

Διάγραμμα εμποτισμού για τη διαφορά
των μέσων $\mu_1 - \mu_2$ δύο ημεθυσμών

Ποιο συχνά στη Στατιστική μας ενδιαφέρει να συγκρίνουμε τους μέσους μ_1 και μ_2 δύο ημεθυσμών.
(Είναι η μια θεραπεία καλύτερη από την άλλη;
Είναι η μια μέθοδος εξεργασίας μετέττου καλύτερη από την άλλη;)

Εστω X_1, \dots, X_n τ.δ. από την $N(\mu_1, \sigma_1^2)$
και Y_1, \dots, Y_m τ.δ. ανεξάρτητο από το πρώτο
από την $N(\mu_2, \sigma_2^2)$. Εστω σ_1^2, σ_2^2 γνωστά.

Μας ενδιαφέρει να φτιάξουμε Δ. Ε. για
την παράμετρο $\boxed{\mu_1 - \mu_2}$.

Θα βασιστούμε στην εκτιμήτρια

$$\boxed{\bar{X} - \bar{Y}}$$

Οδηγίτζια συνάρτηση

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

Αν. Θεώρημα 2 $(\sum_{i=1}^n k_i X_i)$ ($k_1=1, k_2=-1$)

$$\left. \begin{array}{l} \bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \\ \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right) \end{array} \right\} \rightarrow \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

↑
προσσχή

$$P\left(-z_{\alpha/2} \leq \text{οδυσίγρια} \leq z_{\alpha/2}\right) \geq 1 - \alpha$$

Μετα από πράξεις:

Δ.Ε

για το $\mu_1 - \mu_2$:

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

Αν τα n και m είναι μεγάλα,

τότε η οδυσίγρια $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$

Προσεγγιστικά

και τα τ.δ. X_1, \dots, X_n και Y_1, \dots, Y_m μπορούν να ακολουθούν οποιαδήποτε κατανομή

Δ.Ε. ιδιο :

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

Εστω τώρα $X_1, \dots, X_n \sim N(\mu_1, \sigma_1^2)$ και

$Y_1, \dots, Y_m \sim N(\mu_2, \sigma_2^2)$ ανεξάρτητα και

σ_1^2, σ_2^2 αγνώστα.

α' η περίπτωση

$$\sigma_1^2 = \sigma_2^2 = \sigma^2.$$

$$\text{Τότε } \text{Var}(\bar{X} - \bar{Y}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \sigma^2 \left(\frac{1}{n} + \frac{1}{m} \right)$$

Αφού σ^2 αγνώστο θα εκτιμηθεί από τα
δύο δείγματα τα οποία έχουν κοινή διασπορά.

Η εκτίμηση του σ^2 ορίζεται ως

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

↑
pooled

: σταθμικός μέσος

από των
 S_1^2, S_2^2

δειγματικών
διασπορών

Εναλλακτικά,

$$S_p^2 = \frac{(n-1) \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} + (m-1) \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{m-1}}{n-1 + m-1} =$$
$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n+m-2}$$

Ευχρηστική: $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$ *

Δ.Ε.

μια αναμέτρηση

$$\bar{X} - \bar{Y} \pm t_{n+m-2, \alpha/2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

* Απόδειξη

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}}} = \frac{\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/n + \sigma^2/m}}}{\sqrt{\frac{S_p^2}{\sigma^2}}} =$$

$$\frac{Z}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{(n+m-2)\sigma^2}}} = \frac{Z}{\sqrt{\left\{ \underbrace{\frac{(n-1)S_1^2}{\sigma^2}}_{\chi_{n-1}^2} + \underbrace{\frac{(m-1)S_2^2}{\sigma^2}}_{\tilde{\chi}_{m-1}^2} \right\} / (n+m-2)}}$$

Θεώρημα

$$= \frac{Z}{\sqrt{(\chi_{n-1}^2 + \tilde{\chi}_{m-1}^2) / (n+m-2)}} \sim t_{n+m-2}$$

ορίσμος της t

αντίστροφο: χ_{n+m-2}^2 (Θεώρημα 3)

Αν n, m μεγάλα ή έστω $n+m-2$ μεγάλο

Το συμείο $t_{n+m-2, \alpha/2}$ προσεγγίζεται από το

$$z_{\alpha/2}$$

β' περίπτωση

σ_1^2, σ_2^2 άγνωστα αλλά ίσα

Τα σ_1^2, σ_2^2 συμπίπτουν { εξαρτάται } από τα

$$S_1^2, S_2^2$$

$$\Delta. \Xi. \quad \bar{X} - \bar{Y} \pm t_{K, \alpha/2} \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$$

$$\text{όπου } \theta. \epsilon. \quad K = \frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m} \right)^2}{\frac{\left(\frac{S_1^2}{n} \right)^2}{n-1} + \frac{\left(\frac{S_2^2}{m} \right)^2}{m-1}}$$

