Σύγκριση μέσων τιμών δύο κανονικών πληθυσμών

Ανεξάρτητα δείγματα

$H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$ ή $\mu_1 > \mu_2$ ή $\mu_1 < \mu_2$.

1η περίπτωση

Γνωστές διακυμάνσεις

→ στατιστική συνάρτηση ελέγχου κάτω από την Ho

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

H_1	κρίσιμη περιοχή
$\mu_1 < \mu_2$	$Z_0 < -Z_{1-a}$
$\mu_1 > \mu_2$	$Z_0 > Z_{1-a}$
$\mu_1 \neq \mu_2$	$Z_0 < -Z_{1-a/2}$ ή $Z_0 > Z_{1-a/2}$

2η περίπτωση

Άγνωστες ίσες διακυμάνσεις

→ σ.σ ελέγχου υπό την Ho

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

H_1	κρίσιμη περιοχή
$\mu_1 < \mu_2$	$t < -t_{1-a,\, n_1+n_2-2}$
$\mu_1 > \mu_2$	$t > t_{1-a,\, n_1+n_2-2}$
$\mu_1 \neq \mu_2$	$t < -t_{1-a/2,\, n_1+n_2-2}$ ή $t > t_{1-a/2,\, n_1+n_2-2}$

3η περίπτωση

Άγνωστες άνισες διακυμάνσεις — Μεγάλα δείγματα

σ.σ ελέγχου

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

H_1	κρίσιμη περιοχή
$\mu_1 < \mu_2$	$Z_0 < -Z_{1-a}$
$\mu_1 > \mu_2$	$Z_0 > Z_{1-a}$
$\mu_1 \neq \mu_2$	$Z_0 > Z_{1-a/2}$ ή $Z_0 < -Z_{1-a/2}$

Ασκηση 1

Μετρήθηκε η πίεση σε δύο ανεξάρτητες ομάδες παιδιών, αποτελούμενες από 8 και 10 παιδιά αντίστοιχα. Στην Α ομάδα τα παιδιά έχουν υπερτασικούς γονείς ενώ στη Β ομάδα δεν παρουσιάζουν υπέρταση. Υποθέτουμε ότι οι τιμές ακολουθούν κανονική κατανομή με κοινή διασπορά σ^2. Υπάρχει διαφορά στη μέση τιμή πίεσης στις δύο ομάδες σε ε.σ $\alpha=10\%$

ΟΜΑΔΑ Α : 100, 102, 96, 106, 110, 120, 112, 90

ΟΜΑΔΑ Β: 104, 88, 100, 98, 102, 92, 96, 100, 96, 97

Λύση:

$$\bar{x}_1 = \frac{\sum_{i=1}^{8} x_i}{n_1} = \frac{836}{8} = 104,5$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{10} x_i}{n_2} = \frac{973}{10} = 97,3$$

$$S_1^2 = \frac{1}{n_1-1}\left(\sum_{i=8}^{8} x_i^2 - \frac{\left(\sum_{i=1}^{8} x_i\right)^2}{n_1}\right) = \frac{1}{7}\left(88.000 - \frac{836^2}{8}\right) =$$

$$91,143$$

Ομοίως

$$S_2^2 = 22,236$$

Τελικά

$$S = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 52,38$$

συγχωνευμένη δειγματική διασπορά

$H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

Η τιμή του στατ. κριτηρίου:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{104,5 - 97,3}{\sqrt{52,38} \cdot \sqrt{\frac{1}{8} + \frac{1}{10}}} = 2,1$$

Η κρίσιμη τιμή είναι $t_{n_1+n_2-2; 1-\alpha/2} = t_{16; 0,95} = 1,746$

Είναι $t > t_{16, 0,95}$ δηλ. είμαστε στην κ.π και απορρίπτουμε H_0

Άρα υπάρχει διαφορά στην αρτηριακή πίεση ανάμεσα στις δύο ομάδες

Άσκηση 2

Διεξήχθη μια έρευνα για το εισόδημα των εργαζομένων στα μεγάλα αστικά κέντρα και στην επαρχία

Δείγμα 1 (Αστικά κέντρα)
$n_1 = 50$
$\overline{X}_1$ 750
$S_1^2 = 10^2$

Δείγμα 2 (επαρχία)
$n_2 = 72$
$\overline{X}_2 = 770$
$S_1^2 = 12^2$

Είναι το μέσο εισόδημα των εργαζομένων στα μεγάλα αστικά κέντρα μικρότερο από των εργαζομένων στην επαρχία;

