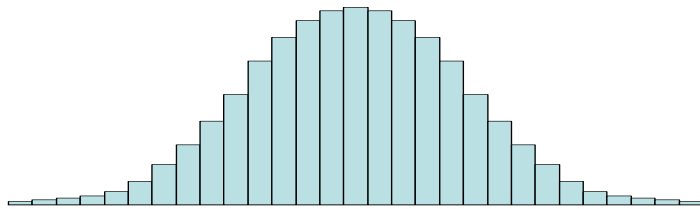


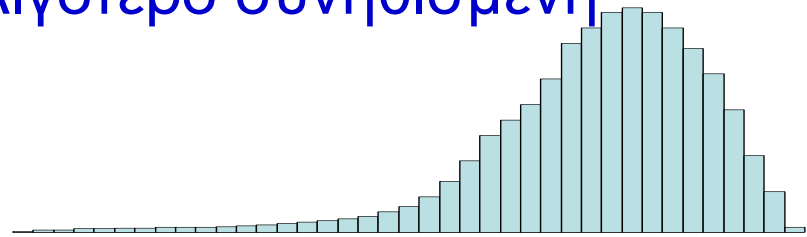
Παραδείγματα ερωτήσεων

Κατανομή Ποσοτικών Δεδομένων

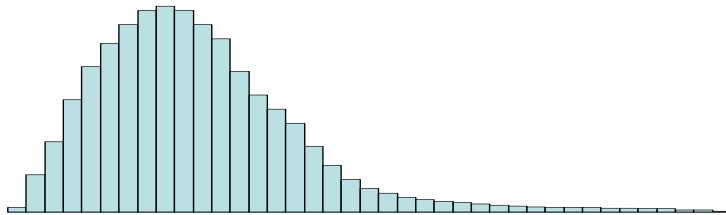
Συμμετρική, σχήμα καμπάνας
Κανονική (Gaussian) κατανομή



Αρνητικά λοξή
Λιγότερο συνηθισμένη



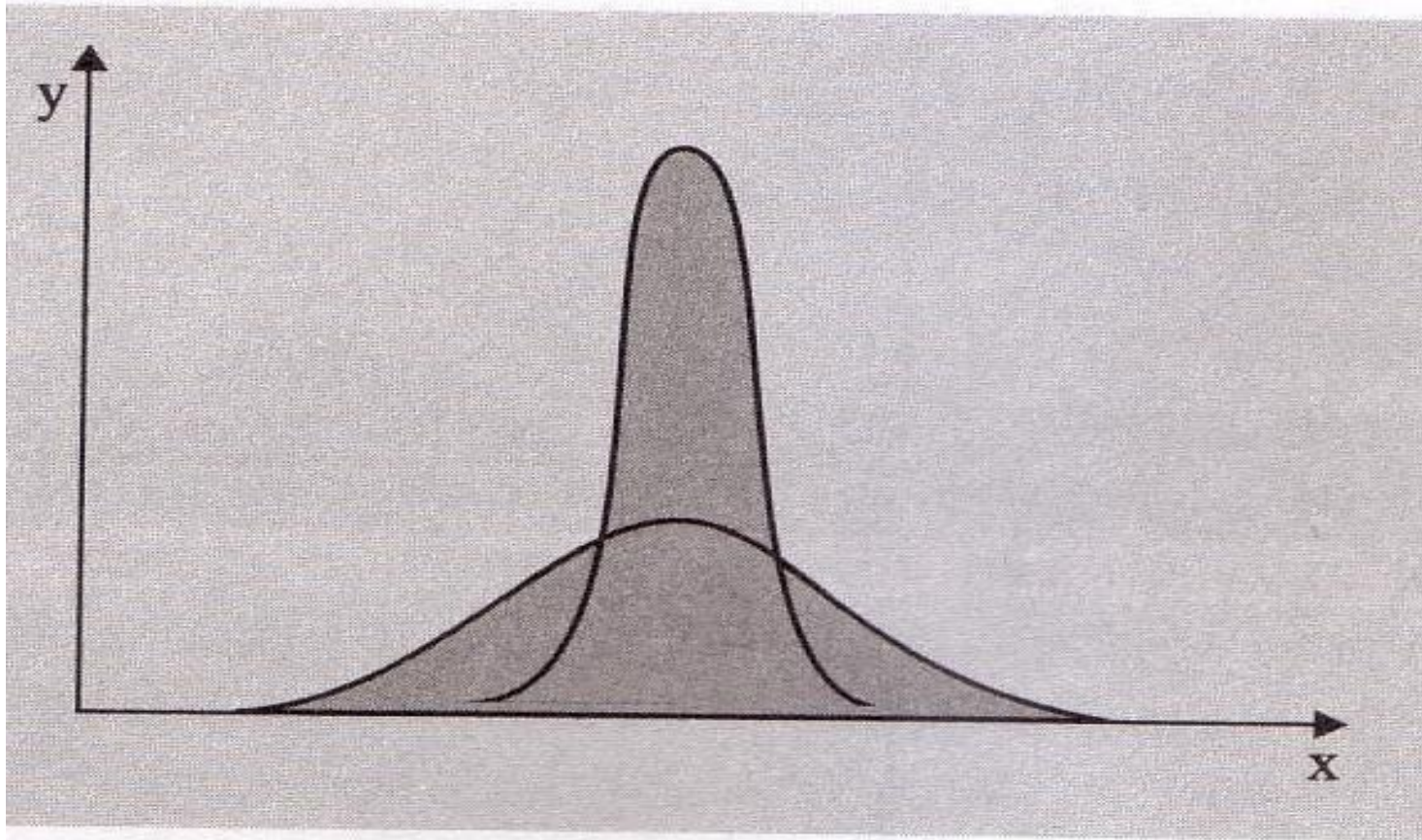
Θετικά λοξή
Σε εργαστηριακά δεδομένα
πχ. CD4 counts in HIV+



