Μια χημική ανάλυση που ακολουθεί κανονική κατανομή έδωσε τα παρακάτω αποτελέσματα:

(i) Να εξετασθεί αν $\sigma^2_A=\sigma^2_B$ για α=0.02
(ii) Να εξετασθεί αν οι μέσες τιμές είναι ίσες για α=0.05
(iii) Να βρεθεί 95% δ.ε. για τη διαφορά μέσων τιμών υποθέτοντας ότι οι διασπορές είναι ίσες.

A	B
93.08	93.95
91.36	93.42
91.60	92.20
91.91	92.46
92.79	92.73
92.80	93.31
91.03	92.94
	93.66

(i)

meanA= 92.08
s^2_A= 0.650
n_A= 7
meanB= 93.08
s^2_B= 0.366
n_B= 8

H_0: $\sigma_A^2/\sigma_B^2=1$

H_1: $\sigma_A^2/\sigma_B^2<>1$

s_A^2/s_B^2= 1.777 **$v_A=6$, $v_B=7$** **$F_{vA,vB;a/2}$= 7.19**
$F_{vB,vA;a/2}$= 8.26 **$1/F_{vB,vA;a/2}$= 0.121**
$1/F_{vB,vA;a/2}$ $<s_A^2/s_B^2<$ $F_{vA,vB;a/2}$

Άρα μπορούμε να δεχθούμε ότι οι διασπορές δεν διαφέρουν

(ii)
Πρέπει να ξαναγίνει έλεγχος για την ισότητα των διασπορών για a=0.05 πριν εφαρμόσουμε την δοκιμασία "t"

H_0: $\mu_A-\mu_B=0$ **s=σταθμισ. διακ.= 0.705**

H_1: $\mu_A-\mu_B\neq 0$ **t= -2.741** **$t_{15-2;0,025}$= 2.160**

Άρα οι μέσοι διαφέρουν σημαντικά

(iii)

$\mu_A-\mu_B$ = 92,08-93,08 +/- 2,16*0,705*0,518 = -1,0 +/- 0,789

Δύο δίαιτες Α,Β εφαρμόσθηκαν σε δύο σειρές ποντικών επί μια εβδομάδα.
Υποθέτοντας ότι η αύξηση του βάρους τους ακολουθεί την κανονική κατανομή
(i) να εξετασθεί αν $\sigma^2_A=\sigma^2_B$ για α=0.05,
(ii) να εξετασθεί εάν οι δύο δίαιτες διαφέρουν στην αύξηση του βάρους
(iii) να δοθεί 95% δ.ε. για τη διαφορά $\mu_A-\mu_B$

A Αύξ. Βάρ. [gr]
78.1
72.4
76.2
74.3
77.4
78.4
76.0
76.1

B Αύξ. Βάρ. [gr]
79.1
81.0
77.3
79.1
80.0
79.1
79.1
77.3
80.2
79.1

meanA= 76.1
s^2_A= 4.65
n_A= 7
meanB= 79.1
s^2_B= 1.51
n_B= 9

H_0: $\sigma_A^2/\sigma_B^2=1$

H_1: $\sigma_A^2/\sigma_B^2>1$

s_A^2/s_B^2= 3.08 $\nu_A=6$, $\nu_B=8$ **$F_{\nu A,\nu B;a}$=** 3.58

Άρα μπορούμε να δεχθούμε ότι οι διασπορές δεν διαφέρουν

H_0: $\mu_A-\mu_B=0$
H_1: $\mu_A-\mu_B\neq 0$

$t_{16-2;0,025}$= 2.145

s=Σταθμ. Διακ.= $$s=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$ = 1.83

$$t=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{s\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$ = **-3.253**

$|t| > t_{16-2;0,025}$= 2.145 **→ οι μέσοι διαφέρουν στατ. σημαντικά**

Διάστημα εμπιστοσύνης για την διαφορά:

$$\bar{x}-\bar{y}-t_{\nu 1+n2,a/2}\left(\frac{s}{\sqrt{n}}\right)\le \mu_A-\mu_B\le \bar{x}-\bar{y}+t_{\nu 1+n2,a/2}\left(\frac{s}{\sqrt{n}}\right)$$

$\bar{x}-\bar{y}=$ **-3.0** **-3.998** $\le \mu_A-\mu_B \le$ **-2.040**

$t_{\nu 1+n2,a/2}\left(\frac{s}{\sqrt{n}}\right)=$ **0.9788**

Έστω, ότι ένα δείγμα ελέγχου χρησιμοποιήθηκε σε ένα πείραμα ελέγχου ποιότητας μιας μεθόδου για 30 συνεχόμενες ημέρες. Η θεωρητική (ισχυριζόμενη) συγκέντρωση της γλυκόζης στο δείγμα αυτό είναι 1120 mg/L.
Ο πειραματικός μέσος είναι 1110 mg/L και η τυπική απόκλιση 25 mg/L.
Είναι η μέση τιμή των δεδομένων σημαντικά διαφορετική από την θεωρητική τιμή;

H_0 : $\bar{y} = \mu$

H_1 : $\bar{y} \neq \mu$

$\mu =$ **1120**

$\bar{y} =$ **1110**

$s =$ **25**

$$t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}} = -2.19$$

Ο πίνακας της κατανομής "t" για β.ε.= 29 και a/2=0.025 δίνει κρίσιμο t=2.045.
Άρα η μέση τιμή των μετρήσεων της γλυκόζης είναι σημαντικά διαφορετική από την θεωρητική.

