ΠΡΟΓΡΑΜΜΑ - SCHEDULE

ΔΕΥΤΕΡΑ MONDAY	ΤΡΙΤΗ TUESDAY	ΤΕΤΑΡΤΗ WEDNESDAY	ΠΕΜΠΤΗ THURSDAY	ΠΑΡΑΣΚΕΥΗ FRIDAY
8				
9				
10				
11				
12				
1 μ.μ.				
2				
3				
4				
5				
6				
7				
8				
9				
10				

ΕΙΔΙΚΕΣ ΣΗΜΕΙΩΣΕΙΣ   SPECIAL NOTES

DAY CLASS LESSON PROF. PRESENTATION DATE	DAY CLASS LESSON PROF. PRESENTATION DATE
DAY CLASS LESSON PROF. PRESENTATION DATE	DAY CLASS LESSON PROF. PRESENTATION DATE

Ποσοτικές Μέθοδοι ΙΙ | 2008-2009
(ΣΤΑΤΙΣΤΙΚΗ)

1. Τι εκφράζει ο μέσος αριθμητικός;

Αν σε όλα τα μαθήματα ενός μαθητή έμπαινε ένας βαθμός αυτός θα ήταν ο απλός μέσος αριθμητικός

2. Τι θα συμβεί στο μέσο αριθμητικό αν σε κάθε παρατήρηση προσθέσουμε μια ποσότητα θετική ή αρνητική καθώς και αν πολλαπλασιάσουμε τις παρατηρήσεις με μη μηδενική ποσότητα;

$y_i = x_i + a \rightarrow \bar{y} = \bar{x} + a$

$y_i = 6x_i \rightarrow \bar{y} = 6\bar{x}$

$y_i = a + 6x_i \rightarrow \bar{y} = a + 6\bar{x}$

3. Βρείτε τη θέση των τεταρτημορίων: 3, 9, 7, 3, 4, 8, 11 και τα ίδ

Τα βάζουμε σε αύξουσα σειρά: 3, 3, 4, 7, 8, 9, 11 $\nu = 7$

1η 2η 3η 4η 5η 6η 7η

Θέση Q_1

$\frac{K(\nu+1)}{4}$ για το Q_1 θέτουμε $K=1$ $\frac{7+1}{4} = 2$η θέση Επομένως $Q_1 = 3$

Θέση Q_2 $(= M)$

$\frac{K(\nu+1)}{4} = \frac{2(7+1)}{4} = 4$η θέση Επομένως $Q_2 = M = 7$

Θέση Q_3

$\frac{K(\nu+1)}{4}$ για το Q_3 θέτουμε $K=3$ $\frac{3(7+1)}{4} = 6$η θέση Επομένως $Q_3 = 9$

4. Βρείτε τα τεταρτημόρια αν: 3, 4, 7, 8, 9, 11 $\nu = 6$

Υπολογισμός Q_1 — στρογγυλοποιώ στον πλησιέστερο ακέραιο

$$\frac{K(\nu+1)}{4} = \frac{6+1}{4} = 1,75 \simeq 2^{\eta} \text{ θέση} \Leftrightarrow Q_1 = 4$$

Υπολογισμός $Q_2 = M$ — Στο κέντρο δύο διαδοχικών μετρήσεων

$$\frac{K(\nu+1)}{4} = \frac{2(6+1)}{4} = 3,5^{\eta} \text{ θέση}$$ Άρα το ημιάθροισμά τους.

$$M = \frac{7+8}{2} = 7,5 = Q_2$$

Υπολογισμός Q_3

$$\frac{K(\nu+1)}{4} = \frac{3(6+1)}{4} = 5,25 \simeq 5^{\eta} \text{ θέση} \Leftrightarrow Q_3 = 9$$

5. Υπολογίστε διάμεσο και τεταρτημόρια αν:

$\nu = \Sigma f_i = 100$

x_i	f_i	F_i
0	5	5
1	15	20
2	45	65
3	20	85
4	10	95
5	5	100

$Q_1 \rightarrow \frac{\nu+1}{4} = 25,25^{\eta}$ θέση. Επομένως $Q_1 = 2$

$Q_2 = M \rightarrow \frac{2(100+1)}{4} = 50,5^{\eta}$ θέση. Επομένως $M = 2$

$Q_3 = \frac{3(100+1)}{4} = 75,75^{\eta}$ θέση. Άρα $Q_3 = 3$

ΥΠΟΛΟΓΙΣΜΟΣ ΧΩΡΙΣ ΤΑΞΕΙΣ ↑

6. Υπολογίστε διάμεσο και τεταρτημόρια, αν:

