Άσκηση 1. Έστω $X_1, \ldots, X_n$ τυχαίο δείγμα από κατανομή με διασπορά σ^2 και θεωρήστε τον εκτιμητή αντικατάστασης της διασποράς $T_n := \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. Να δείξετε ότι

$$T_{\text{jack}} = S^2$$

όπου T_{jack} είναι ο jackknife εκτιμητής της διασποράς και $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.
Hint για την απόδειξη: στη διάλεξη δείξαμε τη σχέση

$$n\widehat{\sigma}^2 = \frac{1}{n}\left\{(n-1)\sum_{i=1}^n X_i^2 - \sum_{i=1}^n \sum_{j\neq i} X_i X_j\right\}.$$

Άσκηση 2. Θεωρήστε το αρχείο δεδομένων `income.txt` που είναι διαθέσιμο στο e-class (κατηγορία Έγγραφα/Υλικό Παπασταμούλη/datasets), το οποίο καταγράφει τα εισοδήματα για τυχαίο δείγμα 25 ατόμων από τρία διαφορετικά χωριά A, B και C. Ο συντελεστής Gini χρησιμοποιείται από τους οικονομολόγους για μέτρηση της ανισοκατανομής του εισοδήματος σε έναν πληθυσμό και δίνεται από τον τύπο

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2\bar{x}}.$$

Τιμές κοντά στο 0 εκφράζουν ισοκατανομή του εισοδήματος, ενώ τιμές κοντά στο 1 εκφράζουν ανισοκατανομή του εισοδήματος. Για κάθε χωριό χρησιμοποιήστε jackknife για να

1. εκτιμήσετε την μεροληψία του δειγματικού συντελεστή Gini
2. εκτιμήσετε το τυπικό σφάλμα του δειγματικού συντελεστή Gini
3. υπολογίσετε 95% ασυμπτωτικό διάστημα εμπιστοσύνης ίσων ουρών για τον πληθυσμιακό συντελεστή Gini.

Να συγκεντρώσετε τα αποτελέσματά σας σε εναν 3×5 πίνακα όπου οι στήλες αντιστοιχούν σε: σημειακή εκτίμηση, εκτίμηση μεροληψίας, εκτίμηση τυπικού σφάλματος, άνω και κάτω όριο του ΔΕ, για κάθε ένα από τα τρία χωριά A, B και C.
Χρησιμοποιήστε δικές σας εντολές και να συγκρίνετε την απαντησή σας και με το πακέτο `boot`.

Άσκηση 3. Έστω X (Y, αντίστοιχα) η τυχαία μεταβλητή που παριστάνει τον πληθυσμό σε χιλιάδες μιας πόλης των Ηνωμένων Πολιτειών Αμερικής το έτος 1920 (1930, αντίστοιχα). Έστω (X_i, Y_i), $i = 1, \ldots, n$, τυχαίο δείγμα n το πλήθος πόλεων των Ηνωμένων Πολιτειών Αμερικής τα έτη 1920 και 1930, αντίστοιχα. Ενδιαφέρει η εκτίμηση του συνολικού πληθυσμού, έστω d, των ΗΠΑ το 1930, με βάση το συγκεκριμένο δείγμα και γνωρίζοντας ότι το 1920 ο πληθυσμός των ΗΠΑ ήταν ίσος με a. Αν οι ΗΠΑ έχουν συνολικά k το πλήθος πόλεις, τότε

$$\mathrm{E}(X) = a/k, \quad \mathrm{E}(Y) = d/k$$

και, επομένως, ο συνολικός πληθυσμός d το 1930 θα είναι

$$d = a\theta,$$

όπου $\theta = \frac{\mathrm{E}(Y)}{\mathrm{E}(X)}$. Να θεωρήσετε τα δεδομένα του Πίνακα 1 και τον εκτιμητή αντικατάστασης του λόγου μέσων τιμών θ, δηλαδή

$$T_n = \frac{\bar{Y}_n}{\bar{X}_n}$$

όπου $\bar{X}_n$ και $\bar{Y}_n$ οι δειγματικοί μέσοι κατά το 1920 και 1930, αντίστοιχα. Να εκτιμήσετε τη μεροληψία και το τυπικό σφάλμα του T_n μέσω jackknife. Να υπολογίσετε και ένα 95% προσεγγιστικό διάστημα εμπιστοσύνης για το d το 1930.

Πίνακας 1: Μέγεθος (σε χιλιάδες κατοίκους) 49 πόλεων των ΗΠΑ το 1920 (x_i) και το 1930 (y_i).

x_i	y_i	x_i	y_i	x_i	y_i
138	143	76	80	67	67
93	104	381	464	120	115
61	69	387	459	172	183
179	260	78	106	66	86
48	75	60	57	46	65
37	63	507	634	121	113
29	50	50	64	44	58
23	48	77	89	64	63
30	111	64	77	56	142
2	50	40	60	40	64
38	52	136	139	116	130
46	53	243	291	87	105
71	79	256	288	43	61
25	57	94	85	43	50
298	317	36	46	161	232
74	93	45	53	36	54
50	58				