνετε αριθμητικά τις τρεις τιμές. Συμφωνεί η ανισότητα $W \geq LR \geq LM$;

(ε') Δώστε τις αποφάσεις (όλες σε $\chi^2_{1,0,05} = 3,841$).

Λύση

Προετοιμασία: Από τα προηγούμενα:

- $\hat{p}_{ML} = \bar{X} = 0,27$.
- $I_1(p) = \frac{1}{p(1-p)}, I_n(p) = \frac{n}{p(1-p)}.$
- Score: $S(p) = \frac{n(\bar{X} - p)}{p(1-p)}.$

$$\bullet \ell(p) = (\sum x_i) \ln p + (n - \sum x_i) \ln(1 - p) = 27 \ln p + 73 \ln(1 - p)$$

(α) LR.

$$\ell(\hat{p}) = 27 \ln(0,27) + 73 \ln(0,73) = 27 \cdot (-1,3093) + 73 \cdot (-0,3147) = -58,3270$$

$$\ell(p_0) = 27 \ln(0,20) + 73 \ln(0,80) = 27 \cdot (-1,6094) + 73 \cdot (-0,2231) = -59,7425$$

$$LR = 2[\ell(\hat{p}) - \ell(p_0)] = 2[-58,3270 - (-59,7425)] = 2 \cdot 1,4155 = \boxed{2,8310}$$

(β) Wald.

$$W = (\hat{p} - p_0)^2 \cdot I_n(\hat{p}) = (0,27 - 0,20)^2 \cdot \frac{100}{0,27 \cdot 0,73}$$

$$= 0,0049 \cdot \frac{100}{0,1971} = 0,0049 \cdot 507,36 = \boxed{2,486}$$

(γ) LM.

$$S(p_0) = \frac{n(\bar{X} - p_0)}{p_0(1 - p_0)} = \frac{100 \cdot 0,07}{0,20 \cdot 0,80} = \frac{7}{0,16} = 43,75$$

$$I_n(p_0) = \frac{100}{0,20 \cdot 0,80} = \frac{100}{0,16} = 625$$

$$LM = \frac{[S(p_0)]^2}{I_n(p_0)} = \frac{43,75^2}{625} = \frac{1914,0625}{625} = \boxed{3,0625}$$

(δ) Σύγκριση.

Έλεγχος	Στατιστική	Απόφαση ($\alpha = 0,05$ με κρίσιμη 3,841)
LR	2,831	Δεν απορρίπτω
Wald	2,486	Δεν απορρίπτω
LM (Score)	3,063	Δεν απορρίπτω

Παρατηρούμε:

$$LM (3,06) > LR (2,83) > W (2,49)$$

Δηλαδή η συνήθης ανισότητα $W \geq LR \geq LM$ δεν ισχύει εδώ — στην παρούσα περίπτωση η σειρά αντιστρέφεται! Αυτό απλώς υπενθυμίζει ότι η $W \geq LR \geq LM$ είναι «κανόνας του παλιού καιρού», όχι θεώρημα.

(ε) Συμπέρασμα: Και οι τρεις έλεγχοι συμφωνούν στην απόφαση: δεν απορρίπτουμε την H_0 . Τα δεδομένα δεν είναι αρκετά για να αμφισβητήσουμε τον ισχυρισμό $p = 0,20$ — παρότι το παρατηρούμενο $\hat{p} = 0,27$ είναι αρκετά υψηλότερο.

Έλεγχος συνέπειας — από Άσκ. 1 της προηγούμενης ενότητας

Στην προηγούμενη ενότητα αποδείξαμε ότι ο $\hat{p} = \bar{X}$ είναι αποτελεσματικός (πιάνει το CRLB), με ασυμπτωτική διασπορά $p(1 - p)/n$. Η κλασική στατιστική Z για αναλογίες είναι

$$Z = \frac{\bar{X} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0,07}{\sqrt{0,16/100}} = \frac{0,07}{0,04} = 1,75.$$

Τότε $Z^2 = 3,0625 = LM$. Αυτό δεν είναι σύμπτωση: η LM στατιστική για Bernoulli στην παρούσα μορφή είναι ακριβώς το τετράγωνο του κλασικού Z -test για ανα-

λογία (με τη διασπορά υπολογισμένη υπό την H_0).

Δύο εκδοχές του Z-test για αναλογίες

- **Z με p_0 στον παρονομαστή** (LM-style): $Z = (\bar{X} - p_0) / \sqrt{p_0(1 - p_0)/n}$. Συνηθίζεται σε ελέγχους υποθέσεων.
- **Z με $\hat{p}$ στον παρονομαστή** (Wald-style): $Z = (\bar{X} - p_0) / \sqrt{\hat{p}(1 - \hat{p})/n}$. Συνηθίζεται σε διαστήματα εμπιστοσύνης.