Ομοιόμορφη κατανομή
Ίδια πιθανότητα οποιασδήποτε τιμής
μέσα στο εύρος



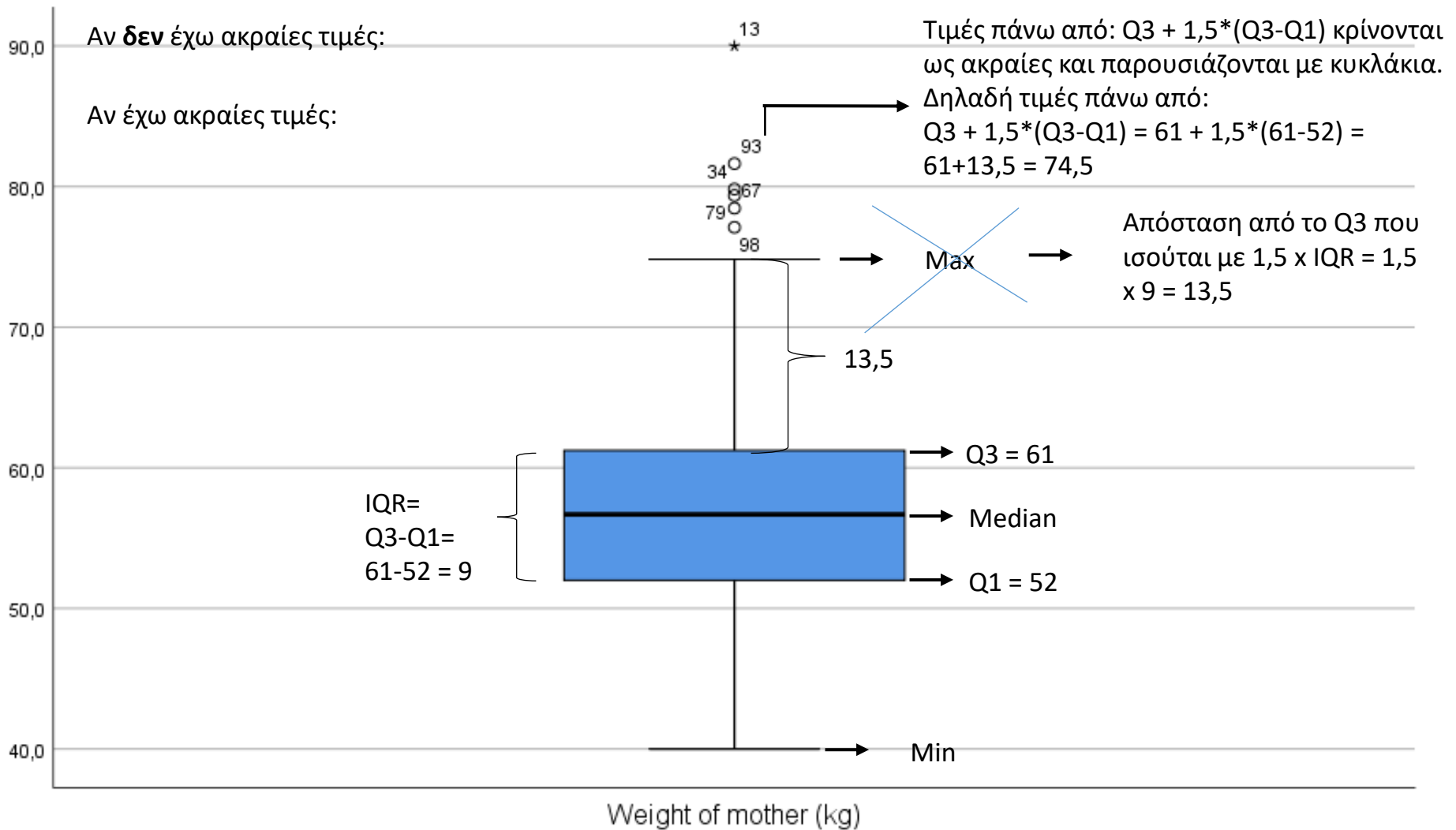
Αυτές οι κατανομές έχουν την ίδια περίπου θέση αλλά διαφορετική διασπορά



