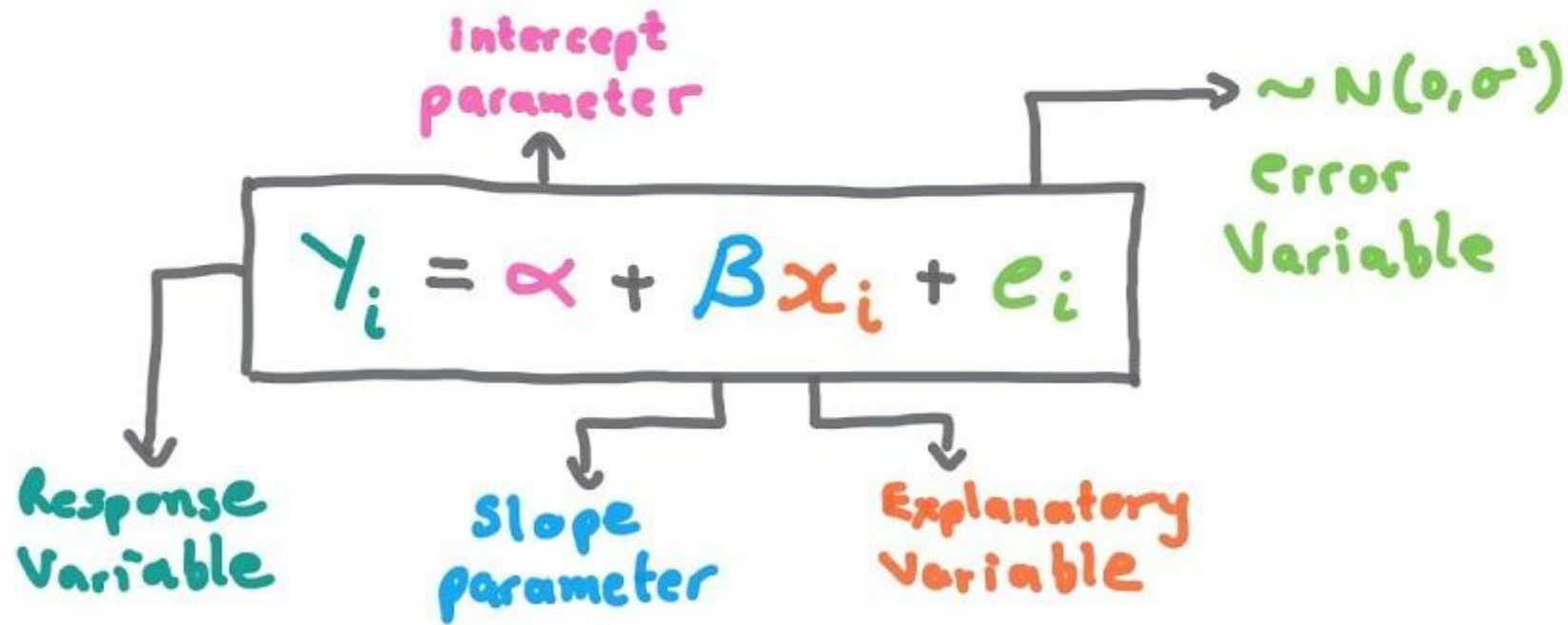


Regression Model

- model response & explanatory variables
- model bivariate data points



- Η **εκτίμηση της ζήτησης** για το προϊόν μιας επιχείρησης είναι μια ουσιαστική και συνεχής διαδικασία. Εξάλλου, οι αποφάσεις για είσοδο σε νέα αγορά, οι αποφάσεις σχετικά με:

- την παραγωγή,
- τον προγραμματισμό της παραγωγικής ικανότητας και τις επενδύσεις σε πάγια στοιχεία (ενεργητικό),
- καθώς και οι στρατηγικές τιμολόγησης και επενδύσεων

εξαρτώνται από την εκτίμηση της ζήτησης.

