

Απαντήσεις θεμάτων Εργαστηρίου Στατιστικής 12 Οκτωβρίου 2007

Θέμα 1^ο : Στα δεδομένα του αρχείου coronary artery data.sav (δες αντίστοιχη εικόνα) εφαρμόσαμε Independent-Samples T-Test το οποίο μας έδωσε τα εξής αποτελέσματα.

	time	group	var	var
1	1014	health		
2	684	health		
3	810	health		
4	990	health		
5	840	health		
6	978	health		
7	1002	health		
8	1110	health		
9	864	diseas		
10	636	diseas		
11	638	diseas		
12	708	diseas		
13	786	diseas		
14	600	diseas		
15	1320	diseas		
16	750	diseas		
17	594	diseas		
18	750	diseas		
19				
20				
21				

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
Treadmill time in seconds	healthy	8	928,50	138,121	48,833
	disease	10	764,60	213,750	67,594

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Treadmill time in seconds	Equal variances assumed	,137	,716	1,873	16	,080	163,90	87,524	-21,642	349,442
	Equal variances not assumed			1,966	15,439	,068	163,90	83,388	-13,398	341,198

- (Α) Ποιοι είναι οι δύο ανεξάρτητοι πληθυσμοί στους οποίους αναφέρεται αυτή η στατιστική δοκιμασία; **(0,5μ)**
Απ. Οι δύο ανεξάρτητοι πληθυσμοί είναι το σύνολο όλων των υγιών και το σύνολο όλων των ασθενών
- (Β) Ποια είναι η υπόθεση που ελέγχουμε με την παραπάνω δοκιμασία; **(0,5μ)**
Απ. Ελέγχουμε την υπόθεση H_0 : η μέση τιμή της μεταβλητής time στον πληθυσμό των υγιών είναι ίση με τη μέση τιμή της στον πληθυσμό των ασθενών.
- (Γ) Γίνεται δεκτή ή όχι και γιατί; (πλήρης αιτιολόγηση) **(1,5μ)**
Απ. Από τη στήλη Sig. παρατηρούμε ότι $p=0,716 > 0,05$ άρα από το Levene's test δέχομαι ότι οι διασπορές είναι ίσες. Στην στήλη Sig. (2-tailed) διαβάζουμε $p=0,08 > 0,05$ άρα η υπόθεση H_0 γίνεται δεκτή ή ισοδύναμα η παρατηρούμενη διαφορά μεταξύ των δύο μέσων τιμών δεν είναι αρκετά μεγάλη ώστε να απορρίψουμε την ισότητα τους.

Θέμα 2^ο : Στα δεδομένα του αρχείου dietstudy.sav (δες αντίστοιχη εικόνα) εφαρμόσαμε Paired-Samples T-Test στα ζευγάρια των μεταβλητών basetg-finaltg και basewt-finalwt και πήραμε τα εξής αποτελέσματα.

	id	basetg	finaltg	basewt	finalwt
1	1	159	194	120	109
2	2	93	122	124	116
3	3	130	158	129	123
4	4	174	154	123	120
5	5	148	93	132	125
6	6	148	90	126	122
7	7	85	101	138	131
8	8	180	99	140	131
9	9	92	183	157	152
10	10	89	82	160	155
11	11	204	100	109	104
12	12	182	104	172	165
13	13	110	72	123	117
14	14	88	108	158	151
15	15	134	110	158	147
16	16	84	81	155	147
17					

Paired Samples Statistics

Pair		Mean	N	Std. Deviation	Std. Error Mean
1	Triglyceride Pre - Triglyceride Post	131,25	16	40,739	10,185
2	Weight Pre - Weight Post	139,00	16	18,529	4,632

Paired Samples Correlations

Pair		N	Correlation	Sig.
1	Triglyceride Pre & Triglyceride Post	16	,111	,682
2	Weight Pre & Weight Post	16	,993	,000

