Μάθημα 1ο

1ο κεφάλαιο

04/10/

Τυχαία μεταβλητή: Συνάρτηση με πεδίο ορισμού το δειγματικό χώρο ενός πειράματος και πεδίο τιμών το $\mathbb{R}$

Συμβολισμός: $X: \Omega \to \mathbb{R}$ και όταν την ορίσουμε, μετά χρησιμοποιούμε τα αποτελέσματα.

πχ

Ρίχνω ένα ζάρι, κερδίζω τόσες χιλιάδες € όσες ο αριθμός που έφερε

$1 \mapsto 1000 \; x_1$
$2 \mapsto 2000 \; x_2$
$\vdots$
$6 \mapsto 6000 \; x_6$

$\Omega = \{\underset{\omega_1}{1}, \underset{\omega_2}{2}, 3, 4, 5, \underset{\omega_6}{6}\}$

$\underset{\text{τμ}}{X}: \Omega \to \mathbb{R}$ με $x_1, x_2, \dots x_6$

-> εκφράζει πόσα λεφτά θα πάρω.

ω: αρχικό αποτέλεσμα

X: αυτό που με ενδιαφέρει, που έχω αντιστοιχίσει.

Γενικά

1) x_i = αριθμοί

2) Σε κάθε αποτέλεσμα του πειράματος αντιστοιχεί μόνο ένα x_i

πχ

Ρίχνω ένα νόμισμα 3 φορές

$\Omega = \{\underset{\omega_1}{ΓΓΓ}, \underset{\omega_2}{ΓΓΚ}, \underset{\omega_3}{ΓΚΓ}, \underset{\omega_4}{ΚΚΓ}, \underset{\omega_5}{ΚΓΓ}, \underset{\omega_6}{ΚΓΚ}, \underset{\omega_7}{ΚΚΚ}, \underset{\omega_8}{ΓΚΚ}\}$

Πόσες φορές μπορεί να έρθει κορώνα? $\{\underset{x_1}{0}, \underset{x_2}{1}, \underset{x_3}{2}, \underset{x_4}{3}\}$

X:	$\omega_1 \to 0$	$\omega_5 \to 1$
πόσες Κ	$\omega_2 \to 1$	$\omega_6 \to 2$
ήρθαν	$\omega_3 \to 1$	$\omega_7 \to 3$
	$\omega_4 \to 2$	$\omega_8 \to 2$

Ενώ αρχικά ίσως να έχω ισοπίθανα ενδεχόμενα, στην αντιστοίχηση δεν είναι απαραίτητο να διατηρηθεί.

Τυχαία Μεταβλητή

Διακριτή

αποτελέσματα είναι συγκεκριμένα
χωρισμένα το ένα απ' το άλλο.
τιμές της είναι ένα πεπερασμένο
(άπειρο) αριθμήσιμο σύνολο. Μπορώ
δηλαδή να τα πω ένα-ένα με το όνομά
τους **πχ** αριθμός παιδιών.

Συνεχής

Τα αποτελέσματα μπορεί να
είναι οποιοδήποτε νούμερο
Οι τιμές είναι οτιδήποτε
μέσα σε ένα διάστημα $[\alpha, \beta]$
πχ ύψος, βάρος.

Συνάρτηση Πιθανότητας της τυχαίας μεταβλητής X (Διακριτή): ποιες είναι οι πιθανότητες να γίνει αυτό που με ενδιαφέρει.

πχ Πόσες κορώνες θα έρθουν αν ρίξω ένα νόμισμα τρεις φορές?

$$\underbrace{P(x=0)=\frac{1}{8}}_{P_1},\quad \underbrace{P(x=1)=\frac{3}{8}}_{P_2},\quad \underbrace{P(x=2)=\frac{3}{8}}_{P_3},\quad \underbrace{P(x=3)=\frac{1}{8}}_{P_4}$$

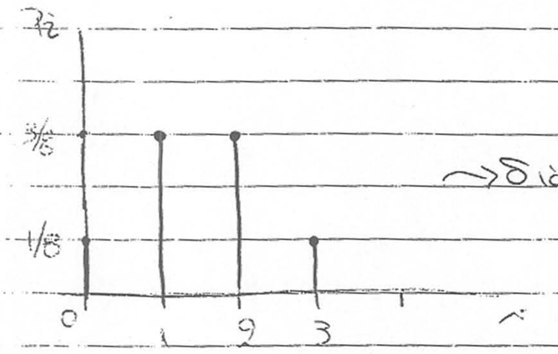
Συνάρτηση πιθανότητας: $f: X \to \mathbb{R}$

$$f(x_i)=P(X=x_i)=P_i$$

x_i	0	1	2	3
P_i	1/8	3/8	3/8	1/8

$X=\{\underset{x_1}{0}, \underset{x_2}{1}, \underset{x_3}{2}, \underset{x_4}{3}\}$

↳ πίνακας κατανομής πιθανότητας



→ διάγραμμα

$$\underbrace{\{\omega_1, \omega_2, \ldots, \omega_n\}}_{\text{ενδεχόμενα}} \xrightarrow{X = \text{τυχαία μεταβλητή}} \{x_1, x_2, \ldots, x_k\} \xrightarrow{f = \text{συνάρτηση πιθανότητας}} \{P_1, P_2, \ldots, P_k\}$$

Άρα $\quad 0 \le P_i \le 1 \quad$ και $\quad \Sigma P_i = 1$

Χ: διακριτή τυχαία μεταβλητή.

Συνάρτηση πιθανότητας: $P_i = P(X=x_i) = f(x_i)$

Αθροιστική συνάρτηση πιθανότητας: $F(x_i) = P(X \le x_i)$

πχ $F(x_3) = P(X \le x_3) = P(X=x_1) + P(X=x_2) + P(X=x_3) = p_1 + p_2 + p_3$

Με ενδιαφέρει η πιθανότητα μέχρι το συγκεκριμένο που θέλω
(Μας αθροίζει όλες τις πιθανότητες μέχρι αυτή που με ενδιαφέρει

πχ * Ρίχνω ένα ζάρι. Αν φέρω άρτιο κερδίζω 50€, αν φέρω 1 χάνω 10€, αν φέρω 5 χάνω 100€

$\Omega = \{1,2,3,4,5,6\}$

$1 \mapsto -10$
$2 \mapsto 50$
$3 \mapsto -10$
$4 \mapsto 50$
$5 \mapsto -100$
$6 \mapsto 50$

X: κέρδος παίκτη
$x: \{\underset{x_1}{-100}, \underset{x_2}{-10}, \underset{x_3}{50}\}$

*	P_i ($=f_i$)	F
$-100 = x_1$	1/6	1/6
$-10 = x_2$	2/6	3/6
$+50 = x_3$	3/6	1

Ποια είναι η πιθανότητα να χάσει 10€ $P(X=-10) = P_2 = 2/6$
Ποια είναι η πιθανότητα να χάσει? $P(X<0) = p_1 + p_2 = \frac{3}{6} = \frac{1}{2}$
← προσθέτω όλα όσα αντιστοιχούν σε

ακριβώς ίσο: απλή πιθανότητα
<> : βρίσκω τα όρια και προσθέτω τις προηγούμενες

Ιδιότητες Αθροιστικής συνάρτησης πιθανότητας: Η πιθανότητα που ψάχνω να είναι μικρότερο ή ίσο από κάτι (αν δεν έχω ≤ πρέπει να κατασκευάσω) $P(X \le k)$

a-1 a a+1 β-1 β β

• $P(X<a) = P(X \le a-1) = F$ στη θέση $a-1$.

• $P(X>a) = 1 - P(X \le a)$

• $P(X \ge a) = 1 - P(X \le a-1)$

• $P(a \le X \le \beta) = 1 - P(X \le a-1) - P(X \ge \beta+1) = 1 - P(X \le a-1) - (1 - P(X \le \beta))$

• $P(a < X < \beta) = 1 - P(X \le a) - P(X \ge \beta) = 1 - P(X \le a) - (1 - P(X \le \beta - 1))$

Πχ

Έχω Χ διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας

$$P(X=x)=\begin{cases}\left(\frac{1}{2}\right)^x, & x=1,2,3,4,5\\ a, & x=6\\ 0, & \text{για κάθε άλλο } x\end{cases}$$

i) Να βρεθεί ο α

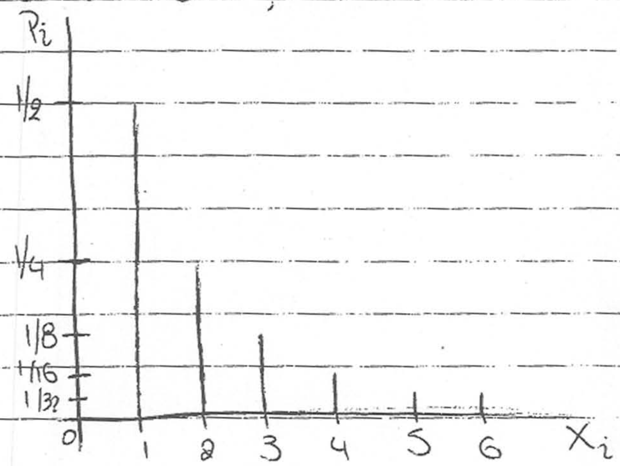
ii) Να γίνει πίνακας κατανομής πιθανότητας και το αντίστοιχο διάγραμμα.

Λύση

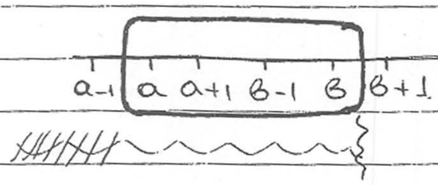
i) $\frac{1}{2}+\left(\frac{1}{2}\right)^2+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+a=1 \Rightarrow a=0,03125$

ii)

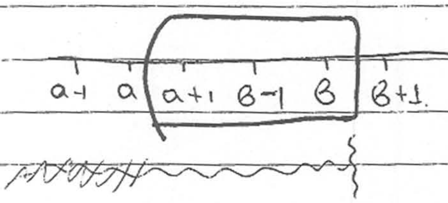
X_i	P_i
1	1/2
2	1/4
3	1/8
4	1/16
5	1/32
6	0,03125 ≃ 1/32



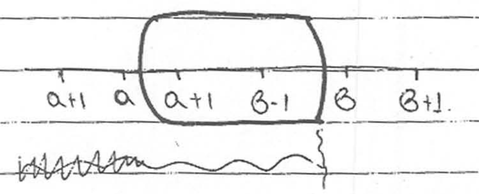
$P(a \le x \le \beta) = P(x \le \beta) - P(x \le a-1)$



$P(a < x \le \beta) = P(x \le \beta) - P(x \le a)$



$P(a < x < \beta) = P(x \le \beta-1) - P(x \le a)$



Άσκηση

Δίνεται ο πίνακας κατανομής πιθανότητας

x_i	P_i	F
0	1/6	1/6
1	1/6	1/3
2	1/8	11/24
3	1/4	17/24
4	7/24	1

Να βρεθούν οι πιθανότητες

i) $P(x \leq 1) = F_1 = 1/3$

ii) $P(1 \leq x \leq 3) = F_3 - F_0 = \frac{17}{24} - \frac{1}{6} = \frac{13}{24}$

iii) $P(x \geq 2) = 1 - F_1 = 1 - \frac{1}{3} = \frac{2}{3}$

Αναμενόμενη τιμή της X ή αριθμητικός μέσος πληθυσμού είν ένας αριθμός που εκφράζει το μέσο όρο των τιμών που θα βρίσκαμε εάν κάναμε πάρα πολλούς υπολογισμούς

Συμβολισμός $E(x) = x_1 \cdot P_1 + x_2 \cdot P_2 + \dots + x_r \cdot P_r$ (μέση τιμή)

Παρατήρηση Το $E(x)$ δεν είναι υποχρεωτικά κάποια τιμή της

πχ *

Ποιο είναι το αναμενόμενο κέρδος του παίκτη?

x_i	P_i	$x_i \cdot P_i$
-100	1/6	-100/6
-10	2/6	-20/6
50	3/6	150/6
Σύνολο		30/6

$E(x) = \frac{30}{6} = 5$ €

Ιδιότητες Αναμενόμενης τιμής

- $E(a \cdot X + B) = a \cdot E(x) + B$
- $E(a) = a$

Αν ξέρω $E(x)$ και ψάχνω σε ένα άλλο που συνδέονται γραμμ κάνω αντικατάσταση

Διακύμανση της X

• Υπολογίζω τη μεταβλητή x^2 (ταχί υψώνονται στο τετράγωνο με ίδιες πιθανότητες P_i)

• Βρίσκω την αναμενόμενη τιμή της x^2

$$\boxed{E(x^2) = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + \ldots + x_k^2 \cdot p_k}$$

• Η διακύμανση της $X = \boxed{E(x^2) - [E(x)]^2 = V(x)}$

$$\boxed{\text{τυπική απόκλιση} = \sqrt{\text{Διακύμανσης}}}$$

Ιδιότητες Διακύμανσης

• $V(a) = 0$

• $V(a \cdot X + \beta) = a^2 \cdot V(x)$

πχ

Ο μέσος όρος των μηνιαίων πωλήσεων ενός καταστήματος είναι 25.000 με τυπική απόκλιση 4000. Αν το μηνιαίο κέρδος είναι το 30% των εσόδων μείον 6.000 € σταθερό κόστος, να βρεθεί το αναμενόμενο κέρδος και η τυπική απόκλιση του κέρδους

Λύση

X πωλήσεις (έσοδα)

$E(x) = 25000$

$V(x) = (\text{τυπική απόκλιση})^2 = 4000^2 = 16.000.000$

$Y = \text{κέρδος} = 0,3 \cdot X - 6.000$

Άρα $E(Y) = E(0,3 \cdot X - 6.000) = 0,3\, E(x) - 6000 = 0,3 \cdot 25000 - 6.000$

$V(Y) = V(0,3X - 6.000) = 0,3^2 \cdot V(x) = 0,09 \cdot 16000.000 = 1.440.000$

Άρα η τυπική απόκλιση του Y είναι 1200

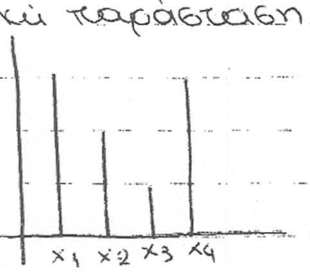
Μάθημα 2ο 11/10/201

X διακριτή τυχαία μεταβλητή

• Συνάρτηση πιθανότητας

$f(x_i)=P(X=x_i)$

γραφική παράσταση



Ιδιότητες:

$0\le f(x_i)\le 1$

$\Sigma f(x_i)=1$ ⇝ γιατί γραφική παράσταση το άθροισμα των μηκών μου κάνει 1.

• Αθροιστική συνάρτηση

$F(x_i)=P(X\le x_i)$

X συνεχής τυχαία μεταβλητή

$P(X=x_i)=0$ εξετάζω τις ανισότητες, δηλαδή διαστήματα και ό συγκεκριμένη τιμή.