Μισθοί	f_i	F_M	
10-20	1	1	$0-1^{\eta}$ μέτρηση
20-30	6	7	$2^{\eta}-7^{\eta}$ μέτρηση
30-40	9	16 (F_{Q_1-1})	$8^{\eta}-16^{\eta}$ μέτρηση
40-50 (L_{Q_1})	31 (f_{Q_1})	47 (F_{M-1})	$17^{\eta}-47^{\eta}$ μέτρηση
50-60 (L_M)	42 (f_M)	89 (F_{Q_3-1})	$48^{\eta}-89^{\eta}$ μέτρηση
60-70 (L_{Q_3})	32 (f_{Q_3})	121	$90^{\eta}-121^{\eta}$ μέτρηση
70-80	17	138	$122^{\eta}-138^{\eta}$ μέτρηση
80-90	10	148	$139^{\eta}-148^{\eta}$ μέτρηση
90-100	2	150	$149^{\eta}-150^{\eta}$ μέτρηση
	150		

Τύπος Q_1

$$Q_1 = L_{Q_1} + \delta \frac{\left(\frac{\nu}{4} - F_{Q_1-1}\right)}{f_{Q_1}} = 40 + \frac{10\left(\frac{150}{4} - 16\right)}{31} = 46,935\ldots$$

Τύπος Q_3

$$Q_3 = L_{Q_3} + \delta \frac{\left(\frac{3\nu}{4} - F_{Q_3-1}\right)}{f_{Q_3}} = 60 + \frac{10\left(\frac{450}{4} - 89\right)}{32} = 67,344\ldots$$

Τύπος Διαμέσου

$$M = L_M + \frac{\delta\left(\frac{\nu}{2} - F_{M-1}\right)}{f_M} = \frac{50 + 10\left(\frac{150}{2} - 47\right)}{42} = 56,666\ldots$$

- L_M: κάτω όριο
- δ: πλάτος $(50-60=10)$
- f_M: συχνότητα διαμέσου

} του διαστήματος όπου εντοπίζεται η διάμεσος

- F_{M-1}: αθροιστική συχνότητα του προηγούμενου διαστήματος από αυτό που εντοπίζεται η M.
- $\nu = \Sigma f_i = 150$

7. Υπολογίστε την επικρατούσα τιμή στις παρακάτω περιπτώσεις

α) 1, 2, 7, 5, 2, 4, 3, 2, 11 $T_0 = 2$

β) 1, 2, 7, 9, 4, 3, 11 $T_0 =$ δεν υπάρχει

γ) 2, 7, 9, 2, 3, 4, 3, 2, 3 $T_0 = 2$ και $T_0 = 3$

δ)

x_i	f_i
0	5
1	15
2	45
3	20
4	10
5	5

$45 \rightarrow T_0 = 2$

ε) σε διαστήματα

$$T_0 = L_0 + \delta \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

- L_0 = κάτω όριο διαστήματος
- δ = πλάτος διαστήματος
- Δ_1 = MAX συχνότητα − Προηγούμενη συχνότητα
- Δ_2 = MAX συχνότητα − Επόμενη συχνότητα

8. Αν ισχύει $Q_1 = 17.604$, $Q_2 = M = 22.140$, $Q_3 = 29.025$ υπολογίστε την τεταρτημοριακή απόκλιση.

$$Q_D = Q_3 - Q_1$$

$QD = 29.025 - 17.604 = 11.421$ €

9. Υπολογίστε τη διακύμανση σε πληθυσμό και δείγμα.

ΠΛΗΘΥΣΜΟΣ

$$Var(x) = \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}$$

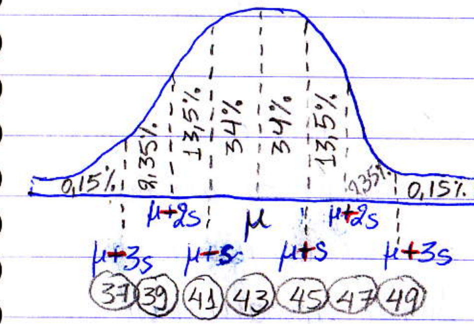
ΔΕΙΓΜΑ

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

π.χ. $x_i: 0, 1, 2, 3 \dots 100$ με $\bar{x} = 2{,}3$

$$s^2 = \frac{(0-2{,}3)^2 + (2-2{,}3)^2 + \dots}{100-1}$$

10. Μια επιχείρηση ρούχων θέλει να βρεί σε τι μεγέθη τη συμφέρει να παράγει έναν αριθμό (π.χ 10.000) παντελονιών. Δίνεται ότι: $\mu = 43cm$ και $S = 2cm$



