5.3.2 Άγνωστες διακυμάνσεις πληθυσμού. Μεγάλα δείγματα

Συνήθως, όταν αγνοούμε τις μ_1, μ_2 αγνοούμε και τις σ_1, σ_2. Στην περίπτωση αυτή μπορούμε, όταν n_1, n_2 >30 να αντικαταστήσουμε τις σ_1, σ_2 με τις δειγματικές τυπικές αποκλίσεις s_1 και s_2, αντίστοιχα και να ακολουθήσουμε την ίδια διαδικασία ελέγχου όπως και προηγουμένως.

5.3.3 Άγνωστες διακυμάνσεις κανονικών πληθυσμών. Μικρά δείγματα. Ισότητα διασπορών

Οι δύο κατανομές είναι κανονικές και μπορούμε ακόμη να κάνουμε την υπόθεση την υπόθεση

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

Στην περίπτωση αυτή μπορούμε να συνδυάσουμε τις πληροφορίες από την μεταβλητότητα του πρώτου δείγματος με αυτές του δεύτερου δείγματος για να εκτιμήσουμε την κοινή διακύμανση. Έτσι ορίζουμε:

$$S^2=\frac{\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2+\sum_{i=1}^{n_2}(X_{2i}-\overline{X}_2)^2}{n_1+n_2-2}$$

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

Η s^2 ως συνάρτηση των s_1 και s_2 είναι τυχαία μεταβλητή. Οι βαθμοί ελευθερίας της ν είναι n_1+n_2-2. Αποδεικνύεται ότι η τυχαία μεταβλητή

$$t=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

ακολουθεί την κατανομή student με $v=n_1+n_2-2$ βαθμούς ελευθερίας. Επομένως, στην περίπτωση αυτή ο έλεγχος της ισότητας των δύο μέσων τιμών θα γίνει σταδιακά ως εξής:

α. Υποθέσεις που γίνονται δεκτές: (i) Οι δύο κατανομές πληθυσμού είναι κανονικές και έχουν την ίδια διακύμανση, και (ii) Τα δείγματα είναι τυχαία και ανεξάρτητα.

β. Η μηδενική και η εναλλακτική υπόθεση:

$$\begin{array}{lll} H_0: & & \mu_1=\mu_2 \\ H_\varepsilon: & (i) & \mu_1\neq\mu_2 \\ & (ii) & \mu_1>\mu_2 \end{array}$$

γ. Το κριτήριο αποφάσεως: Υπολογίζουμε τα παρακάτω στατιστικά:

$$s=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}, \text{ και}$$

$$t_\pi=\frac{\bar{x}_1-\bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

δ. Η απόφαση: Η μηδενική υπόθεση θα απορριφθεί, σε επίπεδο σημαντικότητας α, για τον δίπλευρο και τον μονόπλευρο έλεγχο, αντίστοιχα, αν:

$$(i) \quad |t_\pi| > t_{n_1+n_2-2,\alpha/2}$$

$$(ii) \quad t_\pi > t_{n_1+n_2-2,\alpha}$$

Παράδειγμα 5.6

Για να ελέγξουμε αν η παραγωγικότητα του ανειδίκευτου εργάτη είναι διαφορετική στα δύο φύλα, πήραμε τυχαίο δείγμα n_1=5 ανδρών και τυχαίο δείγμα n_2=5 γυναικών και υπολογίσαμε τον αριθμό των προϊόντων X που παράγουν σε. μια ημέρα. Αν τα αποτελέσματα των μετρήσεων ήταν τα ακόλουθα:

Γυναίκες	14	17	12	13	14
Άνδρες	6	7	8	9	10

και μπορούμε να υποθέσουμε ότι οι κατανομές των X_1 και X_2 είναι κανονικές με ίδια διακύμανση, να ελεγχθεί η υπόθεση της ίσης παραγωγικότητας σε επίπεδο σημαντικότητας α = 0.10.

α. Υποθέσεις που γίνονται δεκτές: X_1, X_2 προέρχονται από κανονικές κατανομές με $\sigma_1=\sigma_2=\sigma$ και τα δύο δείγματα είναι τυχαία και ανεξάρτητα.
β. Η μηδενική και η εναλλακτική υπόθεση:

$$H_0: \ \mu_1=\mu_2$$
$$H_\varepsilon: \ \mu_1\neq\mu_2$$

γ. Το κριτήριο αποφάσεως: Για α=0.10, α/2=0.05 και $v=n_1+n_2-2=8$ έχουμε $t_{8,0.05}=1.86$. Επομένως αν

$$|t|=\frac{|\bar{x}_1-\bar{x}_2|}{s\sqrt{\frac{1}{5}+\frac{1}{5}}}>1.86$$

θα απορρίψουμε την H_0 υπέρ της H_ε.
δ. Η απόφαση: Από τα δεδομένα της δειγματοληψίας υπολογίζουμε:

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} = \frac{14+17+12+13+14}{5} = 14, \quad \bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2} = \frac{6+7+8+9+10}{5} = 8$$

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2}{n_1 - 1} = \frac{(14-14)^2 + (17-14)^2 + (12-14)^2 + (13-14)^2 + (14-14)^2}{4} = 3.5$$

$$s_2^2 = \frac{\sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2}{n_2 - 1} = \frac{(6-8)^2 + (7-8)^2 + (8-8)^2 + (9-8)^2 + (10-8)^2}{4} = 2.5$$

και

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{4(3.5)+4(2.5)}{8} = 3$$

Επειδή $s = \sqrt{3} = 1.73$ και

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{1.73\sqrt{\frac{1}{5}+\frac{1}{5}}} = \frac{|14-8|}{1.73\sqrt{\frac{1}{5}+\frac{1}{5}}} = 5.5 > 1.86$$

απορρίπτουμε την H_0 υπέρ της H_ε. Δηλαδή στο επίπεδο σημαντικότητας $\alpha = 0.10$ η υπόθεση της ίσης απόδοσης ανδρών και γυναικών απορρίπτεται.