- Η συνάρτηση της «εκτιμώμενης-πρόβλεψης» ζήτησης παρέχει στους managers ένα ακριβή τρόπο πρόβλεψης της μελλοντικής ζήτησης για το προϊόν της εταιρείας και τις «ελαστικότητες της ζήτησης» που επιτρέπει στους διευθυντές να γνωρίζουν εκ των προτέρων τις συνέπειες των αλλαγών στις τιμές, τις τιμές των ανταγωνιστών, τις διακυμάνσεις στο εισόδημα των καταναλωτών ή τις αναμενόμενες αλλαγές σε οποιονδήποτε από τους άλλους παράγοντες επηρεάζουν τη ζήτηση.
- Αυτό το κεφάλαιο θα σας παρέχει μια απλοποιημένη έκδοση της **απλής** και **πολλαπλής παλινδρόμησης**, τεχνικές που ανήκουν σε ένα πεδίο που ονομάζεται «Οικονομετρία», το οποίο εστιάζει στη χρήση στατιστικών τεχνικών και οικονομικών θεωριών για την αντιμετώπιση οικονομικών προβλημάτων.
- Μέχρι το τέλος αυτού του κεφαλαίου, θα είστε σε θέση να κάνετε απλή εκτίμηση της ζήτησης ή τουλάχιστον να μπορείτε να διαβάζετε και να κατανοείτε τις εκτυπώσεις και τα reports του υπολογιστή που σας παρουσιάζονται.

Στις επόμενες σελίδες, θα μελετήσουμε την ανάλυση παλινδρόμησης:

Ανάλυση παλινδρόμησης (regression)

- Η **ανάλυση παλινδρόμησης** είναι μια στατιστική τεχνική για την εύρεση της καλύτερης σχέσης μεταξύ **εξαρτημένης μεταβλητής** και επιλεγμένων **ανεξάρτητων μεταβλητών**.
- **Εξαρτημένη μεταβλητή**: εξαρτάται από την τιμή *άλλων μεταβλητών*. Είναι το πρωταρχικό ενδιαφέρον για τους ερευνητές να προσδιορίσουν αυτές τις άλλες μεταβλητές.
- **Ανεξάρτητη(ες) (επεξηγητική ή ερμηνευτική) μεταβλητή(ες)**: χρησιμοποιείται για να εξηγήσει τις μεταβολές (δηλ. την διακύμανση) της εξαρτημένης μεταβλητής.
- Αυτή η ανάλυση παλινδρόμησης χρησιμοποιείται συχνά από οικονομολόγους για την εκτίμηση της ζήτησης για ένα αγαθό ή μια υπηρεσία.



Ανάλυση παλινδρόμησης (regression)

➤ Υπάρχουν δύο τύποι ανάλυσης παλινδρόμησης (regression):

1. Ανάλυση απλής παλινδρόμησης:

Σε αυτήν γίνεται χρήση μιας ανεξάρτητης μεταβλητής:

$$Y = a + bX + \mu$$

Όπου:

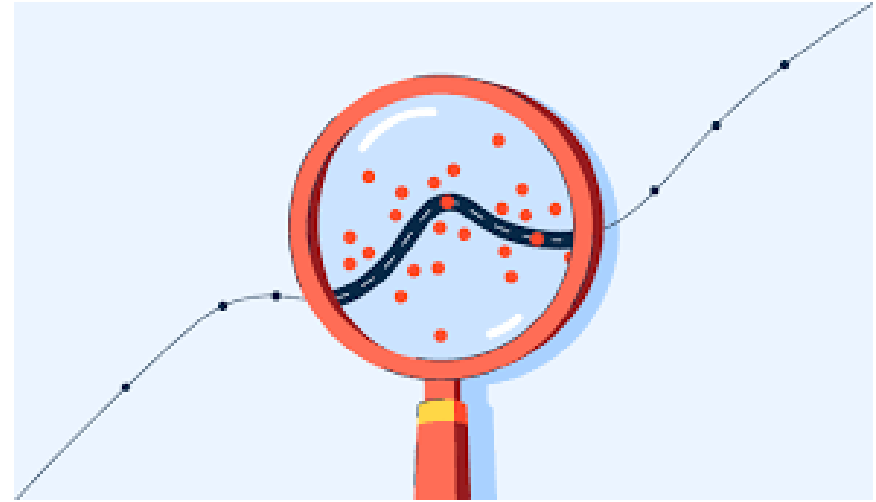
Y: εξαρτημένη μεταβλητή,

a: σταθερός όρος, είναι το σημείο τομής με τον κάθετο άξονα, **Y**-τομή

b: κλίση = συντελεστής παλινδρόμησης = παράμετρος που μετράει τον αντίκτυπο της ανεξάρτητης μεταβλητής πάνω στην εξαρτημένη

