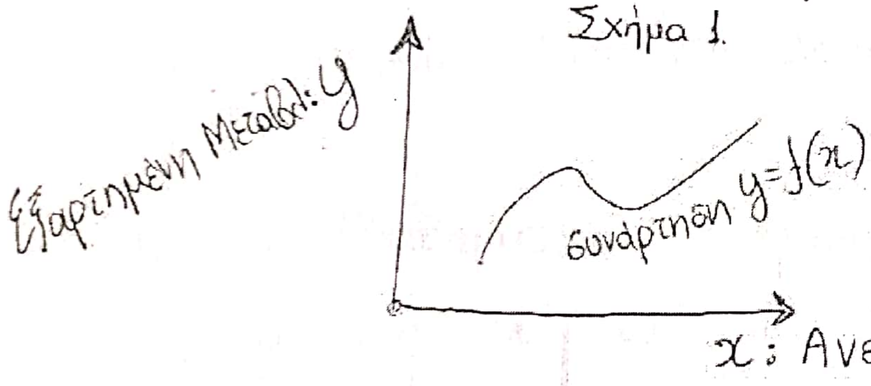
Τρίτη, 19/02/2013

2ο Μάθημα "ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ"

1. Στά **Μαθηματικά** η σχέση μεταξύ τής ανεξάρτητης μεταβλητής x καί τής εξαρτημένης $y(=f(x))$ δίνεται από τήν συνάρτηση $f(x)$ (Σχ. 1)

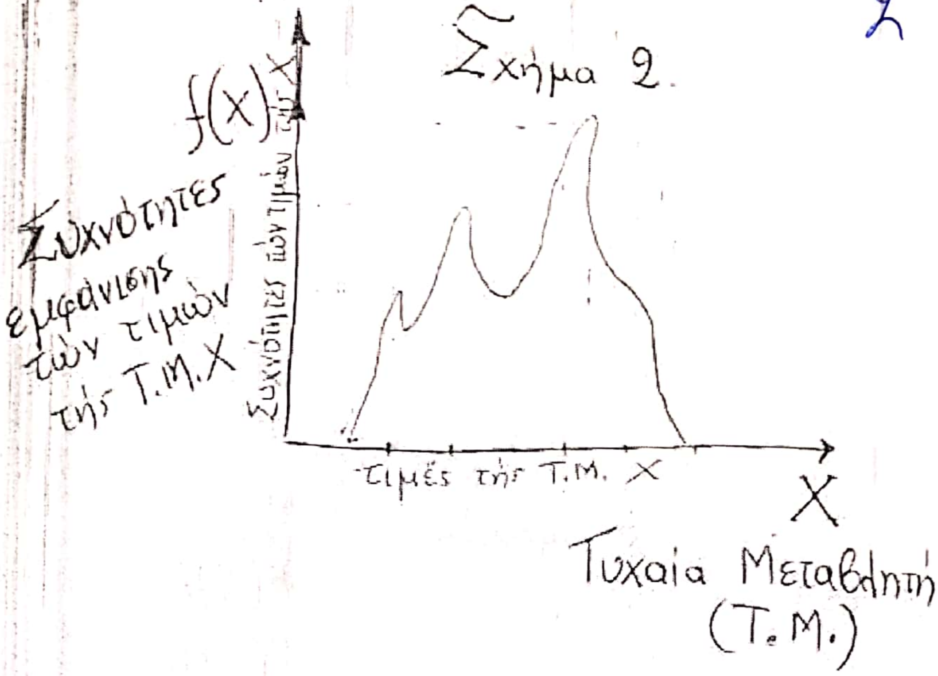
Σχήμα 1.



2. Στή **Στατιστική** η ανεξάρτητη μεταβλητή ονομάζεται καί ΕΙΝΑΙ **Τυχαία Μεταβλητή** καί συμβολίζεται μέ X ενώ η εξαρτημένη μεταβλητή εκφράζει τόν αριθμό εμφανίσεων (στό δείγμα ή πληθυσμό) κάθε τιμής τής ανεξάρτητης μεταβλητής X, δηλ. τις συχνότητες (εμφάνισης) κάθε τιμής τής Α. μεταβλητής X

