Διάστημα εμπιστοσύνης της μ_1 - μ_2

δύο ανεξάρτητων τ.μ. X_1 και X_2

Μέσες τιμές: μ_1 και μ_2

Διασπορές: σ_1^2 και σ_2^2

Δείγμα μεγέθους n_1, n_2 από τον πληθυσμό της X_1, X_2

Δειγματικές μέσες τιμές: $\overline{x}_1$ και $\overline{x}_2$

Δειγματικές διασπορές: s_1^2 και s_2^2

Θέλουμε να εκτιμήσουμε τη διαφορά των μέσων τιμών μ_1 και μ_2
Σημειακή εκτίμηση της διαφοράς μ_1 - μ_2: $\overline{x}_1 - \overline{x}_2$

<u>Γνωστές διασπορές</u> σ_1^2 **<u>και</u>** σ_2^2

Υποθέτουμε

είτε $n_1 > 30$ και $n_2 > 30$

ή X_1 και X_2 ακολουθούν κανονική κατανομή

➔ εκτιμήτρια $\bar{x}_1 - \bar{x}_2$ ακολουθεί κανονική κατανομή

μέση τιμή $$\mu_{\bar{x}_1-\bar{x}_2} \equiv \mathrm{E}\left[\bar{x}_1 - \bar{x}_2\right] = \mu_1 - \mu_2$$

διασπορά $$\sigma^2_{\bar{x}_1-\bar{x}_2} \equiv \mathrm{Var}\left[\bar{x}_1 - \bar{x}_2\right] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Αν X_1 και X_2 *ομοσκεδαστικές* ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Εκτιμήτρια	$\bar{x}$	$\Rightarrow$	$\bar{x}_1 - \bar{x}_2$
Μέση τιμή εκτιμήτριας	μ	$\Rightarrow$	$\mu_1 - \mu_2$
Διασπορά εκτιμήτριας	$\frac{\sigma^2}{n}$	$\Rightarrow$	$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

(1-α)% δ.ε. της μ_1 - μ_2: $\boxed{(\bar{x}_1 - \bar{x}_2) \pm z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Ερμηνεία του δ.ε.:

Αν **περιέχει** το 0 ➔ μ_1 και μ_2 **δε** διαφέρουν
(για το επίπεδο εμπιστοσύνης που χρησιμοποιήσαμε και με βάση τα συγκεκριμένα δεδομένα)

Αν είναι **θετικό** ➔ η τ.μ. X_1 είναι κατά μέσο όρο μεγαλύτερη από τη X_2 κατά ένα ποσό που κυμαίνεται στα όρια του διαστήματος που εκτιμήσαμε

[αντίστοιχα για αρνητικό δ.ε.]

Παράδειγμα

Μελετήσαμε το χρόνο ανάφλεξης σε δύο δείγματα από δύο υλικά ταπετσαρίας Α και Β μεγέθους $n_1=30$ και $n_2=20$.

Είναι ο χρόνος ανάφλεξης του υλικού Α κατά μέσο όρο διαφορετικός από αυτόν του υλικού Β;

Θεωρούμε πως η διασπορά είναι κοινή και γνωστή: $\sigma^2 = 4s^2$

Συνθήκες:

Είναι τα 2 δείγματα μεγάλα;

Ακολουθεί ο χρόνος ανάφλεξης για τα δύο υλικά κανονική κατανομή;

Υλικό 1		Υλικό 2
Φύλλο	Μίσχος	Φύλλο
5	1.	8
3	2.	38
86440	3.	44445
985441	4.	566688
99500	5.	069
8732	6.	07
6533	7.	5
64	8.	
40	9.	