X: ανεξάρτητη (ή επεξηγηματική) μεταβλητή, που χρησιμοποιείται για να εξηγήσει τη μεταβολή της εξαρτημένης μεταβλητής

μ: τυχαίο σφάλμα



Ανάλυση παλινδρόμησης (regression)

➤ Υπάρχουν δύο τύποι ανάλυσης παλινδρόμησης (regression):

1. Ανάλυση πολλαπλής παλινδρόμησης:

Σε αυτήν γίνεται χρήση περισσότερων ανεξάρτητων μεταβλητών:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k + \mu$$

Όπου:

Y: εξαρτημένη μεταβλητή,

a: σταθερός όρος, είναι το σημείο τομής με τον κάθετο άξονα, γ-τομή

b_1 b_2 b_k : συντελεστές παλινδρόμησης = παράμετροι που μετράνε τον «αντίκτυπο» της κάθε ανεξάρτητης μεταβλητής

X_k πάνω στην εξαρτημένη Y.

X_1 X_2 ... X_k : k ανεξάρτητες (ή ερμηνευτικές) μεταβλητές, που χρησιμοποιούνται για να εξηγήσουν τις μεταβολές της εξαρτημένης μεταβλητής Y.

μ : τυχαίο σφάλμα



Για το υπολογισμό των παραπάνω μοντέλων – εξισώσεων χρησιμοποιείται η μέθοδος των **Ελαχίστων Τετραγώνων**

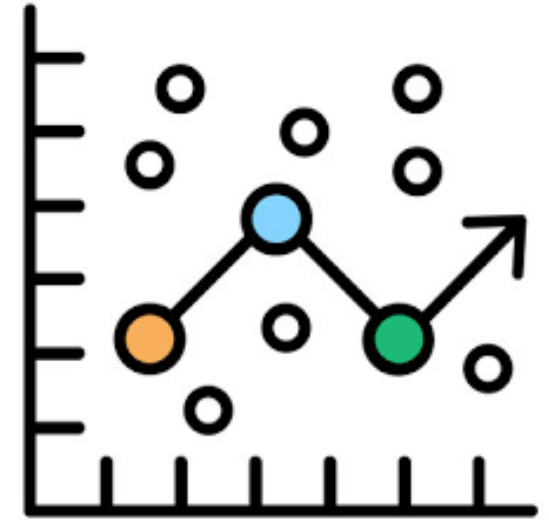
Οι ανεξάρτητες – ερμηνευτικές μεταβλητές πρέπει να είναι **ανεξάρτητες μεταξύ τους** (να μην έχουν ισχυρή συσχέτιση)

Το σφάλμα μ (λέγεται και κατάλοιπο) ακολουθεί την κανονική κατανομή με (αριθμητικό) μέσο 0

Πώς γίνεται η ανάλυση παλινδρόμησης;

Υπάρχουν ορισμένα βήματα για τη διεξαγωγή ανάλυσης παλινδρόμησης.

1. Προσδιορισμός των σχετικών μεταβλητών
2. Συλλογή των δεδομένων (data) για τις μεταβλητές
3. Καθορισμός του μοντέλου παλινδρόμησης (*regression*) (τον μαθηματικό τύπο)
4. Υπολογισμός των παραμέτρων (συντελεστές)
5. Ερμηνεία των αποτελεσμάτων
6. Στατιστική αξιολόγηση των αποτελεσμάτων (έλεγχος στατιστικής σημασίας μοντέλου)
7. Χρησιμοποίηση – αξιοποίηση των αποτελεσμάτων στη λήψη αποφάσεων (**πρόβλεψη** με χρήση μοντέλου παλινδρόμησης = αποτελέσματα).

