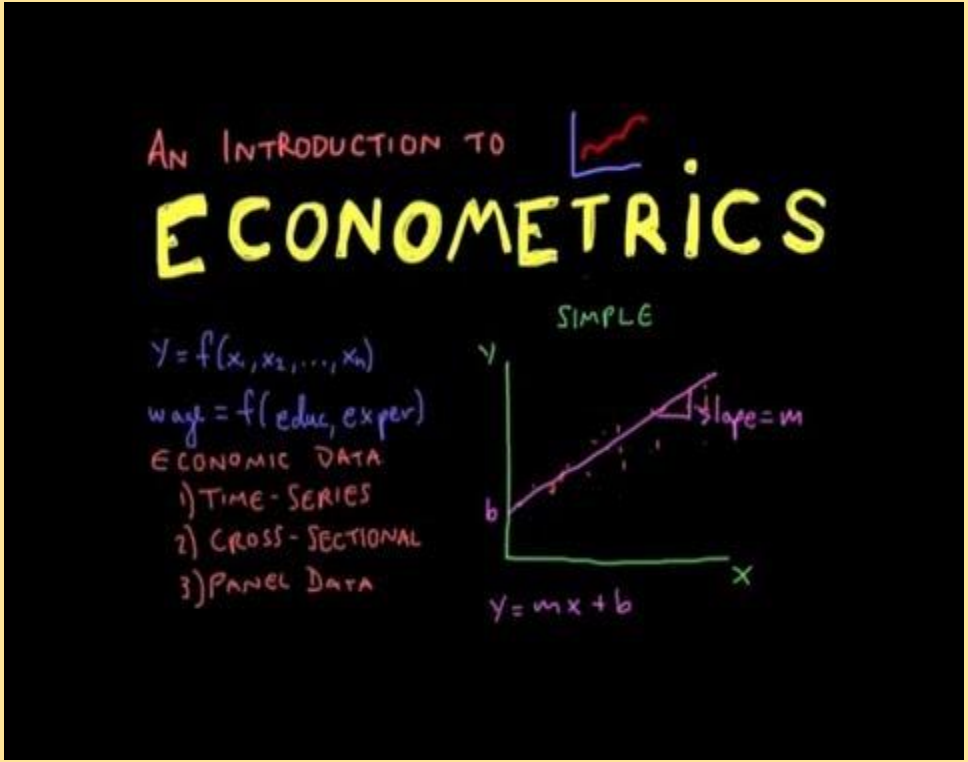
Οικονομικό Πανεπιστήμιο Αθηνών
Μάθημα: “*Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα*”
Συνάντηση: 5η



Δρ. Τρύφωνας Λεμοντζόγλου

Ξεκινήσαμε με τις φόρμουλες υπολογισμού των συντελεστών της παλινδρόμησης για το **απλό γραμμικό υπόδειγμα**, δηλ. για την περίπτωση που το μοντέλο μας έχει **1 μόνο ανεξάρτητη μεταβλητή** (**k=1**).

$$\Psi_i = \alpha + \beta * X_i + \varepsilon_i$$

*τα α και β **ΔΕΝ** μπορούν να υπολογιστούν*

όπου
***Ψ**: οι τιμές της εξαρτημένης μεταβλητής*
***X**: οι τιμές της ανεξάρτητης μεταβλητής*
***α**: ο (άγνωστος) σταθερός όρος της εξίσωσης*
***β**: ο (άγνωστος) συντελεστής διεύθυνσης της εξίσωσης*
***ε**: οι τιμές του διαταρακτικού όρου (όλοι οι άλλοι "τυχαίοι παράγοντες" που αγνοεί το μοντέλο μας)*
***i**: ο αριθμός των παρατηρήσεων*

