Z-test

Elias Zintzaras, M.Sc., Ph.D.

Professor in Biomathematics-Biometry
Department of Biomathematics
School of Medicine
University of Thessaly

Institute for Clinical Research and Health Policy Studies
Tufts University School of Medicine
Boston, MA, USA

Theodoros Mprotsis, MSc, PhD
Teacher & Research Fellow
(http://biomath.med.uth.gr)
University of Thessaly
Email: tmprotsis@uth.gr

Αξιολόγηση ποσοστών άσθματος: Νοσοκομειακό δείγμα έναντι γενικού πληθυσμού

Έλεγχος για ένα ποσοστό

Προϋποθέσεις

- Τυχαίο δείγμα
  - Κάθε δείγμα πρέπει να λαμβάνεται τυχαία από τον αντίστοιχο πληθυσμό του
- Αρκετά μεγάλο δείγμα (Κανονικότητα)
  - $n \geq 30$
- Δυαδικό αποτέλεσμα (Επιτυχία/Αποτυχία)
  - Για παράδειγμα, στο πλαίσιο των ποσοστών άσθματος, κάθε άτομο στο δείγμα είτε έχει ιστορικό άσθματος (επιτυχία) είτε δεν έχει ιστορικό άσθματος (αποτυχία)

Αξιολόγηση ποσοστών άσθματος: Νοσοκομειακό δείγμα έναντι γενικού πληθυσμού

Παράδειγμα

Σε ένα νοσοκομείο συλλέχθηκε ένα τυχαίο δείγμα από $n_1 = 215$ γυναίκες από τους καταλόγους νοσηλευομένων και βρέθηκε ότι $r = 39$ από αυτές έχουν ιστορικό άσθματος (δηλ. το παρατηρούμενο ποσοστό άσθματος είναι $p = \frac{39}{215} = 18\%$). Είναι γνωστό ότι ο επιπολασμός (ποσοστό) της νόσου είναι $P = 15\%$.

Ερώτηση

Συμφωνεί το ποσοστό των γυναικών που βρέθηκε με άσθμα με αυτό του γενικού πληθυσμού;

Υπόθεση

$\boldsymbol{H_0}$: Το ποσοστό ατόμων με άσθμα στο δείγμα του νοσοκομείου είναι **ίσο** με το ποσοστό επιπολασμού του άσθματος στον γενικό πληθυσμό

$$H_0 : p = P$$

$\boldsymbol{H_\alpha}$: Το ποσοστό ατόμων με άσθμα στο δείγμα του νοσοκομείου είναι **διαφορετικό** με το ποσοστό επιπολασμού του άσθματος στον γενικό πληθυσμό

$$H_\alpha : p \neq P$$

Z-test τύπος

Για να ελέγξουμε αν το ποσοστό των περιστατικών δεν διαφέρει από τον επιπολασμό της νόσου χρησιμοποιούμε το z-test:

Ο τύπος για το z-test statistic είναι:

$$z = \frac{p - P}{SE}$$

$$SE = \sqrt{\frac{P(1-P)}{n}}$$

Βήμα 1: Υπολογισμός του ποσοστού στο δείγμα

$$p = \frac{r}{n_1} = \frac{39}{215} = 0.1814$$ που είναι 18.14%

$$P = 15\%$$

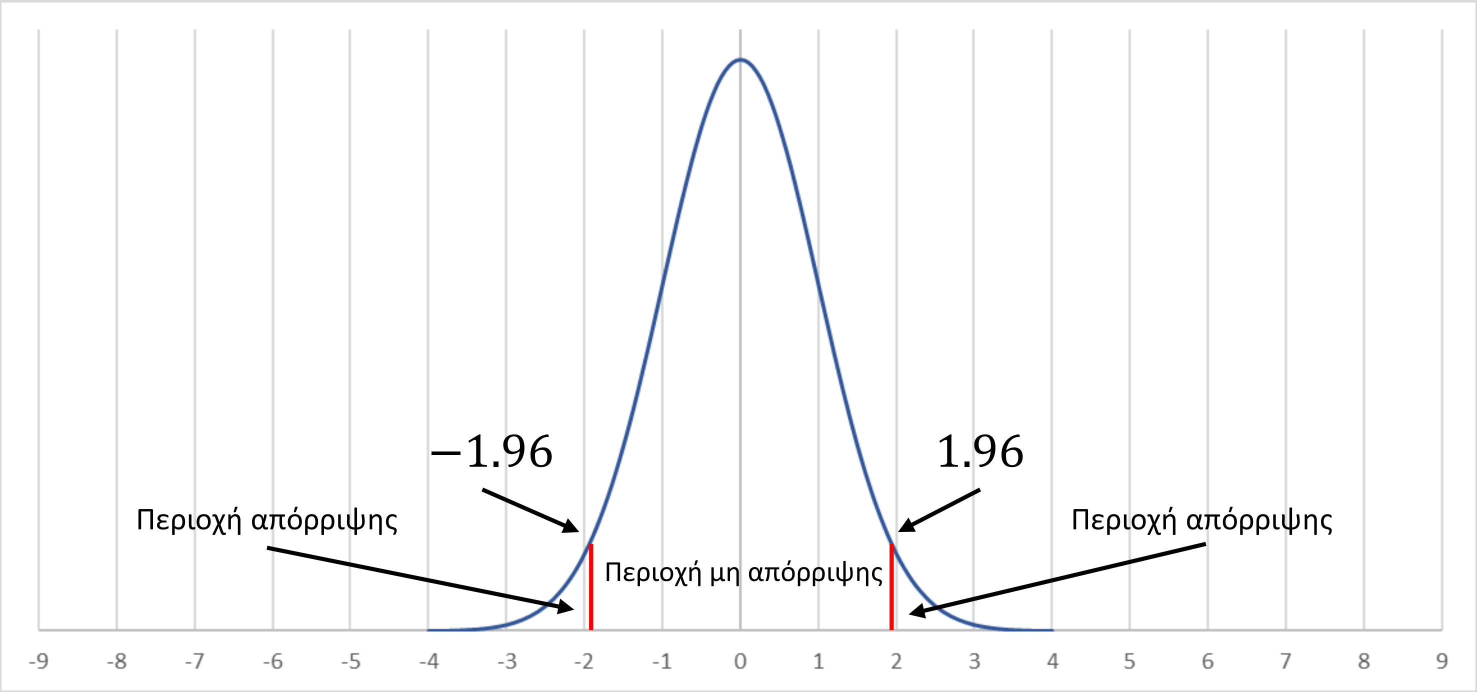
Βήμα 2: Εισαγωγή δεδομένων από Βήμα 1 στον τύπο του z-test