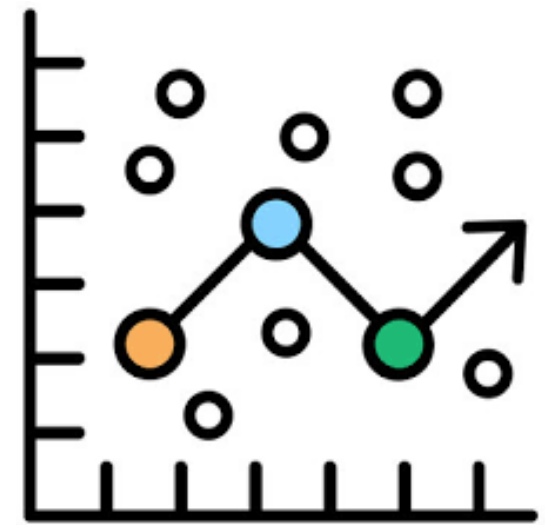


Προσδιορισμός μεταβλητών και συλλογή δεδομένων:

- Εδώ, προσπαθούμε να απαντήσουμε στο ερώτημα **ποιες μεταβλητές πρέπει να συμπεριληφθούν** στην εξίσωση παλινδρόμησης, ποιες μεταβλητές είναι σημαντικές ?

Η οικονομική θεωρία, η διαθεσιμότητα δεδομένων και άλλες συνθήκες μας οδηγούν στο *ποιες μεταβλητές* θα συμπεριληφθούν στο μοντέλο.

- Ο ρόλος της οικονομικής θεωρίας: Κατά την εκτίμηση της ζήτησης για ένα συγκεκριμένο προϊόν, η οικονομική θεωρία μας βοηθά να προσδιορίσουμε όλους τους παράγοντες που μπορεί να επηρεάσουν αυτή τη ζήτηση. Επίσης η οικονομική θεωρία βοηθά **στο είδος της σχέσης** μεταξύ της ζητούμενης ποσότητας (Q_d) που είναι η εξαρτημένη μεταβλητή και των ανεξάρτητων-ερμηνευτικών μεταβλητών.
 - π.χ., αναμένουμε αρνητικό πρόσημο για τον συντελεστή P (τιμή) λόγω της αρνητικής σχέσης μεταξύ P τιμής και Q_d ζητούμενης ποσότητας (όταν ανεβαίνει η τιμή μειώνεται η ζήτηση).
 - π.χ., εάν το υπό εξέταση αγαθό είναι κανονικό αγαθό, αναμένουμε ένα θετικό πρόσημο για τον συντελεστή εισοδήματος λόγω της θετικής σχέσης μεταξύ ζήτησης και εισοδήματος.



Στην πράξη, ωστόσο, η διαθεσιμότητα δεδομένων και το κόστος απόκτησης νέων δεδομένων μπορεί να καθορίσουν τι δεδομένα θα συμπεριληφθούν.

- Ορισμένες μεταβλητές είναι εύκολο να βρεθούν, να μετρηθούν και να ποσοτικοποιηθούν, όπως οι τιμές, ο αριθμός των καταναλωτών και ίσως και το εισόδημα.
- Για άλλες μεταβλητές είναι δύσκολο να ληφθούν δεδομένα
- Ορισμένες μεταβλητές είναι δύσκολο να ποσοτικοποιηθούν, όπως η *τοποθεσία* (αστική, προαστιακή, αγροτική) ή οι *γεύσεις* και *προτιμήσεις* (όπως, αντιπαθητική, αδιάφορη, ...) ή χρησιμοποιούνται εικονικές (δυναμικές) μεταβλητές (1 αν συμβεί το συμβάν και 0 διαφορετικά) ή (1 αν αρέσει, 0 διαφορετικά).

❑ Οι κύριοι τύποι δεδομένων που χρησιμοποιούνται στην παλινδρόμηση είναι:

1. **Διαστρωματικά:** παροχή πληροφοριών σχετικά με τις μεταβλητές για μια δεδομένη χρονική περίοδο (διαφορετικά άτομα, αγαθά, εταιρείες, χώρες ...) δηλαδή δεδομένα από διαφορετικές περιοχές, «κοινωνικά στρώματα», κλπ

2. **Χρονοσειρές:** πρόκειται για μεταβλητές σε διάφορες χρονικές περιόδους (έτη, μήνες, ημερήσια,...) δηλαδή έχουν χρονική αναφορά