Ένα εργαστήριο ξεκινά τη χρήση μιας μεθόδου προσδιορισμού της ανοσοσφαιρίνης Α του ορού. Δείγματα έφθασαν από δύο περιοχές της χώρας. Τυχαία δείγματα υγιών ατόμων κάθε περιοχής εξετάσθηκαν για να διαπιστωθεί αν το διάστημα αναφοράς γι αυτές τις περιοχές είναι το ίδιο.

	Περιοχή Α	**Περιοχή Β**
mean (mg/L)	2260	2650
standard deviation (mg/L)	584	473
number of samples	33	29

$H_0 : \mu_A = \mu_B$

$H_1 : \mu_A \neq \mu_B$

Unpaired t-test

1) Δοκιμασία F για να ελέγξουμε αν οι διακυμάνσεις είναι στατιστικά "ίσες".

$$F = \frac{(584)^2}{(473)^2} = \mathbf{1.52}$$

Ο πίνακας για την κατανομή F (a=0.05) δίνει για v_A=33-1 και v_B=29-1 μια κρίσιμη τιμή F=1.86. 1.56<1.86 άρα οι διακυμάνσεις δεν διαφέρουν σημαντικά με a=0.05. Έτσι μπορούμε να εφαρμόσουμε την δοκιμασία t για τον έλεγχο των μέσων τιμών.

2) Δοκιμασία t για τον έλεγχο των μέσων τιμών

$$s = \sqrt{\frac{(33-1)(584)^2 + (29-1)(473)^2}{33+29-2}} = 535 \qquad t = \frac{2260 - 2650}{535 \cdot \sqrt{1/33 + 1/29}} = -2.86$$

Ο πίνακας της κατανομής t με β.ε.=v_A+v_B =60 δίνει μια κρίσιμη τιμή t=2.00. Επειδή |-2.86| > 2.00 συμπεραίνουμε ότι οι μέσοι είναι σημαντικά διαφορετικοί και ότι το εργαστήριο θα πρέπει να ορίσει διαφορετικά διαστήματα αναφοράς για τις δύο περιοχές.

Μια ομάδα 20 ασθενών υποφέρει από αϋπνίες. Διαθέτουμε ένα φάρμακο του οποίου την αποτελεσματικότητα θέλουμε να ελέγξουμε. Στην ομάδα χορηγούμε πρώτα placebo και μετρούμε τη διάρκεια του ύπνου σε h. Ακολούθως χορηγούμε το φάρμακο και ξαναμετρούμε τη διάρκεια του ύπνου.

t-test κατά ζεύγη

H_0: Δεν υπάρχει στατ. σημαντ. διαφ. στην διάρκεια του ύπνου με placebo και με το φάρμακο, $\delta_0=0$

H_1: Υπάρχει στατ. σημαντ. διαφ. στην διάρκεια του ύπνου με placebo και με το φάρμακο, $\delta_0\neq0$

	x[h]	y[h]	δ[h]
Ασθενής	Διάρκεια χωρίς φάρμ.	Διάρκεια με φάρμ.	Διαφορά
1	5.28	5.83	0.55
2	5.69	4.69	-1.00
3	4.81	4.43	-0.38
4	5.90	6.49	0.59
5	6.50	7.18	0.68
6	5.14	5.34	0.20
7	5.66	6.95	1.29
8	5.20	5.19	-0.01
9	5.35	5.20	-0.15
10	4.56	4.34	-0.22
11	5.97	5.29	-0.68
12	4.94	5.67	0.73
13	5.24	6.26	1.02
14	5.92	6.61	0.69
15	5.17	6.11	0.94
16	6.43	8.08	1.65
17	5.03	5.62	0.59
18	4.34	5.37	1.03
19	5.51	5.69	0.18
20	4.88	4.88	0.00

mean=	**5.376**	**5.761**	**0.385**
variance=	**0.330**	**0.914**	**0.454**
sd=	**0.574**	**0.956**	**0.674**
se=			**0.151**

$$t = \frac{\left(\overline{x} - \overline{y}\right) - \delta_0}{s_d / \sqrt{n}} = \mathbf{2.555}$$

$t_{a=0.05,df=19} = 1.729$ $\quad t > t_{cr} \rightarrow$ απορρίπτουμε την H_0

Διάστημα εμπιστοσύνης:

μέση δ - t*se < $\mu_2 - \mu_1$ < μέση δ + t*se

$0.385 - 1.729 * 0.151 < \mu_2 - \mu_1 < 0.385 + 1.729 * 0.151$

$0.124 < \mu_2 - \mu_1 < 0.646$

x_A = 74 από 250 ασθενείς που πήραν το φάρμακο Α βελτίωσαν την κατάστασή τους.
x_B = 92 από 250 ασθενείς που πήραν το φάρμακο Β βελτίωσαν την κατάστασή τους.
i) Μπορούμε να δεχθούμε ότι και τα δύο φάρμακα έχουν την ίδια αποτελεσματικότητα; (a=0,05)
ii) Δώστε ένα 98% δ.ε. για τη διαφορά του αποτελέσματος των φαρμάκων.

