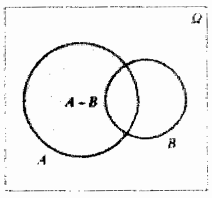
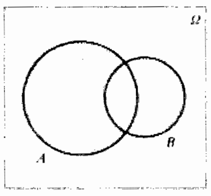
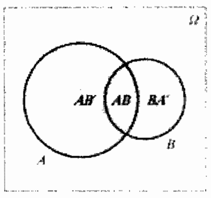
3.4. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Πείραμα τύχης - Η έννοια του τυχαίου	Σε μια εκτέλεση ενός πειράματος τύχης το αποτέλεσμα δεν καθορίζεται με βάση την αρχή της αιτιότητας (όπως στα αιτιοκρατικά φαινόμενα και πειράματα) αλλά αποδίδεται στην *τύχη*. Η έννοια του *τυχαίου* συνδέεται με το *πολυσύνθετο* και το *περιορισμένο* της γνώσης των αιτίων που προκαλούν το αποτέλεσμα. Χαρακτηριστικό ενός πειράματος τύχης είναι ότι μπορεί να επαναληφθεί υπό τις ίδιες συνθήκες όσες φορές θέλουμε (θεωρητικά άπειρες φορές) και ότι σε μια εκτέλεσή του δε μπορούμε να προβλέψουμε *με βεβαιότητα* το αποτέλεσμα που θα εμφανισθεί, όμως μπορούμε να καταγράψουμε όλα τα *δυνατά* αποτελέσματά του.
Δειγματικός χώρος Ω ενός πειράματος τύχης	Το σύνολο των δυνατών αποτελεσμάτων του.
Δειγματικό σημείο	Κάθε στοιχείο ενός δειγματικού χώρου, δηλαδή κάθε δυνατό αποτέλεσμα ενός πειράματος τύχης.
Ενδεχόμενα του δειγματικού χώρου ενός πειράματος τύχης	Υποσύνολα του δειγματικού χώρου.
Πραγματοποίηση ενδεχομένου	Σε μια εκτέλεση ενός πειράματος τύχης, ένα ενδεχόμενο πραγματοποιείται (εμφανίζεται) όταν το αποτέλεσμα του πειράματος είναι στοιχείο του.
Απλό και σύνθετο ενδεχόμενο	Αν ένα ενδεχόμενο αποτελείται από μόνο ένα δειγματικό σημείο ονομάζεται απλό ενδεχόμενο ενώ αν αποτελείται από περισσότερα από ένα δειγματικά σημεία ονομάζεται σύνθετο ενδεχόμενο.
Βέβαιο ενδεχόμενο, Ω	Πραγματοποιείται πάντα.
Αδύνατο ενδεχόμενο, $\varnothing$	Δεν πραγματοποιείται ποτέ.
Το ενδεχόμενο A συνεπάγεται το ενδεχόμενο B, $A \subseteq B$	Όταν πραγματοποιείται το A πραγματοποιείται και το B.
Ίσα ενδεχόμενα $A = B$	Όταν πραγματοποιείται το A πραγματοποιείται και το B και αντιστρόφως.
Τομή ενδεχομένων $A \cap B$ ή AB	Πραγματοποιείται όταν πραγματοποιηθεί και το A και το B.
Ένωση ενδεχομένων $A \cup B$	Πραγματοποιείται όταν πραγματοποιηθεί ή το A ή το B (ή και τα δύο), ή αλλιώς, όταν πραγματοποιηθεί τουλάχιστον ένα από τα A, B.
Συμπλήρωμα A' ενδεχομένου A	Πραγματοποιείται όταν δεν πραγματοποιηθεί το A.
Ξένα ή ασυμβίβαστα ή αμοιβαίως αποκλειόμενα ενδεχόμενα A, B	Ενδεχόμενα τα οποία δεν έχουν κοινά δειγματικά σημεία ($AB = \varnothing$) ή αλλιώς, η πραγματοποίηση του ενός αποκλείει την πραγματοποίηση του άλλου.
Διαφορά $A - B$ ή AB'	Πραγματοποιείται όταν πραγματοποιηθεί το A αλλά όχι το B. 
Συμμετρική διαφορά $A \Delta B = AB' \cup BA'$	Πραγματοποιείται όταν πραγματοποιηθεί ακριβώς ένα από τα A, B. 

$(A \cup B)'$	Πραγματοποιείται όταν δεν πραγματοποιείται ούτε το A ούτε το B.
$(AB)'$	Πραγματοποιείται όταν από τα A, B πραγματοποιηθεί το πολύ ένα.
Ιδιότητες των πράξεων μεταξύ ενδεχομένων	$A \cup \varnothing = A$, $A\varnothing = \varnothing$ $A \cup A = A$, $AA = A$ $A \cup \Omega = \Omega$, $A\Omega = A$ $A \cup A' = \Omega$, $AA' = \varnothing$, $(A')' = A$, $\Omega' = \varnothing$, $\varnothing' = \Omega$ $A \cup B = B \cup A$, $AB = BA$ $A \cup (B \cup \Gamma) = (A \cup B) \cup \Gamma$, $A(B\Gamma) = (AB)\Gamma$ $A \cup (B\Gamma) = (A \cup B)(A \cup \Gamma)$, $A(B \cup \Gamma) = (AB) \cup (A\Gamma)$ *Αν* $A \subseteq B$ *τότε* $AB = A$, $A \cup B = B$ και $A - B = AB' = \varnothing$ Τύποι De Morgan: $(A \cup B)' = A'B'$, $(AB)' = A' \cup B'$ $(A_1 \cup A_2 \cup \ldots \cup A_\nu)' = A_1'A_2' \ldots A_\nu'$ $(A_1A_2 \ldots A_\nu)' = A_1' \cup A_2' \cup \ldots \cup A_\nu'$
Η ένωση $A \cup B$ ως ένωση ξένων ενδεχομένων	$A \cup B = AB' \cup B = A \cup BA' = AB' \cup AB \cup BA'$  *Επίσης* $A = AB' \cup AB$ και $B = BA \cup BA'$
Ο κλασικός ορισμός της πιθανότητας *(Laplace, 1812)*	Αν ο Ω είναι πεπερασμένος και όλα τα απλά ενδεχόμενά του είναι ισοπίθανα, τότε $P(A) = \dfrac{N(A)}{N(\Omega)} = \dfrac{\text{πλήθος στοιχείων του A}}{\text{πλήθος στοιχείων του } \Omega}$
Ο στατιστικός ορισμός της πιθανότητας *(Richard von Mises, 1919)*	$P(A) = \lim\limits_{\nu \to +\infty} \dfrac{\nu_A}{\nu}$, όπου ν_A ο αριθμός εμφανίσεων του ενδεχομένου Α σε ν επαλήψεις του πειράματος
Ο αξιωματικός ορισμός της πιθανότητας *(Kolmogorov, 1933)*	1. $P(A) \geq 0$, $\forall$ ενδεχόμενο Α του Ω 2. $P(\Omega) = 1$ 3. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$, για $A_1, A_2, \ldots$ ξένα ανά δύο ενδεχόμενα.
Άλλες ιδιότητες της πιθανότητας *(προκύπτουν από τα τρία αξιώματα)*	(α) $P(\varnothing) = 0$ (β) $P(A_1 \cup A_2 \cup \ldots \cup A_\nu) = P(A_1) + P(A_2) + \ldots + P(A_\nu)$ για $A_1, A_2, \ldots, A_\nu$ ξένα ανά δύο ενδεχόμενα (γ) Αν $A = \{a_1, a_2, \ldots\}$, τότε $P(A) = P(\{a_1\}) + P(\{a_2\}) + \ldots$ (δ) $P(A) \leq 1$ (ε) $P(A') = 1 - P(A)$ (στ) $P(AB') = P(A) - P(AB)$ (ζ) Αν $A \subseteq B$ τότε $P(A) \leq P(B)$ (η) $P(A \cup B) = P(A) + P(B) - P(AB)$ (θ) $P(A \cup B) \leq P(A) + P(B)$, για οποιαδήποτε ενδεχόμενα A, B (ι) $P(A_1 \cup A_2 \cup \ldots \cup A_\nu) \leq P(A_1) + P(A_2) + \ldots + P(A_\nu)$ για οποιαδήποτε ν ενδεχόμενα $A_1, A_2, \ldots, A_\nu$.

