

ANOVA Table

Πίνακας 1 ANOVA (Analysis of Variance)

Source	SS	df	MS	Number of obs	=	3,173
Model	9509.31256	4	2377.32814	F(4, 3168)	=	115.66
Residual	65117.2616	3,168	20.5546911	Prob > F	=	0.0000
				R-squared	=	0.1274
				Adj R-squared	=	0.1263
Total	74626.5741	3,172	23.5266627	Root MSE	=	4.5337

Model (SS): Ισούται με το άθροισμα των τετραγώνων των αποκλίσεων των υπολογισμένων τιμών της εξαρτημένης μεταβλητής από τον μέσο τους και δηλώνει την μεταβλητότητα της εξαρτημένης που εξηγείται από τις μεταβολές των ανεξάρτητων μεταβλητών.

Residual (SS): Ισούται με το άθροισμα των τετραγώνων των καταλοίπων από τη γραμμή παλινδρόμησης και δηλώνει το μέρος της μεταβλητότητας της εξαρτημένης που δεν εξηγείται.

Total (SS): Το άθροισμα των ανωτέρω δύο.

Model (df): Ισούται με την αφαίρεση του συνόλου των ανεξάρτητων μεταβλητών μειωμένο κατά 1 (σταθερός όρος) και δηλώνει τους βαθμούς ελευθερίας (Degrees of Freedom) ως προς τις ανεξάρτητες μεταβλητές.

Residual (df): Ισούται με την αφαίρεση του αριθμού παρατηρήσεων μειωμένο από τον αριθμό μεταβλητών.

MS: Υπολογίζεται ως (SS/df) και ορίζεται ως ο μέσος του αθροίσματος των τετραγώνων των αποκλίσεων των υπολογισμένων τιμών της εξαρτημένης μεταβλητής από το μέσο τους.

Number of obs: Αριθμός παρατηρήσεων.

F: Υπολογίζεται ως $[Model(SS)/Residual(MS)]$ και είναι ένα τεστ το οποίο έχει μηδενική υπόθεση ότι όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

Prob>F: Ισούται με την πιθανότητα $P[F(m-1, n-m) \geq F_{stat=27,16}] \leq 0,001$. Με απλά λόγια, η πιθανότητα δείχνει αν θα απορριφθεί η μηδενική υπόθεση του F-test, που σημαίνει ότι το μοντέλο περιέχει μεταβλητές που μπορούν να εξηγήσουν την εξαρτημένη μεταβλητή.

R-squared: Ισούται με $Model(SS)/Total(SS)$ και δηλώνει το ποσοστό που εξηγούν οι ανεξάρτητες μεταβλητές την εξαρτημένη.

Root MSE: Υπολογίζει τη ρίζα των μέσων τετραγωνικών σφαλμάτων των καταλοίπων, ήτοι Residual (MS).

Πίνακας 2 (Statistics)

gr_gdpl1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ggdl	-.0078247	.0016884	-4.63	0.000	-.0111352 -.0045142
gfcfl	.1523012	.0077919	19.55	0.000	.1370235 .167579
fdi_infl1	.0056147	.0051178	1.10	0.273	-.00442 .0156493
lfpr1	-.0463196	.0099666	-4.65	0.000	-.0658612 -.0267779
_cons	2.74164	.4967217	5.52	0.000	1.767711 3.715569

Στην πρώτη στήλη βλέπουμε ότι εμφανίζονται η εξαρτημένη, οι 4 ανεξάρτητες μεταβλητές και τέλος ο σταθερός όρος.

coef.: Υπολογίζει το συντελεστή που υπάρχει σε κάθε μια από τις μεταβλητές και τον σταθερό όρο της παλινδρόμησης και δηλώνει πόσο θα μεταβληθεί η εξαρτημένη μεταβλητή αν μεταβάλλω κατά μια μονάδα την ανεξάρτητη μεταβλητή.

Std. Err.: Αποτελεί την τυπική απόκλιση και είναι ένα μέτρο που ποσοτικοποιεί τη διακύμανση των στοιχείων του δείγματος. Όσο πιο κοντά στο μηδέν (0) βρίσκεται δηλώνει ότι τα στοιχεία του δείγματος είναι πολύ κοντά στο μέσο.

t: Η τιμή που υπολογίζει το πρόγραμμα πρέπει να συγκριθεί με μια τιμή από τον στατιστικό πίνακα, η οποία είναι $t_{T-2, \alpha/2}$ όπου T: ο αριθμός των παρατηρήσεων και α : επίπεδο σημαντικότητας. Σε περίπτωση που ισχύει ότι η υπολογισμένη τιμή του προγράμματος είναι μεγαλύτερη ή ίση με την τιμή από τον πίνακα, τότε η ανεξάρτητη μεταβλητή σχετίζεται με την εξαρτημένη μεταβλητή που σημαίνει ότι είναι σημαντική για την ερμηνεία της εξαρτημένης.

