



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Μάθημα: Εισαγωγή στην Γραμμική Παλινδρόμηση

Διδάσκουσα: Ευδοκία Ξεκαλάκη

Εργασία

Μη πειραματικές μελέτες έχουν δείξει ότι η χαμηλή πρόσληψη ή η χαμηλή συγκέντρωση ρετινόλης στο πλάσμα αίματος θα μπορούσε να σχετίζεται με αυξημένο κίνδυνο ανάπτυξης ορισμένων τύπων καρκίνου. Τα στοιχεία που περιλαμβάνονται στο αρχείο <http://stat-athens.aueb.gr/~exek/data-plasmaretinol-B-v1.xls> προέρχονται από μία βιοϊατρική έρευνα επί 314 ασθενών με στόχο την διερεύνηση ενδεχόμενης στατιστικά σημαντικής συνεισφοράς προσωπικών χαρακτηριστικών και διατροφικών παραγόντων στον καθορισμό του επιπέδου Y της ρετινόλης στο πλάσμα αίματος.

Οι μεταβλητές που συνδέονται με τα παραπάνω χαρακτηριστικά και παράγοντες είναι οι εξής:

AGE: ηλικία (έτη)

SEX: φύλο (1=άρρεν, 2=θήλυ).

SMOKSTAT: χρήση καπνού (1=ποτέ, 2=διέκοψε, 3=ναι)

CALORIES: ημερήσια κατανάλωση θερμίδων

FAT: ημερήσια κατανάλωση λίπους (grams)

ALCOHOL: αριθμός καταναλισκομένων αλκοολούχων ποτών ανά εβδομάδα

CHOLESTEROL: κατανάλωση χοληστερόλης (mg ανά ημέρα).

RETDIET: διατροφική κατανάλωση ρετινόλης (mcg ανά ημέρα)

RETPLASMA: περιεκτικότητα πλάσματος αίματος σε ρετινόλη (ng/ml)

Μεταχειρισθείτε τα στοιχεία αυτά ως στοιχεία αναφερόμενα σε έναν πληθυσμό μεγέθους 314 ατόμων και με την βοήθεια του πίνακα τυχαίων αριθμών που παρατίθεται στο τέλος του αρχείου ή μιας γεννήτριας τυχαίων αριθμών (π.χ. <http://www.randomizer.org/form.htm>), να επιλέξετε ένα απλό τυχαίο δείγμα 30 ατόμων.

1. Να περιγράψετε τα βήματα τα οποία ακολουθήσατε για την επιλογή του τυχαίου δείγματός σας και να σημειωθούν στον πίνακα τυχαίων αριθμών οι τυχαίοι αριθμοί που επιλέξατε ή να επισυναφθεί στο γραπτό μία εκτύπωση της τυχαίας ακολουθίας αριθμών που παρήχθη από την γεννήτρια τυχαίων αριθμών που χρησιμοποιήσατε.

2. Με βάση τα στοιχεία του δείγματος που επιλέξατε, να απαντήσετε στα ερωτήματα που ακολουθούν αναπτύσσοντας τα συμπεράσματά σας.
- Να προτείνετε ένα μοντέλο για την περιγραφή του τρόπου με τον οποίο οι μεταβλητές αυτές ενδέχεται να συνεισφέρουν στην διαμόρφωση της μέσης περιεκτικότητας του αίματος σε ρετινόλη. Αιτιολογήστε την επιλογή σας.
 - Να εφαρμόσετε το μοντέλο που επιλέξατε στα στοιχεία του δείγματός σας και να σχολιάσετε την εφαρμογή του.
 - Πώς θα μπορούσατε να αξιοποιήσετε το μοντέλο αυτό για να μελετήσετε την σχέση μεταξύ διατροφικών παραγόντων και συγκεντρώσεως ρετινόλης στο πλάσμα αίματος
 - ανδρών καπνιστών
 - γυναικών που έκοψαν το κάπνισμα
 - ανδρών και γυναικών που δεν υπήρξαν ποτέ καπνιστές
3. Αγνοώντας την δομή του πληθυσμού ως προς τα χαρακτηριστικά τα αναφερόμενα στο φύλο και τις καπνιστικές συνήθειες καθώς και ενδεχόμενες αλληλεπιδράσεις των ανεξαρτήτων μεταβλητών:
- ποιες μεταβλητές θα επιλέγατε για την περιγραφή του τρόπου που τα διατροφικά χαρακτηριστικά συνεισφέρουν στην διαμόρφωση των τιμών της μέσης περιεκτικότητας του αίματος σε ρετινόλη;
 - θα μπορούσατε να συμπεράνετε ότι όλα τα χαρακτηριστικά έχουν στατιστικά σημαντική επίδραση;
 - Ποια χαρακτηριστικά θα μπορούσαν να αξιολογηθούν ως μη έχοντα στατιστικά σημαντική συνεισφορά;
 - Χρησιμοποιώντας το μοντέλο με τις μεταβλητές που έχετε αξιολογήσει ως στατιστικά σημαντικές, ποιο είναι ένα 95% διάστημα εμπιστοσύνης για την μέση περιεκτικότητα του πλάσματος αίματος σε ρετινόλη;