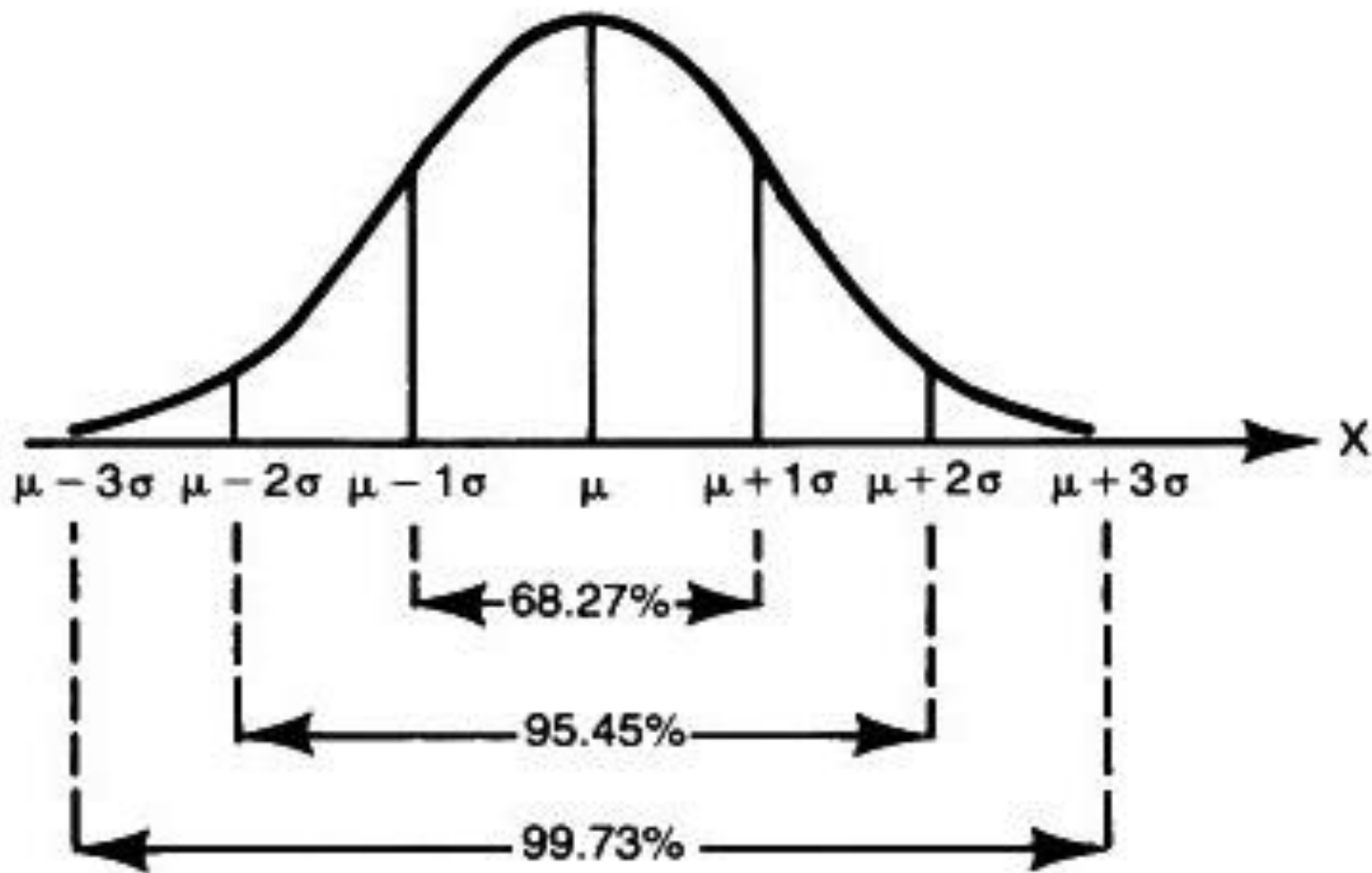
Αντιπροσωπευτικές τιμές

Υπολογίζονται με βάση τα στοιχεία της κατανομής και μπορούν να υποδείξουν τα κύρια χαρακτηριστικά της. Έγκυρες όταν τα δείγματα είναι αντιπροσωπευτικά ή τυχαία του πληθυσμού.

- Τιμές **θέσης**
Επικρατούσα τιμή, μέση τιμή, διάμεσος
- Τιμές βαθμού **διασποράς**
Σταθερή απόκλιση, ακραίες τιμές, εκατοστημόρια



Ιδιότητες Κανονικής Κατανομής (πώς η σταθερή απόκλιση δείχνει τον τρόπο διασποράς των παρατηρήσεων)

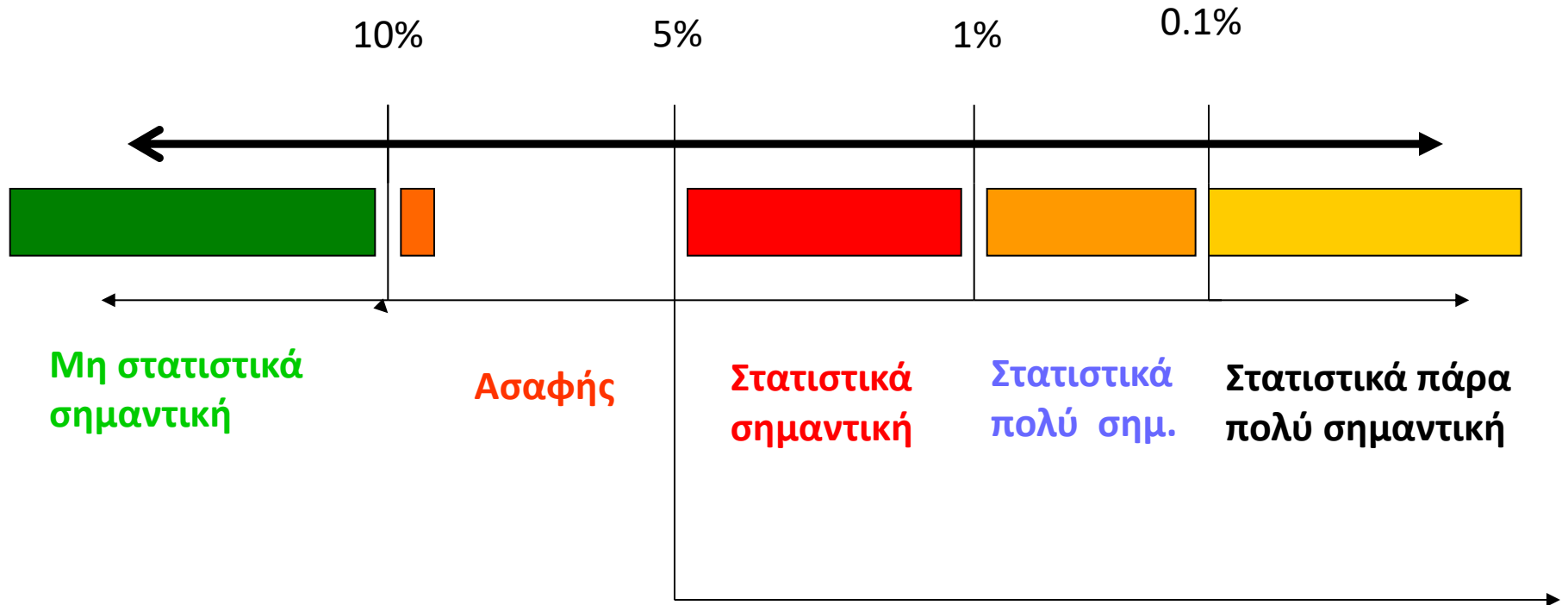


Πιθανό Σφάλμα (Standard Error)

Μέτρο της απόστασης της μέσης τιμής του δείγματος από την πραγματική στο πληθυσμό.

$$SE = \frac{SD}{\sqrt{n}}$$

P-value



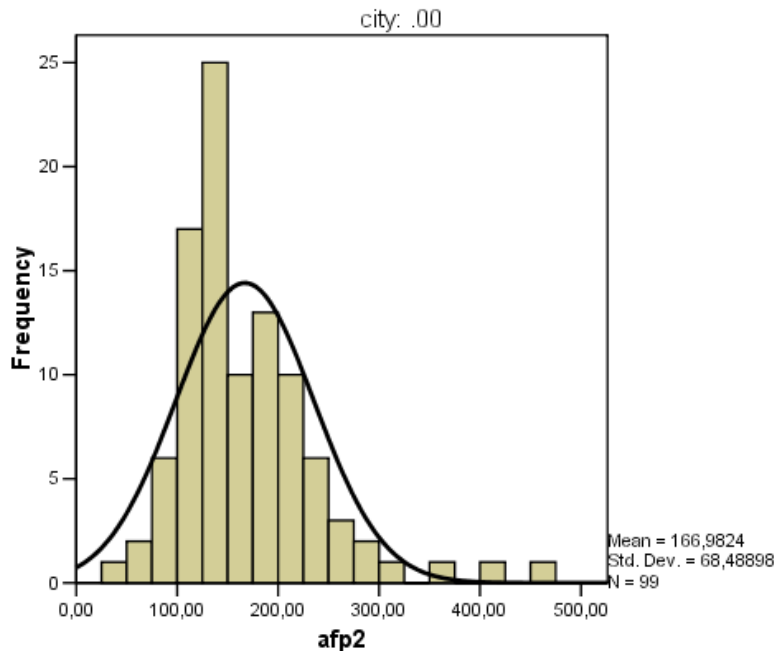
t-test δύο ανεξάρτητων δειγμάτων

Στο t-test για τη σύγκριση 2 μέσων τιμών αρχικά υποθέτουμε ότι δεν υπάρχει πραγματική διαφορά μεταξύ των μ_1 και μ_2 (H_0), δηλαδή $\Delta=0$.

