“Πιθανότητες και Αρχές Στατιστικής” (11η Διάλεξη)

Σωτήρης Νικολετσέας, καθηγητής

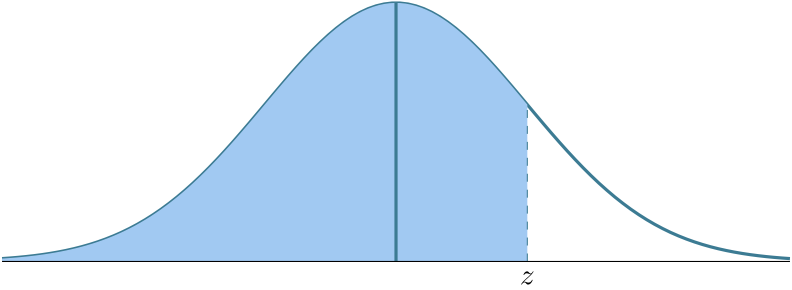
Τμήμα Μηχανικών Η/Υ & Πληροφορικής,
Πανεπιστήμιο Πατρών

Ακαδημαϊκό Έτος 2023 - 2024

Έλεγχος Στατιστικών Υποθέσεων (στατιστικά τεστ) - Σύνοψη

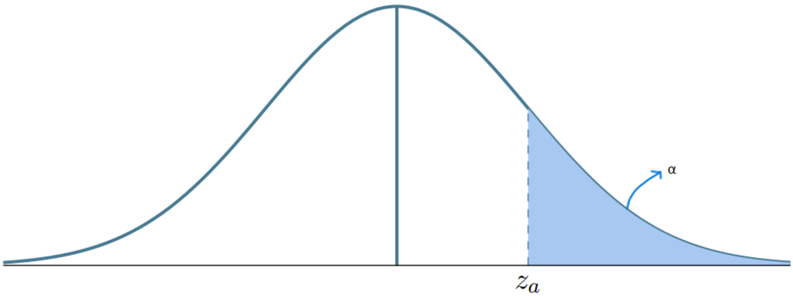
- Η έννοια της στατιστικής υπόθεσης (statistical hypothesis) και πώς ένα δείγμα δεδομένων χρησιμοποιείται για τον έλεγχό της.
- Η μηδενική υπόθεση (null hypothesis H_0) και η εναλλακτική υπόθεση (alternative hypothesis H_1).
- Η p-value
- Τρία ενδεικτικά tests:
  - το Z-test για τη μέση τιμή ενός κανονικού πληθυσμού με γνωστή διασπορά.
  - το t-test για τη μέση τιμή ενός κανονικού πληθυσμού με άγνωστη διασπορά.
  - το test για τις αναλογίες ενός πληθυσμού.
- Συνοπτική αναφορά σε άλλα tests.

Υπενθύμιση βασικών εννοιών (1)



$$\Phi(z) = Pr\{Z \leq z\}$$

Υπενθύμιση βασικών εννοιών (2)



z_a : η τιμή για την οποία

$$Pr\{Z \geq z_\alpha\} = \alpha$$

Ένα ενδεικτικό παράδειγμα

- Μια καπνοβιομηχανία ισχυρίζεται ότι βρήκε έναν νέο τρόπο παραγωγής τσιγάρων με μέση ποσότητα νικοτίνης ανά τσιγάρο $\leq 1.5\ mg$.
- Όμως, ένας ερευνητής του χώρου παραμένει δύσπιστος και πιστεύει ότι η μέση τιμή (μ) νικοτίνης υπερβαίνει τα $1.5\ mg$. Αποφασίζει να ελέγξει, μέσω τυχαίου δείγματος, την υπόθεση $\mu \leq 1.5\ mg$, την οποία και αποκαλούμε μηδενική (null) υπόθεση H_0:

$$H_0 : \ \mu \leq 1.5$$

- Η εναλλακτική (alternative) στη μηδενική υπόθεση συμβολίζεται ως H_1:

$$H_1 : \ \mu > 1.5$$

- Ένα στατιστικό test ελέγχει αν μία στατιστική υπόθεση είναι συμβατή (συνάδει) με το δείγμα, οπότε γίνεται δεκτή ή αν δεν συμβαδίζει με το δείγμα, οπότε απορρίπτεται.

Βασικοί Ορισμοί

- Για την πραγματοποίηση του ελέγχου, επιλέγεται τυχαίο δείγμα τσιγάρων που παρήχθησαν με τη νέα μέθοδο και μετριέται η νικοτίνη σε αυτά.
- Η συμφωνία ή όχι της υπόθεσης με το δείγμα, διαπιστώνεται με την τιμή που λαμβάνει στο δείγμα ένα στατιστικό τεστ.
- Ορισμός: Στατιστική υπόθεση είναι μία πρόταση για έναν πληθυσμό, που συνήθως μεταφράζεται σε μια σχέση ως προς μία παράμετρο του πληθυσμού (π.χ. τη μέση τιμή).
- Ορισμός: Στατιστικό test TS είναι μία στατιστική συνάρτηση της οποίας η τιμή προσδιορίζεται από ένα δείγμα ενός πληθυσμού. Ανάλογα με την τιμή του test, η μηδενική υπόθεση απορρίπτεται ή όχι.