$$SE = \sqrt{\frac{0.15(1-0.15)}{215}} = 0.0244$$

$$z = \frac{0.1814 - 0.15}{0.0244} = 1.2892$$

Πρέπει να ελέγξουμε εάν η **z-τιμή** βρίσκεται εντός της **περιοχής απόρριψης**

Βήμα 3: z-τιμή που σχετίζεται με επίπεδο σημαντικότητας 5%

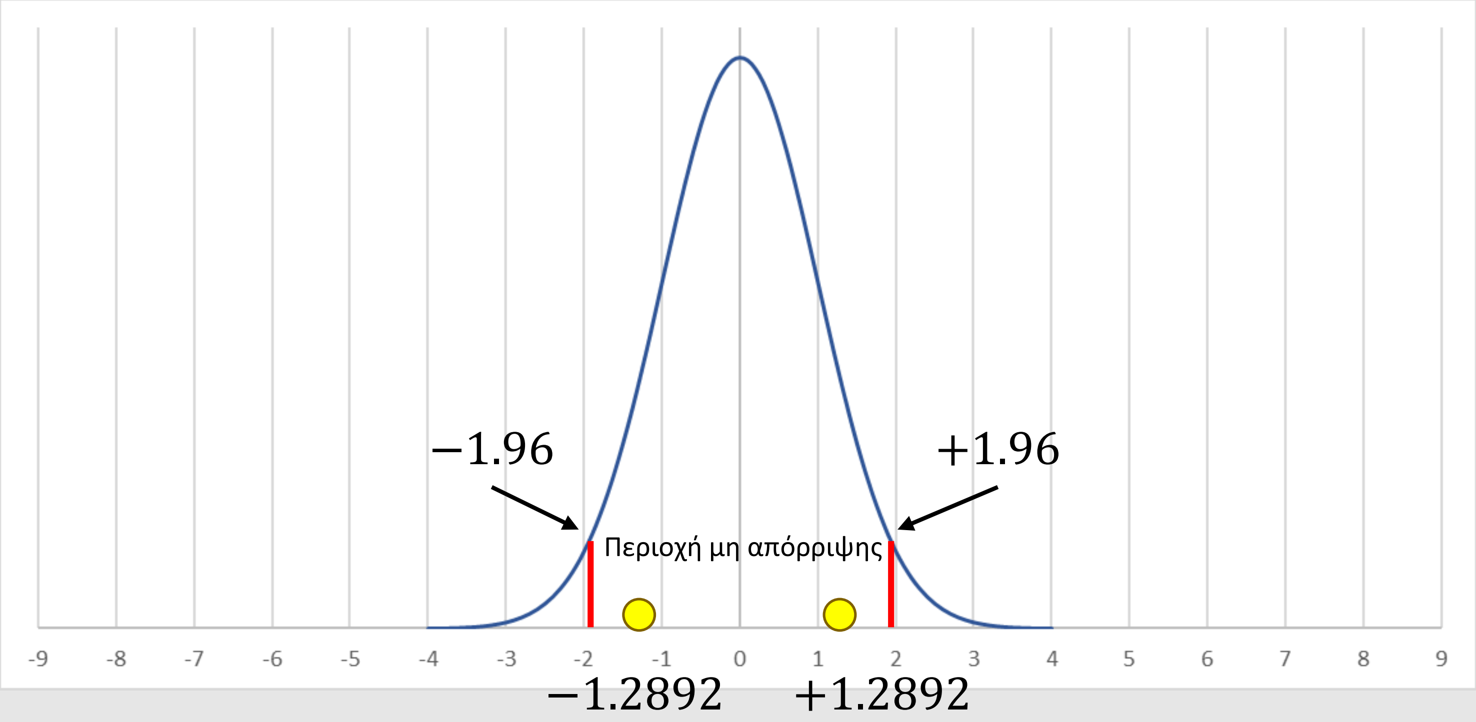


Βήμα 4: Στατιστική σημαντικότητα

Η τιμή $z = 1.2892$ είναι μικρότερη από το 5% κρίσιμο σημείο της τυπικής κανονικής κατανομής, που είναι 1.96

Το αποτέλεσμα ($P \geq \ 0.05$) του z-test υποδεικνύει ότι το παρατηρούμενο ποσοστό άσθματος (18.14%) **δεν διαφέρει στατιστικά σημαντικά** από τον επιπολασμό της νόσου (15%)

Η απουσία στατιστικής σημασίας υποδηλώνει ότι ο αυξημένος επιπολασμός που παρατηρήθηκε στο δείγμα δεν παρέχει επαρκή στοιχεία για να επιβεβαιωθεί ότι το ποσοστό άσθματος μεταξύ των γυναικών σε αυτή τη μελέτη διαφέρει από εκείνο του γενικού πληθυσμού



95% διάστημα εμπιστοσύνης

Το 95% διάστημα εμπιστοσύνης (CI) δίνεται από την παρακάτω σχέση:

$$(p - 1.96 \cdot SE, p + 1.96 \cdot SE)$$

$$SE = \sqrt{\frac{P(1-P)}{n}}$$

95% διάστημα εμπιστοσύνης

$$(0.1814 - 1.96 \cdot 0.0244, 0.1814 + 1.96 \cdot 0.0244)$$

$$(0.1814 - 0.047824, 0.1814 + 0.047824)$$

$$(0.133576,0.229224)$$

$$(0.1336, 0.2292)$$

95% διάστημα εμπιστοσύνης

Το 95% διάστημα εμπιστοσύνης (CI) για το ποσοστό επιπολασμού του άσθματος στο δείγμα των γυναικών είναι

$$(0.1336, 0.2292)$$

Αν υποθέσουμε ότι το δείγμα των γυναικών από το νοσοκομείο είναι αντιπροσωπευτικό όλων των γυναικών στη χώρα τότε μπορούμε να ισχυριστούμε με βεβαιότητα 95% ότι ο επιπολασμός του άσθματος στη χώρα είναι μεταξύ 13.36% και 22.92%