$P > |t|$: Εάν η τιμή του είναι μεγαλύτερη αρκετά από «0.05» αποδεχόμαστε τη μηδενική υπόθεση « H_0 : ο συντελεστής της μεταβλητής (coefficient) είναι μηδέν».

95% conf. Interval: Δηλώνει το διάστημα εμπιστοσύνης των συντελεστών (σταθερός όρος και των ανεξάρτητων), οι οποίοι δεν είναι γνωστοί και αφορούν τον πληθυσμό. Το διάστημα αυτό παρέχει ένα φάσμα εύλογων (πιθανών) τιμών της παραμέτρου, συνοδευόμενο από τον βαθμό εμπιστοσύνης (π.χ. 95%) που έχουμε ότι το διάστημα αυτό περιέχει την πραγματική τιμή της παραμέτρου. Επιθυμούμε διαστήματα εμπιστοσύνης με μικρό φάσμα τιμών.

Descriptive statistics

Percentiles		Smallest		
1%	0	0		
5%	.88	0		
10%	11.068	0	Obs	3,890
25%	26.166	0	Sum of Wgt.	3,890
			Mean	54.53911
50%	44.283		Std. Dev.	49.60079
			Largest	
75%	69.693	616.055		
90%	102.6565	642.875	Variance	2460.239
95%	133.37	715.184	Skewness	4.629635
99%	220.561	789.833	Kurtosis	48.26834

-*Percentiles*: It reports the percentage of observations that show a value up to the specific value.

-*Smallest-largest*: It reports the 4 highest or lowest values of this variable.

-*Obs*: It reports the number of observation of this variable.

-*Mean*: It reports the mean of all values of the variable.

-*Std. Dev.*: It is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance.

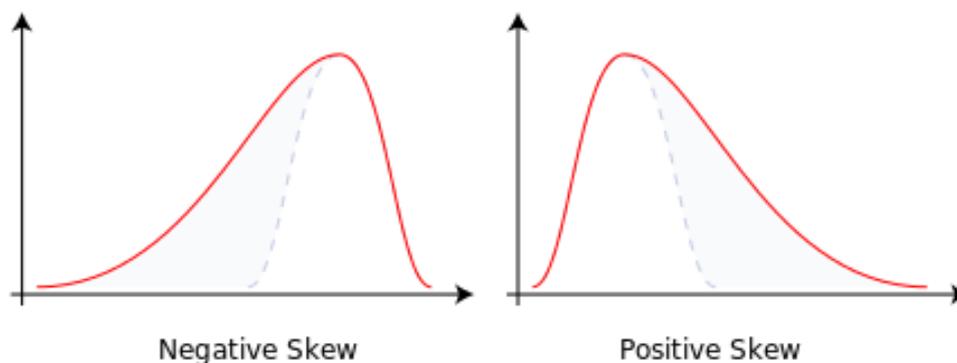
-*Skewness*: It is a measure of the asymmetry of the probability distribution of a random variable about its mean. The skewness value can be positive or negative, or undefined.

Positive skewness:

The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be *right-skewed*, *right-tailed*, or *skewed to the right*.

Negative skewness:

The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be *left-skewed*, *left-tailed*, or *skewed to the left*.



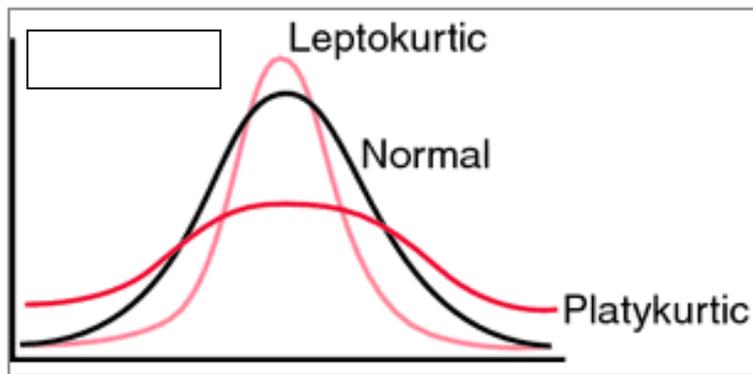
-Kurtosis: is a measure of the "tailedness" of the probability distribution of a random variable. In a similar way to the concept of skewness, kurtosis is a descriptor of the shape of a probability distribution.

Kurtosis is typically measured with respect to the normal distribution. There are the following types of kurtosis:

Mesokurtic distribution: It has excess kurtosis of zero.

Leptokurtic: It has excess kurtosis positive, e.g. t-student.

Platykurtic: It has excess kurtosis negative.



A simple example between two variables

Gross Fixed Capital Formation

Obs	3,912
Sum of Wgt.	3,912
Mean	22.50719
Std. Dev.	10.28805
Variance	105.844
Skewness	6.711935
Kurtosis	102.1554

FDI inflows

Obs	4,274
Sum of Wgt.	4,274
Mean	5.180283
Std. Dev.	14.88742
Variance	221.6352
Skewness	16.103
Kurtosis	380.4069