H_0: $p_A=p_B$
H_1: $p_A \neq p_B$

$p_A = x_A/n = 74/250 = 0,296$ $\qquad$ $p_B = x_B/n = 92/250 = 0,368$

$p_{AB} = (x_A+x_B)/(n+m) = (74+92)/500 = 0,332$

z = -1.709 $\qquad$ Για a=0.05 $\qquad$ **$Z_{a/2}$ = 1.96**

Επειδή $|z| < Z_{a/2}$ μπορούμε να δεχθούμε ότι τα δύο φάρμακα έχουν την ίδια αποτελεσματικότητα

Το διάστημα εμπιστοσύνης 98% (a/2=0,01) για τη διαφορά θα είναι ($Z_{a/2}$=2.33) :

$p_A - p_B +/- Z_{a/2}*sqrt[p_A(1-p_A)/250+p_B(1-p_B)/250)]$ $\qquad$ **(-0,17, 0,026)**

Το ερώτημα μπορεί να απαντηθεί και με την δοκιμασία χ^2

Παρατηρούμενες

Φάρμακο	Βελτίωση	Μη Βελτ.	Σύνολο
A	74	176	250
B	92	158	250
Σύνολο	166	334	500

Θεωρητικές

Φάρμακο	Βελτίωση	Μη Βελτ.	Σύνολο
A	83	167	250
B	83	167	250
Σύνολο	166	334	500

χ^2= 2.922 $\qquad$ με β.ε =1 και a=0.05 το κρίσιμο χ^2=3.841

Αρα αποδεχόμαστε ότι τα δύο φάρμακα είναι το ίδιο αποτελεσματικά

Από 150 ασθενείς, στους n=80 δώσαμε χάπι με ζάχαρη και στους m=70 δώσαμε ασπιρίνες. Μετά από μια εβδομάδα βρέθηκε ότι βελτιώθηκε η υγεία x=48 ασθενών που πήραν χάπι ζάχαρης και y=56 που πήραν ασπιρίνη.

(i) Μπορούμε να δεχθούμε ότι η θεραπεία με ασπιρίνες είναι προτιμότερη σε σ.σ. 0.05;

(ii) Να δοθεί δ.ε. 95% για τη διαφορά των ποσοστών p_1, p_2 (p_1=ποσοστό ασθενών που βελτιώθηκε η υγεία τους με χάπι ζάχαρης, p_2=ποσοστό ασθενών που βελτιώθηκε η υγεία τους με ασπιρίνη).

Ζάχαρη		Ασπιρίνη	
$p_1 = x/n =$	0.60	$p_2 = y/m =$	0.80

Η κοινή αναλογία των επιτυχιών είναι p=(x+y)/(n+m)= 0.693

(i) **H_0: $p_1 = p_2$**

H_1: $p_1 \neq p_2$

Αφού τα δείγματα είναι μεγάλα το στατιστικό που θα χρησιμοποιήσουμε είναι το Ζ. Έτσι:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)(\frac{1}{n}+\frac{1}{m})}} = -2.65$$

Για a=0.05 z_c=1.96. Επειδή $z<-z_c$, συμπεραίνουμε ότι δεχόμαστε την H_1.

(ii) Το δ.ε. για την διαφορά p_1-p_2 θα δίνεται από τον τύπο:

$$p_1 - p_2 \pm z_{a/2}\sqrt{\frac{p_1(1-p_1)}{n}+\frac{p_2(1-p_2)}{m}}$$

Για a/2=0.025 έχουμε Φ=0.475 άρα $z_{a/2}$=1.96.

με $\sqrt{\frac{p_1(1-p_1)}{n}+\frac{p_2(1-p_2)}{m}} =$ 0.072703 έχουμε:

$-0.343 < p_1-p_2 < -0.058$

x_A = 74 από 250 ασθενείς που πήραν το φάρμακο A βελτίωσαν την κατάστασή τους.
x_B = 92 από 250 ασθενείς που πήραν το φάρμακο B βελτίωσαν την κατάστασή τους.
i) Μπορούμε να δεχθούμε ότι και τα δύο φάρμακα έχουν την ίδια αποτελεσματικότητα; (a=0.05)
ii) Δώστε ένα 98% δ.ε. για τη διαφορά του αποτελέσματος των φαρμάκων.

H_0: $p_A = p_B$
H_1: $p_A \neq p_B$

$p_A = x_A/n = 74/250 = 0,296$ $p_B = x_B/n = 92/250 = 0,368$

$p_{AB} = (x_A+x_B)/(n+m) = (74+92)/500 = 0,332$

z = -1.709 Για a=0.05 **$Z_{a/2}$ = 1,96**

Επειδή $|z| < Z_{a/2}$ μπορούμε να δεχθούμε ότι τα δύο φάρμακα έχουν την ίδια αποτελεσματικότητα

Το διάστημα εμπιστοσύνης 98% (a/2=0,01) για τη διαφορά θα είναι ($Z_{a/2}$=2.33) :

$p_A - p_B +/- Z_{a/2}$*sqrt[$p_A(1-p_A)/250+p_B(1-p_B)/250$)] **(-0,17, 0,026)**

Το ερώτημα μπορεί να απαντηθεί και με την δοκιμασία χ^2

Παρατηρούμενες

Φάρμακο	Βελτίωση	Μη Βελτ.	Σύνολο
A	74	176	250
B	92	158	250
Σύνολο	166	334	500

Θεωρητικές

Φάρμακο	Βελτίωση	Μη Βελτ.	Σύνολο
A	83	167	250
B	83	167	250
Σύνολο	166	334	500

χ^2= 2.922 με β.ε =1 και a=0.05 το κρίσιμο χ^2=3.841

Αρα αποδεχόμαστε ότι τα δύο φάρμακα είναι το ίδιο αποτελεσματικά

Σε 346 αυτοκινητιστικά δυστυχήματα με αυτοκίνητα διαφόρων μεγεθών (μικρό, μεσαίο, μεγάλο) έγιναν θανατηφόρα και μη ατυχήματα σύμφωνα με τον πίνακα που ακολουθεί. Μπορούμε να συμπεράνουμε ότι το είδος του ατυχήματος έχει σχέση με το μέγεθος του αυτοκινήτου; (a=0.05)