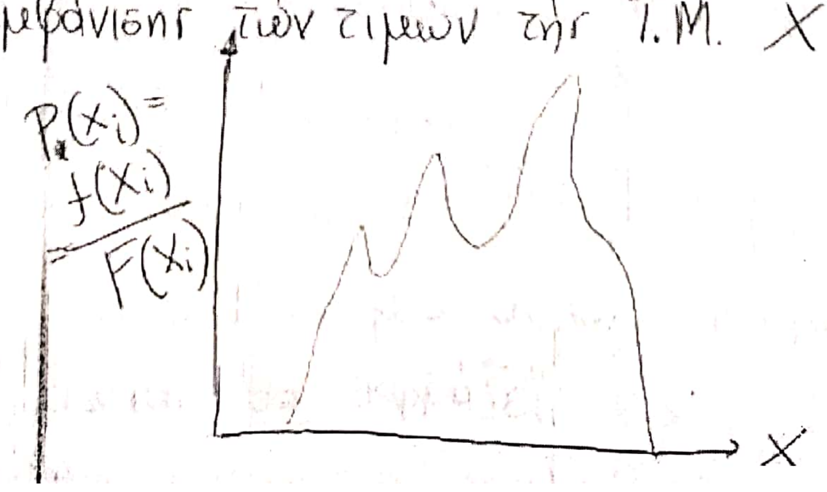
Η εξαρτημένη μεταβλητή $f(x)$ ΔΕΝ ΠΑΙΡΝΕΙ ΑΡΝΗΤΙΚΕΣ ΤΙΜΕΣ.

Συχνά, τό γράφημα των συχνοτήτων τών τιμών τής Α.Μ. X λέγεται Ιστόγραμμα Συχνοτήτων ή Κατανομή Σχετικών Συχνοτήτων.



Σχήμα 2.

3. Στις Πιθανότητες η ανεξάρτητη μεταβλητή εξακολουθεί να είναι μία μεταβλητή πού παίρνει τυχαίες τιμές άρα μία ΤΥΧΑΙΑ ΜΕΤΑΒΛΗΤΗ (Τ.Μ.)
(Αμα ο οριζόντιος άξονας τού σχ. 2 παραμένει αμετάβλητος)
Εάν κάθε μία από τίς συχνότητες $f_i(X)$ $i=1,2,\dots n$ την διαιρέσουμε μέ τό άθροισμα όλων τών (σχετικών) συχνοτήτων $F(X)=\sum_{i=1}^{n} f_i(x)$ (n σύνολο τιμών τής Τ.Μ.Χ) τότε τό Σχ. 2 θά υποστεί μία αλλαγή κλίμακας και ο κάθετος άξονας θά εκφράζει τίς ΠΙΘΑΝΟΤΗΤΕΣ εμφάνισης τών τιμών τής Τ.Μ. Χ



Τρίτη 19/02/2013

2ο ΜΑΘΗΜΑ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

⊙ Η Τ.Μ. X μπορεί νά είναι διακριτή (discrete) εάν παίρνει ένα πεπερασμένο πλήθος τιμών ή συνεχής άν παίρνει τιμές σέ ένα διάστημα (α,β) μί-ωcαcβcτα

⊙ Η συνάρτηση $F_X(x) = P(X \le x)$ $\forall x \in R$ ονομάζεται συνάρτηση αθροιστικής κατανομής και είναι ίση μέ τήν πιθανότητα η Τ.Μ. X νά πάρει όλες τις τιμές (μικρότερες) μέχρι τήν τιμή x

⊙ Γιά διακριτή Τ.Μ. X, συνάρτηση πιθανότητας της Τ.Μ.Χ. ή Κατανομή πιθανότητας λέγεται η συνάρτηση πού δίνει τήν πιθανότητα η Τ.Μ. X νά πάρει τήν τιμή X. και συμβολίζεται μέ $f_x(x) = P(X=x)$

ΚΥΡΙΩΤΕΡΕΣ ΚΑΤΑΝΟΜΕΣ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

ΔΙΑΚΡΙΤΕΣ

1) Bernoulli $B(1,p)$

Τ.Μ X δίτιμη (0,1) ή (αποτυχία, επιτυχία) μέ πιθανότητες $(1-p,\ p)$

$$P(X=x) = p^x(1-p)^{1-x},\quad x=0,1 \quad \text{και } 0<p<1$$

2) Διωνυμική $B(n,p)$

Τ.Μ.X πλήθος επιτυχιών σε n δοκιμές Bernolli

$$P(X=x) = \binom{n}{x} p^x (1-p)^{1-x} \quad x=0,1,\dots,n \quad 0<p<1$$

3) Poisson $P(\lambda)$ Των σπανίων γεγονότων

$$P(X=x) = e^{-\lambda}\cdot\frac{\lambda^{x}}{x!} \qquad x=0,1,2,\dots, \quad \lambda>0$$

παράμετρος

4) Γεωμετρική (Pascal)

5) Αρνητική Διωνυμική (Polya)

6) Υπεργεωμετρική

7) Πολυωνυμική

ΣΥΝΕΧΕΙΣ

1) Ομοιόμορφη $U(\alpha,\beta)$

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \alpha\le x\le \beta\\ 0 & \text{αλλού.}\end{cases}$$

2) Κανονική $N(\mu,\sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}, \quad -\infty<x<+\infty$$