Συνάρτηση πιθανότητας καλείται οποιαδήποτε συνάρτηση πληροί τις:

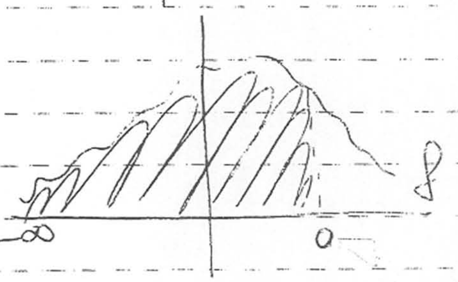
i) $f(x)\ge 0$

ii) $\int_{-\infty}^{+\infty} f(x)\,dx=1$

iii) $P(a\le X\le b)=$ Εμβαδόν $=\int_a^b f(x)\,dx$

Αθροιστική

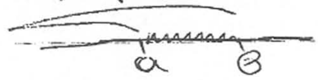
$F(a)=P(X\le a)=$ Εμβαδόν $=\int_{-\infty}^{a} f(x)\,dx$



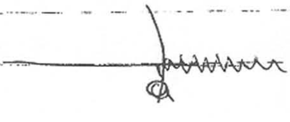
Ιδιότητες

• $0\le F(x)\le 1$

• $P(a\le X\le b)=F(b)-F(a)$ έτοιμα σε πίνακες και εκφράζουν τι γίνεται α -∞ μέχρι ένα σημείο



• $P(X\ge a)=1-F(a)$



πχ

$P(X \leq 4) = F(4)$

$P(X > 2) = 1 - F(2)$

$P(4 \leq X \leq 12) = F(12) - F(4)$

Αν δεν υπάρχει και ισότητα στις ανισότητες, δεν υπάρχει διαφορά, γιατί στη συνεχή μεταβλητή η πιθανότητα να είναι μια συγκεκριμένη τιμή είναι 0.

Έχω διαφορετικούς πίνακες για κάθε περίπτωση (πχ χιλιάδες - εκατομμύρια)

Υπάρχουν κατανομές που καλύπτουν σχεδόν όλες τις περιπτώσεις

Πρόβλημα $\xrightarrow{\text{αντιστοιχώ}}$ θεωρητική κατανομή ⟶ πίνακα με έτοιμα αποτελέσματα

|| Ένα μοντέλο που υπολογίζει τις πιθανότητες και αντιστοιχεί σε διάφορες πραγματικές καταστάσεις

Κατανομές

- Διωνυμική (και κατανομή Bernoulli) } Διακριτές
- Poisson
- Κανονική - συνεχής

Κατανομή Bernoulli: Όταν έχω μια κατάσταση (πείραμα) που έχει μόνο δύο αποτελέσματα (σε περισσότερα δεν δουλεύει) και γίνεται μόνο μία φορά. Τα αποτελέσματα συνήθως ονομάζονται επιτυχία - αποτυχία και συμβολίζονται επιτυχία → 1, αποτυχία → 0

πχ1 νόμισμα: κορώνα γράμματα

πχ2 ισολογισμός εταιρίας < κέρδος / ζημία

πχ3 πρόσληψη προσωπικού < πρόσληψη / όχι

Συμβολισμός πιθανότητας να συμβούν: επιτυχία → p, αποτυχία → q

Ισχύει: $0 \leq p \leq 1$, $0 \leq q \leq 1$, $p + q = 1$

Πίνακας

x_i	Π. Πιθανότητα
1	p
0	q

Συνάρτηση: $f(x) = P(X = x) = p^x \cdot q^{1-x}$

Μέση τιμή: $E(X) = p$ Απόδειξη: $E(X) = 1 \cdot p + 0 \cdot q = p$

Διακύμανση: $V(X) = p \cdot q$

Διωνυμική κατανομή Binomial

- Έχουμε n δοκιμές Bernoulli (ένα πράγμα που έχει δύο μόνο αποτελέσματα, γίνεται πολλές φορές)
- Για κάθε φορά ισχύει p = πιθανότητα επιτυχίας
  $q = 1 - p$ = πιθανότητα αποτυχίας
- κάθε φορά επανάληψης είναι ανεξάρτητη από τις άλλες.

X: πόσες επιτυχίες είχαμε, δηλ. $x = 0, 1, 2, 3, \dots, n$

και $\boxed{P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}}$

Συμβολισμός $B(n, p)$ → πιθανότητα επιτυχίας κάθε φορά
n → αριθμός επαναλήψεων

Μέση τιμή $E(X) = n \cdot p$

Διακύμανση $V(X) = n \cdot p \cdot q$

Παρατήρηση: Αυτά ισχύουν για $p > 0{,}1$ (10%) και $n < 50$

Παράδειγμα

Έχουμε ένα διαγώνισμα πολλαπλής επιλογής με 10 ερωτήσεις. Κάθε ερώτηση έχει 3 απαντήσεις, απ' τις οποίες μία είναι η σωστή. Ένας φοιτητής πηγαίνει σ' εξετάσεις εντελώς αδιάβαστος. Να βρούμε:

i) την πιθανότητα να μην απαντήσει σωστά σε καμία ερώτηση
ii) την πιθανότητα να απαντήσει σωστά 2 ερωτήσεις
iii) την πιθανότητα να κοπεί.

Λύση

10 ερωτήσεις $= n$ — μια ερώτηση:
- απαντάει σωστά (επιτυχία), $p = \frac{1}{3}$
- απαντάει λάθος (αποτυχία), $q = \frac{2}{3}$

Ακολουθεί $B(10, \frac{1}{3})$

X: αριθμός επιτυχιών.

i) $P(X=0) = \binom{10}{0}\left(\frac{1}{3}\right)^0 \cdot \left(\frac{2}{3}\right)^{10} = 1 \cdot 1 \cdot \left(\frac{2}{3}\right)^{10} = 0{,}00003$

ii) $P(X=2) = \binom{10}{2}\left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^8 = \frac{10!}{2!\,8!} \cdot \frac{1}{3^2} \cdot \frac{2^8}{3^8} = \frac{9 \cdot 10}{2} \cdot \frac{2^8}{3^{10}} = 0{,}195 \simeq 20\%$

iii) $P(\text{να αποτύχει}) = P(X \leq 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) \simeq$

κατανομή Poisson

εφαρμόζεται

- Εάν έχουμε προϋποθέσεις διωνυμικής, αλλά p πολύ μικρό ή n πολύ μεγάλο πχ γραμμή παραγωγής και εταιρίες που ελαττωματικά αυξάνονται
- Εάν έχουμε σπάνια γεγονότα που τα εξετάζω για συγκεκριμένη χρονική διάρκεια πχ δολοφονία τον Οκτώβριο στην Πάτρα, πιθανότητα στις εξετάσεις ένα κατάστημα να πουλήσει όλα τα προϊόντα σε μια εβδομάδα

X: αριθμός επιτυχιών

λ: μέσος όρος των επιτυχιών για συγκεκριμένο χρονικό διάστημα

$$\boxed{P(X=x) = e^{-\lambda}\,\frac{\lambda^{x}}{x!}}$$

Παραδείγματα

① Σε μια πόλη διαπράττονται κατά μέσο όρο 5 δολοφονίες το μήνα. Να βρεθεί η πιθανότητα το Φεβρουάριο να γίνουν 4 δολοφονίες

Λύση

$$P(x=4) = e^{-5}\,\frac{5^{4}}{4!} = 0,175 \simeq 17,5\%$$

$$P(x=0) = e^{-5}\cdot\frac{5^{0}}{0!} = e^{-5} = 0,000 \ldots$$ Η πιθανότητα να μην γίνει καμία δολοφονία.

②

Μια εταιρία κατασκευάζει λάμπες. Η πιθανότητα μια λάμπα να είναι ελαττωματική είναι 0,001. Κάνουμε μια δοκιμή στη γραμμή παραγωγής και εξετάζουμε 500 λάμπες. Ποια είναι η πιθανότητα να βρούμε 7 ελαττωματικές λάμπες?

Λύση

μια λάμπα → ελαττωματική $p=0,001$ επιτυχία
μια λάμπα → σωστή αποτυχία

Άρα χρησιμοποιούμε Poisson

λ = μέσος όρος = μέση τιμή $= n\cdot p = 500\cdot 0,001 = 0,5$

500 λάμπες = n $\quad P(x=7) = e^{-0,5}\cdot\dfrac{0,5^{7}}{7!}$ = γνωστό.

Κανονική κατανομή

Συνεχής τυχαία μεταβλητή για την οποία πάντα έχει μέση τιμή $-\infty < \mu \le +\infty$ και τυπική απόκλιση σ, $\sigma > 0$

__Άρα__ μ: προσδιορίζει τη θέση. Αν

$\mu > 0$: θετικά

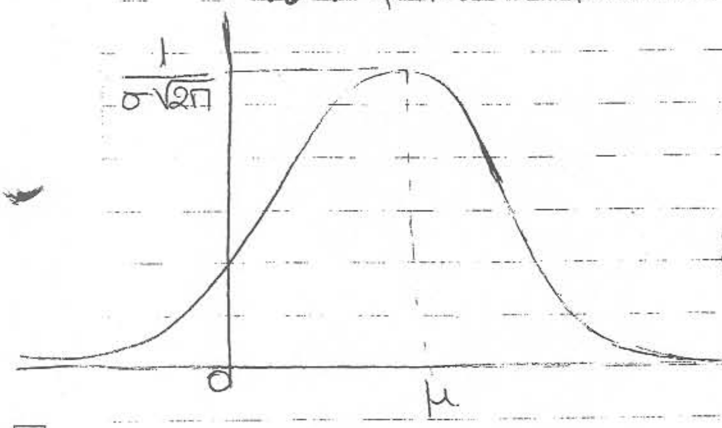
$\mu < 0$: αρνητικά

$\mu = 0$: κέντρο αξόνων, άρα συμμετρική.

σ: προσδιορίζει το ύψος

$\uparrow\sigma$: χαμηλή καμπύλη

$\downarrow\sigma$: ψηλή καμπύλη



__Συνάρτηση__ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

__Συμβολισμός__ $N(\mu, \sigma^2)$

└ διασπορά

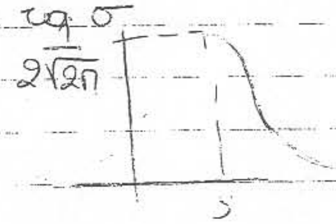
ΟΧΙ τυπική απ.

Ουσιαστικά έχω άπειρες τέτοιες κανονικές κατανομές, ανάλογα τι είναι το μ και το σ

__πχ__ $N(3, 4)$

$\mu = 3$

$\sigma = 2$



__Ιδιότητες__ (κοινές για όλες τις κατανομές)

1) Μορφή καμπύλης

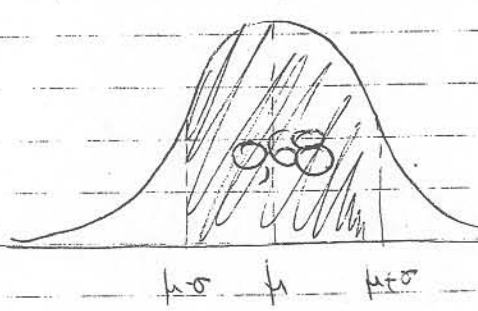
2) Συμμετρικές ως προς την κατακόρυφη ευθεία που διέρχεται από το μ

3) Εμβαδό κάτω από την καμπύλη = 1 (ανεξάρτητα από τον που βρίσκεται η κορυφή. Η θέση της κορυφής καθορίζει το πόσο πλατιά ή λεπτή είναι η

4) $E = P(\mu - \sigma \le x \le \mu + \sigma) = 0,68$

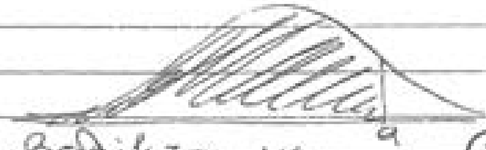
$(\mu - 2\sigma \le x \le \mu + 2\sigma) = 0,954$

$(\mu - 3\sigma \le x \le \mu + 3\sigma) = 0,997$



Παρατήρηση

Για να υπολογίσουμε πιθανότητες για οποιαδήποτε κανονική κατανομή χρησιμοποιούμε τους πίνακες της τυποποιημένης κανονικής κατανομής $N(0,1)$. Ο πίνακας δείχνει την αθροιστική πιθανότητα, δηλαδή την $P(x \leq a)$



Η $N(0,1)$ συμβολίζεται με Z

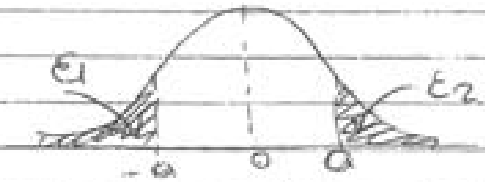
$P(Z \leq 1,25) = 0,89435$

$P(Z \leq 0,17) = 0,56749$

$P(Z \leq 1,7) = 0,95543$

$P(Z \leq -1,13) = P(Z \geq 1,13) = 1 - P(Z \leq 1,13) = 1 - 0,87076 = 0,12924$

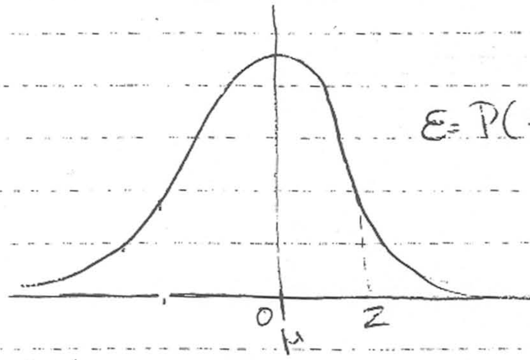
$P(Z < -a)$ = $E_1 = E_2$ (λόγω συμμετρίας) $= P(Z > a) = 1 - P(Z \leq a)$



$P(1,3 \leq Z \leq 2,41) = P(Z \leq 2,41) - P(Z \leq 1,3) = 0,99202 - 0,90320$

<u>Μάθημα 3ο</u> 25/10/15

ΤΥΠΟΠΟΙΗΜΕΝΗ ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ $N(\underset{\mu}{0},\underset{\sigma^2}{1})$



$E = P(x \leq z) \Rightarrow$ απευθείας απ' τον πίνακα

"Μια κανονική κατανομή $N(\mu,\sigma^2)$ να την <u>μετατρέψουμε</u> στην τυποποιημένη, άρα να χρησιμοποιούμε πίνακες για τους υπολογισμούς."

<u>Πχ</u>

Μια τυχαία μεταβλητή X ακολουθεί κανονική κατανομή με μέση $\mu=10$ και διασπορά $\sigma^2=16$. Να βρεθεί $P(x \leq 15)$

<u>Λύση</u>

<u>α' τρόπος</u> $\mu=10$, $\sigma=4$ $\quad P(x \leq 15) = \int_{-\infty}^{15} \frac{1}{4\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-10}{4}\right)^2}$

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

Δεν είναι εύκολος και είναι δύσκολα υπολογίσιμος.