Η έννοια της κρίσιμης περιοχής

- Στο παράδειγμά μας, το στατιστικό τεστ μπορεί να είναι η μέση τιμή της νικοτίνης στο τυχαίο δείγμα τσιγάρων. Διαισθητικά, αν αυτή η δειγματική μέση τιμή είναι "σημαντικά" μεγαλύτερη από 1.5 τότε το test απορρίπτει την H_0.
- Γενικά, ένα test TS πρέπει να συνοδεύεται από ένα σύνολο τιμών του TS για τις οποίες απορρίπτεται η H_0.
- Επομένως, το τεστ ουσιαστικά αποτελείται από την στατιστική συνάρτηση και την κρίσιμη περιοχή C. Δηλαδή ένα test έχει την γενική μορφή:

$$\text{Reject } H_0 \text{ if } TS \text{ is in } \mathbf{C}$$

$$\text{Do not reject } H_0 \text{ if } TS \text{ is not in } \mathbf{C}$$

Παράδειγμα test και κρίσιμης περιοχής

- Για παράδειγμα, αν γνωρίζαμε ότι η τυπική απόκλιση της περιεκτικότητας σε νικοτίνη είναι 0.8 mg, τότε ως test TS μπορούμε να πάρουμε τη δειγματική μέση τιμή $\bar{X}$ και ως κρίσιμη περιοχή:

$$C = \{\bar{X} \geq 1.5 + \frac{1.312}{\sqrt{n}}\}$$

  όπου n το μέγεθος του δείγματος.

- Οπότε για $n = 36$, τότε η H_0 θα απορρίπτεται αν $\bar{X} \geq 1.719$ και δεν θα απορρίπτεται αν $\bar{X} < 1.719$. Με άλλα λόγια όταν η $\bar{X}$ είναι μεγαλύτερη από την υπό έλεγχο τιμή 1.5 δεν απορρίπτεται αυτόματα η H_0, διότι θεωρούμε ότι σε ένα σχετικά μικρό δείγμα 36 τσιγάρων η πιθανότητα απόκλισης μέχρι περίπου 1.7 mg είναι αποδεκτή. Όσο μεγαλύτερο το δείγμα, τόσο πιο “αυστηρό” το test (μεγαλύτερη κρίσιμη περιοχή C).

Οι δύο τύποι σφαλμάτων

- Η απόρριψη της H_0 είναι μια ισχυρή απόφαση ότι η H_0 δεν συμβαδίζει με τα δεδομένα στο δείγμα. Αντίθετα, η απόφαση ότι η H_0 δεν απορρίπτεται (επειδή συνάδει με τα δεδομένα) είναι μια ασθενέστερη απόφαση.
- Για τον λόγο αυτό διακρίνουμε δύο τύπους σφάλματος του test:

	Αποδοχή της H_0	Απόρριψη της H_0
Αληθής η H_0	Ορθή απόφαση (πιθανότητα $1-\alpha$)	Σφάλμα τύπου I (πιθανότητα α)
Αληθής η H_1	Σφάλμα τύπου II (πιθανότητα β)	Ορθή απόφαση (πιθανότητα $1-\beta$)

  Επομένως το σφάλμα τύπου I (απόρριψη αληθούς H_0) είναι σοβαρότερο από το σφάλμα τύπου II (αποδοχή εσφαλμένης H_0).
- Η τιμή της πιθανότητας α καλείται <u>επίπεδο σημαντικότητας (significance level)</u> του test.
- Η τιμή της πιθανότητας $1-\beta$ καλείται δύναμη του test.

Παρατηρήσεις

- Σκοπός των test δεν είναι ο έλεγχος του αν ισχύει ή όχι η H_0, αλλά ο (ασθενέστερος) στόχος διαπίστωσης του αν η αλήθεια της H_0 συμβαδίζει με τα παρατηρούμενα δεδομένα.
- Επομένως, σκόπιμο είναι η H_0 να απορρίπτεται μόνο αν τα δεδομένα είναι πολύ απίθανα όταν η H_0 αληθεύει.
- Αυτό επιτυγχάνεται επιλέγοντας ένα αρκετά χαμηλό επίπεδο σημαντικότητας α ($\alpha = 0.10$, 0.05 ή 0.01), απαιτώντας ουσιαστικά η πιθανότητα απόρριψης της H_0, όταν η H_0 αληθεύει, να είναι το πολύ α, δηλαδή:

$$\alpha = Pr\{\text{απόρριψη της } H_0 \mid H_0 \text{ αληθής}\}$$

  Διαισθητικά, όσο πιο μικρό είναι το α, τόσο πιο "χαλαρό" είναι το test (δηλαδή, "είμαστε διατεθειμένοι" να αποδεχθούμε αποκλίσεις στο δείγμα) και απορρίπτουμε την H_0 μόνο όταν είμαστε αρκετά σίγουροι!

Παρατηρήσεις (II)

- Με άλλα λόγια, αν η H_0 δεν "περνάει" τον έλεγχο ενός "χαλαρού" test, τότε είμαστε αρκετά σίγουροι ότι καλώς απορρίπτεται.
- Ένα μικρό α ουσιαστικά σημαίνει μια ευρύτερη αποδοχή αποκλίσεων, δηλαδή μια μικρότερη περιοχή απόρριψης C.
- Επομένως, είμαστε περισσότερο σίγουροι για την απόφασή μας όταν απορρίπτουμε την H_0 λόγω ασυμφωνίας της με το δείγμα, οπότε ισχύει η εναλλακτική της H_1.
- Για τον λόγο αυτό συνήθως ορίζουμε την υπόθεση που θέλουμε να επιβεβαιώσουμε ως H_1. Αντίθετα, όταν θέλουμε να δείξουμε ότι μία υπόθεση δεν ισχύει, την ορίζουμε ως μηδενική υπόθεση H_0.