Καθώς το διάστημα εμπιστοσύνης $(13.36\%, 22.92\%)$ περιέχει 15%, συμπεραίνουμε ότι ο επιπολασμός του δείγματος **δεν διαφέρει στατιστικά σημαντικά** από τον επιπολασμό του πληθυσμού στο επίπεδο του 5%

Z-Test με διόρθωση

Για σχετικά μικρά δείγματα το παρακάτω z-test δίνει πιο ακριβή αποτελέσματα :

$$z_c = \frac{|p - P| - \frac{1}{2n}}{SE}$$

$$z_c = \frac{|0.1814 - 0.15| - \frac{1}{2 \cdot 215}}{0.0244}$$

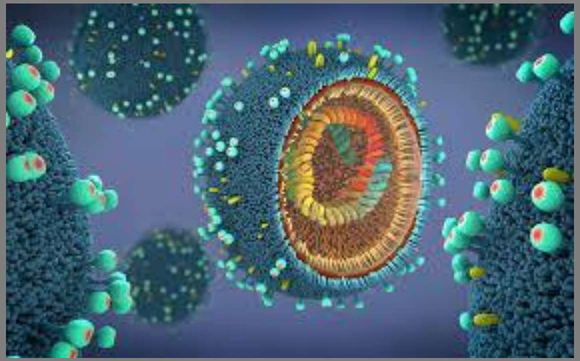
Στο παράδειγμά μας, $z_c = 1.1937$ είναι λίγο μικρότερο από το $z = 1.2892$ (που βρήκαμε χωρίς τη διόρθωση) επειδή το μέγεθος του δείγματος είναι μεγάλο

Practice

1. A study of diseases in native Americans (Kizer et al. 2006) found 381 obese or overweight patients in 449 patients. In the general population of the USA, the percentage obese or overweight is 65%.
   The researchers wanted to determine if the percentage of obesity/overweight native Americans was different than that of the general population
2. Kim et al. (2004) studied the measles-rubella vaccination-rates in Korea. They compared the proportion of children with measles antibodies to the World Health Organization (WHO) target proportion (for children aged 5 to 9 years old: 10%)
   The aim of the study was to test if the proportion of Korean children with the measles antibody in the population was 10% or lower (i.e., better). In the study, 55 children out of 972 had the antibody present

Σύγκριση ποσοστών γρίπης μεταξύ εμβολιασμένων με πραγματικό εμβόλιο και ομάδας placebo

Έλεγχος για τη σύγκριση δύο ποσοστών



Προϋποθέσεις

- Ανεξάρτητα δείγματα
  - Τα δύο δείγματα θα πρέπει να είναι ανεξάρτητα το ένα από το άλλο, πράγμα που σημαίνει ότι η επιλογή των ατόμων στο ένα δείγμα δεν επηρεάζει την επιλογή των ατόμων στο άλλο
- Τυχαίο δείγμα
  - Κάθε δείγμα πρέπει να λαμβάνεται τυχαία από τον αντίστοιχο πληθυσμό του
- Αρκετά μεγάλο δείγμα (Κανονικότητα)
  - $n \geq 30$
- Δυαδικό αποτέλεσμα (Επιτυχία/Αποτυχία)
  - Στο παράδειγμα που ακολουθεί, στο πλαίσιο των ποσοστών γρίπης, κάθε άτομο στο δείγμα είτε προσβλήθηκε από γρίπη (επιτυχία) είτε δεν προσβλήθηκε από γρίπη (αποτυχία)

Σύγκριση ποσοστών γρίπης μεταξύ εμβολιασμένων με πραγματικό εμβόλιο και ομάδας placebo

Παράδειγμα

Είκοσι ($p_1 = 20$) από τα 240 ($n_1 = 240$) άτομα που εμβολιάσθηκαν με πραγματικό εμβόλιο προσβλήθηκαν από γρίπη σε σύγκριση με τα 80 ($p_2 = 80$) από τα 220 ($n_2 = 220$) που εμβολιάσθηκαν με placebo.