<u>β' τρόπος</u>

Αν θέσουμε $\frac{x-\mu}{\sigma} = z$ τότε η κανονική κατανομή που έχει αντιστοιχεί στην $N(0,1)$ και βρίσκω την πιθανότητα από τον

(θα φτιάξουμε $\frac{x-\mu}{\sigma} = \frac{x-10}{4}$)

$$P(x \leq 15) = P\left(\frac{x-10}{4} \leq \frac{15-10}{4}\right) =$$

$$= P(z \leq 1,25)$$

$$= 0,89435$$

$$\simeq 89\%$$

Παράδειγμα 1

Μια εταιρία κατασκευάζει μπαταρίες με τις προδιαγραφές ο μέσος χρόνος διάρκειας της μπαταρίας σε κατάσταση αναμονής να είναι 191 ώρες και τυπική απόκλιση 5 ώρες, δηλαδή ακολουθούν την N(191,25). Να βρεθούν

i) η πιθανότητα μια μπαταρία να διαρκέσει λιγότερο από 180 ώρες

ii) η πιθανότητα μια μπαταρία να διαρκέσει περισσότερο από 195 ώρες

iii) η πιθανότητα μια μπαταρία να διαρκέσει μεταξύ 180 και 195 ωρών

Λύση

$\mu = 191$ άρα θα χρησιμοποιήσω $\frac{x-\mu}{\sigma} = \frac{x-191}{5} = z$

$\sigma = 5$

$\sigma^2 = 25$

i) $P(x \leq 180) = P\left(\frac{x-191}{5} \leq \frac{180-191}{5}\right)$

$= P(z \leq -2,2) = P(z \geq 2,2) = 1 - P(z \leq 2,2) =$

$= 1 - 0,98610 = 0,0139 \simeq 1,3\%$

ii) $P(x \geq 195) = 1 - P(x \leq 195) = 1 - P\left(\frac{x-191}{5} \leq \frac{195-191}{5}\right) =$

$= 1 - P\left(z \leq \frac{4}{5}\right) = 1 - P(z \leq 0,8)$

$= 1 - 0,78814 = 0,21186 \simeq 21,1\%$

iii) $P(180 \leq x \leq 195) = P(x \leq 195) - P(x \leq 180) =$

$= 0,78814 - 0,0139 =$

$= 0,77424 \simeq 77,4\%$

Παράδειγμα 2

Η βαθμολογία σε κάποιες εξετάσεις ακολουθεί κανονική κατανομή N(550,9

Να βρεθεί η πιθανότητα ένας φοιτητής να περάσει.

Λύση

$\mu = 550$

$\sigma^2 = 900$

$\sigma = \sqrt{900} = 30$

$P(x \geq 500) = P\left(\frac{x-550}{30} \geq \frac{500-550}{30}\right) =$

$= P\left(z \geq -\frac{5}{3}\right) = P(z \geq -1,66) = 1 - P(z \geq 1,66) =$

$= 1 - (1 - P(z \leq 1,66)) = P(z \leq 1,66) =$

$= 0,95154 \simeq 95\%$

Δειγματοληψία

	πληθυσμός	τυχαίο δείγμα
Συμβολισμός	N	n
μέση τιμή	μ	$\bar{x}$
διασπορά	σ^2	s^2
τυπική απόκλιση	σ	s

Βασικά Προβλήματα δειγματοληψίας

i) Αν γνωρίζουμε ένα μέγεθος του πληθυσμού μπορούμε να βρούμε τι θα συμβαίνει σε ένα τυχαίο δείγμα του?

ii) Γνωρίζοντας τι συμβαίνει σε ένα δείγμα του πληθυσμού μπορώ να κάνω εκτίμηση πως θα συμπεριφερθεί ολόκληρος ο πληθυσμός?

Μέθοδοι Λύσης

- να απαντήσουμε με πιθανότητα
- να δώσουμε απάντηση με διάστημα εμπιστοσύνης
- να απαντήσω με έλεγχο υπόθεσης.

κανονική κατανομή πληθυσμού: δείχνει πως συμπεριφέρεται όλος ο πληθυσμ
κανονική κατανομή τυχαίου δείγματος: δείχνει πως συμπεριφέρεται ένα δείγ
Δειγματική κατανομή: δείχνει πως συμπεριφέρονται όλα τα δείγματα τ

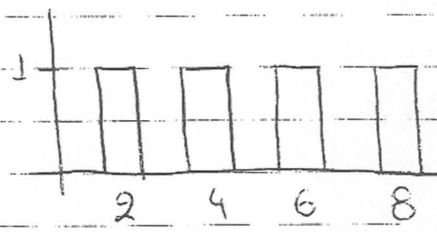
πχ

Έχουμε ένα πληθυσμό με $N=4$ άτομα με στοιχεία: 2, 4, 6, 8 και μας ενδιαφέρει ο μέσος όρος

i) κατανομή του πληθυσμού

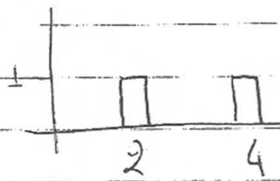
$$\mu = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$$

$$\sigma^2 = 5$$



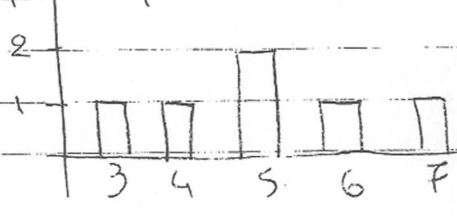
ii) Ένα δείγμα ατόμων, το $\{2,4\}$ να γίνει κατανομή δείγματος

$$\bar{x} = \frac{2+4}{2} = \frac{6}{2} = 3$$



iii) Δειγματική κατανομή παίρνουμε όλα τα δείγματα δύο ατόμων και θα το μέσο όρο

δείγματα	$\bar{x}$
(2, 4)	3
(2, 6)	4
(2, 8)	5
(4, 6)	5
(4, 8)	6
(6, 8)	7



Η τυπική απόκλιση $\sigma_{\bar{x}}$ της δειγματικής κατανομής ονομάζεται τυπικό σφάλμα του μέσου (standard error of mean), μας δείχνει πόσο καλά ο δειγματικός μέσος $\bar{x}$ προσεγγίζει τον πραγματικό μέσο όρο.

Πρόβλημα: "Γνωρίζουμε πως συμπεριφέρεται ένας πληθυσμός και ψάχνουμε τι συμβαίνει σε ένα τυχαίο δείγμα"

Θεώρημα

Κεντρικό Οριακό Θεώρημα (ΚΟΘ)

1) Αν ένας πληθυσμός ακολουθεί την κατανομή $N(\mu, \sigma^2)$ τότε η δειγματική κατανομή του μέσου $\bar{x}$ για δείγμα μεγέθους n προσεγγίζει την κανονική κατανομή με μέσο τον ίδιο $\mu_{\bar{x}} = \mu$ και $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

2) Αν δεν ξέρουμε τι κατανομή ακολουθεί ο αρχικός πληθυσμός αλλά το δείγμα είναι μεγάλο τότε η δειγματική κατανομή του μέσου $\bar{x}$ για δείγμα μεγέθους n προσεγγίζει την κανονική κατανομή με μέσο τον ίδιο $\mu_{\bar{x}} = \mu$ και $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

3) Αν γνωρίζουμε ότι ο πληθυσμός ακολουθεί κανονική κατανομή με μέσο μ αλλά δεν ξέρουμε τη διασπορά, η δειγματική κατανομή του μέσου προσεγγίζει την t-κατανομή.

Πρόβλημα ①

Οι βαθμοί σε ένα test ακολουθούν κανονική κατανομή με $\mu = 500$ και τυπ. απόκλιση $\sigma = 100$. Αν πάρουμε ένα τυχαίο δείγμα 25 ατόμων, ποια είναι η πιθανότητα η μέση τιμή του δείγματος να είναι πάνω από 540?

Απάντηση

πληθυσμός $N(500, 100^2)$. Άρα η δειγματική κατανομή του μέσου ακολουθεί κανονική κατανομή με $\mu_{\bar{x}} = 500$ και $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{5} = 20$

$$P(\bar{x} > 540) = P\left(\frac{\bar{x} - 500}{20} > \frac{540 - 500}{20}\right) = P(z > 2) =$$

$$= 1 - P(z < 2) = 1 - 0{,}97725 =$$

$$= 0{,}02275$$

Πρόβλημα (2)

Η ημερήσια παραγωγή ενός εργοστασίου έχει $\mu=60$ και $\sigma=8$

i) σε ένα δείγμα 4 ημερών πόση είναι η πιθανότητα η μέση τιμή να 'ναι μεγαλύτερη

ii) σε ένα δείγμα 64 ημερών πόση είναι η πιθανότητα η μέση τιμή να 'ναι μικρότερη από

Λύση

i) πληθυσμός $\mu=60$ $\sigma=8$

δείγμα 4 ημερών (μικρό), και επειδή δε γνωρίζουμε τι κατανομή ακ ο πληθυσμός, δεν έχω συμπέρασμα

ii) δείγμα 64 ημερών (μεγάλο), άρα από ΚΟΘ η δειγματική κατανομή του μέσου ακολουθεί κανονική κατανομή με $\mu_{\bar{x}}=60$ και $\sigma_{\bar{x}}=\frac{8}{\sqrt{64}}=\frac{8}{8}$

$$P(\bar{x}<62)=P\left(\frac{\bar{x}-60}{1}<\frac{62-60}{1}\right)=P(z<2)=0,97725 \simeq 97,7\%$$

Πρόβλημα (3)

Το ύψος των μαθητών ακολουθεί κανονική κατανομή με μέση τιμή 16
Ένα τυχαίο δείγμα 15 μαθητών έχει τυπική απόκλιση $s=10$cm. Να β
η πιθανότητα το μέσο ύψος των μαθητών του δείγματος να είναι το πολύ 16

Λύση

$N(\mu=160, \sigma=;)$

άρα η δειγματική κατανομή του μέσου ακολουθεί την t-κατανομή

$s=10$

β. ελ. $= n-1=15-1=14$

$\mu=160$

$\frac{s}{\sqrt{n}}=\frac{10}{\sqrt{15}}=2,58$

t-κατανομή.
S: τυπική απόκλιση δ
βαθμοί ελευθερίας $=\nu$
$\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ ακολουθεί τ κατανομή βαθμούς ελ

$$P(\bar{x}<165)=P\left(\frac{\bar{x}-160}{2,58}<\frac{165-160}{2,58}\right)$$

$$=P(t<1,94)$$

= ανάμεσα στο 97,5% κ' 98,75%

Πινακάκι t-κατανομής → 14 β. ελ. → Το 1,94 βρίσκεται ανάμεσ στο 1,761 και 2,145 που αντιστοιχούν στο 5 και στο 2,5.
Απαντάμε με πιθανότητα δηλαδή στο

$2,145=\frac{2,5}{2}=1,25\%$

$1,761=\frac{5}{2}=2,5\%$

Άρα βρίσκεται ανάμεσα στο
$100\%-1,25\%=98,75\%$
και στο
$100\%-2,5\%=97,5\%$

Άσκηση για σπίτι

Η επίδοση των μαθητών στα μαθηματικά ακολουθεί κανονική κατανομή με μέσο όρο 80. Παίρνουμε δείγμα 20 ατόμων. Να βρεθεί η πιθανότητα η μέση τιμή του δείγματος να είναι μικρότερη από 82 αν γνωρίζουμε ότι η τυπική απόκλιση του δείγματος είναι 5 μονάδες

Απάντηση

Πληθυσμός: κανονική κατανομή, $\mu = 80$, $\sigma^2 = ;$

Δείγμα $n = 20$, $s = 5$ $\quad \dfrac{s}{\sqrt{n}} = \dfrac{5}{\sqrt{20}} = 1,11$

Άρα θα χρησιμοποιήσουμε t-κατανομή με $20 - 1 = 19$ βαθμούς ελευθερίας

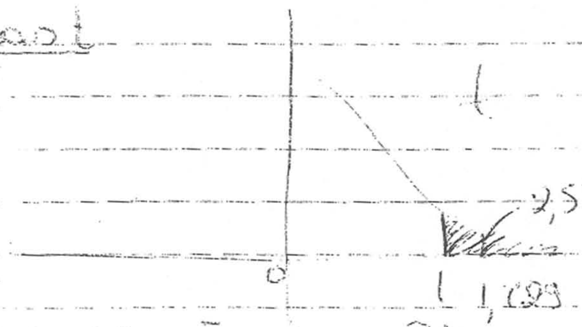
$$P(\bar{x} \le 82) = P\left(\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \le 82\right) = P\left(\frac{\bar{x}-80}{1,11} \le \frac{82-80}{1,11}\right) = P(t \le 1,8)$$

είναι ανάμεσα στο $\alpha = 0,05$ και $\alpha = 0,025$

$\frac{\alpha}{2} = 0,025 \quad \frac{\alpha}{2} = 0,0125$

Άρα η p είναι ανάμεσα στο $1 - 0,025 = 0,975 = 97,5\%$ και $1 - 0,0125 = 0,9875 = 98,75\%$

Πίνακας t

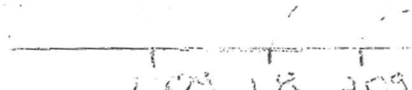


$t_{19;\,0,05} = 1,729$ Τι σημαίνει; Δείχνει την τιμή στον άξονα ώστε πάνω να έχει πιθανότητα $\frac{\alpha}{2}$. Για να έχω την πιθανότητα για μικρότερες τιμές χρησιμοποιώ το $1 - \frac{\alpha}{2}$.

β' τρόπος

$> 1,729 \Rightarrow P = \alpha/2 = 0,05/2$

$> 2,093 \Rightarrow P = \frac{\alpha}{2} = 0,025/2$



$P(t \ge 1,8)$ είναι ανάμεσα στο $\frac{0,05}{2}$ και στο $\frac{0,025}{2}$

$P(t \le 1,8) = 1 - \frac{0,05}{2}$ και $1 - \frac{0,025}{2}$

Μάθημα 4ο 1/11/2012

κατηγορίες

① Ξέρουμε στοιχεία για ολόκληρο τον πληθυσμό και ψάχνουμε τι θα συμβαίνει σε ένα τυχαίο δείγματου και συγκεκριμένα τι θα συμβαίνει με τη **μέση** τιμή

Τότε: Χρησιμοποιούμε δειγματική κατανομή του μέσου

→ Είναι κάτι που θα μου δείξει τι μπορεί να συμβαίνει σε μικρό κομμάτι αν ξέρω τι γίνεται στον πληθυσμό.

πχ1 Ξέρουμε τα ποσοστά των κομμάτων στις εκλογές. Μπορώ να βρω τι θα γίνει εκλογές του ΤΕΙ της Πάτρας?

πχ2 Μετράμε τις πωλήσεις κινητών μέσω εταιρίες σε όλη την Ελλάδα. Τι γίνεται στα καταστήματα της Αχαΐας?

Ανάλογα με τα δεδομένα έχουμε 3 τρόπους επίλυσης

- (ΚΟΘ) Αν ο πληθυσμός ακολουθεί κανονική κατανομή με γνωστή μέση τιμή μ και διασπορά σ^2 **τότε** η δειγματική κατανομή του μέσου ακολ. κανονική κατανομή με μέσο μ και διασπορά $\frac{\sigma^2}{n}$ όπου n=μέγεθος δείγ.
- (ΚΟΘ) Αν δε ξέρουμε τι κατανομή ακολουθεί ο πληθυσμός αλλά το δείγμα είναι μεγάλο (εξαρτάται από το μέγεθος του πληθυσμού, δηλαδή ανάλογα μετράμε) **τότε** ισχύει το προηγούμενο, δηλαδή το δείγμα ακολουθεί κανονική κατανομή με μέσο μ και διασπορά $\frac{\sigma^2}{n}$.
- Αν ο πληθυσμός ακολουθεί κανονική κατανομή με γνωστή μέση τιμή και άγνωστη διασπορά **τότε** θα χρησιμοποιήσουμε τη διασπορά S^2 του δείγματος. Η δειγματική κατανομή του μέσου ακολουθεί την t-κατα με n-1 βαθμούς ελευθερίας **συγκεκριμένα** $\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$

Άσκηση

κοιτάζω τον πινακάκι Z.