41-43 = 0,34 × 10.000 = 3400

43-45 = 0,34 × 10.000 = 3400

39-41 = 0,135 × 10.000 = 1350

45-47 = 0,135 × 10000 = 1350

37-39 = 0,0235 × 10.000 = 235

47-49 = 0,0235 × 10000 = 235

11. Υπολογίστε τον συντελεστή ασυμμετρίας αν:

$Q_1 = 46$

$Q_3 = 68$

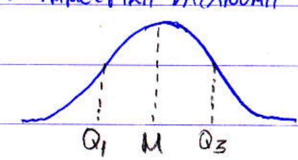
$M = 56$

$$S_K = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{68 + 46 - 2 \cdot 56}{68 - 46} = 0{,}09$$

ΘΕΤΙΚΟΣ → ΘΕΤΙΚΗ ΑΣΥΜΜΕΤΡΙΑ

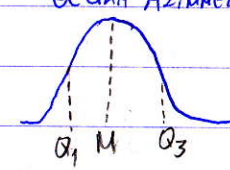
$S_K = [-1, 1]$ 0 = συμμετρική −1 = πλήρως αρνητική 1 = πλήρως θετική

ΣΥΜΜΕΤΡΙΚΗ ΚΑΤΑΝΟΜΗ



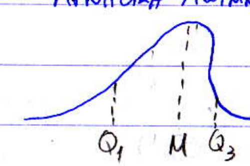
$\bar{x} - M = 0$

ΘΕΤΙΚΗ ΑΣΥΜΜΕΤΡΙΑ



$\bar{x} - M > 0$

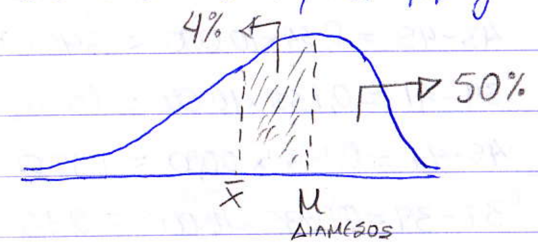
ΑΡΝΗΤΙΚΗ ΑΣΥΜΜΕΤΡΙΑ



$\bar{x} - M < 0$

12. Είναι δυνατόν περισσότεροι από τους μισούς εργαζόμενους μιας τράπεζας να έχουν μηνιαίο μισθό μεγαλύτερο από το μέσο μισθό όλων των εργαζόμενων στην τράπεζα; Εξηγήστε.

Είναι δυνατόν αν οι μισθοί εμφανίζουν αρνητική ασυμμετρία



50% + 4% το ποσοστό των εργαζομένων που αμοίβεται πάνω από τον μέσο μισθό.
Δηλαδή: αρνητική ασυμμετρία.

13. Δύο μαθητές πάνε για συστατική και ο καθηγητής πρέπει να επιλέξει ποιος αξίζει. Ο ένας πέρασε την περίοδο Α και ο άλλος τη Β. ΠΕΡΙΟΔΟΣ

A	B
$X_A = 100$ μόρια	$X_B = 160$ μόρια
$\mu_A = 80$ μονάδες	$\mu_B = 130$ μονάδες ← μέση επίδοση περιόδου
$\sigma_A = 10$	$\sigma_B = 30$ ← διασπορά της βαθμολογίας

Υπολογίζουμε τις τυποποιημένες αποκλίσεις:

$$Z_A = \frac{X_A - \mu_A}{\sigma_A} = \frac{100-80}{10} = 2$$ → μονάδες πάνω από τον Μ.Ο

Επομένως καλύτερος ο Α.

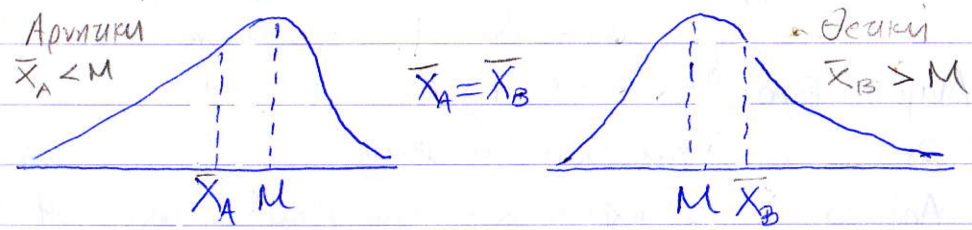
$$Z_B = \frac{160-130}{30} = 1$$

Το Z_i εκφράζει πόσες μονάδες τυπικής απόκλισης η i τιμή της μεταβλητής x βρίσκεται πάνω ή κάτω του μέσου αριθμητικού.