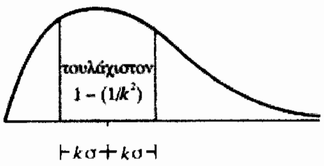
4.6. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Δεσμευμένη πιθανότητα του A δοθέντος του B	$P(A\|B) = \dfrac{P(AB)}{P(B)}$, $P(B) > 0$ (δηλαδή $P(B) \neq 0$).
Ιδιότητες της δεσμευμένης πιθανότητας	1. $P(A\|B) \geq 0$ 2. $P(\Omega\|B) = 1$ 3. $P(A_1 \cup A_2 \cup \ldots \| B) = P(A_1\|B) + P(A_2\|B) + \ldots$ για $A_1, A_2, \ldots$ ξένα ανά δύο ενδεχόμενα.
Άλλες ιδιότητες της δεσμευμένης πιθανότητας	$P(\varnothing\|B) = 0$ $P(A'\|B) = 1 - P(A\|B)$ $P(A\Gamma'\|B) = P(A\|B) - P(A\Gamma\|B)$ Αν $\Gamma \subseteq A$ τότε $P(\Gamma\|B) \leq P(A\|B)$ $P(A \cup \Gamma\|B) = P(A\|B) + P(\Gamma\|B) - P(A\Gamma\|B)$ Όταν $B \subseteq A$ τότε $P(A\|B) = 1$.
Ο πολλαπλασιαστικός τύπος	$P(AB) = P(A)P(B\|A) = P(B)P(A\|B)$ αν $P(A) > 0$, $P(B) > 0$. Γενικότερα: αν $P(A_1A_2 \ldots A_{\nu-1}) > 0$, τότε $P(A_1A_2 \ldots A_\nu) = P(A_1)P(A_2\|A_1) \ldots P(A_\nu\|A_1A_2 \ldots A_{\nu-1})$.
Το θεώρημα ολικής πιθανότητας	Αν $\{B_1, B_2, \ldots, B_\nu\}$ μια διαμέριση του Ω με $P(B_i) > 0$, $i = 1, 2, \ldots, \nu$, τότε $P(A) = P(A\|B_1)P(B_1) + P(A\|B_2)P(B_2) + \ldots + P(A\|B_\nu)P(B_\nu)$.
Ο τύπος του Bayes	Αν $\{B_1, B_2, \ldots, B_\nu\}$ μια διαμέριση του Ω με $P(B_i) > 0$, $i = 1, 2, \ldots, \nu$, τότε $P(B_i\|A) = \dfrac{P(A\|B_i)P(B_i)}{P(A)}$, $i = 1, 2, \ldots, \nu$ όπου $P(A)$ υπολογίζεται από το θεώρημα ολικής πιθανότητας.
Ανεξάρτητα ενδεχόμενα A, B	$P(AB) = P(A)P(B)$. Αν A, B ανεξάρτητα και $P(A) > 0$, $P(B) > 0$ τότε $P(A\|B) = P(A)$ και $P(B\|A) = P(B)$.
Εξαρτημένα ενδεχόμενα A, B	$P(AB) \neq P(A)P(B)$.
Ανεξάρτητα ενδεχόμενα A, B, Γ	$P(AB) = P(A)P(B)$, $P(A\Gamma) = P(A)P(\Gamma)$ $P(B\Gamma) = P(B)P(\Gamma)$ και $P(AB\Gamma) = P(A)P(B)P(\Gamma)$.
Ανεξαρτησία και συμπληρωματικά ενδεχόμενα	Αν A, B ανεξάρτητα τότε είναι ανεξάρτητα και τα ζεύγη α) A, B' β) A', B γ) A', B'.
Ανεξάρτητα ενδεχόμενα και ξένα ενδεχομένα	Αν A, B **ξένα** (με $P(A) > 0$ και $P(B) > 0$) τότε $P(A\|B) = 0 \neq P(A)$ και $P(B\|A) = 0 \neq P(B)$ συνεπώς τα Α, Β είναι **εξαρτημένα**. Αν A, B **ξένα** τότε: $P(AB) = 0$ και $P(A \cup B) = P(A) + P(B)$. Αν A, B **ανεξάρτητα** τότε: $P(AB) = P(A)P(B)$ και $P(A \cup B) = P(A) + P(B) - P(A)P(B)$.
Ανεξάρτητα πειράματα $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\nu$ με δειγματικούς χώρους, αντίστοιχα $\Omega_1, \Omega_2, \ldots, \Omega_\nu$	Τα πειράματα $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\nu$ είναι ανεξάρτητα αν η έκβαση οποιουδήποτε από αυτά, καθώς και η έκβαση οποιασδήποτε ομάδας από αυτά, δεν επηρεάζει την έκβαση των υπολοίπων. Για ανεξάρτητα πειράματα $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\nu$ ισχύει ότι αν $A_1 \subseteq \Omega_1, \ldots, A_\nu \subseteq \Omega_\nu$ τότε $P(A_1A_2 \ldots A_\nu) = P(A_1)P(A_2) \ldots P(A_\nu)$. *(Το $A_1A_2 \ldots A_\nu$ πραγματοποιείται όταν κατά την εκτέλεση του πειράματος ε_1 πραγματοποιηθεί το A_1, ... και κατά την εκτέλεση του πειράματος ε_ν πραγματοποιηθεί το A_ν).*

5.6. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Τυχαία μεταβλητή X στο δειγματικό χώρο Ω	Μια πραγματική συνάρτηση που αντιστοιχίζει τα στοιχεία του δειγματικού χώρου στο σύνολο των πραγματικών αριθμών τέτοια ώστε για κάθε διάστημα πραγματικών αριθμών I, το σύνολο $\{\omega \in \Omega: X(\omega) \in I\}$ να είναι ενδεχόμενο του Ω. Το σύνολο τιμών μιας τυχαίας μεταβλητής X συμβολίζεται με R_X.
Συνάρτηση κατανομής F τυχαίας μεταβλητής X	Η πραγματική συνάρτηση με τύπο $F(x) = P(X \leq x) = P(\{\omega \in \Omega: X(\omega) \leq x\}), -\infty < x < +\infty$
Ιδιότητες της συνάρτησης κατανομής F	α) $0 \leq F(x) \leq 1$, για κάθε $x \in \mathbb{R}$. β) Είναι αύξουσα συνάρτηση στο $\mathbb{R}$. γ) $\lim_{x \to +\infty} F(x) = 1$ και $\lim_{x \to -\infty} F(x) = 0$. δ) Είναι δεξιά συνεχής.
Υπολογισμός πιθανοτήτων μέσω της F	$P(X \leq b) = F(b)$ και $P(X > b) = 1 - F(b)$ $P(X < b) = F(b-)$ και $P(X \geq b) = 1 - F(b-)$ $P(a < X \leq b) = F(b) - F(a)$ $P(a < X < b) = F(b-) - F(a)$ $P(a \leq X < b) = F(b-) - F(a-)$ $P(a \leq X \leq b) = F(b) - F(a-)$ $P(X = a) = F(a) - F(a-)$.
Διακριτή τυχαία μεταβλητή	Το σύνολο τιμών της είναι πεπερασμένο ή αριθμησίμως άπειρο.
Συνάρτηση πιθανότητας f διακριτής τ.μ. X με σύνολο τιμών R_X	$f: \mathbb{R} \to [0, 1]$ με $f(x) = \begin{cases} 0, & \text{για κάθε } x \notin R_X \\ P(X = x), & \text{για κάθε } x \in R_X. \end{cases}$
Ιδιότητες της συνάρτησης πιθανότητας f μιας διακριτής τ.μ. X με $R_X = \{x_1, x_2, ..., x_\nu, ...\}$	α) $f(x) = 0$ για κάθε $x \notin R_X$ β) $f(x_i) \geq 0$ για κάθε $i = 1, 2, ...$ γ) $f(x_1) + f(x_2) + ... + f(x_\nu) + ... = 1$.
Σχέση συνάρτησης πιθανότητας και συνάρτησης κατανομής μιας διακριτής τ.μ. X με $R_X = \{x_1, x_2, ..., x_\nu, ...\}$ και $x_1 < x_2 < ... < x_\nu < ...$	$F(x) = P(X \leq x) = \sum_{i: x_i \leq x} f(x_i), x \in \mathbb{R}$ $f(x_1) = F(x_1)$ και $f(x_k) = F(x_k) - F(x_{k-1}), k = 2, 3, ...$
Πιθανότητα υποσυνόλου A του $R_X = \{x_1, x_2, ..., x_\nu, ...\}$	$P(X \in A) = \sum_{i: x_i \in A} f(x_i)$.
Παράμετροι θέσης μιας διακριτής τ.μ. X με $R_X = \{x_1, x_2, ..., x_\nu, ...\}$	α) Μέση τιμή, $\mu = E(X)$ $\mu = E(X) = \sum_{x \in R_X} x f(x)$ β) p-ποσοστιαίο σημείο, x_p, $0 < p < 1$ $P(X < x_p) \leq p \leq P(X \leq x_p)$ ή $F(x_p-) \leq p \leq F(x_p)$ γ) διάμεσος, $\delta = x_{0.5}$ δ) πρώτο, δεύτερο και τρίτο τεταρτημόριο $x_{0.25}, x_{0.5}, x_{0.75}$, αντίστοιχα.
Μέση τιμή συνάρτησης διακριτής τ.μ.	$E[g(X)] = \sum_{x \in R_X} g(x) f(x)$

