

## סטטיסטיקה 3- שיעור 10

בדיקת תוספת ותרומה ייחודית במונחי  $R^2$

ברגרסיה רבת משתנים אנו מוסיפים משתנים בלתי תלויים למודל על מנת להסביר טוב יותר את המשתנה התלוי.

אנו נתעניין בפרופורציית השונות המוסברת  $R^2$  שהוסיפו משתנים אלו למשתנים שכבר היו במשוואה ונרצה לבדוק האם התוספת הזו מובהקת.

דוג':

$$N=100, r^2_{Y1}=0.1, R^2_{YX_1X_2}=0.3, R^2_{YX_1X_2X_3}=0.4, R^2_{YX_1X_2X_3X_4}=0.6, R^2_{YX_1X_2X_3X_4X_5}=0.7$$

תוספת במונחי  $R^2$

מה התוספת של  $2X$  על  $3X$  על  $1X$ ?

$$R^2_{YX_1X_2X_3} - r^2_{Y1} = 0.4 - 0.1 = 0.3$$

כלומר תוספת של 30% לפרופורציית השונות המוסברת.

מהי התוספת של  $5X$  על כל השאר?

$$R^2_{YX_1X_2X_3X_4X_5} - R^2_{YX_1X_2X_3X_4} = 0.7 - 0.6 = 0.1$$

מובהקות התוספת:

תוספת של  $2X$  על  $3X$  על  $1X$ :

הנחות:

1. דגימה מקרית ותצפיות בלתי-תלויות:  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  עבור כל  $i \neq j$ .

2.  $\varepsilon_i \sim N(0, \sigma^2 e)$  עבור כל קומבינציה קבועה של האים.

3. אין מתאם מלא בין המשתנים הבלתי תלויים.

השערות:

$H_0 : \beta_2 = \beta_3 = 0$  כאשר המשתנה  $x_1$  נמצא במשוואה.

אחרת  $H_1 :$

חישוב סטטיסטי:

$$F = \frac{(R_{all}^2 - R_{part}^2) / (K_{all} - K_{part})}{(1 - R_{all}^2) / (n - K_{all} - 1)} \stackrel{H_0}{\sim} F_{k_{all} - k_{part}, n - k_{all} - 1}$$

לפי הדוג' למעלה: .

$$F = \frac{(0.4 - 0.1) / (3.1)}{(1 - 0.4) / (100 - 3 - 1)} = 24$$

כלל הכרעה: נדחה  $H_0$  כאשר:  $F > F(k_{all} - k_{part}, n - k_{all} - 1)(0.05)$

בדוג':  $F > F_{2,96}(0.05)$

מסקנה: נדחה את  $H_0$  ברמת מובהקות 0.05, כלומר התוספת של 2X 3X על 1X מובהקת באוכלוסייה.

תרומה ייחודית במונחי  $R^2$

כאשר מדובר בתוספת של משתנה, כאשר כל השאר בפנים (1 על השאר) זהו מקרה פרטי של תוספת שנקרא תרומה ייחודית של המשתנה.

בדוג': התוספת של 5X על כל השאר (4 X 1) היא התרומה הייחודית שלו.

תרומה ייחודית אפשר לבדוק או באמצעות מבחן F למובהקות התוספת או באמצעות מבחן t למובהקות המקדם החלקי. (שיפוע  $\beta$ ) כי כאשר יש דרגות חופש 1 במונה של F, מתקיים קשר בין התפלגות F להתפלגות t.

נבדקות אותן השערות ולכן גם ה sig והמסקנה זהה.

נוכל לבדוק מובהקות המקדם החלקי באמצעות מבחן t.

מבחן t

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{S.E.(\hat{\beta}_j)} \stackrel{H_0}{\sim} t_{n-k-1}$$

## שיטות להרצת רגרסיה:

1. ENTER – החוקר מכניס את כל המשתנים הבלתי תלויים בשלב אחד.
2. היררכית- החוקר מחליט להכניס את הנתונים בסדר מסויים, ומחליט אילו יכנסו בכל שלב, לכן, יכול להיות שיכנסו משתנים ב"ת שאינם מובהקים. כמו כן, ייתכן ש  $R^2$  יגדל, אבל לא בטוח שהתוספת של המשתנים שהוכנסו מובהקת. חוקר המריץ בשיטה היררכית, כאשר על סמך ידע מסויים או תיאוריה הוא רוצה לבדוק משתנים מסויים (שלא יכנסו למשוואה או למודל בשיטת רגרסיה בצעדים).
3. רגרסיה בצעדים- המחשב מתחיל בפיתוח מודל רגרסיה עם מנבא יחיד שהוא בעל המתאם הגבוה ביותר עם Y ומובהק. בשלב השני, הוא בוחן מי מהב"ת שנשארו בחוץ הוא בעל מתאם חלקי גבוה ביותר עם Y, או t גבוה ביותר בערך מוחלט ומובהק והוא ייכנס למודל. בשלב הבא הוא ממשיך ובודק מי מאלה שנשארו בחוץ הוא בעל הt הגבוה ביותר בערך מוחלט ומובהק. אם אין משתנה כזה המחשב עוצר. בשיטה זו ייתכן ומשתנים שנכנסו למודל יוצאו ממנו בשלבים מאוחרים יותר. ברגרסיה בצעדים על פעם נכנס רק משתנה 1 בכל שלב וזוהי דרך למנוע מולטי קולינאריות (מתאם בין הב"ת). בנוסף, לחוקר אין שליטה על אילו משתנים יכנסו, המחשב שולט.

