

## Οδηγίες για τις εργασίες στο μάθημα των Γενικευμένων Γραμμικών Μοντέλων

- Μην ξεπεράσετε τις 20 σελίδες A4 (font size 12, single spacing)
- Απαντήστε στα ερωτήματα με τη σειρά δίνοντας τις αντίστοιχες επικεφαλίδες (Ερώτημα 1, 2, κλπ)
- Προτιμάται η ελληνική γλώσσα αλλά οι εργασίες είναι δεκτές και στα αγγλικά
- Αν χρειαστεί να αποκλείσετε άτομα από το dataset της εργασίας εξηγήστε τους λόγους και περιγράψτε τα βήματα της διαδικασίας αποκλεισμού (π.χ. Αρχικό dataset 100 άτομα, εξαιρέθηκαν 5 γιατί είχαν missing values σε βασικές μεταβλητές, κλπ). Όλη η υπόλοιπη εργασία θα πρέπει να βασίζεται στο τελικό δείγμα της μελέτης που προέκυψε.
- Στη διερευνητική ανάλυση (Exploratory Data Analysis – EDA) αποφύγετε την επανάληψη της ίδιας πληροφορίας σε πίνακες και γραφήματα. Χρησιμοποιείτε γραφήματα που βοηθούν στην κατανόηση των δεδομένων λαμβάνοντας υπόψη σας τα κύρια ερωτήματα της εργασίας
- Οι πίνακες και τα γραφήματα πρέπει να έχουν αντίστοιχα legends και αρίθμηση. Πρέπει να είναι self-explanatory (να μπορεί κάποιος να καταλάβει το βασικό τους μήνυμα χωρίς να πρέπει να ανατρέξει στο κείμενο). Πρέπει να χρησιμοποιούνται κατάλληλα variable labels και value labels και όχι τα ονόματα των μεταβλητών. Στις ποσοτικές μεταβλητές πρέπει να δίνονται και οι μονάδες. Για αποτελέσματα μοντέλων συνιστάται η χρήση της εντολής reformat (δείτε το αρχείο «Reformat and Foreach.pdf»). Στο ίδιο αρχείο περιλαμβάνονται οδηγίες για τη χρήση της εντολής foreach η οποία μπορεί να φανεί χρήσιμη για επαναλαμβανόμενες αναλύσεις ίδιου τύπου με διαφορετικές μεταβλητές)
- Παραδείγματα από κώδικα Stata που χρησιμοποιήσατε μπορείτε να δώσετε σε παράρτημα στο τέλος (αν το κρίνετε απαραίτητο) και να αναφέρεστε σε αυτά μέσα στο κυρίως κείμενο της εργασίας.
- Για τη δημιουργία πολυπαραγοντικών μοντέλων αποφύγετε τις αυτοματοποιημένες μεθόδους (δηλαδή με τις εντολές sw του Stata) και προσπαθήστε να ακολουθήσετε τον αλγόριθμο του Collett που περιγράφεται πιο κάτω:

### **Collett's Model Selection Approach**

*This approach assumes that all variables are considered to be on an equal footing, and there is no a priori reason to include any specific variables (like treatment).*

*Approach:*

*(1) Fit a univariate model for each covariate, and identify the predictors significant at some level  $p_1$ , say 0.20.*

*(2) Fit a multivariate model with all significant univariate predictors, and use backward selection to eliminate nonsignificant variables at some level  $p_2$ , say 0.10.*

*(3) Starting with final step (2) model, consider each of the non-significant variables from step (1) using forward selection, with significance level  $p_3$ , say 0.10.*

*(4) Do final pruning of main-effects model (omit variables that are non-significant, add any that are significant), using stepwise regression with significance level  $p_4$ . At this stage, you may also consider adding interactions between any of the main effects currently in the model, under the hierarchical principle.*

*Collett recommends using a likelihood ratio test for all variable inclusion/exclusion decisions.*