

Κατάλογος Διπλωματικών Εργασιών Ακαδημαϊκού Έτους 2022-2023 (21<sup>η</sup> και 22<sup>η</sup> σειρά) του

Π.Μ.Σ. «Εφαρμοσμένη Στατιστική»

15<sup>η</sup>/15.05.2023 Συνέλευση Τμήματος

	Τίτλος Θέματος	Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία
1	<p>Ελληνικά: Μοντέλα ευπάθειας στην ανάλυση επιβίωσης.</p> <p>Αγγλικά: Frailty models in survival analysis.</p> <p>Κατεύθυνση: Β</p> <p>Ονοματεπώνυμο προτείνοντος: Ανδρουλάκης Εμμανουήλ</p> <p>Βαθμίδα: Επίκουρος Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Μια από τις βασικές υποθέσεις στα μοντέλα επιβίωσης, είναι ότι ο υπό μελέτη πληθυσμός είναι ομοιογενής, αναφορικά με κάποιες μετρήσιμες μεταβλητές. Αποτέλεσμα της ομοιογένειας αυτής, είναι ότι οι πειραματικές μονάδες του πληθυσμού διατρέχουν τον ίδιο κίνδυνο. Σε πολλές εφαρμογές όμως, η υπόθεση αυτή είτε παραβιάζεται είτε είναι ακατάλληλη.</p> <p>Ένας από τους τρόπους περιγραφής της ανομοιογένειας ενός πληθυσμού είναι μέσω της εισαγωγής στο μοντέλο, μιας μη παρατηρήσιμης τυχαίας μεταβλητής, καλούμενη ως ευπάθεια, η οποία αποτελεί τον παράγοντα τυχαίων επιδράσεων και ακολουθεί μια συγκεκριμένη κατανομή. Η ευπάθεια αντιπροσωπεύει μη μετρήσιμους παράγοντες κινδύνου που διαφοροποιούν τις πειραματικές μονάδες ως προς το χρόνο επιβίωσής τους. Στην παρούσα εργασία, θα γίνει μια περιγραφή του πλαισίου των μοντέλων ευπάθειας, των κυριότερων κατανομών που χρησιμοποιούνται και των βασικών μεθοδολογιών ανάλυσης, με έμφαση στην περίπτωση όπου έχουμε παρατηρήσεις οργανωμένες σε συστάδες. Επίσης, θα γίνει μια εφαρμογή των τεχνικών που θα παρουσιαστούν σε πραγματικά δεδομένα για το λέμφωμα τύπου Non-Hodgkin.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b></p> <ol style="list-style-type: none"> <li>1. Balan, T. A., &amp; Putter, H. (2020). A tutorial on frailty models. <i>Statistical methods in medical research</i>, 29(11), 3424-3454.</li> <li>2. Fan, J., &amp; Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. <i>The Annals of Statistics</i>, 30(1), 74-99.</li> <li>3. Govindarajulu, U. S., &amp; D'Agostino Sr, R. B. (2020). Review of current advances in survival analysis and frailty models. <i>Wiley Interdisciplinary Reviews: Computational Statistics</i>, 12(6), e1504.</li> <li>4. Hanagal, D. D. (2019). <i>Modeling Survival Data Using Frailty Models</i>. Springer Nature.</li> <li>5. Kosorok, M. R., Lee, B. L., &amp; Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. <i>The Annals of Statistics</i>, 32(4), 1448-1491.</li> <li>6. Wienke, A. (2011), <i>Frailty Models in Survival Analysis</i>, Chapman and Hall/CRC.</li> </ol>
	Τίτλος Θέματος	Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία
2	<p>Ελληνικά: Πληροφοριακή αποκοπή στην ανάλυση δεδομένων επιβίωσης.</p> <p>Αγγλικά: Informative censoring in survival data analysis.</p> <p>Κατεύθυνση: Β</p> <p>Ονοματεπώνυμο προτείνοντος: Ανδρουλάκης Εμμανουήλ</p> <p>Βαθμίδα: Επίκουρος Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Από τις βασικές υποθέσεις στην ανάλυση δεδομένων επιβίωσης με αποκομμένες (ή λογοκριμένες/διακεκομμένες) παρατηρήσεις, είναι η λεγόμενη μη-πληροφοριακή αποκοπή. Βάσει αυτής, η κατανομή των χρόνων επιβίωσης δεν παρέχει πληροφορία σχετικά με την κατανομή των χρόνων αποκοπής και αντίστροφα. Η παραβίαση αυτής της υπόθεσης, οδηγεί σε μεροληπτικά αποτελέσματα, καθώς οι τυπικές μέθοδοι που θεωρούν την αποκοπή ως μη-πληροφοριακή, οδηγούν σε υπερεκτίμηση ή υποεκτίμηση της συνάρτησης επιβίωσης.</p> <p>Για να αντιμετωπιστεί το συγκεκριμένο πρόβλημα, έχουν αναπτυχθεί αρκετές και διαφορετικές προσεγγίσεις στη διεθνή βιβλιογραφία. Στην παρούσα εργασία, θα γίνει μια επιλεκτική</p>