Παρατηρήσεις (III)

- Στο παράδειγμά μας, αν ο έλεγχος γίνονταν από την ίδια την καπνοβιομηχανία (και όχι τον ερευνητή) το test θα ήταν:

  $H_0 : \mu \geq 1.5$ έναντι της

  $H_1 : \mu < 1.5$

  οπότε η απόρριψη της H_0 θα ήταν μια ισχυρή ένδειξη για τον ισχυρισμό $\mu < 1.5\ mg$.

Παρατηρήσεις (IV)

- Παράδειγμα: Σε μία δίκη, όπου στο νομικό μας σύστημα ισχύει το τεκμήριο της αθωότητας, πρέπει να είμαστε σίγουροι ότι κάποιος είναι ένοχος. Επομένως, το test θα είναι:

$$H_0 : \text{Ο κατηγορούμενος είναι αθώος} \quad \text{έναντι της}$$
$$H_1 : \text{Ο κατηγορούμενος είναι ένοχος}$$

και καλό είναι να επιλέξουμε ένα πολύ μικρό επίπεδο σημαντικότητας α.

Η γενική φιλοσοφία ενός test

- Έστω ότι ελέγχουμε μία υπόθεση σχετικά με μια πληθυσμιακή παράμετρο θ, ως προς τις τιμές της σχετικά με μια περιοχή R. Η γενική μορφή του test θα είναι:

$$H_0 : \text{Η } \theta \text{ είναι στην } R \quad \text{έναντι της}$$
$$H_1 : \text{Η } \theta \text{ δεν είναι στην } R$$

- Προκειμένου να κατασκευάσουμε test σημαντικότητας α, πρέπει αρχικά να βρούμε μια σημειακή εκτιμήτρια της παραμέτρου θ, ώστε το test να απορρίπτει την H_0 όταν η θ είναι "αρκετά μακριά" από την R.

- Για να προσδιορίζουμε "πόσο μακριά", πρέπει να βρούμε την κατανομή της θ όταν η H_0 αληθεύει. Με τον τρόπο αυτό, μπορούμε να ορίσουμε μια κατάλληλη κρίσιμη περιοχή ώστε η πιθανότητα η εκτιμήτρια θ να μην είναι στην περιοχή αυτή (όταν η H_0 αληθεύει), να είναι το πολύ α.

Τύποι ελέγχων

Διακρίνουμε 3 είδη ελέγχων:

- $H_0 : \theta = \theta_0$, με εναλλακτική υπόθεση $H_1 : \theta \neq \theta_0$. Ο έλεγχος καλείται δίπλευρος και υπάρχουν 2 περιοχές απόρριψης με την περιοχή αποδοχής ανάμεσά τους.
- $H_0 : \theta \leq \theta_0$, εναντίον της $H_1 : \theta > \theta_0$. Ο έλεγχος καλείται μονόπλευρος από δεξιά.
- $H_0 : \theta \geq \theta_0$, εναντίον της $H_1 : \theta < \theta_0$, οπότε μιλάμε για μονόπλευρο από αριστερά έλεγχο.

Α. Το Z-test για τη μέση τιμή κανονικού πληθυσμού με γνωστή διασπορά

- Έστω δείγμα X_1,X_2,...,X_n από κανονική κατανομή με άγνωστη μέση τιμή μ και γνωστή διασπορά σ^2. Θέλουμε να ελέγξουμε τη μηδενική υπόθεση ότι η μέση τιμή έχει μια ορισμένη τιμή μ_0, έναντι της εναλλακτικής υπόθεσης ότι η μέση τιμή δεν έχει αυτή την ορισμένη τιμή:

  $$H_0 : \mu = \mu_0$$

  έναντι της

  $$H_1 : \mu \neq \mu_0$$

  για συγκεκριμένη τιμή μ_0.

- Μια φυσική επιλογή εκτιμήτριας είναι η δειγματική μέση τιμή:

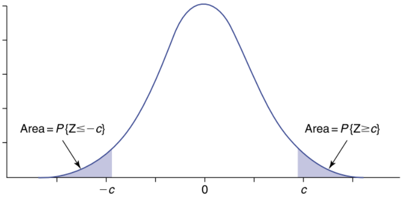
  $$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

  και είναι λογικό να απορρίπτουμε την H_0 (ότι $\mu = \mu_0$) όταν $\bar{X}$ είναι μακριά της μ_0, οπότε η κρίσιμη περιοχή C θα έχει τη μορφή:

  $$C = \{X_1, X_2, ..., X_n : |\bar{X} - \mu_0| \geq c\}$$

  για κατάλληλη τιμή της σταθεράς c.

A. Το Z-test για τη μέση τιμή κανονικού πληθυσμού με γνωστή διασπορά (συνέχεια)



- Για test με επίπεδο σημαντικότητας α, η σταθερά c πρέπει να είναι τέτοια ώστε, όταν $\mu = \mu_0$, η πιθανότητα το $\bar{X}$ να αποκλίνει από την μ κατά c ή περισσότερο να είναι ίση με α:
  $Pr\{|\bar{X} - \mu_0| \geq c\} = \alpha \quad$, όταν $\mu = \mu_0 \qquad (1)$
- Γνωρίζουμε όμως πως, όταν $\mu = \mu_0$, τότε $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$, επομένως η τυπική κανονική κατανομή Z, που ορίζεται ως,
  $$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0)$$
  έχει κατανομή $N(0, 1)$.