Ξέρουμε ότι το IQ ακολουθεί κανονική κατανομή με μέση τιμή 100 και τυπική απόκ. Έχουμε ένα τυχαίο δείγμα 100 ατόμων. Να βρεθεί η πιθανότητα η μέση τιμή του δείγματος να είναι μικρότερη του 90.

Απάντηση

πληθυσμός: κανονική κατανομή με $\mu=100$, $\sigma=15 \Rightarrow \sigma^2=15^2=225$

Ψάχνω για ένα δείγμα, ακολουθεί δειγματική κατανομή $\mu_{\bar{x}}=100$ $\sigma^2_{\bar{x}}=\frac{15^2}{\underbrace{100}_{\text{πλήθος}}} \Rightarrow \sigma_{\bar{x}}=\frac{15}{10}=$

μετατροπή σε ζ τιμή $\frac{x-\mu}{\sigma}$

$$P(\bar{x}<90)=P\left(\frac{\bar{x}-100}{1,5}<\frac{90-100}{1,5}\right)=P(Z<-6,6)=1-P(Z<6,6)=1-0,99 \simeq 1\%$$

Στον πίνακα Z φτάνει μέχρι το 3,59. Από 'κει και πάνω οι τιμές είναι ίδιες (=0,9983).

Πχ

Θέλω $P(t \geq 2)$ όταν έχει 4 βαθμούς ελευθερίας

$0{,}1/2$ $\quad$ $0{,}05/2$

$1{,}533 \quad 2 \quad 2{,}132$ $\quad$ Άρα $\frac{0{,}10}{2}$ και $\frac{0{,}05}{2}$ $\Rightarrow$ 5% και 2,5%

Η z δείχνει πιθανότητες, η t δείχνει τον άξονα και βρίσκω τις πιθανότητες

(2) Ζητάμε τη μέση τιμή μιας μεταβλητής (γενικά). Εξετάζουμε δύο δείγματα. Ψάχνουμε τη διαφορά των μέσων $\bar{x}_A - \bar{x}_B$

Δειγματική κατανομή της διαφοράς των μέσων (εξετάζουμε [illegible]

- Αν το δείγμα A ακολουθεί κανονική κατανομή με μέση τιμή μ_A και διασπορά σ_A^2 $(A \sim N(\mu_A, \sigma_A^2))$ και το δείγμα B ακολουθεί κανονική κατανομή με μέση τιμή μ_B και διασπορά σ_B^2 $(B \sim N(\mu_B, \sigma_B^2))$ **Τότε** η διαφορά $\bar{X}_A - \bar{X}_B$ ακολουθεί κανονική κατανομή με $\mu = \mu_A - \mu_B$ και $\sigma^2 = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$
- Αν οι αρχικοί πληθυσμοί δεν είναι κανονικοί αλλά τα δείγματα μεγάλα, **τότε** ισχύει το ίδιο

Πχ

Εξετάζουμε τους μαθητές δύο σχολών A και B ως προς το βάρος και υποθέτουμε ότι η κατανομή των βαρών και στις δύο σχολές είναι κανονική με $\mu_A = 72kg$ και $\sigma_A^2 = 4$ και $\mu_B = 71\ kg$ και $\sigma_B^2 = 3$. Παίρνουμε ένα τυχαίο δείγμα 40 μαθητών από τη σχολή A και 50 μαθητών από τη σχολή B. Να βρεθεί η πιθανότητα η διαφορά των μέσων στα δείγματα να είναι μεταξύ 1 και 2 κιλών

Απάντηση

Σχολή A: $\mu_A = 72$ $\sigma_A^2 = 4$ $n_A = 40$, Σχολή B $\mu_B = 71$ $\sigma_B^2 = 3$ $n_B = 50$

Διαφορά των μέσων $\bar{x}_A - \bar{x}_B$ ακολουθεί κανονική κατανομή με $\mu = \mu_A - \mu_B = 72 - 71 = 1$ και $\sigma^2 = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{4}{40} + \frac{3}{50} = 0{,}16$

Άρα $\sigma = \sqrt{0{,}16} = 0{,}4$

$P(1 < \bar{x}_A - \bar{x}_B < 2) = P\left(\frac{1-1}{0{,}4} < z < \frac{2-1}{0{,}4}\right) = P(0 < z < 2{,}5) =$

$= P(z < 2{,}5) - P(z < 0) = 0{,}99379 - 0{,}5 = 0{,}49379$

③ Ξέρουμε ότι όταν ο πληθυσμός ψάχνουμε ποσοστό σε ένα δείγμα του π άτομα τότε χρησιμοποιούμε δειγματική κατανομή του ποσοστού.

• Αν ο πληθυσμός ακολουθεί κανονική κατανομή και γνωρίζουμε ποιο ποσοστό του p έχει μια ιδιότητα, θεωρούμε δείγμα n ατόμων. Μας ενδιαφέρει το ποσοστό του δείγματος με την ίδια ιδιότητα. Ακολουθεί κανονική κατανομή με μέση τιμή $\mu = p$ και διασπορά $\sigma^2 = \frac{p\cdot(1-p)}{n}$

ΠΧ

Το 10% των ατόμων παρακολουθεί ένα συγκεκριμένο δελτ ειδήσεων. Παίρνουμε ένα τυχαίο δείγμα 500 ανθρώπ Να βρεθεί η πιθανότητα το δείγμα να περιλαμβάνει τηλεθεατές αυτού του δελτίου σε ποσοστό πάνω από 1!

Απάντηση

πληθυσμός
10% βλέπει συγκεκριμένο δελτίο
$\parallel$
$p = 0{,}10$

Δείγμα $n = 500$

Η δειγματική κατανομή του ποσοστού $\bar{p}$ είναι

κανονική με μέση τιμή $\mu = p = 0{,}10$ και

διασπορά $\sigma^2 = \frac{p(1-p)}{n} = \frac{0{,}10\ 0{,}90}{500} = 0{,}00018$

Άρα $\sigma = \sqrt{0{,}00018} = 0{,}0134$

$$P(\bar{p} \geq 0{,}15) = 1 - P(\bar{p} \leq 0{,}15)$$
$$= 1 - P\left(Z \leq \frac{0{,}15 - 0{,}10}{0{,}0134}\right)$$
$$= 1 - P(Z \leq 3{,}73$$
$$= 1 - 0{,}999 = 1‰$$

2ο κεφάλαιο

Διάστημα εμπιστοσύνης (Δ.Ε)

Δ.Ε για το μέσο

Ξέρουμε ένα δείγμα (πλήρως) και ψάχνουμε τη μέση τιμή όλου του πληθυσμού

ΔΕΝ ΓΙΝΕΤΑΙ γιατί δεν έχω ασφαλές συμπέρασμα.

Τι γίνεται?

θα έχουμε δηλαδή ένα ποσοστό λάθους

Ξέρουμε ένα δείγμα και θα εκτιμήσουμε ένα διάστημα μέσα στο οποίο θα βρίσκεται η μέση τιμή όλου του πληθυσμού.

α = Επίπεδο σημαντικότητας: Η πιθανότητα να έχουμε δυνατή λάθος διάστημα, δηλαδή η πιθανότητα η μέση τιμή να είναι έξω από το διάστημα που υπολογίσαμε (ποσοστό λάθους)

α = 5%, 3%, 1%, 1‰ → πιο σίγουρο αποτέλεσμα

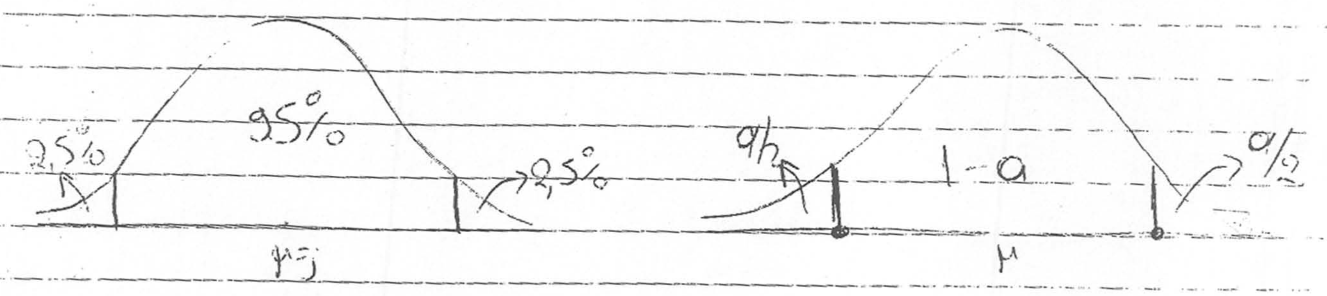
↳ όταν η άσκηση δεν μου δίνει α θεωρώ το 5%

γ = Βαθμός εμπιστοσύνης = 1 − α: Η πιθανότητα η μέση τιμή να είναι μέσα στο διάστημα που βρήκαμε

Περιπτώσεις

① Έχουμε κανονικό πληθυσμό με γνωστή διασπορά σ^2 και ψάχνουμε Δ.Ε για τη μέση τιμή του. Εάν ξέρουμε δείγμα n ατόμων, βρίσκουμε τη μέση τιμή του $\bar{x}$ και $ΔΕ = \left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$

$P(-z_{\frac{\alpha}{2}} \leq z \leq z_{\frac{\alpha}{2}})$



ΠΧ

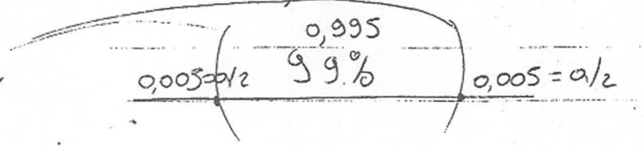
Το εβδομαδιαίο ποσό που χρειάζεται μια οικογένεια για το super market ακολουθεί κανονική κατανομή με διασπορά 144 €. Έχω ένα δείγμα 36 οικογενειών για τις οποίες βρήκαμε ότι ξοδεύουν κατά μέσο όρο 70 € την εβδομάδα. Να υπολογιστεί Δ.Ε για το μέσο ποσό που ξοδεύει μια οικογένεια με ε.σ 1%

Απάντηση

πληθυσμός: εβδομαδιαίο ποσό για ψώνια $\sim N(\,;\,, \sigma^2 = 144)$

δείγμα $n = 36$ οικογένειες $\bar{x} = 70$

$a = 1\% = 0{,}01$



$z_{\frac{a}{2}} = z_{0{,}005} = 2{,}58$

$$\Delta E = \left(\bar{x} - z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \ \bar{x} + z_{\frac{a}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$= \left(70 - 2{,}58 \cdot \frac{12}{\sqrt{36}}, \ 70 + 2{,}58 \cdot \frac{12}{\sqrt{36}} \right)$$

$$= (70 - 5.16, \ 70 + 5.16)$$

$$= (64.84, \ 75.16)$$

Με 99% σιγουριά η μέση τιμή του πληθυσμού θα είναι στο διάστημα (64.84, 75.16)

Μάθημα 5ο — 8/11/2012

Άσκηση (1)

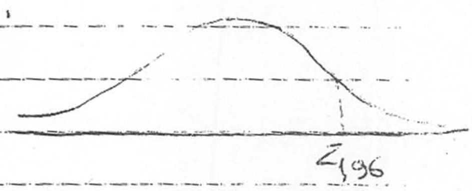
Η βαθμολογία στα μαθηματικά ακολουθεί κανονική κατανομή με διακύμανση 9. Εξετάσαμε ένα δείγμα 16 σπουδαστών και βρήκαμε μέση τιμή 6,3. Να βρεθεί διάστημα εμπιστοσύνης 95% για τη μέση τιμή όλων των σπουδαστών.

Λύση

δείγμα: $n=16$, $\bar{x}=6,3$

πληθυσμός: $\sigma^2=9$, ΔΕ για το μέσο, κανονικότητα

$\alpha=5\%=0,05$, $\frac{\alpha}{2}=0,025$, $Z_{\frac{\alpha}{2}}=1,96$



$1-0,025=0,975$

$$ΔΕ=\left[6,3-1,96\cdot\frac{3}{4},\ 6,3+1,96\cdot\frac{3}{4}\right]$$

$$ΔΕ=[6,3-1,88,\ 6,3+1,88]$$

$$ΔΕ=[4,42,\ 8,18]$$

Άσκηση (2)

Οι πωλήσεις μιας βιομηχανίας ακολουθούν κανονική κατανομή με τυπική απόκλιση 12000 €. Έχουμε δείγμα 25 πωλήσεων με μέση τιμή 81000€. Να βρεθεί 99% διάστημα εμπιστοσύνης για τη μέση τιμή των πωλήσεων.

Λύση

δείγμα: $n=25$, $\bar{x}=81000$

πληθυσμός: $\sigma=12000$, ΔΕ για το μέσο, κανονικότητα

$\alpha=1\%=0,01$ $\alpha/2=0,005$ $Z_{\alpha/2}=2,58$

$$ΔΕ=\left[81.000-2,58\cdot\frac{12000}{5},\ 81.000+2,58\cdot\frac{12.000}{5}\right]$$

$$ΔΕ=[74.808,\ 87.192]$$

Άσκηση (3)

Από ένα κανονικό πληθυσμό παίρνουμε δείγμα 16 ατόμων και βρίσκουμε μέση τιμή 14,5 και τυπική απόκλιση 5. Να βρεθεί 90% ΔΕ για το μέσο του πληθυσμού.

Λύση

δείγμα: $n=16$, $\bar{x}=14,5$, $s=5$, $s^2=25$ $\alpha=10\%=0,1$, $\alpha/2=0,05$, βε $=n-1$

πληθυσμός: κανονικότητα, ΔΕ για το μέσο   Άρα $t_{15;0,05}=1,753$

$$ΔΕ=\left[14,5-1,753\cdot\frac{5}{4},\ 14,5+1,753\cdot\frac{5}{4}\right]=[14,5-2,2,\ 14,5+2,2]=$$

$$=[12,3,\ 16,7]$$

Άσκηση 4

Ένας πληθυσμός έχει άγνωστη διασπορά αλλά ακολουθεί κανονική κατανομή. Σε ένα τυχαίο δείγμα 18 ατόμων η μέση τιμή είναι 14,5 και η διασπορά 1. Να βρεθεί 95% διάστημα εμπιστοσύνης για τη μέση τιμή του πληθυσμού.

Λύση

Δείγμα: $n=18$, $\bar{x}=14,5$, $S^2=1$

πληθυσμός: κανονικότητα, ΔΕ για το μέσο

$a=5\%=0,05$, $a/2=0,025$

$B.E.=n-1=17$

$t_{17,\,0.025}=2,110$

$$\Delta E=\left[\bar{x}-t\cdot\frac{S}{\sqrt{n}},\ \bar{x}+t\cdot\frac{S}{\sqrt{n}}\right]=\left[14,5-2,11\cdot\frac{1}{\sqrt{18}},\ 14,5+2,11\cdot\frac{1}{\sqrt{18}}\right]=$$

$$=[14,5-0,5,\ 14,5+0,5]=[14,\ 15]$$

Διάστημα εμπιστοσύνης για τη διαφορά των μέσων δύο κανονικών πληθυσμών με γνωστή διακύμανση.