		<p>ανασκόπηση και σύγκριση βασικών τεχνικών διαχείρισης του φαινομένου της πληροφοριακής αποκοπής στην ανάλυση δεδομένων επιβίωσης.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b></p> <ol style="list-style-type: none"> <li>1. Arboretti, R., Fontana, R., Pesarin, F., &amp; Salmaso, L. (2018). Nonparametric combination tests for comparing two survival curves with informative and non-informative censoring. <i>Statistical methods in medical research</i>, 27(12), 3739-3769.</li> <li>2. Collett, D. (2003). <i>Modelling survival data in medical research</i>, Chapman &amp; Hall/CRC.</li> <li>3. Danieli, C., Remontet, L., Bossard, N., Roche, L., &amp; Belot, A. (2012). Estimating net survival: the importance of allowing for informative censoring. <i>Statistics in medicine</i>, 31(8), 775-786</li> <li>4. Rebolj Kodre, A., &amp; Pohar Perme, M. (2013). Informative censoring in relative survival. <i>Statistics in medicine</i>, 32(27), 4791-4802.</li> <li>5. Scharfstein, D. O., &amp; Robins, J. M. (2002). Estimation of the failure time distribution in the presence of informative censoring. <i>Biometrika</i>, 89(3), 617-634.</li> </ol> <p>Siannis, F., Copas, J., &amp; Lu, G. (2005). Sensitivity analysis for informative censoring in parametric survival models. <i>Biostatistics</i>, 6(1), 77-91.</p>
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
3	<p>Ελληνικά: Ποινικοποιημένη Λογιστική παλινδρόμηση και εφαρμογές.</p> <p>Αγγλικά: Penalized Logistic regression and applications.</p> <p>Κατεύθυνση: Β</p> <p>Ονοματεπώνυμο προτείνοντος: Ανδρουλάκης Εμμανουήλ</p> <p>Βαθμίδα: Επίκουρος Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Στην παρούσα εργασία θα μελετηθεί το μοντέλο της Λογιστικής παλινδρόμησης, στο οποίο έχει επιβληθεί κάποιος συγκεκριμένος περιορισμός για τις εκτιμήσεις των παραμέτρων του. Αυτό επιτυγχάνεται με τη χρήση κατάλληλων συναρτήσεων ποινής. Η διαδικασία της ποινικοποίησης, οδηγεί στη συρρίκνωση των συντελεστών των εξηγηματικών μεταβλητών, θέτοντας ορισμένους ίσους με μηδέν. Κατά αυτόν τον τρόπο, επιτυγχάνεται ταυτόχρονα η εκτίμηση των παραμέτρων του μοντέλου και η επιλογή των σημαντικών μεταβλητών. Επιπλέον, στο πλαίσιο της εργασίας θα περιγραφούν και θα συγκριθούν ορισμένες από τις κυριότερες συναρτήσεις ποινής που έχουν προταθεί στη διεθνή βιβλιογραφία, εφαρμόζοντας στη συνέχεια το ποινικοποιημένο μοντέλο Λογιστικής παλινδρόμησης σε δεδομένα από το χώρο της Υγείας.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b></p> <ol style="list-style-type: none"> <li>1. Fan, J., &amp; Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. <i>Journal of the American statistical Association</i>, 96(456), 1348-1360.</li> <li>2. Fan, J., &amp; Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. <i>Statistica Sinica</i>, 101-148.</li> <li>3. Gareth, J., Daniela, W., Trevor, H., &amp; Robert, T. (2013). <i>An introduction to statistical learning: with applications in R</i>. Springer.</li> <li>4. Hastie, T., Tibshirani, R., Friedman, J. H., &amp; Friedman, J. H. (2009). <i>The elements of statistical learning: data mining, inference, and prediction</i> (Vol. 2, pp. 1-758). New York: Springer.</li> <li>5. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. <i>Journal of the Royal Statistical Society: Series B</i>, 58(1), 267-288.</li> <li>6. Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. <i>Annals of Statistics</i>, 38(2), 894-942.</li> </ol>

		Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. <i>Journal of the Royal Statistical Society: Series B</i> , 67(2), 301-320.