Ερώτηση

Υπάρχει ένδειξη ότι το εμβόλιο ήταν αποτελεσματικό;

Υπόθεση

$\boldsymbol{H_0}$: Ο εμβολιασμός δεν επηρεάζει το ποσοστό της γρίπης˙ Τα ποσοστά των κρουσμάτων γρίπης στην ομάδα που εμβολιάστηκε (p_1) και στην ομάδα εικονικού φαρμάκου (p_2) **είναι ίδια**

$$H_0: p_1 = p_2$$

$\boldsymbol{H_\alpha}$: Ο εμβολιασμός επηρεάζει το ποσοστό της γρίπης˙ Τα ποσοστά των κρουσμάτων γρίπης στην ομάδα που εμβολιάστηκε (p_1) και στην ομάδα εικονικού φαρμάκου (p_2) **δεν είναι ίδια**

$$H_\alpha: p_1 \neq p_2$$

Z-test τύπος

Για να ελέγξουμε αν το ποσοστό προσβολής από γρίπη όταν γίνεται εμβολιασμός είναι διαφορετικό από το ποσοστό με το placebo, χρησιμοποιείται το z-test.

Ο τύπος για το z-test statistic είναι:

$$z = \frac{p_1 - p_2}{SE}$$

$$SE = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Βήμα 1: Υπολογισμός ποσοστών

$$p_1 = \frac{p_1}{n_1} = \frac{20}{240} = 0.083$$ που είναι 8.3%

$$p_2 = \frac{p_2}{n_2} = \frac{80}{220} = 0.364$$ που είναι 36.4%

Βήμα 2: Υπολογισμός συνολικού ποσοστού

$$p = \frac{p_1 + p_2}{n_1 + n_2}$$

$$p = \frac{20 + 80}{240 + 220}$$

$$p = 0.217$$

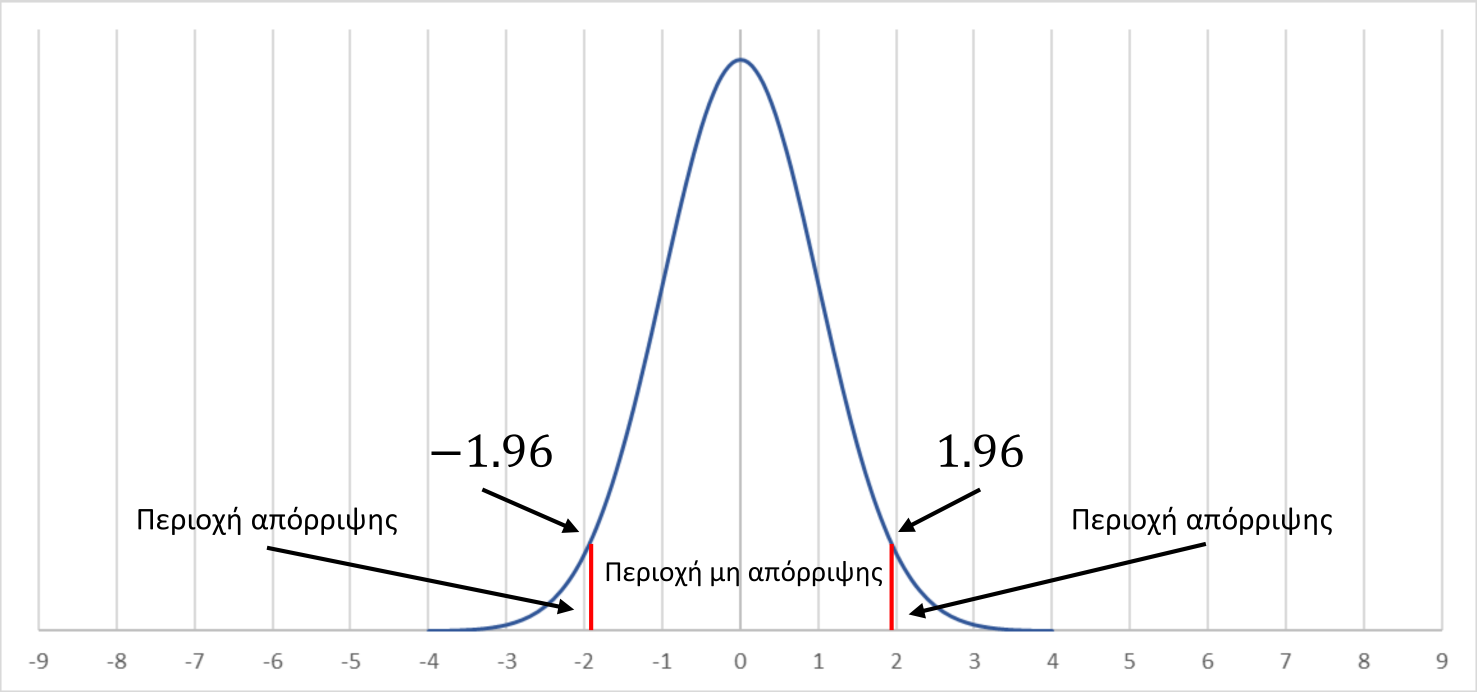
Βήμα 3: Εισαγωγή αριθμών από Βήμα 1 και Βήμα 2 στον τύπο του test