Confidence interval for population difference of means $\mu_1 - \mu_2$

n and m large and σ_1^2 σ_2^2 known	n and m large and σ_1^2 σ_2^2 unknown	n and m small and σ_1^2 σ_2^2 known	n and m small and σ_1^2 σ_2^2 unknown but equal to σ^2
$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$	$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$	$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$	$\bar{X} - \bar{Y} \pm t_{n+m-2, \alpha/2} \sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}}$ <p>where the pooled estimator of the common variance is</p> $S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$
Any distribution for the populations	Any distribution for the populations	Populations Normally distributed	Populations Normally distributed

Εξήγηση βαθμών ελεύθερης

Από το $\sum_{i=1}^n (X_i - \mu)^2$ το οποίο έχει η

βαθμὸς ελευθερίας με το συνολικό ου 01

$X_i - \mu$ $i=1, \dots, n$ αποκρίσεις είναι ελεύθερες

να πάρουν ου τιμές θέλουν,

όταν πάρει το άθροισμα

$$\sum_{i=1}^n (X_i - \bar{X})^2$$

χάνουμε ένα βαθμὸ ελευθερίας με έχουμε $n-1$

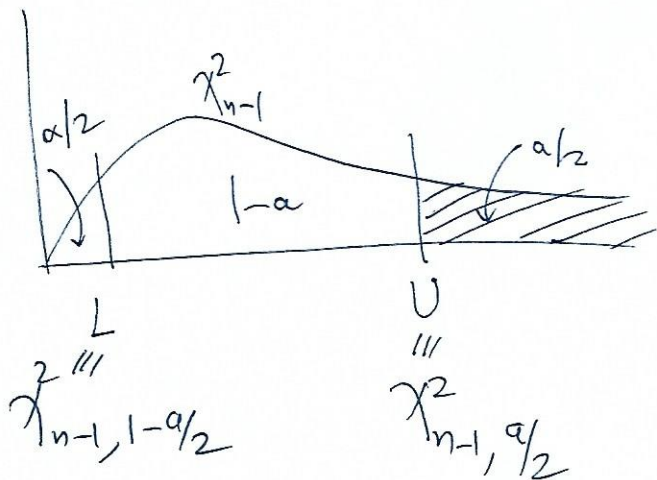
διότι υπάρχει μια γραμμική σχέση που ισχύει

ανάμεσα στις αποκρίσεις $X_i - \bar{X}$, $i=1, \dots, n$.

Συγκεκριμένα,

$$\sum_{i=1}^n (X_i - \bar{X}) = 0 \quad \left(\begin{array}{l} \text{δίου } \sum_{i=1}^n X_i - n\bar{X} = 0 \\ \text{εξ ορισμού} \end{array} \right)$$

Άρα οι $n-1$ ανηγώσεις $X_i - \bar{X}$ είναι ελεύθερες
να πάρουν οτιδήποτε θέλουν αλλά μία ανόγκη
πρέπει να जुड़े σε σχέση με τις υπολογιστές $n-1$
 $n-X$. $X_n - \bar{X} = -\sum_{i=1}^{n-1} (X_i - \bar{X})$



οπου $P(X_{n-1}^2 > \chi_{n-1, \alpha/2}^2) = \alpha/2$

Με βάση αυτό το συμπέρασμα

$L \equiv \chi_{n-1, 1-\alpha/2}^2$ γιατί έχει στα δεξιά του
 $1-\alpha/2$ πιθανότητα

$$P\left(\chi^2_{n-1, 1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1, \alpha/2}\right) = 1-\alpha \Rightarrow$$

⋮

$$P\left(\frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}}\right) = 1-\alpha$$

Δ.Ε.
για το σ^2
με σ.ε. $1-\alpha$

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}}\right)$$

Δ.Ε.
για το σ :

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}}}\right)$$

Διάστημα εμπιστοσύνης για το λόγο των διασπορών

Δύο ηυθρομίων

Έστω X_1, \dots, X_n τ.δ. από την $N(\mu_1, \sigma_1^2)$ και Y_1, \dots, Y_m από την $N(\mu_2, \sigma_2^2)$ ανεξάρτητα από το πρώτο, Έστω $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ αγνώστα.