Συνεχής τ.μ. και συνάρτηση πυκνότητας f συνεχούς τ.μ.	Μια τ.μ. X λέγεται συνεχής αν υπάρχει μια μη αρνητική πραγματική συνάρτηση f τέτοια ώστε $P(X \in A) = \int_A f(x)dx$ για κάθε (μετρήσιμο) υποσύνολο A των πραγματικών. Η συνάρτηση f ονομάζεται συνάρτηση πυκνότητας της X.
Ιδιότητες της συνάρτησης πυκνότητας f συνεχούς τ.μ.	α) $f(x) \geq 0$, για κάθε $x \in (-\infty, +\infty)$. β) $\int_{-\infty}^{+\infty} f(x)dx = 1$.
Σχέση συνάρτησης πυκνότητας και συνάρτησης κατανομής συνεχούς τ.μ.	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$ $f(x) = F'(x)$ (στα σημεία συνέχειας της f).
Υπολογισμός πιθανοτήτων για συνεχείς τ.μ.	$P(a < X < \beta) = P(a \leq X < \beta) = P(a < X \leq \beta) = P(a \leq X \leq \beta) = \int_a^{\beta} f(x)dx = F(\beta) - F(a)$ $P(X = x) = 0$ για κάθε $x \in (-\infty, +\infty)$.
Παράμετροι θέσης συνεχούς τ.μ.	α) Μέση τιμή, $\mu = E(X)$ $\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$ β) p-ποσοστιαίο σημείο, x_p, $0 < p < 1$ $\int_{-\infty}^{x_p} f(x)dx = p$ ή $F(x_p) = p$ γ) διάμεσος, $\delta = x_{0.5}$ δ) πρώτο, δεύτερο και τρίτο τεταρτημόριο $x_{0.25}$, $x_{0.5}$, $x_{0.75}$, αντίστοιχα.
Μέση τιμή συνάρτησης συνεχούς τ.μ.	$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$
Ιδιότητες της μέσης τιμής τ.μ. *(διακριτής ή συνεχούς)*	α) $E[a_1g_1(X) + a_2g_2(X) + \dots + a_kg_k(X)] = a_1E[g_1(X)] + a_2E[g_2(X)] + \dots + a_kE[g_k(X)]$ β) $E(aX + \beta) = aE(X) + \beta$ γ) $E(a_1X_1 + a_2X_2 + \dots + a_kX_k) = a_1E(X_1) + a_2E(X_2) + \dots + a_kE(X_k)$.
Παράμετροι διασποράς τ.μ. *(διακριτής ή συνεχούς)*	*α) Διακύμανση,* $\sigma^2 = Var(X)$ $\sigma^2 = Var(X) = E[(X-\mu)^2]$ *β) Τυπική απόκλιση,* $\sigma = \sqrt{Var(X)}$ $\sigma = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}$ *γ) Ενδοτεταρτημοριακό πλάτος ή εύρος* $x_{0.75} - x_{0.25}$
Εναλλακτικός τύπος υπολογισμού της διακύμανσης τ.μ. *(διακριτής ή συνεχούς)*	$\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$
Ιδιότητες της διακύμανσης τ.μ. *(διακριτής ή συνεχούς)*	α) $\sigma^2_{aX+\beta} = Var(aX + \beta) = a^2Var(X)$ και $\sigma_{aX+\beta} = \|a\|\sigma_X$ β) Αν $X_1, X_2, \dots, X_k$ *ανεξάρτητες* τ.μ. $Var(a_1X_1 + a_2X_2 + a_kX_k) = a_1^2Var(X_1) + a_2^2Var(X_2) + \dots + a_k^2Var(X_k)$

Ροπές τ.μ. *(διακριτής ή συνεχούς)*	α) Ροπή r τάξης γύρω από το 0: $\mu'_r = E(X^r)$, $r = 1, 2, \ldots$ β) Κεντρική ροπή r τάξης: $\mu_r = E[(X-\mu)^r]$, $r = 1, 2, \ldots$ γ) Παραγοντική ροπή r τάξης: $\mu_{(r)} = E[X(X-1)\ldots(X-r+1)]$, $r = 1, 2, \ldots$
Μέτρα λοξότητας και κύρτωσης τ.μ. *(διακριτής ή συνεχούς)*	α) Συντελεστής λοξότητας $\beta = \dfrac{\mu_3}{(\sigma^2)^{3/2}}$. β) Συντελεστής κύρτωσης $\gamma = \dfrac{\mu_4}{\sigma^4}$.
Τυποποιημένη τ.μ. *(διακριτή ή συνεχής)*	$Z = (X-\mu)/\sigma$, όπου $\sigma^2 = Var(X)$ και $\mu = E(X)$. $E(Z) = 0$ και $Var(Z) = 1$.
Ανισότητα Chebyshev *(ισχύει τόσο για διακριτές όσο και για συνεχείς τ.μ.)*	$P(\lvert X-\mu\rvert \geq c) \leq \dfrac{\sigma^2}{c^2}$ όπου $\sigma^2 = Var(X)$, $\mu = E(X)$ και $c > 0$. Για $c = k\sigma$ παίρνουμε $P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \dfrac{1}{k^2}$. 
Ομοιόμορφη κατανομή	α) Συνεχής $f(x) = \begin{cases} \dfrac{1}{\beta-\alpha}, & \alpha \leq x \leq \beta \\ 0, & \text{αλλού} \end{cases}$ (σε υποδιαστήματα του ίδιου πλάτους αντιστοιχούν ίσες πιθανότητες) $\mu = E(X) = (\alpha+\beta)/2$ και $\sigma^2 = Var(X) = (\beta-\alpha)^2/12$ β) Διακριτή με $R_X = \{x_1, x_2, \ldots, x_\nu, \ldots\}$ $f(x) = P(X = x) = \begin{cases} \dfrac{1}{\nu}, & x = x_1, x_2, \ldots, x_\nu \\ 0, & x \neq x_1, x_2, \ldots, x_\nu \end{cases}$ (οι δυνατές τιμές $x_1, x_2, \ldots, x_\nu$ είναι ισοπίθανες) $\mu = E(X) = \dfrac{x_1 + x_2 + \ldots + x_\nu}{\nu}$, $\sigma^2 = Var(X) = \dfrac{x_1^2 + x_2^2 + \ldots + x_\nu^2}{\nu} - \left(\dfrac{x_1 + x_2 + \ldots + x_\nu}{\nu}\right)^2$.
Ανεξάρτητες τ.μ. X, Y	$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$ για οποιαδήποτε (μετρήσιμα) υποσύνολα A, B των πραγματικών.
Ανεξάρτητες τ.μ. $X_1, X_2, \ldots, X_\nu$	$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_\nu \in A_\nu) = P(X_1 \in A_1)\, P(X_2 \in A_2) \ldots P(X_\nu \in A_\nu)$ για οποιαδήποτε (μετρήσιμα) υποσύνολα $A_1, A_2, \ldots, A_\nu$ των πραγματικών.
Τυχαίο δείγμα μεγέθους ν	ν ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_\nu$ οι οποίες ακολουθούν την ίδια κατανομή.
Δειγματικός μέσος	$\bar{X} = \dfrac{1}{\nu} \sum_{i=1}^{\nu} X_i$
Μέση τιμή και Διακύμανση δειγματικού μέσου	$E(\bar{X}) = \mu_{\bar{x}} = \mu$, $Var(\bar{X}) = \sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{\nu}$.
Στατιστικός πληθυσμός	Η κατανομή των (δυνατών) τιμών μιας τυχαίας μεταβλητής.