Πίνακας συνάφειας

	Μικρό αυτ.	Μεσαίο αυτ.	Μεγάλο αυτ.	Σύνολο
Θανατηφόρα	67	26	16	109
Μη Θανατηφόρα	128	63	46	237
Σύνολο	195	89	62	346

H_0 **:** Το είδος του ατυχήματος ανεξάρτητο από το μέγεθος του αυτοκινήτου

H_1 **:** Το είδος του ατυχήματος εξαρτάται από το μέγεθος του αυτοκινήτου

Δοκιμασία "χ^2" ως test ανεξαρτησίας

Οι αναμενόμενες τιμές υπολογίζονται από τα αντίστοιχα επιμέρους σύνολα.

e11= 61.43 e12= 28.04 e13= 19.53

e21= 133.57 e22= 60.96 e23= 42.47

$X^2 = (o11-e11)^2/e11+(o21-e21)^2/e21+(o12-e12)^2/e12+(o22-e22)^2/e22+(o13-e13)^2/e13+(o23-e23)^2/e23$

$\mathbf{X^2= 1.885}$

β.ε. = 2

a= 0.05

$\mathbf{X_{2;0,05}^2= 5.991}$ Άρα δεν απορρίπτουμε την μηδενική υπόθεση

Ισχυρή και ασθενής δόση του παυσίπονου Ibuprofen, Aspirin και Placebo χορηγήθηκαν σε 4 ομάδες ασθενών για μια οδοντιατρική πράξη. Οι απαντήσεις για την αποτελεσματικότητα του παυσίπονου ήταν: "καθόλου", "μικρή", "μέτρια", "καλή" και "πολύ καλή"

H_0 : Τα αναλγητικά δεν διαφέρουν ως προς την αποτελεσματικότητά τους
H_1 : Τα αναλγητικά διαφέρουν ως προς την αποτελεσματικότητά τους

	Κατηγορία "j"					
Ομάδα "i"	**Καθόλου**	**Μικρή**	**Μέτρια**	**Καλή**	**Πολύ καλή**	**Σύνολο**
Ισχ. Δ. Ibu	1	3	5	25	52	86
Ασθ. Δ. Ibu	0	6	10	17	61	94
Aspirin	1	4	11	25	47	88
Placebo	0	18	10	37	32	97
Σύνολο	2	31	36	104	192	365

$e_{11}= 0.471$ $e_{12}= 7.304$ $e_{13}= 8.482$ $e_{14}= 24.504$ $e_{15}= 45.238$
$e_{21}= 0.515$ $e_{22}= 7.984$ $e_{23}= 9.271$ $e_{24}= 26.784$ $e_{25}= 49.447$
$e_{31}= 0.482$ $e_{32}= 7.474$ $e_{33}= 8.679$ $e_{34}= 25.074$ $e_{35}= 46.290$
$e_{41}= 0.532$ $e_{42}= 8.238$ $e_{43}= 9.567$ $e_{44}= 27.638$ $e_{45}= 51.025$

Επειδή δεν είναι όλες οι αναμενόμενες αναλογίες > 5, συμπτύσσουμε τις κατηγορίες "καθόλου" και "μικρή" σε "λίγη"

	Κατηγορία "j"				
Ομάδα "i"	**Λίγη**	**Μέτρια**	**Καλή**	**ολύ καλ**	**Σύνολο**
Ισχ. Δ. Ibu	4	5	25	52	86
Ασθ. Δ. Ibu	6	10	17	61	94
Aspirin	5	11	25	47	88
Placebo	18	10	37	32	97
Σύνολο	33	36	104	192	365

$e_{11}= 7.775$ $e_{12}= 8.482$ $e_{13}= 24.504$ $e_{14}= 45.238$
$e_{21}= 8.499$ $e_{22}= 9.271$ $e_{23}= 26.784$ $e_{24}= 49.447$
$e_{31}= 7.956$ $e_{32}= 8.679$ $e_{33}= 25.074$ $e_{34}= 46.290$
$e_{41}= 8.770$ $e_{42}= 9.567$ $e_{43}= 27.638$ $e_{44}= 51.025$

β.ε. = (4-1)x(4-1) = 9

$\mathbf{X^2 = 33 > \chi_{0.95}^2 = 16.92}$

Άρα απορρίπτουμε την H_0

Μετρήθηκε η ποσότητα πρωτεΐνης (gr/100 ml) στο αίμα ατόμων που ζουν σε διαφορετικές περιοχές Α, Β, Γ και οι τιμές που βρέθηκαν δίνονται στον πίνακα που ακολουθεί. Μπορούμε να ισχυρισθούμε ότι η ποσότητα πρωτείνης στο αίμα είναι η ίδια και στις τρεις περιοχές;

ANOVA

α/α	Α, n_1=7	Β, n_2=8	Γ, n_3=9
1	7.64	7.67	7.98
2	7.07	7.58	7.91
3	7.43	7.04	7.11
4	7.57	6.69	7.65
5	7.74	7.32	8.17
6	7.63	7.12	8.28
7	8.06	7.46	7.21
8		7.21	7.41
9			6.37

κ=3 **n=24**

-0.33904	-0.53992	-0.82369
1.602762	0.076195	0.396213

$\bar{x}_i =$ 7.591 7.261 7.566

$\bar{x} =$ 7.47

$s_i^2 =$ 0.091 0.101 0.370

Υποθέτουμε ότι τα δείγματα προέρχονται από κανονικούς πληθυσμούς

H_0 : $\mu_A = \mu_B = \mu_\Gamma$
H_1 : τουλάχιστον ένας μέσος διαφέρει

$$s_\alpha^2 = \sum_1^\kappa n_i(\bar{x}_i - \bar{x})^2 = 0.534$$

$$F = \frac{s_\alpha^2/(\kappa-1)}{s_\upsilon^2/(n-\kappa)} = 1.3304$$

$$s_\upsilon^2 = \sum_1^\kappa (n_i - 1)s_i^2 = 4.214$$

Ο πίνακας IV δίνει $F_{2,21,0.05}$ = 3.47 > F. Άρα δεχόμαστε την μηδενική υπόθεση.