## רגרסיה רבת משתנים

דוגמה 1: ניבוי היקף פעולות סניף הבנק ביצוא באמצעות 5 משתנים - היקף פעולות בבורסה, יבוא, חסכונות, עמלות והמחאות נוסעים (כל המשתנים באלפי ש"ח ליום). בניית המודל לפי מס' שיטות.

### סטטיסטיקה תיאורית ומטריצת מתאמים בין כל המשתנים

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
Y Export	114	108.2458	73.36699	.00	312.38
X1 Stock Market	114	234.0752	73.88643	100.00	435.45
X2 Import	114	238.9597	63.21238	106.00	415.00
X3 Saving	114	167.5979	51.38949	80.00	300.00
X4 Commission	114	140.2607	60.02344	30.00	289.00
X5 Traveler Checks	114	2.5240	.22241	2.00	2.89

צריך לבדוק שאין מתאם ששווה 1/1-



Correlations<sup>a</sup>

		Y Export	X1 Stock Market	X2 Import	X3 Saving	X4 Commission	X5 Traveler Checks
Y Export	Pearson Correlation Sig. (2-tailed)	1 .	.126 .182	.337** .000	.283** .002	.137 .147	.395** .000
X1 Stock Market	Pearson Correlation Sig. (2-tailed)	.126 .182	1 .	.189* .044	.342** .000	.019 .841	.094 .318
X2 Import	Pearson Correlation Sig. (2-tailed)	.337** .000	.189* .044	1 .	.161 .086	.207* .027	.282** .002
X3 Saving	Pearson Correlation Sig. (2-tailed)	.283** .002	.342** .000	.161 .086	1 .	.109 .249	.119 .206
X4 Commission	Pearson Correlation Sig. (2-tailed)	.137 .147	.019 .841	.207* .027	.109 .249	1 .	.281** .002
X5 Traveler Checks	Pearson Correlation Sig. (2-tailed)	.395** .000	.094 .318	.282** .002	.119 .206	.281** .002	1 .

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

a. Listwise N=114

בניית המודל לפי מס' שיטות:

שיטת ENTER

Regression

Variables Entered/Removed<sup>a</sup>

Model	Variables Entered	Variables Removed	Method
1	X5 Traveler Checks, X1 Stock Market, X4 Commission, X2 Import, X3 Saving	.	Enter

- a. All requested variables entered.
- b. Dependent Variable: Y Export

Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.506 <sup>a</sup>	.256	.221	64.75156

פרופורציית שונות מוסברת

- a. Predictors: (Constant), X5 Traveler Checks, X1 Stock Market, X4 Commission, X2 Import, X3 Saving
- b. Dependent Variable: Y Export

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	155428.3	5	31085.659	7.414	.000 <sup>a</sup>
	Residual	452818.5	108	4192.764		
	Total	608246.8	113			

- a. Predictors: (Constant), X5 Traveler Checks, X1 Stock Market, X4 Commission, X2 Import, X3 Saving
- b. Dependent Variable: Y Export

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-259.207	71.279		-3.637	.000
	X1 Stock Market	-2.00E-02	.089	-.020	-.225	.823
	X2 Import	.257	.103	.221	2.490	.014
	X3 Saving	.312	.128	.219	2.447	.016
	X4 Commission	-2.57E-02	.107	-.021	-.240	.811
	X5 Traveler Checks	103.806	29.454	.315	3.524	.001

a. Dependent Variable: Y Export

מתוקן

יש משתנים לא מובהקים  
( $5\% < \text{sig}$ )

$$\frac{B}{std.Error}$$

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{S.E.(\hat{\beta}_j)} \stackrel{H_0}{\sim} t_{n-k-1}$$

## שיטה היררכית

### Regression

Variables Entered/Removed<sup>a</sup>

Model	Variables Entered	Variables Removed	Method
1	X5 Traveler Checks, X2 Impoft	.	Enter
2	X3 Saying, X4 Commission, X1 Stock Market	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y Export

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.460 <sup>a</sup>	.212	.198	65.72130	.212	14.911	2	111	.000
2	.506 <sup>b</sup>	.256	.221	64.75156	.044	2.117	3	108	.102

a. Predictors: (Constant), X5 Traveler Checks, X2 Import

b. Predictors: (Constant), X5 Traveler Checks, X2 Import, X3 Saving, X4 Commission, X1 Stock Market

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	128805.7	2	64402.831	14.911	.000 <sup>a</sup>
	Residual	479441.1	111	4319.290		
	Total	608246.8	113			
2	Regression	155428.3	5	31085.659	7.414	.000 <sup>b</sup>
	Residual	452818.5	108	4192.764		
	Total	608246.8	113			

a. Predictors: (Constant), X5 Traveler Checks, X2 Import

b. Predictors: (Constant), X5 Traveler Checks, X2 Import, X3 Saving, X4 Commission, X1 Stock Market

c. Dependent Variable: Y Export

יגדל

Msres יקטן כי הוספנו עוד הסבר באמצעות עוד משתנה

תוספת מובהקת

תוספת לא מובהקת

גרגסיה רב משתנית, חישוב תרומה ייחודית, מובהקות התוספת