ΣΥΣΧΕΤΙΣΗ & ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Γραμμική Συσχέτιση

$$\boldsymbol{r} = \frac{\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}})(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})}{\sqrt{[\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}})^2][\sum_{i=1}^{n}(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})^2]}} \text{ ή}$$

$$\boldsymbol{r} = \frac{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{X_i}\boldsymbol{Y_i} - \sum_{i=1}^{n}\boldsymbol{X_i}\sum_{i=1}^{n}\boldsymbol{Y_i}}{\sqrt{\left[\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{X_i}^2 - \left(\sum_{i=1}^{n}\boldsymbol{X_i}\right)^2\right]\left[\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{Y_i}^2 - \left(\sum_{i=1}^{n}\boldsymbol{Y_i}\right)^2\right]}}, \qquad -1 \le r \le 1$$

Έλεγχος Στατιστικής Σημαντικότητας Συντελεστή Συσχέτισης ρ:

Παράμετρος Πληθυσμού	**Ελεγχοσυνάρτηση**	**Υποθέσεις**	**Περιοχή Απόρριψης H_0**
Συντελεστής Συσχέτισης ($\boldsymbol{\rho}$)	$T = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim t_{n-2}$	$H_0: \rho = 0$ vs $H_1: \rho \neq 0$	$\lvert T \rvert \ge t_{n-2,\alpha/2}$

Απλή Γραμμική Παλινδρόμηση

<u>Γραμμή Παλινδρόμησης Πληθυσμού</u>: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ όπου $\varepsilon_i \sim N.i.d.(0, \sigma^2)$

<u>Ευθεία ΕΤ</u>: $\hat{Y}_i = b_0 + b_1 X_i$, Κατάλοιπα: $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$

$$\boldsymbol{b_1} = \frac{\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}})(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})}{\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}})^2} \text{ ή } \boldsymbol{b_1} = \frac{\sum_{i=1}^{n}\boldsymbol{X_i}\boldsymbol{Y_i} - \boldsymbol{n}\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Y}}}{\sum_{i=1}^{n}\boldsymbol{X_i}^2 - \boldsymbol{n}\overline{\boldsymbol{X}}^2} \text{ ή } \boldsymbol{b_1} = \frac{\boldsymbol{\Sigma_{xy}}}{\boldsymbol{\Sigma_{x^2}}}$$

$$\boldsymbol{b_0} = \overline{\boldsymbol{Y}} - \boldsymbol{b_1}\overline{\boldsymbol{X}}$$

όπου:

$$x = X - \overline{X} \text{ και } y = Y - \overline{Y}$$

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

Κάποια χρήσιμα αναπτύγματα είναι τα εξής:

- Σ_{x^2} ή $\Sigma_{xx} = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} {X_i}^2 - n\overline{X}^2$

- Σ_{y^2} ή $\Sigma_{yy} = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} {Y_i}^2 - n\overline{Y}^2$

- $\Sigma_{xy} = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}$

$$\sum_{i=1}^{n}\left(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}}\right)^2 = \sum_{i=1}^{n}\left(\widehat{\boldsymbol{Y}}_i - \overline{\boldsymbol{Y}}\right)^2 + \sum_{i=1}^{n}\left(\boldsymbol{Y_i} - \widehat{\boldsymbol{Y}}_i\right)^2 \quad \text{ή} \quad \boldsymbol{TSS = RSS + ESS}$$

όπου:

$\boldsymbol{TSS}$ = Συνολικό Άθροισμα Τετραγώνων ($Total\ Sum\ of\ Squares$)

$\boldsymbol{RSS}$ = Άθροισμα των Τετραγώνων της Παλινδρόμησης ($Regression\ Sum\ of\ Squares$)

$\boldsymbol{ESS}$ = Άθροισμα των Τετραγώνων των Καταλοίπων ($Error\ Sum\ of\ Squares$)

$$\boldsymbol{RSS = b_1 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)} \quad \text{ή} \quad \boldsymbol{RSS = b_1^2 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

Συντελεστής Προσδιορισμού

$$\boldsymbol{R^2 = \frac{RSS}{TSS}} \quad \text{ή} \quad \boldsymbol{R^2 = 1 - \frac{ESS}{TSS}}, \qquad 0 \le R^2 \le 1$$

Εκτιμητής της Διακύμανσης των Καταλοίπων

$$\boldsymbol{\widehat{\sigma}^2 = \frac{ESS}{n-2}}$$

Η εκτίμηση του ΔΕ δίνεται από:

$$\boldsymbol{\frac{(n-2)\widehat{\sigma}^2}{x^2_{n-2,a/2}} \le \sigma^2 \le \frac{(n-2)\widehat{\sigma}^2}{x^2_{n-2,1-a/2}}}$$

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

ΔΕ και ΕΥ για $\boldsymbol{\beta_0}$ και $\boldsymbol{\beta_1}$

Παράμετρος Πληθυσμού	$\mathbf{100(1-\alpha)\%}$ **Δ. Ε.**
Συντελεστής Σταθεράς $(\boldsymbol{\beta_0})$	$b_0 \pm t_{n-2,a/2}\, Se(b_0)$ $Se(b_0) = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)}$
Συντελεστής Κλίσης $(\boldsymbol{\beta_1})$	$b_1 \pm t_{n-2,a/2}\, Se(b_1)$ $Se(b_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$

