Hellenic Republic
INTERNATIONAL HELLENIC UNIVERSITY
University Center for International Programmes of Studies

Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας

Πρόγραμμα Μεταπτυχιακών Σπουδών:

Διοίκηση Επιχειρήσεων και Οργανισμών για Στελέχη

Ποσοτικές Μέθοδοι για Στελέχη Επιχειρήσεων
Quantitative Methods for Managers
Ανάλυση Διακύμανσης- ANOVA

by
Dr. Efstathios Dimitriadis
Mathematic
Ph.D in Applied Statistics
M.Sc in Statistics and Demography
M.Sc in Quality Assurance

One –Way ANOVA

- Η μονόδρομη ανάλυση της διακύμανσης (ANOVA = Analysis of Variance) χρησιμοποιείται για τον έλεγχο του ισχυρισμού ότι περισσότεροι από δύο αριθμητικοί μέσοι πληθυσμών είναι ίσοι.
- Η ANOVA δεν ελέγχει αν ένας μέσος όρος είναι μικρότερος από τον άλλο αλλά μόνο αν είναι όλοι ίσοι ή τουλάχιστον ένας είναι διαφορετικός.
- Είναι μια επέκταση του t-τεστ δύο ανεξάρτητων δειγμάτων.

One –Way ANOVA

- Η μεταβλητή ***απόκρισης (εξαρτημένη ποσοτική μεταβλητή)*** είναι η μεταβλητή που συγκρίνεται
- Η μεταβλητή ***παράγοντας (ποιοτική μεταβλητή)*** είναι η μεταβλητή που χρησιμοποιείται για τον καθορισμό των ομάδων
- Υποθέτουμε ότι έχουμε ***k δείγματα (γκρουπ)***
- Στην ***one-way*** **ANOVA** κάθε τιμή ταξινομείται με ακριβώς έναν τρόπο.

Παραδείγματα περιλαμβάνουν συγκρίσεις ως προς το φύλο, τη φυλή, το πολιτικό κόμμα, το χρώμα κ.λ.π

One –Way ANOVA

- **Συνθήκες ή Υποθέσεις**
  - Τα δεδομένα λαμβάνονται με τυχαία δειγματοληψία
  - Υποτίθεται ότι οι Διακυμάνσεις όλων των δειγμάτων είναι ίσες.
  - Τα κατάλοιπα κατανέμονται κανονικά
- ***Μηδενική Υπόθεση H_0***: Οι μέσες τιμές όλων των δειγμάτων είναι ίσες $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$
- ***Εναλλακτική Υπόθεση H_1 :*** Μια τουλάχιστον μέση τιμή είναι διαφορετική

Παράδειγμα ANOVA

Ας υποθέσουμε ότι το Εθνικό Συμβούλιο Ασφάλειας Μεταφορών (NTSB) θέλει να εξετάσει την ασφάλεια των συμπαγών αυτοκινήτων, των μεσαίων αυτοκινήτων και των αυτοκινήτων πλήρους μεγέθους. Συλλέγει τρία δείγματα για κάθε τύπο αυτοκινήτου. Χρησιμοποιώντας τα υποθετικά δεδομένα που παρέχονται παρακάτω, ελέγξτε εάν η μέση πίεση που ασκείται στο κεφάλι του οδηγού κατά τη διάρκεια δοκιμής σύγκρουσης είναι ίση για κάθε τύπο αυτοκινήτου. Χρησιμοποιήστε α = 5%.

	Compact cars	Midsize cars	Full-size cars
	643	469	484
	655	427	456
	702	525	402
$\bar{x}$	666.67	473.67	447.33
S	31.18	49.17	41.68

A= Compact cars, B=Midsize cars C= Full-size cars

$\overline{x}_A = 666{,}67$ $\overline{x}_B = 473{,}67$ $\overline{x}_C = 447{,}33$

$s_A^2 = 31{,}18$ $s_B^2 = 49{,}17$ $s_C^2 = 41{,}68$

1. Ορίστε την Μηδενική και την Εναλλακτική υπόθεση

Η μηδενική υπόθεση: Η μέση πίεση είναι ίση στους τρεις τύπους αυτοκινήτων.