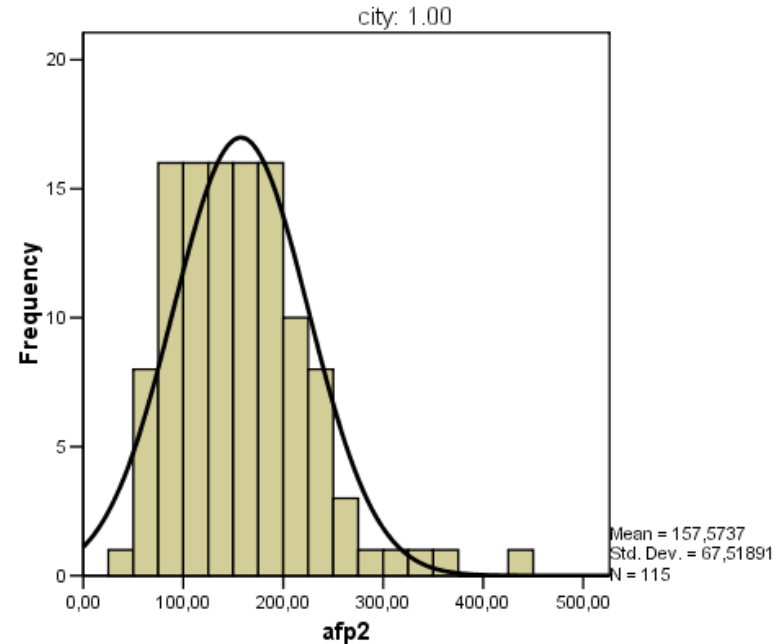
- ✓ Αν μετά τη στατιστική δοκιμασία καταλήξουμε σε πολύ μικρή p-value ($< 0,05$), απορρίπτουμε την H_0 και αποδεχόμαστε την H_A .
- ✓ Αν όχι (p-value $> 0,05$), τότε αποτυγχάνουμε να απορρίψουμε την H_0 .

Παράδειγμα 1: φετο-πρωτεΐνης

Histogram



Histogram



Ομάδα A: $n_1=99$ έγκυες γυναίκες στη Σαγκάη

$\bar{\chi}_1=167$ $\mu\text{mol/L}$, $SD_1=68.5$ $\mu\text{mol/L}$

Ομάδα B: $n_2=115$ έγκυες γυναίκες στη Βοστώνη

$\bar{\chi}_2=157,6$ $\mu\text{mol/L}$, $SD_2=67,5$ $\mu\text{mol/L}$

Παράδειγμα φετο-πρωτεΐνης

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
afp2	Equal variances assumed	.035	.853	1.010	212	.314	9.40877	9.31862	-8.96025	27.77780
	Equal variances not assumed			1.009	206.389	.314	9.40877	9.32861	-8.98281	27.80036

Τα στατιστικά πακέτα εκτιμάνε το 95% ΔΕ για τη διαφορά των δύο μέσων τιμών.

95% Δ.Ε: (-8,9 έως 27,8).

Είμαστε 95% σίγουροι ότι το διάστημα αυτό συμπεριλαμβάνει την πραγματική διαφορά στις μέσες τιμές ορμόνης, στους δύο διαφορετικούς πληθυσμούς.

Προσέξτε ότι το παραπάνω διάστημα περιλαμβάνει το 0, οπότε συμβαδίζει με τον έλεγχο υπόθεσης στο επίπεδο 5%.

Έλεγχος χ^2 και συγχυτικός παράγοντας

Ανάμεσα σε 100 άτομα με καρκίνο του πνεύμονα (cases) οι 61 είναι καπνιστές και πίνουν τουλάχιστον 1 φλιτζάνι καφέ την ημέρα, 7 είναι μη καπνιστές και πίνουν τουλάχιστον 1 φλιτζάνι καφέ την ημέρα, ενώ 17 δεν πίνουν καφέ και δεν καπνίζουν.

Σε 105 μη-καρκινοπαθείς (controls) οι 15 είναι καπνιστές και πίνουν τουλάχιστον 1 φλιτζάνι καφέ την ημέρα, 18 είναι μη καπνιστές και πίνουν τουλάχιστον 1 φλιτζάνι καφέ την ημέρα, ενώ 64 δεν πίνουν καφέ και δεν καπνίζουν.

Υπάρχει σχέση ανάμεσα στον καρκίνο του πνεύμονα και στην κατανάλωση καφέ;

Κατανάλωση καφέ και καρκίνος πνεύμονα

	≥ 1 φλ/ μέρα	Δεν πίνουν καφέ	Σύνολο
Καρκίνος	$61+7=68$	32	100
Υγιείς	$15+18=33$	72	105

$$X^2 = \frac{\left(|ad - bc| - \frac{n}{2} \right)^2 n}{(a+b)(c+d)(a+c)(b+d)} = 26$$

B.E.=1

Αξιολόγηση χ^2

BE	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83



26

10,83 < 26

0,1% > p

Συμπέρασμα: Υπάρχει στατιστικά σημαντική σχέση μεταξύ καρκίνου του πνεύμονα και καθημερινής κατανάλωσης καφέ

Διαστρωμάτωση σύμφωνα με τις καπνιστικές συνήθειες

Μη Καπνιστές

	>=1 φλ/ μέρα	Δεν πίνουν καφέ	Σύν.
Καρκ.	7	17	24
Υγιείς	18	64	82

$$\chi^2=0,2$$

$$BE=1$$

Καπνιστές

	>=1 φλ/ μέρα	Δεν πίνουν καφέ	Σύν.
Καρκίνος	61	15	76
Υγιείς	15	8	23

$$\chi^2=1,5$$

$$BE=1$$

Αξιολόγηση χ^2

BE	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83



$$0,2 < 2,71$$

$$1,5 < 2,71$$

$$p > 0,10$$

Συμπέρασμα: α) Δεν υπάρχει σχέση μεταξύ κατανάλωσης καφέ και καρκίνου του πνεύμονα