$$SE = \sqrt{0.217(1 - 0.217)\left(\frac{1}{240} + \frac{1}{220}\right)} = 0.038$$

$$z = \frac{0.083 - 0.364}{0.038} = -7.30$$

Πρέπει να ελέγξουμε εάν η **z-τιμή** βρίσκεται εντός της **περιοχής απόρριψης**

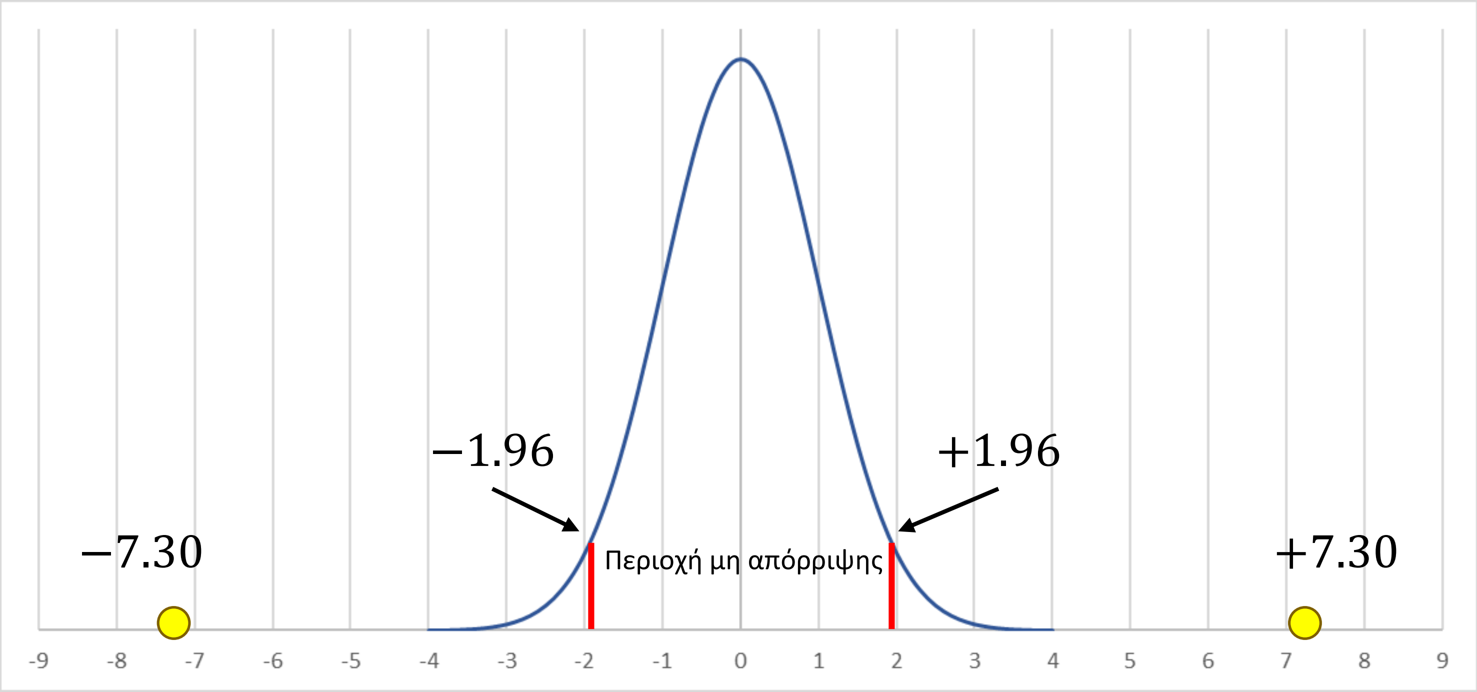
Βήμα 4: z-τιμή που σχετίζεται με επίπεδο σημαντικότητας 5%



