

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΛΕΞΑΚΗΣ ΠΑΡΑΣΚΕΥΑΣ

«Έλεγχοι υποθέσεων για την εκθετική κατανομή»

ΣΤΗΝ ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ : ΣΤΑΤΙΣΤΙΚΗ

ΕΠΙΒΛΕΠΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ :

Βόντα Φιλία, Καθηγήτρια Ε.Μ.Π

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

Βόντα Φιλία, Καθηγήτρια Ε.Μ.Π

Παπανικολάου Βασίλειος, Ομότιμος Καθηγητής Ε.Μ.Π

Καρώνη Χρυσής, Ομότιμη Καθηγήτρια Ε.Μ.Π

[ΑΘΗΝΑ]

ΕΥΧΑΡΙΣΤΙΕΣ

Ιδιαίτερες και θερμές ευχαριστίες θα ήθελα να απευθύνω προς την επιβλέπουσα κ. Βόντα Φιλία, Καθηγήτρια Ε.Μ.Π. για την πολύτιμη αρωγή της αφενός στην επιλογή της θεματικής της παρούσας διπλωματικής εργασίας μου και αφετέρου στην καθοδήγηση της κατά τη διάρκεια εκπόνησής της. Επιπρόσθετα θα ήθελα να ευχαριστήσω τον κ.Παπανικολάου Βασίλειο, Ομότιμο Καθηγητή Ε.Μ.Π. και την κ.Καρώνη Χρυσήδα, Ομότιμη Καθηγήτρια Ε.Μ.Π. για τη συμμετοχή τους στην τριμελή επιτροπή και το χρόνο που αφιέρωσαν στην διπλωματική μου εργασία.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για τη στήριξη που μου παρείχαν καθ' όλη τη διάρκεια των προπτυχιακών σπουδών μου.

.....

ΑΛΕΞΑΚΗΣ ΠΑΡΑΣΚΕΥΑΣ

Διπλωματούχος Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

© (2023) Εθνικό Μετσόβιο Πολυτεχνείο. All rights Reserved. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σ' αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία πραγματεύεται τον έλεγχο υποθέσεων για την εκθετικότητα. Δηλαδή, ελέγχουμε με στατιστικές μεθόδους αν ένα σύνολο δεδομένων ακολουθεί την εκθετική κατανομή με συγκεκριμένη παράμετρο. Σκοπός της διπλωματικής, είναι η κατανόηση της πρακτικής διαδικασίας που ακολουθείται για την εξαγωγή έγκυρων συμπερασμάτων που βασίζονται σε στατιστικούς ελέγχους για τη συγκεκριμένη υπόθεση. Ο έλεγχος υποθέσεων αποτελεί ένα σημαντικό εργαλείο στον χώρο της στατιστικής και της επιστήμης των δεδομένων, καθώς επιτρέπει την αξιολόγηση της στατιστικής σημαντικότητας και την καθοδήγηση στη λήψη αποφάσεων βασισμένων σε παρατηρήσεις.

Στο πρώτο κεφάλαιο της εργασίας, γίνεται μία θεωρητική περιγραφή της εκθετικής κατανομής, αναφέρεται η ιστορία της κατανομής και παρουσιάζεται η χρησιμότητά της σε διάφορους επιστημονικούς κλάδους με αναφορά σε κάποιες πρακτικές εφαρμογές. Επίσης στο πρώτο Κεφάλαιο δίνονται ορισμοί για βασικές έννοιες της Στατιστικής, της Εκτιμητικής και των Ελέγχων Υποθέσεων. Στα επόμενα δύο κεφάλαια περιλαμβάνεται η θεωρητική περιγραφή των ελέγχων που χρησιμοποιήθηκαν στην εργασία καθώς και η χρησιμότητά τους, γίνεται η κατηγοριοποίησή τους και περιγράφεται η συμβολή των επιστημόνων που τους ανακάλυψαν. Στη συνέχεια, ακολουθεί η μαθηματική περιγραφή των ελέγχων. Στο τέταρτο κεφάλαιο πραγματοποιείται πρακτική εφαρμογή των ελέγχων σε προσομοιωμένα δεδομένα που γνωρίζουμε εκ των προτέρων ότι ακολουθούν την εκθετική κατανομή με σκοπό την κατανόηση της διαδικασίας που ακολουθείται για να υλοποιηθούν οι έλεγχοι. Στο τελευταίο κεφάλαιο, επεξεργάζονται και αναλύονται πραγματικά δεδομένα και ελέγχεται η υπόθεση της εκθετικότητας σε αυτά, καθώς εξάγονται και συμπεράσματα για τον συγκεκριμένο έλεγχο.

Για την επεξεργασία και ανάλυση των δεδομένων χρησιμοποιήθηκαν το πρόγραμμα Excel και το στατιστικό πακέτο R. Σε όλη την έκταση της εργασίας, όπου αυτό κρίνεται χρήσιμο για την πληρέστερη κατανόηση, παρουσιάζονται διαγράμματα διαφόρων ειδών .

Λέξεις κλειδιά:

Εκθετική κατανομή, ελεγχος συνάρτησης, έλεγχος υποθέσεων, μηδενική υπόθεση, p-value, διαγράμματα, στατιστική σημαντικότητα, στατιστικό πακέτο R

Abstract

This thesis focuses on hypothesis testing for exponentiality (the problem being whether a data set follows the exponential distribution with a given parameter). The purpose of the thesis, is to understand the practical procedure followed to draw valid conclusions based on statistical tests for the given hypothesis. Hypothesis testing is an important tool in the field of statistics and data science, as it allows for the evaluation of statistical significance and guidance in making decisions based on observations.

In the first chapter of the thesis, a theoretical description of the exponential distribution is given, the history of the distribution is mentioned and the usefulness of the exponential distribution in various scientific disciplines is presented with reference to some practical applications. Also in the first chapter definitions of basic concepts of Statistics, Estimation and Hypothesis Testing are given. The next two Chapters give a theoretical description of the tests used in the paper and their usefulness, their categorization and the contribution of the mathematicians – statisticians who proposed them. Then, the mathematical description of the tests is presented. In Chapter 4, a practical application of the tests to simulated data that we know they follow the exponential distribution is carried out in order to understand the procedure followed to implement the tests. In the last chapter, real data are processed and analyzed and the assumption of exponentiality is tested on them, and conclusions are drawn for this particular test.

The Excel program and the statistical package R were used to process and analyze the data. Throughout the paper, where it is considered useful for a fuller understanding, diagrams/graphs of various kinds are presented.

Key words:

Exponential distribution, test function, hypothesis testing, null hypothesis, p-value, graphs, statistical significance, Statistical package R

EΙΚΟΝΕΣ

EΙΚΟΝΑ 1 : Διάγραμμα P-P Plot.....	34
EΙΚΟΝΑ 2 : Διάγραμμα Q-Q Plot.....	37
EΙΚΟΝΑ 3 : Ιστόγραμμα προσομοιωμένων δεδομένων.....	51
EΙΚΟΝΑ 4 : Ιστόγραμμα προσέγγισης εκθετικής κατανομής.....	52
EΙΚΟΝΑ 5 : Θηκοδιάγραμμα προσομοιωμένων δεδομένων.....	53
EΙΚΟΝΑ 6 : Διάγραμμα Q-Q Plot προσομοιωμένων δεδομένων.....	54
EΙΚΟΝΑ 7 : Γραφική παράσταση εμπειρικής συνάρτησης κατανομής (Kolmogorov).....	58
EΙΚΟΝΑ 8 : Γραφική παράσταση εμπειρικής συνάρτησης κατανομής (Lilliefors).....	61
EΙΚΟΝΑ 9 : Απόσπασμα πραγματικών δεδομένων.....	70
EΙΚΟΝΑ 10 : Ιστόγραμμα πραγματικών δεδομένων.....	73
EΙΚΟΝΑ 11 : Ιστόγραμμα προσέγγισης εκθετικής κατανομής.....	74
EΙΚΟΝΑ 12 : Διάγραμμα Q-Q Plot πραγματικών δεδομένων.....	75
EΙΚΟΝΑ 13 : Γραφική παράσταση εμπειρικής συνάρτησης κατανομής πραγματικών δεδομένων (Kolmogorov).....	76

ΠΙΝΑΚΕΣ ΥΠΟΛΟΓΙΣΜΟΥ

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 1: Παράδειγμα υπολογισμού P-P Plot.....	33
ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 2: Παράδειγμα υπολογισμού Q-Q Plot.....	36
ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 3: Υπολογισμός ελεγχουσυνάρτησης Hartley's F-Max Test.....	55
ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 4: Υπολογισμός ελεγχουσυνάρτησης KUSK.....	56
ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 5 : Υπολογισμός ελεγχουσυνάρτησης Kolmogorov..	57

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 6 : Υπολογισμός ελεγχουσυνάρτησης Lilliefors.....	59
--	-----------

ΠΙΝΑΚΕΣ ΤΙΜΩΝ ΕΛΕΓΧΟΣΥΝΑΡΤΗΣΕΩΝ

ΠΙΝΑΚΑΣ 1 : Αντίστροφα ποσοστιαία σημεία της X^2 κατανομής.....	82
ΠΙΝΑΚΑΣ 2 : Κρίσιμες τιμές της ελεγχουσυνάρτησης Kolmogorov.....	83
ΠΙΝΑΚΑΣ 3 : p-ποσοστιαία σημεία της ελεγχουσυνάρτησης Lilliefors για τον έλεγχο καλής προσασμογής της εκθετικής κατανομής.....	84
ΠΙΝΑΚΑΣ 4 : Κρίσιμες τιμές για τον έλεγχο Deshpande για σταθερό $\alpha=0.05..$	85
ΠΙΝΑΚΑΣ 5 : Κρίσιμες τιμές για τον έλεγχο Gini.....	85
ΠΙΝΑΚΑΣ 6 : Κρίσιμες τιμές για τον έλεγχο New Better than Used.....	86
ΠΙΝΑΚΑΣ 7 : Κρίσιμες τιμές για την κατανομή F για $\alpha = 0.05$.....	87
ΠΙΝΑΚΑΣ 8 : Κρίσιμες τιμές για την κατανομή F για $\alpha = 0.10$.....	88
ΠΙΝΑΚΑΣ 9 : Κρίσιμες τιμές για το $W(E)$ στατιστικό του ελέγχου για την εκθετικότητα (Shapiro-Wilk).....	89

Περιεχόμενα

ΠΡΟΛΟΓΟΣ	5
ΚΕΦΑΛΑΙΟ 1 Ιστορία και χρησιμότητα της εκθετικής κατανομής	6
1.1 Περιγραφή εκθετικής κατανομής.....	6
1.2 Χρησιμότητα εκθετικής κατανομής	7
1.3 Ιστορία εκθετικής κατανομής.....	9
1.4 Βασικοί ορισμοί στη Στατιστική	11
1.5 Βασικοί ορισμοί στην Εκτιμητική	12
1.6 Έλεγχοι υποθέσεων	13
1.6.1 Ιστορία.....	13
1.6.2 Βασικοί Ορισμοί	13
1.6.3 Ιδιότητες Στατιστικών ελέγχων	16
ΚΕΦΑΛΑΙΟ 2 Έλεγχοι καλής προσαρμογής για την εκθετική κατανομή – Ιστορικά στοιχεία	17
2.1 Εισαγωγή	17
2.2 Μονότονες συναρτήσεις κινδύνου (Monotonic hazard rates)	18
2.2.1 KSL Test (Kolmogorov – Smirnov – Lilliefors Test).....	18
2.2.2 The Gini Statistic	19
2.2.3 The Lorenz Statistic.....	20
2.2.4 The Pietra Statistic.....	21
2.2.5 Desphande’s Test.....	22
2.2.6 WE Test.....	22
2.2.7 New Better Than Used Test.....	23
2.2.8 EPS Test.....	24
2.2.9 Hartley’s F-Max Test	25
2.2.10 Cox and Oakes Score Test.....	25
2.2.11 Wong and Wong’s Extremal Quotient Test	26

2.3 Μη μονότονες συναρτήσεις κινδύνου (Non-Monotonic hazard rates)	27
2.3.1 Gnedenko 's F-Test	28
2.3.2 Harris modification of Gnedenko 's F-Test	28
2.3.3 Lin and Mudholkar 's Bivariate F-Test	29
2.3.4 Cox and Oakes Score Test.....	30
2.3.5 . Desphande's Test.....	31
ΚΕΦΑΛΑΙΟ 3 Έλεγχοι καλής προσαρμογής για την εκθετική κατανομή	32
3.1 Οι γραφικοί έλεγχοι P-P Plot και Q-Q Plot	32
3.2 Ο γενικός έλεγχος καλής προσαρμογής χ^2	37
3.3 Λοξότητα και Κύρτωση (Skewness and Kurtosis) : KUSK	38
3.4 KSL Test.....	39
3.5 The Gini Statistic (Gastwirth and Gail, 1978).....	42
3.6 The Lorenz Statistic (Gail and Gastwirth, 1978)	43
3.7 The Pietra Statistic (Gail and Gastwirth , 1978).....	43
3.8 Desphande's Test (Deshpande, 1983)	43
3.9 Shapiro and Wilk Test (Shapiro and Wilk , 1972)	44
3.10 New Better Than Used Test (Hollander and Proschan , 1972)	45
3.11 EPS Test (Epstein , 1960)	45
3.12 Hartley's F-Max Test (Hartley, 1950).....	46
3.13 Cox and Oakes Score Test (Cox and Oakes , 1984).....	46
3.14 Wong and Wong's Extremal Quotient Test (Wong and Wong, 1979)	47
3.15 WE Test (Hahn and Shapiro, 1967).....	47
3.16 Gnedenko 's F-Test (Gnedenko, 1969)	47
3.17 Harris modification of Gnedenko 's F-Test (Harris, 1976)	48
3.18 Lin and Mudholkar 's Bivariate F-Test (Lin and Mudholkar, 1980)	48
ΚΕΦΑΛΑΙΟ 4 Ανάλυση προσομοιωμένων δεδομένων με χρήση του Excel και του στατιστικού πακέτου R.....	50
4.1 Εισαγωγή	50
4.2 Προσομοίωση δεδομένων και διαγράμματα.....	50
4.3 Έλεγχοι Kolmogorov , Lilliefors, Hartley και KUSK.....	55

4.3.1 Έλεγχος Hartley's F-Max	55
4.3.2 Έλεγχος KUSK.....	56
4.3.3 Έλεγχος Kolmogorov.....	57
4.3.4 Έλεγχος Lilliefors.....	59
4.4 Έλεγχοι εκθετικότητας με χρήση του στατιστικού πακέτου R.....	61
4.4.1 Gnedenko 's F-Test	62
4.4.2 Harris modification of Gnedenko 's F-Test	62
4.4.3 The Gini Statistic	62
4.4.4 The Lorenz Statistic.....	63
4.4.5 The Pietra Statistic.....	63
4.4.6 Dehsbande's Test.....	63
4.4.7 Shapiro and Wilk Test	63
4.4.8 New Better than Used Test	64
4.4.9 EPS Test.....	64
4.4.10 Cox and Oakes Score Test.....	64
4.4.11 Wong and Wong's Test.....	65
4.4.12 WE Test.....	65
ΚΕΦΑΛΑΙΟ 5 Επεξεργασία και έλεγχος εκθετικότητας πραγματικών δεδομένων με χρήση του Excel και του στατιστικού πακέτου R.....	66
5.1 Εισαγωγή	66
5.2 Προετοιμασία δεδομένων.....	67
5.2.1 Τακτοποίηση δεδομένων με χρήση του Excel	67
5.2.2 Εισαγωγή των δεδομένων στην R	71
5.3 Στατιστική ανάλυση των δεδομένων για την εκθετικότητα	72
5.3.1 Γραφικός έλεγχος	72
5.3.2 Ανάλυση μέσω στατιστικών ελέγχων.....	77
5.4 Συμπεράσματα	80
ΠΑΡΑΡΤΗΜΑ : ΠΙΝΑΚΕΣ	82
ΠΗΓΕΣ	90
Βιβλιογραφία :.....	90

Επιστημονικά άρθρα :	90
Πηγές διαδικτύου :	91

ΠΡΟΛΟΓΟΣ

Η σημερινή εποχή , χαρακτηρίζεται από πολλούς ως εποχή της πληροφορίας. Άλλοι υποστηρίζουν ότι όποιος διαθέτει τα δεδομένα , κατέχει και κυρίαρχη θέση στον κόσμο . Η άντληση , η σωστή επεξεργασία και η ανάλυση των δεδομένων αποτελούν ίσως τον πιο σημαντικό παράγοντα για τη λήψη μίας ορθολογικής απόφασης σε έναν εξαιρετικά πολύπλοκο κόσμο ο οποίος χαρακτηρίζεται από αβεβαιότητα και που η ανθρώπινη διαίσθηση έχει μεγάλη πιθανότητα να αποτύχει. Πλέον κράτη, μεγάλοι οργανισμοί και εταιρείες βασίζονται τις αποφάσεις τους στη στατιστική ανάλυση δεδομένων και στην εξαγωγή έγκυρων συμπερασμάτων από αυτή. Στο πλαίσιο αυτό η παρούσα διπλωματική εργασία πραγματεύεται ένα σημαντικό κομμάτι της ανάλυσης δεδομένων : τον έλεγχο υποθέσεων για την εκθετική κατανομή.

Μέσα από αυτήν την διπλωματική εργασία , θα καταλάβουμε τη χρησιμότητα της μελέτης της εκθετικής κατανομής σε ποικίλους επιστημονικούς κλάδους (μηχανική , φυσική , οικονομικά κ.α.) , θα αποκτήσουμε μία διαισθητική αντίληψη για την εκθετική κατανομή και θα κατανοήσουμε βασικές έννοιες για τους ελεγχους υποθέσεων.

Θα περιγράψουμε αναλυτικά τα βήματα για την υλοποίηση συγκεκριμένων ελέγχων υποθέσεων που έχουν προταθεί για την εκθετική κατανομή, χρησιμοποιώντας το στατιστικό πακέτο R και το πρόγραμμα Excel σε προσομοιωμένα δεδομένα

Τέλος, θα χρησιμοποιήσουμε αυτά που περιγράψαμε πριν για την επεξεργασία και ανάλυση δεδομένων από τον πραγματικό κόσμο. Πιο συγκεκριμένα, θα εξετάσουμε δεδομένα που αφορούν μεγάλους σεισμούς ανά τον κόσμο.

ΚΕΦΑΛΑΙΟ 1

Ιστορία και χρησιμότητα της εκθετικής κατανομής

1.1 Περιγραφή εκθετικής κατανομής

Η εκθετική κατανομή είναι μια συνεχής κατανομή πιθανότητας που χρησιμοποιείται για τη μοντελοποίηση του χρόνου μεταξύ των γεγονότων σε μια διαδικασία Poisson, μια διαδικασία στην οποία τα γεγονότα συμβαίνουν συνεχώς και ανεξάρτητα με σταθερό μέσο ρυθμό. Η εκθετική κατανομή έχει μελετηθεί και χρησιμοποιηθεί ευρέως σε διάφορους τομείς, όπως οι τηλεπικοινωνίες, η επιστήμη των υπολογιστών, η μηχανική και η φυσική. Μια σημαντική ιδιότητα σχετικά με την εκθετική κατανομή είναι η ιδιότητα απώλειας μνήμης, η οποία δηλώνει ότι η πιθανότητα ενός γεγονότος να συμβεί σε μια συγκεκριμένη χρονική στιγμή είναι ανεξάρτητη από το χρόνο που έχει περάσει από το τελευταίο συμβάν. Αυτή η ιδιότητα είναι χρήσιμη για τη μοντελοποίηση καταστάσεων στις οποίες ο ρυθμός εμφάνισης των γεγονότων είναι σταθερός με την πάροδο του χρόνου. Η εκθετική κατανομή έχει χρησιμοποιηθεί ευρέως σε διάφορα πεδία για τη μοντελοποίηση του χρόνου μεταξύ των γεγονότων.

1.2 Χρησιμότητα εκθετικής κατανομής

Στη διαχείριση ουρών αναμονής :

η εκθετική κατανομή χρησιμοποιείται συχνά για να μοντελοποιήσει τους χρόνους μεταξύ των αφίξεων μεταξύ των πελατών που φτάνουν σε μια ουρά, καθώς και τους χρόνους εξυπηρέτησης των πελατών μόλις φτάσουν στο μπροστινό μέρος της ουράς. Η ουρά ενός διακομιστή είναι η ουρά όπου εξυπηρετείται ένας πελάτης κάθε φορά. Ο αριθμός των πελατών στην ουρά ακολουθεί κατανομή Poisson και ο χρόνος εξυπηρέτησης μοντελοποιείται με εκθετική κατανομή. Ο χρόνος που ξοδεύει ένας πελάτης στο σύστημα (δηλαδή ο χρόνος από την άφιξη έως την αναχώρηση) μοντελοποιείται ως το άθροισμα του χρόνου μεταξύ άφιξης και του χρόνου εξυπηρέτησης, που είναι και οι δύο εκθετικές τυχαίες μεταβλητές. Αναλύοντας την εκθετική κατανομή των χρόνων εξυπηρέτησης, μπορούν να υπολογιστούν διάφορα μέτρα απόδοσης του συστήματος όπως η χρήση, ο μέσος αναμονής, ο αριθμός των πελατών στην ουρά και ο χρόνος απασχόλησης.

Στις τηλεπικοινωνίες :

Η εκθετική κατανομή χρησιμοποιείται συχνά για τη μοντελοποίηση του χρόνου μεταξύ των βλαβών του εξοπλισμού ή των συστημάτων, καθώς και του χρόνου που απαιτείται για την επισκευή ή την αποκατάσταση από βλάβες. Στα τηλεπικοινωνιακά συστήματα, η εκθετική κατανομή χρησιμοποιείται επίσης για τη μοντελοποίηση της διάρκειας μιας κλήσης, που είναι ο χρόνος που ένας πελάτης είναι συνδεδεμένος στο δίκτυο. Η διάρκεια της κλήσης ακολουθεί την εκθετική κατανομή σε πολλές περιπτώσεις.

Στην επιστήμη των υπολογιστών :

Μια κοινή χρήση της εκθετικής κατανομής στην επιστήμη των υπολογιστών είναι στη μοντελοποίηση των χρόνων μεταξύ αφίξεων γεγονότων, όπως αιτήματα σε διακομιστή ή πακέτα που φτάνουν σε έναν κόμβο δικτύου.

Μια άλλη εφαρμογή της εκθετικής κατανομής στην επιστήμη των υπολογιστών είναι η ανάλυση αλγορίθμων. Για παράδειγμα, η εκθετική κατανομή μπορεί να χρησιμοποιηθεί για τη μοντελοποίηση του χρόνου εκτέλεσης ενός αλγορίθμου, όπου η παράμετρος-ρυθμός αντιπροσωπεύει τον μέσο χρόνο που χρειάζεται για να ολοκληρωθεί ο αλγόριθμος. Αυτό χρησιμοποιείται συχνά στην ανάλυση μέσης περίπτωσης, όπου ο στόχος είναι να κατανοήσουμε πώς αποδίδει ο αλγόριθμος κατά μέσο όρο, παρά στα χειρότερα ή καλύτερα σενάρια.

Η εκθετική κατανομή χρησιμοποιείται επίσης στη δοκιμή φορτίου (load testing), η οποία είναι η διαδικασία δοκιμής ενός συστήματος κάτω από ένα προσομοιωμένο μεγάλο φορτίο (heavy load) προκειμένου να αξιολογηθεί η απόδοσή του και να εντοπιστούν τα σημεία συμφόρησης.

Στη μηχανική :

Μια κοινή χρήση της εκθετικής κατανομής στη μηχανική είναι στη μηχανική αξιοπιστίας, η οποία είναι η μελέτη του χρόνου που μπορεί να λειτουργήσει ένα σύστημα ή ένα εξάρτημα πριν από την αστοχία. Ο χρόνος μεταξύ των αστοχιών ενός συστήματος ή ενός στοιχείου μοντελοποιείται συχνά ως εκθετική τυχαία μεταβλητή. Αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν για την πρόβλεψη της διάρκειας ζωής ενός συστήματος ή ενός εξαρτήματος και για τη λήψη αποφάσεων σχετικά με τη συντήρηση, την αντικατάσταση και το σχεδιασμό.

Μια άλλη εφαρμογή της εκθετικής κατανομής στη μηχανική είναι στα συστήματα ελέγχου, τα οποία χρησιμοποιούνται για τον έλεγχο της συμπεριφοράς διαφόρων φυσικών συστημάτων, όπως αεροσκαφών, αυτοκινήτων και βιομηχανικών διεργασιών. Η εκθετική κατανομή χρησιμοποιείται για τη μοντελοποίηση της σταθεράς χρόνου ενός συστήματος ελέγχου, η οποία αντιπροσωπεύει το χρόνο που χρειάζεται για να ανταποκριθεί το σύστημα σε μια αλλαγή βήματος στην είσοδο.

Στη φυσική :

Μοντελοποίηση της διάσπασης ασταθών σωματιδίων ή ραδιενεργών ισotόπων. Η πιθανότητα ότι ένα ασταθές σωματίδιο θα διασπαστεί μέσα σε ένα δεδομένο χρονικό διάστημα δίνεται από την εκθετική κατανομή, με τον ρυθμό διάσπασης (επίσης γνωστό ως χρόνος ημιζωής) του σωματιδίου να είναι η παράμετρος της κατανομής.

Εκτός από τις διαδικασίες διάσπασης, η εκθετική κατανομή χρησιμοποιείται επίσης για τη μοντελοποίηση άλλων φαινομένων στη φυσική, όπως η κατανομή σωματιδίων στην αέρια φάση ενός συστήματος στη στατιστική μηχανική και η κατανομή της διάρκειας ζωής των σωματιδίων που παράγονται σε συγκρούσεις υψηλής ενέργειας. Μια άλλη εφαρμογή της εκθετικής κατανομής στη φυσική είναι στη μελέτη της κίνησης Brown, που είναι η τυχαία κίνηση των σωματιδίων σε ένα ρευστό. Σε αυτό το πλαίσιο, η εκθετική κατανομή χρησιμοποιείται για τη μοντελοποίηση του χρόνου μεταξύ των συγκρούσεων ενός σωματιδίου με άλλα σωματίδια στο ρευστό. Επίσης στη γεωφυσική είναι ιδιαίτερα χρήσιμη στον κλάδο της σεισμολογίας για τη μοντελοποίηση του χρόνου μεταξύ σεισμών.

Είναι σημαντικό να σημειωθεί ότι ενώ η εκθετική κατανομή παρέχει ένα καλό μοντέλο για πάρα πολλά συστήματα, μπορεί να μην είναι ακριβής για άλλα, ανάλογα με τα χαρακτηριστικά και τις υποθέσεις του συστήματος και πολλές φορές να μην ανταποκρίνεται επαρκώς σε πιο περίπλοκα συστήματα της πραγματικής ζωής. Σε τέτοιες περιπτώσεις μπορούν να χρησιμοποιηθούν άλλες κατανομές όπως υπερεκθετικές, κατανομές Erlang, κατανομή Weibull.

1.3 Ιστορία εκθετικής κατανομής

Η εκθετική κατανομή είναι μια θεμελιώδης έννοια στη θεωρία πιθανοτήτων και τη στατιστική, και η προέλευση της μπορεί να εντοπιστεί πίσω στον 18ο αιώνα με το έργο του Daniel Bernoulli, ο οποίος ήταν από τους πρώτους που μελέτησε το πρόβλημα της εξαγωγής της κατανομής πιθανότητας του χρόνου μεταξύ των γεγονότων σε μια διαδικασία Poisson. Η διαδικασία Poisson, η οποία είναι ένας τύπος διαδικασίας στην οποία τα γεγονότα συμβαίνουν ανεξάρτητα με σταθερό μέσο ρυθμό, είναι μια βασική έννοια στην ανάπτυξη της εκθετικής κατανομής.

Η εκθετική κατανομή εισήχθη για πρώτη φορά από τον Γάλλο μαθηματικό Siméon Denis Poisson το 1837, στο έργο του "Recherches sur la probabilité des jugements en matière criminelle et en matière civile" (Έρευνα σχετικά με την πιθανότητα δικαστικών αποφάσεων σε ποινικές και αστικές υποθέσεις). Ο Poisson χρησιμοποίησε την εκθετική κατανομή για να μοντελοποιήσει το χρόνο

μεταξύ σπάνιων γεγονότων, όπως ο χρόνος μεταξύ των ατυχημάτων σε ένα εργοστάσιο ή ο χρόνος μεταξύ των σφαλμάτων στη μετάδοση ενός μηνύματος.

Ήταν ο Siméon Denis Poisson και ταυτόχρονα ο Pierre Simon Laplace που ανέπτυξαν περαιτέρω τη θεωρία των διαδικασιών Poisson και καθιέρωσαν τη σύνδεση μεταξύ της κατανομής Poisson και της εκθετικής κατανομής. Αναγνώρισαν ότι ο χρόνος μεταξύ των γεγονότων σε μια διαδικασία Poisson ακολουθεί μια εκθετική κατανομή και χρησιμοποίησαν αυτό το αποτέλεσμα για να εξαγάγουν την κατανομή πιθανότητας του αριθμού των γεγονότων σε μια δεδομένη χρονική περίοδο. Το έργο του Poisson και του Laplace έθεσε τα θεμέλια για τη σύγχρονη κατανόηση της εκθετικής κατανομής.

Επομένως, μπορεί να ειπωθεί ότι η εκθετική κατανομή επινοήθηκε από τον Daniel Bernoulli και αναπτύχθηκε περαιτέρω από τους Siméon Denis Poisson και Pierre-Simon Laplace τον 18ο και 19ο αιώνα, και η συνεισφορά τους συνεχίζει να αναγνωρίζεται ως σημαντικό ορόσημο αφενός στη θεωρία πιθανοτήτων και αφετέρου στη στατιστική.