14. Δίνονται οι κατανομές της διάρκειας ζωής (σε ώρες λειτουργίας) δύο τύπων λαμπτήρων με τα εξής χαρακτηριστικά:

	Τύπος Α	Τύπος Β
ΕΥΡΟΣ →	400 h	400 h
ΜΕΣΟΣ ΑΡΙΘΜΗΤΙΚΟΣ →	900 h	900h
ΑΣΥΜΜΕΤΡΙΑ →	Αρνητική	Θετική

Αν οι δύο τύποι λαμπτήρων έχουν ίδια τιμή στην αγορά ποιον θα επιλέγατε και γιατί;



X_A = περισσότεροι από τους μισούς έχουν διάρκεια ζωής πάνω από 900h λόγω αρνητικής ασυμμετρίας.

X_B = περισσότεροι από τους μισούς έχουν διάρκεια ζωής κάτω από 900h.

ΕΠΟΜΕΝΩΣ ΜΑΣ ΣΥΜΦΕΡΕΙ ο Α.

15. Ένας καθηγητής φυσικής αγωγής μελετά το βάρος των μαθητών της Α' Γυμνασίου του σχολείου που θα υπηρετεί. Από ένα δείγμα ν = 200 μαθητών προέκυψε μια εμπειρική κατανομή απόλυτα συμμετρική καθώς και οι εξής εκτιμήσεις.

$\sum_{i=1}^{200} x_i = 10.000 kg$ $\qquad s^2 = 9 \ (kg)^2$

Ο καθηγητής ζητά να προβείτε στις εξής αναλύσεις:

1. Να εκτιμήσετε το συντελεστή μεταβλητότητας και το διάμεσο βάρος.
2. Πόσα παιδιά από τα 200 έχουν βάρος το πολύ 53 kg.

1. $CV = s/\bar{x}$ → $s^2 = 9 \leftrightarrow s = 3$

→ $\bar{x} = \frac{10.000}{200} = 50\,kg$

Άρα $CV = \frac{3}{50} = 0,06$

Επειδή η κατανομή είναι συμμετρική

Διάμεσο βάρος $M = \bar{x} = 50\,kg$

2. Στο $\bar{x} + s$ ισχύει $50 + 3 = 53\,kg$

Από το $\bar{x} - 3s$ μέχρι το $\bar{x} + s$ σχηματίζεται ένα ποσοστό 84%

Επομένως το 84% των παιδιών έχουν βάρος το πολύ 53 kg

$200 * 0,84 = 168$ παιδιά

16. Μια έρευνα μεταξύ παντρεμένων ζευγαριών με αντικείμενο το βαθμό ικανοποίησης από τη μέχρι σήμερα συμβίωση, έδωσε τα εξής αποτελέσματα:

ΕΡΩΤΩΜΕΝΟΣ	ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ	ΜΕΣΟΣ ΒΑΘΜΟΣ ΙΚΑΝΟΠΟΙΗΣΗΣ
Άνδρας χωρίς παιδιά	58 u_1	8,4 $\bar{x}_1$
Γυναίκα χωρίς παιδιά	83 u_2	7,7 $\bar{x}_2$
Άνδρας με παιδιά	72 u_3	5,9 $\bar{x}_3$
Γυναίκα με παιδιά	65 u_4	6,3 $\bar{x}_4$

1. Ποιος είναι ο μέσος βαθμός ικανοποίησης για το σύνολο του δείγματος;
2. Ποιος είναι ο μέσος βαθμός ικανοποίησης για τους άνδρες;

1. Γενικός μέσος όρος:

$$\bar{x} = \frac{u_1\bar{x}_1 + u_2\bar{x}_2 + u_3\bar{x}_3 + u_4\bar{x}_4}{u_1 + u_2 + u_3 + u_4} = 7,053$$

2. $$\bar{x}_{ΑΝΔΡΩΝ} = \frac{u_1\bar{x}_1 + u_3\bar{x}_3}{u_1 + u_3} = 7,015$$

17. Τα παρακάτω δεδομένα αναφέρονται στα συμβόλαια ασφαλειών ζωής που επέτυχε ένα δείγμα 24 ασφαλιστών σε ένα μήνα:

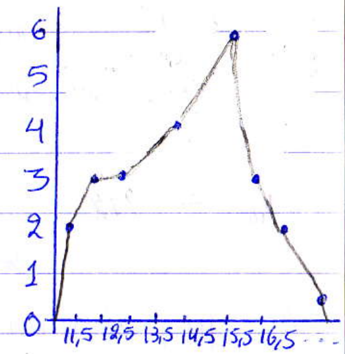
14 11 12 13 15 15 12 17 16 14 13 15

14 15 12 15 14 15 13 16 11 18 17 16

α) Παρουσιάστε τα παραπάνω δεδομένα σε κατάλληλη κατανομή συχνοτήτων και κατασκευάστε το ιστόγραμμα και την πολυγωνική γραμμή.