3. **Pooled** (Πάνελ): Συνδυασμοί διαστρωματικών δεδομένων και χρονοσειρών

- Τα δεδομένα για μελέτες που σχετίζονται με χώρες, περιοχές ή κλάδους είναι άμεσα διαθέσιμα και αξιόπιστα.
- Τα δεδομένα για την ανάλυση συγκεκριμένων κατηγοριών προϊόντων μπορεί να είναι πιο δύσκολο να ληφθούν.
- Μία προσέγγιση είναι να αγοράσετε τα δεδομένα από παρόχους δεδομένων (εφόσον υπάρχουν), ή να πραγματοποιήσετε έρευνα καταναλωτών, ομάδες εστίασης κ.λπ.

Προδιαγραφές του μοντέλου:

- Εδώ εκφράζεται η σχέση μεταξύ της εξαρτημένης μεταβλητής (ας πούμε ζήτηση για ένα προϊόν, Q_d) και των παραγόντων που την επηρεάζουν (οι ανεξάρτητες ή επεξηγηματικές - ερμηνευτικές μεταβλητές) σε εξίσωση παλινδρόμησης (regression).
- Η εκτίμηση (υπολογισμός) της εξίσωσης παλινδρόμησης περιλαμβάνει την αναζήτηση της καλύτερης γραμμικής σχέσης μεταξύ των μεταβλητών.

Για λόγους επεξήγησης, ας υποθέσουμε ότι έχουμε λάβει διαστρωματικά δεδομένα για φοιτητές 30 τυχαίων επιλεγμένων πανεπιστημιούπολεων κατά τη διάρκεια ενός συγκεκριμένου μήνα, με την ακόλουθη εξίσωση.

$$Q_d = a + b_1P + b_2T - b_3P_c + b_4L + \mu$$

Όπου:

Q_d : Ζητούμενη ποσότητα πίτσας (μέσος αριθμός κομματιών πίτσας ανά άτομο ανά μήνα)

P : Μέση τιμή κομματιού πίτσας (σε cents)

T : Ετήσιο επίδομα φοιτητών ως μορφή εισοδήματος (σε χιλιάδες \$)

P_c : Τιμή των κουτιών αναψυκτικών (σε cents) (επειδή καταναλώνονται μαζί με πίτσα)

L : Τοποθεσία πανεπιστημιούπολεων (1 εάν αστική περιοχή, 0 διαφορετικά)

a : Σταθερός όρος (ή τομή με το κάθετο άξονα Y)

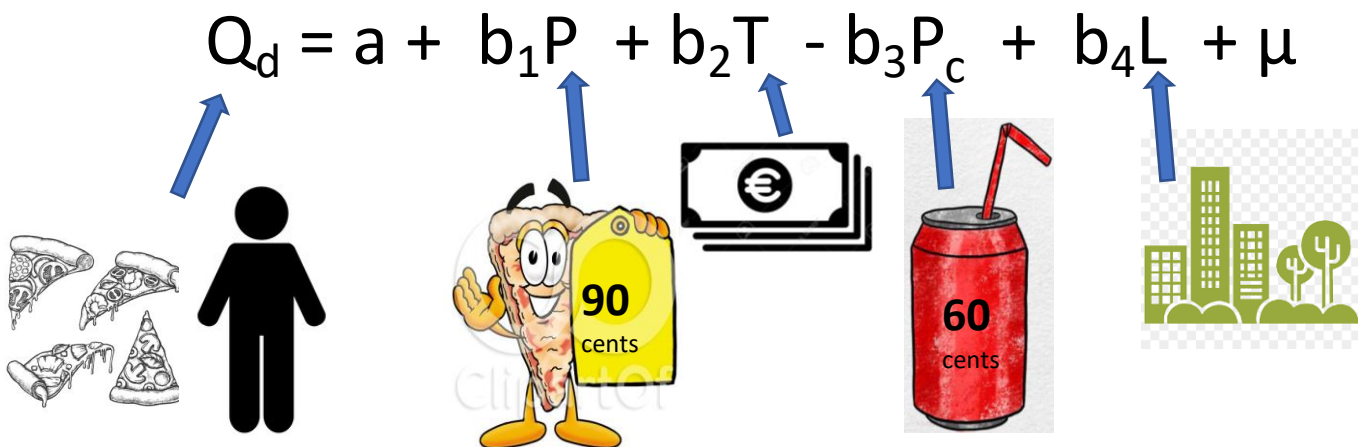
b_i : Συντελεστές ανεξάρτητων μεταβλητών προς εκτίμηση

μ : Όρος τυχαίου σφάλματος που αντιστοιχεί σε όλες τις άλλες παραλειπόμενες μεταβλητές (που μπορεί να είναι άπειρες...)