Λίγα χρόνια αργότερα, το 1845, ένας Βρετανός αστρονόμος και μαθηματικός, ο John Couch Adams, χρησιμοποίησε την εκθετική κατανομή για να μοντελοποιήσει την κατανομή των σφαλμάτων στις αστρονομικές παρατηρήσεις. Το 1875, ο Γερμανός μαθηματικός Friedrich Bessel ερμήνευσε ανεξάρτητα την εκθετική κατανομή για τον ίδιο σκοπό. Το έργο του Bessel ακολούθησε το έργο πολλών άλλων μαθηματικών, συμπεριλαμβανομένου του Γάλλου μαθηματικού Paul Lévy, του Γερμανού μαθηματικού Richard von Mises και του Αμερικανού μαθηματικού Leonard Savage, οι οποίοι συνέβαλαν σημαντικά στα θεωρητικά θεμέλια της εκθετικής κατανομής.

Η εκθετική κατανομή μελετήθηκε εκτενώς από τον Paul Lévy στις αρχές του 20ου αιώνα. Ο Lévy χρησιμοποίησε τη κατανομή για να μελετήσει το χρόνο μεταξύ των αφίξεων τηλεγραφημάτων σε ένα ταχυδρομείο. Ο Lévy γεννήθηκε στο Παρίσι το 1886 και πέθανε το 1971 στην ίδια πόλη. Ήταν μέλος της Γαλλικής Ακαδημίας Επιστημών και τιμήθηκε με το Grand Prix de l'Académie des Sciences το 1948.

Ο Paul Lévy ήταν Γάλλος μαθηματικός που συνέβαλε σημαντικά στον τομέα της θεωρίας πιθανοτήτων. Είναι γνωστός για το έργο του στη θεωρία των στοχαστικών διαδικασιών, συμπεριλαμβανομένης της διαδικασίας Lévy, η οποία είναι μια στοχαστική διαδικασία συνεχούς χρόνου με σταθερές και ανεξάρτητες

προσαυξήσεις. Ο Λένυ είναι επίσης γνωστός για την εργασία του στην εκθετική κατανομή. Έχει διακριθεί για τη γενίκευση της εκθετικής κατανομής, για την κατηγορία κατανομών Levy. Συνοπτικά, ενώ η εκθετική κατανομή είναι μια συγκεκριμένη κατανομή πιθανότητας, η κατανομή Λένυ είναι μια γενίκευση της εκθετικής κατανομής που περιλαμβάνει μια ευρύτερη κατηγορία κατανομών.

1.4 Βασικοί ορισμοί στη Στατιστική

Πληθυσμός

Ορίζεται σαν ένα βασικό σύνολο στοιχείων που πρόκειται να μελετηθεί ως προς μία ή περισσότερες χαρακτηριστικές ιδιότητες. Είναι το σύνολο δηλαδή των απλών στοιχείων για τα οποία μας ενδιαφέρει να εξάγουμε κάποια συμπεράσματα.

Μεταβλητές

Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε ένα πληθυσμό λέγονται μεταβλητές. Τις μεταβλητές τις διακρίνουμε σε ποσοτικές και σε ποιοτικές ή κατηγορικές. Οι ποσοτικές μεταβλητές διακρίνονται ανάλογα με το πλήθος των τιμών που μπορούν να πάρουν, σε διακριτές και συνεχείς.

Διακριτή θεωρείται η μεταβλητή που λαμβάνει ένα σύνολο τιμών στις οποίες μπορούν να αντιστοιχισθούν ένα προς ένα, στοιχεία του συνόλου των φυσικών αριθμών. Αντίθετα, μία μεταβλητή χαρακτηρίζεται ως συνεχής εάν παίρνει, θεωρητικά τουλάχιστον, κάθε τιμή στο διάστημα (α, β) . Ποιοτικές μεταβλητές είναι εκείνες για τις οποίες δεν είναι δυνατόν να ληφθούν πλήρως καθοριστικές τιμές αλλά μόνο ενδεικτικές ή συγκριτικές.

Δείγμα

Είναι ένα σύνολο μονάδων που επιλέγονται από τον πληθυσμό. Το δείγμα επιλέγεται για να μελετηθεί και να εξαχθούν έγκυρα συμπεράσματα για τον πληθυσμό, χωρίς να μελετηθεί όλος ο πληθυσμός γιατί είναι πολύ μεγάλος ή πολυέξοδη και χρονοβόρα διαδικασία η συλλογή πληροφοριών από όλο τον

πληθυσμό. Το δείγμα πρέπει να είναι αντιπροσωπευτικό του πληθυσμού και γι' αυτό παίζει ρόλο το μέγεθός του και η μέθοδος δειγματοληψίας που χρησιμοποιείται.

Πείραμα

Είναι οποιαδήποτε διαδικασία ή μελέτη, η οποία έχει ως αποτέλεσμα τη συλλογή των δεδομένων, η έκβαση της οποίας είναι άγνωστη.

Συνάρτηση πυκνότητας πιθανότητας

Έστω ότι X είναι μία συνεχής τυχαία μεταβλητή. Αν υπάρχει μία μη αρνητική συνάρτηση f ορισμένη για κάθε πραγματικό αριθμό $X \in R$ η οποία έχει την εξής ιδιότητα : για κάθε σύνολο B με πραγματικούς αριθμούς ισχύει

$$P[X \in B] = \int_B f(x) dx$$

τότε η συνάρτηση f ονομάζεται συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X .

1.5 Βασικοί ορισμοί στην Εκτιμητική

Θεωρία εκτίμησης

Ο κλάδος της στατιστικής που ασχολείται με την εκτίμηση των παραμέτρων με βάση δεδομένα που έχουμε από το δείγμα. Οι παράμετροι περιγράφουν μια υποκείμενη φυσική ρύθμιση με τέτοιο τρόπο ώστε οι τιμές τους να επηρεάζουν την κατανομή των δεδομένων μέτρησης. Μία εκτιμήτρια συνάρτηση, η οποία είναι συνάρτηση ενός τυχαίου δείγματος X_1, X_2, \dots, X_n από κάποια κατανομή F , επιχειρεί να χρησιμοποιήσει τις μετρήσεις για να προσεγγίσει τις άγνωστες παραμέτρους.

Εκτιμήτρια συνάρτηση

Μια συνάρτηση των γνωστών δεδομένων που χρησιμοποιείται για την εκτίμηση μιας άγνωστης παραμέτρου. Μια εκτίμηση είναι το αποτέλεσμα της πραγματικής εφαρμογής της συνάρτησης σε ένα συγκεκριμένο σύνολο δεδομένων. Για παράδειγμα, ο μέσος όρος μπορεί να χρησιμοποιηθεί ως εκτιμητής της μέσης τιμής.

Αναμενόμενη τιμή

Το άθροισμα (διακριτό ή συνεχές) των πιθανοτήτων κάθε πιθανού αποτελέσματος ενός πειράματος πολλαπλασιασμένο με την αντίστοιχη τιμή του. Η έννοια είναι, διαισθητικά, μια γενίκευση του σταθμισμένου μέσου όρου όλων των πιθανών αποτελεσμάτων μιας συγκεκριμένης διαδικασίας ή πειράματος και μπορεί να θεωρηθεί ως ο αριθμητικός μέσος όρος ενός μεγάλου αριθμού ανεξάρτητων πραγματοποιήσεων του πειράματος. Η αναμενόμενη τιμή της τυχαίας μεταβλητής X γράφεται συνήθως ως $E(X)$.

1.6 Έλεγχοι υποθέσεων

1.6.1 Ιστορία

Ενώ ο έλεγχος υποθέσεων έγινε ευρέως γνωστός στις αρχές του 20ου αιώνα, οι πρώιμες μορφές χρησιμοποιήθηκαν το 1700. Η πρώτη χρήση αποδίδεται στον John Arbuthnot (1710) και ακολούθησε ο Pierre-Simon Laplace (δεκαετία 1770), στην ανάλυση της αναλογίας των ανθρώπινων φύλων κατά τη γέννηση.

1.6.2 Βασικοί Ορισμοί

Στατιστική υπόθεση (statistical hypothesis)

Ονομάζεται κάθε υπόθεση που αφορά την κατανομή μιας τυχαίας μεταβλητής X . Συγκεκριμένα η υπόθεση (συμβολιζόμενη συνήθως με H) είναι δυνατόν να αφορά (α) την άγνωστη παράμετρο έστω θ (ή μια συνάρτηση του θ) της κατανομής της τυχαίας μεταβλητής X ή (β) τη συναρτησιακή μορφή της κατανομής της τυχαίας μεταβλητής. Στην πρώτη περίπτωση έχουμε πρόβλημα ελέγχου παραμετρικής υπόθεσης (parametric test) ενώ στη δεύτερη έχουμε πρόβλημα ελέγχου καταλληλότητας ή καλής προσαρμογής κατανομής (goodness of fit test).

Εναλλακτική υπόθεση (alternative hypothesis H_1)

Λαμβάνεται να είναι η άποψη που έχει διατυπωθεί, ενώ η μηδενική υπόθεση (null hypothesis H_0) λαμβάνεται να είναι η άρνηση της άποψης που έχει διατυπωθεί.

Σφάλμα τύπου I (type I error)

Διαπράττουμε σφάλμα τύπου I όταν απορρίπτουμε την H_0 ενώ είναι στην πραγματικότητα αληθής.

Σφάλμα τύπου II (type II error)

Διαπράττουμε σφάλμα τύπου II όταν αποδεχόμαστε την H_0 ενώ στην πραγματικότητα η H_1 είναι αληθής.

Επίπεδο σημαντικότητας (level of significance)

Το επίπεδο σημαντικότητας ενός στατιστικού ελέγχου υποθέσεων είναι μια σταθερή πιθανότητα (η οποία προεπιλέγεται) λανθασμένης απόρριψης της μηδενικής υπόθεσης H_0 , όταν αυτή είναι όντως αληθινή.

Είναι η πιθανότητα σφάλματος τύπου I και καθορίζεται από τον ερευνητή σε σχέση με τις συνέπειες ενός τέτοιου σφάλματος. Δηλαδή, θέλουμε να καταστήσουμε το επίπεδο σημαντικότητας όσο το δυνατόν μικρότερο, προκειμένου να προστατευθεί η μηδενική υπόθεση και να αποτρέψει, στο μέτρο του δυνατού, τον εξεταστή από ακούσιους ψευδείς ισχυρισμούς. Το επίπεδο σημαντικότητας συνήθως συμβολίζεται με α .

Επίπεδο σημαντικότητας = P (σφάλμα τύπου I) = α . Συνήθως, το επίπεδο σημαντικότητας επιλέγεται να είναι 0.05 (ή ισοδύναμα, 5%).

Στατιστική ελεγχουσυνάρτηση (test statistic)

Ονομάζεται η στατιστική συνάρτηση η οποία χρησιμοποιείται για τον έλεγχο της υπό μελέτην υπόθεσης.

Χωρίο απορρίψεως (Rejection Region)

Το χωρίο απορρίψεως ή αλλιώς κρίσιμη περιοχή είναι ένα σύνολο τιμών της στατιστικής συνάρτησης, για το οποίο η μηδενική υπόθεση απορρίπτεται σε έναν έλεγχο υποθέσεων. Πιο συγκεκριμένα αν ο χώρος τιμών της στατιστικής συνάρτησης χωριστεί σε δύο περιοχές τότε η μια θα οδηγεί σε απόρριψη της μηδενικής υπόθεσης ενώ η άλλη σε αποδοχή. Έτσι, αν η παρατηρούμενη τιμή της στατιστικής συνάρτησης ανήκει στην κρίσιμη περιοχή τότε υποδηλώνεται ότι απορρίπτουμε την μηδενική υπόθεση. Αλλιώς δεν απορρίπτουμε την μηδενική υπόθεση.

Ισχύς Στατιστικού ελέγχου (Power of a Test)

Η ισχύς ενός στατιστικού ελέγχου υποθέσεων μετρά την ικανότητα του τεστ να απορρίπτει τη μηδενική υπόθεση όταν είναι στην πραγματικότητα ψευδής -

δηλαδή, να πάρει μια σωστή απόφαση. Με άλλα λόγια, η ισχύς ενός ελέγχου υποθέσεων είναι η πιθανότητα να μην συμβαίνει ένα σφάλμα τύπου II.

Υπολογίζεται αφαιρώντας την πιθανότητα σφάλματος τύπου II από τη μονάδα και συνήθως εκφράζεται ως: $\text{Ισχύς} = 1 - P(\text{σφάλματος τύπου II}) = 1 - \beta$. Η μέγιστη ισχύς που μπορεί να έχει ένας έλεγχος είναι 1, το ελάχιστο είναι 0. Ιδανικά θέλουμε έναν έλεγχο να έχει μεγάλη ισχύ, κοντά στο 1, ειδικά όσο απομακρυνόμαστε από την μηδενική υπόθεση.

P-Value ή πιθανότητα σημαντικότητας (significance probability)

Είναι η μικρότερη τιμή του επιπέδου σημαντικότητας α για την οποία η μηδενική υπόθεση H_0 απορρίπτεται. Αν η $p\text{-value} < \alpha$ απορρίπτουμε την H_0 . Γενικά όσο μικρότερη είναι η $p\text{-value}$ ενός ελέγχου τόσο ισχυρότερα απορρίπτουμε την μηδενική υπόθεση.

Έλεγχοι καλής προσαρμογής (goodness of fit tests)

Τέτοιου τύπου έλεγχοι είναι οι έλεγχοι που προσπαθούν να απαντήσουν σε υποθέσεις με ερώτημα: «αποτελούν οι παρατηρήσεις μας ένα δείγμα από κάποια συγκεκριμένη κατανομή;». Σχεδιάζονται για να συγκρίνουν το δείγμα με τον τύπο του δείγματος που θα περιμέναμε να έχουμε από την κατανομή που υποθέτουμε κάτω από τη μηδενική υπόθεση, ώστε να μπορούμε να δούμε αν η υποτεθείσα συνάρτηση κατανομής προσαρμόζεται στα δεδομένα του δείγματος. Ο έλεγχος αυτός χρησιμοποιείται σε περιπτώσεις προβλημάτων στα οποία ενδιαφερόμαστε να εξετάσουμε αν τα δεδομένα προέρχονται από μια ορισμένη κατανομή, δηλαδή, για τον έλεγχο υποθέσεων της μορφής

$$H_0: F_x(x) = F_0(x)$$

$$H_1: F_x(x) \neq F_0(x)$$

όπου $F_x(x) = P(X \leq x)$ είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής X , και $F_0(x)$ γνωστή συνάρτηση κατανομής.

Ο αρχαιότερος και γνωστότερος τέτοιος έλεγχος καλής προσαρμογής είναι ο έλεγχος χ^2 ο οποίος προτάθηκε από τον Pearson το 1900. Αργότερα, ο Kolmogorov (1933, 1941) επινόησε έναν εναλλακτικό έλεγχο για τον ίδιο σκοπό και ο Smirnov (1939, 1948) τον επέκτεινε για την περίπτωση του ελέγχου της υπόθεσης ότι δύο ανεξάρτητα δείγματα μπορούν να υποτεθούν ότι προέρχονται από την ίδια κατανομή.

1.6.3 Ιδιότητες Στατιστικών ελέγχων

Για να είναι ένας έλεγχος κατάλληλος, ή τουλάχιστον καταλληλότερος από τους υπόλοιπους, θα πρέπει να ικανοποιεί κάποια κριτήρια σε ένα βαθμό. Τέτοια κριτήρια είναι τα εξής :

- Αμεροληψία: η ιδιότητα, η αναμενόμενη τιμή της δειγματοληπτικής κατανομής μιας εκτιμήτριας να είναι ίση με την άγνωστη πραγματική τιμή της παραμέτρου του πληθυσμού.
- Συνέπεια : Η συνέπεια αναφέρεται στη λογική και την αριθμητική συνοχή. Μια εκτιμήτρια ονομάζεται συνεπής εάν συγκλίνει κατά πιθανότητα στην προς εκτίμηση παράμετρο καθώς αυξάνεται το μέγεθος του δείγματος.
- Αποδοτικότητα: η αποδοτικότητα είναι ένα μέτρο της ποιότητας μιας εκτιμήτριας, ενός πειραματικού σχεδιασμού ή μιας διαδικασίας ελέγχου υποθέσεων. Ουσιαστικά, μια πιο αποδοτική εκτιμήτρια χρειάζεται λιγότερα δεδομένα εισόδου ή παρατηρήσεις από μια λιγότερο αποδοτική για να επιτύχει το επιθυμητό όριο σφάλματος.

Στην παρούσα εργασία θα ασχοληθούμε με τον έλεγχο καλής προσαρμογής της εκθετικής κατανομής. Θα αναφέρουμε και θα αναλύσουμε διάφορους ελέγχους καλής προσαρμογής που έχουν προταθεί για την εκθετική κατανομή και θα εξετάσουμε – δοκιμάσουμε κάποιους από αυτούς σε ένα σετ δεδομένων, για να ελεγχθεί η υπόθεση αν το συγκεκριμένο σετ ακολουθεί την εκθετική κατανομή.

ΚΕΦΑΛΑΙΟ 2

Έλεγχοι καλής προσαρμογής για την εκθετική κατανομή – Ιστορικά στοιχεία

2.1 Εισαγωγή

Στη Στατιστική, ο έλεγχος υποθέσεων για την εκθετική κατανομή (Exponentiality Test) αποτελεί ένα πολύ σημαντικό έλεγχο υποθέσεων και χρησιμοποιείται για να εξαχθεί το συμπέρασμα αν ένα σύνολο δεδομένων ακολουθεί την εκθετική κατανομή ή όχι. Η εκθετική κατανομή παίζει βασικό ρόλο στην ανάλυση επιβίωσης. Οι έλεγχοι καλής προσαρμογής για την εκθετική κατανομή έχουν καθοριστικό ρόλο σε διάφορους επιστημονικούς τομείς, καθώς στις φυσικές επιστήμες, ορισμένες από τις κοινώς χρησιμοποιούμενες κατανομές, όπως οι κατανομές Γάμμα και Weibull, είναι μεταφρασμένες εκδόσεις των εκθετικών κατανομών.

Πριν αναφερθούμε σε συγκεκριμένους ελέγχους πρέπει να αναφέρουμε δύο σημαντικούς γενικούς ελέγχους καλής προσαρμογής που μας επιτρέπουν να ελέγχουμε αν κάποια δεδομένα ακολουθούν μία συγκεκριμένη κατανομή (όχι αναγκαστικά την εκθετική):

α) Οι γραφικοί έλεγχοι P-P plot (Probability-Probability plot ή Percent-Percent plot ή P value plot) και Q-Q plot (Quantile-Quantile plot). Είναι δύο γραφήματα τα οποία μας βοηθούν να ελέγξουμε αν κάποια δεδομένα προέρχονται από κάποια συγκεκριμένη κατανομή (π.χ. από την εκθετική).

β) Ο έλεγχος καλής προσαρμογής X^2 (έλεγχος για οποιαδήποτε συνεχή ή διακριτή κατανομή). Είναι ο αρχαιότερος και περισσότερο γνωστός έλεγχος καλής προσαρμογής και προτάθηκε το 1900 από τον Karl Pearson.

Κάποιοι έλεγχοι είναι αποτελεσματικότεροι για δεδομένα που ακολουθούν φαινόμενα – συναρτήσεις με μονότονες συναρτήσεις κινδύνου (monotonic

hazard rates) ενώ άλλοι για δεδομένα που ακολουθούν μη μονότονες συναρτήσεις κινδύνου (non monotonic hazard rates). Παρακάτω ορίζουμε τα δύο είδη συναρτήσεων κινδύνου, και αναφέρουμε τους κύριους ελέγχους εκθετικής κατανομής που αντιστοιχούν σε κάθε κατηγορία με ιστορικά στοιχεία. Επισημαίνεται πως ο κάθε έλεγχος γενικότερα, δεν είναι κατάλληλος για όλες τις μορφές δεδομένων και για αυτό το λόγο καλή πρακτική είναι κάθε έλεγχος να χρησιμοποιείται σε συνδυασμό με άλλους ελέγχους για την πληρέστερη ανάλυση των δεδομένων και τα πιο ακριβή συμπεράσματα.

2.2 Μονότονες συναρτήσεις κινδύνου (Monotonic hazard rates)

Η πιθανότητα να συμβεί ένα γεγονός, αυξάνεται όσο περνάει ο χρόνος. Παράδειγμα : Έστω η θανατηφόρα ασθένεια A. Από τη στιγμή που ένα άτομο διαγνωσθεί με αυτή, δεδομένου πως όποιος έχει αυτή την ασθένεια πεθαίνει μέσα σε ένα χρονικό διάστημα, όσο περνάει ο χρόνος η πιθανότητα θανάτου του αυξάνεται.

2.2.1 KSL Test (Kolmogorov – Smirnov – Lilliefors Test)

Ο **Andrey Kolmogorov** (1930) ήταν εξέχων Ρώσος μαθηματικός και πιθανοθεωρητικός. Θεωρείται ευρέως ως ένας από τους πιο σημαντικούς μαθηματικούς του 20ου αιώνα και συνέβαλε σημαντικά σε μια σειρά πεδίων, όπως η θεωρία πιθανοτήτων, η τοπολογία, και η μαθηματική λογική.

Ο Kolmogorov έλαβε το προπτυχιακό του πτυχίο από το Πανεπιστήμιο της Μόσχας το 1925 και το διδακτορικό του από το ίδιο ίδρυμα το 1929.

Ο **Nikolai Smirnov** (1900) ήταν επίσης ένας εξέχων Ρώσος μαθηματικός. Είναι περισσότερο γνωστός για τη συμβολή του στη θεωρία πιθανοτήτων, τη μαθηματική στατιστική και τη θεωρία των στοχαστικών διαδικασιών.

Ο Smirnov σπούδασε στο Κρατικό Πανεπιστήμιο της Μόσχας, όπου πήρε το πτυχίο του στα μαθηματικά το 1924. Συνέχισε τις σπουδές του στο ίδιο πανεπιστήμιο, απέκτησε το διδακτορικό του το 1929.

Ο **Hubert Lilliefors** (1918) ήταν Αμερικανός στατιστικός. Είναι περισσότερο γνωστός για τη συμβολή του στην ανάπτυξη στατιστικών ελέγχων, ιδιαίτερα του τεστ Lilliefors, το οποίο χρησιμοποιείται για να αξιολογήσει εάν ένα δείγμα δεδομένων προέρχεται από μια κανονική κατανομή. Ο Lilliefors έλαβε το προπτυχιακό του πτυχίο από το Dartmouth College το 1940 και το διδακτορικό του στη στατιστική από το Πανεπιστήμιο της Βόρειας Καρολίνας στο Chapel Hill το 1951.

Ο έλεγχος KSL για την εκθετικότητα, γνωστός και ως έλεγχος Kolmogorov-Smirnov-Lilliefors, είναι ένας στατιστικός έλεγχος που χρησιμοποιείται για να προσδιοριστεί εάν ένα σύνολο δεδομένων ακολουθεί την εκθετική κατανομή. Ο έλεγχος προτάθηκε για πρώτη φορά από τους Ρώσους μαθηματικούς Andrey Kolmogorov και Nikolai Smirnov τη δεκαετία του 1930 και αργότερα τροποποιήθηκε από τον Αμερικανό στατιστικό Hubert Lilliefors το 1960 ("On the Kolmogorov Smirnov Test for the exponential distribution with mean unknown". JASA 64).

Ο έλεγχος KSL για την εκθετικότητα βασίζεται στον έλεγχο Kolmogorov-Smirnov, ο οποίος είναι ένας συνηθισμένος έλεγχος για την αξιολόγηση της καλής προσαρμογής μιας κατανομής. Η δοκιμή KSL τροποποιεί τη δοκιμή Kolmogorov-Smirnov προσαρμόζοντας τις κρίσιμες τιμές του στατιστικού ελέγχου για να ληφθεί υπόψη το μέγεθος του δείγματος και οι εκτιμώμενες παραμέτροι της εκθετικής κατανομής. Αυτή η τροποποίηση καθιστά τον έλεγχο KSL πιο ισχυρό από τον έλεγχο Kolmogorov-Smirnov για την ανίχνευση αποκλίσεων από την εκθετική κατανομή.

Από την ανάπτυξή του, ο έλεγχος KSL έχει γίνει ένα ευρέως χρησιμοποιούμενο εργαλείο στη στατιστική ανάλυση, ιδιαίτερα στους τομείς των οικονομικών, της μηχανικής και της ιατρικής. Ο έλεγχος έχει αποδειχθεί ότι έχει καλή ισχύ και ακρίβεια στην ανίχνευση αποκλίσεων από την εκθετική κατανομή σε ένα ευρύ φάσμα εφαρμογών.

2.2.2 The Gini Statistic

Ο **Joseph L. Gastwirth** (1940) είναι στατιστικός και καθηγητής στατιστικής στο Πανεπιστήμιο George Washington. Απέκτησε το διδακτορικό του στη στατιστική από το Πανεπιστήμιο του Stanford το 1965.

Ο Gastwirth έχει συνεισφέρει σημαντικά στη στατιστική θεωρία και πρακτική, ιδιαίτερα στους τομείς των στατιστικών ελέγχων και των στατιστικών μεθόδων για νομικά ζητήματα και ζητήματα δημόσιας πολιτικής. Εκτός από την εργασία του για την στατιστική συνάρτηση Gini, έχει δημοσιεύσει ευρέως σε θέματα όπως στατιστικά στοιχεία στο δικαστήριο, στατιστικές μεθόδους για τον εντοπισμό απάτης και χρήση στατιστικής στην περιβαλλοντική ρύθμιση.

Η στατιστική συνάρτηση Gini είναι ένα μέτρο ανισότητας που χρησιμοποιείται ευρέως στα οικονομικά, αλλά έχει βρει εφαρμογές και στη στατιστική και άλλους τομείς. Αυτή η στατιστική συνάρτηση πήρε το όνομά της από τον δημιουργό της, τον Ιταλό στατιστικό Corrado Gini. Οι Joseph L. Gastwirth και Gail M. Rosen πρότειναν έναν έλεγχο καλής προσαρμογής ανεξάρτητο της παραμέτρου κλίμακας της εκθετικής κατανομής με βάση τη στατιστική συνάρτηση Gini στην εργασία τους του 1978 στο *Journal of the Royal Statistical Society*. Η εργασία έδειξε ότι η στατιστική συνάρτηση Gini είναι μια χρήσιμη εναλλακτική λύση στο κλασικό έλεγχο καλής προσαρμογής χ^2 για την εκθετικότητα, καθώς δεν απαιτεί την υπόθεση μιας συγκεκριμένης παραμέτρου κλίμακας και ονομάστηκε έλεγχος Gini.

Έκτοτε, η στατιστική συνάρτηση Gini έχει χρησιμοποιηθεί σε διάφορες στατιστικές εφαρμογές, συμπεριλαμβανομένου του ελέγχου για ισότητα δύο πληθυσμών και για τη μέτρηση του βαθμού εξάρτησης μεταξύ των μεταβλητών.

2.2.3 The Lorenz Statistic

Η καμπύλη Lorenz και η σχετική στατιστική συνάρτηση Lorenz εισήχθησαν από τον Αμερικανό οικονομολόγο Max O. Lorenz το 1905. Ο Lorenz ενδιαφέρθηκε να αναπτύξει έναν τρόπο μέτρησης του βαθμού ανισότητας στην κατανομή του πλούτου και το έργο του έθεσε τα θεμέλια για το πεδίο της ανάλυσης της εισοδηματικής ανισότητας. Από την αρχική εργασία του Lorenz, η καμπύλη Lorenz και η στατιστική συνάρτηση Lorenz έχουν χρησιμοποιηθεί ευρέως στους τομείς της οικονομίας, της κοινωνιολογίας και της δημόσιας πολιτικής για τη μελέτη της ανισότητας εισοδήματος και πλούτου και για την αξιολόγηση της αποτελεσματικότητας διαφόρων πολιτικών που στοχεύουν στη μείωση της

ανισότητας. Ωστόσο, η στατιστική συνάρτηση Lorenz εφαρμόστηκε για πρώτη φορά στο πρόβλημα του ελέγχου της εκθετικότητας από τον **Joseph L. Gastwirth**

Το 1978, οι Gail και Gastwirth πρότειναν τη χρήση της στατιστικής Lorenz ως έλεγχο καλής προσαρμογής για την εκθετική κατανομή. Έδειξαν ότι κάτω από την μηδενική υπόθεση της εκθετικότητας, η στατιστική Lorenz έχει ασυμπτωτική κατανομή χ^2 με έναν βαθμό ελευθερίας. Αυτό σημαίνει ότι η στατιστική Lorenz μπορεί να χρησιμοποιηθεί για να ελεγχθεί η υπόθεση ότι ένα συγκεκριμένο σύνολο δεδομένων ακολουθεί μια εκθετική κατανομή. Οι Gail και Gastwirth παρείχαν επίσης πίνακες κρίσιμων τιμών για τη στατιστική Lorenz, οι οποίοι μπορούν να χρησιμοποιηθούν για τον προσδιορισμό του μεγέθους του ελέγχου.

2.2.4 The Pietra Statistic

Το στατιστικό ή δείκτης Pietra πήρε το όνομά του από τον Richard Pietra, ο οποίος το εισήγαγε 1979. Ο Pietra ήταν καθηγητής βιολογίας στο Πανεπιστήμιο του Ιλινόις και το έργο του επικεντρώθηκε στη χρήση στατιστικών μεθόδων στην οικολογική και περιβαλλοντική έρευνα. Ο δείκτης Pietra αναπτύχθηκε αρχικά ως μέτρο της διακριτικής δύναμης των διαγνωστικών τεστ στην ιατρική έρευνα, αλλά έκτοτε έχει προσαρμοστεί για χρήση σε άλλους τομείς, συμπεριλαμβανομένης της οικολογίας και της στατιστικής. Ο δείκτης είναι ένα ευρέως χρησιμοποιούμενο εργαλείο για την αξιολόγηση της ακρίβειας και της απόδοσης των διαγνωστικών ελέγχων και εξακολουθεί να είναι μια σημαντική έννοια στον τομέα της βιοστατιστικής.