H_0: $\mu_A = \mu_B = \mu_C$

Επειδή η μηδενική υπόθεση υποθέτει ότι όλες οι μέσες τιμές είναι ίσες, εμείς πρέπει να την απορρίψουμε αν έστω μια μέση τιμή δεν είναι ίση. Έτσι η εναλλακτική υπόθεση είναι:

H_1: Τουλάχιστον μία μέση πίεση δεν είναι στατιστικά ίση.

2. Καθορίστε και υπολογίστε το κατάλληλο στατιστικό έλεγχο

Ο στατιστικός έλεγχος στην ANOVA είναι η αναλογία της ***μεταξύ*** των δειγμάτων διακύμανσης και της ***εντός*** των δειγμάτων διακύμανσης. Ακολουθεί την κατανομή F.

Το συνολικό άθροισμα τετραγώνων- η συνολική διακύμανση των δεδομένων, είναι το άθροισμα της διακύμανσης μεταξύ των δειγμάτων και εντός των δειγμάτων.

Συνολικό Άθροισμα Τετραγώνων: SST = $\sum_{i=1}^{r} \sum_{j=1}^{k} (x_{ij} - \bar{\bar{x}})^2$

όπου r το πλήθος των γραμμών στον πίνακα, k το πλήθος των δειγμάτων, $\bar{\bar{x}}$ ο γενικός μέσος όρος, και χ_{ij} είναι η i^{th} παρατήρηση στην j^{th} στήλη.

$$\bar{\bar{x}} = \frac{(643 + 655 + 702 + 469 + 427 + 525 + 484 + 456 + 402)}{9} = 529{,}22$$

SST=(643-529,22)^2+(655-529,22)^2+……….+(402-529,22)^2=**96303,55**

Άθροισμα Τετραγώνων μεταξύ των δειγμάτων: SSB= $\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2$

Είναι η διακύμανση των δεδομένων μεταξύ των διαφορετικών δειγμάτων, όπου n_j παριστάνει το μέγεθος του j δείγματος.

SSB $=3*(666,67-529,22)^2+3*(473,67-529,22)^2+3*(447,33-529,22)^2=$**86049,55**

Άθροισμα Τετραγώνων εντός των δειγμάτων: SSW= $\sum_{i=1}^{r} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2$

Είναι η διακύμανση των δεδομένων εντός των ανεξάρτητων δειγμάτων.

SSW$=(643-666,67)^2+(655-666,67)^2+(702-666,67)^2+$
$+(469-473,67)^2+(427-473,67)^2+(525-473,67)^2+$
$+(484-447,33)^2+(456-447,33)^2+(402-447,33)^2=$**10254**

SST=SSB+SSW

Ολικά Μέσα Τετράγωνα: $MST = \frac{SST}{N-1}$ $MST = \frac{96303{,}55}{9-1} = 12037{,}94$

Μέσα Τετράγωνα μεταξύ των δειγμάτων: $MSB = \frac{SSB}{k-1}$

$$MSB = \frac{86049{,}55}{3-1} = 43024{,}78$$

Μέσα Τετράγωνα εντός των δειγμάτων: $MSW = \frac{SSW}{N-k}$ $MSW = \frac{10254}{9-3} = 1709$

Ο στατιστικός έλεγχος μπορεί τώρα να πραγματοποιηθεί. Για την ANOVA ο στατιστικός έλεγχος εκφράζεται από το λόγο MSB προς MSW. Επιπλέον, αυτός ο λόγος είναι γνωστό ότι ακολουθεί μια κατανομή F. Ως εκ τούτου,

$$F = \frac{MSB}{MSW} \qquad F = \frac{43024{,}78}{1709} = 25{,}17$$

The basic one-way ANOVA table

Source	SS	df	MS	F	p
Between	SSB	k-1	MSB	$\frac{MSB}{MSW}$	
Within	SSW	N-k	MSW		
Total	**SST**	**N-1**			

