22/4/2020 / 8o online μάθημα

ΕΛΕΓΧΟΣ Ho: $\sigma^2=\sigma_0^2$ (γνωστός - θετικός) για κανονικό πληθυσμό

$$X^2=\frac{(n-1)S^2}{\sigma_0^2}\overset{Ho}{\sim}X^2_{n-1}$$

Περιοχή απόρριψης

$X^2 > X^2_{a,n-1}$ για $Ha:\ \sigma^2>\sigma_0^2$

$X^2 < X^2_{1-a,n-1}$ για $Ha:\ \sigma^2<\sigma_0^2$

$$\left\{\begin{array}{l} X^2 < X^2_{1-a/2,n-1} \\ X^2 > X^2_{a/2,n-1} \end{array}\right. \quad \text{για } Ha:\ \sigma^2\neq\sigma_0^2$$

ΕΛΕΓΧΟΣ ΙΣΟΤΗΤΑΣ ΔΥΟ ΔΙΑΚΥΜΑΝΣΕΩΝ ΚΑΝΟΝΙΚΩΝ

$$\left.\begin{array}{l} X_1,\dots,X_n \quad \text{τ.δ.} \quad N(\mu_1,\sigma_1^2) \\ Y_1,\dots,Y_w \quad \text{τ.δ.} \quad N(\mu_2,\sigma_2^2) \end{array}\right\} \text{ανεξάρτητα}$$

$Ho:\ \sigma_1^2=\sigma_2^2$

$Ha:\ \sigma_1^2>\sigma_2^2$ ή $\sigma_1^2<\sigma_2^2$ ή $\sigma_1^2\neq\sigma_2^2$

$$\frac{(n-1)S_1^2}{\sigma_1^2}\sim X^2_{n-1}$$

$$\frac{(w-1)S_2^2}{\sigma_2^2}\sim X^2_{w-1}$$

$$F=\frac{\frac{S_1^2}{\sigma_1^2}(n-1)/(n-1)}{\frac{S_2^2}{\sigma_2^2}(w-1)/(w-1)}=\frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}$$

Παρατήρηση: Σε προηγούμενο μάθημα μιλήσαμε τόσο για την κατασκευή Δ.Ε για τη διαφορά των μέσων τιμών καν. πληθ. $\mu_1-\mu_2$ με ανεξ. δείγμ. όσο για τον έλεγχο της υπόθεσης της ισότητας δύο μέσων τιμών καν. πληθ. με ανεξ. δείγμ. υπό την προϋπόθεση υπόθεση ότι οι άγνωστες διακυμ. σ_1^2, σ_2^2 είναι ίσες. Δηλαδή υπό την υπόθεση ότι $\sigma_1^2=\sigma_2^2=\sigma^2$. Τα παραπάνω προϋποθέτουν αρχικά να ελεγχθεί η υπόθεση της ισότητας των διακυμ. Ο έλεγχος αυτής της υπόθ. ~~θα~~ γίνεται όπως θα δούμε με το F τεστ.

SOS Πριν κάνω το Δ.Ε για τη διαφορά $\mu_1-\mu_2$ ή τον έλεγχο της Ho $\mu_1=\mu_2$ πρέπει να κάνω τον έλεγχο $\sigma_1^2=\sigma_2^2$

Αν αυτό δεν απορριφθεί τότε προχωράω κανονικά στο ΔΕ και στον έλεγχο μέσων με όσα έμαθα στα προηγούμενα μαθήματα. Αν απορριφθεί αλλά δεν ... όλα έχω μάθει σε αυτό το μάθημα δεν μπορώ να κάνω ΔΕ μέσων και δεν μπορώ να κάνω τον έλεγχο μέσων.