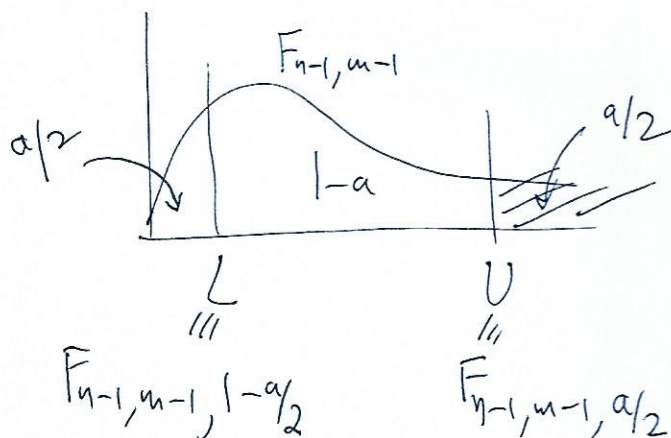
Θέλουμε να κατασκευάσουμε Δ.Ε. για το λόγο

$$\frac{\sigma_1^2}{\sigma_2^2}$$

Οδυσζίζρια :
συνάρτηση

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1, m-1}$$

$$P(L \leq \text{οδυσζίζρια} \leq U) = 1 - \alpha$$



$$P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq F_{n-1, m-1, \alpha/2}\right) = 1 - \alpha$$

$$P\left(\frac{S_1^2}{F_{n-1, m-1, \alpha/2} S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{F_{n-1, m-1, 1-\alpha/2} S_2^2}\right) = 1 - \alpha$$

Δ.Ε για το $\frac{\sigma_1^2}{\sigma_2^2}$ με σ.ε $1 - \alpha$

$$\left(\frac{S_1^2}{F_{n-1, m-1, \alpha/2} S_2^2}, \frac{S_1^2}{F_{n-1, m-1, 1-\alpha/2} S_2^2} \right)$$

✓ Θέλουμε διάστημα για το $\frac{\sigma_2^2}{\sigma_1^2}$ αντιστρέφουμε

τα όρια του διαστήματος. (και αντιστρέφουμε)

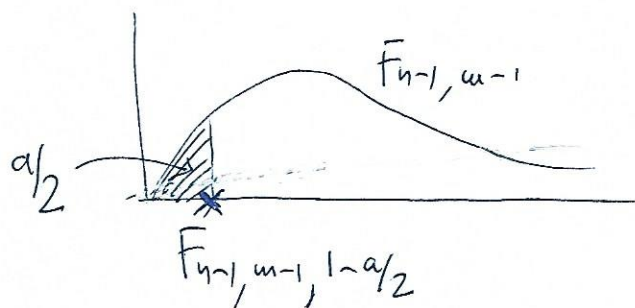
Παρατήρηση

Ο πίνακας της F κατανομής δίνει σημεία για μικρές πιθανότητες στα δεξιά (π.χ. 0.05, 0.005) αλλά όχι για μεγάλες πιθανότητες στα δεξιά τα οποία επίσης χρειάζονται στα διαστήματα

λογος οτι αν

$$F \sim F_{n-1, u-1} \quad \text{τότε} \quad \frac{1}{F} \sim F_{u-1, n-1}$$

Αρα



$$P(F_{n-1, u-1} \leq F_{n-1, u-1, 1-\alpha/2}) = \frac{\alpha}{2}$$

$$\Rightarrow P\left(\frac{1}{F_{n-1, u-1}} \geq \frac{1}{F_{n-1, u-1, 1-\alpha/2}}\right) = \frac{\alpha}{2}$$

$$\Rightarrow P(F_{u-1, n-1} \geq \frac{1}{F_{n-1, u-1, 1-\alpha/2}}) = \frac{\alpha}{2}$$

αλλα και

$$P(F_{u-1, n-1} \geq F_{u-1, n-1, \alpha/2}) = \frac{\alpha}{2}$$

Αρα $\frac{1}{F_{n-1, u-1, 1-\alpha/2}} = F_{u-1, n-1, \alpha/2}$ η

εναλλακτικά,

$$F_{n-1, u-1, 1-\alpha/2} = \frac{1}{F_{u-1, n-1, \alpha/2}}$$

υποδεικνύει
στον
πίνακα.

Παράδειγμα

Εστω ότι θέλουμε να συγκρίνουμε δύο εκπαιδευτικές μεθόδους Α και Β. Παίρνουμε τ.δ. μαθητών που εκπαιδεύθηκαν με τη μέθοδο Α και από τυχαίο δείγμα μαθητών που εκπαιδεύθηκαν με τη μέθοδο Β και τους βαθμολογούμε σε κάποια κοινή εξέταση. (σε κάποια κλίμακα)

Δίνονται οδοντιατρικός μέσος του πρώτου δείγματος 3.6 με διασπορά 4.14 και μεγέθος δείγματος 21. Επίσης ο δοντιατρικός μέσος του δεύτερου δείγματος 5.6 με διασπορά 7.26 και μεγέθος δείγματος 10.