6.4. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Δοκιμή Bernoulli	Πείραμα σε κάθε εκτέλεση του οποίου εμφανίζεται ακριβώς ένα από δύο αμοιβαίως αποκλειόμενα δυνατά αποτελέσματα. Το ένα ονομάζεται επιτυχία και το άλλο αποτυχία.
Η κατανομή Bernoulli με παράμετρο p	Είναι η κατανομή της δίτιμης τυχαίας μεταβλητής, έστω X, που εκφράζει τον αριθμό των επιτυχιών σε μια δοκιμή Bernoulli. Συμβολίζεται με $b(p)$ και έχει • συνάρτηση πιθανότητας $f(x) = P(X = x) = p^x(1-p)^{1-x} = p^x q^{1-x}$, $x = 0, 1$ • μέση τιμή $\mu = E(X) = p$ • διακύμανση $\sigma^2 = Var(X) = p(1-p) = pq$.
Η διωνυμική κατανομή με παραμέτρους ν και p	Είναι η κατανομή της τυχαίας μεταβλητής, έστω X, που εκφράζει τον αριθμό των επιτυχιών σε ν ανεξάρτητες δοκιμές Bernoulli με ίδια πιθανότητα επιτυχίας p. Συμβολίζεται με $B(\nu, p)$ και έχει • συνάρτηση πιθανότητας $f(x) = P(X = x) = \binom{\nu}{x} p^x(1-p)^{\nu-x}$, $x = 0, 1, 2, \ldots, \nu$ • μέση τιμή $\mu = E(X) = \nu p$ • διακύμανση $\sigma^2 = Var(X) = \nu p(1-p)$ • πιο πιθανή τιμή $x_0 = [(\nu+1)p]$ όταν $(\nu+1)p$ δεν είναι ακέραιος ή $x_0 = (\nu+1)p$ και $x_0' = (\nu+1)p - 1$ όταν $(\nu+1)p$ είναι ακέραιος.
Η κατανομή Poisson με παράμετρο λ	Συμβολίζεται με $P(\lambda)$ και είναι γνωστή και ως κατανομή των σπάνιων ενδεχομένων. Έστω X τ.μ. με $X \sim P(\lambda)$. Η X έχει • συνάρτηση πιθανότητας $f(x) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x}$, $x = 0, 1, 2, \ldots$ • μέση τιμή $\mu = E(X) = \lambda$ • διακύμανση $\sigma^2 = Var(X) = \lambda$ • πιο πιθανή τιμή $x_0 = [\lambda]$ όταν το λ δεν είναι ακέραιος ή $x_0 = \lambda$ και $x_0' = \lambda - 1$ όταν το λ είναι ακέραιος.
Προσέγγιση της διωνυμικής κατανομής από την κατανομή Poisson	Αν $\nu \to +\infty$ και $p \to 0$ έτσι ώστε $\nu p \to \lambda$ τότε $$\binom{\nu}{x} p^x(1-p)^{\nu-x} \to e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$ Πρακτικά, η προσέγγιση της διωνυμικής $B(\nu, p)$ από την $P(\nu p)$, είναι ικανοποιητική αν $\nu \geq 10$ και $p \leq 10/\nu$ ώστε η μέση τιμή $\lambda = \nu p$ να παίρνει μέτριες τιμές (μικρότερες του 10).
Η διαδικασία Poisson με ρυθμό λ	Αν X_t ο αριθμός εμφανίσεων ενός ενδεχομένου σε χρόνο t (ή σε μήκος t ή σε επιφάνεια t ή σε όγκο t) τότε κάτω από ορισμένες προϋποθέσεις η συνάρτηση πιθανότητας της X_t δίνεται από τον τύπο $$P(X_t = x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots$$
Πολυωνυμική δοκιμή	Πείραμα σε κάθε εκτέλεση του οποίου εμφανίζεται ακριβώς ένα από $k \geq 2$ αμοιβαίως αποκλειόμενα δυνατά αποτελέσματα $$r_i, \; i = 0, 1, 2, \ldots, k$$
Η πολυωνυμική κατανομή με παραμέτρους $\nu, p_1, p_2, \ldots, p_k$	Αν X_i ο αριθμός εμφανίσεων του αποτελέσματος r_i, $i = 1, 2, \ldots, k$, σε ν ανεξάρτητες επαναλήψεις μιας πολυωνυμικής δοκιμής, όπου η πιθανότητα p_i εμφάνισης του αποτελέσματος r_i παραμένει σε κάθε επανάληψη σταθερή για όλα τα $i = 1, 2, \ldots, k$, τότε $$P(X_1 = \nu_1, X_2 = \nu_2, \ldots, X_k = \nu_k) = \frac{\nu!}{\nu_1! \cdot \nu_2! \cdot \ldots \cdot \nu_k!} p_1^{\nu_1} p_2^{\nu_2} \ldots p_k^{\nu_k}$$ $E(X_i) = \nu p_i$ και $Var(X_i) = \nu p_i(1-p_i)$.

7.4. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Η κανονική κατανομή με παραμέτρους μ και σ

Συμβολίζεται με $N(\mu, \sigma^2)$. Έστω X τ.μ. με $X \sim N(\mu, \sigma^2)$. Η X έχει

- συνάρτηση πυκνότητας

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty, \ -\infty < \mu < +\infty, \ \sigma > 0$$

- μέση τιμή $E(X) = \mu$
- διακύμανση $Var(X) = \sigma^2$.

Η τυποποιημένη κανονική κατανομή

Είναι κανονική κατανομή με μέση τιμή $\mu = 0$ και διακύμανση $\sigma^2 = 1$. Συμβολίζεται με Z, δηλαδή, $Z \sim N(0, 1)$. Η συνάρτηση πυκνότητάς της συμβολίζεται με $\varphi(z)$ και δίνεται από τον τύπο

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < +\infty$$

και η συνάρτηση κατανομής της συμβολίζεται με $\Phi(z)$ και δίνεται από τον τύπο

$$\Phi(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt, \quad -\infty < z < +\infty$$

Υπολογισμός πιθανοτήτων κανονικής τ.μ.

α) Τυποποιημένη κανονική $Z \sim N(0, 1)$:

- $P(Z \le z) = \Phi(z)$ και $P(Z \le -z) = \Phi(-z) = 1 - \Phi(z)$ (δίνονται από τον πίνακα της τυποποιημένης κανονικής)
- $P(\alpha \le Z \le \beta) = \Phi(\beta) - \Phi(\alpha)$
- $P(Z > \alpha) = 1 - P(Z \le \alpha) = 1 - \Phi(\alpha)$
- $P(-\alpha \le Z \le \alpha) = 2\Phi(\alpha) - 1$

β) Κανονική με $X \sim N(\mu, \sigma^2)$

Αν $X \sim N(\mu, \sigma^2)$ τότε $Z = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$

- $P(\alpha \le X \le \beta) = \Phi\left(\dfrac{\beta-\mu}{\sigma}\right) - \Phi\left(\dfrac{\alpha-\mu}{\sigma}\right)$
- $P(X \le \beta) = \Phi\left(\dfrac{\beta-\mu}{\sigma}\right)$
- $P(X \ge \alpha) = 1 - \Phi\left(\dfrac{\alpha-\mu}{\sigma}\right)$.