Συγκεκριμένα, ο **Gastwirth** και οι συνεργάτες του έχουν εφαρμόσει τον δείκτη Pietra στην ανάλυση περιβαλλοντικών και οικολογικών δεδομένων, όπου έχει χρησιμοποιηθεί για την αξιολόγηση της αποτελεσματικότητας των προγραμμάτων περιβαλλοντικής παρακολούθησης και για την αξιολόγηση της επίδρασης των ρύπων σε διάφορα οικοσυστήματα.

2.2.5 Deshpande's Test

Ο **Jayant Deshpande** (1948) είναι ένας γνωστός Ινδός στατιστικός που έχει συμβάλει σημαντικά στον τομέα της ανάλυσης αξιοπιστίας. Έλαβε το διδακτορικό του στη στατιστική από το Πανεπιστήμιο Savitribai Phule Pune University το 1965. Η έρευνα του Deshpande έχει επικεντρωθεί σε διάφορες πτυχές της ανάλυσης αξιοπιστίας, συμπεριλαμβανομένων στατιστικών μεθόδων για ανάλυση δεδομένων διάρκειας ζωής, επιταχυνόμενων μοντέλων χρόνων ζωής και μοντελοποίησης ανάπτυξης αξιοπιστίας.

Ο έλεγχος *Deshpande's Test* προτάθηκε από τον Deshpande το 1983 για τον έλεγχο της εκθετικότητας κατανομών με αύξουσες συναρτήσεις κινδύνου ("A class of Tests for exponentiality against increasing failure rate average alternatives". *Biometrika* 70)

Ο έλεγχος του Deshpande είναι απλός, γρήγορος και δεν περιλαμβάνει πολύπλοκους υπολογισμούς, ωστόσο σύμφωνα με τους Dipankar Bandyopadhyay και Asit P. Basu ο έλεγχος έχει κάποιες παραλείψεις και στο άρθρο τους "A Note on Tests for Exponentiality by Deshpande" το 1989 προτείνουν ορισμένες βελτιώσεις. Ο έλεγχος έχει χρησιμοποιηθεί ευρέως στον τομέα της ανάλυσης αξιοπιστίας και έχει αποδειχθεί αποτελεσματικός σε πολλές εφαρμογές.

2.2.6 WE Test

Ο **Samuel Sanford Shapiro** (1930) ήταν ένας διακεκριμένος Αμερικανός στατιστικός. Ο Shapiro έλαβε το πτυχίο του από το City College της Νέας Υόρκης και το διδακτορικό του στη μαθηματική στατιστική από το Πανεπιστήμιο Rutgers το 1963. Η έρευνα του Shapiro επικεντρώθηκε στην ανάπτυξη στατιστικών μεθόδων για την ανάλυση δεδομένων και τον έλεγχο υποθέσεων. Είναι περισσότερο γνωστός για τη συν-ανάπτυξη της δοκιμής Shapiro-Wilk για

κανονικότητα με τον Martin Wilk το 1965. Η δοκιμή Shapiro-Wilk χρησιμοποιείται ευρέως στη στατιστική ανάλυση για να ελεγχθεί εάν ένα σύνολο δεδομένων ακολουθεί μια κανονική κατανομή.

Ο **Martin Wilk** (1922) ήταν ένας Αμερικανός στατιστικός που συνέβαλε σημαντικά στον τομέα της στατιστικής. Έλαβε το διδακτορικό του στη μαθηματική στατιστική από το Πανεπιστήμιο της Βόρειας Καρολίνας στο Chapel Hill το 1952. Η έρευνα του Wilk επικεντρώθηκε στην ανάπτυξη στατιστικών μεθόδων για την ανάλυση δεδομένων και τον έλεγχο υποθέσεων. Ο Wilk συνέβαλε σημαντικά στην ανάπτυξη στατιστικών μεθόδων για την ανάλυση κατηγορικών δεδομένων, την πολυμεταβλητή ανάλυση και τις μεθόδους σχεδίασης πιθανοτήτων. Στο βιβλίο του "Probability Plotting Methods for the Analysis of Data", πρότεινε την τροποποίηση του τεστ Shapiro-Wilk για κανονικότητα που θα μπορούσε να χρησιμοποιηθεί για να ελεγχθεί εάν ένα σύνολο δεδομένων ακολουθεί μια εκθετική κατανομή.

Ο έλεγχος WE για την εκθετικότητα προτάθηκε από τον Martin Wilk και τον Samuel Shapiro το 1972 ("An analysis of variance test for the exponential distribution", *Technometrics* 14). Στο άρθρο προτάθηκε μια τροποποίηση του ελέγχου Shapiro-Wilk για την κανονικότητα. Ο έλεγχος WE βασίζεται στο γεγονός ότι εάν ένα σύνολο δεδομένων ακολουθεί μια εκθετική κατανομή, τότε ο λογάριθμος των δεδομένων θα ακολουθεί μια κανονική κατανομή. Ο έλεγχος WE χρησιμοποιείται συνήθως σε τομείς όπως η μηχανική, τα οικονομικά και η βιολογία για να ελεγχθεί εάν τα δεδομένα ακολουθούν την εκθετική κατανομή.

2.2.7 New Better Than Used Test

Ο **Myles Hollander** (1941) είναι Αμερικανός ακαδημαϊκός στατιστικός, έλαβε το διδακτορικό του στη Στατιστική από το πανεπιστήμιο του Στάνφορντ το 1965. Έχει συνεισφέρει σημαντικά σε τομείς όπως η μη παραμετρική στατιστική, η βιοστατιστική και η θεωρία αξιοπιστίας. Είναι ομότιμος καθηγητής Στατιστικής στο Florida State University.

Ο **Frank Proschan** (1921) έλαβε το διδακτορικό του στα μαθηματικά από το Πανεπιστήμιο του Στάνφορντ το 1959. Εργάστηκε ως στατιστικός για διάφορους οργανισμούς, συμπεριλαμβανομένου του Εθνικού Γραφείου

Προτύπων. Ο Proschan είναι αρκετά γνωστός για την εργασία του στη θεωρία αξιοπιστίας.

Ο έλεγχος "New Better Than Used" (BTU) προτάθηκε από τον Hollander και τον Proschan το 1972 ("Testing whether new is better than used". *Annals of Math.Stat* 43) είναι ένας στατιστικός έλεγχος που χρησιμοποιείται κυρίως στον τομέα της μηχανικής αξιοπιστίας και της ανάλυσης επιβίωσης. Αυτά τα πεδία περιλαμβάνουν τη μελέτη του χρόνου μέχρι την αποτυχία ή του χρόνου μέχρι ένα συγκεκριμένο συμβάν.

Το όνομα " New Better Than Used " δίνει έμφαση στη σύγκριση μεταξύ των νέων και των ως τώρα χρησιμοποιούμενων δεδομένων. Ο όρος «New» αναφέρεται στις πιο πρόσφατες παρατηρήσεις ή δεδομένα που αναλύονται. Αντιπροσωπεύει τα δεδομένα από μια συγκεκριμένη χρονική περίοδο που ενδιαφέρει τον έλεγχο της εκθετικότητας. Ο όρος "Used" αναφέρεται σε ιστορικά ή προηγούμενα δεδομένα που χρησιμοποιούνται ως αναφορά για την αξιολόγηση του αναμενόμενου αριθμού γεγονότων στη νέα χρονική περίοδο. Αντιπροσωπεύει τα δεδομένα από προηγούμενες χρονικές περιόδους, συνήθως πριν από τα νέα δεδομένα ενδιαφέροντος.

2.2.8 EPS Test

Ο **Benjamin Epstein** (1918) έλαβε το διδακτορικό του στη Στατιστική από το Πανεπιστήμιο του Ιλινόις το 1941. Η κύρια συνεισφορά του Epstein είναι ο έλεγχος εκθετικότητας και η μαθηματική μοντελοποίηση για συστήματα ασφαλείας.

Ο έλεγχος EPS είναι ένας έλεγχος καλής προσαρμογής (goodness-of-fit test) που εισήχθη από τον Benjamin Epstein το 1960 («Tests for the Validity of the Assumption That the Underlying Distribution of Life Is Exponential». *Technometrics* 2) και εξετάστηκε από τους Fercho και Ringer το 1972 ("Small sample power of some tests of the constant failure rate", *Technometrics* 14). Ο έλεγχος EPS περιλαμβάνει τη σύγκριση των παρατηρούμενων χρόνων μεταξύ συμβάντων με τους αναμενόμενους χρόνους μεταξύ συμβάντων κάτω από την υπόθεση της εκθετικότητας. Το τεστ EPS έχει χρησιμοποιηθεί σε διάφορες

εφαρμογές που σχετίζονται με την ανάλυση επιβίωσης και την ανάλυση αξιοπιστίας.

2.2.9 Hartley's F-Max Test

Ο **Herman Otto Hartley** (1912) ήταν Γερμανοαμερικανός στατιστικός. Το 1934, σε ηλικία 22 ετών, ο Hartley απέκτησε διδακτορικό στα μαθηματικά από το Πανεπιστήμιο του Βερολίνου. Επίσης απέκτησε άλλα δύο διδακτορικά στη μαθηματική στατιστική από το Πανεπιστήμιο του Κέιμπριτζ το 1940 και από το University College του Λονδίνου το 1954. Έχει συνεισφέρει σημαντικά σε πολλούς τομείς της στατιστικής, του μαθηματικού προγραμματισμού και της βελτιστοποίησης.

Ο έλεγχος F-max αρχικά προτάθηκε από τον Herman Otto Hartley το 1950 ("The Maximum F-Ratio as a short-cut test for heterogeneity of variance", *Biometrika* 37) και στη συνέχεια συζητήθηκε από τους Fercho και Ringer το 1972 ("Small sample power of some tests of the constant failure rate", *Technometrics* 14). Βασίζεται στην κατανομή F, και χρησιμοποιείται για να ελεγχθούν υποθέσεις σχετικά με τις διακυμάνσεις. Μπορεί επίσης να χρησιμοποιηθεί για τον έλεγχο της εκθετικότητας μιας κατανομής. Σε αυτό το πλαίσιο, χρησιμοποιείται για τη σύγκριση της διακύμανσης των παρατηρούμενων δεδομένων με τη διακύμανση μιας εκθετικής κατανομής με τον ίδιο μέσο όρο.

2.2.10 Cox and Oakes Score Test

Ο **David Cox** (1924) είναι ένας εξέχων Βρετανός στατιστικός που έχει συνεισφέρει σημαντικά στον τομέα της στατιστικής. Ο Cox απέκτησε το πτυχίο του στα μαθηματικά από το Πανεπιστήμιο του Μπέρμιγχαμ το 1944 και το διδακτορικό του στη στατιστική από το Πανεπιστήμιο του Κέιμπριτζ το 1949. Αφού ολοκλήρωσε τις σπουδές του, εντάχθηκε στη σχολή του Πανεπιστημίου του Μπέρμιγχαμ, όπου πέρασε το μεγαλύτερο μέρος της καριέρας του. Ο Cox έχει συνεισφέρει σημαντικά σε διάφορους τομείς της στατιστικής, συμπεριλαμβανομένης της ανάλυσης επιβίωσης, της ανάλυσης παλινδρόμησης και του πειραματικού σχεδιασμού. Είναι ίσως περισσότερο γνωστός για την

εργασία του στο μοντέλο των αναλογικών κινδύνων, το οποίο έχει γίνει ο ακρογωνιαίος λίθος της ανάλυσης επιβίωσης.

Ο David Oakes (1943) είναι ένας εξέχων Καναδός στατιστικός που έχει συνεισφέρει σημαντικά στον τομέα της βιοστατιστικής. Ο Oakes απέκτησε το πτυχίο του στα μαθηματικά από το Πανεπιστήμιο McGill το 1964 και το διδακτορικό του στη στατιστική από το Πανεπιστήμιο της Βόρειας Καρολίνας στο Chapel Hill το 1969. Αφού ολοκλήρωσε τις σπουδές του, εντάχθηκε στη σχολή του Πανεπιστημίου του Ρότσεστερ, όπου πέρασε τα περισσότερα της καριέρας του. Ο Oakes έχει συνεισφέρει σημαντικά σε διάφορους τομείς της βιοστατιστικής, όπως η ανάλυση επιβίωσης, η επιδημιολογία και οι κλινικές δοκιμές. Είναι ίσως περισσότερο γνωστός για την εργασία του στο μοντέλο αναλογικών κινδύνων του Cox.

Ο έλεγχος βαθμολογίας Cox και Oakes εισήχθη για πρώτη φορά από τους David Cox και David Oakes στο βιβλίο τους το 1984, "Analysis of Survival Data".

Ο έλεγχος σκορ (score) των Cox και Oakes είναι ένα ισχυρό και ευέλικτο εργαλείο για την αξιολόγηση της καλής προσαρμογής των εκθετικών μοντέλων και έχει χρησιμοποιηθεί ευρέως σε ποικίλες εφαρμογές, όπως η ανάλυση επιβίωσης, η ανάλυση αξιοπιστίας και η θεωρία αναμονής.

Από την εισαγωγή του, ο έλεγχος score Cox and Oakes έχει υποστεί αρκετές τροποποιήσεις και βελτιώσεις. Για παράδειγμα, έχουν αναπτυχθεί επεκτάσεις για το χειρισμό λογοκριμένων δεδομένων, καθώς και για τον έλεγχο για πιο γενικές κατηγορίες κατανομών πέρα από την εκθετική. Συνολικά, ο έλεγχος score Cox και Oakes παραμένει ένα σημαντικό και ευρέως χρησιμοποιούμενο εργαλείο για την ανάλυση δεδομένων επιβίωσης και αξιοπιστίας και έχει συμβάλει σημαντικά στην κατανόηση αυτών των σημαντικών τομέων μελέτης.

2.2.11 Wong and Wong's Extremal Quotient Test

Ο Samuel Wong (1939) ήταν ένας εξέχων Κινεζοαμερικανός στατιστικός που συνέβαλε σημαντικά στον τομέα της μη παραμετρικής στατιστικής. Απέκτησε το διδακτορικό του στη Στατιστική από το Πανεπιστήμιο της Καλιφόρνια στο

Μπέρκλεϋ το 1966. Η έρευνα του Wong επικεντρώθηκε στη στατιστική συμπερασματολογία, μη παραμετρικές μεθόδους και ανάλυση επιβίωσης. Ανέπτυξε αρκετούς στατιστικούς ελέγχους, συμπεριλαμβανομένου του γενικευμένου ελέγχου Wilcoxon και του ελέγχου ακραίου πηλίκου, που έχουν χρησιμοποιηθεί ευρέως στον τομέα της στατιστικής.

Ο Wing-Kam Fung Wong (1941) είναι ένας Κινεζοαμερικανός στατιστικός που έχει συμβάλει σημαντικά στους τομείς της μη παραμετρικής στατιστικής, της θεωρίας αξιοπιστίας και του ποιοτικού ελέγχου. Απέκτησε το διδακτορικό του στη Στατιστική από το Πανεπιστήμιο της Καλιφόρνια στο Μπέρκλεϋ το 1966.

Ο έλεγχος ακραίου πηλίκου (Extreme Ratio Test / EXQT) αναπτύχθηκε από τους στατιστικούς S. P. Wong και P. G. Wong το 1979 (“ An extremal quotient test for exponential distributions”. *Metrika* 26) και έκτοτε έχει χρησιμοποιηθεί σε διάφορους τομείς, συμπεριλαμβανομένων των τομέων μηχανικής, οικονομικών και περιβαλλοντικών θεμάτων. Ο έλεγχος βασίζεται στην αναλογία της μεγαλύτερης και της μικρότερης διατεταγμένης στατιστικής παρατήρησης του δείγματος.

Εάν ο λόγος είναι μεγαλύτερος από μία κρίσιμη τιμή, τότε η μηδενική υπόθεση ότι τα δεδομένα ακολουθούν την εκθετική κατανομή απορρίπτεται.

2.3 Μη μονότονες συναρτήσεις κινδύνου (Non-Monotonic hazard rates)

Η πιθανότητα να συμβεί ένα γεγονός, ανάλογα με την χρονική στιγμή που βρισκόμαστε, άλλοτε αυξάνεται, άλλοτε μειώνεται και άλλοτε μένει σχετικά σταθερή.

Παράδειγμα : Η πιθανότητα ενός θηρευτή να επιβιώσει τις πρώτες μέρες της ζωής του στη φύση είναι πολύ μικρή, μετά από ένα σημείο, οι πιθανότητες αυξάνονται (όταν ο θηρευτής έχει πλέον αναπτύξει τα απαραίτητα χαρακτηριστικά για να επιβιώσει), και τέλος μειώνονται καθώς φτάνει προς το τέλος της ζωής του και είναι πιο ευάλωτος.

2.3.1 Gnedenko 's F-Test

Ο Boris Vladimirovic Gnedenko (1912) ήταν ένας σπουδαίος Σοβιετικός (Ουκρανός) μαθηματικός, μαθητής του Andrey Kolmogorov. Αποφοίτησε από το Κρατικό Πανεπιστήμιο της Μόσχας το 1935 και στη συνέχεια συνέχισε να εργάζεται στο Μαθηματικό Ινστιτούτο Steklov στη Μόσχα, όπου πέρασε το μεγαλύτερο μέρος της καριέρας του. Είναι γνωστός για το έργο του στη θεωρία πιθανοτήτων και στη στατιστική. Η πιο διάσημη συμβολή του Gnedenko στη θεωρία πιθανοτήτων είναι η θεωρία των ακραίων τιμών. Ο Gnedenko συνέβαλε επίσης σημαντικά στη θεωρία των στοχαστικών διαδικασιών, οι οποίες είναι μαθηματικά μοντέλα που χρησιμοποιούνται για την περιγραφή τυχαίων φαινομένων που εξελίσσονται με την πάροδο του χρόνου.

Ο έλεγχος προτάθηκε από τον Gnedenko το 1969 ("Mathematical Models of Reliability Theory". Academic Press). Με τον έλεγχο επίσης ασχολήθηκαν οι Fercho και Ringer το 1972 ("Small sample power of some tests of the constant failure rate". Technometrics 14) και οι Lin και Mudholkar το 1980 ("A test of exponentiality based on a bivariate F distribution". Technometrics 22). Για την υλοποίηση του ελέγχου τα δεδομένα διατάσσονται και χωρίζονται σε δύο ομάδες.

Ο έλεγχος Gnedenko έχει χρησιμοποιηθεί ευρέως σε διάφορους τομείς όπως στην ανάλυση αξιοπιστίας, στη θεωρία ουρών και σε άλλα πεδία όπου οι εκθετικές κατανομές χρησιμοποιούνται συνήθως για τη μοντελοποίηση του χρόνου μεταξύ των γεγονότων.

2.3.2 Harris modification of Gnedenko 's F-Test

Ο **Carl M. Harris** είναι Αμερικανός μαθηματικός και στατιστικός. Είναι καθηγητής στατιστικής στο Τμήμα Βιομηχανικής Μηχανικής και Επιχειρησιακής Έρευνας στο πανεπιστήμιο της Syracuse στη Νέα Υόρκη. Έχει συνεισφέρει σημαντικά και έχει δημοσιεύσει αρκετά επιστημονικά άρθρα σε διάφορους τομείς των μαθηματικών. Είναι γνωστός για το έργο του στα πεδία της θεωρίας πιθανοτήτων, των στοχαστικών διαδικασιών και της στατιστικής συμπερασματολογίας.

Ο έλεγχος εισήχθη για πρώτη φορά από τον Carl Harris το 1976 (“A note on testing for exponentiality”. Naval Research Logistics Quarterly 23) ως τροποποίηση του αρχικού F-test για την εκθετικότητα που προτάθηκε από τον Boris Gnedenko το 1969. Η τροποποίηση του Harris στο F-test περιλαμβάνει τη χρήση διαφορετικής στατιστικής ελεγχουσυνάρτησης που λαμβάνει υπόψη τη μεταβλητότητα των δεδομένων. Συγκεκριμένα, η στατιστική ελεγχουσυνάρτηση βασίζεται στο εύρος του δείγματος και στο μέσο όρο του δείγματος των δεδομένων και συγκρίνεται με μια κατανομή αναφοράς για τον προσδιορισμό της p-value της ελέγχου.

Η τροποποίηση Harris του F-test έχει αποδειχθεί ότι είναι πιο ισχυρή από το αρχικό F-test, ειδικά όταν το μέγεθος του δείγματος είναι μικρό ή όταν τα δεδομένα παρουσιάζουν λοξότητα (skewness). Έκτοτε έχει γίνει ένα ευρέως χρησιμοποιούμενο εργαλείο στην ανάλυση δεδομένων επιβίωσης, μηχανικής αξιοπιστίας και άλλων πεδίων όπου χρησιμοποιούνται συνήθως εκθετικές κατανομές.

2.3.3 Lin and Mudholkar ‘s Bivariate F-Test

Ο **Govind S. Mudholkar** είναι μαθηματικός και στατιστικός. Ερευνητής και καθηγητής στο Πανεπιστήμιο του Rochester στη Νέα Υόρκη. Έχει δημοσιεύσει πάνω από 100 επιστημονικά άρθρα και έχει συνεισφέρει σε ποικίλους τομείς των Μαθηματικών. Κάποιοι από τους τομείς έρευνας του είναι η αντίστροφη κατανομή Gauss, η πολυμεταβλητή κανονική κατανομή και η κανονική κατανομή.

Ο **Ching – Chuong Lin** είναι μαθηματικός , καθηγητής και ερευνητής στο Πανεπιστήμιο του Maryland.

Ο έλεγχος Bivariate F-Test (διμεταβλητής F) προτάθηκε από τον Govind Mudholkar και τον Ching Lin το 1980 (“A test of exponentiality based on a bivariate F distribution”. Technometrics 22). Στην ουσία ο έλεγχος αποτελεί ένα συνδυασμό των δύο προηγούμενων ελέγχων, του Gnedenko και του Harris. Συνήθως εφαρμόζεται κατά την ανάλυση της κοινής κατανομής δύο μεταβλητών για να προσδιοριστεί εάν τα δεδομένα ακολουθούν εκθετική κατανομή. Ο

έλεγχος διμεταβλητής F αξιολογεί εάν η κοινή κατανομή των δύο μεταβλητών μπορεί να μοντελοποιηθεί επαρκώς από την εκθετική κατανομή. Ο έλεγχος βασίζεται στη σύγκριση της παρατηρούμενης κοινής κατανομής με τη θεωρητική εκθετική κατανομή. Ο έλεγχος χρησιμοποιείται σε διάφορους τομείς, όπως η ανάλυση δεδομένων επιβίωσης και μηχανικής αξιοπιστίας.

2.3.4 Cox and Oakes Score Test

Ο David Cox (1924) ήταν ένας εξέχων Βρετανός στατιστικός που έχει συνεισφέρει σημαντικά στη στατιστική.

Ο Cox έχει συνεισφέρει σημαντικά σε διάφορους τομείς της στατιστικής, συμπεριλαμβανομένης της ανάλυσης παλινδρόμησης, στους πειραματικούς σχεδιασμούς και στην ανάλυση επιβίωσης. Η πιο γνωστή του εργασία εισάγει το μοντέλο των αναλογικών κινδύνων, το οποίο αποτελεί κεντρικό πυρήνα της ανάλυσης επιβίωσης.

Ο David Oakes (1943) είναι γνωστός στατιστικός με ιδιαίτερη συνεισφορά στον κλάδο της βιοστατιστικής.

Η συνεισφορά του Oakes στη βιοστατιστική σχετίζεται με διάφορους τομείς του κλάδου όπως η επιδημιολογία, η ανάλυση επιβίωσης και οι κλινικές δοκιμές. Η ερευνητική του εργασία πάνω στο μοντέλο αναλογικών κινδύνων σε συνεργασία με τον David Cox, είναι εξαιρετικά σημαντική για την ανάλυση επιβίωσης.

Ο έλεγχος τάξεων (rank) Cox και Oakes εισήχθη για πρώτη φορά από τους David Cox και David Oakes στο βιβλίο τους το 1984, "Analysis of Survival Data". Ο έλεγχος τάξεων Cox και Oakes αποτελεί ισχυρό και ευέλικτο εργαλείο για την αξιολόγηση της καλής προσαρμογής των δεδομένων στην εκθετική κατανομή και χρησιμοποιείται εκτενώς σε διάφορες εφαρμογές, όπως η θεωρία αναμονής, η ανάλυση αξιοπιστίας και η ανάλυση επιβίωσης. Από την ανακάλυψή του μέχρι και σήμερα, ο έλεγχος Cox and Oakes έχει τροποποιηθεί και βελτιωθεί. Ένα παράδειγμα αποτελεί η επέκτασή του με σκοπό την διαχείριση λογοκριμένων δεδομένων, καθώς και για τον έλεγχο για διαφορετικές κατηγορίες κατανομών πέρα από την εκθετική κατανομή. Ο έλεγχος Cox και Oakes αποτελεί σημαντικό εργαλείο και χρησιμοποιείται

εκτενώς στην ανάλυση δεδομένων επιβίωσης και αξιοπιστίας και συνεχίζει να συμβάλει στην κατανόηση αυτών των σημαντικών περιοχών.

2.3.5 . Desphande's Test

Ο Δρ Jayant Deshpande (1948) είναι γνωστός Ινδός στατιστικός με ιδιαίτερη συμβολή στον κλάδο της ανάλυσης αξιοπιστίας.

Ο Deshpande επικεντρώθηκε σε διάφορους κλάδους της ανάλυσης αξιοπιστίας, συμπεριλαμβανομένων στατιστικών μεθόδων για ανάλυση δεδομένων διάρκειας ζωής, επιταχυνόμενων δοκιμών ζωής και μοντελοποίησης ανάπτυξης και αξιοπιστίας.

Ο έλεγχος του Deshpande προτάθηκε από τον Deshpande για τον έλεγχο δεδομένων με αυξανόμενο ρυθμό κινδύνου με μηδενική υπόθεση την εκθετικότητα.

Ο έλεγχος του Deshpande αποτελεί ένα γρήγορο και απλό εργαλείο χωρίς να περιλαμβάνει πολύπλοκους υπολογισμούς, ωστόσο σύμφωνα με κάποιους συναδέλφους του όπως οι Dipankar Bandyopadhyay και Asit P. Basu ο έλεγχος έχει κάποιες αδυναμίες και προτείνουν κάποιες βελτιώσεις σε άρθρο που έχουν δημοσιεύσει αναλύοντας τον συγκεκριμένο έλεγχο. Ο έλεγχος έχει χρησιμοποιηθεί ευρέως στον τομέα της ανάλυσης αξιοπιστίας και έχει αποδειχθεί αποτελεσματικός σε διάφορες εφαρμογές.

ΚΕΦΑΛΑΙΟ 3

Έλεγχοι καλής προσαρμογής για την εκθετική κατανομή

Σε αυτό το κεφάλαιο, θα παρουσιάσουμε τη μαθηματική περιγραφή για τους ελέγχους καλής προσαρμογής που αναφέραμε στο προηγούμενο κεφάλαιο.

Έστω X_1, X_2, \dots, X_n τυχαίο δείγμα από πληθυσμό με συνάρτηση πυκνότητας πιθανότητας $f_x(\cdot)$. Εξετάζουμε τη μηδενική υπόθεση :

$$H_0: f_x(\cdot) = \lambda e^{-\lambda x}, \lambda > 0 \text{ και } x \geq 0.$$

Οι κανονικοποιημένες αποστάσεις, οι οποίες χρησιμοποιούνται σε διάφορους ελέγχους, ορίζονται ως εξής:

$$D_i = (N - i + 1)(X_{(i)} - X_{(i-1)}) \quad (3.1)$$

όπου $i=1,2,\dots,N$, $X_{(0)} = 0$ και $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(N)}$ είναι το διατεταγμένο τυχαίο δείγμα.

3.1 Οι γραφικοί έλεγχοι P-P Plot και Q-Q Plot

P-P Plot

Το διάγραμμα P-P Plot είναι ένα διάγραμμα πιθανοτήτων για την αξιολόγηση του πόσο στενά συμφωνούν δύο σύνολα δεδομένων ή για την αξιολόγηση του πόσο στενά ταιριάζει ένα σύνολο δεδομένων σε ένα συγκεκριμένο μοντέλο. Λειτουργεί με την απεικόνιση των δύο αθροιστικών συναρτήσεων κατανομής η μία έναντι της άλλης- εάν είναι παρόμοιες, τα δεδομένα θα εμφανίζονται σχεδόν σε ευθεία γραμμή.

Έστω δύο κατανομές πιθανότητας, με αθροιστικές συναρτήσεις κατανομής F και G αντίστοιχα. Έχουμε το διάγραμμα $(F(z), G(z))$ με τη μεταβλητή z να λαμβάνει

τιμές στο $(-\infty, +\infty)$. Αφού η αθροιστική συνάρτηση κατανομής λαμβάνει τιμές στο $[0,1]$ το εύρος του γραφήματος θα είναι το τετράγωνο $[0,1] \times [0,1]$.

Η γραμμή σύγκρισης είναι η ευθεία 45° από το $(0,0)$ στο $(1,1)$ και οι κατανομές είναι ίσες εάν και μόνο εάν η γραφική παράσταση εμπίπτει σε αυτή τη γραμμή. Ο βαθμός απόκλισης διευκολύνει τον οπτικό εντοπισμό του πόσο διαφορετικές είναι οι κατανομές, αλλά λόγω του σφάλματος δειγματοληψίας, ακόμη και δείγματα που προέρχονται από πανομοιότυπες κατανομές δεν θα εμφανίζονται εντελώς πανομοιότυπα.