Α. Το Z-test για τη μέση τιμή κανονικού πληθυσμού με γνωστή διασπορά (συνέχεια)

- Επομένως, η σχέση (1) γίνεται:

$$\begin{array}{c} Pr\{|Z| \geq \sqrt{n}\frac{c}{\sigma}\} = \alpha \quad \text{ή} \\ 2Pr\{Z \geq \sqrt{n}\frac{c}{\sigma}\} = \alpha \quad \text{ή} \\ Pr\{Z \geq \sqrt{n}\frac{c}{\sigma}\} = \frac{\alpha}{2} \end{array}$$

- Αλλά το $z_{\frac{a}{2}}$ ορίζεται ως εξής:

$$Pr\{Z \geq z_{\frac{a}{2}}\} = \frac{a}{2}$$

οπότε,

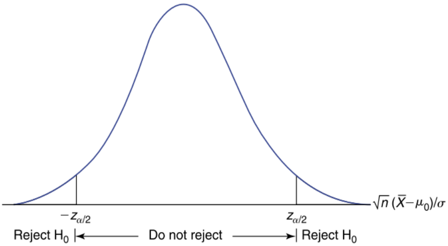
$$\sqrt{n}\frac{c}{\sigma} = z_{\frac{a}{2}} \text{ και } c = z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}}$$

- Τελικά έχουμε ότι το test έχει επίπεδο σημαντικότητας α αν απορρίπτουμε την H_0 όταν:

$$|\bar{X} - \mu_0| \geq z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}}$$

Α. Το Z-test για τη μέση τιμή κανονικού πληθυσμού με γνωστή διασπορά (συνέχεια)

- Τελικά δηλαδή το Z-test σημαντικότητας α είναι:
  Απόρριψη της H_0 αν $\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| \geq z_{\frac{\alpha}{2}}$
  Μη απόρριψη της H_0 αλλιώς



Ένα παράδειγμα Z-test

Ένα σήμα έντασης μ εκπέμπεται από πομπό και λαμβάνεται στον δέκτη ως σήμα με τιμή $N(\mu, 16)$, δηλαδή η λαμβανόμενη τιμή είναι κανονικά κατανεμημένη, με μέση τιμή μ και διασπορά 16. Με άλλα λόγια, το σήμα υπόκειται σε τυχαίο θόρυβο κανονικής κατανομής $N(0, 16)$. Δοθέντος ότι 20 ανεξάρτητες λήψεις του σήματος έχουν μέση τιμή 11.6 θέλουμε να ελέγξουμε την υπόθεση, με επίπεδο σημαντικότητας $\alpha = 0.05$, το εκπεμπόμενο σήμα να έχει ένταση $\mu = 10$.

Λύση: Το Z-test θα είναι:
$H_0 : \mu = 10$
εναντίον της εναλλακτικής:
$H_1 : \mu \neq 10$
Η τιμή του test θα είναι:
$TS = \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{20}}{4}|11.6 - 10| = 1.79 < 1.96 = z_{0.025}$
Επομένως η H_0 δεν απορρίπτεται θεωρώντας ότι η παρατηρούμενη απόκλιση από την υποτιθέμενη τιμή θα συμβαίνει, όταν η H_0 αληθεύει, τουλάχιστον στο 5% των περιπτώσεων.

Το "σωστό" επίπεδο σημαντικότητας

- Στο προηγούμενο παράδειγμα, αν επιλέγαμε μεγαλύτερο επίπεδο σημαντικότητας π.χ. $\alpha = 0.1$ (αντί για 0.05), τότε επειδή $z_{\frac{a}{2}} = z_{0.05} = 1.645$ θα έπρεπε να απορρίψουμε την H_0.
- Το "σωστό" επίπεδο σημαντικότητας α εξαρτάται από την συγκεκριμένη περίπτωση.
  - Αν η απόρριψη της H_0 (ενώ είναι αληθής) έχει σημαντικό κόστος και συνέπειες, θα πρέπει να είμαστε "συντηρητικοί", να αποδεχόμαστε αποκλίσεις και άρα να επιλέξουμε ένα "πιο χαλαρό" test με μικρό α. Για παράδειγμα, έστω H_1 η υπόθεση ότι μία νέα μέθοδος παραγωγής είναι καλύτερη από την τρέχουσα, ενώ H_0 η υπόθεση ότι δεν είναι καλύτερη. Επειδή μία αλλαγή μεθόδων είναι μία πολύ σημαντική ενέργεια, θα πρέπει η πιθανότητα να απορρίψουμε την H_0 (ενώ είναι αληθής) να είναι πολύ μικρή, οπότε επιλέγουμε πολύ μικρό a.
  - Παρομοίως, μικρό α επιλέγουμε όταν έχουμε μία καταρχήν ισχυρή πεποίθηση ότι η H_0 αληθεύει, οπότε θέλουμε πολύ ισχυρή ένδειξη ότι δεν ισχύει αν την απορρίψουμε.