Η ταξινόμηση θα γίνει σε τιμές μεμονωμένες όχι σε διαστήματα λόγω μικρού αριθμού διαφορετικών τιμών αρχικά καθώς και για την αποφυγή λάθους που γίνεται από τα κατανεμημένα δεδομένα.

x_i ΣΥΜΒΟΛΑΙΑ	f_i ΑΣΦΑΛΙΣΕΙΣ	$f_i x_i$	$f_i(x_i - 14,292)^2$
[10,5 11 11,5)	2	22	⋮
[11,5 12 12,5)	3	36	⋮
[12,5 13 13,5)	3	39	⋮
[13,5 14 14,5)	4	56	⋮
[14,5 15 15,5)	6	90	⋮
[15,5 16 16,5)	3	48	⋮
[16,5 17 17,5)	2	34	⋮
[17,5 18 18,5)	1	18	⋮
	24	343	82,958



β) Εκτιμήστε το σημείο μέγιστης συχνότητας, το μέσο αριθμητικό και την τυπική απόκλιση.

Παρατηρούμε ότι $T_0 = 15$ γιατί $f_{i_{MAX}} = 6 \leftrightarrow x_i = 15$

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{343}{24} = 14,292$$

$$S = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{82,958}{24-1}} = 1,899$$

γ) Τι συμπεράσματα προκύπτουν αν ο διευθυντής πωλήσεων ισχυριζόταν ότι οι παραπάνω πωλήσεις ασφαλειών είναι αποτέλεσμα της πρόσφατης διαφημιστικής εκστρατείας της εταιρίας και ότι αντίστοιχο δείγμα ασφαλιστών επέτυχε, πριν τη διαφημιστική εκστρατεία, μέσο μηνιαίο αριθμό πωλήσεων 12 συμβόλαια με τυπική απόκλιση 4 συμβόλαια.

$\bar{x}_{ΠΡΙΝ} = 12$ $\quad S_{ΠΡΙΝ} = 4$ $\quad$ Η διασπορά θέλουμε να μικρύνει

$\bar{x}_{ΜΕΤΑ} = 14,292$ $\quad S_{ΜΕΤΑ} = 1,899$ $\quad$ αλλά πρέπει να λάβουμε υπόψιν το CV και όχι το S:

$CV_{ΠΡΙΝ} = 4/12 = 0,33 \to 33\%$ $\qquad CV_{ΜΕΤΑ} = \frac{1,899}{14,292} = 0,133 \to 13,3\%$

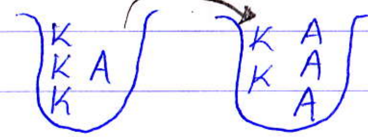
Άρα μίκρυνε η διασπορά, πράγμα που σημαίνει ότι η εκστρατεία ήταν αποτελεσματική.

Τύποι

- Εξαρτημένα ενδεχόμενα $P(A \cap B) = P(A/B)P(B)$
- Ανεξάρτητα $P(A/B) = P(A) \leftrightarrow P(A \cap B) = P(A)P(B)$

18. Έχουμε δύο κουτιά με κόκκινες και άσπρες μπάλες. Ποια η πιθανότητα και οι δύο εξαγωγές να είναι άσπρη;

πολλαπλασιασμός



A_1: άσπρη στην πρώτη εξαγωγή

A_2: άσπρη στη δεύτερη εξαγωγή

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2/A_1) = \frac{1}{4}\frac{4}{6} = \frac{1}{6}$$

19. Από τους πελάτες ενός βενζινάδικου:

A_1: το 70% ζητάει αμόλυβδη απλή → το 80% γεμίζει

A_2: το 10% αμόλυβδη Super → το 70% γεμίζει

A_3: το 20% Super → το 90% γεμίζει

α) Ποια η πιθανότητα ο επόμενος πελάτης να ζητήσει αμόλυβδη απλή και να γεμίσει;

β) Να γεμίσει.

γ) Αν γεμίσει, να ζητήσει αμόλυβδη απλή (δοθέντος ότι θα γεμίσει.)