Παρατήρηση: Εάν έχουμε μεγάλα δείγματα, αλλά δεν ξέρουμε την κανονικότητα θα πάρουμε τον ίδιο τύπο (δεν παίζει ρόλ η κανονικότητα αφού έχω μεγάλα δείγματα)

$$\left[(\bar{x}_1-\bar{x}_2)-Z_{a/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ (\bar{x}_1-\bar{x}_2)+Z_{a/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

Άσκηση

Σε δύο χώρες πωλείται ένα προϊόν με διαφορετική μέση τιμή. Η διακύμανση της τιμής του στην χώρα A είναι 200 και στη χώρα B είναι 550. Πήραμε δύο τυχαία δείγματα 45 και 55 τιμών αντίστοιχα από τις δύο χώρες για τα οποία βρήκαμε μέσες τιμές 120 και 100 αντίστοιχα. Να υπολογιστεί 99% διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών στις δύο χώρες.

Λύση

πληθυσμός δείγμα

Χώρα A: $\sigma^2=200$, $n_A=45$, $\bar{x}_A=120$

Χώρα B: $\sigma^2=550$, $n_B=55$, $\bar{x}_B=100$

Επειδή τα δείγματα είναι μεγάλα μπορώ να χρησιμοποιήσω τον τύπο για την κανονικότητα (με το Z)

$\bar{X}_A - \bar{X}_B = 20$, $\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{14,44} = 3,8$

$a = 1\% = 0,01 \Rightarrow a/2 = 0,005 \rightarrow Z_{a/2} = 0,995 = 2,58$

$$ΔΕ = \left[20 - 2.58 \cdot 3.8 \,,\; 20 + 2.58 \cdot 3.8\right]$$

$$= [10.196 \,,\; 29.804]$$

Άσκηση

Ρωτήσαμε 100 καταναλωτές και οι 22 από αυτούς προτιμούν ένα προϊόν Α. Να βρεθεί σε επίπεδο σημαντικότητας 5% το διάστημα εμπιστοσύνης για το ποσοστό όλων των καταναλωτών που θα προτιμούσαν το προϊόν Α.

Λύση

δείγμα: $n = 100$, $p = \frac{22}{100} = 0,22$ $q = 1 - p = 0,78$

$a = 5\% = 0,05$, $a/2 = 0,025 \Rightarrow Z_{a/2} = 1.96$

$$ΔΕ = \left[0,22 - 1.96\sqrt{\frac{0,22 \cdot 0,78}{100}} \,,\; 0.22 + 1.96 \cdot \sqrt{\frac{0,22 \cdot 0,78}{100}}\right] =$$

$$= [0,22 - 1.96 \cdot 0,041 \,,\; 0,22 + 1,96 \cdot 0,041] =$$

$$= [0,22 - 0,08 \,,\; 0,22 + 0,08]$$

$$= [0.14 \,,\; 0.3]$$

Έλεγχος Υποθέσεων για την παράμετρο ενός πληθυσμού

Προϋποθέσεις → κανονικότητα πληθυσμού
→ τυχαίο δείγμα

Έλεγχος: κάνουμε εμείς μια υπόθεση και θα ελέγξουμε την πιθανότητα να είναι σωστή.

Διατύπωση: H_0 = μηδενική υπόθεση
H_1 = εναλλακτική υπόθεση

- Δεν πρέπει να υπάρχει περίπτωση το H_0 και το H_1 να συμβαίνουν ταυτόχρονα
- H_0 να είναι πάντα μονοσήμαντα καθορισμένη

a = επίπεδο σημαντικότητας (εκφράζει την πιθανότητα να κάνω λάθος)
10%, 5%, 3%, 1%, 1‰ → πιο σίγουρο αποτέλεσμα

Έλεγχος για το μέσο ενός πληθυσμού

κάν. πληθ. με γνωστή σ → Ζ-τ
καν. πληθ. με άγνωστη → t-τ

Διατυπώσεις:
αμφίπλευρος έλεγχος: $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
μονόπλευροι έλεγχοι: $H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$

Έλεγχος για το μέσο ενός κανονικού πληθυσμού με γνωστή διακύμανση

- Θα χρησιμοποιήσουμε z-τιμο
Δείγμα: $\bar{x}, n$
πληθυσμός: σ^2
υπολογίζουμε: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ Υπολογίζουμε και το $z_{a/2}$

αμφίπλευρος: $H_0: \mu = \mu_0$ και $H_1: \mu \neq \mu_0$

Z

$-z_{a/2}$ $z_{a/2}$

Υπολογίζουμε και το $Z_{a/2}$

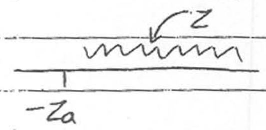
Συγκρίνουμε το Z με το $Z_{a/2}$
- αν $|z| < Z_{a/2}$ τότε αποδεχόμαστε την H_0
- αν $|z| > Z_{a/2}$ τότε απορρίπτουμε την H_0

μονόπλευρος: 1) $H_0: \mu=\mu_0$ και $H_1: \mu>0$

Υπολογίζω το Z_a
Συγκρίνω το Z με το Z_a

- αν $Z<Z_a$ τότε αποδέχομαι την H_0
- αν $Z>Z_a$ τότε απορρίπτω την H_0

2) H_0 $\mu=\mu_0$ και H_1: $\mu<0$



Υπολογίζω το Z_a
Συγκρίνω το Z με το $-Z_a$
αν $Z>-Z_a$ τότε αποδέχομαι την H_0
αν $Z<-Z_a$ τότε απορρίπτω την H_0

Άσκηση

Ψάχνουμε εάν οι φοιτητές διαβάζουν την ημέρα πριν την εξέταση. Ένα δείγμα 100 φοιτητών έδειξε ότι διαβάζουν κατά μέσο όρο 2,8 ώρες πριν τις εξετάσεις. Εάν οι ώρες διαβάσματος ακολουθούν κανονική κατανομή με τυπική απόκλιση μισή ώρα να ελεγχθεί σε επίπεδο σημαντικότητας 5% εάν ο μέσος όρος όλων των φοιτητών είναι 2.5 ώρες διαβάσματος πριν τις εξετάσεις.

Λύση

πληθυσμός: κανονικότητα, $\sigma=0{,}5$
δείγμα $n=100$, $\bar{x}=2.8$
έλεγχος $\alpha=5\%$ εάν το $\mu=2.5$
Διατύπωση H_0 $\mu=2.5$ H_1 $\mu\neq2.5$

$\alpha=5\% \Rightarrow \alpha/2=0{,}025 \Rightarrow Z_{\alpha/2}=1.96$

$$Z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\frac{2.8-2.5}{\frac{0{,}5}{\sqrt{100}}}=\frac{0{,}3}{\frac{0{,}5}{10}}=\frac{3}{0{,}5}=6$$

Άρα απορρίπτω την H_0

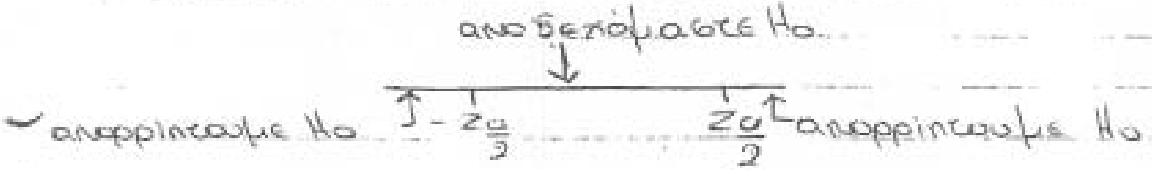
Μάθημα 6ο 22/11/201

1) Έλεγχος για το μέσο, όταν ακολουθείται κανονική κατανομή με άγνωστη μέση τιμή και γνωστή διασπορά

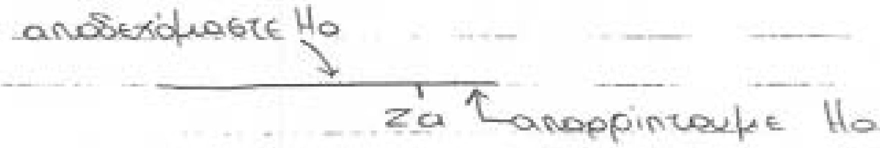
• Υπολογίζουμε $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$

• Συγκρίνουμε το z που βρήκαμε με μια τιμή του πίνακα (εξα από το τι έλεγχο έχουμε)

(A) αμφίπλευρος: $H_0: \mu=\mu_0$ και $H_1: \mu \neq \mu_0$
τότε συγκρίνουμε με το $z_{\frac{a}{2}}$



(B) μονόπλευρος 1) $H_0: \mu=\mu_0$ και $H_1: \mu > \mu_0$
τότε συγκρίνουμε με το z_a



2) $H_0: \mu=\mu_0$ και $H_1: \mu < \mu_0$
τότε συγκρίνουμε με το $-z_a$

Παράδειγμα 1

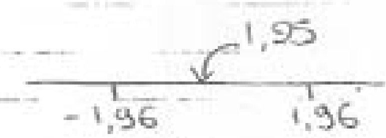
Πήραμε ένα τυχαίο δείγμα n ατόμων από έναν κανονικό πληθυσμό με δ. $\sigma^2=144$. Τους ρωτήσαμε για το μηνιαίο ποσό που ξοδεύουν για διασκέδαση. Ο μέσος των 9 ατόμων του δείγματος είναι 60€. Να ελέγξετε σε επίπεδο σημαντικότητας εάν η μέση τιμή του πληθυσμού είναι 55

Λύση

$\sigma^2=144 \Rightarrow \sigma=12$, $n=9$, $\bar{x}=60$

Αμφίπλευρος $H_0: \mu=55$ κ' $H_1: \mu \neq 55$

$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{60-55}{\frac{12}{\sqrt{9}}} = \frac{5}{4} = 1{,}25$



Άρα αποδεχόμαστε

$a=5\%=0{,}05$ $1-\frac{a}{2}=1-0{,}025=0{,}975$ $z_{\frac{a}{2}}=1{,}96$

Παράδειγμα ②

Ένα τυχαίο δείγμα 25 φοιτητών μας έδειξε ότι ο μέσος όρος εργασίας τους είναι 28 ώρες την εβδομάδα. Γνωρίζουμε ότι η κατανομή του χρόνου εργασίας είναι κανονική με τυπική απόκλιση 10. Να εξεταστεί εάν οι φοιτητές εργάζονται συνολικά λιγότερο από 32 ώρες την εβδομάδα.

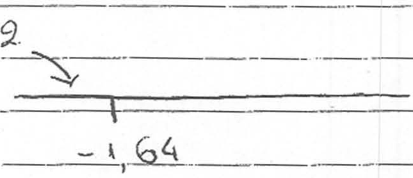
Λύση

$n=25$, $\bar{x}=28$, $\sigma=10$

$\alpha=0{,}05$ $\quad 1-\alpha=0{,}95$ $\quad Z_\alpha=1{,}64$

Μονόπλευρος: $H_0: \mu=32$ κ' $H_1: \mu<32$

$$Z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\frac{28-32}{\frac{10}{\sqrt{25}}}=\frac{-4}{2}=-2$$



Άρα απορρίπτουμε την H_0, δηλαδή οι φοιτητές εργάζονται λιγότερο από 32 ώρες την εβδομάδα, με περιθώριο λάθους 5%.

Παράδειγμα ③ Παλιό Θέμα

Ένα δείγμα 18 ναυτικών ρωτήθηκε για το μέσο χρόνο υπηρεσίας του πριν συνταξιοδοτηθεί. Ο μέσος όρος βρέθηκε 25 χρόνια. Γνωρίζουμε ότι ο χρόνος υπηρεσίας των ναυτικών ακολουθεί κανονική κατανομή με τυπική απόκλιση 3,5 χρόνια. Να ελεγχθεί εάν ο μέσος χρόνος υπηρεσίας είναι μεγαλύτερος από 23 χρόνια.

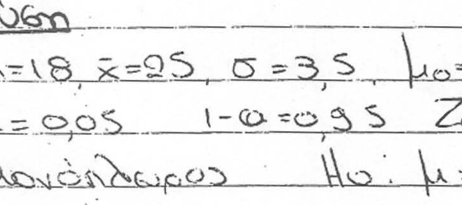
Λύση

$n=18$, $\bar{x}=25$, $\sigma=3{,}5$, $\mu_0=23$

$\alpha=0{,}05$ $\quad 1-\alpha=0{,}95$ $\quad Z_\alpha=1{,}64$

Μονόπλευρος $H_0: \mu=23$ κ' $H_1: \mu>23$

$$Z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\frac{25-23}{\frac{3{,}5}{\sqrt{18}}}=\frac{25-23}{\frac{3{,}5}{4{,}2}}=\frac{2}{0{,}8}=2{,}43$$



Άρα απορρίπτουμε την H_0 οπότε ο μέσος χρόνος υπηρεσίας είναι μεγαλύτερος από 23 χρόνια.

2) Έλεγχος για το μέσο, όταν ακολουθείται κανονική κατανομή με άγνωστη μέση τιμή και άγνωστη διασπορά.

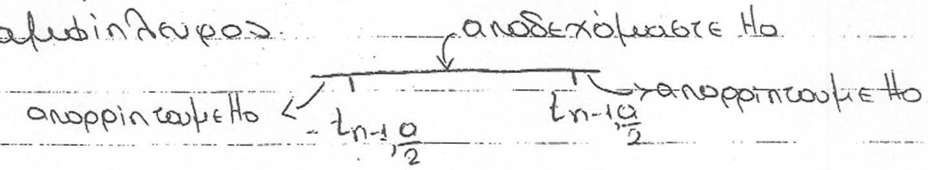
α) το δείγμα είναι μεγάλο (>30), τότε χρησιμοποιούμε ως σ, την τυπική απόκλιση του δείγματος και κάνουμε ό,τι κάναμε στον πρώτο έλεγχο.

β) το δείγμα είναι μικρό (<30), τότε θα χρησιμοποιούμε την t-κατανομή

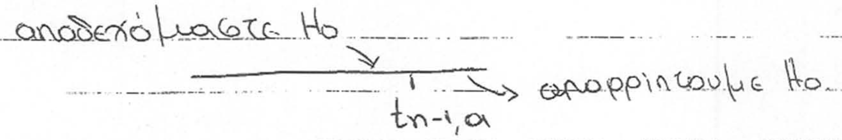
• Υπολογίζουμε $t = \dfrac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}}$

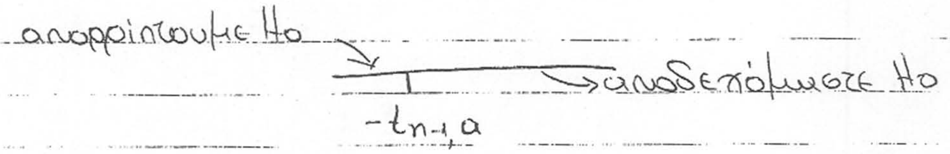
• Συγκρίνουμε ανάλογα με τον έλεγχο που έχουμε, με μια τιμή του πίνακα t με n-1 βαθμούς ελευθερίας.

i) αμφίπλευρος



ii) μονόπλευρος





Παράδειγμα ①

Μια βιομηχανία θέλει να ελέγξει εάν η διανεμητική τιμή ενός προϊόντος της είναι γύρω στα 600 €. Ένα τυχαίο δείγμα 10 σημείων πώλησης έδωσε μέση τιμή 598,8 και τυπική απόκλιση 5,6. Γνωρίζοντας ότι οι τιμές ακολουθούν κανονική κατανομή να ελεγχθεί ο ισχυρισμός της βιομηχανίας σε επίπεδο σημαντικότητας 1%.