Σύγκριση της περιεκτικότητας σε αιμοσφαιρίνη (Hb) στο αίμα ασθενών με κυτταρική ασθένεια Hb SS, Hb Sb και Hb SC. Υπάρχουν διαφορές μεταξύ των ασθενειών ως προς την περιεκτικότητα σε Hb;

Η περιεκτικότητα σε Hb είναι ο μόνος παράγοντας που θα κρίνει την διαφορά.
Άρα πρόκειται για ανάλυση με έναν παράγοντα (One Way Analysis of Variance).

H_0 = Οι ασθένειες δεν διαφέρουν ως προς την περιεκτικότητα σε Hb
H_1 = Οι ασθένειες διαφέρουν ως προς την περιεκτικότητα σε Hb

n	Hb SS g[100ml]	Hb Sb g[100ml]	Hb SC g[100ml]
1	7.2	8.1	10.7
2	8.4	11.1	12.0
3	9.1	9.2	13.3
4	7.7	11.9	11.3
5	8.5	10.0	12.1
6	9.8	12.0	13.8
7	8.0	10.4	11.5
8	8.6	12.1	12.3
9	10.1	10.6	13.9
10	8.1	10.9	11.6
11	8.7		12.6
12	10.3		11.7
13	8.3		12.6
14	9.1		11.8
15	8.4		13.3
16	9.1		

mean=	8.71	10.63	12.30	total mean=	10.49		
Sum of dev	10.70	14.84	12.42	SS within=	37.96	MS withinn=	1.00
				SS between=	100.03	MS between=	50.02

F= 50.1

Τέσσερα διαφορετικά είδη λιπάσματος A,B,C,D χρησιμοποιήθηκαν το καθένα σε εκτάσεις ίδιου εμβαδού, σπαρμένες με μια ποικιλία σιταριού. Η απόδοση ανά έκταση kg δίνεται στον πίνακα που ακολουθεί. Είναι τα τέσσερα λιπάσματα ίδια; Αν όχι, πιο λίπασμα είναι καλύτερο; (a=0.05)

ANOVA

H_0 : $\mu_A = \mu_B = \mu_\Gamma$

H_1 : τουλάχιστον ένας μέσος διαφέρει

A, 6	B, 5	C, 4	D, 6
50	55	70	54
55	57	68	71
57	59	68	67
56	60	66	66
59	57		61
53			53

mean x :	**55.0**	**57.6**	**68.0**	**62.0**
total mean :	**60.1**			
variance :	**10.00**	**3.80**	**2.67**	**53.60**

s_α^2 : 458.61

s_υ^2 : 341.21

F = 7.616

$F_{3,17;0,05}$: 3,20

$F > F_{3,17;0,05}$

Σε 12 αυτοκίνητα τεσσάρων διαφορετικών τύπων (Α,Β,Γ,Δ) βάλαμε 10 lt βενζίνης.
Η απόσταση σε km που διανύθηκε από κάθε αυτοκίνητο ήταν:

ANOVA

Απόσταση σε km			
Α	**Β**	**Γ**	**Δ**
72	84	92	80
80	88	104	76
88	104	104	84

Με την προϋπόθεση ότι ισχύουν οι βασικές υποθέσεις της α.δ., μπορούμε να συμπεράνουμε ότι η κατανάλωση είναι ίδια σε όλα τα αυτοκίνητα;

80.00	92.00	100.00	80.00	Total mean=	88.00
128.00	224.00	96.00	32.00		

H_0 : Η κατανάλωση είναι ίδια σε όλα τα αυτοκίνητα
H_1 : Τουλάχιστον ένας τύπος διαφέρει στην κατανάλωση

$$s_\alpha^2 = \sum_1^\kappa n_i(\overline{x}_i - \overline{x})^2 \qquad s_\upsilon^2 = \sum_1^\kappa (n_i - 1)s_i^2$$

$s_\alpha^2 =$ 864.00 MS μεταξύ= 288.00

$s_\upsilon^2 =$ 480.00 MS εντός= 60.00

$$F = \frac{\text{τετράγωνα μεταξύ ομάδων}}{\text{τετράγωνα εντός ομάδων}} = \frac{s_\alpha^2/(\kappa-1)}{s_\upsilon^2/(n-\kappa)} = \mathbf{4.80}$$

***$F_{3,8;0.05}$* = 4.07**

$F > F_{3,8;0.05}$ **άρα η H_0 απορρίπτεται**

Για να προσδιορισθεί η επίδραση του ναρκωτικού εφεδρίνη στους παλμούς της καρδιάς, χορηγήθηκε το ναρκωτικό σε έναν ασθενή για 6 ημέρες και σε διαφορετικές δόσεις. Τα αποτελέσματα συνοψίζονται στο πίνακα που ακολουθεί.