Paired Samples Test

Pair		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
1	Triglyceride Pre - Triglyceride Post	15,56	51,794	12,949	-12,04	43,16	1,202	15	,248
2	Weight Pre - Weight Post	6,81	2,257	,564	5,61	8,02	12,071	15	,000

(A) Ποιες είναι οι δύο υποθέσεις που ελέγχουμε με την παραπάνω δοκιμασία; (0,5μ)

Απ. Ελέγχουμε τις υποθέσεις H_0 : η μέση τιμή της μεταβλητής basetg είναι ίση με την μέση τιμή της μεταβλητής finaltg σε όλο τον πληθυσμό και

H_1 : η μέση τιμή της μεταβλητής basewt είναι ίση με την μέση τιμή της μεταβλητής finalwt σε όλο τον πληθυσμό και

(B) Γίνονται δεκτές ή όχι και γιατί; (πλήρης αιτιολόγηση) (2μ)

Απ. Η H_0 γίνεται δεκτή διότι $p=0,248>0,05$ ενώ η H_1 απορρίπτεται διότι $p=0,000<0,05$

Θέμα 3^ο : Με τις μεταβλητές lifeexpf και lifeexpm του αρχείου world95.sav (δες αντίστοιχη εικόνα) δημιουργήσαμε διάγραμμα διασποράς το οποίο και επεξεργαστήκαμε. Θεωρήσαμε ως ανεξάρτητη τη μεταβλητή lifeexpm και ως εξαρτημένη την μεταβλητή lifeexpf.

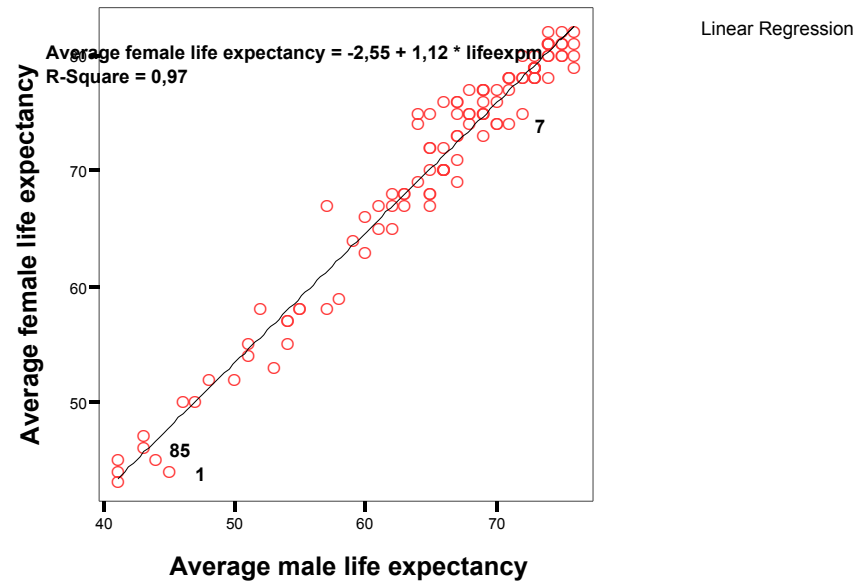
(Α) Ο αριθμός 1 στο σχήμα σε ποια καταχώρηση αντιστοιχεί; **(0,5μ) Απ. Στην πρώτη καταχώρηση δηλαδή στο Αφγανιστάν.**

(Β) Πως ονομάζεται η εξίσωση που εμφανίζεται στο σχήμα; **Απ. Εξίσωση της ευθείας γραμμικής παλινδρόμησης της μεταβλητής lifeexpf πάνω στην μεταβλητή lifeexpm (0,5μ)**

(Γ) Αν σε κάποια χώρα που δεν υπάρχει στο δείγμα μας οι άνδρες ζούνε κατά μέσο όρο 65 χρόνια, πόσα περιμένουμε να ζούνε οι γυναίκες εκεί; **(1μ)**

Απ. Η εξίσωση είναι $lifeexpf = -2,55 + 1,12 * lifeexpm$. Για $lifeexpm = 65$ χρόνια υπολογίζουμε $lifeexpf = 70,25$ χρόνια.