Ορίζουμε: $P(A_1) = 0,7$ $\quad P(Γ/A_1) = 0,8$

$P(A_2) = 0,1$ $\quad P(Γ/A_2) = 0,7$

$P(A_3) = 0,2$ $\quad P(Γ/A_3) = 0,9$

α) $P(A_1 \cap Γ) = P(A_1)P(Γ/A_1) = 0,7 \times 0,8 = 0,56$

β) Αφορά τα ενδεχόμενα: $Γ$, $A_1 \cap Γ$, $A_2 \cap Γ$, $A_3 \cap Γ$

$$P(Γ) = P[(A_1 \cap Γ) \cup (A_2 \cap Γ) \cup (A_3 \cap Γ)] = P(A_1 \cap Γ) + P(A_2 \cap Γ) + P(A_3 \cap Γ)$$
$$= P(A_1)P(Γ/A_1) + P(A_2)P(Γ/A_2) + P(A_3)P(Γ/A_3) = 0,81$$

γ) $P(A_1/Γ) = \frac{P(A_1 \cap Γ)}{P(Γ)} = 0,6913$

20. Από έρευνα Εθνικού ΕΚΚΕ σε τυχαίο δείγμα 300 γυναικών με αντικείμενο απασχόλησης τη μητρότητα, προέκυψαν τα εξής:

Απασχόληση	ΑΡ. ΠΑΙΔΙΩΝ 0	1	2	3	4
Εργάζεται	80	75	40	5	0
Δεν εργάζεται	5	20	50	20	5

α) Αν επιλέξουμε στην τύχη μια γυναίκα:

I. Ποια η πιθανότητα να εργάζεται;

$$P(E) = \frac{80+75+40+5}{300} = 0,667$$

II. Ποια η πιθανότητα να μην έχει παιδιά

$$P(\Pi') = \frac{80+5}{300} = 0,283$$

III Ποια η πιθανότητα να εργάζεται και να έχει 2 παιδιά

$$P = \frac{40}{300} = 0,133$$

IV Ποια η πιθανότητα να εργάζεται ή να μην έχει παιδιά

$$P(E \cup \Pi') = P(E) + P(\Pi') - P(E \cap \Pi')$$
$$= 0,667 + 0,283 - \frac{80}{300} = 0,683$$

β) Ποια η P μια εργαζόμενη γυναίκα να μην έχει παιδιά;

$$P(\Pi'/E) = \frac{P(\Pi' \cap E)}{P(E)} = \frac{\overset{80/300}{0,266}}{0,667} = 0,4$$

γ) Ποια η P μια γυναίκα χωρίς παιδιά να εργάζεται

$$P(E/\Pi') = \frac{P(E \cap \Pi')}{P(\Pi')} = \frac{0,266}{0,283} = 0,940$$

ΑΝΑΜΕΝΟΜΕΝΗ ΤΙΜΗ = ΜΕΣΟΣ ΑΡΙΘΜΗΤΙΚΟΣ = $\Sigma x_i P(x_i)$

21. Η εμπειρική κατανομή πιθανοτήτων του αριθμού των παιδιών ανά οικογένεια, είναι η εξής σύμφωνα με την Ε.Σ.Υ.Ε.:

ΑΡΙΘΜΟΣ ΠΑΙΔΙΩΝ X	ΠΙΘΑΝΟΤΗΤΑ P	$x_i P(x_i)$	$(x_i - 1,84)^2 P(x_i)$
0	0,15	0	
1	0,25	0,25	
2	0,35	0,70	
3	0,15	0,45	
4	0,07	0,28	
5	0,02	0,10	
6	0,01	0,06	
ΣΥΝΟΛΑ	1	1,84	1,5944

α) Ποια η πιθανότητα μια οικογένεια να έχει 3 ή περισσότερα παιδιά;

$$P(x \geq 3) = P(3) + P(4) + P(5) + P(6)$$
$$= 0,15 + 0,07 + 0,02 + 0,01 = 0,25 \text{ ή } 25\%$$

β) Ποια η πιθανότητα να μην έχει κανένα παιδί;

$$P(x=0) = P(0) = 0,15$$

γ) Ποια η πιθανότητα να έχει το πολύ 2 παιδιά;

$$P(x \leq 2) = P(0) + P(1) + P(2) = 0,15 + 0,25 + 0,35 = 0,75$$

↑ Προσθέσαμε και τα μηδέν!