Χρησιμοποιήσαμε τη **Μέθοδο των Ελαχίστων Τετραγώνων (OLS)** για να υπολογίσουμε τους **συντελεστές της παλινδρόμησης** (δηλ. *του εκτιμητές των τιμών του α και του β*)

$$\widehat{\Psi}_i = \hat{a} + \hat{\beta} * X_i$$ → *γραμμή παλινδρόμησης*

όπου

$\widehat{\Psi}_i$: *οι* ***εκτιμώμενες*** *τιμές του Ψ*

X_i : *οι τιμές της ανεξάρτητης μεταβλητής*

$\hat{a}$: *η* ***εκτίμηση*** *για τον πραγματικό όρο α*

$\hat{\beta}$: *η* ***εκτίμηση*** *για τον πραγματικό όρο β*

Δρ. Τρύφωνας

Οι **φόρμουλες υπολογισμού** που προέκυψαν ήταν οι εξής:

$$\hat{\beta} = \frac{n\Sigma\Psi_i X_i - \Sigma X_i \Sigma \Psi_i}{n\Sigma X_i^2 - (\Sigma X_i)^2}$$

$$\hat{\alpha} = \overline{\psi_i} - \hat{\beta}\,\overline{X_i}$$

→ *συντελεστές παλινδρόμησης*

Έπειτα, δώσαμε την **ερμηνεία των συντελεστών** της παλινδρόμησης στο απλό γραμμικό μοντέλο (k=1), ενώ μιλήσαμε και για τη σημασία του **συντελεστή προσδιορισμού** (βλ. **R-squared**).

Για την περίπτωση του απλού γραμμικού υποδείγματος είπαμε πως ισχύει: **r^2 = R^2** (*δηλ το τετράγωνο του συντελεστή συσχέτισης ισούται με τον συντελεστή προσδιορισμού*)

Τι θα συμβεί τώρα αν προσθέσουμε στο μοντέλο μας **μια ακόμη ανεξάρτητη μεταβλητή**, δηλ. **έναν νέο πιθανό προσδιοριστικό παράγοντα** (**k=2**) ; ; ;

$$Y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \; i = 1,2,\ldots,n$$

Όπως και πριν, οι πραγματικοί συντελεστές α , β1 και β2 ΔΕΝ μπορούν να υπολογιστούν. Ωστόσο, μπορούμε να υπολογίσουμε τους εκτιμητές των α , β1 και β2 (βλ. συντελεστές παλινδρόμησης).

Δες πίσω τις νέες φόρμουλες για τον υπολογισμό των συντελεστών της παλινδρόμησης στο υπόδειγμα με **k=2** ανεξάρτητες μεταβλητές →

$$\hat{\beta}_1 = \frac{\sum_1^n (X_{1i}-\overline{X_1})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{2i}-\overline{X_2})^2 - \sum_1^n (X_{2i}-\overline{X_2})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})}{\sum_1^n (X_{1i}-\overline{X_1})^2 * \sum_1^n (X_{2i}-\overline{X_2})^2 - [\sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})]^2}$$

$$\hat{\beta}_2 = \frac{\sum_1^n (X_{2i}-\overline{X_2})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})^2 - \sum_1^n (X_{1i}-\overline{X_1})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})}{\sum_1^n (X_{1i}-\overline{X_1})^2 * \sum_1^n (X_{2i}-\overline{X_2})^2 - [\sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})]^2}$$

$$\hat{\alpha} = \hat{\beta}_0 = \overline{\Psi} - \hat{\beta}_1 * \overline{X_1} - \hat{\beta}_2 * \overline{X_2}$$

$$\hat{\beta}_1 = \frac{S_{X1\,\psi} * S_{X2\,X2} - S_{X2\,\psi} * S_{X1\,X2}}{S_{X1\,X1} * S_{X2\,X2} - (S_{X1\,X2})^2}$$

$$\hat{\beta}_2 = \frac{S_{X2\,\psi} * S_{X1\,X1} - S_{X1\,\psi} * S_{X1\,X2}}{S_{X1\,X1} * S_{X2\,X2} - (S_{X1\,X2})^2}$$

		country	tot_educ	mil_exp	BOM	fem_emp
1	2017	Austria	11	0,8	42	52
2	2017	Belgium	12	0,9	40	45
3	2017	Denmark	15	1,1	54	54
4	2017	Finland	12	1,4	45	51
5	2017	France	10	1,9	60	46
6	2017	Germany	11	1,2	35	53
7	2017	Ireland	13	0,3	38	52
8	2017	Luxembourg	9	0,6	41	51
9	2017	Netherlands	12	1,2	51	55
10	2017	Norway	16	1,6	56	58
11	2017	Sweden	16	1,0	55	57
12	2017	United Kingdom	14	1,8	48	55