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

Όμως όταν $H_0: \sigma_1^2 = \sigma_2^2$

τότε $$F = \frac{S_1^2}{S_2^2} \overset{H_0}{\sim} F_{n-1,m-1}$$

<u>Περιοχή απόρριψης</u>

$F \geq F_{\alpha,n-1,m-1}$ ή $F \leq F_{1-\alpha,n-1,m-1}$ ή $F \geq F_{\alpha/2,n-1,m-1}$

$F \leq F_{1-\alpha/2,n-1,m-1}$

Υπενθύμιση: $$F_{1-\alpha,n_1,n_2} = \frac{1}{F_{\alpha,n_2,n_1}}$$

Τέλος: έλεγχος διωνυμικών p (ανεξάρτητα)

$H_0: p_1 = p_2$

$H_a: p_1 > p_2$ ή $p_1 < p_2$ ή $p_1 \neq p_2$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \underset{\text{προσ.}}{\overset{H_0}{\sim}} N(0,1)$$

όπου $$p_1 = \frac{X_1}{n_1} = \frac{\text{πλήθος επιτυχιών σε } n_1}{n_1}$$

$$p_2 = \frac{X_2}{n_2} = \frac{\text{πλήθος επιτυχιών σε } n_2}{n_2}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Θα εξηγηθεί με ένα παράδειγμα.

Ποιες οι περιοχές απόρριψης;

Γραμμική Παλινδρόμηση - Συσχέτιση

Συχνά δύο ή περισσότερες ποσοτικές μεταβλητές εξετάζονται μαζί με σκοπό να εξετασθεί η ύπαρξη σχέσης μεταξύ τους και να προσδιοριστεί η μαθηματική έκφραση της σχέσης αυτής.
Στο πλαίσιο αυτό συχνά λέμε ότι θέλουμε να προβλέψουμε την τιμή μιας μεταβλητής από τις τιμές των άλλων. Η στατιστική μεθοδολογία που χρησιμοποιείται με αυτόν τον σκοπό καλείται ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.
Ο όρος χρησιμοποιήθηκε πρώτη φορά από τον Galton το 1900.
Στα πλαίσια αυτού του μαθήματος θα περιοριστούμε στην απλή περίπτωση δύο μεταβλητών (ποσοτικών)

Ειδικότερα, έστω X και Y δύο τ.μ. για τις οποίες θέλουμε από τις διαθέσιμες τιμές τους $(x_1, y_1), \ldots, (x_n, y_n)$ να δημιουργήσουμε ένα μοντέλο πρόβλεψης των τιμών της Y (καλείται εξαρτημένη) από τις τιμές της X (καλείται ανεξάρτητη). Όπως είναι λογικό οι τ.μ. X, Y σπάνια συνδέονται με μια συναρτησιακή σχέση δηλαδή με μια τέλεια σχέση και ο στατιστικός θέλει να την προσδιορίσει - προσεγγίσει. Όταν οι τ.μ. συνδέονται γραμμικά τότε κάνουμε λόγο για το μοντέλο της απλής γραμμικής παλινδρόμησης.
Το μοντέλο αυτό χρησιμοποιείται αν διαπιστωθεί με κάποιο τρόπο ότι συνδέονται οι X, Y γραμμικά και θέλουμε έπειτα να προσδιορίσουμε τη γραμμική σχέση.
Επομένως αρχικά πριν χρησιμοποιήσουμε το μοντέλο της ΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ θα πρέπει να βρούμε, ψάξουμε τρόπους διερεύνησης της ύπαρξης γραμμικής σχέσης μεταξύ των X, Y στη βάση η το πλήθος δεδομένων (x_i, y_i) $i=1, \ldots, n$.

$1^{ος}$ ΤΡΟΠΟΣ : ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ

Έστω X, Y οι προς μελέτη τ.μ. και έστω τ.δ. $X_1, \ldots, X_n$ το οποίο δημιουργήθηκε από την μέτρηση της μεταβλητής X σε n πραγματικές μονάδες και $Y_1, \ldots, Y_n$ το τυχαίο δείγμα που δημιουργείται από τη μέτρηση της Y στις ίδιες πειραματικές μονάδες.

Κατασκευή διαγράμματος διασποράς

ΠΑΡΑΔΕΙΓΜΑΤΑ (προβολή)