Παράδειγμα

Παρακάτω παρουσιάζεται ένα παράδειγμα χάραξης γραφικής παράστασης P-P Plot για 20 παρατηρήσεις :

Για τη συμπλήρωση της στήλης ECDF χρησιμοποιείται ο τύπος:

$$(Rank_i - 0.5) / N$$

Για τη συμπλήρωση της στήλης Data CDF χρησιμοποιείται ο τύπος (συνάρτηση κατανομής της $E(\lambda)$ κατανομής) :

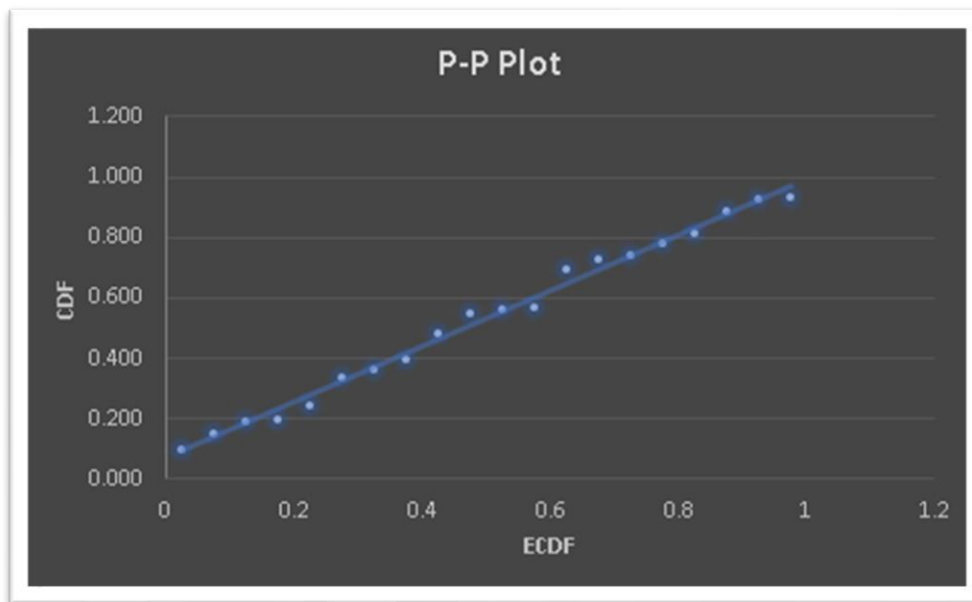
$$1 - e^{-\hat{\lambda} x_i}$$

όπου $\hat{\lambda}$ ο εκτιμητής της παραμέτρου της εκθετικής κατανομής ($\hat{\lambda} = 1/\bar{X}$) και x_i η παρατήρηση με τάξη (Rank)=i.

Lamba est.		Mean Est.	
3.358		0.298	

Rank	Data	ECDF	Data CDF	Rank	Data	ECDF	Data CDF
1	0.030	0.025	0.097	11	0.245	0.525	0.561
2	0.049	0.075	0.151	12	0.253	0.575	0.572
3	0.064	0.125	0.194	13	0.357	0.625	0.699
4	0.067	0.175	0.201	14	0.390	0.675	0.730
5	0.084	0.225	0.246	15	0.400	0.725	0.739
6	0.123	0.275	0.338	16	0.453	0.775	0.782
7	0.135	0.325	0.364	17	0.503	0.825	0.815
8	0.151	0.375	0.397	18	0.646	0.875	0.886
9	0.196	0.425	0.482	19	0.773	0.925	0.925
10	0.236	0.475	0.548	20	0.801	0.975	0.932

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 1: Παράδειγμα υπολογισμού P-P Plot



ΕΙΚΟΝΑ 1 : Διάγραμμα P-P Plot

Όπως παρατηρούμε, τα σημεία βρίσκονται πολύ κοντά στην ευθεία $y=x$, άρα η πιθανότητα τα δεδομένα μας να ακολουθούν εκθετική κατανομή με $\lambda=3.358$ είναι υψηλή.

Q-Q Plot

Το διάγραμμα Q-Q (quantile-quantile plot) είναι ένα διάγραμμα πιθανοτήτων, μια γραφική μέθοδος για τη σύγκριση δύο κατανομών πιθανότητας με την απεικόνιση των ποσοστημορίων τους το ένα έναντι του άλλου. Ένα σημείο (x, y) στο διάγραμμα αντιστοιχεί σε ένα από τα ποσοστημόρια της δεύτερης κατανομής (συντεταγμένη y) που απεικονίζεται έναντι του ίδιου ποσοστημορίου της πρώτης κατανομής (συντεταγμένη x). Αυτό ορίζει μια παραμετρική καμπύλη όπου η παράμετρος είναι ο δείκτης του διαστήματος των ποσοστημορίων.

Εάν οι δύο κατανομές που συγκρίνονται είναι παρόμοιες, τα σημεία του διαγράμματος Q-Q θα βρίσκονται κατά προσέγγιση πάνω στην ευθεία $y = x$. Εάν οι κατανομές σχετίζονται γραμμικά, τα σημεία του διαγράμματος Q-Q θα βρίσκονται κατά προσέγγιση πάνω σε μια γραμμή, αλλά όχι απαραίτητα πάνω στην ευθεία $y = x$.

Ένα διάγραμμα Q-Q χρησιμοποιείται για τη σύγκριση των σχημάτων των κατανομών, παρέχοντας μια γραφική εικόνα του πώς ιδιότητες όπως η θέση, η κλίμακα και η λοξότητα είναι παρόμοιες ή διαφορετικές στις δύο κατανομές. Τα διαγράμματα Q-Q μπορούν να χρησιμοποιηθούν για τη σύγκριση συλλογών δεδομένων ή θεωρητικών κατανομών. Η χρήση των διαγραμμάτων Q-Q για τη σύγκριση δύο δειγμάτων δεδομένων μπορεί να θεωρηθεί ως μια μη παραμετρική προσέγγιση για τη σύγκριση των υποκείμενων κατανομών τους.

Τα διαγράμματα Q-Q χρησιμοποιούνται συνήθως για τη σύγκριση ενός συνόλου δεδομένων με ένα θεωρητικό μοντέλο. Αυτό μπορεί να παρέχει μια αξιολόγηση της καλής προσαρμογής που είναι γραφική, αντί να περιορίζεται στις τιμές μιας αριθμητικής συνοπτικής ελεγχοσυνάρτησης.

Το κύριο βήμα για την κατασκευή ενός διαγράμματος Q-Q είναι ο υπολογισμός ή η εκτίμηση των ποσοστημορίων που πρόκειται να απεικονιστούν. Εάν ένας ή και οι δύο άξονες σε ένα διάγραμμα Q-Q βασίζονται σε μια θεωρητική κατανομή με συνεχή αθροιστική συνάρτηση κατανομής (CDF), όλα τα ποσοστημόρια είναι μοναδικά καθορισμένα και μπορούν να προκύψουν από την αντιστροφή της CDF. Εάν μια θεωρητική κατανομή πιθανοτήτων με ασυνεχή CDF είναι μία από τις δύο κατανομές που συγκρίνονται, ορισμένα από τα ποσοστημόρια μπορεί να μην είναι καθορισμένα, οπότε μπορεί να σχεδιαστεί ένα παρεμβαλλόμενο ποσοστημόριο. Εάν το διάγραμμα Q-Q βασίζεται σε δεδομένα, χρησιμοποιούνται διάφοροι εκτιμητές ποσοστημορίων.

Πιο γενικά, δεδομένων δύο αθροιστικών συναρτήσεων κατανομής πιθανότητας F και G , με σχετικές συναρτήσεις ποσοστημορίων F^{-1} και G^{-1} (η αντίστροφη συνάρτηση της CDF είναι η συνάρτηση ποσοστημορίου), το διάγραμμα Q-Q σχεδιάζει το q -οστό ποσοστημόριο της F έναντι του q -οστού ποσοστημορίου της G για ένα εύρος τιμών του q . Έτσι, το διάγραμμα Q-Q είναι μια παραμετρική καμπύλη στο $[0,1]$ και τιμές στο πραγματικό επίπεδο R^2 .

Για να ελέγξουμε αν ένα σετ δεδομένων με N παρατηρήσεις ακολουθεί μία συγκεκριμένη κατανομή, αρχικά βρίσκουμε τα σημεία p_i διαμερίζοντας το διάστημα $[0,1]$ σε N μέρη. Έπειτα, χρησιμοποιώντας την αντίστροφη συνάρτηση της θεωρητικής αθροιστικής συνάρτησης κατανομής προς εξέταση υπολογίζουμε τα θεωρητικά ποσοστημόρια. Στη συνέχεια, είτε μπορούμε να χαράξουμε το γράφημα Q-Q χρησιμοποιώντας τα δεδομένα και τα αντίστοιχα ποσοστημόρια, είτε αρχικά να κανονικοποιήσουμε τα δεδομένα και μετά να κάνουμε το ίδιο.

Παράδειγμα

Χρησιμοποιούμε τα ίδια δεδομένα με το προηγούμενο παράδειγμα.
 Για τη συμπλήρωση της στήλης ECDF στα σημεία p_i χρησιμοποιείται ο τύπος :

$$(Rank_i - 0.5) / N$$

Για τη συμπλήρωση της στήλης Quantile χρησιμοποιείται ο τύπος:

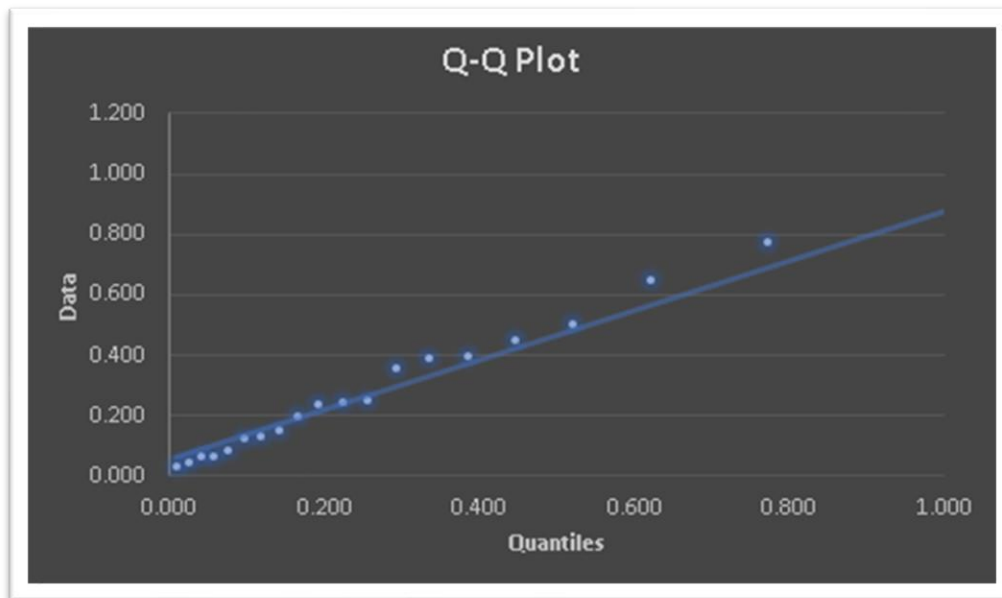
$$(-\ln(1 - p_i)) / \hat{\lambda}$$

όπου $\hat{\lambda}$ ο εκτιμητής της παραμέτρου της εκθετικής κατανομής ($\hat{\lambda} = 1/\bar{X}$) και p_i το i -οστό στοιχείο της στήλης ECDF.

Lamba est.		Mean Est.	
3.358		0.298	

Rank	Data	Pi -ECDF	Quantile	Rank	Data	Pi -ECDF	Quantile
1	0.030	0.025	0.008	11	0.2454	0.5250	0.2217
2	0.049	0.075	0.023	12	0.2528	0.5750	0.2548
3	0.064	0.125	0.040	13	0.3575	0.6250	0.2921
4	0.067	0.175	0.057	14	0.3900	0.6750	0.3347
5	0.084	0.225	0.076	15	0.4002	0.7250	0.3845
6	0.123	0.275	0.096	16	0.4531	0.7750	0.4442
7	0.135	0.325	0.117	17	0.5027	0.8250	0.5191
8	0.151	0.375	0.140	18	0.6463	0.8750	0.6193
9	0.196	0.425	0.165	19	0.7730	0.9250	0.7714
10	0.236	0.475	0.192	20	0.8010	0.9750	1.0986

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 2: Παράδειγμα υπολογισμού Q-Q Plot



ΕΙΚΟΝΑ 2 : Διάγραμμα Q-Q Plot

Όπως παρατηρούμε, τα σημεία, όπως και πριν βρίσκοντα πολύ κοντά στην ευθεία $y=x$, άρα η πιθανότητα τα δεδομένα μας να ακολουθούν εκθετική κατανομή με $\lambda=3.358$ είναι υψηλή.

3.2 Ο γενικός έλεγχος καλής προσαρμογής X^2

Ο έλεγχος καλής προσαρμογής X^2 είναι κατάλληλος για να ελεγχθεί η υπόθεση ότι μία μεταβλητή X ακολουθεί είτε μία διακριτή, είτε μία συνεχή κατανομή.

Έστω ότι X_1, X_2, \dots, X_n είναι τυχαίο δείγμα μίας τυχαίας μεταβλητής X με αθροιστική συνάρτηση κατανομής (α.σ.κ.) $F(x)$, η οποία είναι άγνωστη. Θέλουμε να ελέγξουμε την υπόθεση :

$$H_0: F(x) = F_0(x)$$

όπου $F_0(x)$ είναι μία ειδική αθροιστική συνάρτηση κατανομής, η οποία μπορεί να είναι είτε συνεχής είτε διακριτή.

Υπό τη μηδενική υπόθεση διαμερίζουμε το πεδίο τιμών των X_i , σε k ξένα μεταξύ τους σύνολα A_1, A_2, \dots, A_k . Έστω N_i , το πλήθος των παρατηρήσεων που ανήκουν

στο σύνολο A_i , ($i=1,2,\dots,k$). Τότε κάτω από τη μηδενική υπόθεση το διάνυσμα $(N_1, N_2, \dots, N_k)^T$ ακολουθεί την πολυωνυμική κατανομή με παραμέτρους n και p_1, p_2, \dots, p_k όπου :

$$p_i = \{X_i \in A_i | H_0\} \quad (i=1,2,\dots,k)$$

Για μεγάλο n , κάτω από τη μηδενική υπόθεση η ελεγχοσυνάρτηση είναι :

$$T = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \quad (3.2)$$

Ακολουθεί προσεγγιστικά την κατανομή X^2 με $(k-1)$ βαθμούς ελευθερίας. Άρα με τη βοήθεια της κατανομής X^2 με $k-1$ βαθμούς ελευθερίας μπορούμε να υπολογίσουμε προσεγγιστικά την p -value του αμφίπλευρου ελέγχου. Συγκεκριμένα υπολογίζουμε την τιμή της παραπάνω ελεγχοσυνάρτησης με βάση τις παρατηρήσεις μας και εν-συνεχεία η p -value του ελέγχου είναι η πιθανότητα δεξιά της τιμής αυτής, με βάση την κατανομή X^2 με $(k-1)$ βαθμούς ελευθερίας. Απαραίτητη προϋπόθεση για τα παραπάνω είναι :

Όλες οι αναμενόμενες συχνότητες $np_i \geq 5$

Στην περίπτωση που η παραπάνω προϋπόθεση δεν ισχύει ενώνουμε γειτονικές κλάσεις.

Παρατηρήσεις :

Θεωρήσαμε ότι γνωρίζουμε τα p_i , αφού καθορίζονται πλήρως από την κατανομή F_0 (για εμάς την εκθετική). Στην περίπτωση που η F_0 εξαρτάται από άγνωστες παραμέτρους, εκτιμούμε αυτές τις παραμέτρους από τα δεδομένα και χρησιμοποιούμε τις τιμές των εκτιμητών στη θέση των αγνώστων παραμέτρων.

Το μειονέκτημα του εν λόγω ελέγχου είναι ότι για την υλοποίησή του χρειάζεται να προβούμε σε ομαδοποίηση δεδομένων (διαμερίζοντας το πεδίο τιμών του τυχαίου δείγματος σε k όπως είπαμε ξένες ομάδες), με συνέπεια την απώλεια πληροφορίας στην περίπτωση των συνεχών παρατηρήσεων. Επίσης η εν λόγω διαμέριση αρκετές φορές είναι αυθαίρετη.

3.3 Λοξότητα και Κύρτωση (Skewness and Kurtosis) : KUSK

Η ελεγχοσυνάρτηση που προτείνεται εδώ είναι:

$$K = (\hat{\beta}_1 + 0.5) / \hat{\beta}_2 \quad (3.3)$$

όπου: $\hat{\beta}_1 = \hat{\mu}_3^2 / \hat{\mu}_2^3$ (ο δειγματικός συντελεστής λοξότητας)
 $\hat{\beta}_2 = \hat{\mu}_4 / \hat{\mu}_2^2$ (ο δειγματικός συντελεστής κύρτωσης).

Όταν η μηδενική υπόθεση είναι αληθής, το $(\beta_1 + 0.5) / \beta_2$ παίρνει την τιμή 0.5. Οι κατώτερες και οι ανώτερες κρίσιμες τιμές για το K λαμβάνονται χρησιμοποιώντας προσομοιώσεις. Για μικρά μεγέθη δείγματος, αυτός ο έλεγχος θα είναι παραπλανητικός, καθώς και τα δύο, $\hat{\beta}_1$ και $\hat{\beta}_2$ είναι ευαίσθητα σε απόμακρα (ή άτυπα) σημεία (outliers).

Τα $\hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$ είναι οι κεντρικές ροπές δεύτερης, τρίτης και τέταρτης τάξης αντίστοιχα. Δηλαδή έχουμε :

$$\hat{\mu}_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

$$\hat{\mu}_3 = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N}$$

$$\hat{\mu}_4 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{N}$$

3.4 KSL Test

Αρχικός έλεγχος Kolmogorov για ένα δείγμα (Kolmogorov , 1933)

Αρχικά παρουσιάζουμε τον αρχικό έλεγχο του Kolmogorov ο οποίος είναι κατάλληλος για οποιαδήποτε συνεχή κατανομή:

Έστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από έναν πληθυσμό με συνεχή αλλά άγνωστη αθροιστική συνάρτηση κατανομής $F(x)$. Θέλουμε να ελέγξουμε μία από τις εξής τρεις υποθέσεις:

A. Τη μηδενική υπόθεση

$H_0 : F(x) = F_0(x), \forall x \in R$, έναντι της εναλλακτικής

$H_1 : F(x) \neq F_0(x)$, για κάποιο $x \in R$.

B. Τη μηδενική υπόθεση

$H_0 : F(x) = F_0(x), \forall x \in R$, έναντι της εναλλακτικής

$H_{1+} : F(x) > F_0(x)$, για κάποιο $x \in R$.

Γ. Τη μηδενική υπόθεση

$H_0 : F(x) = F_0(x), \forall x \in R$, έναντι της εναλλακτικής

$H_{1-} : F(x) < F_0(x)$, για κάποιο $x \in R$,

όπου $F_0(x)$ είναι μία, γνωστή, ειδική αθροιστική συνάρτηση κατανομής. Αν θέλουμε να εξετάσουμε την εκθετικότητα, η $F_0(x)$ είναι η εκθετική αθροιστική συνάρτηση κατανομής.

Η στατιστική συνάρτηση για τον έλεγχο των παραπάνω υποθέσεων βασίζεται στην εμπειρική αθροιστική συνάρτηση κατανομής, καθώς ως αμερόληπτη εκτιμήτρια της αθροιστικής συνάρτησης κατανομής δε θα πρέπει να αποκλίνει σημαντικά από την τελευταία. Για καθένα από τα τρία διαφορετικά σύνολα υποθέσεων, η στατιστική ελεγχοσυνάρτηση ορίζεται ως εξής:

A. Χρησιμοποιείται η στατιστική συνάρτηση:

$$\sqrt{n}D_n = \sqrt{n} \sup |F_n(x) - F_0(x)|$$

η οποία στην ουσία μετρά πόσο αποκλίνει η εμπειρική αθροιστική συνάρτηση κατανομής $F_n(x)$ από την αθροιστική συνάρτηση κατανομής $F_0(x)$ και αναζητά τη μέγιστη κατακόρυφη απόσταση μεταξύ των γραφημάτων των $F_n(x)$ και $F_0(x)$. Επομένως γίνεται εύκολα αντιληπτό ότι απορρίπτεται η μηδενική υπόθεση H_0 για μεγάλες τιμές της στατιστικής αυτής συνάρτησης.

Έχουμε ότι :

$$K_n = \sqrt{n} \max(D_n^+, D_n^-) \quad (3.4)$$

με : $D_n^+ = \sup (F_n(x) - F_0(x))$

και $D_n^- = \sup (F_0(x) - F_n(x))$

Απορρίπτουμε την H_0 αν και μόνο αν $K_n \geq c_\alpha$

όπου c_α τέτοιο ώστε :

$$P(K_n \geq c_a) \approx a$$

Η τιμή του c_a/\sqrt{n} δίνεται από πίνακες.

Β. Χρησιμοποιείται η στατιστική συνάρτηση:

$$\sqrt{n}D_n^+ = \sqrt{n} \sup (F_n(x) - F_0(x))$$

και η μηδενική υπόθεση απορρίπτεται για μεγάλες τιμές του. Προκύπτει ότι απορρίπτουμε την H_0 αν και μόνο αν $\sqrt{n}D_n^+ \geq c_a$, όπου c_a τέτοιο ώστε :

$$P(K_n \geq c_a) \approx a$$

Γ. Χρησιμοποιείται η στατιστική συνάρτηση:

$$\sqrt{n}D_n^- = \sqrt{n} \sup (F_0(x) - F_n(x))$$

και η μηδενική υπόθεση απορρίπτεται για μεγάλες τιμές του. Προκύπτει ότι απορρίπτουμε την H_0 αν και μόνο αν $\sqrt{n}D_n^- \geq c_a$, όπου c_a τέτοιο ώστε :

$$P(K_n \geq c_a) \approx a$$

Η τροποποίηση του Lilliefors για την εκθετική κατανομή (Lilliefors ,1969)

Έστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα μεγέθους n από την τυχαία μεταβλητή X της οποίας η συνάρτηση κατανομής είναι : $F_X(x), x \in R$

Ελέγχουμε την υπόθεση :

$$H_0: F_X(x) = \begin{cases} 1 - e^{-\frac{x}{\mu}}, & x > 0, \mu \text{ άγνωστη παράμετρος} \\ 0 & , \text{ διαφορετικά} \end{cases}$$

H_1 : η κατανομή της X δεν είναι εκθετική

Μετασχηματίζουμε τα δεδομένα χρησιμοποιώντας το μετασχηματισμό: (η εκτιμήτρια μεγίστης πιθανοφάνειας του μ είναι \bar{X})

$$Z_i = \frac{X_i}{\bar{X}}, i = 1, 2, \dots, n$$

$$\text{με } \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Έτσι, η στατιστική συνάρτηση που θα χρησιμοποιηθεί, αντί να εκφράζει ένα μέτρο της απόστασης μεταξύ της εμπειρικής συνάρτησης κατανομής $F_n(x)$ από τη συνάρτηση κατανομής κάτω από τη μηδενική υπόθεση $F_0(x)$, θα εκφράζει την απόσταση μεταξύ της εμπειρικής συνάρτησης κατανομής $F_n^*(z)$ των μετασχηματισμένων δεδομένων Z_1, Z_2, \dots, Z_n από τη συνάρτηση κατανομής

$$F^*(z) = 1 - e^{-z}, z > 0$$

Στη συγκεκριμένη περίπτωση, η στατιστική συνάρτηση του ελέγχου είναι η:

$$T = \sup |F^*(z) - F_n^*(z)| \quad (3.5)$$

Απορρίπτουμε τη μηδενική υπόθεση H_0 σε επίπεδο σημαντικότητας α αν η τιμή της στατιστικής συνάρτησης T υπερβαίνει το $(1-\alpha)$ -ποσοστιαίο σημείο της κατανομής της.

Ο Lilliefors μελέτησε την κατανομή της στατιστικής συνάρτησης T και προσδιόρισε κατά προσέγγιση τα ποσοστιαία σημεία της, αλλά η πραγματική κατανομή της συνάρτησης αυτής μελετήθηκε αργότερα από τον Durbin το 1975.

3.5 The Gini Statistic (Gastwirth and Gail, 1978)

Ο έλεγχος του Gini χρησιμοποιεί την ακόλουθη στατιστική συνάρτηση :

$$G = \frac{\sum_{i=1}^{N-1} i(N-i)(X_{(i+1)} - X_{(i)})}{(N-1) \sum_{i=1}^N X_i} = \frac{\sum_{i=1}^{N-1} iD_{i+1}}{(N-1) \sum_{i=1}^N X_i} \quad (3.6)$$

όπου D_{i+1} δίνεται από τη σχέση (3.1).

Οι Gastwirth και Gail (1978) παρέχουν έναν πίνακα με προσεγγιστικές κρίσιμες τιμές και την ακόλουθη κανονική προσέγγιση :

$$G^* = \frac{G - E(G|H_0)}{\sqrt{Var(G|H_0)}}$$

όπου $E(G|H_0) = 0.5$ και $Var(G|H_0) = 1/[12*(N-1)]$

Υπό την προϋπόθεση της εκθετικότητας, το G^* ακολουθεί μια ασυμπτωτικά τυπική κανονική κατανομή, ακόμη και για δείγματα τόσο μικρά όσο ίσα με 10. Έχει καλή ισχύ για εναλλακτικές υποθέσεις από τις κατανομές Weibull, Uniform

(ομοιόμορφη) και Γάμμα. Το Gini στατιστικό μπορεί επίσης να προσαρμοστεί στα δεδομένα που είναι λογοκριμένα στη διατεταγμένη τ.μ. $X_{(R)}$, όπου $R \leq n$.

3.6 The Lorenz Statistic (Gail and Gastwirth, 1978)

Οι Gail και Gastwirth διαπίστωσαν ότι η στατιστική συνάρτηση Lorenz παρέχει ένα ισχυρό τεστ εκθετικότητας. Η ελεγχοσυνάρτηση του ελέγχου είναι :

$$L_N(p) = \frac{\sum_{i=1}^{[Np]} X_{(i)}}{N * \bar{X}} \quad (3.7)$$

όπου $0 < p < 1$ και $[Np]$ είναι ο μεγαλύτερος ακέραιος αριθμός μικρότερος ή ίσος με Np . Οι συγγραφείς παρέχουν κατώτερες και ανώτερες κρίσιμες τιμές και συνιστούν τον καθορισμό $p = 0,5$.

3.7 The Pietra Statistic (Gail and Gastwirth , 1978)

Η ελεγχοσυνάρτηση του ελέγχου είναι η ακόλουθη:

$$P = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{2N \bar{X}} \quad (3.8)$$

Οι συγγραφείς (Gail και Gastwirth) παρέχουν κατώτερες και ανώτερες κρίσιμες τιμές.

3.8 Deshpande's Test (Deshpande, 1983)

Αυτή η διαδικασία προτάθηκε για τον έλεγχο της εκθετικότητας έναντι κατανομών με αυξανόμενους ρυθμούς αποτυχίας (increasing failure rate distributions). Η ελεγχοσυνάρτηση του ελέγχου υπολογίζεται ως εξής: Πολλαπλασιάζουμε τα X_i , $i = 1, 2, \dots, n$ με b ($b = 0.5$ ή 0.9 εδώ) και διατάσσουμε τα X_1, \dots, X_n και bX_1, \dots, bX_n μαζί σε αύξουσα τάξη μεγέθους. Υπολογίζουμε την ποσότητα:

$$S = \sum_{i=1}^n R_i - 0.5n(n+1) - n \quad (3.9)$$

όπου R_i είναι η τάξη του X_i . Κρίσιμες τιμές για μονόπλευρους ελέγχους, που βρέθηκαν με προσομοίωση για αυτό το Wilcoxon-τύπου στατιστικό, παρέχονται από τον συγγραφέα για $b = 0.5$ και 0.9 , όταν $n \leq 15$. Ο συγγραφέας προτείνει να χρησιμοποιείται το 0.5 , όταν η εναλλακτική κατανομή υπάρχει υποψία ότι ανήκει στην ευρύτερη κατηγορία «η νέα είναι καλύτερη από αυτήν που χρησιμοποιείται» και το 0.9 , όταν η εναλλακτική περιορίζεται στην κατηγορία με αυξανόμενο μέσο ρυθμό αποτυχίας. Αφού δεν γνωρίζουμε εκ των προτέρων την εναλλακτική κατανομή, κρίσιμες τιμές για αμφίπλευρο έλεγχο για $n = 20$ βρέθηκαν με προσομοίωση και χρησιμοποιήθηκαν σ' αυτή τη μελέτη. Ο Deshrande δίνει επίσης την ακόλουθη κανονική προσέγγιση στον έλεγχο: το $n^{1/2} [J.b - M(F)]$ είναι ασυμπτωτικά κανονικά κατανομημένο με μέση τιμή 0 και διασπορά $4c$, όπου κάτω από την υπόθεση της εκθετικότητας, $M(F) = (b+1)^{-1}$ και

$$c = \frac{1}{4} \left[1 + \frac{b}{b+2} + \frac{1}{2b+1} + \frac{2(1-b)}{b+1} - \frac{2b}{b^2+b+1} - \frac{4}{(b+1)^2} \right]$$

3.9 Shapiro and Wilk Test (Shapiro and Wilk, 1972)

Έστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από μία συνεχή συνάρτηση κατανομής $F(x)$. Ο έλεγχος Shapiro-Wilk για τη μηδενική υπόθεση :

$$H_0: F(x) = 1 - e^{-\frac{x-a}{b}}, x \geq a$$

χρησιμοποιεί το ακόλουθο στατιστικό:

$$W(E) = \frac{n(\bar{X} - X_{\min})^2}{[(n-1)s]^2} = \frac{n(\bar{X} - X_{\min})^2}{(n-1) \sum_{i=1}^n (X_i - \bar{X})^2} \quad (3.10)$$

όπου s είναι η δειγματική τυπική απόκλιση, X_{\min} η μικρότερη από τις τ. μ. και n το μέγεθος του δείγματος.