ΠΟΛΥ ΣΗΜΑΝΤΙΚΗ διότι

1) Πολλά φυσικά φαινόμενα ακολουθούν κανονική κατανομή

2) Για μεγάλα δείγματα ισχύουν τα θεωρήματα σύγκλισης και πολλές κατανομές συγκλίνουν στην κανονική κατανομ

3) Η τυποποιημένη (ή τυπική) κανονική κατανομή στην οποία μετασχηματίζεται εύκολα οποιαδήποτε κανονική κατα είναι αναλυτικά τυποποιημένη σε πίνακες τιμών της Τ.Μ.,

Τρίτη 19/02/2013

2ο ΜΑΘΗΜΑ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

- Οι πίνακες τής τυπικής κανονικής κατανομής $N(0,1)$ μάς δίνουν τίς πιθανότητες γιά οποιοδήποτε διάστημα τιμών τής τυπικής Τ.Μ. $\cancel{Z}$.
- Η Τ.Μ. Z προκύπτει από οποιαδήποτε Τ.Μ. X μέ τόν μετασχηματισμό $Z=\frac{X-\mu}{\sigma}$

καί η $Z \xrightarrow[\text{τήν κατανομή}]{\text{ακολουθεί}} N(0,1)$

δηλ $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$

3) Εκθετική $E(\lambda)$

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases} \qquad \lambda>0.$$

4) Student t_v με v βαθμούς ελευθερίας

$$f(x)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}} \qquad -\infty<x<\infty$$

5) Χι-τετράγωνο με v βαθμούς ελευθερίας χ^2_v

$$f(x)=\left(\frac{1}{2}\right)^{v/2}\frac{1}{\Gamma\left(\frac{v}{2}\right)}x^{\frac{v}{2}-1}\cdot e^{-\frac{x}{2}} \qquad x>0.$$

⊙ Γνωρίζοντας τίς παραμέτρους ενός πληθυσμού μπορούμε νά υπολογίσουμε τήν πιθανότητα η Τ.Μ. X (τής οποίας γνωρίζουμε τήν κατανομή πιθανότητας) νά παίρνει τιμές εντός η εκτός ενός διαστήματος

⊙ Όμως δέν γνωρίζουμε τίς παραμέτρους του πληθυσμού ΑΛΛΑ γνωρίζουμε τίς παραμέτρους του δείγματος

ΑΠΟΔΕΙΚΝΥΕΤΑΙ ΟΤΙ

* Οι παράμετροι τού πληθυσμού συνδέονται μέ τίς παραμέτρους τού δείγματος *

1) Η μέση τιμή τής δειγματικής μέσης τιμής $\overline{X}$ (σαν Τ.Μ.) ισούται μέ τήν μέση τιμή τού πληθυσμού.

$$\mu_{\overline{X}} = E(\overline{X}) = \mu.$$

2) Η διασπορά τής δειγματικής μέσης τιμής $\overline{X}$ (σαν Τ.Μ.)

$$\sigma_{\overline{X}}^2 = E\left[(\overline{X}-\mu)^2\right] = \frac{\sigma^2}{n}$$

3) Η μέση τιμή τής δειγματικής διασποράς S^2 (σαν Τ.Μ.)

$$S^2 = \frac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + \dots + (x_n-\overline{x})^2}{n}$$

$$E(S^2) = \mu_{S^2} = \frac{n-1}{n}\sigma^2$$

καί εάν $n \geq 30$

$$E(S^2) \approx \sigma^2$$

Τρίτη 19/02/2013

2ο ΜΑΘΗΜΑ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

⊙ Άρα σύμφωνα μέ τά 1) καί 2)

η Τ.Μ. $\bar{X}$ (δειγματική μέση τιμή) ακολουθεί τήν κανονική κατανομή μέ μέση τιμή μ καί (τυπική απόκλιση $\frac{\sigma}{\sqrt{n}}$) διασπορά $\frac{\sigma^2}{n}$ δηλ. $\bar{X} \rightsquigarrow N\left(\mu, \frac{\sigma^2}{n}\right)$ *

καί άρα η $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightsquigarrow N(0,1)$

όταν • η διασπορά τού πληθυσμού γνωστή

ή • δείγμα μεγάλο $n \geq 30$ καί μπορούμε νά ισχυριστούμε ότι $\sigma = S$

⊙ Εάν θ είναι μία άγνωστη παράμετρος τού πληθυσμού καί $\theta_1 < \theta < \theta_2$

καί α μία μικρή πιθανότητα τότε

ορίζεται $P(\theta_1 < \theta < \theta_2) = 1-\alpha$

καί τό διάστημα $100(1-\alpha)\%$ λέγεται διάστημα εμπιστοσύνης γιά τήν παράμετρο θ καί η πιθανότητα $1-\alpha$ συντελεστής εμπιστοσύνης.