β) το κάπνισμα δρα ως συγχυτικός παράγοντας της σχέσης καφέ-καρκίνου πνεύμονα

	>=1 φλ/ μέρα	Δεν πίνουν καφέ	Σύνολο
Καρκίνος	68	32	100
Υγιείς	33	72	105

$$\text{Odds Ratio (}\Sigma\text{.}\Lambda\text{.)} = \frac{68 * 72}{32 * 33} = \frac{4896}{1056} = 4,6$$

Διαστρωμάτωση σύμφωνα με τις καπνιστικές συνήθειες

Μη Καπνιστές

	≥ 1 φλ/ μέρα	Δεν πίνουν καφέ	Σύν.
Καρκ.	7	17	24
Υγιείς	18	64	82

$\Sigma\Lambda=1,5$

Καπνιστές

	≥ 1 φλ/ μέρα	Δεν πίνουν καφέ	Σύν.
Καρκίνος	61	15	76
Υγιείς	15	8	23

$\Sigma\Lambda=2,2$

Αξιολόγηση εργαστηριακών ευρημάτων

Έστω ότι είναι γνωστό ότι 0,4% του πληθυσμού πάσχει από μια συγκεκριμένη νόσο.

Ένα διαγνωστικό τεστ για αυτή τη νόσο δίνει θετικό αποτέλεσμα στο 92% των πασχόντων και στο 1% των μη-πασχόντων.

Ποια είναι η ευαισθησία και η ειδικότητα αυτού του διαγνωστικού τεστ;

Να βρεθεί η πιθανότητα ένα άτομο επιλεγμένο τυχαία από τον πληθυσμό να έχει θετικό αποτέλεσμα στο τεστ αυτό;

Ποια είναι η θετική διαγνωστική αξία;

Ευαισθησία, ειδικότητα

Τι είναι το 0,4 %

Απάντηση: Επιπολασμός

Τι μας δίνει το 92%

Απάντηση: Ευαισθησία $P(T^+|D^+) = 0,92$

«Θετικό αποτέλεσμα στο 1% των μη-πασχόντων»

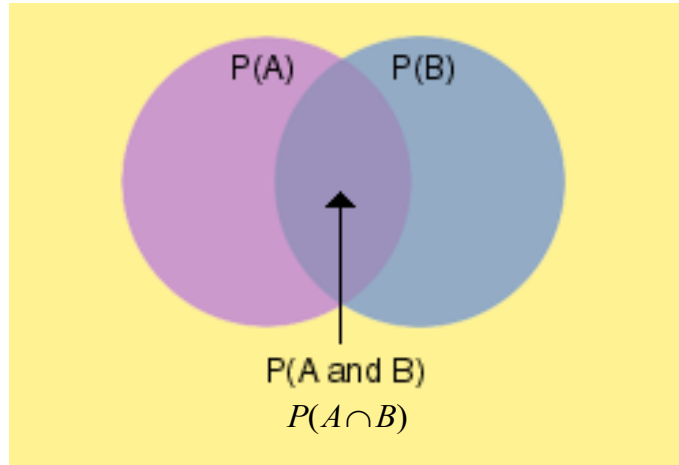
$$P(T^+|D^-) = 0,01$$

Ειδικότητα $P(T^-|D^-) = ?$

$$P(T^+|D^-) + P(T^-|D^-) = 1 \Rightarrow$$

$$P(T^-|D^-) = 1 - P(T^+|D^-) = 1 - 0,01 = 0,99$$

P(A | B)



- **P(A | B)**: η πιθανότητα να συμβεί το ενδεχόμενο A δεδομένου ότι έχει συμβεί το ενδεχόμενο B

→ μεταξύ αυτών που έχουν το B, πόσοι έχουν και το A;

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Σχετικά με την ερώτηση 1 (δεν είναι πλήρης απάντηση)

Ένα άτομο από τον πληθυσμό αν έχει θετικό τεστ, αυτό ή θα είναι ορθώς θετικό ή εσφαλμένα θετικό.

$$P(T^+ \cap D^+) + P(T^+ \cap D^-) = P(T^+ | D^+) \times P(D^+) + P(T^+ | D^-) \times P(D^-)$$

Η πιθανότητα ορθώς θετικού είναι: $P(\text{να είναι ασθενής και να έχει θετικό τεστ}) = P(T^+ \cap D^+) =$

$$P(T^+ / D^+) * P(D^+) = 0,92 \times 0,004 = 0,0037$$

Η πιθανότητα ψευδώς θετικού είναι: $P(\text{να είναι υγιής και να έχει θετικό τεστ}) = P(T^+ / D^-) * P(D^-)$

$$= P(T^+ / D^-) * (1 - P(D^+)) = 0,01 * 0,996 = 0,0100$$

Άρα η $P(T^+) = P(\text{ορθώς θετικού}) + P(\text{ψευδώς θετικού}) = 0,0037 + 0,01 = 0,0137$

Θετική Διαγνωστική Αξία

$$P(D^+|T^+) = \frac{P(D^+ \cap T^+)}{P(T^+)} = \frac{P(T^+|D^+) \times P(D^+)}{P(T^+)} =$$

$$\frac{P(T^+|D^+) \times P(D^+)}{P(T^+|D^+) \times P(D^+) + P(T^+|D^-) \times P(D^-)} =$$

$$\frac{0,92 * 0,004}{0,0137} = 0,269 =$$

$$\frac{\text{ευαισθησία} * \text{επιπολασμός}}{(\text{ευαισθησία} * \text{επιπολασμός}) * [(1 - \text{ειδικότητα}) * (1 - \text{επιπολασμός})]}$$

Μέτρα θέσης/διασποράς – Φυσιολογικές τιμές

Αν ένα εργαστηριακό μέγεθος δεν κατανέμεται κανονικά: ποια είναι τα καταλληλότερα μέτρα (τιμές) θέσης και διασποράς που περιγράφουν την κατανομή του;

Πώς υπολογίζονται και τι εκφράζουν τα 95% όρια αξιοπιστίας (i) μιας μέσης τιμής κανονικά κατανεμημένου ποσοτικού μεγέθους και (ii) μιας αναλογίας;

Σε 100 υγιή άτομα μετρήθηκε ένα κανονικά κατανεμημένο βιοχημικό μέγεθος. Η μέση τιμή ήταν 200 μονάδες και το πιθανό σφάλμα της μέσης τιμής 2 μονάδες. Ποια είναι τα όρια των φυσιολογικών τιμών του μεγέθους και τι εκφράζουν;