Μονόπλευρος έλεγχος σε ε.σ $a = 0,05$

$H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 < \mu_2$

Στατιστικό ελέγχου

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{-20}{\sqrt{\frac{100}{50} + \frac{144}{72}}} = \frac{-20}{\sqrt{2+2}} = -10$$

Κριτική τιμή ελέγχου

$Z_{0.95} = 1,65$

Επειδή $Z_0 < -Z_{0.95}$ η μηδενική υπόθεση απορρίπτεται

Συμπέρασμα

Για $a = 0,05$ το εισόδημα των εργαζομένων στα αστικά κέντρα είναι σημαντικά μικρότερο από των εργαζομένων στην επαρχία.

Σύγκριση διακυμάνσεων από δύο κανονικούς πληθυσμούς

$H_0: \sigma_1^2 = \sigma_2^2$ αντιστοιχεί σε $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$

Στατιστική συνάρτηση ελέγχου κάτω από την H_0:

$$F_0 = \frac{S_1^2}{S_2^2} \sim F_{n-1, m-1}$$

H_1	κρίσιμη περιοχή
$\frac{\sigma_1^2}{\sigma_2^2} > 1$	$F_0 > F_{n-1, m-1, \alpha}$
$\frac{\sigma_1^2}{\sigma_2^2} < 1$	$F_0 < F_{n-1, m-1, 1-\alpha}$
$\frac{\sigma_1^2}{\sigma_2^2} \neq 1$	$F_0 > F_{n-1, m-1, \alpha/2}$ ή $F_0 < F_{n-1, m-1, 1-\alpha/2}$

Ασκηση 3

Από δύο πληθυσμούς που ακολουθούν την κανονική κατανομή ελήφθησαν δύο δείγματα (ένα από κάθε πληθυσμό) μεγέθους $n_1=12$ και $n_2=10$ αντίστοιχα με $S_1^2=14,5$ και $S_2^2=10,8$

Να ελεγχθεί σε επίπεδο σημαντικότητας 5%, αν η μεταβλητότητα του πρώτου πληθυσμού είναι μεγαλύτερη από την μεταβλητότητα του δεύτερου

$H_0: \sigma_1^2=\sigma_2^2$ vs $H_1: \sigma_1^2>\sigma_2^2$

$$f_0=\frac{S_1^2}{S_2^2}=\frac{14,5}{10,8}=1,34$$

$$f_{11,9,0.05}=4.53$$

$$f_0=1,34<f_{11,9,0,05}$$

Συμπέρασμα Δεν απορρίπτεται η H_0

Άσκηση 4

Ένα γκρουπ 5 ασθενών πήρε ένα φάρμακο Α. Έπειτα οι ασθενείς ζυγίστηκαν 42, 39, 38, 60, 41 κιλά. Ένα δεύτερο γκρουπ 7 ασθενών στο ίδιο νοσοκομείο πήρε ένα φάρμακο Β και ζυγίστηκε 38, 42, 56, 64, 68, 69, 62 κιλά. Υπάρχει διαφορά μεταξύ των δύο φαρμάκων;

$H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

επίπεδο σημαντικότητας 5%

Πριν συγκρίνουμε τις μέσες τιμές πρέπει να εξετάσουμε την ισότητα των διασπορών με ένα F-test

$H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$

$$S_1^2 = \frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}}{n_1 - 1} = 82,5$$

$$S_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{n_2 - 1} = 154,33$$

$$F_0 = \frac{S_1^2}{S_2^2} = 0,535$$

$$F_{4,6,0.975} = 6,23$$

Αφού $F_0 < F_{4,6,0.975} = 6,23$ αποδεχόμαστε την H_0, δλδ οι διακυμάνσεις είναι ίσες

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-2)S_2^2}{n_1+n_2-1} = \frac{330+926}{10}$$

$$= 125,6$$

$$t_0 = \frac{|44-57|}{\sqrt{125.6\left(\frac{1}{7}+\frac{1}{5}\right)}} = 1,98$$

$$t_{10,0.975} = 2.228$$

Άρα $t_0 < t_{10,0.975}$

<u>Συμπέρασμα</u>

Αποδεχόμαστε H_0 δλδ τα φάρμακα δεν παρουσιάζουν στατιστικά σημαντική διαφορά.