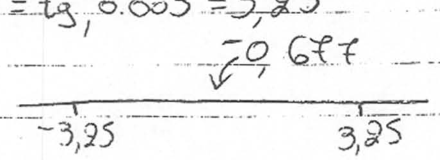
Λύση

κανονικότητα, έλεγχος για το μέσο $H_0: \mu=600$ κ' $H_1: \mu \neq 600$

$\sigma = ;$, $n=10$ (μικρό), $\bar{x}=598,8$ $S=5,6$

Υπολογίζουμε $t = \dfrac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}} = \dfrac{598,8-600}{\frac{5,6}{\sqrt{10}}} = -0,677$

$t_{n-1,\frac{a}{2}} = t_{9,0.005} = 3,25$



Άρα αποδεχόμαστε την Ho, δηλαδή η διανεμητική τιμή του προϊόντος είναι 600€ με περιθώριο λάθους 1%.

Άσκηση

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τις τιμές 42 52 40 74 Να ελεγχθεί σε επίπεδο σημαντικότητας 5% εάν η μέση τιμή του πληθυσμού είναι μεγαλύτερη από 50

Λύση

κανονικότητα, $\sigma = ;$, $n = 4$ (μικρό), $\bar{x} = \frac{\Sigma x_i}{n} = 52$

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = 242,66 \Rightarrow s = 15,57$$

στον πληθυσμό διαιρώ με n δηλ. $\sigma^2 = \frac{\Sigma (x_i - \bar{x})}{n}$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{42+52+40+74}{4} = \frac{208}{4} = 52$$

$$s^2 = \frac{\Sigma (x_i - \bar{x})}{n-1} =$$

$$= \frac{(42-52)^2 + (52-52)^2 + (40-52)^2 + (74-52)^2}{4-1} =$$

$$= \frac{(-10)^2 + 0^2 + (-12)^2 + 22^2}{3} = \frac{100+144+484}{3} = \frac{728}{3} = 242,66$$

ή

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
42	-10	100
52	0	0
40	12	144
74	22	484
208		728

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = \frac{728}{3} = 242,66$$

$$s = \sqrt{242,66} = 15,57$$

$H_0: \mu = 50$ κ' $H_1: \mu \neq 50$

0,257

2,353

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{52-50}{\frac{15,57}{\sqrt{4}}} = 0,257$$

$$t_{n-1,a} = t_{3,\,0.05} = 2,353$$

Άρα αποδεχόμαστε την H_0, οπότε η μέση τιμή του πληθυσμού δεν είναι μεγαλύτερη από 50, με περιθώριο λάθους 5%

3) Έλεγχος για τη διαφορά των μέσων τιμών δύο κανονικών πληθυσμών (ανεξάρτητα δείγματα)

αμφίπλευρος: $H_0: \mu_1=\mu_2$ κ' $H_1: \mu_1 \neq \mu_2$
μονόπλευρος: $H_0: \mu_1=\mu_2$ κ' $H_1: \mu_1 < \mu_2$
$H_0: \mu_1=\mu_2$ κ' $H_1: \mu_1 > \mu_2$

α) Αν έχουμε γνωστή διασπορά χρησιμοποιούμε z-τιμές
- Υπολογίζουμε $z = \dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$
- Συγκρίνουμε με την αντίστοιχη z-τιμή.

β) i) Αν έχουμε άγνωστη διασπορά και μικρά δείγματα χρησιμοποιούμε t-τιμές
- Υπολογίζουμε $t = \dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$ με β.ε $= n_1+n_2-2$
- Συγκρίνουμε με την αντίστοιχη t-τιμή

ii) Αν έχουμε άγνωστη διασπορά και τα δύο δείγματα είναι μεγ (όχι απαραίτητα) τότε χρησιμοποιούμε πάλι z-τιμές με την τυπική απόκλιση του δείγματος που, σ' αυτή την περίπτωση, εκφράζει την τυπική απόκλιση του πληθυσμού

Παράδειγμα 1

Δύο πολυκαταστήματα θέλουν να συγκρίνουν τις μέσες ημερήσιες πωλήσεις. Οι πωλήσεις του πρώτου ακολουθούν κανονική κατανομή με τυπική απόκλιση και οι πωλήσεις του δεύτερου ακολουθούν κανονική κατανομή με τυπική απόκλιση 18. Ένα δείγμα 100 ημερών από το πρώτο έδωσε μέση τιμή 300 κ' ένα δείγμα 75 ημερών από το δεύτερο έδωσε μέση τιμή 280

Λύση

Δυο πληθυσμοί (ανεξάρτητοι), κανονικοί, $\sigma_1=30$, $\sigma_2=18$ ⇒ z-τιμές

1ο δείγμα $n_1=100$ $\bar{x}_1=300$ — $a=5\% \Rightarrow a/2=0,025$

2ο δείγμα: $n_2=75$, $\bar{x}_2=280$ — $1-a/2=0,975$

$z_{a/2} = 1,96$

$H_0: \mu_1=\mu_2$ και $H_1: \mu_1 \neq \mu_2$

$$z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{300-280}{\sqrt{\frac{30^2}{100}+\frac{18^2}{75}}} = 5,48$$

Άρα απορρίπτουμε την H_0, οπότε έχουν δι στις πωλήσεις τους, με περιθώριο λάθους 5%

Παράδειγμα (2) παλιό θέμα

Θέλουμε να συγκρίνουμε την κατανάλωση καυσίμου σε δύο τύπους μηχανών. Για τον τύπο A κάναμε 50 δοκιμές και διαπιστώσαμε μέση κατανάλωση 18 και για τον τύπο B κάναμε 75 δοκιμές και βρήκαμε μέση κατανάλωση σ' 21. Η τυπική απόκλιση του πρώτου είναι 3 και του δεύτερου 4. Να γίνει η σύγκριση σε επίπεδο σημαντικότητας 4%

Λύση

Κ.Ο.Θ. επειδή τα δείγματα είναι μεγάλα πάω κατευθείαν σε z-τιμές

A μηχανή: $n_1=50$, $\bar{x}_1=18$, $\sigma_1=s_1=3$

B μηχανή: $n_2=75$, $\bar{x}_2=21$, $\sigma_2=s_1=4$

H_0: $\mu_1=\mu_2$ και H_1: $\mu_1 \neq \mu_2$

$$z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}=\frac{18-21}{\sqrt{\frac{3^2}{50}+\frac{4^2}{75}}}=\frac{-3}{\sqrt{\frac{9}{50}+\frac{16}{75}}}=\frac{-3}{\sqrt{0,18+0,21}}=\frac{-3}{\sqrt{0,39}}=\frac{-3}{0,6}=-4,78$$

$a=4\%=0,04 \Rightarrow a/2=0,02 \Rightarrow 1-a/2=0,98 \Rightarrow z_{a/2}=2,06$

−4,78

−2,06 2,06

Άρα απορρίπτουμε την H_0, οπότε οι δύο μηχανές έχουν διαφορά στην κατανάλωση με περιθώριο λάθους 4%

Παράδειγμα (3) παλιό θέμα

Ένας χρηματιστηριακός αναλυτής θέλει να εξετάσει εάν οι εμπορικές μετοχές αποδίδουν περισσότερο από τις βιομηχανικές. Ένα δείγμα 18 εμπορικών παρουσιάσε μέση απόδοση 3,5 και τυπική απόκλιση 1,3 και ένα δείγμα 20 βιομηχανικών παρουσίασε μέση απόδοση 2,7 και τυπική απόκλιση 1,5. Θεωρώντας κανονικές τις αποδόσεις των μετοχών να γίνει ο έλεγχος.

Λύση

Δύο πληθυσμοί (ανεξάρτητοι), κανονικότητα, μικρά δείγματα ⇒ t-τιμές

εμπορικές $n_1=18$, $\bar{x}_1=3,5$, $S_1=1,3$

βιομηχανικές $n_2=20$, $\bar{x}_2=2,7$, $S_2=1,5$

$n_1+n_2-2=18+20-2=36$ β.ε.

H_0: $\mu_1=\mu_2$ κ' H_1: $\mu_1>\mu_2$

$a=0,05$

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}=\frac{3,5-2,7}{\sqrt{\frac{1,3^2}{18}+\frac{1,5^2}{20}}}=\frac{0,8}{0,454}=1,762$$

$t_{36,0,05}=1,69$

1,762

1,69

Απορρίπτουμε την H_0, οπότε οι εμπορικές έχουν μεγαλύτερη απόδοση από τις βιομηχανικές με περιθώριο λάθους 5%

Μάθημα 7ο 29/11

4) Έλεγχος υποθέσεων για ποσοστά.

(A) Εάν έχουμε έλεγχο για ένα ποσοστό γνωρίζοντας πλήρως ένα δε του πληθυσμού (κανονικός)

Έστω ένα δείγμα μεγέθους n στο οποίο ένα ποσοστό του $\hat{p}$ έχει μ ιδιότητα (πχ αγοράζει μια μάρκα τσιγάρων) Θέλουμε →του δείγμ να ελέγξουμε τι ποσοστό του πληθυσμού έχει την ίδια ιδιότητα, δηλ. εάν το ποσοστό p είναι p_0 (ή διάφορο ή μεγαλύτερο ή μικρότερο) →του πληθυσμού

Ο έλεγχος γίνεται υπολογίζοντας μια Z-τιμή. $Z = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

Παράδειγμα

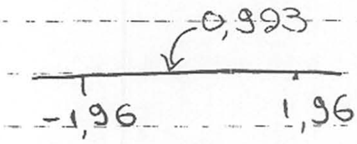
Ρωτήσαμε 400 ανθρώπους για το ποιο δελτίο ειδήσεων παρακολουθούν το 27% μας απάντησε ότι παρακολουθεί το STAR. Να ελέγξετε εάν μπορούμε να πούμε ότι το 25% όλων των τηλεθεατών παρακολουθεί το STAR.

Λύση

Δείγμα μεγέθους 400 στο οποίο ένα ποσοστό του $\hat{p}=27\%$ παρακολ. στο STAR. Θέλουμε να ελέγξουμε εάν το 25% όλων παρακολουθε

$H_0: p=25\%=0,25$ κ' $H_1: p \neq 25\% \neq 0,25$

$a=5\%=0,05 \Rightarrow a/2=0,025 \Rightarrow Z_{a/2}=1,96$



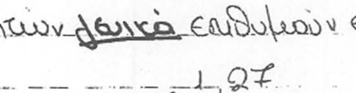
$$Z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}=\frac{0,27-0,25}{\sqrt{\frac{0,25(1-0,25)}{400}}}=\frac{0,02\cdot 20}{\sqrt{0,25\cdot 0,75}}=\frac{0,4\cdot 10\cdot 10}{5\sqrt{75}}=\frac{8}{\sqrt{75}}=0,923$$

Άρα αποδεχόμαστε την Ho, δηλαδή μπορούμε να υποθέσουμε ότι του πληθυσμού παρακολουθεί ειδήσεις του STAR.

Άσκηση I παλιό θέμα

Ένας καθηγητής ρώτησε 121 μαθητές εάν θέλουν να γράψουν επαναλ. διαγώνισμα. Το 30% απάντησε καταφατικά ενώ το 70% απάντησε αρνητ. Να ελέγξετε εάν πάνω από το 25% των μαθητών γενικά επιθυμούν επαν.

Λύση



$n=121$, $\hat{p}=30\%=0,3$, $a=0,05 \Rightarrow Z_a=1,64$

$H_0: p=0,25$ κ' $H_1: p>0,25$

$$Z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}=\frac{0,30-0,25}{\sqrt{\frac{0,25(1-0,25)}{121}}}=\frac{0,05\cdot 11}{\frac{5}{10}\cdot\frac{\sqrt{75}}{10}}=\frac{55}{5\sqrt{75}}=\frac{11}{75}=1,27$$

Άρα αποδεχόμαστε Ho, δηλαδή το 25% των μαθητών επιθυμούν επαναλ.

Άσκηση ΙΙ

Σε τυχαίο δείγμα 400 ατόμων οι 100 δήλωσαν ότι αγοράζουν συγκεκριμένο αποσμητικό. Μπορούμε να δεχθούμε ότι στο σύνολο των καταναλωτών έχει μερίδιο αγοράς λιγότερο από 30%

Λύση

Δείγμα $n=400$, $\hat{p} = \frac{100}{400} = 0{,}25$

πληθυσμός: λιγότερο από 30%

$H_0: p=0{,}3$ ή $H_1: p<0{,}3$

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0{,}25-0{,}3}{\sqrt{\frac{0{,}3\cdot 0{,}7}{400}}} = \frac{-0{,}05\cdot 20}{\sqrt{0{,}3\cdot 0{,}7}} = \frac{-1}{0{,}458} = -2{,}18$$

$-2{,}18$ ↙ $-1{,}64$

Απορρίπτουμε την Ho άρα λιγότερο από 30% αγοράζουν το αποσμητικό

(B) Έλεγχος για διαφορά ποσοστών στους αντίστοιχους πληθυσμούς (Δεν έχω ποσοστό πληθυσμού)

Δείγμα 1: n_1, p_1 Υπολογίζουμε $z = \frac{p_1+p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Δείγμα 2: n_2, p_2

όπου $\hat{p} = \frac{p_1\cdot n_1 + p_2\cdot n_2}{n_1+n_2}$

Άσκηση Ι παλιό θέμα

Ένα προϊόν παράγεται με δύο διαφορετικές μεθόδους. Παίρνουμε τυχαίο δείγμα χιλίων τεμαχίων από αυτά που παράγονται με την Α μέθοδο και βρίσκουμε 150. Παίρνουμε τυχαίο δείγμα 500 τεμαχίων από αυτά που παράγονται με τη Β μέθοδο και βρίσκουμε 60 ελαττωματικά. Μπορούμε να πούμε σε επίπεδο σημαντικότητας 1% ότι οι δύο μέθοδοι διαφέρουν ως προς το ποσοστό των ελαττωματικών προϊόντων?

Λύση

Δείγμα 1: $n_1=1000$, $p_1 = \frac{150}{1000} = 0{,}15$

Δείγμα 2: $n_2=500$, $p_2 = \frac{60}{500} = 0{,}12$

$1{,}57$ ↙ $-2{,}58$ $2{,}58$

$a=0{,}01 \Rightarrow a/2 = 0{,}005$, $1-0{,}005 = 0{,}995 \Rightarrow Z_{a/2} = 2{,}58$

$$\hat{p} = \frac{p_1\cdot n_1 + p_2\cdot n_2}{n_1+n_2} = \frac{1000\cdot 0{,}15 + 500\cdot 0{,}12}{1000+500} = \frac{150+60}{1500} = \frac{210}{1500} = 0{,}14$$

$$z = \frac{\hat p_1 - \hat p_2}{\sqrt{\hat p(1-\hat p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0,15-0,12}{0,14(1-0,14)\left(\frac{1}{1000}+\frac{1}{500}\right)} = 1,57$$

Άρα αποδεχόμαστε Ho, δηλαδή δεν υπάρχει διαφορά μεταξύ των μεθόδων όσον αφορά την παραγωγή ελαττωματικών προϊόντων

Άσκηση II

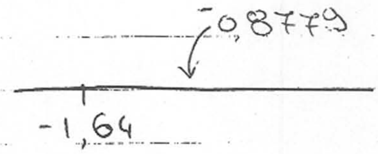
Ένα τυχαίο δείγμα 200 βιομηχανικών επιχειρήσεων δίνει την πληροφορία ότι 12% από αυτές έχει αυξημένα έξοδα διαφήμισης. Ένα τυχαίο δείγμα ίσου μεγέθους εμπορικών επιχειρήσεων μας πληροφορεί ότι το 15% από αυτές έχει αυξημένα έξοδα διαφήμισης. Να ελέγξετε εάν σε επίπεδο σημαντικότητας 5% το ποσοστό των εμπορικών επιχειρήσεων που ξοδεύει μεγάλα ποσά για διαφήμιση ε[ίναι] μεγαλύτερο από αυτό των βιομηχανικών

Λύση

Δείγμα βιομηχανικών: $n_1=200$, $p_1=0,12$

Δείγμα εμπορικών: $n_2=200$, $p_2=0,15$

$a=0,05 \Rightarrow 1-a=0,95 \Rightarrow Z_a=1,64$



$H_0: p_1=p_2$ και $H_1: p_1<p_2$

$$\hat p = \frac{p_1\cdot n_1 + p_2\cdot n_2}{n_1+n_2} = \frac{0,12\cdot 200+0,15\cdot 200}{200+200} = \frac{24+30}{400} = \frac{54}{400} = 0,135$$

$$z = \frac{p_1-p_2}{\sqrt{\hat p(1-\hat p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0,12-0,15}{\sqrt{0,135(1-0,135)\left(\frac{1}{200}+\frac{1}{200}\right)}} = \frac{-0,03}{\frac{\sqrt{0,135\cdot 0,865}}{\sqrt{100}}} = \frac{-0,03\cdot 10}{\sqrt{0,116775}} =$$

$$= \frac{-0,3}{0,3417} = -0,8779$$

Άρα αποδεχόμαστε Ho, δηλαδή δεν έχουν διαφορά τα έξοδα για διαφ[ήμιση] των δύο επιχειρήσεων.