Παρατηρήστε ότι στην παρούσα άσκηση:

$$Z_{\text{Wald}}^2 = \frac{(\bar{X} - p_0)^2}{\hat{p}(1 - \hat{p})/n} = \frac{0,07^2}{0,1971/100} = \frac{0,0049}{0,001971} = 2,486 = W$$

Άρα τα δύο « Z^2 » αντιστοιχούν στα δύο πιο φυσικά μέλη της τριάδας. Το LR είναι κάπου ανάμεσα.

Σύνδεση με την Οικονομετρία — Παράδειγμα στο logit/probit

Στην εκτίμηση binary choice μοντέλων (logit, probit) στο πλαίσιο της οικονομετρίας:

- Ο έλεγχος σημαντικότητας ενός συντελεστή β_j από Stata/EViews/R εμφανίζεται ως «z statistic» — αυτό είναι το $\sqrt{W}$, δηλαδή ένας Wald test.
- Ο έλεγχος «overall significance» (αν όλοι οι συντελεστές μαζί είναι 0) γίνεται με LR test.
- Οι έλεγχοι ετεροσκεδαστικότητας / specification (RESET, White) είναι συχνά LM tests.

Όλα αυτά είναι ακριβώς η ίδια θεωρία που μόλις εφαρμόσατε σε μια απλή Bernoulli. Η μετάβαση από εδώ στην Οικονομετρία είναι, μαθηματικά, σχεδόν αόρατη — προστίθεται μόνο μια λίστα από regressors.

5 Σύνοψη Τύπων

Σύνοψη ενότητας

Δομή ελέγχου: H_0 vs H_1 · σφάλματα I (α) και II (β)· ισχύς = $1 - \beta$. Απορρίπτω αν $p\text{-value} \leq \alpha$.

Κλασικοί έλεγχοι μέσης τιμής:

- Z-test (γνωστή σ): $Z = (\bar{X} - \mu_0) / (\sigma / \sqrt{n}) \sim N(0, 1)$.
- t-test (άγνωστη σ): $T = (\bar{X} - \mu_0) / (S / \sqrt{n}) \sim t_{n-1}$.
- χ^2 -test για διασπορά: $W = (n - 1)S^2 / \sigma_0^2 \sim \chi_{n-1}^2$.
- Welch t-test (δύο δείγματα, ανεξάρτητα): $T = (\bar{X} - \bar{Y}) / \sqrt{S_X^2/n_X + S_Y^2/n_Y}$.
- Paired t-test: $T = \bar{D} / (S_D / \sqrt{n})$, με $D_i = X_i - Y_i$.

Δυσκρότητα ελέγχου-διαστήματος: ΔΕ $(1 - \alpha)$ = συλλογή των τιμών θ_0 που δεν απορρίπτονται από αμφίπλευρο έλεγχο επιπέδου α .

Neyman-Pearson: MP έλεγχος βασίζεται στον λόγο πιθανοφάνειας $\Lambda = L(\theta_1)/L(\theta_0)$.

Η τριάδα LR / Wald / LM ($H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$):

$$\begin{aligned} LR &= 2[\ell(\hat{\theta}_{ML}) - \ell(\theta_0)]. \\ W &= (\hat{\theta}_{ML} - \theta_0)^2 \cdot I_n(\hat{\theta}_{ML}). \\ LM &= \frac{[S(\theta_0)]^2}{I_n(\theta_0)}. \end{aligned}$$

Όλες $\xrightarrow{d} \chi_1^2$ υπό H_0 · απορρίπτω αν στατιστική $> \chi_{1,\alpha}^2 = 3,841$ σε $\alpha = 0,05$.

Ιδιότητες:

- Ασυμπτωτικά ισοδύναμα — ίδια απόφαση για μεγάλα n .
- Σε πεπερασμένα δείγματα διαφέρουν αριθμητικά.
- Συνήθης (αλλά όχι θεώρημα) ανισότητα: $W \geq LR \geq LM$.
- Στην κανονική με γνωστή σ^2 όλα ισούνται με $Z^2 = n(\bar{X} - \mu_0)^2/\sigma^2$.

Υπολογιστική επιλογή:

- LM: μόνο θ_0 , δεν χρειάζεται MLE — ιδανικό για ελέγχους ετεροσκεδαστικότητας.
- Wald: μόνο MLE, δεν χρειάζεται $\ell(\theta_0)$ — βάση του « t -statistic» σε όλα τα λογισμικά.
- LR: χρειάζεται και τα δύο — πιο ακριβές γενικά αλλά πιο ακριβό υπολογιστικά.

Σύνδεση με Οικονομετρία:

- « t -statistic» σε OLS = Wald test.
- Έλεγχοι μοντέλου (Breusch-Pagan, Breusch-Godfrey, White) = LM tests.
- Σύγκριση restricted vs unrestricted μοντέλου = LR test.