World95 - SPSS Data Editor								
File Edit View Data Transform Analyze Graphs Utilities Window Help								
1 : country Afghanistan								
	country	populatn	density	urban	religion	lifeexpf	lifeexpm	literacy
1	Afghanistan	20500	25,0	18	Muslim	44	45	29
2	Argentina	33900	12,0	86	Catholic	75	68	95
3	Armenia	3700	126,0	68	Orthodox	75	68	98
4	Australia	17800	2,3	85	Protstnt	80	74	100
5	Austria	8000	94,0	58	Catholic	79	73	99
6	Azerbaijan	7400	86,0	54	Muslim	75	67	98



Θέμα 4^ο : Στο παρακάτω αρχείο θέλοντας να βρούμε αν οι μεταβλητές staff και smoke είναι ανεξάρτητες, εφαρμόσαμε τη δοκιμασία X^2 , η οποία έδωσε τα παρακάτω αποτελέσματα. (Α) Πως ερμηνεύονται οι αριθμοί στις γραμμές με τίτλο «expected value»; **(1μ) Απ. Είναι οι συχνότητες που θα είχαμε στα αντίστοιχα κελιά αν οι μεταβλητές staff και smoke ήταν ανεξάρτητες.** (Β) Ποια είναι η υπόθεση που ελέγχουμε με τη δοκιμασία X^2 ; **(1μ) Απ. Ελέγχουμε την υπόθεση H_0 : Οι μεταβλητές staff και smoke είναι ανεξάρτητες.** (Γ) Γίνεται δεκτή η υπόθεση ή όχι; (αιτιολογήστε αναλυτικά) **(1μ) Στην γραμμή Pearson Chi-Square και στη στήλη Asymp. Sig. (2-sided) διαβάζουμε $p=0,172$ άρα η υπόθεση γίνεται δεκτή ή ισοδύναμα οι παρατηρούμενες συχνότητες δεν διαφέρουν αρκετά από τις αναμενόμενες ώστε να απορριφθεί η ανεξαρτησία των δύο μεταβλητών.**

staff	smoke	count	
1	Senior Managers	None	4,00
2	Senior Managers	Light	2,00
3	Senior Managers	Medium	3,00
4	Senior Managers	Heavy	2,00
5	Junior Managers	None	4,00
6	Junior Managers	Light	3,00
7	Junior Managers	Medium	7,00
8	Junior Managers	Heavy	4,00
9	Senior Employees	None	25,00
10	Senior Employees	Light	10,00
11	Senior Employees	Medium	12,00
12	Senior Employees	Heavy	4,00
13	Junior Employees	None	18,00
14	Junior Employees	Light	24,00
15	Junior Employees	Medium	33,00
16	Junior Employees	Heavy	13,00
17	Secretaries	None	10,00
18	Secretaries	Light	6,00
19	Secretaries	Medium	7,00
20	Secretaries	Heavy	2,00
21			

Smoking * Staff Group Crosstabulation

			Staff Group					
			Senior Managers	Junior Managers	Senior Employees	Junior Employees	Secretaries	Total
Smoking	None	Count	4	4	25	18	10	61
		Expected Count	3,5	5,7	16,1	27,8	7,9	61,0
	Light	Count	2	3	10	24	6	45
		Expected Count	2,6	4,2	11,9	20,5	5,8	45,0
	Medium	Count	3	7	12	33	7	62
		Expected Count	3,5	5,8	16,4	28,3	8,0	62,0
	Heavy	Count	2	4	4	13	2	25
		Expected Count	1,4	2,3	6,6	11,4	3,2	25,0
Total	Count		11	18	51	88	25	193
	Expected Count		11,0	18,0	51,0	88,0	25,0	193,0

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16,442 ^a	12	,172
Likelihood Ratio	16,348	12	,176
Linear-by-Linear Association	,000	1	,994
N of Valid Cases	193		

a. 7 cells (35,0%) have expected count less than 5. The minimum expected count is 1,42.