Προσοχή !

(1) Όταν στη ρίζα έχω κλάσμα κάνω αριθμητή-παρονομαστή, επάνω σ[το] κλάσμα, κάνω το σύνθετο απλό και μετά πράξεις χωρίς προσεγγίσεις από την

(2) Όταν βάζω μέχρι 4 δεκαδικά γράφω όλα και κάνω προσέγγιση ι[σ]την τελική πράξη ή μόνο εάν έχω άπειρα δεκαδικά

Συσχέτιση δύο μεταβλητών

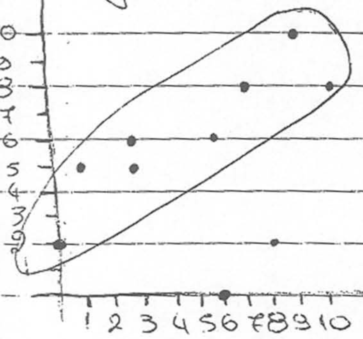
Συσχέτιση: Σχέση των τιμών τους ανεξάρτητα από το εάν είναι εξαρτημένα ή ανεξάρτητα δείγματα

Γραφικά χρησιμοποιούμε το διάγραμμα διασποράς (ονομάζεται σκεδό διάγραμμα γιατί δεν σχεδιάζουμε καμπύλη, αλλά σημεία. Στο SPSS είναι το scatterdiag

Παράδειγμα

Έχουμε τους βαθμούς στη στατιστική και στο marketing 10 φοιτητών

Στατιστική	9	8	6	10	3	1	7	3	6	0
Marketing	10	2	0	8	5	5	8	6	6	2

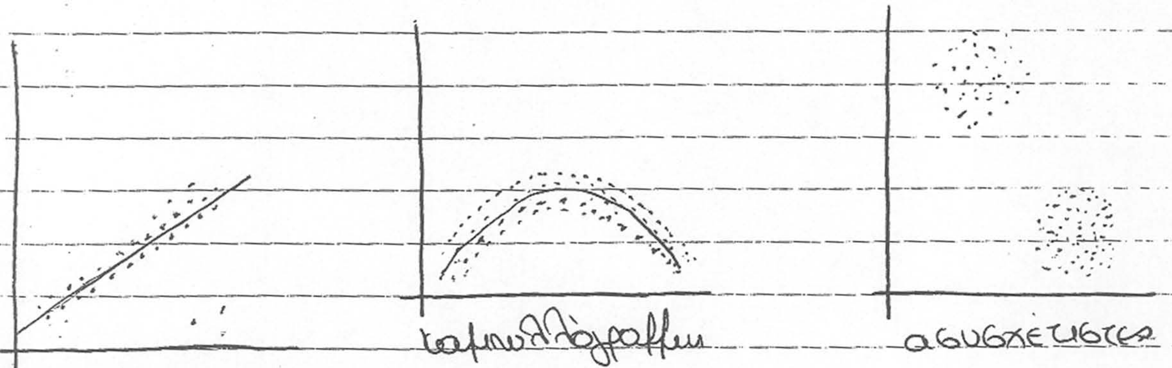


Βασικό ερώτημα

Όλες ή σχεδόν όλες οι τιμές συγκεντρώνονται γύρω από μια καμπύλη?

Απάντηση

Εάν ναι τότε οι μεταβλητές είναι συσχετισμένες. Εάν όχι τότε οι μεταβλητές είναι ασυσχέτιστες



γραμμική συσχέτιση — καμπυλόγραμμη συσχέτιση — ασυσχέτιστες

Παλινδρόμηση

Παλινδρόμηση: πόσο κοντά πέφτουν / μεταβάλλονται τα σημεία σε μια γραμμή

εξίσωση παλινδρόμησης: η εξίσωση της καμπύλης γύρω από την οποία συγκεντρώνονται τα σημεία συσχετισμένων μεταβλητών

γραμμική συσχέτιση -> γραμμική παλινδρόμηση -> εξίσωση ευθείας παλινδρόμησης

καμπυλόγραμμη συσχέτιση -> καμπυλόγραμμη παλινδρόμηση -> εξίσωση καμπύλης παλινδρόμησης (δευτεροβάθμια)

Συντελεστής συσχέτισης: αριθμός που δείχνει πόσο συνδέονται οι τιμές των δύο μεταβλητών.

Γραμμική Συσχέτιση

συντελεστής συσχέτισης = Συντελεστής Pearson = r

$$r = \frac{Cor(X,Y)}{S_X \cdot S_Y}$$

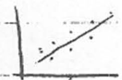
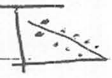
όπου S_X, S_Y: τυπικές αποκλίσεις

$$Cor(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

συνδιακύμανση

Correlation

Συμπεράσματα

- Ο r είναι ανεξάρτητος από μονάδες μέτρησης
- Ο r είναι αξιόπιστο **μόνο** στη γραμμική συσχέτιση (στην καμπυλόγραμμη χρησιμοποιούμε άλλο συντελεστή)
- $-1 \leq r \leq 1$
- Αν $r>0$ τότε έχουμε θετική γραμμική συσχέτιση (αύξουσα) 
- Αν $r<0$ τότε έχουμε αρνητική γραμμική συσχέτιση (φθίνουσα) 
- Αν $|r|$ είναι από 0 έως 0,2 έχουμε ασήμαντη συσχέτιση
  από 0,2 έως 0,4 έχουμε μέτρια συσχέτιση
  από 0,4 έως 0,7 έχουμε σημαντική συσχέτιση
  από 0,7 έως 1 έχουμε ισχυρή συσχέτιση

Μόνο στην ισχυρή μπορούμε να βρούμε ευθεία με μικρό ποσοστό λάθους

- Ο r αναφέρεται στο δείγμα.

Συντελεστής pearson r: αναφέρεται στο δείγμα (υπολογίζεται ακριβώς)
Βαθμός συσχέτισης ρ: αναφέρεται στον πληθυσμό (έλεγχος υπόθεσης)
Ho: $\rho=0$ κ' H1: $\rho\neq0$, δηλαδή αμφίπλευρος έλεγχος (γίνεται χρησιμοποιώντας
ασυσχέτιστα συσχετισμένα την κατανομή t)

Υπολογίζουμε $t=\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$ και συγκρίνουμε με $t_{n-2,\alpha/2}$

Αν μας ενδιαφέρει αν έχουμε θετική ή αρνητική συσχέτιση κάνω μονόπλευρο έλεγχο Ho: $\rho=0$ κ' H1: $\rho>0$ (θετική) Ho: $\rho=0$ κ' H1: $\rho<0$ (αρνητική)
πχ. μας ενδιαφέρει αν έχουμε θετική συσχέτιση.
Υπολ. $t=\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$ και συγκρίνω με $t_{n-2,\alpha}$

Παράδειγμα

Από ένα δείγμα μεγέθους n=8 βρήκαμε συντελεστή pearson=-0,83. Να ελεγχθεί σε επίπεδο σημαντικότητας 5% αν υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών και να προσδιοριστεί το είδος της.

Λύση

n=8, r=-0,83

1) Υπάρχει γραμμική συσχέτιση?

Ho: $\rho=0$ κ' H1: $\rho\neq0$ $t=\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}=\dfrac{-0,83}{\sqrt{\dfrac{1-(-0,83)^2}{8-2}}}=-3,656$

$t_{n-2,\alpha/2}=t_{6,0.025}=2,447$

-3,656

-2,447 2,447

Άρα απορρίπτουμε την Ho, δηλαδή υπάρχει γραμμική συσχέτιση.

2) Επειδή ο συντελεστής pearson είναι αρνητικός υποψιαζόμαστε πως έχω αρνητική συσχέτιση

Ho: $\rho=0$ κ' H1: $\rho\neq0$

$t=-3,656$ και $t_{n-2,\alpha}=t_{6,0.05}=1,943$

-3,656

-1,943

Άρα απορρίπτουμε Ho, δηλαδή έχουμε αρνητική γραμμική συσχέτιση.

Αν ο pearson είναι εκτός ορίων δηλαδή μεγαλύτερος του 1 ή μικρότερος απ' το -1, έχω κάνει λάθος πράξεις.

Απλή Γραμμική Παλινδρόμηση

Έχουμε δύο μεταβλητές X, Y.

- Πρέπει να βρω ποια είναι η λογική σχέση μεταξύ των δύο μεταβλητών, δηλαδή να αποφασίσω ποια μεταβλητή εξαρτάται από την άλλη.

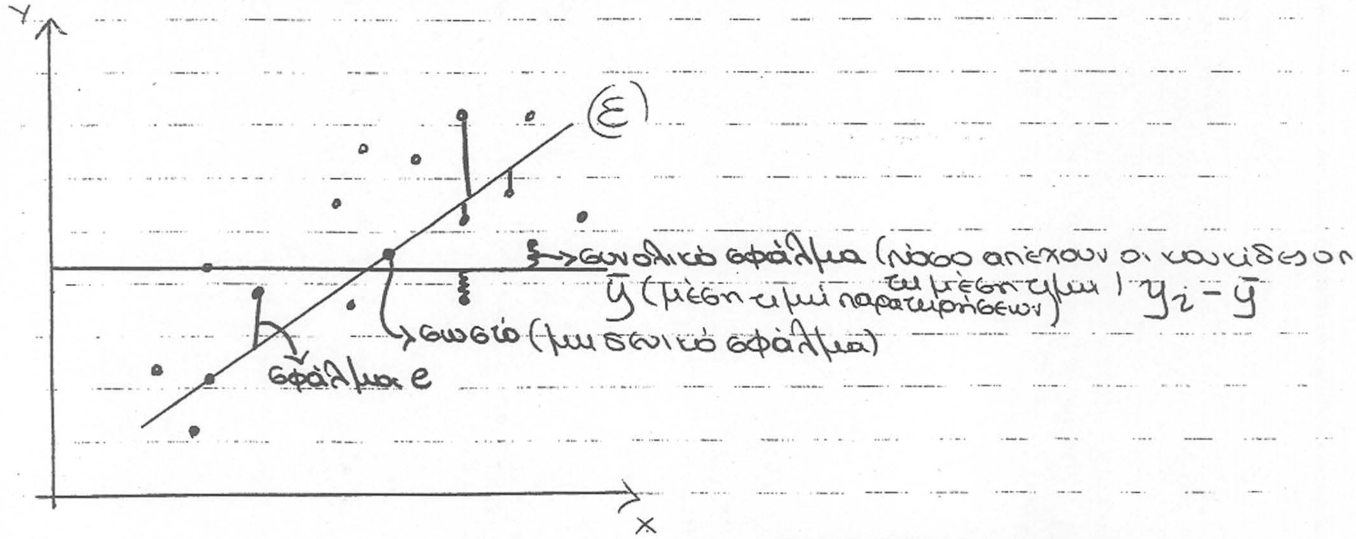
__Πχ__ 1. Με ενδιαφέρει η σχέση μεταξύ των διαφημιστικών δαπανών μιας εταιρ. και των πωλήσεών της. Οι πωλήσεις εξαρτώνται από τις δαπάνες για διαφήμ.

2. Με ενδιαφέρει η σχέση μεταξύ του εισοδήματος των καταναλωτών κ. η κατανάλωση ενός προϊόντος. Η κατανάλωση εξαρτάται από το εισόδημα.

X : ανεξάρτητη μεταβλητή → οριζόντιο άξονα

Y : εξαρτημένη μεταβλητή → κατακόρυφο άξονα

⚠ Στο διάγραμμα διασποράς στη συσχέτιση δε παίζει ρόλο πως θα τοποθετήσω τις μεταβλητές στον άξονα. Αντίθετα στην παλινδρόμηση ΟΠΩΣΔΗΠΟΤΕ τοποθετώ την ανεξάρτητη στον άξονα X και την εξαρτημένη στον άξονα Y.



- Ψάχνουμε την "καλύτερη" (ε). Έστω $\boxed{(\varepsilon): y = b_1 \cdot x + b_0}$

__Κριτήριο__: Θέλω να ελαχιστοποιηθούν τα σφάλματα, δηλαδή οι αποκλίσεις από τα πραγματικά σημεία.

σφάλμα $= e = y - y_i$ → παρατήρηση

πόσο απέχει η ευθεία – την πραγματική τιμή = e

Κάποια σφάλματα είναι θετικά, κάποια είναι αρνητικά και κάποια εί μηδέν. Για να μην έχω πρόβλημα με τα πρόσημα υπολογίζουμε τα τετράγωνα τω σφαλμάτων και προσπαθούμε να ελαχιστοποιήσουμε το άθροισμά τους (μέθοδ ελαχίστων τετραγώνων).

SSE = sum square errors = άθροισμα τετραγώνων σφαλμάτων $= \Sigma e^2 = \Sigma(y-y$

$= \Sigma y_i^2 - b_0 \cdot \Sigma y_i - b_1 \cdot \Sigma(x_i y_i)$

Διασπορά σφαλμάτων $S_e^2 = \frac{SSE}{n-2}$

SST = sum square total errors = άθροισμα τετραγώνων συνολικών σφαλμάτων =
$= \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

- Συντελεστής προσδιορισμού $R^2 = 1 - \frac{SSE}{SST}$

- $0 \le R^2 \le 1$
- Ο R^2 μετράει την προσαρμοστικότητα του μοντέλου.
  - → Αν $R^2 \to 1$ τότε έχουμε καλή προσαρμογή
  - → Αν $R^2 \to 0$ τότε δεν έχουμε καλή προσαρμογή
- Στην απλή γραμμική παλινδρόμηση ισχύει $R^2 = r^2$

- Υπολογισμός ευθείας

b_1 = συντελεστής διεύθυνσης της ευθείας = συντελεστής παλινδρόμησης

$$= \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2}$$

Γενική Ερμηνεία (την προσαρμόζω στην άσκηση)

Εκφράζει πόσο θα μεταβληθεί η τιμή της εξαρτημένης μεταβλητής εάν η ανεξάρτητη ~~μεταβληθεί~~ αυξηθεί κατά μία μονάδα.