Η ακριβής μηδενική κατανομή του $W(E)$ δεν είναι γνωστή, αλλά δεν εξαρτάται από τα α και β . Οι Shapiro και Wilk παρουσίασαν πίνακες με τα άνω και κάτω εκατοστημόρια του $W(E)$ που ελήφθησαν με προσομοίωση, για n από 3 έως 100.

3.10 New Better Than Used Test (Hollander and Proschan , 1972)

Αυτή η διαδικασία, εφαρμόζεται συνήθως σε μονόπλευρες εναλλακτικές υποθέσεις (η νέα υπόθεση είναι καλύτερη από την χρησιμοποιούμενη ή η νέα υπόθεση είναι χειρότερη από την χρησιμοποιούμενη). Σε αυτή την εργασία, δεδομένου ότι έχει υποτεθεί ότι δεν γνωρίζουμε την εναλλακτική υπόθεση, ο έλεγχος ήταν αμφίπλευρος. Η ελεγχοσυνάρτηση του ελέγχου είναι:

$$T = \sum_{i>j>k} G(X_{(i)}, X_{(j)} + X_{(k)}) \quad (3.11), \text{ όπου}$$

$$G(a, b) = \begin{cases} 1 & , \quad \text{αν } a > b \\ 0.5 & , \quad \text{αν } a = b \\ 0 & , \quad \text{αν } a < b \end{cases}$$

Οι συγγραφείς παρέχουν ένα πίνακα με κατά προσέγγιση κατώτερες και ανώτερες κρίσιμες τιμές και την ακόλουθη Κανονική προσέγγιση:

$$T^* = \frac{T - E(T|H_0)}{\sqrt{VAR(T|H_0)}}, \text{ όπου}$$

$$E(T|H_0) = n(n - 1)(n - 2)/8 \text{ και}$$

$$VAR(T|H_0) = \{1.5n(n - 1)(n - 2) \left[\left(\frac{5}{2592} \right) (n - 3)(n - 4) + (n - 3) \left(\frac{7}{432} \right) + \frac{1}{48} \right] \}$$

Όταν η μηδενική υπόθεση είναι αληθής και το n τείνει στο άπειρο, το T^* έχει μια ασυμπτωτικά Κανονική Κατανομή με μέση τιμή 0 και διασπορά 1.

3.11 EPS Test (Epstein , 1960)

Η ελεγχοσυνάρτηση του ελέγχου είναι όπου τα D_i δίνονται στη σχέση (3.1)

$$EPS = 2n[\ln(\sum_{i=1}^n D_i/n) - \frac{1}{n} \sum_{i=1}^n \ln(D_i)]/[1 + (n+1)/6n] \quad (3.12)$$

Δεδομένου ότι η μηδενική υπόθεση είναι αληθής, η ελεγχουσυνάρτηση κατανέμεται προσεγγιστικά σαν μία X^2 με $n-1$ βαθμούς ελευθερίας. Η υπόθεση απορρίπτεται για μεγάλες τιμές του EPS. Αυτή η διαδικασία είναι ισχυρή έναντι εναλλακτικών Γάμμα και Weibull.

3.12 Hartley's F-Max Test (Hartley, 1950)

Αυτός ο έλεγχος, προέκυψε από έναν έλεγχο για την ομοιογένεια των διασπορών. Το στατιστικό του ελέγχου είναι:

$$HARTF = \text{Max}(W_i)/\text{Min}(W_i), \text{ όπου } 1 \leq i \leq K \quad (3.13)$$

$$W_i = \sum_{j=(i-1)R+1}^{iR} D_j$$

όπου τα D_i δίνονται στη σχέση (3.1), $K =$ ο αριθμός των ομάδων και $R =$ το μέγεθος κάθε ομάδας. Δεδομένου ότι η μηδενική υπόθεση είναι αληθής, το HARTF έχει μία F_{Max} κατανομή με $2R$ και K βαθμούς ελευθερίας. Η υπόθεση απορρίπτεται για μεγάλες τιμές του HARTF. Όταν $n = 20$, συστήνεται να χρησιμοποιείται $K = 2$ και $R = 10$.

3.13 Cox and Oakes Score Test (Cox and Oakes, 1984)

Αυτή η διαδικασία βασίζεται στη συνάρτηση Score :

$$U = d + \sum \ln(X_i) - d \sum_{i=1}^n \frac{X_i \ln(X_i)}{\sum_{i=1}^n X_i} \quad (3.14)$$

όπου το πρώτο άθροισμα λαμβάνεται πάνω σε όλα τα μη λογοκριμένα (παρατηρήσιμα) σημεία και d είναι ο αριθμός των μη λογοκριμένων σημείων. Για την παρούσα εργασία, όλα τα σημεία είναι παρατηρήσιμα (δηλ. $d = n$). Με τη χρήση του πίνακα πληροφορίας, μία ασυμπτωτικά τυπική κανονική απόκλιση

μπορεί να υπολογιστεί. Η υπόθεση της εκθετικότητας απορρίπτεται και για τα δύο, για μεγάλες και για μικρές τιμές της ελεγχουσυνάρτησης. Μία χρήσιμη ιδιότητα αυτού του ελέγχου είναι η ικανότητα να χειρίζεται λογοκριμένα δεδομένα. Οι συγγραφείς ισχυρίζονται ότι ο έλεγχος είναι χρήσιμος έναντι εναλλακτικών υποθέσεων οι οποίες προσδιορίζουν μονότονες συναρτήσεις κινδύνου.

3.14 Wong and Wong's Extremal Quotient Test (Wong and Wong, 1979)

Αυτός ο έλεγχος, ο οποίος προτάθηκε από τους Wong και Wong, βασίζεται σε μία ποσότητα γνωστή ως το ακραίο πηλίκο:

$$Q = \frac{X(n)}{X(1)} \quad (3.15)$$

όπου $X(1)$ και $X(n)$ είναι η μικρότερη και η μεγαλύτερη διατεταγμένη τυχαία μεταβλητή του δείγματος, αντίστοιχα. Οι συγγραφείς δίνουν κρίσιμες τιμές για αυτόν τον έλεγχο, ο οποίος απορρίπτει τη μηδενική υπόθεση για μεγάλες τιμές του Q .

3.15 WE Test (Hahn and Shapiro, 1967)

Η ελεγχουσυνάρτηση του ελέγχου είναι :

$$WE = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(\sum_{i=1}^n X_i)^2} = \frac{(n-1)S^2}{n^2 \bar{X}^2} \quad (3.16)$$

όπου S^2 είναι η δειγματική διασπορά και \bar{X} είναι ο δειγματικός μέσος. Κρίσιμες τιμές για αυτόν τον έλεγχο δίνονται από τους συγγραφείς.

3.16 Gnedenko 's F-Test (Gnedenko, 1969)

Για αυτόν το έλεγχο, τα n δεδομένα είναι διατεταγμένα και χωρίζονται σε δύο ομάδες, με την ομάδα ένα να περιέχει τα πρώτα R και την ομάδα δύο τα υπόλοιπα $n-R$ δεδομένα. Η ελεγχουσυνάρτηση του ελέγχου είναι:

$$Q(R) = \frac{\sum_{i=1}^R D_i/R}{\sum_{i=R+1}^n D_i/(n-R)} \quad (3.17)$$

όπου τα D_i δίνονται στη σχέση (3.1).

Αν η μηδενική υπόθεση της εκθετικότητας είναι αληθινή, τότε το $Q(R)$ ακολουθεί μια F κατανομή με $2R$ και $2(n-R)$ βαθμούς ελευθερίας. Η υπόθεση απορρίπτεται για μικρές και για μεγάλες τιμές του $Q(R)$. Οι Fercho και Ringer συνιστούν να τεθεί $R = n/2$ και ισχυρίζονται ότι ο έλεγχος είναι κατάλληλος και για την Weibull κατανομή ως εναλλακτική και για τις Γάμμα κατανομές με μονότονους ρυθμούς κινδύνου.

3.17 Harris modification of Gnedenko 's F-Test (Harris, 1976)

Η ελεγκοσυνάρτηση του ελέγχου είναι :

$$Q'(R) = \frac{(\sum_{i=1}^R D_i + \sum_{i=n-R+1}^n D_i)/2R}{\sum_{i=R+1}^{n-R} D_i/(n-2R)} \quad (3.18)$$

όπου τα D_i δίνονται στη σχέση (3.1).

Το $Q'(R)$ κατανέμεται σαν μία F με $4R$ και $2(n-2R)$ βαθμούς ελευθερίας, με δεδομένο ότι η μηδενική υπόθεση είναι αληθής. Η υπόθεση απορρίπτεται για μικρές και για μεγάλες τιμές του $Q'(R)$. Αυτή η διαδικασία φαίνεται να είναι ισχυρή έναντι της λογαριθμοκανονικής κατανομής (της οποίας η συνάρτηση κινδύνου έχει σχήμα U) και ασθενέστερη για μονότονες συναρτήσεις κινδύνου. Ο Harris συνιστά να τεθεί $R = n/4$.

3.18 Lin and Mudholkar 's Bivariate F-Test (Lin and Mudholkar, 1980)

Αυτός ο έλεγχος, είναι ουσιαστικά ένας συνδυασμός των παραπάνω ελέγχων (3.16 και 3.17). Έστω οι ποσότητες :

$$F_L = \frac{\sum_{i=1}^R D_i/R}{\sum_{i=R+1}^{n-R} D_i/(n-2R)} \quad (3.19)$$

$$F_U = \frac{\sum_{i=n-R+1}^n D_i/R}{\sum_{i=R+1}^{n-R} D_i/(n-2R)} \quad (3.20)$$

όπου τα D_i δίνονται στη σχέση (3.1).

Κάτω από τη μηδενική υπόθεση, οι F_L και F_U από κοινού ακολουθούν μια διμεταβλητή (bivariate) F κατανομή. Απόρριψη της εκθετικότητας θα συμβεί εάν υπάρχει F_L ή F_U που δεν είναι μέσα σε κάποιο διάστημα (a, b) . Αυτό το διάστημα προσδιορίζεται χρησιμοποιώντας το επόμενο θεώρημα από τους Hewett και Bulgren (1971):

Για κάθε $0 \leq a \leq b < +\infty$, $P(a \leq F_L \leq b, a \leq F_U \leq b | H_0) \leq [P(a \leq F \leq b)]^2$, όπου F είναι η F του Snedecor τυχαία μεταβλητή με $2R$ και $2(n-2R)$ βαθμούς ελευθερίας. Το δεξί μέλος της ανισότητας ορίζεται ίσο με $1 - \alpha$ (όπου α είναι το επιθυμητό σφάλμα τύπου I) και υποθέτοντας ίσες ουρές πιθανότητας για την F , τα a και b μπορούν εύκολα να βρεθούν. Αυτή η διαδικασία φαίνεται να είναι ισχυρή έναντι εναλλακτικών με μη μονότονες συναρτήσεις κινδύνων (π.χ. λογαριθμοκανονική). Οι Lin και Mudholkar συνιστούν να χρησιμοποιηθεί $R = n/10$.

ΚΕΦΑΛΑΙΟ 4

Ανάλυση προσομοιωμένων δεδομένων με χρήση του Excel και του στατιστικού πακέτου R

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο, αρχικά θα προσομοιώσουμε δεδομένα τα οποία γνωρίζουμε ότι ακολουθούν την εκθετική κατανομή και θα εφαρμόσουμε τους ελέγχους μας σε αυτά. Αυτό θα γίνει με χρήση του στατιστικού πακέτου R.

Χρησιμοποιώντας αυτά τα δεδομένα, θα παρουσιάσουμε διάφορα διαγράμματα με χρήση της R για την οπτικοποίηση των δεδομένων καθώς και για έναν πρώτο γραφικό έλεγχο. Έπειτα, θα παραθέσουμε πίνακες με τους υπολογισμούς των τιμών των ελεγχουσυναρτήσεων των ελέγχων που επιλέχθηκαν (για τους πίνακες αυτούς έγινε χρήση του προγράμματος Excel) οι οποίοι θα συνοδεύονται με διαγράμματα όπου χρειαστεί.

4.2 Προσομοίωση δεδομένων και διαγράμματα

Αρχικά παράγουμε τα δεδομένα στην R με χρήση της εντολής :

```
>Data=rexp(40,3)
```

Με αυτή την εντολή λαμβάνουμε ένα δείγμα μεγέθους 40, το οποίο προέρχεται από την εκθετική κατανομή με $\lambda=3$. (Οι τιμές θα παρουσιαστούν παρακάτω στους πίνακες)

Για να κατανοήσουμε τη μορφή του δείγματος χρησιμοποιούμε δύο διαγράμματα. Το ιστόγραμμα σχετικών συχνοτήτων (relative frequency histogram) και θηκοδιάγραμμα (boxplot).

Ιστόγραμμα Σχετικών Συχνοτήτων

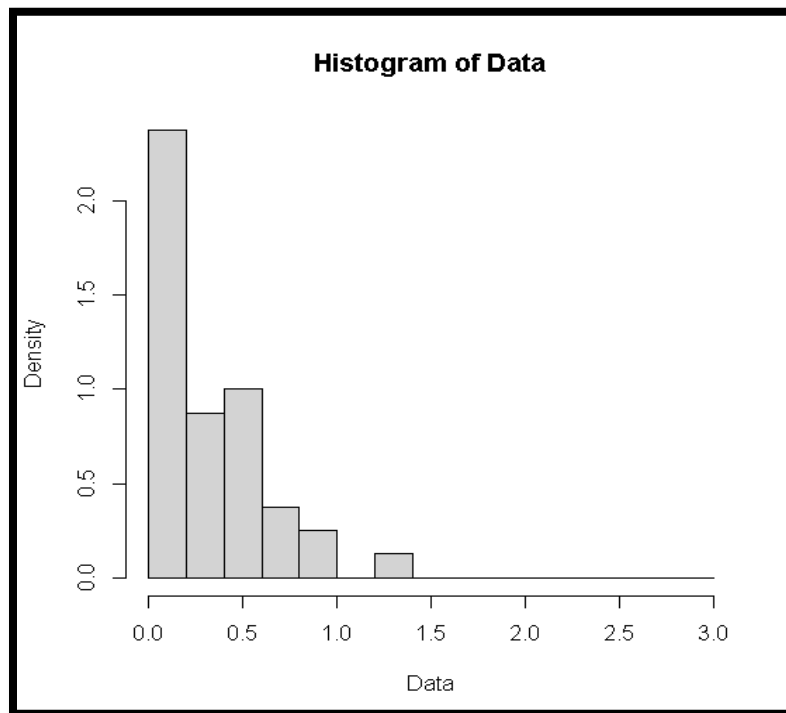
Το ιστόγραμμα σχετικών συχνοτήτων, χωρίζει το δείγμα σε κλάσεις και δημιουργεί διαδοχικά ορθογώνια για κάθε κλάση. Το ύψος των ορθογωνίων αντιστοιχεί στη σχετική συχνότητα των παρατηρήσεων στην αντίστοιχη κλάση. Κάνουμε το ιστόγραμμα για το δείγμα μας μεγέθους 40, και ακριβώς από κάτω παρουσιάζουμε το ιστόγραμμα για δείγμα από εκθετική κατανομή μεγέθους 1000.

Χρησιμοποιούμε τις εντολές :

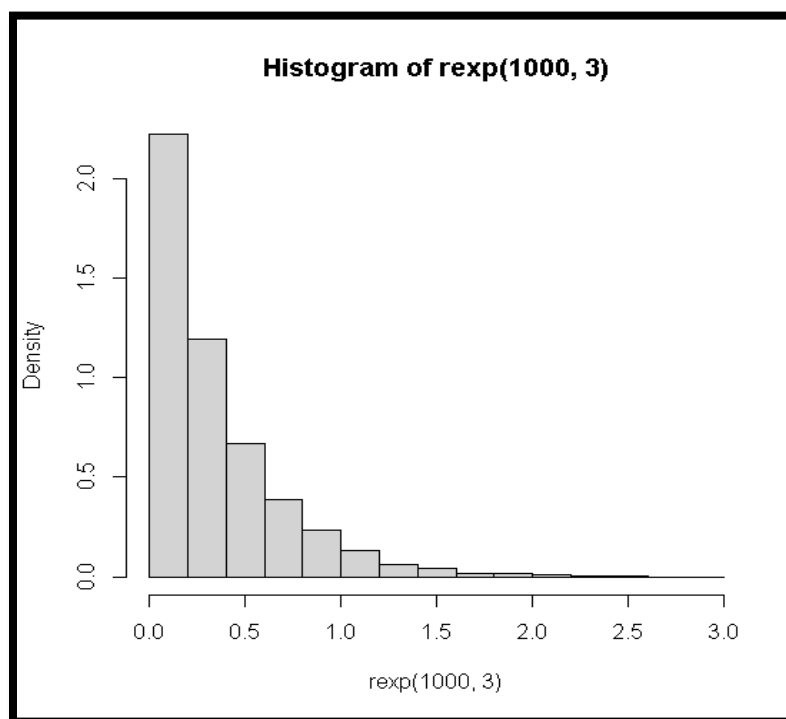
```
> h1=hist(Data,breaks=seq(0,3,by=0.2),probability=T)
```

```
> h2=hist(rexp(1000,3),breaks=seq(0,3,by=0.2),probability=T)
```

και παίρνουμε τα διαγράμματα:



ΕΙΚΟΝΑ 3 : Ιστόγραμμα προσομιωμένων δεδομένων



ΕΙΚΟΝΑ 4 : Ιστόγραμμα προσέγγισης εκθετικής κατανομής

Όπως παρατηρούμε, υπάρχει ομοιότητα στα δύο διαγράμματα παρ' όλο που το πρώτο έχει κατασκευαστεί με 40 δεδομένα ενώ το δεύτερο βασίζεται σε 1000 δεδομένα. Σκοπός μας εδώ η γραφική επιβεβαίωση ότι το προσομοιωμένο δείγμα μας ακολουθεί εκθετική κατανομή με παράμετρο 3.

Θηκοδιάγραμμα

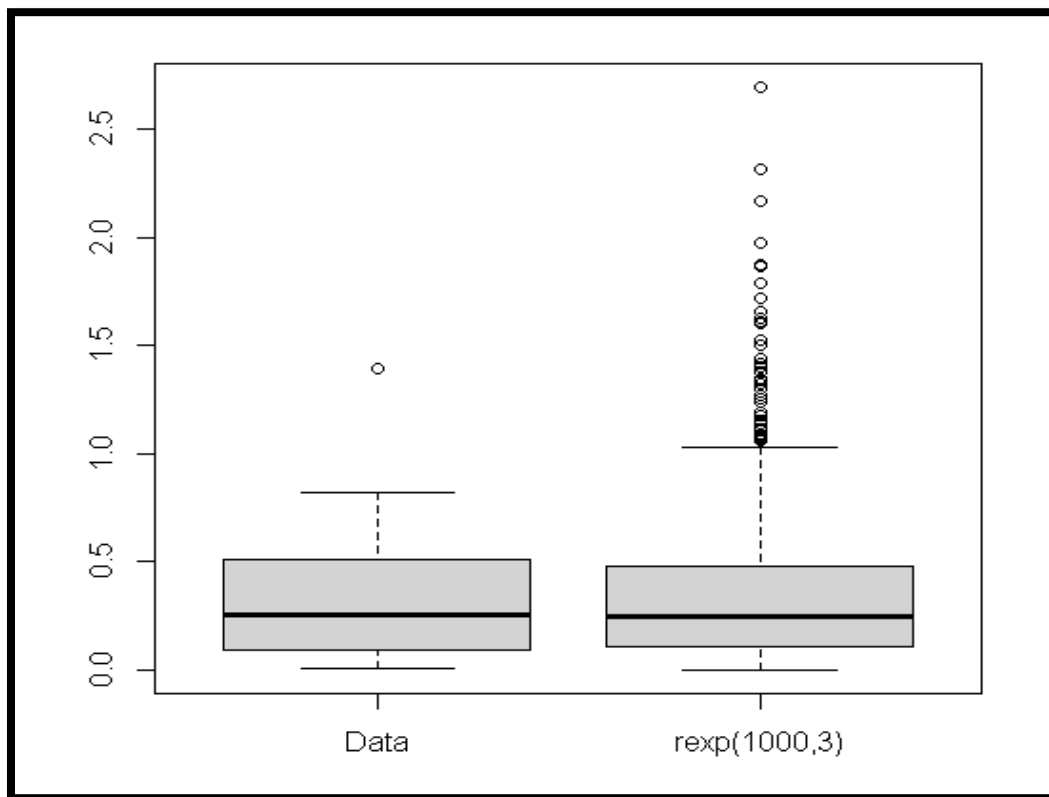
Το θηκοδιάγραμμα είναι ένα ορθογώνιο με κάτω βάση στο πρώτο και άνω βάση στο τρίτο ταταρτημόριο. Για την κατασκευή του φέρνουμε το ευθύγραμμο τμήμα μέσα στο ορθογώνιο το οποίο αντιστοιχεί στη διάμεσο. Έπειτα φέρουμε ευθύγραμμο τμήματα στις δύο οριακές τιμές των παρατηρήσεων που ορίζονται ως το τρίτο (αντίστοιχα το πρώτο) τεταρτημόριο συν (αντίστοιχα μείον) 1.5 φορές το ενδοτεταρτημοριακό εύρος. Αν δεν υπάρχουν παρατηρήσεις τόσο απομακρυσμένες οι γραμμές τοποθετούνται πιο κοντά στο πρώτο και τρίτο τεταρτημόριο. Οι οριζόντιες γραμμές τελειώνουν στην ελάχιστη και μέγιστη παρατήρηση αντίστοιχα αν δεν υπάρχουν έκτροπες παρατηρήσεις. Έκτροπες τιμές (τιμές πιο ακραίες των δύο οριακών, παριστάνονται με κουκίδες. Παρουσιάζουμε το θηκοδιάγραμμα του δείγματος μας, σε σύγκριση με το θηκοδιάγραμμα της εκθετικής κατανομής με βάση δείγμα 1000 και $\lambda=3$.

Προφανώς, το δεύτερο θηκοδιάγραμμα θα έχει περισσότερες έκτροπες τιμές, λόγω του μεγέθους του δείγματος. Η μέση τιμή της εκθετικής κατανομής που υποθέσαμε είναι $1/3$ αλλά παρατηρούμε ότι η διάμεσος είναι μικρότερη του $1/3$ λόγω του ότι η κατανομή είναι λοξή προς τα δεξιά, κάτι που φαίνεται και από τη θέση των έκτροπων τιμών.

Γράφουμε την εντολή :

```
> boxplot(Data, rexp(1000,3), names=c("Data", "rexp(1000,3)"))
```

Και παίρνουμε τα παρακάτω διαγράμματα.



ΕΙΚΟΝΑ 5 : Θηκοδιάγραμμα προσωμειωμένων δεδομένων

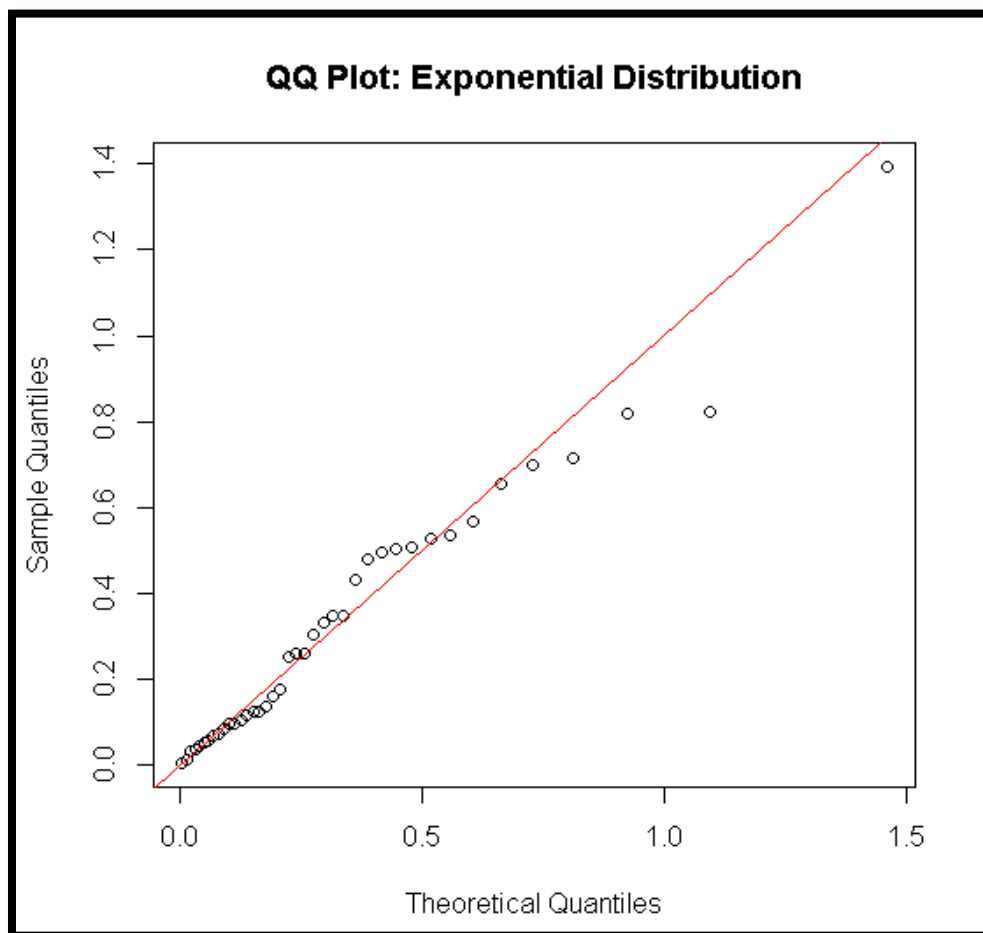
Παρατηρούμε ξανά την ομοιότητα των δύο θηκοδιαγραμμάτων.

Διάγραμμα Q-Q Plot

Αναλύσαμε στο προηγούμενο κεφάλαιο τι είναι το Q-Q Plot. Αφού πήραμε μία γραφική εικόνα του δείγματος μας σε σύγκριση με την εκθετική κατανομή $E(3)$, τώρα χρησιμοποιούμε το Q-Q Plot για να κάνουμε έναν πιο αυστηρό γραφικό έλεγχο. Εδώ θα χρησιμοποιήσουμε την εντολή στην R :

```
> theoretical_quantiles <- qexp(ppoints(length(Data)), rate = 3)
> qqplot(theoretical_quantiles, Data,xlab = "Theoretical Quantiles",ylab =
"Sample Quantiles",main = "QQ Plot: Exponential Distribution")
> abline(a = 0, b = 1, col = "red")
```

Και παίρνουμε το παρακάτω διάγραμμα :



ΕΙΚΟΝΑ 6 : Διάγραμμα Q-Q Plot

Όπως παρατηρούμε, τα σημεία βρίσκονται πολύ κοντά στην ευθεία $y=x$ με μία εξαίρεση. Άρα επιβεβαιώνεται μέσω του Q-Q Plot ότι το δείγμα μας ακολουθεί την εκθετική κατανομή με $\lambda=3$.

4.3 Έλεγχοι Kolmogorov , Lilliefors, Hartley και KUSK

4.3.1 Έλεγχος Hartley's F-Max

Παρουσιάζουμε τους υπολογισμούς για τον έλεγχο Hartley's F-Max:

Rank	X_i	D_i	Rank	X_i	D_i	W_i
1	0.0069	0.2778	21	0.2605	0.1842	2.9802
2	0.0126	0.2200	22	0.2621	0.0309	3.91135
3	0.0320	0.7384	23	0.3060	0.7903	3.80256
4	0.0395	0.2779	24	0.3307	0.4196	2.18279
5	0.0443	0.1718	25	0.3469	0.2586	
6	0.0532	0.3107	26	0.3483	0.0208	HARTF
7	0.0565	0.1117	27	0.4329	1.1850	1.79191
8	0.0711	0.4821	28	0.4797	0.6090	
9	0.0731	0.0649	29	0.4979	0.2175	
10	0.0836	0.3250	30	0.5057	0.0866	
11	0.0959	0.3701	31	0.5096	0.0386	
12	0.0979	0.0573	32	0.5264	0.1511	
13	0.1061	0.2306	33	0.5344	0.0641	
14	0.1171	0.2953	34	0.5683	0.2372	
15	0.1243	0.1875	35	0.6569	0.5314	
16	0.1267	0.0593	36	0.6978	0.2048	
17	0.1362	0.2298	37	0.7151	0.0691	
18	0.1615	0.5821	38	0.8176	0.3074	
19	0.1763	0.3239	39	0.8217	0.0083	
20	0.2513	1.5753	40	1.3924	0.5707	

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 3: Υπολογισμός ελεγχουσυνάρτησης Hartley's F-Max Test

Όπως βλέπουμε, η τιμή της ελεγχουσυνάρτησης, η οποία ορίζεται στη σχέση (3.13) είναι $HARTF=1.79191$. Από τον ΠΙΝΑΚΑ 7 του προηγούμενου Κεφαλαίου της κατανομής F με 20 και 4 βαθμούς ελευθερίας (στους υπολογισμούς, διαλέξαμε $R=10$ και $K=4$) και επίπεδο σημαντικότητας $\alpha=0.05$ έχουμε ότι η

κρίσιμη τιμή του ελέγχου είναι 5.8. Αφού $1.79 < 5.8$, συμπεραίνουμε ότι δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση.