Βήμα 5: Στατιστική σημαντικότητα

Η τιμή $z = -7.30$ είναι μικρότερη από το 5% κρίσιμο σημείο της τυπικής κανονικής κατανομής, που είναι 1.96

Οπότε, υπάρχει ένδειξη ($P < 0.05$) ότι πραγματικός εμβολιασμός ελαττώνει τον κίνδυνο προσβολής από γρίπη, **στατιστικά σημαντικά**



95% διάστημα εμπιστοσύνης

Το 95% διάστημα εμπιστοσύνης (CI) δίνεται από την παρακάτω σχέση:

$$\left((p_1 - p_2) - 1.96 \cdot SE, (p_1 - p_2) + 1.96 \cdot SE\right)$$

$$SE = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}$$

95% διάστημα εμπιστοσύνης

$$\left((p_1 - p_2) - 1.96 \cdot SE, (p_1 - p_2) + 1.96 \cdot SE\right)$$

$$\left((0.083 - 0.364) - 1.96 \cdot 0.038, (p_1 - p_2) + 1.96 \cdot 0.038\right)$$

$$(-0.281 - 0.07448, -0.281 + 0.07448)$$

$$(-0.35548, -0.20652)$$

$$(-0.3555, -0.2065)$$

95% διάστημα εμπιστοσύνης

Το 95% διάστημα εμπιστοσύνης για τη μείωση του ποσοστού της γρίπης λόγω του εμβολίου, ή για τη διαφορά των δύο ποσοστών, είναι

$$(-0.3555, -0.2065)$$

Συνεπώς, η πιθανότητα προσβολής από γρίπη όταν γίνεται πραγματικός εμβολιασμός είναι χαμηλότερος μεταξύ 20.65% και 35.55% σε σχέση με την χορήγηση placebo

Επειδή το 95% διάστημα εμπιστοσύνης **δεν περιέχει το μηδέν,** αυτό υποδηλώνει ότι η διαφορά μεταξύ των δύο ποσοστών **είναι στατιστικά σημαντική** σε επίπεδο σημαντικότητας 5%

Practice

1. Researchers want to test the effectiveness of a new anti-anxiety medication. In clinical testing, 64 out of 200 people taking the medication report symptoms of anxiety. Of the people receiving a placebo, 92 out of 200 report symptoms of anxiety. Is the medication working any differently than the placebo? Test this claim using alpha = 0.05.

2. Suppose a Drug Company develops a new drug, designed to prevent colds. The company states that the drug is equally effective for men and women. To test this claim, they choose a simple random sample of 100 women and 200 men from a population of 100.000 volunteers.

   At the end of the study, 38% of the women caught a cold; and 51% of the men caught a cold. Based on these findings, can we reject the company's claim that the drug is equally effective for men and women? Use a 0.05 level of significance.