95% OA

$$\bar{X} \pm t_{(n-1),5\%} * SE(\bar{X})$$

$$p \pm 1,96 * SE(p),$$

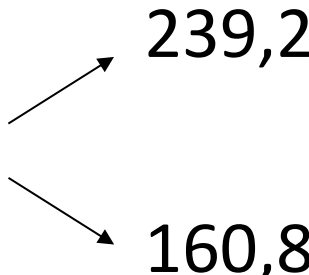
$$SE(p) = \sqrt{\frac{pq}{n}}, n * p > 5, n * q > 5$$

Φυσιολογικές τιμές

$N=100$, $\bar{X}=200$, $SE(\bar{X})=2$, $\Phi T=?$

$$\Phi T = \bar{X} \pm 1,96 * SD(X)$$

$$SE(\bar{X}) = \frac{SD(X)}{\sqrt{n}} \Rightarrow SD(X) = SE(\bar{X}) * \sqrt{n} = 2 * 10 = 20$$

$$\Phi T = 200 \pm 1,96 * 20 = 200 \pm 39,2$$


239,2

160,8

Για να διερευνηθούν οι επιβλαβείς επιδράσεις της δηλητηρίασης με DDT, χορηγήθηκε DDT σε 25 τυχαία επιλεγμένα ποντίκια από μία ομάδα 50 ποντικιών. Τα υπόλοιπα 25 ποντίκια χρησιμοποιήθηκαν ως ομάδα ελέγχου και δεν τους χορηγήθηκε DDT. Στον ακόλουθο Πίνακα δίνονται τα αποτελέσματα των μετρήσεων, με ειδική κλίμακα, της επιθετικότητας που παρουσιάζουν τα ποντίκια (η κατανομή είναι περίπου κανονική).

Ποια στατιστική μέθοδο θα εφαρμόσετε για να εξάγετε συμπέρασμα για τις επιδράσεις του παρασιτοκτόνου, όπως αυτές εκφράζονται με τη μέτρηση της επιθετικότητας;

	\bar{x}	s^2
Ομάδα DDT	17,6	5,3
Ομάδα Ελέγχου	12,8	4,5

- T-test (τυπολόγιο)
- $BE = (n_1 - 1) + (n_2 - 1)$
- Αξιολόγηση στην t κατανομή

- ΠΡΟΣΟΧΗ: $SD = \sqrt{S^2}$
- $SD_1 = 2,30$ $SD_2 = 2,12$

- Προϋποθέσεις



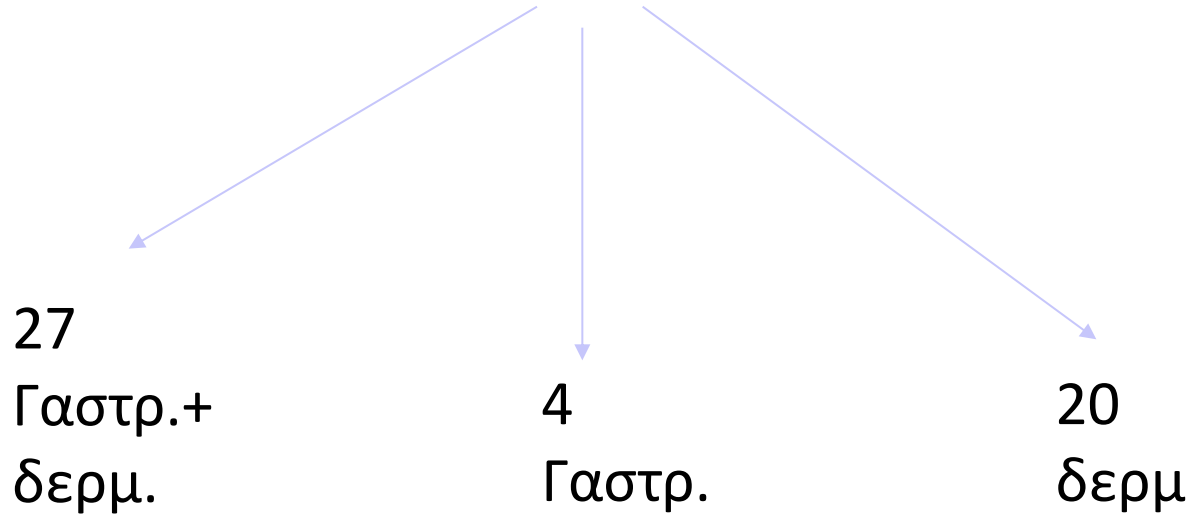
- Αν ένα μέγεθος κατανέμεται κατά προσέγγιση κανονικά, ενδείκνυται η εφαρμογή μιας μη παραμετρικής στατιστικής δοκιμασίας για τον έλεγχο της διαφοράς της θέσης του μεγέθους αυτού στις 2 ομάδες; Δικαιολογείστε.
- Ισχύς διαδικασίας

Σε μια έρευνα για τους κινδύνους μικροβιακών λοιμώξεων κατά την κολύμβηση στη θάλασσα, που έγινε στις Η.Π.Α., τυχαιοποιήθηκαν 1303 άτομα σε 2 ομάδες. Τα 652 άτομα της 1^{ης} ομάδας (κολυμβητές) κολύπησαν με τυποποιημένο τρόπο, την ίδια μέρα, ενώ τα 651 άτομα της 2^{ης} ομάδας (μη-κολυμβητές) κάθισαν στην παραλία. Κατά τη διάρκεια των 7 επόμενων ημερών, κατά την οποία τα άτομα δεν ξανα-κολύπησαν, 27 από τους κολυμβητές έπαθαν γαστρεντερίτιδα και δερματική νόσο, 4 μόνο γαστρεντερίτιδα και 20 μόνο δερματική νόσο. Από τους μη-κολυμβητές 1 έπαθε γαστρεντερίτιδα και δερματική νόσο, 17 μόνο γαστρεντερίτιδα και 8 μόνο δερματική νόσο.

A) Υπάρχει διαφορά στη συχνότητα περιστατικών γαστρεντερίτιδας ανάμεσα στους κολυμβητές και στους μη-κολυμβητές; Ποιες αναλογίες συγκρίνουμε; Καταστρώστε το κατάλληλο πίνακα.