Δειγματικές μέσες τιμές A: $\bar{x}_1 = 5.59\,\mathrm{s}$ B: $\bar{x}_2 = 4.46\,\mathrm{s}$

$\bar{x}_1 - \bar{x}_2 = 1.13\,\mathrm{s}$

$z_{0.025} = 1.96$ και $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 4\,\mathrm{s}^2$

95% δ.ε. της $\mu_1 - \mu_2$

$$\boxed{(\bar{x}_1 - \bar{x}_2) \pm z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$1.13 \pm 1.96 \cdot \sqrt{4 \cdot \left(\frac{1}{30} + \frac{1}{20}\right)} = 1.13 \pm 1.132 \quad \mapsto \quad (-0.002, 2.262)$$

Υπάρχει σημαντική διαφορά μεταξύ των δύο υλικών;

Άγνωστες διασπορές, μεγάλο μέγεθος δείγματος

Όταν το δείγμα είναι μεγάλο:

$s_1^2 \to \sigma_1^2$ και $s_2^2 \to \sigma_2^2$

(1-α)% δ.ε. της $\mu_1 - \mu_2$:

$$\boxed{\left(\bar{x}_1 - \bar{x}_2\right) \pm z_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Άγνωστες διασπορές, μικρό μέγεθος δείγματος

Περίπτωση 1

Οι κατανομές των X_1 και X_2 δε είναι κανονικές ➔
δε μπορούμε να προσδιορίσουμε την κατανομή της διαφοράς $\bar{x}_1 - \bar{x}_2$

για να εκτιμήσουμε το δ.ε. ➔ μη-παραμετρική προσέγγιση.

Περίπτωση 2

Υποθέτουμε

1. Οι κατανομές των X_1 και X_2 είναι κανονικές
2. X_1 και X_2 ομοσκεδαστικές ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

Εκτίμηση της κοινής διασπορά σ^2
Δειγματική κοινή διασπορά :

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Εκτίμηση της διασποράς της $\bar{x}_1 - \bar{x}_2$

$$s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Μετασχηματισμός

$$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

(1-a)% δ.ε. της $\mu_1 - \mu_2$:

$$\boxed{(\bar{x}_1 - \bar{x}_2) \pm t_{n_1 + n_2 - 2, a/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Παράδειγμα

Άγνωστη διασπορά του χρόνου ανάφλεξης για τα δύο υλικά ταπετσαρίας.

Δειγματικές διασπορές: A: $s_1^2 = 4.12 \text{ s}^2$, B: $s_2^2 = 2.11 \text{ s}^2$

Δεχόμαστε ότι η διασπορά δε διαφέρει σημαντικά στα δύο υλικά.
Δεχόμαστε επίσης ότι ο χρόνος ανάφλεξης και για τα δύο υλικά ακολουθεί κανονική κατανομή.

Τα δείγματα είναι σχετικά μικρά ➔ $\bar{x}_1 - \bar{x}_2$ ακολουθεί κατανομή student

Δειγματική κοινή διασπορά: $s_p^2 = \dfrac{29 \cdot 4.12 + 19 \cdot 2.11}{48} = 3.324$ $s_p = 1.823$

Βαθμοί ελευθερίας: $n_1 + n_2 - 2 = 48$ $t_{48, 0.025} = 2.01$

$\bar{x}_1 - \bar{x}_2 = 1.13$

Το 95% δ.ε. είναι

$$\boxed{\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{n_1+n_2-2, a/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$1.13 \pm 2.01 \cdot 1.823 \sqrt{\frac{1}{30} + \frac{1}{20}} = 1.13 \pm 1.058 \quad \mapsto \quad (0.072, 2.188)$$

Παρατηρήσεις

Υπάρχει φανερή αντιστοιχία των περιπτώσεων για το δ.ε. της μέσης τιμής μ και το δ.ε. της διαφοράς δύο μέσων τιμών $\mu_1 - \mu_2$.

Για τη διαφορά μέσων τιμών υπάρχει ακόμα η περίπτωση των άνισων κι αγνώστων διασπορών σε συνδυασμό με μικρά δείγματα.

διασπορές των X_1, X_2	*κατανομές των X_1, X_2*	n_1, n_2	κατανομή της $\bar{x}_1 - \bar{x}_2$	δ.ε.της $\bar{x}_1 - \bar{x}_2$
γνωστές	*κανονική*		$z \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$	$(\bar{x}_1 - \bar{x}_2) \pm z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
γνωστές	*μη κανονική*	*μεγάλα*	$z \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$	$(\bar{x}_1 - \bar{x}_2) \pm z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
γνωστές	*μη κανονική*	*μικρά*	*μη-παραμετρική*	*μη-παραμετρικό*
άγνωστες άνισες ή ίσες		*μεγάλα*	$z \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \mathrm{N}(0,1)$	$(\bar{x}_1 - \bar{x}_2) \pm z_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
άγνωστες και ίσες	*κανονική*	*μικρά*	$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2, a/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
άγνωστες και ίσες	*μη κανονική*	*μικρά*	*μη-παραμετρική*	*μη-παραμετρικό*
άγνωστες και άνισες		*μικρά*	----	----