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
4	<p>Ελληνικά: Συσταδοποίηση δεδομένων κίνησης με αλγορίθμους διανυσματικών δεδομένων.</p> <p>Αγγλικά: Trajectory data clustering with vector-based algorithms.</p> <p>Κατεύθυνση: ΕΔ</p> <p>Ονοματεπώνυμο προτείνοντος: Νικόλαος Πελέκης</p> <p>Βαθμίδα: Αναπληρωτής Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Σκοπός της διπλωματικής εργασίας είναι η μελέτη και ο σχεδιασμός μεθόδων (συμπεριλαμβανομένων τεχνικών μηχανικής μάθησης) διανυσματοποίησης δεδομένων κίνησης και πειραματικής αξιολόγησής τους με αλγορίθμους συσταδοποίησης.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b>          Σχετική εργασία: Xiucheng Li, et al., Deep Representation Learning for Trajectory Similarity Computation  <a href="https://personal.ntu.edu.sg/gaocong/papers/icde18-camera-ready-final.pdf">https://personal.ntu.edu.sg/gaocong/papers/icde18-camera-ready-final.pdf</a>.</p>
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
5	<p>Ελληνικά: Εφαρμογή τεχνικών μηχανικής μάθησης σε μεγάλα χωρικά δεδομένα.</p> <p>Αγγλικά: Application of machine learning techniques in BIG spatial data.</p> <p>Κατεύθυνση: ΕΔ</p> <p>Ονοματεπώνυμο προτείνοντος: Νικόλαος Πελέκης</p> <p>Βαθμίδα: Αναπληρωτής Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Σκοπός της διπλωματικής εργασίας είναι η μελέτη και η πειραματική αξιολόγηση μεθόδων μηχανικής μάθησης σε μεγάλου όγκου χωρικά δεδομένα.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b>          Σχετική εργασία: Sabek et al., Machine Learning Meets Big Spatial Data (Revised), MDM, 2021.  <a href="http://people.csail.mit.edu/ibrahimsabek/pdf/21_tutorial_mdm.pdf">http://people.csail.mit.edu/ibrahimsabek/pdf/21_tutorial_mdm.pdf</a></p>
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
6	<p>Ελληνικά: Συγκριτική αξιολόγηση συστημάτων επεξεργασίας και ανάλυσης ναυτιλιακών δεδομένων κίνησης.</p> <p>Αγγλικά: A benchmark for maritime data processing and analytics.</p> <p>Κατεύθυνση: ΕΔ</p> <p>Ονοματεπώνυμο προτείνοντος: Νικόλαος Πελέκης</p> <p>Βαθμίδα: Αναπληρωτής Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Σκοπός της διπλωματικής εργασίας είναι η συγκριτική/πειραματική μελέτη συστημάτων και μεθόδων επεξεργασίας και ανάλυσης μεγάλου όγκου ναυτιλιακών δεδομένων κίνησης.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b>          Σχετική εργασία: Ray et al., Heterogeneous integrated dataset for Maritime Intelligence, surveillance, and reconnaissance, 2019.  <a href="https://doi.org/10.1016/j.dib.2019.104141">https://doi.org/10.1016/j.dib.2019.104141</a></p>
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