B) Αν θέλατε να διερευνήσετε αν οι κολυμβητές που έπαθαν γαστρεντερίτιδα είχαν μεγαλύτερη πιθανότητα να πάθουν και δερματική νόσο σε σχέση με τους κολυμβητές που δεν έπαθαν γαστρεντερίτιδα, τι πίνακα θα καταστρώνατε;

Κολυμβητές N=652



Μη-Κολυμβητές N=651



Α ερώτημα

Υπάρχει διαφορά στη συχνότητα περιστατικών γαστρεντερίτιδας ανάμεσα στους κολυμβητές και στους μη-κολυμβητές; Ποιες αναλογίες συγκρίνουμε; Καταστρώστε το κατάλληλο πίνακα.

Α ερώτημα

Υπάρχει διαφορά στη συχνότητα περιστατικών γαστρεντερίτιδας ανάμεσα στους κολυμβητές και στους μη-κολυμβητές; Ποιες αναλογίες συγκρίνουμε; Καταστρώστε το κατάλληλο πίνακα.

Γαστρεντερίτιδα	Κολυμβητές	Μη-Κολυμβητές	Σύνολο
Ναι	31 $31/652 = 4,75\%$	18 $18/651 = 2,76\%$	49
Όχι	621	633	1254
Σύνολο	652	651	1303

Β ερώτημα

Αν θέλατε να διερευνήσετε αν οι κολυμβητές που έπαθαν γαστρεντερίτιδα είχαν μεγαλύτερη πιθανότητα να πάθουν και δερματική νόσο σε σχέση με τους κολυμβητές που δεν έπαθαν γαστρεντερίτιδα, τι πίνακα θα καταστρώνατε;

Β ερώτημα

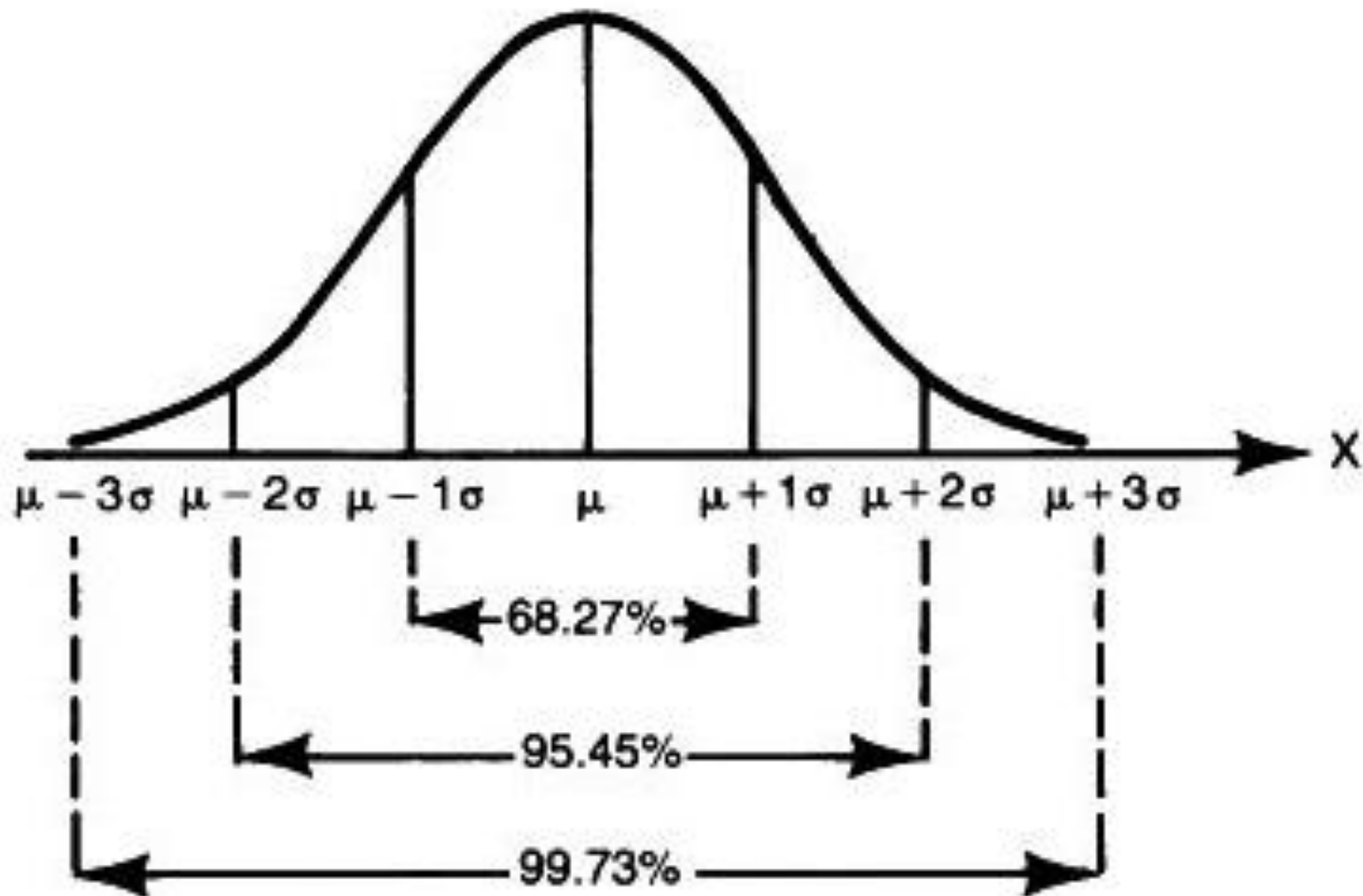
Αν θέλατε να διερευνήσετε αν οι κολυμβητές που έπαθαν γαστρεντερίτιδα είχαν μεγαλύτερη πιθανότητα να πάθουν και δερματική νόσο σε σχέση με τους κολυμβητές που δεν έπαθαν γαστρεντερίτιδα, τι πίνακα θα καταστρώνατε;