Η τιμή p (p-value) ενός test

- Καλούμε p-value ενός test το μικρότερο επίπεδο σημαντικότητας α για το οποίο το test απορρίπτει την υπόθεση H_0.
- Επομένως, μια μικρή p-value σημαίνει ισχυρή ένδειξη ότι η H_0 δεν αληθεύει, αφού ακόμα και ένα "χαλαρό" test που "ανέχεται" μεγάλες αποκλίσεις, την απορρίπτει.
- Στην πράξη, πρώτα υπολογίζουμε την p-value και στη συνέχεια, ανάλογα με την περίπτωση, επιλέγουμε κατάλληλο επίπεδο σημαντικότητας α.
- Μια πολύ μεγάλη p-value υπαινίσσεται ότι η H_0 δεν πρέπει να απορριφθεί ενώ μια πολύ μικρή p-value είναι ένδειξη ότι η H_0 πρέπει να απορριφθεί.

Η τιμή p (p-value)

Είναι $\text{TS} = \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0|$

απόρριψη $H_0 \Leftrightarrow |TS| \geq z_{\alpha/2}$

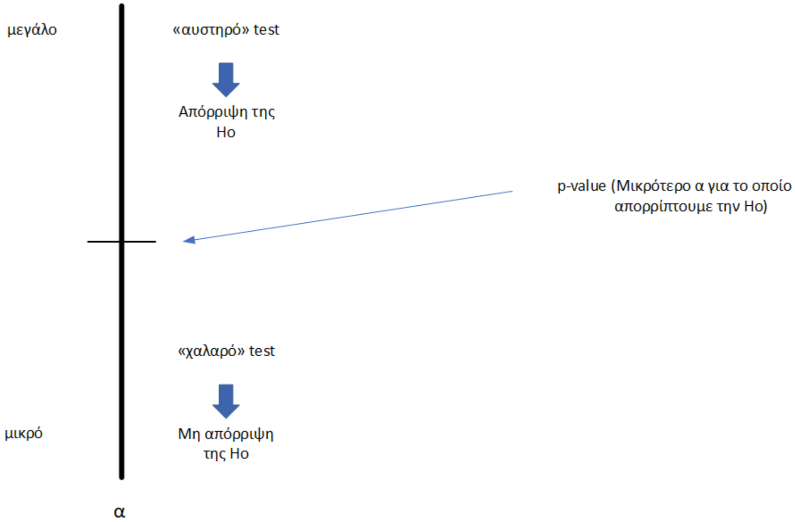
και p-value είναι η μικρότερη τιμή α για την οποία απορρίπτουμε την H_0. Ας βρούμε αυτήν την μικρότερη τιμή.

Άρα $z_{\alpha/2} \leq |TS| \Leftrightarrow$
$\text{Pr}\{\text{Z} \geq z_{\alpha/2}\} \geq Pr\{Z \geq |TS|\} \Leftrightarrow$
$\frac{\alpha}{2} \geq Pr\{Z \geq |TS|\} \Leftrightarrow$
$\alpha \geq 2Pr\{Z \geq |TS|\}$

Άρα η μικρότερη τιμή (δηλαδή η p-value) είναι :

$$\alpha = p\text{-value} = 2Pr\{Z \geq |TS|\}$$

Σχηματική επεξήγηση p-value



Παράδειγμα υπολογισμού της p-value

- Έστω στο προηγούμενο παράδειγμα εκπομπής σήματος, ότι ο μέσος όρος των 20 δειγμάτων είναι 10.8. Η τιμή του test είναι:

$$\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{20}}{4}|10.8 - 10| = 0.894$$

Από τους πίνακες βλέπουμε ότι:

$$2Pr\{Z \geq 0.894\} = 0.371$$

Άρα η p-value είναι 0.371. Επομένως, η H_0 δεν θα απορριφθεί για οποιοδήποτε επίπεδο σημαντικότητας $\alpha \leq 0.371$ (αντιθέτως, θα απορριφθεί για επίπεδα σημαντικότητας $\alpha > 0.371$). Άρα, υπάρχει ισχυρή ένδειξη αληθείας της H_0.

- Αν αντίθετα ο δειγματικός μέσος είναι 7.8 τότε:

$$TS = \frac{\sqrt{20}}{4}2.2 = 2.46 \text{ και}$$
$$p\text{-value} = 2Pr\{Z \geq 2.46\} = 0.014$$

Άρα η H_0 θα απορρίπτεται για όλα τα $\alpha > 0.014$, άρα υπάρχει ισχυρή ένδειξη απόρριψης της H_0.

p-value (Σύνοψη)

Επομένως, η p-value του Z-test υπολογίζεται ως εξής:

- Στην αρχή υπολογίζουμε την τιμή του στατιστικού test:

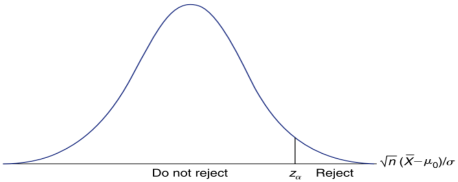
$$TS = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma} = \mathrm{v}$$

- Στη συνέχεια η p-value είναι:

$$p\text{-value} = 2Pr\{Z \geq |\mathrm{v}|\}$$

Μονόπλευρα tests

- Στα δίπλευρα tests ελέγχουμε την υπόθεση $\mu = \mu_0$ και την απορρίπτουμε ουσιαστικά όταν η παρατηρούμενη $\bar{X}$ είτε είναι "αρκετά" μεγαλύτερη είτε είναι "αρκετά" μικρότερη από την μ_0.
- Όμως, σε αρκετές περιπτώσεις απλά ενδιαφερόμαστε για το αν η μ είναι μεγαλύτερη ή ίση (ή μικρότερη ή ίση) από μία τιμή μ_0. Οπότε το μονόπλευρο test είναι:
  $H_0 : \mu \leq \mu_0$
  έναντι της
  $H_1 : \mu > \mu_0$