α) Να εκτιμηθεί η ευθεία y=α+βx+e

β) Να γίνει το διάγραμμα διασποράς και να σχεδιασθεί η ευθεία $y = \alpha + \beta x$ **. Να χρησιμοποιηθεί αυτή η ευθεία για να εκτιμηθούν οι παλμοί της καρδιάς για 3.5 χάπια την ημέρα.**

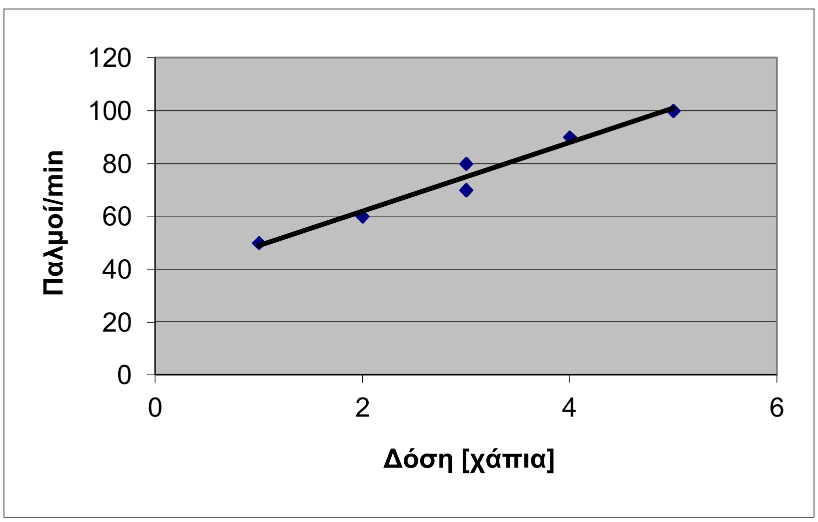
γ) Να υπολογισθούν η διασπορά των σφαλμάτων και ο συντελεστής συσχέτισης. Τι συμπεράσματα βγάζετε;

α/α	Δόση [χάπια]	Παλμοί/min
1	3	70
2	2	60
3	1	50
4	3	80
5	5	100
6	4	90
	3	**75**

α)

$$\left.\begin{array}{l}\beta = 13.00 \\ \\ \alpha = 36.00\end{array}\right\} \; y = 36 + 13x$$

β)



y(3.5)= 81.5

γ)

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \mathbf{0.983} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \mathbf{3.873}$$

Κριτήριο Wald-Wolfowitz ή των Ροών

Να εξετασθεί με την μέθοδο των "ροών" αν η ομολογία μεταξύ των ακολουθιών

M	Q	A	Q	I	T	G	R	P	E	W	I	W	L	A	L	G	T	A	L	M	Q	A	Q	I	T	G	R	P	E	W	I	W	L	A	L	G	T	A	L
|	|	|				|	|			|	|			|	|	|	|			|	|			|			|	|	|	|						|	|	|	|
M	Q	A	I	Q	I	G	R	R	W	W	I	A	D	A	L	G	T	T	A	M	Q	Q	I	I	G	L	R	P	E	W	W	L	A	L	A	G	T	A	L

E E E A A A E E A A E E A A E E E E A A E E A A E A A E E E E A A A A A E E E E

είναι τυχαία ή όχι με a=0.05.

n_1=22 n_2=18 > 10

$$\mu_u = \frac{2n_1 n_2}{n_1 + n_2} + 1 = \textbf{20.80}$$ αναμενόμενο πλήθος ροών

$$\sigma_u^2 = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} = \textbf{9.08} \qquad \sigma_u = \textbf{3.01} \qquad \textbf{u=15}$$

$$Z = \frac{U - \mu_u}{\sigma_u} = \textbf{-1.93}$$

Το |Z| που βρήκαμε είναι < 1.96 που αντιστοιχεί στο a/2=0.025 άρα το πλήθος των παρατηρούμενων ροών είναι τυχαίο. Δεν μπορούμε να αποφανθούμε ότι οι δύο ακολουθίες διαφέρουν σημαντικά ή μοιάζουν σημαντικά.

Έστω ότι έχουμε τα δύο παρακάτω δείγματα με τις συχνότητες f_1 και f_2.
Μπορούμε να ισχυριστούμε ότι τα δύο δείγματα είναι ομογενή σε σ.σ. 5%;

H_0: Τα δείγματα προέρχονται από την ίδια κατανομή (ομογενή)

ΔΟΚΙΜΑΣΙΑ ΟΜΟΓΕΝΕΙΑΣ (Kolmogorov – Smirnov)

x	f_1	f_2	Αθρ. f_1	Αθρ. f_2	$\|F_m(x)-G_n(x)\|$
0	5	6	0.0833	0.1000	0.0167
1	9	2	0.2333	0.1333	0.1000
2	4	5	0.3000	0.2167	0.0833
3	5	7	0.3833	0.3333	0.0500
4	7	8	0.5000	0.4667	0.0333
5	2	4	0.5333	0.5333	0.0000
6	8	4	0.6667	0.6000	0.0667
7	4	9	0.7333	0.7500	0.0167
8	9	6	0.8833	0.8500	0.0333
9	7	9	1.0000	1.0000	0.0000
	60	**60**			

$D_{60,60}$ = 0.1000

Από τον πίνακα VII για a=0,05

$D_{0,05,60,60}$ = 0.2483

$D_{60,60} < D_{0,05,60,60}$
άρα αποδεχόμαστε την μηδενική υπόθεση

Δύο δίαιτες Α και Β εφαρμόσθηκαν επί μια εβδομάδα σε 7 ποντικούς η πρώτη και σε 9 η δεύτερη και το βάρος σε gr που κέρδισε κάθε ποντικός δίνεται στον πίνακα που ακολουθεί. Υποθέτοντας ότι το βάρος που κερδήθηκε ακολουθεί την κανονική κατανομή

α) Να δοθεί 95% δ.ε. για τη διαφορά μ_A-μ_B των μέσων αυξήσεων σε βάρος κάθε εβδομάδα.