- (α) Να βρεθεί Δ.Ε. με σ.ε. 95% για τη διαφορά των μέσων $\mu_1 - \mu_2$ των δυο πληθυσμών.
- (β) Να βρεθεί Δ.Ε. με σ.ε. 90% για τη διασπορά του πρώτου πληθυσμού σ^2 .
- (γ) Να βρεθεί ένα 90% Δ.Ε. για το λόγο των διασπορών των πληθυσμών $\frac{\sigma_1^2}{\sigma_2^2}$.

Σχολιάστε τα αποτελέσματα σε σχέση με περίληψεις.
Τι υποθέση χρειάζεται να γίνει για να κατασκευαστούν αυτά τα διαστήματα;

(α) $n_1 = 21$, $n_2 = 10$ 'μικρά'

Τα σ_1^2, σ_2^2 αγνώστα

Ξέρουμε τις δειγματικές διασπορές $s_1^2 = 4.14$ και $s_2^2 = 7.26$

Επίσης $\bar{x} = 3.6$ και $\bar{y} = 5.6$, $\alpha = 5\%$

Επειδή ο λόγος των δειγματικών διασπορών είναι

$$\frac{s_1^2}{s_2^2} = 0.57 \quad \text{δηλαδή κοντά στο } 1 \quad \text{θα κάνουμε}$$

Την υπόθεση ότι $\sigma_1^2 = \sigma_2^2 = \sigma^2$ $\left(\begin{array}{l} \text{Εμπειρικός κανόνας} \\ \frac{1}{2} \leq \frac{s_1^2}{s_2^2} < 2 \end{array} \right)$

A.E.:

$$\bar{x} - \bar{y} \pm t_{n+m-2, \alpha/2} \cdot S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$S_p^2 = \frac{20 \cdot 4.14 + 9 \cdot 7.26}{29} = 5.10$$

$$3.6 - 5.6 \pm t_{29, 0.025} \sqrt{5.10} \sqrt{\frac{1}{21} + \frac{1}{10}}$$

$$-2 \pm 2.045 \cdot 0.867 = -2 \pm 1.773:$$

$$(-3.773, -0.227)$$

Με εμπιστοσύνη 95% το διάστημα $(-3.773, -0.227)$ περιέχει
τη διαφορά των μέσων $\mu_A - \mu_B$. Το 0 \notin διάστημα άρα
μηνύμε να απορρίψουμε ότι οι μέσοι των γυθισμών
είναι ίσοι, (φαίνεται ότι $\mu_A < \mu_B$ και η Β μέθοδος υπερφέρει.)

6) ΔG. ja 20 σ_1^2 :

$$\left(\frac{(n-1)S_1^2}{\chi_{n-1, \alpha/2}^2}, \frac{(n-1)S_1^2}{\chi_{n-1, 1-\alpha/2}^2} \right)$$

$$\alpha = 0.1 \rightarrow \alpha/2 = 0.05$$

Ano nirana χ^2 : $\chi_{20, 0.05}^2 = 31.410$ na $\chi_{20, 0.95}^2 = 10.851$

$$\left(\frac{20 \cdot 4.14}{31.410}, \frac{20 \cdot 4.14}{10.851} \right); (2.636, 7.630)$$

γ) Δ.Ε για το $\frac{\sigma_1^2}{\sigma_2^2}$:

$$\left(\frac{S_1^2}{S_2^2 \cdot F_{n-1, m-1, \alpha/2}}, \frac{S_1^2}{S_2^2 \cdot F_{n-1, m-1, 1-\alpha/2}} \right)$$

$$\alpha = 0.1, \quad \alpha/2 = 0.05$$

$$F_{20, 9, 0.05} = 2.94 \text{ από πίνακα.}$$

$$F_{20, 9, 0.95} = \frac{1}{F_{9, 20, 0.05}} = \frac{1}{2.39} = 0.418$$

$$\left(\frac{4.14}{7.26 \cdot 2.94}, \frac{4.14}{7.26 \cdot 0.418} \right) = (0.193, 1.364)$$

Με εμπιστοσύνη 90% το Δ.Ε. $(0.193, 1.364)$ περιέχει το
 λογάριθμό διασπορών $\frac{\sigma_1^2}{\sigma_2^2}$. Επειδή το $1 \in \Delta.Ε$ έχουμε τεκμήρια
 να θεωρήσουμε ότι οι διασπορές είναι ίσες.
 Σε όλα τα εξωκείμελα οι ομνύοντες πρέπει να ακολουθούν
 κανονική κατανομή