7	<p>Ελληνικά: Στατιστικά μοντέλα για την ανάλυση της απόδοσης των ομάδων στο Παγκόσμιο Κύπελο ποδοσφαίρου.</p> <p>Αγγλικά: Statistical models for analysing the performance of teams in the football World Cup.</p> <p>Κατεύθυνση: ΕΔ</p> <p>Ονοματεπώνυμο προτείνοντος: Κων/νος Πολίτης</p> <p>Βαθμίδα: Αναπληρωτής Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Τα τελευταία χρόνια, μια πληθώρα στατιστικών εργαλείων και μοντέλων χρησιμοποιούνται για την πρόβλεψη και ανάλυση αποτελεσμάτων σε αθλητικούς αγώνες.</p> <p>Στην εργασία θα μελετηθούν δεδομένα από πρόσφατες διοργανώσεις του Παγκοσμίου Κυπέλου ποδοσφαίρου, τόσο με κλασικές στατιστικές μεθόδους, όσο και με εργαλεία από την εξόρυξη δεδομένων, για να εξετασθεί ποια από τα στατιστικά στοιχεία που συλλέγονται σε έναν αγώνα ποδοσφαίρου είναι τα πλέον καθοριστικά, αφενός μεν για την ανάδειξη του τελικού νικητή σε έναν αγώνα, αφετέρου δε για την συνολική πορεία μιας ομάδας στο Παγκόσμιο Κύπελο.</p>

		<p><b>Ενδεικτική Βιβλιογραφία:</b></p> <ol style="list-style-type: none"> <li>1. Castellano J, Casamichana, D, Lago C. (2012) The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. <i>Journal of Human Kinetics</i>; volume 31, 139-147.</li> <li>2. Castellano J, Perea A, Blanco-Villaseñor A. (2007). Has soccer changed in the last three world championships? <i>J Sports Sci Med</i>, 6: 2.</li> <li>3. Lago-Ballesteros J, Lago C. (2010) Performance in team sports: Identifying the keys to success in soccer. <i>J Hum Kinet</i>, 25: 85-91.</li> <li>4. Lepschy, H, Woll, A and Wäsche, H (2021) Success Factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil. <i>Front. Psychology</i>, March 2021.</li> <li>5. Lawlor J, Low D, Taylor S, Williams AM (2003). The FIFA World Cup 2002: an analysis of successful versus unsuccessful teams. <i>J Sport Sci</i>, 22: 500-520.</li> <li>6. Liu, H., Yi, Q, Giménez, J-V, Gómez, M. A. &amp; Lago-Peñas, C. (2015) Performance profiles of football teams in the UEFA Champions League considering situational efficiency. <i>International Journal of Performance Analysis in Sport</i>, 15:1, 371-390.</li> <li>7. Pollard, R and Reep, C (1997) Measuring the Effectiveness of Playing Strategies at Soccer. <i>JRSS Series D (The Statistician)</i>, Vol. 46, No. 4, pp. 541-550.</li> </ol>
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
8	<p>Ελληνικά: Δείκτες Επιλογής του Βέλτιστου Πλήθους των Συστάδων: Μια Εμπειρική Συγκριτική Μελέτη.</p> <p>Αγγλικά: Indexes for Selecting the Best Number of Clusters: An Empirical Comparative Study.</p> <p>Κατεύθυνση: ΕΔ</p> <p>Ονοματεπώνυμο προτείνοντος: Αθανάσιος Ρακιτζής</p> <p>Βαθμίδα: Επίκουρος Καθηγητής</p> <p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<p>Για την εφαρμογή μεθόδων συσταδοποίησης (<i>clustering methods</i>) είναι συχνά απαραίτητο να επιλεγεί, προτού γίνει η εφαρμογή της μεθόδου, το πλήθος των συστάδων στις οποίες θα καταλήξει η μέθοδος. Πολλές φορές δεν είναι σαφές ποιο μπορεί να είναι αυτό το πλήθος ενώ μια λανθασμένη επιλογή μπορεί να οδηγήσει σε μη αποδεκτή συσταδοποίηση, αν επιλέξουμε έναν πολύ μεγάλο ή έναν πολύ μικρό αριθμό συστάδων. Για το λόγο αυτό έχουν προταθεί διάφοροι δείκτες οι οποίοι προσπαθούν να δώσουν μια εικόνα σχετικά με το ιδανικό πλήθος συστάδων. Στην παρούσα διπλωματική εργασία θα γίνει παρουσίαση των κυριότερων δεικτών επιλογής του πλήθους συστάδων. Επιπρόσθετα, μέσα από μια εμπειρική μελέτη, θα γίνει προσπάθεια να προταθεί ο δείκτης με την καλύτερη απόδοση.