Γραμμικός συνδυασμός ανεξάρτητων κανονικών τυχαίων μεταβλητών

Αν $X_1, X_2, \ldots, X_\nu$ ανεξάρτητες τ.μ. με $X_i \sim N(\mu_i, \sigma_i^2)$ και $\alpha_1, \alpha_2, \ldots, \alpha_\nu, \beta$ πραγματικοί, τότε η τ.μ.

$$\sum_{i=1}^{\nu} \alpha_i X_i + \beta = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_\nu X_\nu + \beta$$

ακολουθεί μια κανονική κατανομή με μέση τιμή

$$\alpha_1\mu_1 + \alpha_2\mu_2 + \ldots + \alpha_\nu\mu_\nu + \beta$$

και διακύμανση

$$\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + \ldots + \alpha_\nu^2\sigma_\nu^2.$$

Επίσης, αν $X_1, X_2, \ldots, X_\nu$ ανεξάρτητες τ.μ. με $X_i \sim N(\mu, \sigma^2)$, τότε

$$\bar{X} = \frac{\sum_{i=1}^{\nu} X_i}{\nu} \sim N\left(\mu, \frac{\sigma^2}{\nu}\right).$$

Το Κεντρικό Οριακό Θεώρημα	Αν $X_1, X_2, \ldots, X_\nu$ ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την ίδια κατανομή με $E(X_i) = \mu$ και $Var(X_i) = \sigma^2$, τότε για μεγάλα ν, κατά προσέγγιση, $$\bar{X} = \frac{\sum_{i=1}^{\nu} X_i}{\nu} \sim N\left(\mu, \frac{\sigma^2}{\nu}\right)$$ ή ισοδύναμα $$S_\nu = \sum_{i=1}^{\nu} X_i \sim N(\nu\mu, \nu\sigma^2).$$
Προσέγγιση της Διωνυμικής κατανομής από την Κανονική (Θεώρημα De Moivre-Laplace)	Αν $X \sim B(\nu, p)$ τότε για μεγάλα ν (θεωρητικά $\nu \to +\infty$), κατά προσέγγιση, $$X \sim N(\nu p, \nu p(1-p)).$$ Η προσέγγιση είναι ικανοποιητική αν $\nu p \geq 5$ και $\nu(1-p) \geq 5$.
Προσέγγιση της κατανομής Poisson από την Κανονική	Αν $X \sim P(\lambda)$ τότε για μεγάλα λ, κατά προσέγγιση, $$X \sim N(\lambda, \lambda).$$ Η προσέγγιση είναι ικανοποιητική αν $\lambda > 10$.
Η κατανομή χ^2 με ν βαθμούς ελευθερίας	Αν Z_i, $i = 1, 2, \ldots, \nu$ ανεξάρτητες τυποποιημένες κανονικές τ.μ., τότε η κατανομή της τ.μ. $$Z_1^2 + Z_2^2 + \ldots + Z_\nu^2$$ ονομάζεται κατανομή χι-τετράγωνο με ν βαθμούς ελευθερίας και συμβολίζεται με χ^2_ν. Για μεγάλα ν, προσεγγίζεται ικανοποιητικά από την $N(\nu, 2\nu)$.
Η κατανομή t ή Student με ν βαθμούς ελευθερίας	Αν Z, S_ν ανεξάρτητες τ.μ. με $Z \sim N(0, 1)$ και $S_\nu \sim \chi^2_\nu$ τότε η κατανομή της τ.μ. $$\frac{Z}{\sqrt{S_\nu/\nu}}$$ ονομάζεται κατανομή t ή κατανομή Student με ν βαθμούς ελευθερίας και συμβολίζεται με t_ν. Για μεγάλα ν, προσεγγίζεται ικανοποιητικά από την $N(0, \nu/(\nu-2))$.
Η κατανομή F με ν και m βαθμούς ελευθερίας	Αν S_ν, S_m ανεξάρτητες τ.μ. με $S_\nu \sim \chi^2_\nu$ και $S_m \sim \chi^2_m$ τότε η κατανομή της τ.μ. $$\frac{S_\nu/\nu}{S_m/m} = \frac{m}{\nu} \cdot \frac{S_\nu}{S_m}$$ ονομάζεται κατανομή F με ν και m βαθμούς ελευθερίας και συμβολίζεται με $F_{\nu,m}$.

9.4. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Πληθυσμός ή στατιστικός πληθυσμός	Πληθυσμό ή στατιστικό πληθυσμό ονομάζουμε την κατανομή των τιμών μιας τυχαίας μεταβλητής, δηλαδή την κατανομή των τιμών που παίρνει ένα κοινό χαρακτηριστικό μιας ομάδας υποκειμένων. Κάθε υποκείμενο επί του οποίου μετράται/παρατηρείται η τιμή ενός κοινού χαρακτηριστικού ονομάζεται *απλό στοιχείο* ή *δειγματοληπτική/πειραματική μονάδα*.
Τυχαίο δείγμα και πραγματοποίηση τυχαίου δείγματος	Τυχαίο δείγμα μεγέθους ν από έναν πληθυσμό ονομάζουμε ν ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_\nu$ που παίρνουν τιμές από τον πληθυσμό αυτό (και έχουν επομένως την ίδια κατανομή). Οι συγκεκριμένες τιμές $x_1, x_2, \ldots, x_\nu$, που έχουμε διαθέσιμες για επεξεργασία μετά τη λήψη του δείγματος αποτελούν μια πραγματοποίηση των $X_1, X_2, \ldots, X_\nu$ και ονομάζονται *δεδομένα* ή *παρατηρήσεις*.
Πίνακας κατανομής συχνοτήτων	**α) Ποσοτικές μεταβλητές** Στην πρώτη στήλη του πίνακα κατανομής συχνοτήτων καταγράφονται σε αύξουσα σειρά οι διαφορετικές τιμές $y_1, y_2, \ldots, y_k$ ($k \le \nu$) από τις $x_1, x_2, \ldots, x_\nu$ που εμφανίσθηκαν στο δείγμα. Στις επόμενες στήλες, για κάθε τιμή y_i, $i = 1, 2, \ldots, k$, καταγράφεται • *η συχνότητά της,* ν_i *(πόσες φορές εμφανίσθηκε)* • *η σχετική συχνότητά της,* $f_i = \nu_i/\nu$ • *η αθροιστική συχνότητά της,* N_i *(το άθροισμα των συχνοτήτων των τιμών που είναι* $\le y_i$*)* • *η αθροιστική σχετική συχνότητά της,* F_i *(το άθροισμα των σχετικών συχνοτήτων των τιμών που είναι* $\le y_i$*)* Αν (έχει) γίνει ομαδοποίηση των τιμών, στην πρώτη στήλη αντί των διαφορετικών τιμών καταγράφονται οι διαφορετικές κλάσεις τιμών. Στις επόμενες στήλες καταγράφεται η συχνότητα, η σχετική συχνότητα, η αθροιστική συχνότητα και η αθροιστική σχετική συχνότητα κάθε κλάσης τιμών. **β) Ποιοτικές μεταβλητές** • Στις ποιοτικές μεταβλητές *κατηγορίας* δεν έχει νόημα η διάταξη των διαφορετικών τιμών $y_1, y_2, \ldots, y_k$ και επομένως δεν έχουν νόημα ούτε οι αθροιστικές ούτε οι αθροιστικές σχετικές συχνότητες αλλά μόνο οι συχνότητες και οι σχετικές συχνότητες. • Στις ποιοτικές μεταβλητές *διάταξης* η διάταξη των διαφορετικών τιμών $y_1, y_2, \ldots, y_k$ έχει νόημα και επομένως έχουν νόημα τόσο οι συχνότητες και οι σχετικές συχνότητες όσο και οι αθροιστικές και οι αθροιστικές σχετικές.
Γραφική παρουσίαση κατανομής συχνοτήτων	**α) Ποσοτικές μεταβλητές** • *Σημειόγραμμα* • *Ραβδόγραμμα συχνοτήτων και σχετικών συχνοτήτων* • *Διάγραμμα συχνοτήτων και σχετικών συχνοτήτων* • *Κυκλικό διάγραμμα συχνοτήτων και σχετικών συχνοτήτων* • *Ιστόγραμμα συχνοτήτων/σχετικών συχνοτήτων/αθροιστικών συχνοτήτων/αθροιστικών σχετικών συχνοτήτων* • *Πολύγωνο συχνοτήτων/σχετικών συχνοτήτων/αθροιστικών συχνοτήτων/αθροιστικών σχετικών συχνοτήτων* • *Φυλλογράφημα* • *Θηκόγραμμα*

β) Ποιοτικές μεταβλητές

- *Ραβδόγραμμα συχνοτήτων και σχετικών συχνοτήτων*
- *Κυκλικό διάγραμμα συχνοτήτων και σχετικών συχνοτήτων*

γ) Κυκλικές μεταβλητές (διεύθυνσης ή κατεύθυνσης)

- *Κυκλικό διάγραμμα διασποράς*
- *Ροδόγραμμα*
- *Κυκλικό ιστόγραμμα*
- *Γραμμικό ιστόγραμμα*

Αριθμητικά περιγραφικά μέτρα *(για συγκεκριμένη πραγματοποίηση $x_1, x_2, \dots, x_\nu$ δείγματος με $y_1, y_2, \dots, y_k$, $k \leq \nu$ διαφορετικές τιμές)*

α) Ποσοτικές μεταβλητές

Μέτρα θέσης

- ***Δειγματικός μέσος,*** $\bar{x}$

  $\bar{x} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i = \frac{1}{\nu}\sum_{i=1}^{k} \nu_i y_i = \sum_{i=1}^{k} f_i y_i$
- ***Κορυφή του δείγματος,*** M_0

  Η τιμή με τη μεγαλύτερη συχνότητα
- ***Διάμεσος του δείγματος,*** δ ή Q_2

  Το πολύ 50% των τιμών του δείγματος είναι μικρότερες από τη διάμεσο και επίσης το πολύ 50% των τιμών του δείγματος είναι μεγαλύτερες από τη διάμεσο.