Γαστρ. Δερματ.	Ναι	Οχι	Σύνολο
Ναι	27	20	47
Οχι	4	601	605
Σύνολο	31	621	652

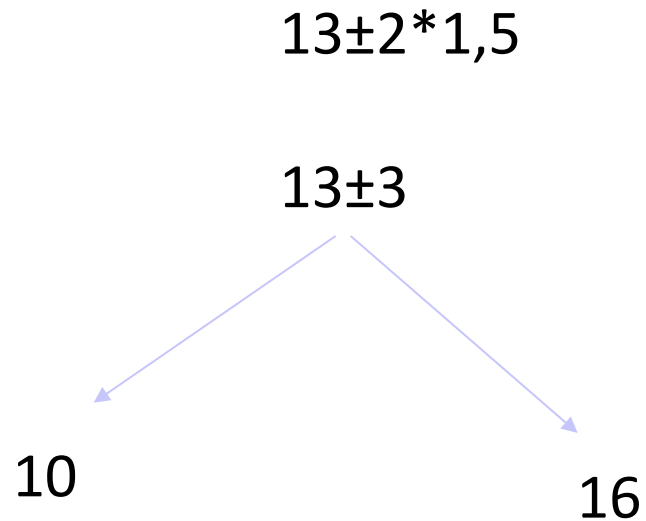
Ο παρακάτω Πίνακας παρουσιάζει την κατανομή της ηλικίας εμμηναρχής από έρευνα που έγινε σε αντιπροσωπευτικό δείγμα 200 γυναικών από ένα πληθυσμό. Ακολουθεί η ηλικία εμμηναρχής περίπου κανονική κατανομή; Δικαιολογήστε. Τι ποσοστό γυναικών έχουν ηλικία εμμηναρχής μεγαλύτερη από 16 έτη; Σε ποια ηλικία έχει το 50% των γυναικών ήδη εμμηναρχή;

Χαρακτηριστικά της κατανομής της ηλικίας εμμηναρχής αντιπροσωπευτικού δείγματος πληθυσμού	
Αριθμός έγκυρων παρατηρήσεων	200
Μέση τιμή	13 έτη
Διάμεση τιμή	12,9 έτη
Σταθερή απόκλιση (μέγιστη & ελάχιστη τιμή)	1,5 έτη (17,5 και 8)

Ιδιότητες Κανονικής Κατανομής



$\bar{X} \pm 2SD \rightarrow 95\% \text{ παρατηρήσεων}$



>16 έτη το ποσοστό 2,5%

Μέχρι τα 12,9 έχει εμμηναρχή (διάμεσος)

Σε μία ομάδα 180 ατόμων μελετήθηκε η σχέση των τιμών της κρεατινίνης του ορού (mmol/liter) με το φύλο (0=άνδρας, 1= γυναίκα), την ηλικία (έτη) και τη λειτουργία του νεφρού (0 = φυσιολογική, 1= παθολογική). Στον Πίνακα που ακολουθεί παρουσιάζονται τα αποτελέσματα πολλαπλής γραμμικής εξάρτησης.

<u>Μεταβλητή</u>	<u>b</u>	<u>SE_b</u>	<u>t</u>	<u>p</u>
Φύλο	-3,80	2,58	1,47	>0,10
Ηλικία	0,120	0,025	4,80	<0,001
Λειτουργία νεφρού	1,88	0,92	2,04	<0,05

Ερμηνεύστε τους συντελεστές μερικής εξάρτησης του μοντέλου. Είναι στατιστικά σημαντικοί;

Για τη διερεύνηση της πιθανής σχέσης του βάρους γέννησης (gr) με τις καπνισματικές συνήθειες της μητέρας, μελετήθηκαν 81 νεογνά και οι μητέρες τους. Τα αποτελέσματα της εξάρτησης του βάρους γέννησης από ορισμένους πιθανούς προγνωστικούς παράγοντες (μεταβλητές) ήταν:

Μεταβλητή	Συντελεστής εξάρτησης (b)	Πιθανό σφάλμα SE (b)
Έτη σχολικής φοίτησης	23,8	9,5
Βάρος μητέρας πριν τον τοκετό (kg)	2,4	4,7
Κάπνισμα (0: όχι, 1: ναι)	-190,8	80,4

Ποιοι από τους παραπάνω παράγοντες επηρεάζουν το βάρος γέννησης του παιδιού;

Ερμηνεύστε τους παραπάνω συντελεστές εξάρτησης.

Μεταβλητή	Συντελεστής εξάρτησης (b)	Πιθανό σφάλμα SE (b)	T-test b/SE(b)
Έτη σχολικής φοίτησης	23,8	9,5	2,51
Βάρος μητέρας πριν τον τοκετό (Kg)	2,4	4,7	0,51
Κάπνισμα (0: όχι, 1: ναι)	-190,8	80,4	-2,36

$$BE = n - (p + 1) = 81 - (3 + 1) = 77$$

Σε μια μελέτη ασθενών-μαρτύρων που έγινε στον Καναδά, σε 10.000 ασθενείς με καρκίνο της κεφαλής και του τραχήλου και σε 15.000 υγιείς, εξετάστηκαν ως πιθανοί προγνωστικοί παράγοντες του νοσήματος ο δείκτης μάζας σώματος (που υπολογίζεται ως βάρος/ανάστημα², με μονάδες μέτρησης kg/m², και χρησιμοποιήθηκε αφού δημιουργήθηκαν 4 κατηγορίες) κατά την διάγνωση της νόσου ή την ένταξη στη μελέτη, και η ηλικία.

Στον Πίνακα φαίνονται τα αποτελέσματα μοντέλου λογαριθμιστικής ανάλυσης με εξαρτημένη μεταβλητή την παρουσία της νόσου (ναι/όχι).

Μεταβλητή	Σχετικός λόγος	95% ΔΕ
Ηλικία (πενταετία)	1,02	1,01 - 1,04
Δείκτης μάζας σώματος (<18,5 kg/m²)	2,13	1,75 – 2,58
Δείκτης μάζας σώματος (18,5-24,9 kg/m²)	1,00	-
Δείκτης μάζας σώματος (25,0-29,9 kg/m²)	0,52	0,44 - 0,60
Δείκτης μάζας σώματος (≥30,0 kg/m²)	0,43	0,33 - 0,57

α) Είναι στατιστικά σημαντική η σχέση των παραπάνω μεταβλητών με το νόσημα που μελετήθηκε; Εξηγήστε.

β) Ερμηνεύστε το σχετικό λόγο για την ηλικία και για τις κατηγορίες του Δείκτη Μάζας Σώματος.

γ) Διατυπώστε σύντομα το βασικό συμπέρασμα από τα αποτελέσματα που παρουσιάζονται.

Αξιολόγηση μαθήματος «Ιατρική Στατιστική» :

- Ανώνυμα
- Αξιολόγηση (1) Μαθήματος & (2) Διδασκόντων/διδασκουσών
- Θα βρείτε τα ερωτηματολόγια στο eclass του μαθήματος (ενότητα «Ερωτηματολόγια») ή μπορείτε να χρησιμοποιήσετε τα παρακάτω QR codes

Αξιολόγηση μαθήματος



Αξιολόγηση διδασκόντων/ουσών



Διαθέσιμα από 13/1/2025-10/2/2025