δ) Ποια η πιθανότητα να έχει τουλάχιστον 2 παιδιά;

$$P(x \geq 2) = 1 - P(x<2) = 1 - P(x \leq 1) = 1 - [P(0) + P(1)]$$

ε) Μέσος αριθμητικός (αναμενόμενη τιμή) και διακύμανση

$$\mu = E(x) = \Sigma x_i P(x_i) = 1,84 \text{ παιδιά}$$

$$\sigma^2 = Var(x) = \Sigma (x_i - \mu)^2 P(x_i) = 1,5944$$

ΤΥΠΟΙ

- Αναμενόμενη τιμή $E(x) = \mu = \Sigma x_i P(x_i)$
- Διακύμανση $Var(x) = \sigma^2 = \Sigma (x_i - \mu)^2 P(x_i)$
- Τυπική απόκλιση $\sigma = \sqrt{\sigma^2}$

22. Ένα συνοικιακό περίπτερο πουλάει μεταξύ άλλων το περιοδικό υπολογιστών Personal Computer. Με βάση τα στοιχεία των πωλήσεων των 5 τελευταίων ετών η κατανομή πιθανοτήτων των μηνιαίων πωλήσεων έχει ως εξής:

ΜΗΝΙΑΙΑ ΖΗΤΗΣΗ X	ΠΙΘΑΝΟΤΗΤΕΣ $P(X=x)$	$x_i P(x_i)$
9	0,20	1,80
10	0,35	3,5
11	0,30	3,3
12	0,15	1,8
		ΣΥΝΟΛΟ → 10,4

α) Ποια η πιθανότητα μηνιαίας ζήτησης τουλάχιστον 10 τευχών;

$P(x \geq 10) = P(10) + P(11) + P(12) = 0,80$

β) Ποια η αναμενόμενη τιμή (μαθηματική ελπίδα) της μηνιαίας ζήτησης;

$E(x) = \Sigma x_i P(x_i) = 10,4$ τεύχη ← Δεν στρογγυλοποιώ!

γ) Πόσα τεύχη πρέπει να προμηθεύεται το περίπτερο κάθε μήνα;

Αφού δεν έχει επιβάρυνση επιστροφής, 12 ή και κάποια παραπάνω.

Αν είχε επιβάρυνση τότε θα έπρεπε 12 για να μην υπάρχει χασούρα.

δ) Να κατασκευαστεί η κατανομή πιθανότητας του μηνιαίου κέρδους

ΜΗΝΙΑΙΟ ΚΕΡΔΟΣ Ψ	$P(\Psi=y)$
18 (2€ × 9 τεύχη)	0,20
20 (2€ × 10 τεύχη)	0,35
22	0,30
24	0,15

Έχω ήδη την κατανομή πιθανότητας ζήτησης (εκφώνηση).

Ψάχνω κατανομή πιθανότητας κέρδους.

Η πιθανότητα η μεταβλητή X να πάρει μια τιμή στο διάστημα x_1 έως x_2 ισούται με $P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx$

ε) Ποια η αναμενόμενη τιμή (μαθηματική ελπίδα) του μηνιαίου κέρδους;

$E(\Psi) = \Sigma \psi \cdot P(\psi) = 18 \cdot 0,20 + \dots + 24 \cdot 0,15 = 20,8€$

23 Στρίβουμε $n = 500$ φορές ένα νόμισμα. Ποια η πιθανότητα στα 500 στριψίματα, οι 177 φορές να είναι κεφάλι;

ΣΗΜΕΙΩΣΗ
$3! = 1 \cdot 2 \cdot 3 = 6$
$1! = 1$
$0! = 1$

Πρόκειται για πρόβλημα διωνυμικής κατανομής

"επιτυχία" → κεφάλι $P = 0,5$

Τύπος Διωνυμικής: $P(x) = \frac{n!}{x!(n-x)!} P^x (1-P)^{n-x}$

Επομένως $P(x=177) = \frac{500!}{177!(500-177)!} 0,5^{177} (1-0,5)^{500-177}$

24. Σε μια παρέα 3 ατόμων ισχύει ότι το 40% έχει κινητό

α) Ποια η πιθανότητα κανένας να μην έχει κινητό;

$P(x=0) = \frac{3!}{0!(3-0)!} 0,40^0 (1-0,40)^{3-0} = 0,216$

β) Ποια η P ένας να έχει κινητό;

$P(x=1) = \frac{3!}{1!(3-1)!} 0,4^1 (1-0,4)^{3-1} = 0,432$

Με τον ίδιο τρόπο $P(x=2) = 0,2880$ και $P(x=3) = 0,064$

γ) Ποια η P τουλάχιστον ένας να έχει κινητό;

$P(x \geq 1) = 1 - P(x=0) = 1 - 0,216 = 0,784$

Για "τουλάχιστον" παίρνουμε μείον το πολύ των προηγούμενη τιμή δ) Ποιά η πιθανότητα το πολύ ένας να έχει κινητό;

$P(x \le 1) = P(0) + P(1) = 0{,}216 + 0{,}432 = 0{,}648$

ε) Ποιος ο αναμενόμενος αριθμός ατόμων που έχουν κινητό;

$E(x) = n \cdot p = 3 \cdot 0{,}4 = 1{,}2$ παιδιά

25. Σε παραγωγή, το αναπόφευκτο ποσοστό ελαττωματικών είναι 0,04. Τα προϊόντα συσκευάζονται σε κουτιά 20 τεμαχίων. Ποιά η κατανομή πιθανότητας των ελαττωματικών προϊόντων αν γνωρίζουμε ότι ακολουθεί το διωνυμικό νόμο;

Από εκφώνηση: $n = 20$ $\quad p = 0{,}04$ → ΔΙΩΝΥΜΙΚΗ!