  Σε αύξουσα διάταξη των $x_1, x_2, \dots, x_\nu$, τη θέση της διαμέσου δίνει ο αριθμός $0.5(\nu+1)$ εφόσον είναι ακέραιος, ενώ αν δεν είναι ακέραιος, τότε η διάμεσος είναι ίση με το ημιάθροισμα των δύο τιμών που οι θέσεις τους είναι οι πλησιέστερες στον αριθμό $0.5(\nu+1)$.
- ***p-ποσοστιαία σημεία του δείγματος,*** x_p, $0 < p < 1$

  Το πολύ $100p$ % των τιμών του δείγματος είναι μικρότερες από το p-ποσοστιαίο σημείο και το πολύ $100(1-p)$% των τιμών του δείγματος είναι μεγαλύτερες από το p-ποσοστιαίο σημείο.

  Σε αύξουσα διάταξη των $x_1, x_2, \dots, x_\nu$, τη θέση του p-ποσοστιαίου σημείου δίνει ο αριθμός $p(\nu+1)$ εφόσον είναι ακέραιος, ενώ αν δεν είναι ακέραιος, τότε το p-ποσοστιαίο σημείο εκτιμάται με παρεμβολή μεταξύ των δύο τιμών που οι θέσεις τους είναι οι πλησιέστερες στον αριθμό $p(\nu+1)$.
- ***Τεταρτημόρια,*** Q_1, Q_2, Q_3

  $Q_1 = x_{0.25}$, $Q_2 = x_{0.5} = \delta$, $Q_3 = x_{0.75}$

Αν (έχει) γίνει ομαδοποίηση των τιμών του δείγματος σε k κλάσεις:

- Η *κορυφή* υπολογίζεται από τον τύπο

  $M_0 = L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} c_i$

  όπου, L_i το κάτω άκρο της επικρατούσας κλάσης, δηλαδή της κλάσης με τη μεγαλύτερη συχνότητα, c_i το πλάτος της επικρατούσας κλάσης, $\Delta_1 = \nu_i - \nu_{i-1}$ και $\Delta_2 = \nu_i - \nu_{i+1}$ όπου ν_i η συχνότητα της επικρατούσας κλάσης.
- Στον τύπο υπολογισμού του *δειγματικού μέσου*

  $\bar{x} = \frac{1}{\nu}\sum_{i=1}^{k} \nu_i y_i = \sum_{i=1}^{k} f_i y_i$

  τα y_i, $i = 1, 2, \dots, k$ είναι οι κεντρικές τιμές των κλάσεων.
- Η *διάμεσος* υπολογίζεται από τον τύπο

  $\delta = L_i + \frac{0.5\nu - N_{i-1}}{\nu_i} c_i$

όπου, L_i το κάτω άκρο της μεσαίας κλάσης, δηλαδή της κλάσης στην οποία ανήκει η διάμεσος, c_i το πλάτος της μεσαίας κλάσης, v_i η συχνότητα της μεσαίας κλάσης και N_{i-1} η αθροιστική συχνότητα της προηγούμενης κλάσης από τη μεσαία.

- Τα *p-ποσοστιαία σημεία* υπολογίζονται από τον τύπο

$$x_p = L_i + \frac{pv - N_{i-1}}{v_i} c_i$$

όπου, L_i το κάτω άκρο της κλάσης στην οποία βρίσκεται το x_p, c_i το πλάτος της, v_i η συχνότητά της και N_{i-1} η αθροιστική συχνότητα της προηγούμενης κλάσης.

Μέτρα μεταβλητότητας/διασποράς

- ***Εύρος***

$$R = x_{max} - x_{min}$$

- ***Ενδοτεταρτημοριακό εύρος***

$$Q_3 - Q_1$$

- ***Δειγματική διακύμανση***

$$s^2 = \frac{1}{v-1}\sum_{i=1}^{v}(x_i - \bar{x})^2 = \frac{1}{v-1}\left(\sum_{i=1}^{v} x_i^2 - v\bar{x}^2\right) = \frac{1}{v-1}\sum_{i=1}^{k}(y_i - \bar{x})^2 v_i = \frac{1}{v-1}\left(\sum_{i=1}^{k} v_i y_i^2 - v\bar{x}^2\right)$$

- ***Δειγματική τυπική απόκλιση***

$$s = \sqrt{\frac{1}{v-1}\sum_{i=1}^{v}(x_i - \bar{x})^2} = \sqrt{\frac{1}{v-1}\left(\sum_{i=1}^{v} x_i^2 - v\bar{x}^2\right)} =$$

$$= \sqrt{\frac{1}{v-1}\sum_{i=1}^{k}(y_i - \bar{x})^2 v_i} = \sqrt{\frac{1}{v-1}\left(\sum_{i=1}^{k} v_i x_i^2 - v\bar{x}^2\right)}$$

- ***Συντελεστής μεταβλητότητας***

$$CV = \frac{s}{|\bar{x}|}100\%$$

Αν (έχει) γίνει ομαδοποίηση των τιμών του δείγματος σε k κλάσεις τα $y_i = 1, 2, ..., k$, είναι οι κεντρικές τιμές των κλάσεων.

Μέτρα λοξότητας και μέτρα κύρτωσης

- ***Συντελεστές ασυμμετρίας του Pearson***

$$\gamma_1 = \frac{\bar{x} - M_o}{s}, \quad \gamma_2 = \frac{3(\bar{x} - \delta)}{s}$$

Αν $\gamma_1 = \gamma_2 = 0$ η κατανομή είναι συμμετρική.

- ***Συντελεστής ασυμμετρίας του Bowley***

$$S_k = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

Αν $S_k = 0$ η κατανομή είναι συμμετρική.

- ***Συντελεστής κύρτωσης του Pearson***

$$\beta_2 = \frac{m_4}{(m_2)^2}$$

Αν $\beta_2 = 3$ η κατανομή είναι μεσόκυρτη, αν $\beta_2 > 3$ είναι πλατύκυρτη και αν $\beta_2 < 3$ είναι λεπτόκυρτη.

β) Ποιοτικές μεταβλητές

Ορίζεται (και έχει νόημα) μόνο η *κορυφή* της κατανομής.