</p> <p><b>Ενδεικτική Βιβλιογραφία:</b></p> <ol style="list-style-type: none"> <li>1. Charrad M., Ghazzali N., Boiteau V., Niknafs A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set., <i>Journal of Statistical Software</i>, 61(6), 1-36., URL <a href="http://www.jstatsoft.org/v61/i06/">http://www.jstatsoft.org/v61/i06/</a>.</li> <li>2. Zhou, S., Xu, Z., &amp; Liu, F. (2016). Method for determining the optimal number of clusters based on agglomerative hierarchical clustering. <i>IEEE transactions on neural networks and learning systems</i>, 28(12), 3007-3017.</li> <li>3. Akhanli, S. E., &amp; Hennig, C. (2020). Comparing clusterings and numbers of clusters by aggregation of calibrated clustering validity indexes. <i>Statistics and Computing</i>, 30, 1523-1544.</li> </ol>
	<b>Τίτλος Θέματος</b>	<b>Σύντομη Περιγραφή / Ενδεικτική Βιβλιογραφία</b>
9	<p>Ελληνικά: Μοντέλα επιβίωσης ασθενών με ηπατίτιδα.</p> <p>Αγγλικά: Survival models for the patients with hepatitis.</p> <p>Κατεύθυνση: Β</p>	<p>Η ηπατίτιδα είναι μια φλεγμονώδης πάθηση του ήπατος. Προκαλείται συνήθως από μια ιογενή λοίμωξη, αλλά υπάρχουν και άλλες πιθανές αιτίες αυτής. Αποτελεί σημαντική απειλή για τη δημόσια υγεία σε παγκόσμια κλίμακα. Στην εργασία αυτή</p>

	<p>Όνοματεπώνυμο προτείνοντος: Γεώργιος Τζαβελάς</p>	<p>κατασκευάζονται μοντέλα επιβίωσης των ασθενών με ηπατίτιδα και μελετώνται τα συμπτώματα και οι παράγοντες κινδύνου που σχετίζονται με την ασθένεια αυτή.</p>
<p>Βαθμίδα: Αναπληρωτής Καθηγητής</p>	<p><b><u>Ενδεικτική βιβλιογραφία</u></b></p>	
<p>Τμήμα: Στατιστικής και Ασφαλιστικής Επιστήμης</p>	<ol style="list-style-type: none"> <li>1. Regression Methods in Biostatistics (2012) E. Vittinghoff, D.V. Glidden, S. C. Shiboski, Ch. E. McCulloch. 2nd Ed. Springer Verlag.</li> <li>2. Applied Survival Analysis (1999) D.W. Hosmer and S.Lemeshow. John Wiley &amp; Sons, Inc. Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data. 2nd Ed. Willey Interscience.</li> <li>3. <i>McCullagh, Peter Nelder J.A. (1989). Generalized Linear Models. 2nd Ed. NY Chapman and Hall.</i></li> <li>4. Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. <i>Biometrics</i>, 34,541-554.</li> <li>5. Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data. 2nd Ed. Willey Interscience.</li> <li>6. Wen-Bin Yang , En-Qiang Chen, Hong-Xia Bi , Lang Bai, Xue-Bin Chen , Ping Feng , Hong Tang (2012) .Different models in predicting the short-term prognosis of patients with hepatitis B virus-related acute-on-chronic liver failure. (11) 3 May-June pp. 311-319</li> </ol>	