3. Κρίσιμη τιμή

Για να υπολογιστεί η κρίσιμη τιμή από την F κατανομή πρέπει να γνωρίζουμε του αριθμητή (MSB) και του παρονομαστή τους βαθμούς ελευθερίας, μαζί με το επίπεδο σημαντικότητας.
Η κατανομή F έχει df1 και df2 βαθμούς ελευθερίας, όπου df1 είναι οι βαθμοί ελευθερίας του αριθμητή και είναι ίσοι με k-1 και df2 είναι οι βαθμοί ελευθερίας του παρονομαστή και είναι ίσοι με N-k.
Στο παράδειγμα, df1 = 3 - 1 = 2 και df2 = 9 - 3 = 6. Έτσι, πρέπει να βρούμε την τιμή της κατανομής $F_{2,6}$ που αντιστοιχεί σε α= 5%.
Χρησιμοποιώντας τον πίνακα της F κατανομής έχουμε: $F_{2,6}$= 5.14.

4. Κανόνας Απόφασης

Απορρίπτεται η μηδενική υπόθεση αν:
F (παρατηρούμενη τιμή) > F (κριτική τιμή).
Στο παράδειγμα 25,17 > 5,14
Έτσι, ***απορρίπτουμε την μηδενική υπόθεση***

Πολλαπλές Συγκρίσεις

- Ένα στατιστικά σημαντικό ***F-test*** μας λέει ότι τουλάχιστον δύο από τους πληθυσμούς έχουν διαφορετική μέση τιμή, αλλά δεν μας λέει ποιός από αυτούς διαφέρει από τους άλλους.
- Χρειαζόμαστε επιπλέον ελέγχους, τους οποίους ονομάζουμε Πολλαπλές Συγκρίσεις, για να συγκρίνουμε όλους τους μέσους.
- Βλέπουμε τις διαφορές των μέσων των πληθυσμών μεταξύ όλων των ζευγών, καθώς επίσης και το διάστημα εμπιστοσύνης για κάθε διαφορά.
- Εάν έχουμε $\boldsymbol{k}$ πληθυσμούς υπάρχουν: $\dfrac{k(k-1)}{2}$ ζεύγη μέσων τιμών προς σύγκριση.

Μονόδρομη ANOVA- One Way ANOVA: Έλεγχοι Πολλαπλών Συγκρίσεων

a. Όταν οι Διακυμάνσεις είναι ίσες

1. LSD
2. Bonferroni
3. Sidak
4. Scheffe
5. Tukay
6. Duncan

b. Όταν οι Διακυμάνσεις δεν είναι ίσες

1. Tamhane’s T2
2. Dunnnett’s T3
3. Games- Howell
4. Dunnett’s C

Πολλαπλές Συγκρίσεις

- First, the ***Bonferroni correction***.
- Instead of using $t_{df,\alpha/2}$ as our multiplier for the confidence interval, we use $t_{df,\alpha/2L}$, where L is the total number of possible pair-wise comparisons (i.e. $L = k(k-1)/2$).
- That is, we divide $\alpha/2$ by the number of tests to be done ($\alpha/2L$).
- This assumes all pair-wise comparisons are independent, which is not the case, so this adjustment is too conservative (intervals will be too wide; i.e. finds too few significant differences).

Πολλαπλές Συγκρίσεις

- Second, we have ***Tukey Intervals***.
- The calculation of Tukey Intervals is quite complicated, but overcomes the problems of the unadjusted pair-wise comparisons finding too many significant differences (i.e. confidence intervals that are too narrow), and the Bonferroni correction finding too few significant differences (i.e. confidence intervals that are too wide).