γ) Μεταβλητές διεύθυνσης και κατεύθυνσης

Μέτρα θέσης

- ***Μέση κατεύθυνση*** $\bar{\vartheta}$ ν κατευθύνσεων $\vartheta_1, \vartheta_2, \ldots, \vartheta_\nu$.

$$\bar{\vartheta} = \tau o\xi\varepsilon\varphi \frac{\sum_{i=1}^{\nu} \eta\mu\,\vartheta_i}{\sum_{i=1}^{\nu} \sigma\upsilon\nu\,\vartheta_i}$$

σε συνδυασμό με το πρόσημο των

$$x_r = \sum_{i=1}^{\nu} \eta\mu\,\vartheta_i \text{ και } y_r = \sum_{i=1}^{\nu} \sigma\upsilon\nu\,\vartheta_i.$$

Μέτρα μεταβλητότητας

- ***Διακύμανση*** ν κατευθύνσεων $\vartheta_1, \vartheta_2, \ldots, \vartheta_\nu$.

$$S^2 = 1 - \bar{r}$$

$$s^2 = 2(1 - \bar{r})$$

$$s_0^2 = -2\ln \bar{r}$$

Όπου,

$$\bar{r} = \frac{r}{\nu} = \frac{\sqrt{x_r^2 + y_r^2}}{\nu}$$

- ***Τυπική απόκλιση*** ν κατευθύνσεων $\vartheta_1, \vartheta_2, \ldots, \vartheta_\nu$.

$$s = \sqrt{2(1 - \bar{r})}$$

$$s_0 = \sqrt{-2\ln \bar{r}}$$

Η μέση διεύθυνση, η διακύμανση και η τυπική απόκλιση ν *διευθύνσεων* $\vartheta_1, \vartheta_2, \ldots, \vartheta_\nu$, υπολογίζεται όπως τα αντίστοιχα μέτρα κατευθύνσεων αφού προηγουμένως οι τιμές $\vartheta_1, \vartheta_2, \ldots, \vartheta_\nu$ μετασχηματισθούν κατάλληλα.

Μέτρα θέσης και μεταβλητότητας γραμμικού μετασχηματισμού των παρατηρήσεων/δεδομένων

Αν $t_i = \alpha x_i + \beta$, $i = 1, 2, \ldots, \nu$

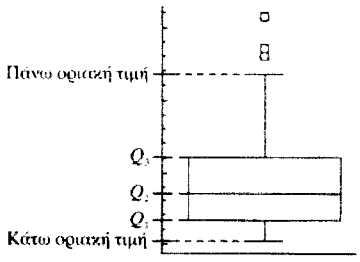
τότε

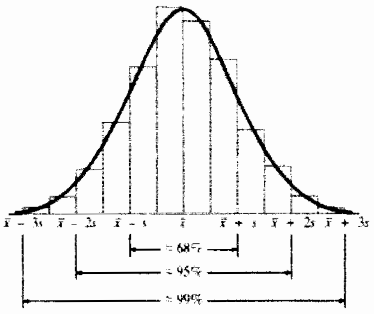
- $\bar{t} = \alpha\bar{x} + \beta$
- $s_t^2 = \alpha^2 s_x^2$
- $s_t = |\alpha| s_x$
- $\delta_t = \alpha\delta_x + \beta$
- $M_{0t} = \alpha M_{0x} + \beta$

Θηκόγραμμα

Πάνω οριακή τιμή: η μεγαλύτερη τιμή του δείγματος που είναι $\leq Q_3 + 1.5(Q_3 - Q_1)$ ή $\leq Q_3 + 3(Q_3 - Q_1)$

Κάτω οριακή τιμή: η μικρότερη τιμή του δείγματος που είναι $\geq Q_1 - 1.5(Q_3 - Q_1)$ ή $\geq Q_1 - 3(Q_3 - Q_1)$



Ο εμπειρικός κανόνας	Αν η κατανομή του δείγματος προσομοιάζει με μια κανονική κατανομή (έχει κωδωνοειδή μορφή), τότε • στο διάστημα $(\bar{x} - s, \bar{x} + s)$ βρίσκεται περίπου το 68% των παρατηρήσεων • στο διάστημα $(\bar{x} - 2s, \bar{x} + 2s)$ βρίσκεται περίπου το 95% των παρατηρήσεων • στο διάστημα $(\bar{x} - 3s, \bar{x} + 3s)$ βρίσκονται περίπου όλες οι παρατηρήσεις (πάνω από το 99%). 
Η ανισότητα *Chebyshev*	Το ποσοστό των τιμών του δείγματος που βρίσκονται στο διάστημα $(\bar{x} - ks, \bar{x} + ks)$ είναι *τουλάχιστον* $1 - (1/k^2)$.

10.3. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Στατιστική συνάρτηση ή Δειγματοσυνάρτηση	Κάθε πραγματική συνάρτηση $T = T(X_1, X_2, \ldots, X_\nu)$ τυχαίου δείγματος $X_1, X_2, \ldots, X_\nu$ από έναν πληθυσμό.
Δειγματοληπτική κατανομή ή Κατανομή δειγματοληψίας	Η κατανομή των τιμών μιας στατιστικής συνάρτησης.
Κατανομή του δειγματικού μέσου $\bar{X}$	• Αν $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, \nu$, τότε $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{\nu}\right)$. • Αν X_i, $i = 1, 2, \ldots, \nu$ από οποιαδήποτε κατανομή με $E(X_i) = \mu$ και $Var(X_i) = \sigma^2$ και ν μεγάλο (≥ 30) τότε $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{\nu}\right)$ κατά προσέγγιση.
Κατανομή του δειγματικού ποσοστού $\hat{P}$	Αν X_i, $i = 1, 2, \ldots, \nu$ τ.μ. Bernoulli με παράμετρο p και ν μεγάλο ($\nu p \geq 5$ και $\nu(1-p) \geq 5$), τότε $\hat{P} \sim N\left(p, \frac{p(1-p)}{\nu}\right)$ κατά προσέγγιση.
Κατανομή της τ.μ. $\frac{(\bar{X}-\mu)\sqrt{\nu}}{\sigma}$	• Αν $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, \nu$, τότε $\frac{(\bar{X}-\mu)\sqrt{\nu}}{\sigma} \sim N(0,1)$. • Αν X_i, $i = 1, 2, \ldots, \nu$, από οποιαδήποτε κατανομή με $E(X_i) = \mu$ και $Var(X_i) = \sigma^2$ και ν μεγάλο (≥ 30) τότε $\frac{(\bar{X}-\mu)\sqrt{\nu}}{\sigma} \sim N(0,1)$ κατά προσέγγιση.
Κατανομή της τ.μ. $\frac{(\bar{X}-\mu)\sqrt{\nu}}{s}$	• Αν X_i, $i = 1, 2, \ldots, \nu$, από οποιαδήποτε κατανομή με $E(X_i) = \mu$ και $Var(X_i) = \sigma^2$ και ν μεγάλο (≥ 30) τότε $\frac{(\bar{X}-\mu)\sqrt{\nu}}{s} \sim N(0,1)$ κατά προσέγγιση. • Αν $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, \nu$, τότε $\frac{(\bar{X}-\mu)\sqrt{\nu}}{s} \sim t_{\nu-1}$.
Κατανομή της τ.μ. $\frac{(\nu-1)S^2}{\sigma^2}$	Αν $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, \nu$, τότε $\frac{(\nu-1)S^2}{\sigma^2} \sim \chi^2_{\nu-1}$.
Κατανομή της τ.μ. $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	Αν $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ είναι ανεξάρτητα και $X_1, X_2, \ldots, X_{\nu_1} \sim N(\mu_1, \sigma_1^2)$ και $Y_1, Y_2, \ldots, Y_{\nu_2} \sim N(\mu_2, \sigma_2^2)$ τότε $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{\nu_1-1;\nu_2-1}$.

Αποτελέσματα για τη διαφορά δύο δειγματικών μέσων $\bar{X}, \bar{Y}$

- Αν $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ είναι ανεξάρτητα και $X_1, X_2, \ldots, X_{\nu_1} \sim N(\mu_1, \sigma_1^2)$ και $Y_1, Y_2, \ldots, Y_{\nu_2} \sim N(\mu_2, \sigma_2^2)$ τότε

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \sim N(0,1).$$

- Αν $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ είναι ανεξάρτητα και $X_1, X_2, \ldots, X_{\nu_1} \sim N(\mu_1, \sigma^2)$ και $Y_1, Y_2, \ldots, Y_{\nu_2} \sim N(\mu_2, \sigma^2)$ τότε

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \sim t_{\nu_1 + \nu_2 - 2}$$

όπου

$$S^2 = \frac{(\nu_1 - 1)S_1^2 + (\nu_2 - 1)S_2^2}{\nu_1 + \nu_2 - 2}.$$

- Αν $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ είναι ανεξάρτητα και $\nu_1, \nu_2 \geq 30$, τότε κατά προσέγγιση

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \sim N(0,1)$$

και

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{\nu_1} + \dfrac{S_2^2}{\nu_2}}} \sim N(0,1).$$