$b_0 = \bar{y} - b_1 \cdot \bar{x}$

Γενική Ερμηνεία

Εκφράζει την τιμή της εξαρτημένης μεταβλητής όταν η ανεξάρτητη μηδενιστεί

- Έλεγχος σημαντικότητας για το b_1 { όχι η τιμή του b_1 αλλά ο χαρακτηρισμός του ως σημαντικός ή όχι

$H_0: b_1 = 0$ κ' $H_1: b \neq 0$

δεν είναι σημαντικό   είναι σημαντικό

Υπολογίζουμε $t = \frac{b_1}{S_{b_1}}$ όπου $S_{b_1} = \frac{\sqrt{\frac{SSE}{n-2}}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}}$

και συγκρίνουμε με $t_{n-2,a/2}$

- Διάστημα εμπιστοσύνης για το b_1

$$\left[b_1 - S_{b_1} \cdot t_{n-2,a/2} \; , \; b_1 + S_{b_1} \cdot t_{n-2,a/2} \right]$$

Άσκηση

Δίνονται τα έξοδα για διαφήμιση και οι αντίστοιχες πωλήσεις μιας επιχείρησης ; δέκα χρόνια (οι πωλήσεις εξαρτώνται από τη διαφήμιση άρα διαφήμιση =X και πωλήσεις
Τα έξοδα δίνονται σε χιλιάδες ευρώ και οι πωλήσεις σε τόνους

1) Να βρεθεί η ευθεία παλινδρόμησης, να υπολογίσετε και να ερμηνεύσετε τους συντελε του μοντέλου.
2) Να γίνει έλεγχος για τη στατιστική σημαντικότητα του συντελεστή παλινδρόμησ
3) 95% Δ.Ε για το συντελεστή παλινδρόμησης
4) Να υπολογιστεί και να ερμηνευθεί ο συντελεστής προσδιορισμού R^2
5) Να υπολογιστεί ο συντελεστής συσχέτισης και να γίνει έλεγχος για τη στατιστική σημαντικότ

x_i =διαφ.	y_i =πωλ.	$x_i \cdot y_i$	x_i^2	y_i^2
6	50	300	36	2500
4	46	184	16	2116
5	55	275	25	3025
6	60	360	36	3600
3	44	132	9	1936
4	41	164	16	1681
5	45	225	25	2025
3	38	114	9	1444
6	55	330	36	3025
8	66	528	64	4356
50	500	2612	272	25.708

Λύση

1) $b_1 = \dfrac{n \cdot \Sigma x_i \cdot y_i - \Sigma x_i \Sigma y_i}{n \cdot \Sigma x_i^2 - (\Sigma x_i)^2} = \dfrac{10 \cdot 2612 - 50 \cdot 50}{10 \cdot 272 - (50)^2}$

$= \dfrac{26120 - 25.000}{2720 - 2500} = \dfrac{1120}{220} = 5,09$

$b_0 = \bar{y} - b_1 \cdot \bar{x} = \dfrac{\Sigma y_i}{n} - b_1 \cdot \dfrac{\Sigma x_i}{n} =$

$= \dfrac{50\cancel{0}}{1\cancel{0}} - 5,09 \cdot \dfrac{5\cancel{0}}{1\cancel{0}} = 24,55$

Άρα (ε): $y = 5,09\,x - 24,55$

Θετικός ο συντελεστής του X άρα αύξουσα άρα έχουμε θετική γραμμική συσχέ

Ερμηνεία

b_1: Αν τα έξοδα για διαφήμιση αυξηθούν κατά 1000€ αναμένουμε οι πωλήσ να αυξηθούν κατά 5,09 τόνους

b_0: Αν τα έξοδα για διαφήμιση μηδενιστούν οι πωλήσεις αναμένονται να είναι

2) Έλεγχος σημαντικότητας για το b_1

H_0: b_1 = δεν είναι σημαντικός H_1: b_1 = σημαντικός

$t = \dfrac{b_1}{S_{b_1}} = \dfrac{5,09}{0,88} = 5,78$ $t_{n-2, a/2} = t_{8, 0.025} = 2,306$

−2,306   2,306   5,78

όπου

$S_{b_1} = \dfrac{\sqrt{\dfrac{SSE}{n-2}}}{\sqrt{\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}}} = \dfrac{\sqrt{\dfrac{137,92}{8}}}{\sqrt{272 - \dfrac{50^2}{10}}} = \dfrac{\sqrt{17,24}}{\sqrt{272-250}} = \dfrac{\sqrt{17,24}}{\sqrt{22}} = \dfrac{4,15210}{4,6904} = 0,88$

$SSE = \Sigma y_i^2 - b_0 \cdot \Sigma y_i - b_1 \cdot \Sigma x_i y_i = 25708 - 24,55 \cdot 500 - 5,09 \cdot 2612 =$

$= 25708 - 12275 - 13295,08 = 137,92$

Άρα απορρίπτουμε H_0, δηλαδή ο συντελεστής b_1 είναι στατιστικά σημαντι

3) 95% Δ.Ε ⇒ $\left[b_1 - S_{b_1} \cdot t_{n-2,\alpha/2} \, , \, b_1 + S_{b_1} \cdot t_{n-2,\alpha/2} \right] =$

$= [5,09 - 0,88 \cdot 2,306 \, , \, 5,09 + 0,88 \cdot 2,306] =$

$= [5,09 - 2,02928 \, , \, 5,09 + 2,02928] =$

$= [3,06 \, , \, 7,12]$

Ερμηνεία

Με 95% πιθανότητα αν αυξήσουμε κατά 1000€ τα έξοδα για διαφήμιση, οι πωλήσεις θα μεταβληθούν από 3 έως 7 τόνους (αυξητικά).

4) $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{137,92}{708} = 1 - 0,1948 = 0,81$

$SST = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 25708 - \frac{500^2}{10} = 25708 - 25.000 = 708$

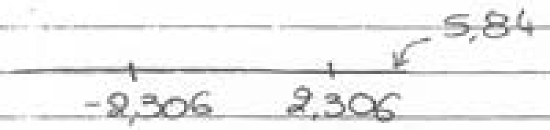
Ερμηνεία

Τα διαφημιστικά έξοδα επηρεάζουν κατά 81% τις πωλήσεις.

5) $r^2 = R^2 \Rightarrow r = +\sqrt{0,81} \Rightarrow r = 0,9$

H_0: δεν είναι σημαντικός κ' H_1: είναι σημαντικός

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0,9}{\sqrt{\frac{1-0,9^2}{10-2}}} = \frac{0,9}{\sqrt{\frac{0,19}{8}}} = \frac{0,9}{\sqrt{0,02375}} = \frac{0,9}{0,1541} = 5,84$$



Απορρίπτουμε την H_0 άρα ο συντελεστής pearson είναι στατιστικά σημα

Μάθημα 9ο 13/12/2012

Άσκηση Ι

Μια εταιρεία κινητής τηλεφωνίας χρησιμοποιεί πωλητές που επισκέπτονται τα σπίτια προκειμένου να αυξήσει τους συνδρομητές της. Ο αριθμός των πωλητών οι αντίστοιχοι νέοι συνδρομητές σε διάστημα 8 εβδομάδων είναι:

πωλητές: 1, 4, 2, 6, 1, 3, 5, 2 συνδρομητές: 12, 16, 15, 18, 10, 17, 16, 13

α) Να προσδιοριστεί η ευθεία γραμμικής παλινδρόμησης και να ερμηνευτούν οι συντελεστές του μοντέλου

β) Να γίνει έλεγχος στατιστικής σημαντικότητας και να κατασκευαστεί 95% Διάστημα Εμπιστοσύνης για τον συντελεστή παλινδρόμησης

γ) Να υπολογιστεί και να ερμηνευτεί ο συντελεστής προσδιορισμού

δ) Να υπολογιστεί ο συντελεστής συσχέτισης και να γίνει έλεγχος στατιστικής σημαντικότητας με $a=1\%$.

ε) Να υπολογιστεί το υπόλοιπο του μοντέλου για την 2η εβδομάδα και να ερμηνευτεί.

στ) Αν η εταιρεία διαθέσει 10 πωλητές, πόσοι προβλέπεται να είναι οι νέοι συνδρομητές?

Λύση

X	Y	$X\cdot Y$	X^2	Y^2	πρόβλεψη	υπόλοιπο
1	12	12	1	144		
4	16	64	16	256	16	0
2	15	30	4	225	13	+2
6	18	108	36	324		
1	10	10	1	100		
3	17	51	9	289		
5	16	80	25	256		
2	13	26	4	169		
24	117	381	96	1763		

α) $$b_1 = \frac{n\cdot\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\cdot\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{8\cdot 381 - 24\cdot 117}{8\cdot 96 - 24^2} = \frac{240}{192} = 1,25$$

$$b_0 = \bar{y} - b_1\cdot\bar{x} = \frac{\Sigma y_i}{n} - b_1\frac{\Sigma x_i}{n} = \frac{117}{8} - 1,25\cdot\frac{24}{8} = 10,875$$

Άρα $\boxed{y = 1,25x + 10,875}$ θετική γραμμική συσχέτιση

Ερμηνεία

b_1: Αν οι πωλητές αυξηθούν κατά $1\cdot 100 = 100$ οι συνδρομητές αναμένεται να αυξηθούν κατά $1,25\cdot 100 = 125$

b_0: Όταν οι πωλητές μηδενιστούν, τότε αναμένεται οι συνδρομητές να είναι 11 (στρογγυλοποίηση του 10,875)

Επειδή πρόκειται για ανθρώπους κάνω μετατροπή στο ×100

β) Έλεγχος σημαντικότητας για το B_1: Ho δεν είναι σημαντικός Η1: είναι σημαντικός:

$$t = \frac{B_1}{S_{B_1}} = \frac{1,26}{0,32} = 3,9$$

$$S_{B_1} = \frac{\sqrt{\frac{SSE}{n-2}}}{\sqrt{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}} = \frac{\sqrt{\frac{14,96}{6}}}{\sqrt{96 - \frac{24^2}{8}}} = \frac{\sqrt{2,4933}}{\sqrt{24}} = \frac{1,5790}{4,8989} = 0,32$$

$$SSE = \Sigma y_i^2 - b_0 \cdot \Sigma y_i - b_1 \Sigma x_i y_i = 1763 - 10,87 \cdot 117 - 1,25 \cdot 381 = 14,96$$

$-2,447$ $\quad 2,447 = t_{n-2,a/2}$ $\quad$ 3,9

$t_{6,0.025} = 2,447$

Απορρίπτω την Ho. Άρα ο B_1 είναι στατιστικά σημαντικός

95% ΔΕ για το $B_1 = [B_1 - S_{B_1} \cdot t_{n-2,a/2},\ B_1 + S_{B_1} \cdot t_{n-2,a/2}] =$
$= [1,25 - 0,32 \cdot 2,447,\ 1,25 + 0,32 \cdot 2,447] = [0,47,\ 2,03]$

Ερμηνεία: Με σιγουριά 95% αν αυξηθούν οι πωλητές κατά 100, θα μεταβληθούν οι συνδρομητές από 47 μέχρι 203

γ) Συντελεστής προσδιορισμού $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{14,96}{51,875} = 0,72$

$$SST = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = 1763 - \frac{(117)^2}{8} = 51,875$$

Ερμηνεία: Το πόσοι συνδρομητές θα μου έρθουν εξαρτάται από το πόσους πωλητές έχω κατά 72%

δ) Συντελεστής συσχέτισης $r = \sqrt{R^2}$ Άρα $r = \sqrt{0,72} = 0,848$

Ερμηνεία: Ισχυρή θετική συσχέτιση ανάμεσα στον αριθμό των συνδρομητών και πωλητών

Έλεγχος στατιστικής σημαντικότητας κατά 1%

Ho: βαθμός σημαντικότητας = 0 κ' Η1: βαθμός σημαντικότητας ≠ 0

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0,848}{\sqrt{\frac{1-0,848^2}{6}}} = \frac{0,848 \cdot \sqrt{6}}{\sqrt{1-0,848^2}} = 3,9$$

$-3,707$ $\quad 3,707 = t_{n-2,a/2}$ $\quad$ 3,9

$a = 0,01 \to t_{6,0.005} = 3,707$ Άρα απορρίπτω την Ho. Είναι οριακά σημαντικό να χρησιμοποιήσω το μοντέλο για την πρόβλεψη.

ε) Τη 2η εβδομάδα έχω 4 πωλητές και 16 συνδρομητές. Άρα με βάση το μοντέλο $y = 1,25 \cdot 4 + 10,87 = 15,87 \approx 16$ Άρα η πραγματική τιμή με την πρόβλεψη συμπίπτουν

Για την 3η εβδομάδα η πραγματικότητα είναι καλύτερη από την πρόβλεψη

στ) Επειδή οι συντελεστές ήταν στατιστικά σημαντικοί μπορώ να κάνω πρόβλ (αλλιώς όχι) $y = 1,25 \cdot 10 + 10,87 = 23,37 \approx 23$

Αν διαθέτω 10 πωλητές αναμένεται να έχω 23 συνδρομητές.

Άσκηση ΙΙ

Μια εταιρεία κατασκευάζει διαφημιστικά μπαλόνια. Οι ώρες που τα μπαλόνια μένουν φουσκωμένα ακολουθούν κανονική κατανομή με $\sigma=12$

α) Από ένα τυχαίο δείγμα 100 μπαλονιών βρέθηκε ότι παραμένουν φουσκωμένα για 155 ώρες. Σε επίπεδο σημαντικότητας 5% να ελεγχθεί η υπόθεση ο μέσος χρόνος που τα μπαλόνια παραμένουν φουσκωμένα είναι μικρότερος από 160 ώρες

β) Σε ένα τυχαίο δείγμα 300 μπαλονιών βρέθηκαν 15 ελαττωματικά. Να ελεγχθεί με $a=5\%$ η υπόθεση ότι το 3% της παραγωγής είναι ελαττωματικά προϊόντα

Λύση

α) $H_0: \mu=160$ κ' $H_1: \mu<160$

Έχω μεγάλο δείγμα και κανονικότητα με γνωστή διασπορά. Άρα θα χρησιμοποιήσω Z κατανομή.

Πληθυσμός: κανονικότητα, $\sigma=12$

δείγμα: $n=100$, $\bar{x}=155$ επίπεδο σημαντικότητας $a=5\%$

$$Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{155-160}{\frac{12}{\sqrt{100}}}=\frac{-5}{1,2}=-4,16$$

$-4,16$ Z_a $-1,64$

$a=0,05$ $1-a=0,95$ $Z_{0,95}=1,64$

Άρα απορρίπτω την H_0. Οπότε τα μπαλόνια παραμένουν φουσκωμένα λιγότερες από 160 ώρες

β) πληθυσμός: κανονικότητα, $\sigma=12$

Δείγμα: $n=300$, ελαττωματικά $=15$

ποσοστό δείγματος $\hat{p}=\frac{15}{300}=0,05$

Έλεγχος $H_0: p=0,03$ $H_1: p\neq 0,03$

Μεγάλο δείγμα ⇒ Z-τιμές

$$Z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}=\frac{0,05-0,03}{\sqrt{\frac{0,03\cdot 0,97}{300}}}=\frac{0,02\cdot\sqrt{300}}{\sqrt{0,03\cdot 0,97}}=2,03$$

$a=0,05$ $a/2=0,025$ $1-a/2=0,975$

$Z_{0,975}=1,96$

$-1,96$ $1,96$ $2,03$

Άρα απορρίπτω την H_0. Άρα το ποσοστό ελαττωματικών μπαλονιών δεν είναι 3%