Μελέτη Περίπτωσης

Μελέτη Περίπτωσης

Μοντέλο πολλαπλής παλινδρόμησης για την ζήτηση κομματιών πίτσας από φοιτητές 30 πανεπιστημιούπολεων



- Ο συντελεστής b_i (η οριακή επίδραση) κάθε μεταβλητής είναι ο συντελεστής αυτής της μεταβλητής στην εξίσωση παλινδρόμησης. Ο συντελεστής του P είναι b_1 (dQ/dP), ο συντελεστής του T είναι b_2 (dQ/dT), κ.λπ..

Μοντέλο για την ζήτηση κομματιών πίτσας από φοιτητές 30 πανεπιστημιούπολεων

Μελέτη Περίπτωσης

$$Q_d = a + b_1P + b_2T - b_3P_c + b_4L + \mu$$

Όπου:

Q_d : Ζητούμενη ποσότητα πίτσας (μέσος αριθμός φετών πίτσας ανά φοιτητή ανά μήνα)

P : Μέση τιμή κομματιού πίτσας (σε cents)

T : Ετήσιο επίδομα φοιτητών ως μορφή εισοδήματος (σε χιλιάδες \$)

P_c : Τιμή των κουτιών αναψυκτικών (σε cents) (επειδή καταναλώνονται μαζί με πίτσα)

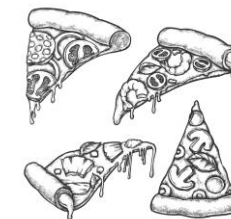
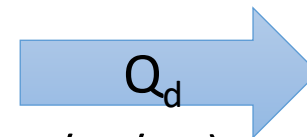
L : Τοποθεσία πανεπιστημιούπολεων (1 εάν είναι αστική περιοχή, 0 διαφορετικά)

a : Σταθερός όρος (ή τομή με το κάθετο άξονα Y)

b_i : ($=b_1, b_2, b_3, b_4$) Συντελεστές ανεξάρτητων μεταβλητών προς εκτίμηση (κλίση)

μ : Όρος τυχαίου σφάλματος που αντιστοιχεί σε όλες τις άλλες παραλειπόμενες μεταβλητές (που μπορεί να είναι άπειρες...)

- Ο συντελεστής b_i (η οριακή επίδραση) κάθε μεταβλητής είναι ο συντελεστής αυτής της μεταβλητής στην εξίσωση παλινδρόμησης. Ο συντελεστής του P είναι b_1 (dQ/dP), ο αντίκτυπος του T είναι b_2 (dQ/dT), κ.λπ..



Μοντέλο για την ζήτηση κομματιών πίτσας από φοιτητές 30 πανεπιστημιούπολεων

$$Q_d = a + b_1P + b_2T - b_3P_c + b_4L + \mu$$

Όπου:

- Ο κάθε συντελεστής b_i (η οριακή επίδραση) κάθε μεταβλητής είναι ο συντελεστής αυτής της μεταβλητής στην εξίσωση παλινδρόμησης.
- Ο συντελεστής του P είναι b_1 (dQ/dP),
- ο συντελεστής του T είναι b_2 (dQ/dT), κ.λπ..