4.3.2 Έλεγχος KUSK

Παρουσιάζουμε τους υπολογισμούς για τον έλεγχο KUSK:

Rank	χ_i	$\chi_i - \bar{\chi}$	$(\chi_i - \bar{\chi})^2$	$(\chi_i - \bar{\chi})^3$	$(\chi_i - \bar{\chi})^4$	Rank	χ_i	$\chi_i - \bar{\chi}$	$(\chi_i - \bar{\chi})^2$	$(\chi_i - \bar{\chi})^3$	$(\chi_i - \bar{\chi})^4$
1	0.0069	-0.3150	0.0992	-0.0312	0.0098	21	0.260	-0.061	0.004	0.000	0.000
2	0.0126	-0.3093	0.0957	-0.0296	0.0092	22	0.262	-0.060	0.004	0.000	0.000
3	0.0320	-0.2899	0.0840	-0.0244	0.0071	23	0.306	-0.016	0.000	0.000	0.000
4	0.0395	-0.2824	0.0797	-0.0225	0.0064	24	0.331	0.009	0.000	0.000	0.000
5	0.0443	-0.2776	0.0771	-0.0214	0.0059	25	0.347	0.025	0.001	0.000	0.000
6	0.0532	-0.2687	0.0722	-0.0194	0.0052	26	0.348	0.026	0.001	0.000	0.000
7	0.0565	-0.2655	0.0705	-0.0187	0.0050	27	0.433	0.111	0.012	0.001	0.000
8	0.0711	-0.2509	0.0629	-0.0158	0.0040	28	0.480	0.158	0.025	0.004	0.001
9	0.0731	-0.2488	0.0619	-0.0154	0.0038	29	0.498	0.176	0.031	0.005	0.001
10	0.0836	-0.2383	0.0568	-0.0135	0.0032	30	0.506	0.184	0.034	0.006	0.001
11	0.0959	-0.2260	0.0511	-0.0115	0.0026	31	0.510	0.188	0.035	0.007	0.001
12	0.0979	-0.2240	0.0502	-0.0112	0.0025	32	0.526	0.204	0.042	0.009	0.002
13	0.1061	-0.2158	0.0466	-0.0100	0.0022	33	0.534	0.212	0.045	0.010	0.002
14	0.1171	-0.2049	0.0420	-0.0086	0.0018	34	0.568	0.246	0.061	0.015	0.004
15	0.1243	-0.1976	0.0391	-0.0077	0.0015	35	0.657	0.335	0.112	0.038	0.013
16	0.1267	-0.1953	0.0381	-0.0074	0.0015	36	0.698	0.376	0.141	0.053	0.020
17	0.1362	-0.1857	0.0345	-0.0064	0.0012	37	0.715	0.393	0.155	0.061	0.024
18	0.1615	-0.1604	0.0257	-0.0041	0.0007	38	0.818	0.496	0.246	0.122	0.060
19	0.1763	-0.1457	0.0212	-0.0031	0.0005	39	0.822	0.500	0.250	0.125	0.062
20	0.2513	-0.0706	0.0050	-0.0004	0.0000	40	1.392	1.071	1.146	1.227	1.313

μ_2	μ_3	μ_4
0.0864	0.0350	0.0395
β_1	β_2	
1.8940	5.2821	
KUSK		
0.4532		

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 4:
Υπολογισμός
ελεγχουσυνάρτησης KUSK

Όπως βλέπουμε η τιμή της ελεγχουσυνάρτησης KUSK, η οποία ορίζεται στη σχέση (3.3), είναι κοντά στο 0.5. Αυτό μας δίνει την πληροφορία ότι το δείγμα μας πιθανόν να προέρχεται από την εκθετική κατανομή.

4.3.3 Έλεγχος Kolmogorov

Παρουσιάζουμε τους υπολογισμούς για τον έλεγχο του Kolmogorov:

Rank	χ_i	$F_n(\chi_i)$	$F_n(\chi_i^-)$	$F_0(\chi_i)$	$ F_n(\chi_i)-F_0(\chi_i) $	$ F_0(\chi_i)-F_n(\chi_i^-) $	Rank	χ_i	$F_n(\chi_i)$	$F_n(\chi_i^-)$	$F_0(\chi_i)$	$ F_n(\chi_i)-F_0(\chi_i) $	$ F_0(\chi_i)-F_n(\chi_i^-) $	KS TEST
1	0.0069	0.0250	0.0000	0.0206	0.0044	0.0206	21	0.2605	0.5250	0.5000	0.5423	0.0173	0.0423	0.0895
2	0.0126	0.0500	0.0250	0.0371	0.0129	0.0121	22	0.2621	0.5500	0.5250	0.5445	0.0055	0.0195	
3	0.0320	0.0750	0.0500	0.0916	0.0166	0.0416	23	0.3060	0.5750	0.5500	0.6007	0.0257	0.0507	
4	0.0395	0.1000	0.0750	0.1118	0.0118	0.0368	24	0.3307	0.6000	0.5750	0.6292	0.0292	0.0542	
5	0.0443	0.1250	0.1000	0.1245	0.0005	0.0245	25	0.3469	0.6250	0.6000	0.6468	0.0218	0.0468	
6	0.0532	0.1500	0.1250	0.1475	0.0025	0.0225	26	0.3483	0.6500	0.6250	0.6482	0.0018	0.0232	
7	0.0565	0.1750	0.1500	0.1558	0.0192	0.0058	27	0.4329	0.6750	0.6500	0.7271	0.0521	0.0771	
8	0.0711	0.2000	0.1750	0.1920	0.0080	0.0170	28	0.4797	0.7000	0.6750	0.7629	0.0629	0.0879	
9	0.0731	0.2250	0.2000	0.1969	0.0281	0.0031	29	0.4979	0.7250	0.7000	0.7754	0.0504	0.0754	
10	0.0836	0.2500	0.2250	0.2218	0.0282	0.0032	30	0.5057	0.7500	0.7250	0.7807	0.0307	0.0557	
11	0.0959	0.2750	0.2500	0.2501	0.0249	0.0001	31	0.5096	0.7750	0.7500	0.7832	0.0082	0.0332	
12	0.0979	0.3000	0.2750	0.2545	0.0455	0.0205	32	0.5264	0.8000	0.7750	0.7938	0.0062	0.0188	
13	0.1061	0.3250	0.3000	0.2727	0.0523	0.0273	33	0.5344	0.8250	0.8000	0.7987	0.0263	0.0013	
14	0.1171	0.3500	0.3250	0.2962	0.0538	0.0288	34	0.5683	0.8500	0.8250	0.8182	0.0318	0.0068	
15	0.1243	0.3750	0.3500	0.3112	0.0638	0.0388	35	0.6569	0.8750	0.8500	0.8606	0.0144	0.0106	
16	0.1267	0.4000	0.3750	0.3161	0.0839	0.0589	36	0.6978	0.9000	0.8750	0.8767	0.0233	0.0017	
17	0.1362	0.4250	0.4000	0.3355	0.0895	0.0645	37	0.7151	0.9250	0.9000	0.8830	0.0420	0.0170	
18	0.1615	0.4500	0.4250	0.3841	0.0659	0.0409	38	0.8176	0.9500	0.9250	0.9139	0.0361	0.0111	
19	0.1763	0.4750	0.4500	0.4107	0.0643	0.0393	39	0.8217	0.9750	0.9500	0.9150	0.0600	0.0350	
20	0.2513	0.5000	0.4750	0.5294	0.0294	0.0544	40	1.3924	1.0000	0.9750	0.9847	0.0153	0.0097	

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 5: Υπολογισμός ελεγχουσυνάρτησης Kolmogorov

Όπως βλέπουμε, η ελεγχουσυνάρτηση για τον έλεγχο του Kolmogorov, η οποία ορίζεται στη σχέση (3.4), είναι ίση με 0.0895. Από τον ΠΙΝΑΚΑ 2 του προηγούμενου Κεφαλαίου έχουμε ότι η κρίσιμη τιμή για δείγμα μεγέθους 40 και επίπεδο σημαντικότητας $\alpha=0.05$ είναι 0.2107. Αφού $0.0895 < 0.2107$ δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση, επομένως μπορούμε να ισχυριστούμε ότι το δείγμα μας προέρχεται από την εκθετική κατανομή με $\lambda=3$.

Μπορούμε χρησιμοποιώντας την R να πάρουμε την ίδια τιμή έτοιμη με την εντολή:

```
> ks.test(Data,"pexp",3)
```

Και παίρνουμε :

Exact one-sample Kolmogorov-Smirnov test

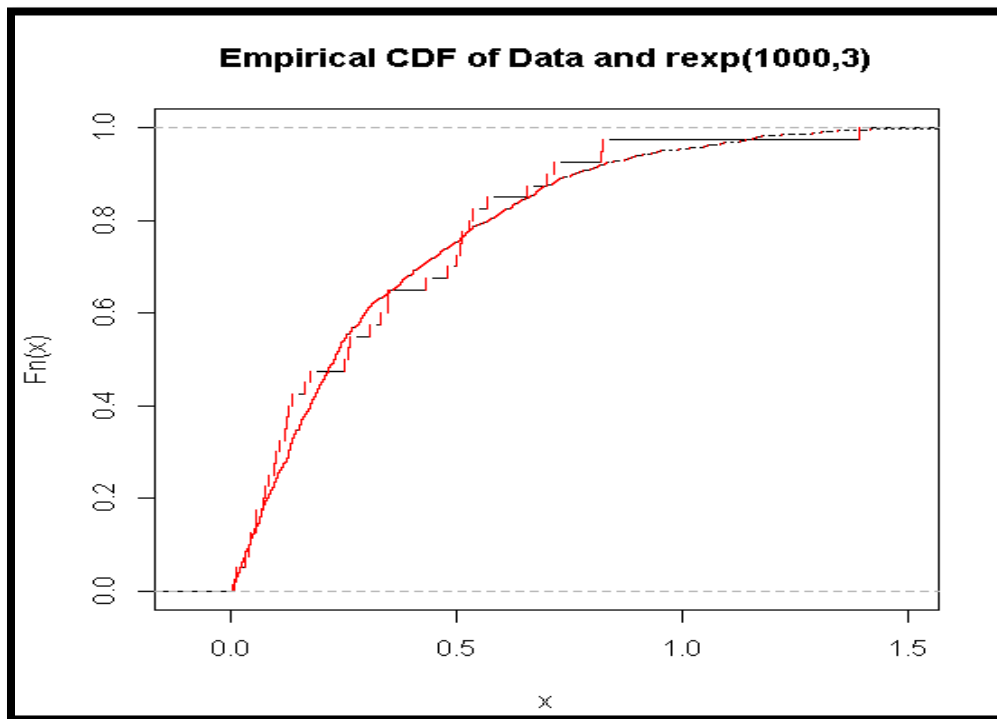
```
data: Data
D = 0.089522, p-value = 0.8773
alternative hypothesis: two-sided
```

Η τιμή D είναι ακριβώς η ίδια που πήραμε πριν. Η p-value είναι αρκετά μεγάλη για να απορρίψουμε τη μηδενική υπόθεση.

Μπορούμε να παραστήσουμε και γραφικά τις διαφορές $|F_n(X_i) - F_0(X_i)|$ και $|F_0(X_i) - F_n(X_{i-})|$. Χρησιμοποιώντας την εντολή :

```
> plot.ecdf(Data, verticals=TRUE, col.p="white", col.v="red", main="Empirical
CDF of Data and rexp(1000,3)")
> plot(ecdf(rexp(1000,3)), add=TRUE, verticals=TRUE, col.p="white",
col.v="red", lty="dotted")
```

Παίρνουμε τη γραφική παράσταση :



ΕΙΚΟΝΑ 7 : Γραφική παράσταση εμπειρικής συνάρτησης κατανομής (Kolmogorov)

Όπως φαίνεται και γραφικά, η εμπειρική αθροιστική συνάρτηση κατανομής βρίσκεται αρκετά κοντά στην αθροιστική συνάρτηση κατανομής της $E(3)$, κάτι που αποτυπώνεται από τον έλεγχο του Kolmogorov.

4.3.4 Έλεγχος Lilliefors

Παρουσιάζουμε τους υπολογισμούς για τον έλεγχο του Lilliefors:

Rank	X_i	Z_i	$F(Z_i)$	$F_n^*(Z_i)$	$F_n^*(Z_i^-)$	$ F(Z_i) - F_n^*(Z_i) $	$ F(Z_i) - F_n^*(Z_i^-) $	Rank	X_i	Z_i	$F(Z_i)$	$F_n^*(Z_i)$	$F_n^*(Z_i^-)$	$ F(Z_i) - F_n^*(Z_i) $	$ F(Z_i) - F_n^*(Z_i^-) $	Lilliefors Test
1	0.0069	0.0216	0.0213	0.0250	0.0000	0.0037	0.0213	21	0.2605	0.8092	0.5548	0.5250	0.5000	0.0298	0.0548	0.0997
2	0.0126	0.0391	0.0383	0.0500	0.0250	0.0117	0.0133	22	0.2621	0.8142	0.5570	0.5500	0.5250	0.0070	0.0320	
3	0.0320	0.0995	0.0947	0.0750	0.0500	0.0197	0.0447	23	0.3060	0.9506	0.6135	0.5750	0.5500	0.0385	0.0635	
4	0.0395	0.1228	0.1156	0.1000	0.0750	0.0156	0.0406	24	0.3307	1.0273	0.6420	0.6000	0.5750	0.0420	0.0670	
5	0.0443	0.1376	0.1286	0.1250	0.1000	0.0036	0.0286	25	0.3469	1.0775	0.6595	0.6250	0.6000	0.0345	0.0595	
6	0.0532	0.1652	0.1523	0.1500	0.1250	0.0023	0.0273	26	0.3483	1.0818	0.6610	0.6500	0.6250	0.0110	0.0360	
7	0.0565	0.1754	0.1609	0.1750	0.1500	0.0141	0.0109	27	0.4329	1.3447	0.7394	0.6750	0.6500	0.0644	0.0894	
8	0.0711	0.2208	0.1981	0.2000	0.1750	0.0019	0.0231	28	0.4797	1.4902	0.7747	0.7000	0.6750	0.0747	0.0997	
9	0.0731	0.2271	0.2031	0.2250	0.2000	0.0219	0.0031	29	0.4979	1.5466	0.7870	0.7250	0.7000	0.0620	0.0870	
10	0.0836	0.2596	0.2287	0.2500	0.2250	0.0213	0.0037	30	0.5057	1.5710	0.7922	0.7500	0.7250	0.0422	0.0672	
11	0.0959	0.2980	0.2577	0.2750	0.2500	0.0173	0.0077	31	0.5096	1.5830	0.7946	0.7750	0.7500	0.0196	0.0446	
12	0.0979	0.3041	0.2622	0.3000	0.2750	0.0378	0.0128	32	0.5264	1.6351	0.8051	0.8000	0.7750	0.0051	0.0301	
13	0.1061	0.3297	0.2808	0.3250	0.3000	0.0442	0.0192	33	0.5344	1.6600	0.8099	0.8250	0.8000	0.0151	0.0099	
14	0.1171	0.3637	0.3049	0.3500	0.3250	0.0451	0.0201	34	0.5683	1.7653	0.8289	0.8500	0.8250	0.0211	0.0039	
15	0.1243	0.3861	0.3203	0.3750	0.3500	0.0547	0.0297	35	0.6569	2.0404	0.8700	0.8750	0.8500	0.0050	0.0200	
16	0.1267	0.3934	0.3253	0.4000	0.3750	0.0747	0.0497	36	0.6978	2.1677	0.8856	0.9000	0.8750	0.0144	0.0106	
17	0.1362	0.4232	0.3450	0.4250	0.4000	0.0800	0.0550	37	0.7151	2.2214	0.8915	0.9250	0.9000	0.0335	0.0085	
18	0.1615	0.5018	0.3946	0.4500	0.4250	0.0554	0.0304	38	0.8176	2.5396	0.9211	0.9500	0.9250	0.0289	0.0039	
19	0.1763	0.5475	0.4216	0.4750	0.4500	0.0534	0.0284	39	0.8217	2.5526	0.9221	0.9750	0.9500	0.0529	0.0279	
20	0.2513	0.7806	0.5418	0.5000	0.4750	0.0418	0.0668	40	1.3924	4.3254	0.9868	1.0000	0.9750	0.0132	0.0118	

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ 6 : Υπολογισμός ελεγχουσυνάρτησης Lilliefors

Όπως βλέπουμε, η ελεγχουσυνάρτηση, η οποία ορίζεται στη σχέση (3.5), για τον έλεγχο του Lilliefors είναι 0.0997. Από τον ΠΙΝΑΚΑ 3 του προηγούμενου Κεφαλαίου έχουμε ότι η κρίσιμη τιμή για δείγμα μεγέθους 40 και επίπεδο σημαντικότητας $\alpha=0.05$ είναι 0.1681. Αφού $0.0997 < 0.1681$ δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση, επομένως μπορούμε να ισχυριστούμε ότι το δείγμα μας προέρχεται από την εκθετική κατανομή με $\lambda=3$.

Μπορούμε χρησιμοποιώντας την R να πάρουμε την ίδια τιμή. Αρχικά, για τον έλεγχο του Lilliefors πρέπει να φορτώσουμε το πακέτο KScorrect. Χρησιμοποιούμε τις εντολές :

```
> install.packages("KScorrect")
> library(KScorrect)
```

Έπειτα για την υλοποίηση του ελέγχου την εντολή :

```
> LcKS(Data,"rexp")
```

Και παίρνουμε το παρακάτω αποτέλεσμα :

```
§D.obs
[1] 0.09968313

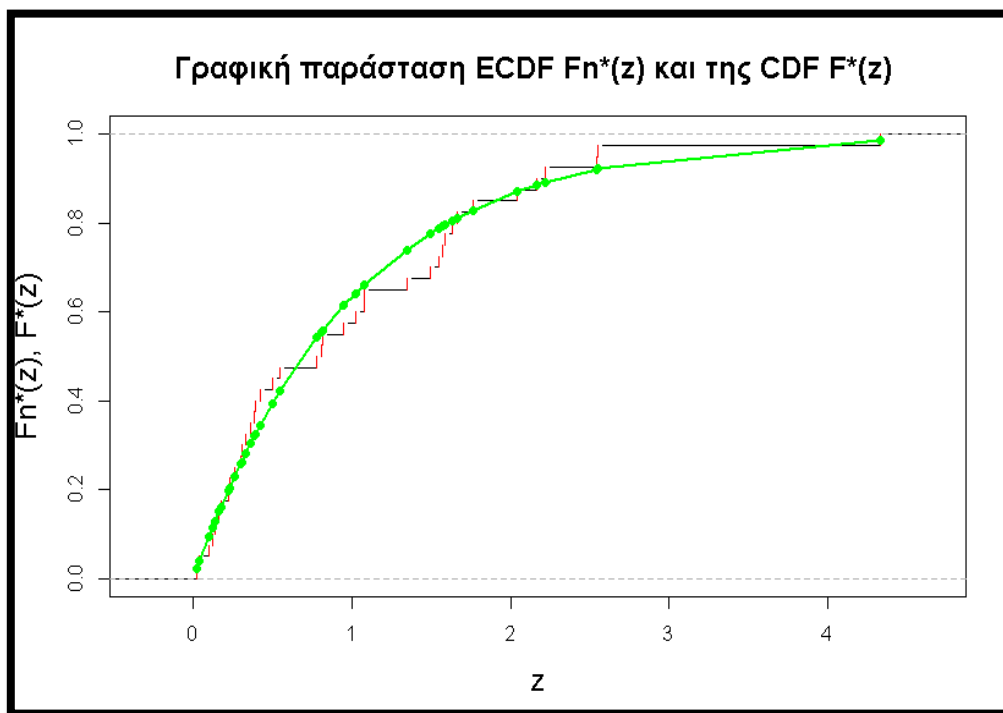
§p.value
[1] 0.6016
```

Όπως υπολογίσαμε και πριν αναλυτικά , η τιμή της ελεγχοσυνάρτησης είναι η ίδια. Από την p-value φαίνεται ότι δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση.

Μπορούμε να παραστήσουμε και γραφικά τις διαφορές $|F_n^*(Z_i) - F^*(Z_i)|$ και $|F^*(Z_i) - F_n^*(Z_{i-})|$. Χρησιμοποιώντας τις εντολές :

```
> DataZ=Data/mean(Data) (Μετασχηματισμός δεδομένων)
> DataY=1-exp(-DataZ) (Παραγωγή συνάρτησης  $F^*(Z_i)$ )
> plot.ecdf(DataZ, verticals=TRUE, col.p="white", col.v="red", xlab="z",
ylab="Fn*(z), F*(z)",cex.lab=1.5, main="Γραφική παράσταση ECDF Fn*(z) και
της CDF F*(z) ", cex.main=1.5)
> lines(DataZ, DataY, type="o", pch=16, lty=1, lwd=2, col="green")
```

παίρνουμε την παρακάτω γραφική παράσταση:



ΕΙΚΟΝΑ 8 : Γραφική παράσταση εμπειρικής συνάρτησης κατανομής (Lilliefors)

4.4 Έλεγχοι εκθετικότητας με χρήση του στατιστικού πακέτου R

Για να υλοποιήσουμε τους ελέγχους, αρχικά θα πρέπει να κατεβάσουμε το αρχείο `exptest_1.2.tar.gz` το οποίο μας παρέχει τους ελέγχους για την εκθετικότητα που χρειαζόμαστε. Έπειτα πρέπει να εγκαταστήσουμε το πακέτο στην R. Στο μενού της R, επιλέγουμε

```
packages>Install packages from local files...
```

και επιλέγουμε το παραπάνω αρχείο. Στην R εμφανίζεται η εντολή :

```
> utils:::menuInstallLocal()
```

Δίνουμε την εντολή στην R να φορτώσει τη βιβλιοθήκη :

```
> library(exptest)
```

Και τώρα είμαστε έτοιμοι να υλοποιήσουμε τους ελέγχους.

4.4.1 Gnedenko 's F-Test

```
> gnedenko.exp.test(Data)
```

```
      Gnedenko's F-test of exponentiality  
  
data:  Data  
Q = 1.1514, p-value = 0.6579
```

4.4.2 Harris modification of Gnedenko 's F-Test

```
> harris.exp.test(Data)
```

```
      Harris modification of Gnedenko's F-test  
  
data:  Data  
Q = 0.66931, p-value = 0.2085
```

4.4.3 The Gini Statistic

```
> gini.exp.test(Data)
```

```
      Test for exponentiality based on the Gini statistic  
  
data:  Data  
Gn = 0.49262, p-value = 0.8732
```

.

4.4.4 The Lorenz Statistic

```
> lorenz.exp.test(Data)
```

```
      Lorenz test for exponentiality  
  
data:  Data  
L = 0.14491, p-value = 0.9571
```

4.4.5 The Pietra Statistic

```
> pietra.exp.test(Data)
```

```
      Test for exponentiality based on the Pietra statistic  
  
data:  Data  
Pn = 0.36574, p-value = 0.974
```

4.4.6 Dehspande's Test

```
> deshpande.exp.test(Data)
```

```
      Deshpande test for exponentiality  
  
data:  Data  
J = 0.68974, p-value = 0.8423
```

4.4.7 Shapiro and Wilk Test

```
> shapiro.exp.test(Data)
```

```
Shapiro-Wilk test for exponentiality
```

```
data: Data  
W = 0.029435, p-value = 0.6325
```

4.4.8 New Better than Used Test

```
> hollander.exp.test(Data)
```

```
Hollander-Proshan test for exponentiality
```

```
data: Data  
T = 0.25388, p-value = 0.8196
```

4.4.9 EPS Test

```
> epstein.exp.test(Data)
```

```
Epstein test for exponentiality
```

```
data: Data  
EPS = 32.489, p-value = 0.7599
```

4.4.10 Cox and Oakes Score Test

```
> co.exp.test(Data)
```

```
Test for exponentiality based on the statistic of Cox and Oakes
```

```
data: Data  
COn = 3.2427, p-value = 0.6893
```

4.4.11 Wong and Wong's Test

```
> ww.exp.test(Data)
```

```
      Wong and Wong's test for exponentiality  
  
data:  Data  
Q = 200.49, p-value = 0.58
```

4.4.12 WE Test

```
> we.exp.test(Data)
```

```
      WE test for exponentiality  
  
data:  Data  
WE = 0.020848, p-value = 0.753
```

Όλες οι p-values είναι πολύ μεγάλες, με μεγαλύτερη αυτή της ελεγχοσυνάρτησης Pietra (0.0974) και ως εκ τούτου η μηδενική υπόθεση της εκθετικότητας δεν μπορεί να απορριφθεί.

ΚΕΦΑΛΑΙΟ 5

Επεξεργασία και έλεγχος εκθετικότητας πραγματικών δεδομένων με χρήση του Excel και του στατιστικού πακέτου R

5.1 Εισαγωγή

Στο προηγούμενο Κεφάλαιο, τα δεδομένα με τα οποία ασχοληθήκαμε ήταν προσομοιωμένα με χρήση του στατιστικού πακέτου R. Τα δεδομένα κατασκευάστηκαν να ακολουθούν την εκθετική κατανομή με συγκεκριμένη παράμετρο.

Η ουσία του προηγούμενου κεφαλαίου δεν ήταν τόσο η επεξεργασία των δεδομένων, αλλά η κατανόηση των ελέγχων, των διαγραμμάτων και παράλληλα για κάποιους ελέγχους η παρουσίαση του τρόπου με τον οποίο υπολογίζεται η τιμή της ελεγχοσυνάρτησης, καθώς και η προετοιμασία του στατιστικού πακέτου R με κατάλληλες βιβλιοθήκες και πακέτα για την υλοποίηση των ελέγχων για την εκθετικότητα.

Σε αυτό το κεφάλαιο, θα ασχοληθούμε με πραγματικά δεδομένα και στόχος είναι η προετοιμασία των δεδομένων για επεξεργασία, καθώς και ο στατιστικός έλεγχος για το αν ακολουθούν την εκθετική κατανομή με συγκεκριμένη παράμετρο.

Για την άντληση των δεδομένων χρησιμοποιήσαμε το Kaggle. Τα δεδομένα που αντλήσαμε από το Kaggle αφορούν την ημερομηνία, την ώρα, την τοποθεσία, το βάθος, το μέγεθος και την πηγή κάθε σεισμού με αναφερόμενο μέγεθος 5,5 ρίχτερ ή μεγαλύτερο από το 1965 μέχρι και το 2016.

Πηγή των δεδομένων μας :

<https://www.kaggle.com/datasets/usgs/earthquake-database>

Η συλλογή των δεδομένων πραγματοποιήθηκε από το Εθνικό Κέντρο Πληροφοριών για τους Σεισμούς (NEIC) των Ηνωμένων Πολιτειών Αμερικής. Το Εθνικό Κέντρο Πληροφοριών για τους Σεισμούς (NEIC) καθορίζει τη θέση και το μέγεθος όλων των σημαντικών σεισμών που συμβαίνουν παγκοσμίως και διανέμει αυτές τις πληροφορίες αμέσως σε εθνικούς και διεθνείς φορείς, επιστήμονες, κρίσιμες εγκαταστάσεις και στο ευρύ κοινό. Το NEIC καταρτίζει και παρέχει στους επιστήμονες και στο κοινό μια εκτενή σεισμική βάση δεδομένων που χρησιμεύει ως βάση για επιστημονική έρευνα μέσω της λειτουργίας σύγχρονων ψηφιακών εθνικών και παγκόσμιων δικτύων σειсмоγράφων και συνεργατικών διεθνών συμφωνιών. Το NEIC είναι το εθνικό κέντρο δεδομένων και το αρχείο πληροφοριών για τους σεισμούς.

Σε αυτήν την εργασία, θα χρησιμοποιήσουμε μόνο τους σεισμούς που έχουν ένταση που είναι ίση ή ξεπερνάει τα 7.5 ρίχτερ. Συγκεκριμένα θα ελέγξουμε την υπόθεση, αν το χρονικό διάστημα που μεσολαβεί για την πραγματοποίηση του επόμενου σεισμού ακολουθεί την εκθετική κατανομή με συγκεκριμένη παράμετρο.

5.2 Προετοιμασία δεδομένων.

5.2.1 Τακτοποίηση δεδομένων με χρήση του Excel

Αρχικά, κατεβάζουμε το αρχείο με τα δεδομένα που παρέχεται από το Kaggle. Στο αρχείο παρατηρούμε πολλές στήλες, οι οποίες δεν μας είναι χρήσιμες για τον συγκεκριμένο έλεγχο.

Διατηρούμε τις στήλες με όνομα "Date" και "Magnitude" διαγράφοντας όλες τις υπόλοιπες με τα βήματα : Home>Delete>Delete Sheet Columns επιλέγοντας όλες τις στήλες που θέλουμε να διαγράψουμε.

Στη συνέχεια, θα επιλέξουμε τα δεδομένα που μας ενδιαφέρουν, δηλαδή θα διαλέξουμε μόνο τις σειρές που αφορούν τους σεισμούς που έχουν ένταση από 7.5 ρίχτερ και περισσότερο. Αυτό το κάνουμε, επιλέγοντας όλα τα δεδομένα στη στήλη "Magnitude" και ύστερα εκτελούμε τα βήματα : Data>Filter.

Στο εικονίδιο που θα εμφανιστεί στο πρώτο κελί των δεδομένων που διαλέξαμε,

επιλέγουμε το βελάκι που εμφανίστηκε και ύστερα : NumberFilters>Less Than. Συμπληρώνουμε την τιμή 7.5, με αυτόν τον τρόπο επιλέγουμε όλες τις τιμές που είναι μικρότερες από 7.5. Με δεξί κλικ στην πρώτη επιλεγμένη τιμή διαλέγουμε την επιλογή Delete Row. Έτσι , διαγράφουμε όλες τις γραμμές που οι σεισμοί τους έχουν ένταση μικρότερη των 7.5 ρίχτερ. Άρα οι γραμμές που έμειναν αφορούν μόνο τους σεισμούς που μας ενδιαφέρουν.