Σύνοψη (Z-test για τη μέση τιμή κανονικού πληθυσμού με γνωστή διασπορά)

Table 9.1 Hypothesis Tests Concerning the Mean μ of a Normal Population with Known Variance σ^2

$X_1, \ldots, X_n$ *are sample data, and* $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

H_0	H_1	Test statistic TS	Significance-level-α test	p value if TS $= v$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}$	Reject H_0 if $\lvert\text{TS}\rvert \geq z_{\alpha/2}$ Do not reject H_0 otherwise	$2P\{Z \geq \lvert v\rvert\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}$	Reject H_0 if $\text{TS} \geq z_\alpha$ Do not reject H_0 otherwise	$P\{Z \geq v\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}$	Reject H_0 if $\text{TS} \leq -z_\alpha$ Do not reject H_0 otherwise	$P\{Z \leq v\}$

B. Το t-test για τη μέση τιμή κανονικού πληθυσμού με άγνωστη διασπορά

- Πολύ πιο συνηθισμένη είναι η περίπτωση όπου η η διασπορά είναι άγνωστη, οπότε υπολογίζουμε την δειγματική διασπορά:
$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$
- Το test πλέον γίνεται:
$$TS = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$
(όπου S η δειγματική τυπική απόκλιση), και θα απορρίπτουμε την H_0 για "μεγάλα" TS.
- Γνωρίζουμε όμως (από την προηγούμενη διάλεξη, διαφάνεια 26) πως, όταν $\mu = \mu_0$, η TS ακολουθεί την κατανομή t με $n-1$ βαθμούς ελευθερίας.

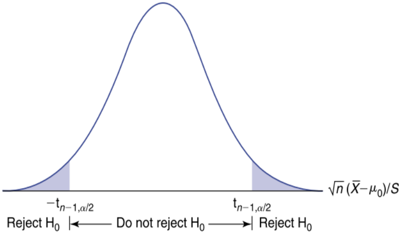
t-test (συνέχεια)

- Επομένως, το δίπλευρο test επιπέδου σημαντικότητας α θα είναι:

$H_0: \ \mu = \mu_0$ εναντίον $H_1: \ \mu \neq \mu_0$

Απόρριψη της H_0 αν $|TS| \geq t_{n-1,\frac{a}{2}}$

Μη απόρριψη της H_0 αλλιώς



t-test (σύνοψη)

Συνοπτικά, το t-test και η p-value φαίνονται στον ακόλουθο πίνακα:

Table 9.2 Hypothesis Tests Concerning the Mean μ of a Normal Population with Unknown Variance σ^2

$X_1, \ldots, X_n$ *are sample data;* $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$, $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$

H_0	H_1	Test statistic TS	Significance-level-α test	p value if TS $= \nu$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}$	Reject H_0 if $\lvert\text{TS}\rvert \geq t_{n-1,\alpha/2}$ Do not reject otherwise	$2P\{T_{n-1} \geq \lvert\nu\rvert\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}$	Reject H_0 if TS $\geq t_{n-1,\alpha}$ Do not reject otherwise	$P\{T_{n-1} \geq \nu\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}$	Reject H_0 if TS $\leq -t_{n-1,\alpha}$ Do not reject H_0 otherwise	$P\{T_{n-1} \leq \nu\}$

T_{n-1} is a t random variable with $n-1$ degrees of freedom, and $t_{n-1,\alpha}$ and $t_{n-1,\alpha/2}$ are such that $P\{T_{n-1} \geq t_{n-1,\alpha}\} = \alpha$ and $P\{T_{n-1} \geq t_{n-1,\alpha/2}\} = \alpha/2$.

Γ. Έλεγχος υποθέσεων για αναλογίες πληθυσμού (populations proportions)

- Ελέγχουμε το ποσοστό ενός πληθυσμού που έχει ένα συγκεκριμένο χαρακτηριστικό, π.χ. το ποσοστό προϊόντων μιας βιομηχανίας που είναι ελαττωματικά, το ποσοστό απόδοσης ενός νέου φαρμάκου, το ποσοστό των κατοίκων μιας πόλης που υποστηρίζει την κατασκευή μιας νέας γέφυρας κλπ.
- Θεωρούμε ότι ο πληθυσμός είναι πολύ μεγάλος και έστω p το άγνωστο ποσοστό του πληθυσμού με το ζητούμενο χαρακτηριστικό. Το test θα είναι:
  $H_0 : p \leq p_0$
  έναντι της εναλλακτικής υπόθεσης
  $H_1 : p > p_0$

Test αναλογίας πληθυσμού (συνέχεια)

- Αν θεωρήσουμε μια τυχαία επιλογή n στοιχείων του πληθυσμού, τότε ο αριθμός X των στοιχείων με το χαρακτηριστικό θα ακολουθεί διωνυμική κατανομή με παραμέτρους n και p, όπου p η άγνωστη "αναλογία" που θέλουμε να ελέγξουμε.
- Ουσιαστικά, θα απορρίπτουμε την H_0 $(p \leq p_0)$ αν το X είναι "αρκετά" μεγάλο. Με απλά λόγια, αν το παρατηρούμενο πλήθος των στοιχείων του πληθυσμού στο δείγμα που έχουν το χαρακτηριστικό είναι x, τότε το test statistic TS είναι το ίδιο το X που ακολουθεί διωνυμική κατανομή $X \sim B(n, p_0)$ και θα απορρίπτουμε την H_0 με p-value :

$$p\text{-value} = Pr\{X \geq x\}$$

Παράδειγμα

Μία βιομηχανία ισχυρίζεται ότι το πολύ 2% των προϊόντων της είναι ελαττωματικά. Ένας πιθανός αγοραστής δοκιμάζει ένα δείγμα 400 προϊόντων και βρίσκει 13 ελαττωματικά (3.25%). Μπορούμε, με επίπεδο σημαντικότητας 5%, να απορρίψουμε τον ισχυρισμό της βιομηχανίας;