πραγματικά δεδομένα: πηγή Παγκόσμια Τράπεζα

ΠΑΡΑΔΕΙΓΜΑ υπολογισμού των συντελεστων της παλινδρόμησης στο μοντέλο με **k=2** ανεξάρτητες μεταβλητές

Ψ: εξαρτημένη ("*γυναικεία εργασία*")

X1: ανεξάρτητη ("*κρατικές δαπάνες για την εκπαίδευση*")

X2: ανεξάρτητη ("*στρατιωτικές δαπάνες*")

Δρ. Τρύφωνας

τιμές ανεξάρτητων

τιμές εξαρτημένης

X1	X2		Ψ
tot_educ	mil_exp	BOM	fem_emp
11	0,8	42	52
12	0,9	40	45
15	1,1	54	54
12	1,4	45	51
10	1,9	60	46
11	1,2	35	53
13	0,3	38	52
9	0,6	41	51
12	1,2	51	55
16	1,6	56	58
16	1	55	57
14	1,8	48	55
$\overline{X_1}$ 12,58	$\overline{X_2}$ 1,15		$\overline{\Psi}$ 52,42

(X1-X1μ)*(Ψ-Ψμ)	(X2-X2μ)^2	(X2-X2μ)*(Ψ-Ψμ)	(X1-X1μ)*(X2-X2μ)
0,66	0,12	0,15	0,55
4,30	0,06	1,86	0,15
3,82	0,00	-0,08	-0,12
0,82	0,06	-0,36	-0,15
16,56	0,56	-4,82	-1,94
-0,92	0,00	0,03	-0,08
-0,18	0,72	0,36	-0,36
5,08	0,30	0,78	1,97
-1,50	0,00	0,13	-0,03
19,08	0,20	2,51	1,54
15,66	0,02	-0,69	-0,51
3,66	0,42	1,68	0,92
67,08	**2,49**	**1,55**	**1,95**
			164,01

(X1-X1μ)^2	(X2-X2μ)^2	(X1-X1μ)*(X2-X2μ)
2,50	0,12	0,55
0,34	0,06	0,15
5,86	0,00	-0,12
0,34	0,06	-0,15
6,66	0,56	-1,94
2,50	0,00	-0,08
0,18	0,72	-0,36
12,82	0,30	1,97
0,34	0,00	-0,03
11,70	0,20	1,54
11,70	0,02	-0,51
2,02	0,42	0,92
56,92	**2,49**	**1,95**
		137,92

$\widehat{\beta_1} =$ **1,189**

(X2-X2μ)*(Ψ-Ψμ)	(X1-X1μ)^2	(X1-X1μ)*(Ψ-Ψμ)	(X1-X1μ)*(X2-X2μ)
0,15	2,50	0,66	0,55
1,86	0,34	4,30	0,15
-0,08	5,86	3,82	-0,12
-0,36	0,34	0,82	-0,15
-4,82	6,66	16,56	-1,94
0,03	2,50	-0,92	-0,08
0,36	0,18	-0,18	-0,36
0,78	12,82	5,08	1,97
0,13	0,34	-1,50	-0,03
2,51	11,70	19,08	1,54
-0,69	11,70	15,66	-0,51
1,68	2,02	3,66	0,92
1,55	**56,92**	**67,08**	**1,95**
			-42,59

(X1-X1μ)^2	(X2-X2μ)^2	(X1-X1μ)*(X2-X2μ)
2,50	0,12	0,55
0,34	0,06	0,15
5,86	0,00	-0,12
0,34	0,06	-0,15
6,66	0,56	-1,94
2,50	0,00	-0,08
0,18	0,72	-0,36
12,82	0,30	1,97
0,34	0,00	-0,03
11,70	0,20	1,54
11,70	0,02	-0,51
2,02	0,42	0,92
56,92	**2,49**	**1,95**
		137,92
	$\widehat{\beta}_2$	**-0,309**

$$\widehat{\beta_1} = \frac{\sum_1^n (X_{1i}-\overline{X_1})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{2i}-\overline{X_2})^2 - \sum_1^n (X_{2i}-\overline{X_2})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})}{\sum_1^n (X_{1i}-\overline{X_1})^2 * \sum_1^n (X_{2i}-\overline{X_2})^2 - [\sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})]^2}$$