Στη συνέχεια, δημιουργούμε μία Τρίτη στήλη με όνομα DataDays (η στήλη των ημερών που πέρασαν ανάμεσα σε δύο σεισμούς).

Για να συμπληρώσουμε το πρώτο κελί αυτής της στήλης, απλά αφαιρούμε τις δύο πρώτες ημερομηνίες από τη στήλη “Dates”. Προφανώς η διαφορά αυτή είναι το πλήθος των ημερών που έχουν μεσολαβήσει ανάμεσα στους δύο πρώτους σεισμούς. Συμπληρώνουμε την υπόλοιπη στήλη χρησιμοποιώντας την αυτοματοποιημένη συμπλήρωση του Excel.

Διόρθωση δεδομένων με τιμή 0:

Θα μπορούσαμε να λάβουμε ως το σετ δεδομένων μας την στήλη “DataDays”. Ωστόσο η στήλη αυτή περιέχει τιμές οι οποίες είναι μηδενικές. Ο χρόνος είναι συνεχές μέγεθος, οπότε είναι δυνατόν να έχουμε οποιαδήποτε πραγματική τιμή στα δεδομένα μας, αφού προφανώς δύο σεισμοί δεν θα πραγματοποιηθούν με ακριβώς “n” ημέρες διαφορά με “n” φυσικό αριθμό. Η στρογγυλοποίηση του πλήθους των ημερών θα έχει ελάχιστη επίδραση στα αποτελέσματα των ελέγχων, εφόσον η μέση τιμή του δείγματός μας είναι λίγο μεγαλύτερη του 85, και στη χειρότερη περίπτωση που όλες οι τιμές στρογγυλοποιήθηκαν προς μία μόνο κατεύθυνση (είτε στο μικρότερο, είτε στο μεγαλύτερο φυσικό αριθμό), το σφάλμα μας στη μέση τιμή θα είναι μόλις 0.5 ημέρες ή 0.56% της μέσης τιμής.

Τα σφάλματα της στρογγυλοποίησης στη συγκεκριμένη περίπτωση, θα ακολουθούν την κανονική κατανομή με μέση τιμή το 0. Οπότε δεν χρειάζεται να ανησυχούμε για τις στρογγυλοποιήσεις.

Εξαίρεση αποτελούν τα δεδομένα τα οποία έχουν τιμή 0. Η τιμή αυτή θέλει προσοχή, γιατί στην ουσία αντιστοιχεί σε ακριβώς μηδενικό χρόνο ανάμεσα σε δύο γεγονότα, πράγμα που σημαίνει ότι δύο σεισμοί πραγματοποιήθηκαν ταυτόχρονα. Αυτό είναι πιθανό να επηρεάσει κάποιους ελέγχους όπως αυτός του “Wong and Wong’s” , ο οποίος για τον υπολογισμό της ελεγχοσυνάρτησης (3.15) χρησιμοποιεί την ελάχιστη τιμή στον παρανομαστή, αν η ελάχιστη τιμή είναι 0, δηλαδή αν δύο γεγονότα συνέβησαν ταυτόχρονα (ενδεχόμενο απίθανο, αφού ο

χρόνος είναι συνεχές μέγεθος), το κλάσμα της ελεγχουσυνάρτησης δεν θα ορίζεται, οπότε ο έλεγχος δεν θα μπορεί να εφαρμοστεί.

Η λύση σε αυτό το πρόβλημα, είναι να αντικαταστήσουμε τα μηδενικά δεδομένα με την τιμή 0.25. Το 0.25 αντιπροσωπεύει τη μέση τιμή των χρόνων που στρογγυλοποιήθηκαν στην τιμή 0 (ημέρες στο διάστημα (0,5)).

Για τη διόρθωση αυτή, δημιουργούμε στο αρχείο Excel ακόμη μία στήλη με όνομα "DataDaysCor". Στο πρώτο κελί της στήλης πληκτρολογούμε την εντολή :

=if(ΑντίστοιχοΚελίDataDays>0, ΑντίστοιχοΚελίDataDays, 0.25)

Γεμίζουμε όλη τη στήλη αυτοματοποιημένα, και η στήλη που δημιουργήθηκε εκτός της τελευταίας τιμής (προφανώς το πλήθος των διαφορών είναι 1 λιγότερο από το πλήθος των σεισμών, οπότε διαγράφουμε την τελευταία τιμή) αποτελεί και το σετ δεδομένων προς εξέταση.

Οφείλουμε να παρατηρήσουμε ένα ακόμη χαρακτηριστικό των δεδομένων που είναι δυνατόν να μας δημιουργήσει πρόβλημα σε κάποιους ελέγχους. Όσοι έλεγχοι χρησιμοποιούν το λογάριθμο των ποσοτήτων D_i (3.1) θα δίνουν άπειρη τιμή για την ελεγχουσυνάρτησή τους, αφού με τη στρογγυλοποίηση που έχουμε κάνει στις ημέρες, δύο ζευγάρια σεισμών είναι δυνατόν να έχουν ακριβώς την ίδια χρονική διαφορά σε ημέρες. Άρα στο διατεταγμένο δείγμα που χρησιμοποιείται για τον υπολογισμό των D_i ποσοτήτων δύο ή περισσότερες συνεχόμενες τιμές μπορεί να είναι ίσες, οπότε κάποια D_i θα είναι 0 και ο λογάριθμός τους θα είναι άπειρος. Τέτοιοι έλεγχοι είναι ο έλεγχος του Epstein και εν μέρει ο έλεγχος του Cox and Oakes (ο πρώτος χρησιμοποιεί το λογάριθμο των D_i ενώ ο δεύτερος τον λογάριθμο των X_i παρατηρήσεων του δείγματος). Όπως και με τις μηδενικές τιμές, ο έλεγχος του Epstein φαίνεται να απορρίπτει το γεγονός πως δύο ζευγάρια σεισμών έχουν ακριβώς την ίδια χρονική διαφορά. Ο χρόνος είναι συνεχές μέγεθος, οπότε είναι λογικό να απορρίπτεται το γεγονός αυτό (ακριβώς ίδια χρονική διαφορά).

Για τον λόγο αυτό, μόνο για τον έλεγχο του Epstein και του Cox and Oakes δημιουργούμε μία έξτρα στήλη στο Excel με όνομα DataDaysCorX ως εξής :

Αρχικά σε μία στήλη (Έστω στήλη K) τοποθετούμε τα δεδομένα της στήλης

DataDaysCor σε αύξουσα σειρά χρησιμοποιώντας την εντολή Sort > from smallest to largest.

Έπειτα στη διπλανή στήλη (DataDaysCorX) εισάγουμε τον κώδικα :

=IF(Κελί Στήλης K(i)=Κελί Στήλης K(i-1),Κελί στήλης K(i)+RAND()*0.6, Κελί στήλης K(i))

Με αυτόν τον τρόπο, καταφέρνουμε να περάσουμε ελάχιστο «θόρυβο» στα δεδομένα μας , με σκοπό να μην έχουμε ίδιες τιμές. (αν έχουμε ίδιες τιμές , τότε προσθέτουμε ένα τυχαίο αριθμό από το διάστημα (0,0.6). Έτσι πετυχαίνουμε να μην έχουμε μηδενικές διαφορές και ο έλεγχος του Epstein μπορεί να λειτουργήσει επιτυχώς.

Χρησιμοποιώντας την εντολή =count(...) του Excel με input τη στήλη που δημιουργήσαμε, λαμβάνουμε ότι το μέγεθος του δείγματός μας είναι 222. Η εικόνα που παίρνουμε από την τακτοποίηση και διόρθωση των δεδομένων είναι αυτή (Παρουσιάζονται μόνο οι πρώτες στήλες):

Date	Magnitude	DataDays	DataDaysCor	Count
1/24/1965	8.2	11	11	222
2/4/1965	8.7	54	54	
3/30/1965	7.6	51	51	
5/20/1965	7.7	43	43	
7/2/1965	7.8	40	40	
8/11/1965	7.6	213	213	
3/12/1966	7.5	180	180	
9/8/1966	7.7	39	39	
10/17/1966	8.1	72	72	
12/28/1966	7.7	3	3	
12/31/1966	7.8	457	457	
4/1/1968	7.5	45	45	
5/16/1968	8.2	0	0.25	
5/16/1968	7.9	12	12	
5/28/1968	7.5	65	65	
8/1/1968	7.6	9	9	
8/10/1968	7.6	173	173	
1/30/1969	7.6	29	29	
2/28/1969	7.8	164	164	
8/11/1969	7.5	102	102	
11/21/1969	7.6	191	191	
5/31/1970	7.9	61	61	
7/31/1970	8	163	163	
1/10/1971	7.7	180	180	

ΕΙΚΟΝΑ 9 : Απόσπασμα πραγματικών δεδομένων

5.2.2 Εισαγωγή των δεδομένων στην R

Για να εισάγουμε τα δεδομένα στην R από αρχείο excel αρχικά πρέπει να εγκαταστήσουμε ένα πακέτο και να χρησιμοποιήσουμε μία βιβλιοθήκη:

```
> install.packages("readxl")  
> library(readxl)
```

Ύστερα χρησιμοποιούμε τη βιβλιοθήκη για να εισάγουμε το αρχείο στην R με την εντολή :

```
> Data=read_excel("C:/Users/Paris/Downloads/Earthquake7.5Times.xlsx")
```

Η εντολή έχει ως input το path του αρχείου στον υπολογιστή, προσέχουμε ότι στους περισσότερους υπολογιστές, στο path χρησιμοποιείται το σύμβολο “\”. Η R δεν το αναγνωρίζει και απλά πρέπει να το αντικαταστήσουμε με το σύμβολο “/”.

Δημιουργούμε το διάνυσμα των δεδομένων το οποίο και θα χρησιμοποιήσουμε για την ανάλυσή μας, εισάγοντας στο διάνυσμα μας τη στήλη που μας ενδιαφέρει:

```
> DataDays=Data$DataDaysCor
```

```
> DataDaysX=Data$DataDaysCorX ( για Epstein και Cox and Oakes )
```

Με τη χρήση του συμβόλου “\$” απλώς εκχωρήσαμε στο νέο διάνυσμα “DataDays” τη στήλη του αρχείου Excel “DataDaysCor”.

5.3 Στατιστική ανάλυση των δεδομένων για την εκθετικότητα

5.3.1 Γραφικός έλεγχος

Αφού τελείωσε η επεξεργασία και η εκχώρηση των δεδομένων στην R, θα υλοποιήσουμε τους ελέγχους που χρησιμοποιήθηκαν στο Κεφάλαιο 4, για να ελέγξουμε την υπόθεση ότι τα δεδομένα μας ακολουθούν την εκθετική κατανομή.

Αρχικά κάνουμε μία εκτίμηση της παραμέτρου λ μέσω της μεθόδου μεγίστης πιθανοφάνειας, χρησιμοποιώντας την εντολή :

```
> LamdaEst=1/mean(DataDays)
```

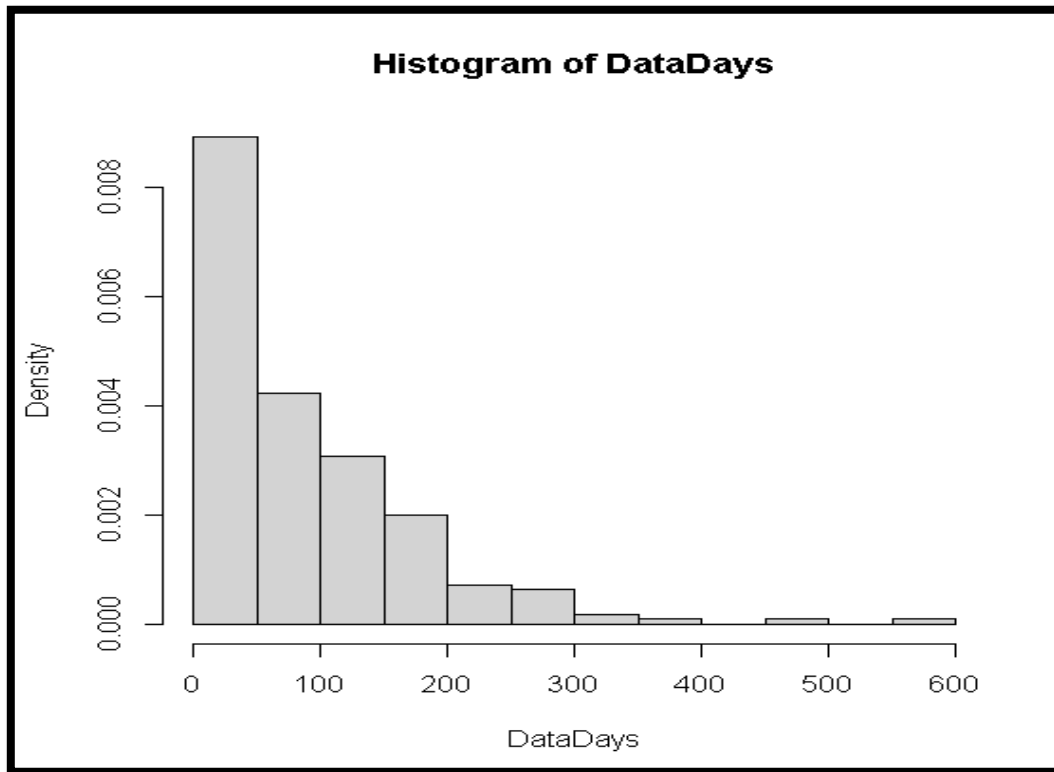
```
> LamdaEst
```

```
[1] 0.01170516
```

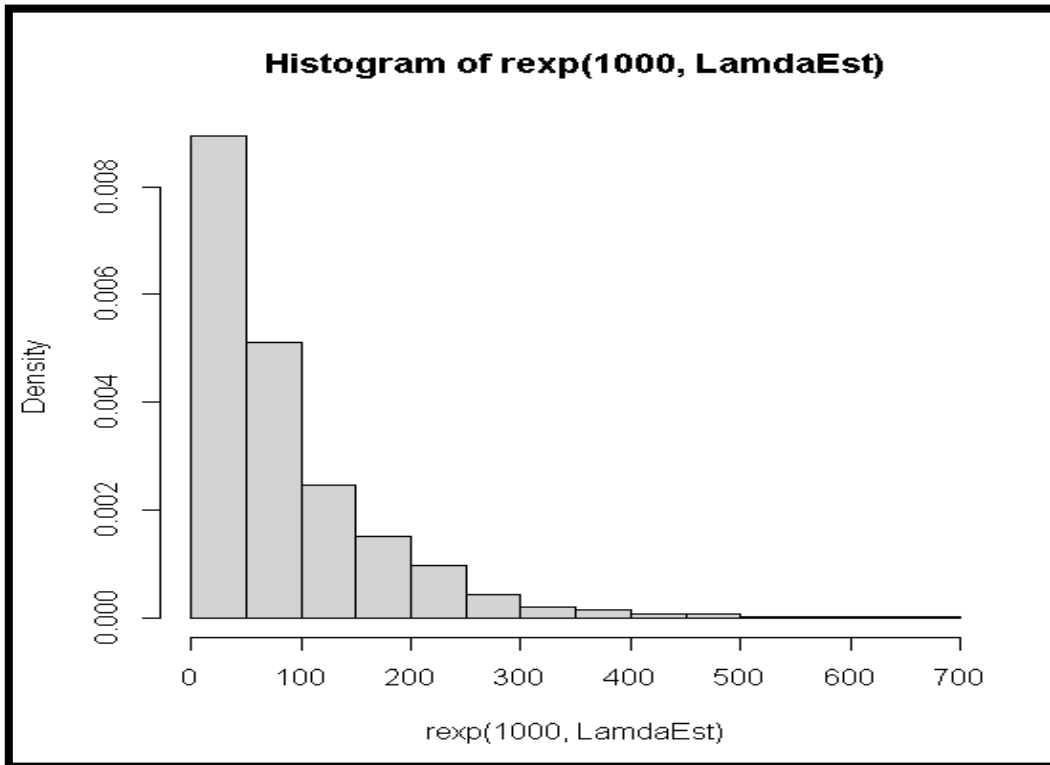
Αρχικά, κατασκευάζουμε τα ιστογράμματα των σχετικών συχνοτήτων των δεδομένων προς εξέταση και των 1000 δεδομένων που προέρχονται από την εκθετική κατανομή με παράμετρο LamdaEst για να συγκρίνουμε τα δεδομένα προς εξέταση με την εκθετική κατανομή.

```
> h1=hist(DataDays,probability=T)
```

```
> h2=hist(rexp(1000,LamdaEst),probability=T)
```



ΕΙΚΟΝΑ 10 : Ιστόγραμμα πραγματικών δεδομένων

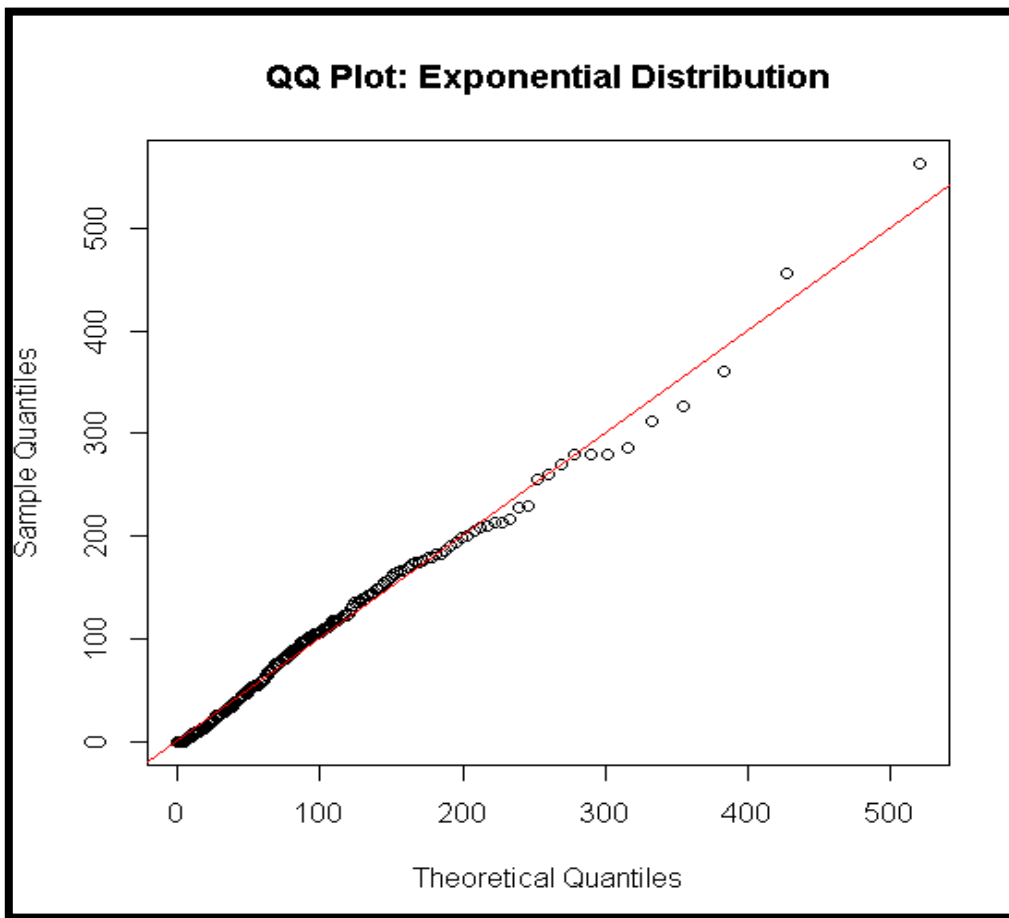


ΕΙΚΟΝΑ 11 : Ιστογράμμα προσέγγισης εκθετικής κατανομής

Όπως μπορούμε να διακρίνουμε εύκολα, η ομοιότητα ανάμεσα στα δύο ιστογράμματα είναι αρκετά μεγάλη, γεγονός που μας δείχνει ότι το δείγμα μας έχει πολλές πιθανότητες να ακολουθεί την εκθετική κατανομή με παράμετρο LamdaEst.

Το επόμενο διάγραμμα που θα χρησιμοποιήσουμε είναι το διάγραμμα Q-Q Plot.

```
> theoretical_quantiles = qexp(ppoints(length(DataDays)), rate = LamdaEst)
> qqplot(theoretical_quantiles, DataDays,xlab = "Theoretical Quantiles",ylab =
"Sample Quantiles",main = "QQ Plot: Exponential Distribution")
> abline(a = 0, b = 1, col = "red")
```



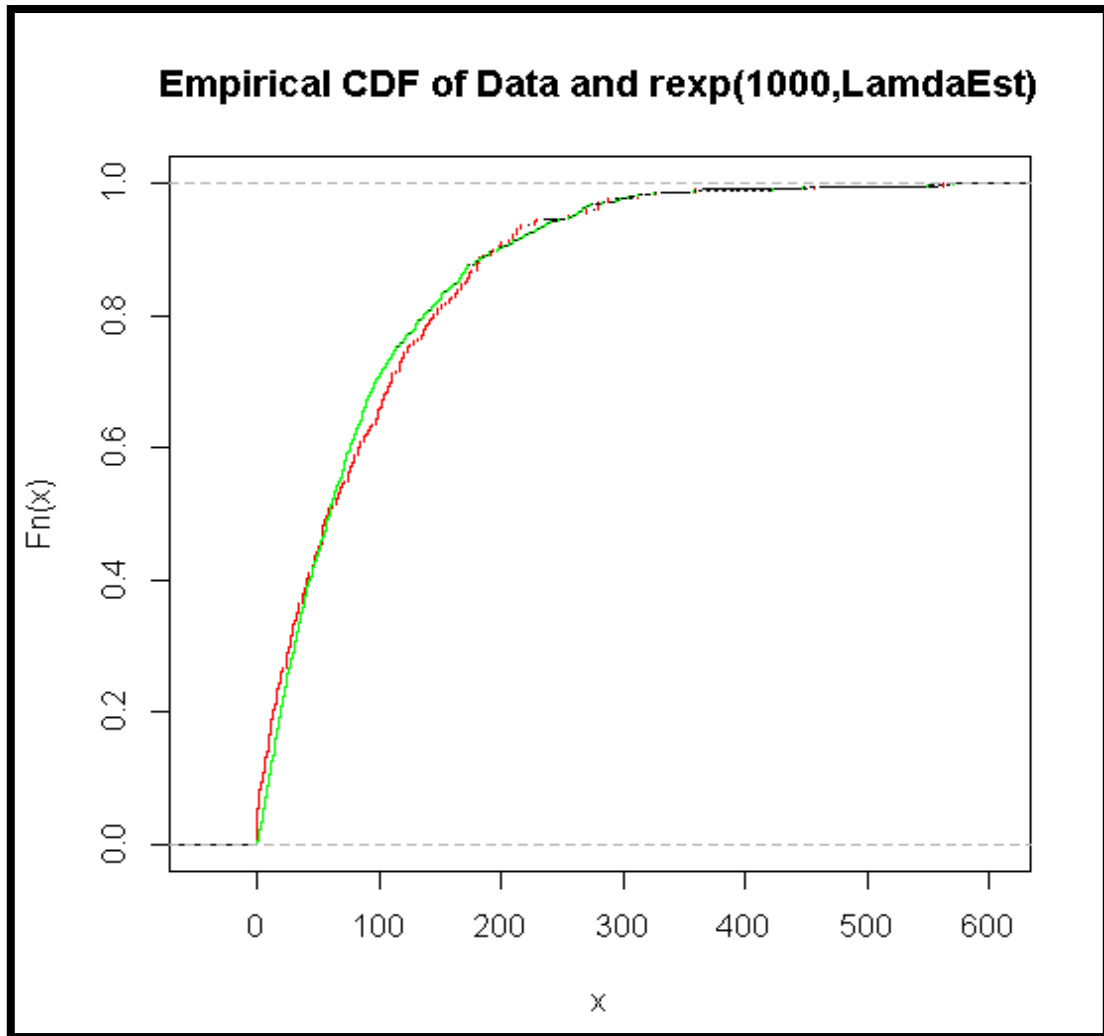
ΕΙΚΟΝΑ 12 : Διάγραμμα Q-Q Plot πραγματικών δεδομένων

Και εδώ, το αποτέλεσμα φαίνεται ίδιο με αυτό που μας έδωσε η σύγκριση των ιστογραμμάτων. Όπως παρατηρούμε, τα σημεία βρίσκονται πολύ κοντά στην ευθεία $y=x$. Άρα επιβεβαιώνεται μέσω του Q-Q Plot ότι το δείγμα μας ακολουθεί την εκθετική κατανομή με $\lambda = \text{LamdaEst}$.

Για την ολοκλήρωση του γραφικού ελέγχου, παρουσιάζουμε τον κώδικα και την γραφική παράσταση των διαφορών που δίνει ο έλεγχος του Kolmogorov-Smirnov. Ο μετασχηματισμός του Lilliefors μας παρέχει αντίστοιχο αποτέλεσμα, οπότε παρουσιάζεται μόνο η γραφική παράσταση του Kolmogorov-Smirnov ελέγχου.

```
> plot.ecdf(DataDays, verticals=TRUE, col.p="white", col.v="red",
main="Empirical CDF of Data and rexp(1000,LamdaEst)")
```

```
> plot(ecdf(rexp(1000,LamdaEst)), add=TRUE, verticals=TRUE, col.p="white",  
col.v="green", lty="dotted")
```



ΕΙΚΟΝΑ 13 : Γραφική παράσταση εμπειρικής συνάρτησης
κατανομής πραγματικών δεδομένων (Kolmogorov)

5.3.2 Ανάλυση μέσω στατιστικών ελέγχων

Όπως και στο Κεφάλαιο 4, θα χρησιμοποιήσουμε την αντίστοιχη βιβλιοθήκη για την υλοποίηση των ελέγχων και στη συνέχεια υλοποιούμε τους ελέγχους:

```
> library(exptest)
> library(KScorrect)

> ks.test(DataDays,"pexp",LamdaEst)
```

```
One-sample Kolmogorov-Smirnov test
```

```
data: DataDays
D = 0.061548, p-value = 0.3696
alternative hypothesis: two-sided
```

Εδώ όπως βλέπουμε από τον έλεγχο Kolmogorov-Smirnov διαπιστώνουμε ότι η $p - value$ είναι πολύ μεγαλύτερη από το 0.05. Όπως βλέπουμε από τον πίνακα των Kolmogorov-Smirnov που δίνεται στο Παράρτημα στον Πίνακα 2, για $n = 222$ και $\alpha = 5\%$, η κρίσιμη τιμή είναι ίση με 0.0911. Εφόσον η τιμή της ελεγχουσυνάρτησης $D = 0.0615$ δεν ξεπερνά το 0.0911, η μηδενική υπόθεση της εκθετικότητας δεν μπορεί να απορριφθεί.

```
> LcKS(DataDays,"pexp")
```

```
§D.obs
[1] 0.06154751
```

```
§p.value
[1] 0.1538
```

Από τον έλεγχο του Lilliefors διαπιστώνουμε ότι η $p - value$ είναι πολύ μεγαλύτερη από το 0.05. Όπως βλέπουμε από τον πίνακα του Lilliefors που δίνεται στο Παράρτημα στον Πίνακα 3, για $n = 222$ και $\alpha = 5\%$, η κρίσιμη τιμή είναι ίση με 0.0721. Εφόσον η τιμή της ελεγχουσυνάρτησης $D = 0.0615$ δεν ξεπερνά το 0.0721, η μηδενική υπόθεση της εκθετικότητας δεν μπορεί να απορριφθεί.

```
> gnedenko.exp.test(DataDays)
```

```
Gnedenko's F-test of exponentiality
```

```
data: DataDays  
Q = 0.9208, p-value = 0.5393
```

```
> harris.exp.test(DataDays)
```

```
Harris modification of Gnedenko's F-test
```

```
data: DataDays  
Q = 0.76003, p-value = 0.04136
```

```
> gini.exp.test(DataDays)
```

```
Test for exponentiality based on the Gini statistic
```

```
data: DataDays  
Gn = 0.51701, p-value = 0.3809
```

```
> lorenz.exp.test(DataDays)
```

```
Lorenz test for exponentiality
```

```
data: DataDays  
L = 0.13408, p-value = 0.7732
```

```
> pietra.exp.test(DataDays)
```

```
Test for exponentiality based on the Pietra statistic
```

```
data: DataDays  
Pn = 0.38149, p-value = 0.37
```



```
> deshpande.exp.test(DataDays)
```

```
      Deshpande test for exponentiality
```

```
data:  DataDays  
J = 0.67472, p-value = 0.04914
```

```
> shapiro.exp.test(DataDays)
```

```
      Shapiro-Wilk test for exponentiality
```

```
data:  DataDays  
W = 0.0044431, p-value = 0.403
```

```
> hollander.exp.test(DataDays)
```

```
      Hollander-Proshan test for exponentiality
```

```
data:  DataDays  
T = 0.26125, p-value = 0.1191
```

```
> epstein.exp.test(DataDaysX)
```

```
      Epstein test for exponentiality
```

```
data:  DataDaysX  
EPS = 231.24, p-value = 0.3046
```

```
> co.exp.test(DataDaysX)
```

```
Test for exponentiality based on the statistic of Cox and Oakes
```

```
data: DataDaysX  
COx = -45.727, p-value = 0.01672
```

```
> ww.exp.test(DataDays)
```

```
Wong and Wong's test for exponentiality
```

```
data: DataDays  
Q = 2252, p-value = 0.439
```

```
> we.exp.test(DataDays)
```

```
WE test for exponentiality
```

```
data: DataDays  
WE = 0.0045606, p-value = 0.783
```

5.4 Συμπεράσματα

Οι γραφικοί έλεγχοι, μας δείχνουν ότι δεν μπορούμε να απορρίψουμε την υπόθεση ότι το σετ δεδομένων ακολουθεί την εκθετική κατανομή με παράμετρο λ Est.