Διάστημα της διαφοράς δύο αναλογιών $p_1 - p_2$

Συγκρίνουμε δύο πληθυσμούς ως προς μια ιδιότητα

Μελετάμε τη διαφορά αναλογιών $p_1 - p_2$

m_1 είναι το πλήθος «επιτυχιών» στο δείγμα μεγέθους n_1 από τον πρώτο πληθυσμό $\quad \hat{p}_1 = \frac{m_1}{n_1}$

m_2 είναι το πλήθος «επιτυχιών» στο δείγμα μεγέθους n_2 από το δεύτερο πληθυσμό $\quad \hat{p}_2 = \frac{m_2}{n_2}$

Εκτίμηση της διαφοράς $p_1 - p_2$: $\hat{p}_1 - \hat{p}_2$

Για μεγάλα n_1 και n_2 ➔ $\hat{p}_1 - \hat{p}_2$ ακολουθεί κανονική κατανομή

μέση τιμή $E[\hat{p}_1 - \hat{p}_2] = p_1 - p_2$ διασπορά $\mathrm{Var}[\hat{p}_1 - \hat{p}_2] = \frac{p_1 \cdot (1 - p_1)}{n_1} + \frac{p_2 \cdot (1 - p_2)}{n_2}$

Μετασχηματισμός σε τυπική κανονική κατανομή

$$z \equiv \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 \cdot (1 - p_1)}{n_1} + \frac{p_2 \cdot (1 - p_2)}{n_2}}} \sim N(0,1)$$

Αντικαθιστούμε

$\hat{p}_1 \rightarrow p_1$

$\hat{p}_2 \rightarrow p_2$

(1-a)% δ.ε. της $p_1 - p_2$: $\boxed{(\hat{p}_1 - \hat{p}_2) \pm z_{a/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}}}$

Παράδειγμα

Θέλουμε να εκτιμήσουμε αν υπάρχει διαφορά στο ποσοστό των μικρών διαμερισμάτων σε δύο περιοχές A και B.

Περιοχή A: $m_1 = 48$ μικρά διαμερίσματα (με ένα ή δύο δωμάτια) σε $n_1 = 120$ διαμερίσματα

Περιοχή B: $m_2 = 60$ μικρά διαμερίσματα σε $n_2 = 100$ διαμερίσματα

Περιοχή A: Δειγματική αναλογία : $\hat{p}_1 = \dfrac{48}{120} = 0.4$

Περιοχή B: Δειγματική αναλογία : $\hat{p}_2 = \dfrac{60}{100} = 0.6$

Σημειακή εκτίμηση της διαφοράς αναλογίας των μικρών διαμερισμάτων στις δύο περιοχές: $\hat{p}_1 - \hat{p}_2 = -0.2$ (σε ποσοστό είναι *–20%*)

95% δ.ε. της $p_1 - p_2$: ($z_{a/2} = 1.96$)

$$\boxed{\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{a/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot \left(1 - \hat{p}_1\right)}{n_1} + \frac{\hat{p}_2 \cdot \left(1 - \hat{p}_2\right)}{n_2}}}$$

$$-0.2 \pm 1.96 \cdot \sqrt{\frac{0.4 \cdot \left(1 - 0.4\right)}{120} + \frac{0.6 \cdot \left(1 - 0.6\right)}{100}} = -0.2 \pm 0.13 \quad \mapsto \quad \left(-0.33, -0.07\right)$$

Το 95% δ.ε. **δεν περιέχει** το 0 ➔ με βάση αυτά τα δείγματα και σ' αυτό το επίπεδο εμπιστοσύνης συμπεραίνουμε ότι η αναλογία μικρών διαμερισμάτων είναι **από** 0.07 **ως** 0.33 **μικρότερη** στην περιοχή A