

ΑΣΚΗΣΕΙΣ
ΑΝΑΛΥΣΗ ΚΑΤΗΓΟΡΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

- Εξετάσθηκαν 368 ασθενείς καπνιστές. Μετά από 2 έτη εξετάσθηκαν πόσοι από αυτούς είχαν επιβιώσει. Επί συνόλου 110 ασθενών που κάπνιζαν πολύ, 96 αναρρώνουν. Επί συνόλου 115 ασθενών που δεν κάπνιζουν, 15 πέθαναν και 2 επιβίωσαν, ενώ 99 ασθενείς ενώ κάπνιζαν λίγο επιβίωσαν. Τέλος συνολικά 34 ασθενείς πέθαναν ενώ 105 επιβίωσαν. Κατασκευάστε τον πίνακα διπλής εισόδου. Υπολογίστε την καλύτερη διαμέριση για τον παραπάνω πίνακα.. Με βάση την διαμέριση υπολογίστε τους αντίστοιχους κινδύνους και συγκρίνατε τους με εκείνους του αρχικού πίνακα. Ποία είναι η εκτίμηση μέγιστης πιθανοφάνειας?
- Να ελέγξετε (ε.σ. 5%) αν οι παρακάτω 45 παρατηρήσεις προέρχονται από την ομοιόμορφη κατανομή στο (0,1) και από την Poisson με άγνωστο λ .

0.00	0.06	0.38	0.38	0.56	0.44	0.44	0.60	0.77
0.17	0.18	0.38	0.34	0.46	0.44	0.53	0.75	0.72
0.14	0.04	0.28	0.32	0.43	0.42	0.56	0.61	0.98
0.04	0.20	0.36	0.54	0.48	0.40	0.79	0.66	0.85
0.16	0.26	0.22	0.51	0.52	0.43	0.71	0.78	0.86

Με βάση τον παραπάνω πίνακα εξετάστε αν τα δεδομένα σας προέρχονται από τριωνυμική κατανομή με πιθανότητες $\pi_1=2\theta^2$, $\pi_2=2\theta(1-\theta)$ και $\pi_3=(1-\theta)^2$.

Αν αναλογικά οι 3 ομάδες κατανέμονται ως 6: 4: 2 πιστεύετε ότι ανταποκρίνονται στην πραγματικότητα με τον καλύτερο δυνατό τρόπο?

- Αναφέρεται δείκτες συνάφειας που ικανοποιούν την προϋπόθεση οι πίνακες να είναι τετραγωνικοί (Αναλύστε τους).
- Έστω ο παρακάτω πίνακας συνάφειας . Υπολογίστε:

TABLE 3.10 Data for Problem 3.4

Race	Party Identification		
	Democrat	Independent	Republican
Black	103	15	11
White	341	105	405

Source: 1991 General Social Survey, National Opinion Research Center.

τον καλύτερο δείκτη συνάφειας (και γιατί), την πιθανοφάνεια, τον αναμενόμενο κίνδυνο (με 95% διαστήματα εμπιστοσύνης), την κατανομή των δεδομένων καθώς και την ανάλυση της. Στην περίπτωση που η φυλή αποτελεί ξεχωριστό πίνακα, αναφέρεται το είδος της δειγματοληψίας (και αναλύστε την), και ελέγξτε την κατανομή των δεδομένων. Ποια είναι καλύτερη και γιατί?