β) Χωρίς την υπόθεση της κανονικότητας ελέγξτε με την δοκιμασία Mann-Whitney αν τα δύο δείγματα είναι ομογενή (a=0.05).

γ) Τα συμπεράσματά σας συμφωνούν ή όχι;

α/α	A	B
1	78.10	79.10
2	72.40	81.00
3	76.20	77.30
4	74.30	79.10
5	77.40	80.00
6	78.40	79.10
7	76.00	79.10
8		77.30
9		80.20
	76.11	**79.13**

α)

$$\mu_1 - \mu_2 = \hat{x} - \hat{y} \pm t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Από τον πίνακα II για a/2=0.025 και β.ε.=7+9-2=14 βρίσκουμε t=2.145.

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \quad 0.504$$

s_1^2= 4.65

s_2^2= 1.51

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \quad 2.85918 \qquad s= 1.691$$

μ_1-μ_2 = 76.11 - 79.13 ± 2.145*1.691*0.504 = -3.02 ± 1.828

-4.848 < μ1 - μ2 < -1192

β)

α/α	A	B		Τάξη
1	78.10	79.10	72.40	1.0
2	72.40	81.00	74.30	2.0
3	76.20	77.30	76.00	3.0
4	74.30	79.10	76.20	4.0
5	77.40	80.00	77.30	5.5
6	78.40	79.10	77.30	5.5
7	76.00	79.10	77.40	7.0
8		77.30	78.10	8.0
9		80.20	78.40	9.0
10			79.10	11.5
11	**76.11**	**79.13**	79.10	11.5
12			79.10	11.5
13			79.10	11.5
14			80.00	14.0
15			80.20	15.0
16			81.00	16.0

T_A= 34 $\qquad N_1$=7, N_2=9

T_B= 102

$T_{min}\{T_A,T_B\}$=34

Ο πίνακας για Tmin=34, N1=7 και N2=9 δίνει περιοχή αποδοχής (40-79).
Το Tmin=34 δεν ανήκει στην περιοχή αποδοχής άρα τα δείγματα μη ομογενή.

Διεξήχθη μια σύγκριση με σκοπό να διαπιστωθεί αν υπάρχει διαφορά μεταξύ των συγκεντρώσεων του αζώτου ουρίας αίματος (BUN, blood urea nitrogen) σε 12 ασθενείς αποδοχείς νεφρικών μοσχευμάτων με σταθερή λειτουργία μοσχεύματος και σε 14 ασθενείς με εκτεταμένες μολύνσεις του ουροποιητικού (a=0.05).

Δοκιμασία Mann-Whitney

H_0: $M_A=M_B$

H_1: $M_A \neq M_B$

	BUN			
α/α	**Μεταμόσχευση (n_A=12)**	**UTI (n_B=14)**	**Τάξη του A**	**Τάξη του B**
1		150		1
2		170		2
3		180		3
4		190		4.5
5	190		4.5	
6		200		7
7		200		7
8	200		7	
9		210		9.5
10	210		9.5	
11		220		12
12	220		12	
13	220		12	
14		230		14
15		240		16.5
16		240		16.5
17	240		16.5	
18	240		16.5	
19	250		19	
20		260		20.5
21	260		20.5	
22	270		22	
23		280		23
24		290		24
25	310		25	
26	320		26	
		T =	**190.5**	**160.5**

Το μικρότερο άθροισμα τάξεων (T=160.5) αντιστοιχεί στο δείγμα UTI με n_B=14.

Αυτό το n θα είναι και το N_1 για τον πίνακα με τις περιοχές αποδοχής για το άθροισμα τάξεων T. Βρίσκουμε την περιοχή 150-228 στην οποία εμπίπτει το παρατηρούμενο T. Έτσι αποδεχόμαστε την μηδενική υπόθεση.

Προκειμένου να διαπιστωθεί αν υπάρχουν σημαντικές διαφορές μεταξύ των συγκεντρώσεων του Καλίου στο πλάσμα και στον ορό, ελήφθησαν δείγματα και των δύο τύπων από 18 εθελοντές.
Η μηδενική υπόθεση δηλώνει, ότι δεν υπάρχει διαφορά μεταξύ των διαμέσων.
Αν ισχύει, τότε το πλήθος των θετικών διαφορών θα πρέπει να είναι "ίσο" με το πλήθος των αρνητικών.

Δοκιμασία Προσήμου

	$[K_{πλ}]$/mM	$[K_{ορ}]$/mM	Πρόσ. Διαφ.
1	4.0	4.2	-
2	3.8	3.8	0
3	3.6	3.7	-
4	3.9	3.8	+
5	4.4	4.5	-
6	4.6	4.4	+
7	4.8	4.9	-
8	4.5	4.7	-
9	4.3	4.5	-
10	4.0	3.9	+
11	4.1	4.1	0
12	4.0	4.1	-
13	3.5	3.6	-
14	3.7	3.7	0
15	3.6	3.7	-
16	4.2	4.2	0
17	4.1	4.0	+
18	4.5	4.5	0
Median	**4.05**	**4.10**	

$N_+ = 4$ $N_- = 9$

N = πλήθος μη μηδενικών διαφορών = 18 - 5 = 13

Από τους πίνακες της διωνυμικής κατανομής βρίσκουμε, ότι η κρίσιμη περιοχή για μέγεθος δείγματος 13 και α=0.05 είναι 2-11.
Το $N_+ = 4 = \min(4, 9)$ εμπίπτει στο διάστημα 2-11, άρα οι διάμεσοι δεν διαφέρουν.