Λύση:
Αν p η πιθανότητα ένα προϊόν να είναι ελαττωματικό, το test θα είναι:

$$H_0 : \ p \leq 0.02 \text{ εναντίον } H_1 : \ p > 0.02$$

Η p-value θα είναι ουσιαστικά η πιθανότητα εμφάνισης ενός τέτοιου αριθμού ελαττωματικών προϊόντων αν το p ήταν 0.02 (η μεγαλύτερη τιμή του αν η H_0 αληθεύει). Άρα:

$$p\text{-value} = Pr\{X \geq 13\} = 0.0619 > 0.05$$

και επομένως τα δεδομένα του δείγματος, αν και προδιαθέτουν αρνητικά, δεν επαρκούν για την απόρριψη της H_0 με επίπεδο σημαντικότητας 5%.

Παράδειγμα (συνέχεια)

Η μεταβλητή X είναι διωνυμική: $X \sim B(400, 0.02)$, και η ακριβής τιμή της $Pr\{X > 13\}$ είναι 0.0619 (η τιμή αυτή μπορεί να υπολογιστεί μέσω κατάλληλου προγράμματος). Μία προσεγγιστική τιμή μπορεί να υπολογιστεί μέσω της κανονικής:

$$\begin{aligned} & Pr\{X \geq 13\} = Pr\{X \geq 12.5\} \qquad \text{(διόρθωση συνέχειας)} \\ = \quad & Pr\{\tfrac{X-8}{\sqrt{400*0.02*0.98}} \geq \tfrac{125-8}{\sqrt{400*0.02*0.98}}\} \\ \simeq \quad & Pr\{Z \geq 1.607\} = 0.054 \end{aligned}$$

(Άρα η προσέγγιση, αν και δεν είναι πολύ ακριβής, οδηγεί στο ίδιο σωστό συμπέρασμα μη απόρριψης της H_0).

Σύνοψη (test αναλογιών πληθυσμού)

Table 9.3 Hypothesis Tests Concerning p, the Proportion of a Large Population that Has a Certain Characteristic

The number of population members in a sample of size n that have the characteristic is X, and B is a binomial random variable with parameters n and p_0.

H_0	H_1	Test statistic TS	p value if TS $= x$
$P \leq p_0$	$p > p_0$	X	$P\{B \geq x\}$
$P \geq p_0$	$p < p_0$	X	$P\{B \leq x\}$
$P = p_0$	$p \neq p_0$	X	$2\,\mathrm{Min}\{P\{B \leq x\}, P\{B \geq x\}\}$

Δ. Άλλα tests στατιστικών υποθέσεων

Δ1. Tests για δύο κανονικούς πληθυσμούς

Συχνά θέλουμε να ελέγξουμε υποθέσεις σχετικά με δύο πληθυσμούς (π.χ την επίδραση ενός φαρμάκου και ενός placebo του φαρμάκου σε δύο διαφορετικά σύνολα ασθενών, που το ένα σύνολο παίρνει το φάρμακο και το άλλο το placebo).

- έλεγχος ισότητας (και γενικά της σχέσης) των μέσων τιμών δύο ανεξάρτητων κανονικών δειγμάτων με γνωστές διασπορές (παρόμοιο με Z-test).

Table 10.2 Tests of Means of Two Normal Populations Having Known Variances when Samples are Independent

The sample mean of a sample of size n from a normal population having mean μ_x and known variance σ_x^2 is $\overline{X}$. The sample mean of a sample of size m from a second normal population having mean μ_y and known variance σ_y^2 is $\overline{Y}$. The two samples are independent.

H_0	H_1	Test statistic TS	Significance-level-α test	p value if $TS = v$
$\mu_x = \mu_y$	$\mu_x \neq \mu_y$	$\frac{\overline{X}-\overline{Y}}{\sqrt{\sigma_x^2/n+\sigma_y^2/m}}$	Reject H_0 if $\lvert TS\rvert \geq z_{\alpha/2}$ Do not reject otherwise	$2P\{Z \geq \lvert v\rvert\}$
$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$\frac{\overline{X}-\overline{Y}}{\sqrt{\sigma_x^2/n+\sigma_y^2/m}}$	Reject H_0 if $TS \geq z_\alpha$ Do not reject otherwise	$P\{Z \geq v\}$

Δ1. Tests για δύο κανονικούς πληθυσμούς

- αν οι διασπορές είναι άγνωστες, τότε υπάρχει αντίστοιχο test για την περίπτωση που τα δύο ανεξάρτητα δείγματα είναι "αρκετά" μεγάλα (συχνά αρκεί μέγεθος ≥ 20, παρόμοιο με Z-test).