Παράμετρος Πληθυσμού	Ελεγχοσυνάρτηση	Υποθέσεις	Περιοχή Απόρριψης $\boldsymbol{H_0}$
Συντελεστής Σταθεράς $(\boldsymbol{\beta_0})$	$t_{b_0} = \frac{b_0 - k}{Se(b_0)} \sim t_{n-2}$	$H_0: \beta_0 = k$ vs $H_1: \beta_0 \neq k$	$\lvert t_{b_0} \rvert \geq t_{n-2,\alpha/2}$
		$H_0: \beta_0 = k$ vs $H_1: \beta_0 > k$	$t_{b_0} \geq t_{n-2,\alpha}$
		$H_0: \beta_0 = k$ vs $H_1: \beta_0 < k$	$t_{b_0} \leq -t_{n-2,\alpha}$
Συντελεστής Κλίσης $(\boldsymbol{\beta_1})$	$t_{b_1} = \frac{b_1 - k}{Se(b_1)} \sim t_{n-2}$	$H_0: \beta_1 = k$ vs $H_1: \beta_1 \neq k$	$\lvert t_{b_1} \rvert \geq t_{n-2,\alpha/2}$
		$H_0: \beta_1 = k$ vs $H_1: \beta_1 > k$	$t_{b_1} \geq t_{n-2,\alpha}$
		$H_0: \beta_1 = k$ vs $H_1: \beta_1 < k$	$t_{b_1} \leq -t_{n-2,\alpha}$

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

<u>Σημείωση</u>: Αν πρέπει να ελεγχθεί η **στατιστική σημαντικότητα** της παραμέτρου β_i θα διενεργηθεί ο έλεγχος της αρχικής υπόθεσης $H_0: \beta_i = 0$, έναντι της εναλλακτικής $H_1: \beta_i \neq 0$ (δηλαδή $k = 0$). Εάν η H_0 απορριφθεί, η παράμετρος β_i είναι στατιστικά σημαντική, ειδάλλως όχι.

F – Έλεγχος Στατιστικής Σημαντικότητας Παλινδρόμησης

Πίνακας ANOVA

Πηγή Μεταβλητότητας	**Αθροίσματα Τετραγώνων**	**Βαθμοί Ελευθερίας**	**Μέσα Τετράγωνα**	$\boldsymbol{F}$
Παλινδρόμηση	RSS	1	RSS	$\boldsymbol{F = \frac{RSS}{ESS/(n-2)}}$
Κατάλοιπα	ESS	$n-2$	$ESS/(n-2)$	
Σύνολο:	TSS	$n-1$		

Ο Έλεγχος Στατιστικής Σημαντικότητας της Παλινδρόμησης εκτελείται ως:

Παράμετρος Πληθυσμού	**Ελεγχοσυνάρτηση**	**Υποθέσεις**	**Περιοχή Απόρριψης H_0**
Συντελεστής Προσδιορισμού (R^2)	$F = \frac{RSS}{ESS/(n-2)} \sim F_{1,n-2}$ ή $F = \frac{R^2}{(1-R^2)/(n-2)} \sim F_{1,n-2}$	$H_0: R^2 = 0$ vs $H_1: R^2 > 0$	$F \geq F_{1,n-2,\alpha}$

<u>Σημείωση</u>: Στην περίπτωση της απλής γραμμικής παλινδρόμησης ο έλεγχος F που πραγματοποιείται για να ελεγχθεί η στατιστική σημαντικότητα της παλινδρόμησης είναι ισοδύναμος με τον έλεγχο $t\ student$ που πραγματοποιείται για τον έλεγχο στατιστικής σημαντικότητας του συντελεστή κλίσης β_1.

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

Προβλέψεις

Παράμετρος Πληθυσμού	$100(1-\alpha)\%$ Δ. Ε.
Μέση Πρόβλεψη $E(Y\|X=x_0)$	$\hat{Y}_0 \pm t_{n-2,a/2}\, Se(\hat{Y}_0)$ $Se(\hat{Y}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right]}$
Ατομική Πρόβλεψη (Y_0)	$\hat{Y}_0 \pm t_{n-2,a/2}\, Se(\hat{Y}_0)$ $Se(\hat{Y}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right]+1}$

Παράμετρος Πληθυσμού	Ελεγχοσυνάρτηση	Υποθέσεις	Περιοχή Απόρριψης H_0
Μέση Πρόβλεψη $E(Y\|X=x_0)$	$t_{\hat{Y}_0} = \frac{\hat{Y}_0 - k}{Se(\hat{Y}_0)} \sim t_{n-2}$	$H_0: E(Y\|X=x_0)=k$ vs $H_1: E(Y\|X=x_0)\neq k$	$\|t_{\hat{Y}_0}\| \geq t_{n-2,\alpha/2}$
		$H_0: E(Y\|X=x_0)=k$ vs $H_1: E(Y\|X=x_0)> k$	$t_{\hat{Y}_0} \geq t_{n-2,\alpha}$
		$H_0: E(Y\|X=x_0)=k$ vs $H_1: E(Y\|X=x_0)< k$	$t_{\hat{Y}_0} \leq -t_{n-2,\alpha}$
Ατομική Πρόβλεψη (Y_0)	$t_{\hat{Y}_0} = \frac{\hat{Y}_0 - k}{Se(\hat{Y}_0)} \sim t_{n-2}$	$H_0: Y_0=k$ vs $H_1: Y_0\neq k$	$\|t_{\hat{Y}_0}\| \geq t_{n-2,\alpha/2}$
		$H_0: Y_0=k$ vs $H_1: Y_0> k$	$t_{\hat{Y}_0} \geq t_{n-2,\alpha}$
		$H_0: Y_0=k$ vs $H_1: Y_0< k$	$t_{\hat{Y}_0} \leq -t_{n-2,\alpha}$

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