Η αρτηριακή πίεση δέκα ασθενών πριν και μετά τη χορήγηση φαρμάκου κατά της πίεσης δίνεται από τον πίνακα που ακολουθεί. Ελέγξτε αν υπάρχει στατιστικά σημαντική διαφορά στην αρτηριακή πίεση πριν και μετά την χορήγηση του φαρμάκου με στάθμη σημαντικότητας a=0.05 (δίπλευρη).
α) με την μέθοδο του προσήμου (sign test)
β) με την μέθοδο Wilcoxon

H_0: Τα δείγματα είναι ομογενή

Δοκιμασία Wilcoxon

α/α	Α(πριν)	Β(μετά)	Διαφορά	Απόλ.Τιμή			
1	12	11	1	1	1	3.0	3.0
2	13	13	0	0	1		
3	17	15	2	2	1		7.0
4	17	14	3	3	1		9.0
5	18	17	1	1	1		3.0
6	15	14	1	1	2	7.0	3.0
7	13	14	-1	1	2		3.0
8	12	10	2	2	2		7.0
9	13	14	-1	1	3	9.0	3.0
10	15	13	2	2			7.0

T_-= 6 **T_+= 39** **n=10-1=9**

T=min{6,39}=6

Για n=9 και a/2=0.025 ο πίνακας XI δίνει T_c=6. Έτσι το T =T_c και απορρίπτουμε οριακά την μηδενική υπόθεση.

Δοκιμασία Προσήμου

α/α	Α(πριν)	Β(μετά)	Πρόσ. Διαφ.
1	12	11	+
2	13	13	0
3	17	15	+
4	17	14	+
5	18	17	+
6	15	14	+
7	13	14	-
8	12	10	+
9	13	14	-
10	15	13	+
Median	**13.00**	**14.00**	

N_+ = 7
N_- = 2

Για N=9 (μη μηδενικές διαφορές) το διάστημα επιστοσύνης είναι 1-8, στο οποίο εντάσσεται το N_-=2. Άρα δεν απορρίπτεται η μηδενική υπόθεση.

Η διάρκεια της εγκυμοσύνης προσδιορίσθηκε σε δείγμα 10 εγκύων
με 2 διαφορετικές μεθόδους, την μέθοδο της τελευταίας περιόδου (LMP),
και του υπερηχογραφήματος (US). Δίνουν ίδια ή διαφορετική διάρκεια;

H_0: $M_{LMP} = M_{US}$
H_1: $M_{LMP} \neq M_{US}$
Αθροίσματα Τάξεων-Wilcoxon

	LMP[d]	US[d]	Διαφορά	Αύξ.σειρά απόλ.	Τάξη	Προσ.
1	275	273	2	1	1	-
2	292	285	7	2	2	+
3	281	270	11	4	3	+
4	284	272	12	7	5	+
5	285	278	7	7	5	+
6	283	276	7	7	5	+
7	290	291	-1	8	7	-
8	294	290	4	11	8	+
9	300	279	21	12	9	+
10	284	292	-8	21	10	+

T_+ = 2+5+8+9+5+5+3+10 = **47**

$$\mathbf{T = T_+ - T_- = 39}$$

T_- = 1+7 = **8**

Αλλιώς, **$\min\{T_+,T_-\}$=8** ίση με την τιμή του πίνακα XI για n=10 και δίπλευρη σ.σ. 0.05.
Η H_0 απορρίπτεται όταν $\min\{T_+,T_-\} \leq$ προς την τιμή του πίνακα XI.
Άρα απορρίπτουμε οριακά την H_0.

Από τέσσερα σημεία ενός θαλασσίου κόλπου πήραμε δείγματα νερού και μετρήσαμε την συγκέντρωση ενός είδους μικροβίων. Οι μετρήσεις έδωσαν τις τιμές του πίνακα που ακολουθεί. Να εξετασθεί μη παραμετρικά και σε σ.σ. 0.05 αν η μόλυνση είναι ομοιόμορφα κατανεμημένη στον κόλπο αυτόν.

H_0: Τα 4 δείγματα προέρχονται από τον ίδιο πληθυσμό
H_1: Τουλάχιστον 1 δείγμα δεν προέρχεται από τον ίδιο πληθυσμό

	κ=4, n=25							
	$n_A=6$	$n_B=6$	$n_\Gamma=6$	$n_\Delta=7$				
α/α	**A**	**B**	**Γ**	**Δ**	**A+B+Γ+Δ**	**Τιμή**	**μ**	**Τάξη**
1	4	2	8	7	2	2	2	1.5
2	8	2	10	8	2			
3	10	7	10	9	4	4	1	3
4	10	7	10	10	7			
5	10	9	10	10	7	7	3	5
6	12	10	15	15	7			
7				15	8			
8					8	8	3	8
9					8			
10					9	9	2	10.5
11					9			
12					10			
13					10			
14					10			
15					10			
16					10	10	10	16.5
17					10			
18					10			
19					10			
20					10			
21					10			
22					12	12	1	22
23					15			
24					15	15	3	24
25					15			

R_A= 82.5
R_B= 40
R_Γ= 98
R_Δ= 104.5

$$H=\frac{12}{n(n+1)}\cdot\sum_{j=1}^{\kappa}\frac{R_j^2}{n_j}-3(n+1)= \quad \mathbf{6.2168}$$

$$C=1-\frac{\left(\mu_1^3+\mu_2^3+\ldots+\mu_\rho^3\right)-\left(\mu_1+\mu_{21}+\ldots+\mu_\rho\right)}{n^3-n}= \quad \mathbf{0.9312}$$

***H'*= 6.6765**

Ο πίνακας ΙΙΙ για β.ε.=κ-1=3 και a=0.05 δίνει X_c=7.81473 δηλ. $H'=6.6765<X_c$.

Άρα αποδεχόμαστε την H_0.