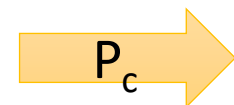
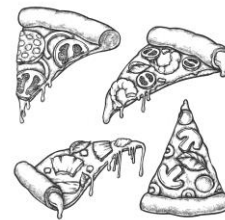
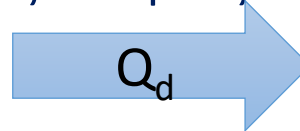
➤ Η ελαστικότητα κάθε μεταβλητής υπολογίζεται ως συνήθως

➤ $E_d = (dQ/dP) \cdot (P/Q) = b_1 \cdot (P/Q)$

➤ $E_T = (dQ/dT) \cdot (T/Q) = b_2 \cdot (T/Q)$

➤ $E_{P_c} = (dQ/dP_c) \cdot (P_c/Q) = b_3 \cdot (P_c/Q)$

➤ $E_L = (dQ/dL) \cdot (L/Q) = b_4 \cdot (L/Q)$



Εκτίμηση (υπολογισμός) των συντελεστών παλινδρόμησης:

Δεδομένης αυτής της συγκεκριμένης διάταξης της εξίσωσης παλινδρόμησης μπορούμε τώρα να εκτιμήσουμε τις τιμές των συντελεστών της ανεξάρτητης μεταβλητής καθώς και τον σταθερό όρο. Χρησιμοποιώντας τη μέθοδο συνηθισμένων ελαχίστων τετραγώνων (OLS).

Συνήθως λογισμικά στατιστικής και οικονομετρίας χρησιμοποιούνται για την εκτίμηση της εξίσωσης παλινδρόμησης χρησιμοποιώντας το excel and πολλά άλλα στατιστικά πακέτα όπως SPSS, SAS, EViews, LimOep, TSP...

Τα αποτελέσματα αναφέρονται συνήθως σε εξίσωση παλινδρόμησης ή σε μορφή πίνακα, που περιέχουν συγκεκριμένες πληροφορίες:

$$Q_d = 26,67 - 0,088P + 0,138T - 0,076P_c - 0,544L$$

(0,018) (0,0087) (0,020) (0,884)

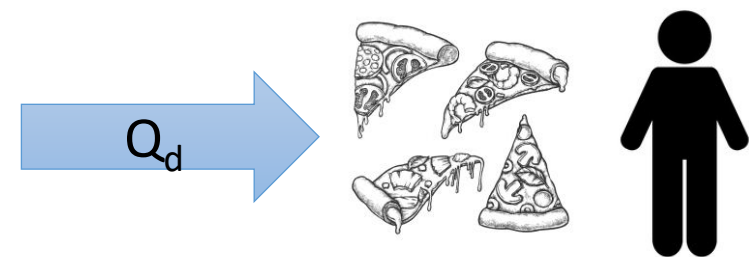
Είναι τα τυπικά σφάλματα των συντελεστών (σε παρένθεση).

$R^2 = 0,717$ (Ο συντελεστής προσδιορισμού)

Προσαρμοσμένο R^2 (Adjusted R^2) = 0,67

SE της «εκτιμήτριας» Q_d (SEE) = 1,64

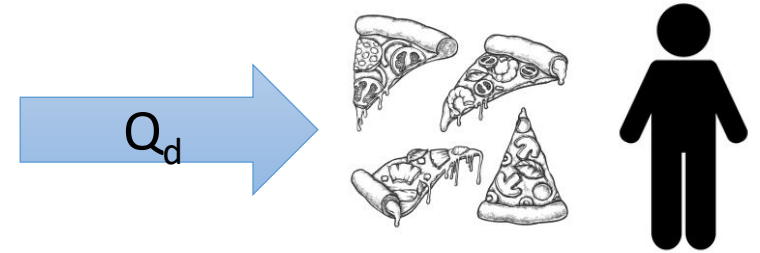
F = 15,8 (F-Statistic)



Ερμηνεία των συντελεστών παλινδρόμησης:

➤ Η ανάλυση των αποτελεσμάτων παλινδρόμησης περιλαμβάνει τα ακόλουθα βήματα:

- Έλεγχος των πρόσημων (+ ή -) και των μεγεθών
- Υπολογισμός συντελεστών ελαστικότητας $b_j = (b_1, b_2, b_3, b_4)$
- Προσδιορισμός στατιστικής σημαντικότητας



➤ Περιλαμβάνει επίσης δύο εργασίες:

- Ερμηνεία συντελεστών
- Στατιστική αξιολόγηση συντελεστών

$$Q_d = 26,67 - 0,088P + 0,138T - 0,076P_c - 0,544L$$

➤ Ποιο είναι το αναμενόμενο μέγεθος και τα πρόσημα του εκτιμώμενου συντελεστή;