- Το ιστόγραμμα των δεδομένων προς εξέταση, έχει μεγάλη ομοιότητα με το ιστόγραμμα της εκθετικής κατανομής με παράμετρο λ Est.
- Το διάγραμμα Q-Q Plot μας δείχνει ότι όλα τα σημεία είναι σχεδόν πάνω στην ευθεία $y=x$, γεγονός που δεν μας επιτρέπει να απορρίψουμε την υπόθεση της εκθετικότητας
- Το διάγραμμα διαφορών για τον έλεγχο του Kolmogorov μας δείχνει ότι οι διαφορές των τιμών των δεδομένων με τις τιμές της εκθετικής κατανομής με παράμετρο λ Est βρίσκονται μέσα στα επιτρεπτά όρια.

Οι στατιστικοί έλεγχοι που εξετάστηκαν μας οδηγούν γενικά στα ίδια συμπεράσματα με κάποιες εξαιρέσεις:

- 11 από τους 14 ελέγχους που υλοποιήθηκαν μας δίνουν p-values μεγαλύτερες από 0.1.
- Από τους 11 αυτούς ελέγχους η ελάχιστη p-value εντοπίζεται στον έλεγχο “Hollander-Proshan” ίση με 0.1191.
- Από τους 11 αυτούς ελέγχους, οι 8 μας δίνουν p-value μεγαλύτερη από 0.35.
- Οι έλεγχοι “Harris Modification of Gnedenko’s F-Test” και “Deshpande’s Test” μας δίνουν p-values 0.041 και περίπου 0.05 αντίστοιχα, γεγονός που δεν μας επιτρέπει να απορρίψουμε την υπόθεση της εκθετικότητας με μεγάλη σιγουριά.
- Μόνο ένας έλεγχος, ο “Cox and Oakes” μας δίνει πολύ χαμηλή p-value και απορρίπτει ισχυρά την μηδενική υπόθεση της εκθετικότητας. Ο έλεγχος Cox and Oakes είναι χρήσιμος έναντι εναλλακτικών υποθέσεων οι οποίες σχετίζονται με μονότονες συναρτήσεις κινδύνου. Πιθανώς αυτός ο έλεγχος να εντοπίζει ότι κάποια εναλλακτική της εκθετικής κατανομής, π.χ. κάποια Weibull, είναι καταλληλότερη για αυτά τα δεδομένα.
- Το τελικό συμπέρασμα είναι, ότι δεν μπορούμε να απορρίψουμε την υπόθεση ότι το χρονικό διάστημα που μεσολαβεί ανάμεσα σε δύο σεισμούς των 7.5 ρίχτερ και άνω ακολουθεί την εκθετική κατανομή με $\lambda = \text{LamdaEst} = 0.0117$.

ΠΑΡΑΡΤΗΜΑ : ΠΙΝΑΚΕΣ

ΠΙΝΑΚΑΣ 1 : Αντίστροφα ποσοστιαία σημεία της χ^2 κατανομής

<https://statisticsbyjim.com/hypothesis-testing/chi-square-table/>

Chi-Square Right-Tail Probability ($\geq \chi^2$)										
DF	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

ΠΙΝΑΚΑΣ 2: Κρίσιμες τιμές της ελεγχουσυνάρτησης Kolmogorov

<https://real-statistics.com/statistics-tables/kolmogorov-smirnov-table/>

$n \setminus \alpha$	0.001	0.01	0.02	0.05	0.1	0.15	0.2
1		0.99500	0.99000	0.97500	0.95000	0.92500	0.90000
2	0.97764	0.92930	0.90000	0.84189	0.77639	0.72614	0.68377
3	0.92063	0.82900	0.78456	0.70760	0.63604	0.59582	0.56481
4	0.87137	0.76855	0.72718	0.62394	0.56522	0.52476	0.49265
5	0.8137	0.6855	0.62718	0.56327	0.50945	0.47439	0.44697
6	0.72479	0.61660	0.57741	0.51926	0.46799	0.43526	0.41035
7	0.67930	0.57580	0.53844	0.48343	0.43607	0.40497	0.38145
8	0.64098	0.54180	0.50654	0.45427	0.40962	0.38062	0.35828
9	0.60846	0.51330	0.47960	0.43001	0.38746	0.36006	0.33907
10	0.58042	0.48895	0.45662	0.40925	0.36866	0.34250	0.32257
11	0.55588	0.46770	0.43670	0.39122	0.35242	0.32734	0.30826
12	0.53422	0.44905	0.41918	0.37543	0.33815	0.31408	0.29573
13	0.51490	0.43246	0.40362	0.36143	0.32548	0.30233	0.28466
14	0.49753	0.41760	0.38970	0.34890	0.31417	0.29181	0.27477
15	0.48182	0.40420	0.37713	0.33760	0.30397	0.28233	0.26585
16	0.46750	0.39200	0.36571	0.32733	0.29471	0.27372	0.25774
17	0.45440	0.38085	0.35528	0.31796	0.28627	0.26587	0.25035
18	0.44234	0.37063	0.34569	0.30936	0.27851	0.25867	0.24356
19	0.43119	0.36116	0.33685	0.30142	0.27135	0.25202	0.23731
20	0.42085	0.35240	0.32866	0.29407	0.26473	0.24587	0.23152
25	0.37843	0.31656	0.30349	0.26404	0.23767	0.22074	0.20786
30	0.34672	0.28988	0.27704	0.24170	0.21756	0.20207	0.19029
35	0.32187	0.26898	0.25649	0.22424	0.20184	0.18748	0.17655
40	0.30169	0.25188	0.23993	0.21017	0.18939	0.17610	0.16601
45	0.28482	0.23780	0.22621	0.19842	0.17881	0.16626	0.15673
50	0.27051	0.22585	0.21460	0.18845	0.16982	0.15790	0.14886
OVER 50	1.94947	1.62762	1.51743	1.35810	1.22385	1.13795	1.07275
	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}

ΠΙΝΑΚΑΣ 3: p-ποσοστιαία σημεία της ελεγχουσυνάρτησης Lilliefors για τον έλεγχο καλής προσαρμογής της εκθετικής κατανομής

(Ευδοκία Ξεκαλάκη (2001))

	$p = .05$.10	.20	.30	.50	.70	.80	.90	.95	.99	.999
$n = 2$.3127	.3200	.3337	.3617	.4337	.5034	.5507	.5934	.6133	.6284	.6317
3	.2299	.2544	.2899	.3166	.3645	.4122	.4508	.5111	.5508	.6003	.6296
4	.2072	.2281	.2545	.2766	.3163	.3685	.4007	.4442	.4844	.5574	.6215
5	.1884	.2052	.2290	.2483	.2877	.3317	.3603	.4045	.4420	.5127	.5814
6	.1726	.1882	.2102	.2290	.2645	.3045	.3320	.3732	.4085	.4748	.5497
7	.1604	.1750	.1961	.2136	.2458	.2838	.3098	.3481	.3811	.4459	.5181
8	.1506	.1646	.1845	.2006	.2309	.2671	.2914	.3274	.3590	.4208	.4913
9	.1426	.1561	.1746	.1897	.2186	.2529	.2758	.3101	.3404	.3995	.4679
10	.1359	.1486	.1661	.1805	.2082	.2407	.2626	.2955	.3244	.3813	.4473
12	.1249	.1364	.1524	.1657	.1912	.2209	.2411	.2714	.2981	.3511	.4132
14	.1162	.1268	.1418	.1542	.1778	.2054	.2242	.2525	.2774	.3272	.3858
16	.1091	.1191	.1332	.1448	.1669	.1929	.2105	.2371	.2606	.3076	.3632
18	.1032	.1127	.1260	.1369	.1578	.1824	.1990	.2242	.2465	.2911	.3441
20	.0982	.1073	.1199	.1303	.1501	.1735	.1893	.2132	.2345	.2771	.3277
22	.0939	.1025	.1146	.1245	.1434	.1657	.1809	.2038	.2241	.2649	.3135
24	.0901	.0984	.1099	.1195	.1376	.1590	.1735	.1954	.2150	.2542	.3010
26	.0868	.0947	.1058	.1150	.1324	.1530	.1670	.1881	.2069	.2447	.2899
28	.0838	.0914	.1021	.1110	.1278	.1477	.1611	.1815	.1997	.2362	.2799
30	.0811	.0885	.0988	.1074	.1236	.1428	.1559	.1756	.1932	.2286	.2709
35	.0754	.0822	.0918	.0997	.1148	.1326	.1447	.1630	.1793	.2123	.2517
40	.0707	.0771	.0861	.0935	.1077	.1243	.1356	.1528	.1681	.1990	.2361
45	.0668	.0729	.0814	.0884	.1017	.1174	.1281	.1443	.1588	.1880	.2231
50	.0636	.0693	.0774	.0840	.0966	.1116	.1217	.1371	.1509	.1787	.2121
60	.0582	.0635	.0708	.0769	.0885	.1021	.1114	.1255	.1381	.1635	.1943
70	.0541	.0589	.0658	.0714	.0821	.0946	.1033	.1164	.1281	.1517	^a
80	.0507	.0553	.0616	.0669	.0769	.0887	.0968	.1090	.1200	.1421	^a
90	.0479	.0522	.0582	.0632	.0726	.0838	.0914	.1029	.1132	.1341	^a
$n = 100$.0455	.0496	.0553	.0600	.0690	.0796	.0868	.0977	.1075	.1274	^a
Approximation	.4550	.4959	.5530	.6000	.6898	.7957	.8678	.9773	1.0753	1.2743	^a
for $n > 100$	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	

ΠΙΝΑΚΑΣ 4: Κρίσιμες τιμές για τον έλεγχο Deshpande για σταθερό $\alpha=0.05$

Deshpande, J.V. (1983). A class of tests for exponentiality against increasing failure rate average alternatives . *Biometrika* 70, 514-518

Table 1. *Monte Carlo estimates of critical values, exact levels of significance and power of the $J_{\frac{1}{2}}$ and $J_{0.9}$ tests; $\alpha = 0.05$.*

<i>n</i>	<i>b = 0.5</i>				<i>b = 0.9</i>			
	Exact α <i>p(b, α, <i>n</i>, 0)</i>	<i>c$_{\alpha}$</i>	<i>p(b, α, <i>n</i>, 1)</i>	<i>p(b, α, <i>n</i>, 2)</i>	Exact α <i>p(b, α, <i>n</i>, 0)</i>	<i>c$_{\alpha}$</i>	<i>p(b, α, <i>n</i>, 1)</i>	<i>p(b, α, <i>n</i>, 2)</i>
5	0.130	16	0.578	0.193	0.073	12	0.233	0.096
	0.043	17	0.340	0.074	0.024	13	0.092	0.032
7	0.057	32	0.506	0.102	0.075	24	0.281	0.102
	0.035	33	0.405	0.068	0.021	25	0.117	0.030
9	0.052	51	0.572	0.100	0.056	40	0.278	0.082
	0.033	52	0.481	0.067	0.021	41	0.139	0.034
11	0.070	73	0.703	0.142	0.106	59	0.452	0.152
	0.048	74	0.631	0.099	0.043	60	0.279	0.066
13	0.059	100	0.724	0.130	0.090	83	0.436	0.124
	0.040	101	0.661	0.094	0.033	84	0.281	0.056
15	0.050	132	0.742	0.119	0.057	111	0.407	0.094
	0.037	133	0.690	0.091	0.025	112	0.266	0.046

ΠΙΝΑΚΑΣ 5: Κρίσιμες τιμές για τον έλεγχο Gini

Gail, M.H and Gastwirth, J.L. (1978). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. *JRSS B* 40, 350-357

<i>n</i>	<i>Probability level</i>			<i>n</i>	<i>Probability level</i>		
	0.950	0.975	0.990		0.950	0.975	0.990
3	0.84189	0.88818	0.92932	12	0.64337	0.66992	0.70020
4	0.77686	0.82288	0.86951	13	0.63725	0.66275	0.69183
5	0.73834	0.77997	0.82501	14	0.63185	0.65641	0.68448
6	0.71307	0.75079	0.79260	15	0.62704	0.65076	0.67792
7	0.69439	0.72931	0.76831	16	0.62273	0.64567	0.67197
8	0.67988	0.71252	0.74921	17	0.61882	0.64107	0.66659
9	0.66821	0.69896	0.73370	18	0.61527	0.63688	0.66168
10	0.65855	0.68768	0.72070	19	0.61201	0.63304	0.65723
11	0.65039	0.67816	0.70972	20	0.60902	0.62952	0.65308

ΠΙΝΑΚΑΣ 6: Κρίσιμες τιμές για τον έλεγχο New Better than Used

Hollander, M. and Proschan, F.(1972). Testing whether new is better than used. *Annals of Math. Stat.* 43, 1136-1146

TABLE 4.1
Critical values of the T_n statistic

$n \backslash \alpha$	Lower Tail					Upper Tail				
	.01	.025	.05	.075	.10	.01	.025	.05	.075	.10
4				0(.067)	1(.103)					
5	0(.018)	1(.028)	2(.040)	3(.072)	4					10(.177)
6	2	5(.028)	7	8(.061)	9(.086)			20(.053)		19(.167)
7	7	11	14(.045)	16(.072)	18(.105)	35		34(.047)		33(.103)
8	15	21	25	28	30	55	54	53	52(.083)	52(.083)
9	27	34	40	44	47	81	80(.019)	78	77	76
10	42	52	60	65	69	114	112	110	109(.067)	107(.107)
11	63	76	86	92	97	155	152(.028)	150(.046)	148(.070)	146
12	89	105	117	125	131	205	201	198(.047)	195	193
13	122	141	157	167	174	264	260	255	252(.072)	249
14	162	185	204	215	223	334	328	322	318	315
15	209	236	259	272	282	415	408	401	396	392
16	265	298	323	338	350	508	499	491	484	480
17	330	368	397	415	429	613	603	593	586	580
18	405	446	480	502	518	732	720	709	700	693
19	490	538	577	601	619	866	852	838	828	820
20	594	642	683	713	732	1015	998	982	970	961
25	1290	1351	1427	1472	1507	2018	1986	1956	1934	1918
30	2320	2463	2574	2651	2706	3521	3461	3411	3375	3349
35	3890	4064	4215	4316	4394	5025	5546	5464	5403	5359
40	5947	6214	6434	6593	6700	6418	8301	8189	8113	8049
45	8665	9040	9341	9543	9686	12012	11864	11709	11600	11502
50	12170	12661	13020	13255	13439	16519	16310	16085	15937	15823

ΠΙΝΑΚΑΣ 7: Κρίσιμες τιμές για την κατανομή F για $\alpha = 0.05$

<https://statisticsbyjim.com/hypothesis-testing/f-table/>

F-table of Critical Values of $\alpha = 0.05$ for F(df1, df2)																			
	DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

ΠΙΝΑΚΑΣ 8: Κρίσιμες τιμές για την κατανομή F για $\alpha = 0.10$

<https://statisticsbyjim.com/hypothesis-testing/f-table/>

		F-table of Critical Values of $\alpha = 0.10$ for F(df1, df2)																		
		DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1		39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2		8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3		5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4		4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5		4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
6		3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7		3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8		3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9		3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10		3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11		3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12		3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13		3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14		3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15		3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16		3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17		3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18		3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19		2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20		2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21		2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22		2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23		2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24		2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25		2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26		2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27		2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28		2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29		2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30		2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40		2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60		2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120		2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞		2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

ΠΙΝΑΚΑΣ 9: Κρίσιμες τιμές για το $W(E)$ στατιστικό του ελέγχου για την εκθετικότητα (Shapiro-Wilk)

<https://mvpprograms.com>

	Alpha (single-tailed values)										
n	0.005	0.010	0.025	0.050	0.100	0.200	0.300	0.400	0.500	0.575	0.690
3	0.2518	0.2538	0.2596	0.2697	0.2915	0.3714	0.4709	0.5926	0.5981	0.5997	0.5999
4	0.1241	0.1302	0.1424	0.1604	0.1881	0.2768	0.3714	0.4581	0.5236	0.5880	0.5937
5	0.0645	0.0695	0.1046	0.1167	0.1442	0.2075	0.2647	0.3062	0.3590	0.4000	0.4192
6	0.0610	0.0665	0.0802	0.0956	0.1179	0.2276	0.2922	0.3099	0.3642	0.3775	0.3901
7	0.0514	0.0591	0.0700	0.0810	0.0966	0.1674	0.2474	0.2462	0.2852	0.2906	0.2926
8	0.0454	0.0512	0.0614	0.0710	0.0852	0.1625	0.2034	0.2497	0.2403	0.2446	0.2426
9	0.0404	0.0442	0.0537	0.0633	0.0781	0.1416	0.2853	0.3008	0.3454	0.3418	0.3433
10	0.0369	0.0404	0.0487	0.0568	0.0678	0.1226	0.2178	0.2628	0.2879	0.3391	0.3701
11	0.0338	0.0380	0.0447	0.0528	0.0616	0.1112	0.1934	0.2268	0.2818	0.3038	0.3314
12	0.0311	0.0358	0.0410	0.0484	0.0567	0.1008	0.1723	0.2018	0.2364	0.2716	0.2978
13	0.0287	0.0337	0.0382	0.0460	0.0528	0.0925	0.1563	0.1829	0.2113	0.2422	0.2642
14	0.0269	0.0317	0.0362	0.0428	0.0486	0.0847	0.1417	0.1647	0.1902	0.2131	0.2318
15	0.0247	0.0296	0.0344	0.0398	0.0466	0.0776	0.1285	0.1485	0.1699	0.1920	0.2123
16	0.0233	0.0280	0.0326	0.0374	0.0436	0.0728	0.1187	0.1355	0.1542	0.1770	0.1931
17	0.0222	0.0264	0.0310	0.0352	0.0412	0.0664	0.1099	0.1257	0.1423	0.1614	0.1764
18	0.0212	0.0250	0.0294	0.0332	0.0388	0.0640	0.1015	0.1164	0.1311	0.1463	0.1600
19	0.0203	0.0238	0.0278	0.0314	0.0368	0.0600	0.0935	0.1071	0.1199	0.1374	0.1482
20	0.0196	0.0227	0.0264	0.0302	0.0353	0.0570	0.0864	0.1002	0.1121	0.1286	0.1369
21	0.0190	0.0217	0.0250	0.0289	0.0337	0.0540	0.0839	0.0948	0.1084	0.1198	0.1288
22	0.0185	0.0208	0.0238	0.0278	0.0323	0.0518	0.0794	0.0894	0.0988	0.1118	0.1212
23	0.0181	0.0201	0.0230	0.0268	0.0310	0.0492	0.0749	0.0836	0.0933	0.1043	0.1142
24	0.0177	0.0194	0.0224	0.0258	0.0298	0.0468	0.0704	0.0788	0.0882	0.0984	0.1071
25	0.0173	0.0189	0.0218	0.0248	0.0286	0.0447	0.0686	0.0749	0.0836	0.0927	0.1000
26	0.0169	0.0182	0.0213	0.0240	0.0274	0.0426	0.0636	0.0712	0.0791	0.0885	0.0948
27	0.0165	0.0177	0.0206	0.0232	0.0264	0.0407	0.0606	0.0676	0.0747	0.0843	0.0896
28	0.0161	0.0172	0.0203	0.0225	0.0256	0.0391	0.0576	0.0649	0.0706	0.0801	0.0859
29	0.0157	0.0166	0.0196	0.0219	0.0249	0.0377	0.0565	0.0621	0.0671	0.0769	0.0822
30	0.0153	0.0164	0.0193	0.0213	0.0242	0.0364	0.0536	0.0593	0.0643	0.0719	0.0786
31	0.0148	0.0160	0.0188	0.0207	0.0236	0.0352	0.0518	0.0569	0.0618	0.0719	0.0753
32	0.0145	0.0156	0.0183	0.0201	0.0229	0.0340	0.0491	0.0547	0.0591	0.0686	0.0722
33	0.0141	0.0152	0.0178	0.0196	0.0223	0.0329	0.0475	0.0527	0.0573	0.0661	0.0691
34	0.0137	0.0148	0.0173	0.0190	0.0217	0.0319	0.0459	0.0507	0.0555	0.0638	0.0660
35	0.0133	0.0144	0.0168	0.0185	0.0211	0.0309	0.0444	0.0488	0.0537	0.0611	0.0639
36	0.0129	0.0141	0.0164	0.0180	0.0205	0.0300	0.0429	0.0470	0.0519	0.0598	0.0608
37	0.0125	0.0136	0.0160	0.0176	0.0200	0.0291	0.0414	0.0454	0.0501	0.0567	0.0576
38	0.0122	0.0135	0.0156	0.0172	0.0195	0.0283	0.0400	0.0440	0.0483	0.0546	0.0553
39	0.0120	0.0133	0.0152	0.0168	0.0190	0.0275	0.0386	0.0426	0.0465	0.0525	0.0531
40	0.0116	0.0131	0.0148	0.0164	0.0185	0.0267	0.0375	0.0414	0.0447	0.0512	0.0510
41	0.0116	0.0129	0.0144	0.0161	0.0182	0.0260	0.0364	0.0402	0.0430	0.0499	0.0493
42	0.0114	0.0127	0.0140	0.0158	0.0178	0.0253	0.0355	0.0389	0.0417	0.0475	0.0482
43	0.0112	0.0125	0.0137	0.0155	0.0174	0.0248	0.0346	0.0379	0.0408	0.0464	0.0471
44	0.0110	0.0123	0.0134	0.0152	0.0170	0.0243	0.0338	0.0369	0.0394	0.0452	0.0460
45	0.0108	0.0121	0.0131	0.0149	0.0166	0.0238	0.0329	0.0359	0.0385	0.0440	0.0448
46	0.0106	0.0119	0.0129	0.0146	0.0162	0.0233	0.0320	0.0349	0.0376	0.0423	0.0438
47	0.0104	0.0117	0.0127	0.0143	0.0158	0.0228	0.0311	0.0340	0.0367	0.0416	0.0427
48	0.0103	0.0115	0.0125	0.0141	0.0156	0.0228	0.0311	0.0340	0.0367	0.0416	0.0426
49	0.0102	0.0113	0.0123	0.0139	0.0155	0.0223	0.0303	0.0332	0.0358	0.0406	0.0416
50	0.0101	0.0111	0.0121	0.0137	0.0152	0.0218	0.0298	0.0317	0.0343	0.0390	0.0398
51	0.0100	0.0109	0.0120	0.0136	0.0151	0.0217	0.0295	0.0314	0.0331	0.0378	0.0383
52	0.0099	0.0107	0.0119	0.0133	0.0148	0.0206	0.0276	0.0303	0.0323	0.0361	0.0373
53	0.0097	0.0106	0.0118	0.0131	0.0145	0.0201	0.0270	0.0298	0.0318	0.0352	0.0363
54	0.0096	0.0104	0.0116	0.0129	0.0143	0.0197	0.0264	0.0288	0.0307	0.0339	0.0343
55	0.0094	0.0103	0.0115	0.0127	0.0139	0.0193	0.0258	0.0282	0.0299	0.0321	0.0343
56	0.0093	0.0102	0.0113	0.0125	0.0137	0.0189	0.0252	0.0275	0.0292	0.0313	0.0333
57	0.0092	0.0101	0.0112	0.0123	0.0135	0.0185	0.0247	0.0268	0.0285	0.0306	0.0324
58	0.0091	0.0100	0.0110	0.0121	0.0133	0.0182	0.0242	0.0262	0.0279	0.0301	0.0318
59	0.0090	0.0098	0.0109	0.0119	0.0131	0.0179	0.0238	0.0257	0.0274	0.0296	0.0312
60	0.0089	0.0096	0.0106	0.0117	0.0129	0.0176	0.0234	0.0252	0.0270	0.0291	0.0306
61	0.0088	0.0095	0.0105	0.0116	0.0127	0.0173	0.0230	0.0247	0.0266	0.0288	0.0301
62	0.0087	0.0094	0.0104	0.0115	0.0126	0.0170	0.0226	0.0242	0.0262	0.0281	0.0296
63	0.0086	0.0091	0.0104	0.0112	0.0123	0.0167	0.0222	0.0238	0.0257	0.0278	0.0291
64	0.0085	0.0090	0.0102	0.0111	0.0121	0.0164	0.0218	0.0234	0.0252	0.0271	0.0288
65	0.0084	0.0089	0.0101	0.0109	0.0119	0.0161	0.0215	0.0230	0.0247	0.0266	0.0281
66	0.0082	0.0088	0.0099	0.0108	0.0117	0.0159	0.0211	0.0225	0.0242	0.0261	0.0276
67	0.0081	0.0087	0.0098	0.0107	0.0115	0.0157	0.0207	0.0221	0.0237	0.0256	0.0271
68	0.0080	0.0086	0.0096	0.0105	0.0114	0.0155	0.0204	0.0217	0.0232	0.0251	0.0266
69	0.0079	0.0085	0.0095	0.0104	0.0113	0.0153	0.0198	0.0213	0.0227	0.0246	0.0261
70	0.0078	0.0084	0.0094	0.0103	0.0111	0.0150	0.0194	0.0209	0.0222	0.0241	0.0256

ΠΗΓΕΣ

Βιβλιογραφία :

- Ίλια Βόντα, Αλέξανδρος Καραγρηγορίου, (2017) Εφαρμοσμένη Στατιστική και Στοιχεία πιθανοτήτων, Εκδόσεις Παρασκήνιο
- Δ.Φουσκάκης, (2013) Ανάλυση Δεδομένων με χρήση της R , Εκδόσεις Τσότρας
- Χαράλαμπος Χ. Δαμιανού. (2016), Μεθοδολογία Δειγματοληψίας, τεχνικές και εφαρμογές , Εκδόσεις Σοφία
- Sheldon Ross, (2011), Βασικές αρχές θεωρίας Πιθανοτήτων, Εκδόσεις Κλειδάριθμος
- Conover, W.J (1999) Practical non parametric Statistics, 3rd Edition, Wiley
- The International Statistical Institute, (2003) "The Oxford Dictionary of Statistical Terms", εκδ. Yadolah Dodge, Oxford University Press

Επιστημονικά άρθρα :

- Deshpande, J.V. (1983). A class of tests for exponentiality against increasing failure rate average alternatives . *Biometrika* 70, 514-518
- Dipankar Bandyopadhyay and Asit P. Basu (1989). A Note on Tests for Exponentiality by Deshpande . *Biometrika* 76, 403-405
- Shapiro and Wilk (1972). Exponentiality test when the sample is censored , *Technometrics* 14 , 355-370
- Lilliefors, H.W.(1969). On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *JASA* 64, 387-389
- Lin, C.C and Mudholkar, G.S (1980). A test of exponentiality based on the bivariate F distribution. *Technometrics* 22, 79-82
- Hollander, M. and Proschan, F.(1972). Testing whether new is better than used. *Annals of Math. Stat.* 43, 1136-1146
- Hartley , H.O (1950). The maximum F-Ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37, 308-312
- Harris, C.M (1976). A note on testing for exponentiality. *Naval Research Logistics Quarterly* 23, 169-175
- Hahn , G.J and Shapiro , S.S. (1967). *Statistical Models in Engineering*. John Wiley and Sons
- Gnedenko, B.V , Belyayev , YU. K. , Solovyev, A.D. (1969). *Mathematical Models of Reliability Theory*. Academic Press

- Gail, M.H and Gastwirth, J.L. (1978). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. JRSS B 40, 350-357
- Gail, M. and Gastwirth, J.L. (1978), A Scale-Free Goodness of Fit Test for the Exponential Distribution Based on the Lorenz Curve. Journal of the American Statistical Association, 73, 787-793
- Epstein, B. (1960). Tests for the validity of the assumption that the underlying distribution of life is exponential (part 1). Technometrics 2
- Cox, D.R and Oakes, D. (1984) Analysis of Survival Data. Chapman and Hall.
- Steven Ascher (1990): A survey of tests for exponentiality, Communications in Statistics - Theory and Methods, 19:5, 1811-1825

Πηγές διαδικτύου :

https://en.wikipedia.org/wiki/Exponential_distribution
https://en.wikipedia.org/wiki/Statistical_hypothesis_testing
https://en.wikipedia.org/wiki/Glossary_of_probability_and_statistics

<https://statistics.columbian.gwu.edu/joseph-gastwirth>
https://en.wikipedia.org/wiki/Hubert_Lilliefors
https://en.wikipedia.org/wiki/Andrey_Kolmogorov
[https://en.wikipedia.org/wiki/Nikolai_Smirnov_\(mathematician\)](https://en.wikipedia.org/wiki/Nikolai_Smirnov_(mathematician))
https://en.wikipedia.org/wiki/Samuel_Sanford_Shapiro
https://en.wikipedia.org/wiki/Martin_Wilk
<https://stat.fsu.edu/HollanderLecture>
https://en.wikipedia.org/wiki/Myles_Hollander
https://fr.wikipedia.org/wiki/Frank_Proschan
https://en.wikipedia.org/wiki/Herman_Otto_Hartley
https://en.wikipedia.org/wiki/Boris_Vladimirovich_Gnedenko
<https://typeset.io/authors/govind-s-mudholkar-342adhxr2h>
<https://rochester.academia.edu/GovindSMudholkar>
[https://en.wikipedia.org/wiki/David_Cox_\(statistician\)](https://en.wikipedia.org/wiki/David_Cox_(statistician))
<https://prabook.com/web/david.oakes/3644762>
<https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800230116>

<https://en.wikipedia.org/wiki/P%E2%80%93plot>
<https://en.wikipedia.org/wiki/Q%E2%80%93plot>

https://en.wikipedia.org/wiki/Chi-squared_test

<https://www.investopedia.com/terms/c/chi-square-statistic.asp>

<https://www.kaggle.com/datasets/usgs/earthquake-database>