Η εργασία θα υποβληθεί σε εκτύπωση την **Δευτέρα 30 Ιουνίου κατά το πέρας της γραπτής εξέτασης του μαθήματος μαζί με το γραπτό της εξέτασης**. Θα υποβληθεί επίσης ηλεκτρονικά στην διεύθυνση oxek@aub.gr, μέχρι την **ίδια ημερομηνία και ώρα**.

Το όνομα του ηλεκτρονικού αρχείου να είναι του τύπου **LinRegr-erg-επίθετο** (π.χ. LinRegr-erg-Georgiou). Στην πρώτη σελίδα να αναγράφεται το όνομα, ο αριθμός μητρώου και τα στοιχεία της εργασίας.

Καθηγήτρια Ευδοκία Ξεκαλάκη

Ιούνιος 2014

Τυχαίοι Αριθμοί στο διάστημα (0, 1)

3407	1440	6960	8675	5649	5793	1514
5044	9859	4658	7779	7986	0520	6697
0045	4999	4930	7408	7551	3124	0527
7536	1448	7843	4801	3147	3071	4749
7653	4231	1233	4409	0609	6448	2900
6157	1144	4779	0951	3757	9562	2354
6593	8668	4871	0946	3155	3941	9662
3187	7434	0315	4418	1569	1101	0043
4780	1071	6814	2733	7968	8541	1003
9414	6170	2581	1398	2429	4763	9192
1948	2360	7244	9682	5418	0596	4971
1843	0914	9705	7861	6861	7865	7293
4944	8903	0460	0188	0530	7790	9118
3882	3195	8287	3298	9532	9066	8225
6596	9009	2055	4081	4842	7852	5915
4793	2503	2906	6807	2028	1075	7175
2112	0232	5334	1443	7306	6418	9639
0743	1083	8071	9779	5973	1141	4393
8856	5352	3384	8891	9189	1680	3192
8027	4975	2346	5786	0693	5615	2047
3134	1688	4071	3766	0570	2142	3492
0633	9002	1305	2256	5956	9256	8979
8771	6069	1598	4275	6017	5946	8189
2672	1304	2186	8279	2430	4896	3698
3136	1916	8886	8617	9312	5070	2720
6490	7491	6562	5355	3794	3555	7510
8628	0501	4618	3364	6709	1289	0543
9270	0504	5018	7013	4423	2147	4089
5723	3807	4997	4699	2231	3193	8130
6228	8874	7271	2621	5746	6333	0345
7645	3379	8376	3030	0351	8290	3640
6842	5836	6203	6171	2698	4086	5469
6126	7792	9337	7773	7286	4236	1788
4956	0215	3468	8038	6144	9753	3131
1327	4736	6229	5965	7215	6458	3937
9188	1516	5279	5433	2254	5768	8718
0271	9627	9442	9217	4656	7603	8826
2127	1847	1331	5122	8332	8195	3322
2102	9201	2911	7318	7670	6079	2676
1706	6011	5280	5552	5180	4630	4747
7501	7635	2301	0889	6955	8113	4364