11.9. Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Εκτιμήτρια	Κάθε στατιστική συνάρτηση που χρησιμοποιείται για την εκτίμηση μιας παραμέτρου ενός πληθυσμού.
Σημειακή εκτίμηση	Η εκτίμηση μιας άγνωστης παραμέτρου ενός πληθυσμού μέσω της τιμής μιας εκτιμήτριας για συγκεκριμένη πραγματοποίηση ενός τυχαίου δείγματος.
Ιδιότητες εκτιμητριών	Αμεροληψία: Σε επαναλαμβανόμενες δειγματοληψίες η εκτιμήτρια κατά μέσο όρο εκτιμά σωστά την άγνωστη παράμετρο (ούτε την υπερεκτιμά ούτε την υποεκτιμά). Αποτελεσματικότητα: Μεταξύ δύο αμερόληπτων εκτιμητριών μιας παραμέτρου, πιο αποτελεσματική είναι αυτή που έχει πιο μικρή διακύμανση. Συνέπεια: Αυξανομένου του μεγέθους του δείγματος η εκτίμηση γίνεται καλύτερη/πιο ακριβής (οι τιμές της εκτιμήτριας συγκλίνουν στην τιμή της παραμέτρου).
Εκτίμηση με διάστημα	Ένα $100(1-\alpha)\%$ διάστημα εμπιστοσύνης $(0<\alpha<1)$ για μια παράμετρο ενός πληθυσμού, είναι ένα διάστημα που υπολογίζεται από ένα τυχαίο δείγμα από τον πληθυσμό και έχει πιθανότητα $1-\alpha$ να περιέχει την πραγματική τιμή της παραμέτρου. Η πιθανότητα $1-\alpha$ ονομάζεται *συντελεστής εμπιστοσύνης* του διαστήματος.
Ερμηνεία ενός $100(1-\alpha)\%$ διαστήματος εμπιστοσύνης σύμφωνα με την ερμηνεία της πιθανότητας ως οριακή σχετική συχνότητα	Σε μεγάλο αριθμό επαναλήψεων του πειράματος «παίρνω ένα τυχαίο δείγμα μεγέθους ν από τον πληθυσμό και κατασκευάζω για μια άγνωστη παράμετρο ένα $100(1-\alpha)\%$ διάστημα εμπιστοσύνης», ποσοστό $1-\alpha$ των δειγμάτων θα δώσουν διάστημα που θα περιέχει την τιμή της παραμέτρου και ποσοστό α των δειγμάτων θα δώσουν διάστημα που δε θα περιέχει την τιμή της παραμέτρου.

Διαστήματα Εμπιστοσύνης (Δ.Ε.) συντελεστή εμπιστοσύνης 1 – α

1. Για τη μέση τιμή μ ενός πληθυσμού με ένα τυχαίο δείγμα μεγέθους ν

Πληθυσμός	Διακύμανση του πληθυσμού σ^2	Μέγεθος του δείγματος ν	$100(1-\alpha)\%$ *Δ.Ε. για τη μέση τιμή, μ, του πληθυσμού*
Κανονικός	Γνωστή	Οτιδήποτε	$\bar{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{\nu}}$ (1)
Οποιοσδήποτε	Γνωστή	Μεγάλο	
Οποιοσδήποτε	Άγνωστη	Μεγάλο	$\bar{X} \pm z_{\alpha/2}\dfrac{S}{\sqrt{\nu}}$ (2)
Κανονικός	Άγνωστη	Οτιδήποτε	$\bar{X} \pm t_{\nu-1,\alpha/2}\dfrac{S}{\sqrt{\nu}}$ (3)
Όχι κανονικός	Γνωστή ή Άγνωστη	Μικρό	?

Σημείωση: (α) Για μεγάλο δείγμα από κανονική κατανομή με άγνωστη διακύμανση, μπορεί να εφαρμοσθεί ή ο τύπος (2) ή ο τύπος (3) (β) Ο τύπος (3) εφαρμόζεται με την προϋπόθεση ότι ο πληθυσμός είναι κανονικός, όμως στην πράξη δίνει καλά αποτελέσματα ακόμη και όταν ο πληθυσμός δεν είναι κανονικός αρκεί να μην απέχει δραματικά από το να είναι κανονικός (σοβαρή ασυμμετρία, περισσότερες από μια κορυφές, κτλ.) και το δείγμα να μην είναι πολύ μικρό. (γ) Το διάστημα (2) είναι συντελεστή $1-\alpha$ κατά προσέγγιση. Επίσης και το διάστημα (1), αν ο πληθυσμός δεν είναι κανονικός, είναι συντελεστή $1-\alpha$ κατά προσέγγιση.

2. Για το διωνυμικό ποσοστό p με ένα τυχαίο δείγμα μεγέθους ν

Αν

$$\hat{p} = \frac{\textit{αριθμός επιτυχιών στο δείγμα}}{\nu}$$

και

$$\nu\hat{p} \geq 5 \text{ και } \nu(1-\hat{p}) \geq 5$$

το διάστημα

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{\nu}}$$

είναι ένα κατά προσέγγιση 100 (1 – α)% διάστημα εμπιστοσύνης για το διωνυμικό ποσοστό p.

3. Για τη διακύμανση σ^2 ενός κανονικού πληθυσμού με ένα τυχαίο δείγμα μεγέθους ν

$$\left[\frac{\nu-1}{\chi^2_{\nu-1;\alpha/2}}S^2, \frac{\nu-1}{\chi^2_{\nu-1;1-\alpha/2}}S^2\right]$$

4. Για τη διαφορά $\mu_1 - \mu_2$ των μέσων τιμών δύο πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους ν_1 και ν_2 αντίστοιχα

Σημείωση: Το διάστημα (2) είναι συντελεστή 1 – α κατά προσέγγιση. Επίσης και το διάστημα (1), αν οι πληθυσμοί δεν είναι κανονικοί, είναι συντελεστή 1 – α κατά προσέγγιση.

Πληθυσμοί	Διακύμανσεις των πληθυσμών σ_1^2, σ_2^2	Μεγέθη των δειγμάτων ν_1, ν_2	100(1 – α)% *Δ.Ε. για τη διαφορά* $\mu_1 - \mu_2$ *των μέσων τιμών των πληθυσμών*
Κανονικοί	Γνωστές	Οτιδήποτε	$(\bar{X}-\bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{\nu_1}+\frac{\sigma_2^2}{\nu_2}}$ (1)
Οποιοιδήποτε	Γνωστές	Μεγάλα	
Οποιοιδήποτε	Άγνωστες	Μεγάλα	$(\bar{X}-\bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{\nu_1}+\frac{S_2^2}{\nu_2}}$ (2)
Κανονικοί	Άγνωστες *και ίσες*	Οτιδήποτε	$(\bar{X}-\bar{Y}) \pm t_{\nu;\alpha/2}S\sqrt{\frac{1}{\nu_1}+\frac{1}{\nu_2}}$ (3) Όπου, $\nu = \nu_1 + \nu_2 - 2$ και $S = \sqrt{\frac{(\nu_1-1)S_1^2+(\nu_2-1)S_2^2}{\nu_1+\nu_2-2}}$
Όχι κανονικοί	Γνωστές ή Άγνωστες (ίσες ή άνισες)	Μικρά	?

5. Για τη διαφορά $p_1 - p_2$ δύο διωνυμικών ποσοστών με δύο ανεξάρτητα δείγματα μεγέθους ν_1 και ν_2 αντίστοιχα

Αν

$$\hat{p}_i = \frac{\text{αριθμός επιτυχιών στο δείγμα } i}{\nu}, \quad i = 1, 2$$

και

$$\nu_i \hat{p}_i \geq 5 \text{ και } \nu_i(1 - \hat{p}_i) \geq 5, \, i = 1, 2$$

το διάστημα

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{\nu_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{\nu_2}}$$

είναι ένα κατά προσέγγιση $100(1 - \alpha)\%$ διάστημα εμπιστοσύνης για τη διαφορά $(p_1 - p_2)$ δύο διωνυμικών ποσοστών.

6. Για το λόγο σ_1^2/σ_2^2 των διακυμάνσεων δύο κανονικών πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους ν_1 και ν_2 αντίστοιχα

$$\left[\frac{1}{F_{\nu_1 - 1; \nu_2 - 1; \alpha/2}} \frac{S_1^2}{S_2^2}, \, F_{\nu_2 - 1; \nu_1 - 1; \alpha/2} \frac{S_1^2}{S_2^2}\right]$$