➤ Είναι αυτά τα πρόσημα των συντελεστών συμβατά / σύμφωνα με την οικονομική θεωρία και είναι λογικά ?
για παράδειγμα όταν αυξάνεται η τιμή P , η ζήτηση Q_d για την κομμάτια πίτσας μειώνεται (αρνητικό πρόσημο)

T: Το πρόσημο για το εισόδημα εξαρτάται από το αν η πίτσα είναι κανονικό ή κατώτερο αγαθό. (+,-)

P_c : το αναμενόμενο πρόσημο για το P_c είναι (-) λόγω συμπληρωματικής σχέσης των αναψυκτικών με την πίτσα (καταναλώνονται μαζί) (όταν P_c αυξάνεται, η ζήτηση για πίτσα μειώνεται)

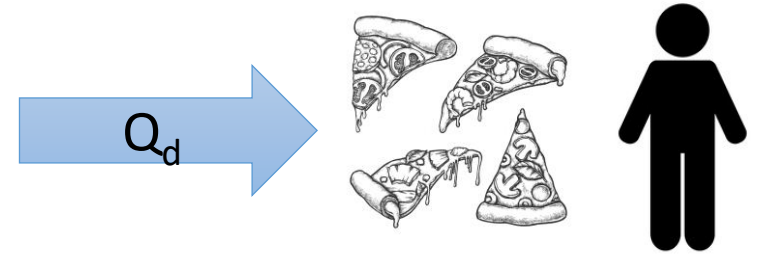
L: Το αναμενόμενο πρόσημο είναι (-) επειδή στις αστικές περιοχές ($L = 1$) οι μαθητές έχουν ποικιλία εστιατορίων (περισσότερα υποκατάστατα), άρα >> θα καταναλώσουν λιγότερη πίτσα από ό,τι οι αντίστοιχοι σε άλλες περιοχές.

➤ Έλεγχος και με βάση την Οικονομική θεωρία

Ερμηνεία των συντελεστών παλινδρόμησης:

Όσον αφορά το μέγεθος, μπορούμε να δούμε ότι κάθε εκτιμώμενος συντελεστής μας λέει πόσο θα αλλάξει η ζήτηση για κομμάτια πίτσας σε σχέση με μια μοναδιαία μεταβολή (μεταβολή μιας μονάδας) σε καθεμία από τις ανεξάρτητες μεταβλητές.

$$Q_d = 26,67 - 0,088P + 0,138T - 0,076P_c - 0,544L$$



$b_1 = -0,088$ αύξηση μιας μονάδας στο P **μειώνει το Q_d κατά 0,088 μονάδες.**

$b_2 = 0,138$ για μια αύξηση 1000 \$ στα δίδακτρα, **η ζήτηση Q_d αυξάνει κατά 0,138 μονάδες.**

$b_3 = -0,076$ για μια αύξηση μιας μονάδας στην τιμή των αναψυκτικών (+1 cent), **η ζήτηση Q_d μειώνεται κατά 0,076 μονάδες.**

b_4 : οι φοιτητές σε αστικές περιοχές θα αγοράσουν περίπου **μισά** λιγότερα (0,544) από φοιτητές σε μη αστικές περιοχές.

➤ Το μέγεθος των συντελεστών παλινδρόμησης μετριέται με την ελαστικότητα κάθε μεταβλητής.

Εάν $P=100$ (cents), $T=14$ (\$000), $P_c=110$ (cents), $L=1$

Τότε υπολογίζουμε: $Q_d = 26.67 - 0.088(100) + 0.138(14) - 0.076(110) - 0.544(1) = 10.898$

➤ $E_d = b_1 \cdot (P/Q) = -0.088 (100) / 10.898 = -0,807$ αρκετά ανελαστικό

➤ $E_T = b_2 \cdot (T/Q) = 0.138 (14) / 10.898 = 0,177$ ελαστικό

➤ $E_{P_c} = b_3 \cdot (P_c/Q) = 0,076 (110) / 10.898 = -0,767$ αρκετά ανελαστικό

➤ $E_L = b_4 \cdot (L/Q) = -0,544 (1) / 10.898 = -0,05$ ελαστικό

Μελέτη Περίπτωσης

Τέλος