=1,189

$$\widehat{\beta_2} = \frac{\sum_1^n (X_{2i}-\overline{X_2})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})^2 - \sum_1^n (X_{1i}-\overline{X_1})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})}{\sum_1^n (X_{1i}-\overline{X_1})^2 * \sum_1^n (X_{2i}-\overline{X_2})^2 - [\sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})]^2}$$

= - 0,309

$$\widehat{\alpha} = \widehat{\beta_0} = \overline{\Psi} - \widehat{\beta_1} * \overline{X_1} - \widehat{\beta_2} * \overline{X_2}$$

=52,4167- (1,1892*12,5888) + (0,3088*1,15)

=37,4460+0,35512=37,80

$$\widehat{\beta_1} = \frac{\sum_1^n (X_{1i}-\overline{X_1})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{2i}-\overline{X_2})^2 - \sum_1^n (X_{2i}-\overline{X_2})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})}{\sum_1^n (X_{1i}-\overline{X_1})^2 * \sum_1^n (X_{2i}-\overline{X_2})^2 - [\sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})]^2}$$

=1,189 *θετική επίδραση*

$$\widehat{\beta_2} = \frac{\sum_1^n (X_{2i}-\overline{X_2})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})^2 - \sum_1^n (X_{1i}-\overline{X_1})(\Psi_i-\overline{\Psi}) * \sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})}{\sum_1^n (X_{1i}-\overline{X_1})^2 * \sum_1^n (X_{2i}-\overline{X_2})^2 - [\sum_1^n (X_{1i}-\overline{X_1})(X_{2i}-\overline{X_2})]^2}$$

= - 0,309 *αρνητική επίδραση*

Ερμηνεία συντελεστων: *Αν αυξηθούν οι κρατικές δαπάνες για την εκπαίδευση κατά 1 μονάδα (δηλ κατά 1% ως ποσοστό επί του ΑΕΠ) τότε θα αυξηθεί το ποσοστό των εργαζόμενων γυναικών κατά **1,189 μονάδες** (δηλ κατά 1,189%)*
*Αν αυξηθούν οι στρατιωτικές δαπάνες κατά 1 μονάδα (δηλ κατά 1% ως ποσοστό επί του ΑΕΠ) τότε θα μειωθεί το ποσοστό των εργαζόμενων γυναικών κατά **0,309 μονάδες** (δηλ κατά 0,309%)*

$$\hat{\alpha} = \hat{\beta}_0 = \overline{\Psi} - \hat{\beta}_1 * \overline{X}_1 - \hat{\beta}_2 * \overline{X}_2$$

= 37,80

Ερμηνεία συντελεστων: *Ακόμη και αν ένα κράτος δεν δαπανήσει τίποτα για την εκπαίδευση και τις στρατιωτικές δαπάνες (δηλ για X1=X2=0) θα υπάρχει ένα επίπεδο γυναικείας εργασίας κοντά στο 38%*

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{\Psi}_i - \overline{\psi})^2}{\sum_{i=1}^{n} (\Psi_i - \overline{\Psi})^2} = 1 - \frac{\sum_{i=1}^{n} (\Psi_i - \hat{\Psi}_i)^2}{\sum_{i=1}^{n} (\Psi_i - \overline{\Psi})^2}$$

= 0,46 (46%)

Ερμηνεία: *Περίπου το 46% της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής μπορεί να εξηγηθεί από την κίνηση των τιμών των ανεξάρτητων μεταβλητών χ1 και χ2*

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{\Psi}_i - \bar{\psi})^2}{\sum_{i=1}^{n}(\Psi_i - \bar{\Psi})^2} = 1 - \frac{\sum_{i=1}^{n}(\Psi_i - \hat{\Psi}_i)^2}{\sum_{i=1}^{n}(\Psi_i - \bar{\Psi})^2}$$

= 0,46 (46%)

Κάθε φορά που προσθέτουμε στο μοντέλο μια νέα ανεξάρτητη μεταβλητή **η τιμή του συντελεστή προσδιορισμού αυξάνει**, δίχως να λαμβάνει υπόψη την στατιστική σημαντικότητα του παράγοντα που προσθέτουμε.