Table 10.3 Tests of Means of Two Normal Populations Having Unknown Variances when Samples are Independent and Sample Sizes are Large

The sample mean and sample variance of a sample of size n from a normal population having mean μ_x and unknown variance σ_x^2 are, respectively, $\overline{X}$ and S_x^2. The sample mean and sample variance of a sample of size m from a second normal population having mean μ_y and unknown variance σ_y^2 are, respectively, $\overline{Y}$ and S_y^2. The two samples are independent, and both n and m are at least 20.

H_0	H_1	Test statistic TS	Significance-level-α test	p value if $TS = v$
$\mu_x = \mu_y$	$\mu_x \neq \mu_y$	$\frac{\overline{X}-\overline{Y}}{\sqrt{S_x^2/n+S_y^2/m}}$	Reject H_0 if $\lvert TS\rvert \geq z_{\alpha/2}$ Do not reject otherwise	$2P\{Z \geq \lvert v\rvert\}$
$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$\frac{\overline{X}-\overline{Y}}{\sqrt{S_x^2/n+S_y^2/m}}$	Reject H_0 if $TS \geq z_\alpha$ Do not reject otherwise	$P\{Z \geq v\}$

Δ1. Tests για δύο κανονικούς πληθυσμούς

- αν, αντιθέτως, τα δείγματα είναι μικρά αλλά οι (άγνωστες) διασπορές είναι ίσες, μπορούμε να ορίσουμε άλλο test (που βασίζεται στην κατανομή t).

Table 10.4 Tests of Means of Two Normal Populations Having Unknown Though Equal Variances when Samples are Independent

The sample mean and sample variance, respectively, of a sample size n from a normal population having mean μ_x and variance σ^2 are $\overline{X}$ and S_x^2. And the sample mean and sample variance of a sample of size m from a second normal population having mean μ_y and variance σ^2 are $\overline{Y}$ and S_y^2. The two samples are independent.

$$S_p^2 = \frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}$$

H_0	H_1	Test statistic TS	Significance-level-α test	p value if $TS = v$
$\mu_x = \mu_y$	$\mu_x \neq \mu_y$	$\frac{\overline{X}-\overline{Y}}{\sqrt{S_p^2(1/n+1/m)}}$	Reject H_0 if $\lvert TS\rvert \geq t_{n+m-2,\alpha/2}$ Do not reject otherwise	$2P\{T_{n+m-2} \geq \lvert v\rvert\}$
$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$\frac{\overline{X}-\overline{Y}}{\sqrt{S_p^2(1/n+1/m)}}$	Reject H_0 if $TS \geq t_{n+m-2,\alpha}$ Do not reject otherwise	$P\{T_{n+m-2} \geq v\}$

Δ2. Test αναλογιών δύο πληθυσμών

- Πρόκειται για τον έλεγχο της σχέσης των αναλογιών p_1, p_2 κάποιου χαρακτηριστικού σε δύο πληθυσμούς.
- Παράδειγμα: Μια βιομηχανία πιστεύει ότι μία νέα μέθοδος παραγωγής μειώνει τον αριθμό των ελαττωματικών προϊόντων. Για να ελέγξει αυτήν την υπόθεση, παράγει δύο διαφορετικά δείγματα, ένα με την παλιά μέθοδο και ένα με τη νέα, και μετράει τα ελαττωματικά προϊόντα σε κάθε δείγμα.
- Τα tests αυτά είναι παρόμοια με το Z-test και το t-test.

Δ3. Chi-Square Goodness-of-fit Test

Έστω μεγάλος πληθυσμός κάθε μέλος του οποίου έχει μία τιμή που είναι 1 ή 2 ή 3 ή ... ή k. Έστω πιθανότητες $p_1, p_2, ..., p_k$.

Παράδειγμα: Είναι γνωστό ότι η κατανομή στον πληθυσμό της ομάδας αίματος A είναι 41%, της ομάδας B είναι 9%, της ομάδας AB είναι 4% και της ομάδας 0 είναι 46%. Θέλουμε να ελέγξουμε αν η κατανομή αυτή είναι διαφορετική σε άτομα που έχουν μία σοβαρή ασθένεια.

- Στο test αυτό, παίρνουμε ένα δείγμα 200 ατόμων που έχουν την ασθένεια, μετράμε την κατανομή κάθε ομάδας αίματος στο δείγμα αυτό και ελέγχουμε αν το δείγμα στοιχειοθετεί την ίδια ή διαφορετική κατανομή στον γενικό πληθυσμό.
- Το test αυτό ονομάζεται έτσι γιατί βασίζεται στην κατανομή χ^2_{k-1} ($k-1$ βαθμοί ελευθερίας).

Δ4. Chi-Square Independence Test

Έλεγχος της ανεξαρτησίας δύο χαρακτηριστικών.

Παράδειγμα: Έστω το χαρακτηριστικό X : marital status ενός πληθυσμού (δηλαδή αν ένα μέλος του είναι single, married, widowed/divorced) και το χαρακτηριστικό Y : το επίπεδο κατάθλιψης (ήπιο, κανονικό, μεγάλο) των μελών του πληθυσμού. Θέλουμε να ελέγξουμε την υπόθεση ότι τα χαρακτηριστικά X και Y είναι στοχαστικά ανεξάρτητα.

- Το αντίστοιχο test βασίζεται στην κατανομή $\chi^2_{(r-1)\cdot(s-1)}$ με $(r-1)\cdot(s-1)$ βαθμούς ελευθέριας, όπου r, s το πλήθος των διαφορετικών τιμών που παίρνουν τα χαρακτηριστικά X, Y αντίστοιχα.