Για το λόγο αυτό, συχνά κοιτάμε την τιμή του **διορθωμένου συντελεστή προσδιορισμού**:

$$Adjusted\ R^2 = 1 - \frac{(n-1)}{(n-k-1)} * (1 - R^2)$$

= 1-(11/9)*(1-0,46)=0,34

εκτιμώμενες τιμές του Ψ

σφάλματα

*Include	✔	✘			
Transform					
Groups	X1	X2	Y	Pred Y	Residual
Data	11 12 15 12 10 11 13 9 12 16 16 14	0.8 0.9 1.1 1.4 1.9 1.2 0.3 0.6 1.2 1.6 1 1.8	52 45 54 51 46 53 52 51 55 58 57 55	50.550512 51.729136 55.265007 51.729136 49.371889 50.550512 52.90776 48.193265 51.729136 56.443631 56.443631 54.086384	1.449488 -6.729136 -1.265007 -0.729136 -3.371889 2.449488 -0.90776 2.806735 3.270864 1.556369 0.556369 0.913616
P-value:	0.0141318				
Average:	12.583333		52.416667	52.416667	0
n:	12		12	12	12
S:	2.274696		3.918681	2.681011	2.858014
Skewness:	0.231846		-0.668978	0.231846	-1.246053
Normality:**	0.5668		0.4284	0.5668	0.1695
Outliers:					-6.729136163982446

Συντελεστές Συσχέτισης

Correlation matrix (pearson)

	Y	X_1	X_2
Y	1	0.684162	0.0755781
X_1	0.684162	1	0.1638
X_2	0.0755781	0.1638	1

Ο πίνακας με τις τιμές του συντελεστή γραμμικής συσχέτισης.

Παρατηρείτε κάτι ; ; ;

συντελεστής προσδιορισμού

$$R\text{-Squared} = \frac{SS_{regression}}{SS_{total}}$$

Αθροίσματα Τετραγώνων

ANOVA table

Source	DF	Sum of Square	Mean Square	F Statistic	P-value
Regression (between $\hat{y}_i$ and $\bar{y}$)	1	79.066008 → *SSR*	79.066008	8.799714	0.0141318
Residual (between y_i and $\hat{y}_i$)	10	89.850659 → *SSE*	8.985066		
Total (between y_i and $\bar{y}$)	11	168.916667 → *SST*	15.356061		

Έλεγχος Από Κοινού Σημαντικότητας των Συντελεστών της Παλινδρόμησης

Αθροίσματα Τετραγώνων

ANOVA table

Source		DF	Sum of Square	Mean Square	F Statistic	P-value
Regression (between $\hat{y}_i$ and $\bar{y}$)	*k-1*	1	79.066008 → *SSR*	79.066008	8.799714	0.0141318
Residual (between y_i and $\hat{y}_i$)	*n-k*	10	89.850659 → *SSE*	8.985066		
Total (between y_i and $\bar{y}$)	*n-1*	11	168.916667 → *SST*	15.356061		

Ho: $\beta 1 = \beta 2 = 0$

H1: ένα τουλάχιστον από τα β είναι στατιστικά σημαντικό

Δρ. Τρύφωνας

Έλεγχος Στατιστικής Σημαντικότητας των Συντελεστών της Παλινδρόμησης

Ho: β=0
H1: β είναι στατιστικά σημαντικό

Συντελεστές Παλινδρόμησης

	Coeff	SE	t-stat	lower $t_{0.025}(9)$	upper $t_{0.975}(9)$	Stand Coeff	p-value	VIF
b	37.807654	5.536749	6.828493	25.282658	50.332651	0	0.0000765594	
X1	1.189204	0.424	2.804724	0.230048	2.148359	0.690303	0.0205534	1.02757
X2	-0.308814	2.027152	-0.152339	-4.89455	4.276922	-0.0374939	0.88228	1.02757

Κανόνας με τη χρήση του p-value: Αν p-value<0,05 (5%) τότε απορρίπτω την Ho → δηλ ο συντελεστής είναι στατιστικά σημαντικός