ΑΡΙΘΜΟΣ X ΕΛΑΤΤΩΜΑΤΙΚΩΝ	$P(X=x)$
0	0,4420
1	0,3683
2	0,1458
3	0,0364
4	0,0065
5	0,0009
6	0,0001
7	0,0000

ΕΚΕΙ ΠΟΥ ΜΗΔΕΝΙΖΕΤΑΙ ΣΤΑΜΑΤΑΜΕ

α) Αγοράζει κάποιος 1 κουτί: ποια η πιθανότητα να έχει μέσα 3 ελαττωματικά;

$P(x=3) = 0{,}0364$

β) το πολύ 2 ελαττωματικά;

$P(x \le 2) = P(0) + P(1) + P(2) = 0{,}9561$

γ) τουλάχιστον 3

$P(x \ge 3) = 1 - P(x \le 2) = 1 - 0{,}9561 = 0{,}0439$

δ) Αγοράζει κάποιος 1000 κουτιά: Πόσα από αυτά θα έχουν 2 ελαττωματικά;

$P(2) = 0{,}1458$ Επομένως $1000\,P(2) = 145{,}8$ κουτιά

Πιθανότητα εξαιρετικά μικρή → Poison!

Δίνεται p και n → Διωνυμική!

26. Σε ένα τραπεζικό κατάστημα, ύστερα από παρατηρήσεις εβδομάδων φτάνουν κατά Μ.Ο. 3 πελάτες το λεπτό. Ποια η P σε ένα λεπτό:

Κατανομή Poison $P(x) = \dfrac{e^{-\lambda} \cdot \lambda^{x}}{x!}$

με $\lambda = E(x) = n \cdot p$

α) να φτάσουν 0 πελάτες

$$P(x=0) = \frac{e^{-3} \cdot 3^{0}}{0!} = 0{,}0498$$

β) να φτάσουν 1 πελάτες

$$P(x=1) = \frac{e^{-3} \cdot 3^{1}}{1!} = 0{,}1494$$

γ) να φτάσουν το πολύ 5 πελάτες

$P(x \le 5) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$

δ) να φτάσουν τουλάχιστον 5 πελάτες

$P(x \ge 5) = 1 - P(x \le 4)$

ε) Κάντε τον πίνακα της Poison

X	$P(X=x)$
0	0,498
1	0,1494
2	⋮
3	⋮
⋮	
12	0,0001

→ Όλα βγαίνουν με τον τύπο $P(x) = \dfrac{e^{-\lambda} \cdot \lambda^{x}}{x!}$

→ Σταματάω όταν μηδενίζεται η P.

στ) Ποια η πιθανότητα να έλθουν 4 πελάτες στα 3 λεπτά;

Πρέπει να αλλάξει το λ ώστε να συμφωνεί με τα ζητούμενα.

$\lambda' = 9$ πελάτες / τρίλεπτο $\left\{ \begin{array}{l} 3 \text{ πελάτες} \to 1 \text{ λεπτό} \\ 9 \text{ πελάτες} \to 3 \text{ λεπτά} \end{array} \right.$

Επομένως $P(4) = \dfrac{e^{-9} \cdot 9^{4}}{4!}$

27. Πότε είναι προτιμότερο να επιλέξουμε Poison / Διωνυμική;

$n \le 20$ (μικρό), $p \ge 0,05$ (μεγάλο) } Διωνυμική Κατανομή. Η Poison είναι προσέγγιση της διωνυμικής. Οπότε πρώτα πάμε στη διωνυμική κι αν δεν υπάρχουν τα στοιχεία πάμε

$n \ge 20$ (μεγάλο), $p \le 0,05$ (μικρό) } Poison Κατανομή στην Poison.

Στην Poison $\mu = \bar{x} = E(x) = S^2$ (μέσος αρ. = διακύμανση)

28. Έχει παρατηρηθεί ότι σε αίθουσα εξετάσεων με 50 φοιτητές οι 10 αντιγράφουν. Επιλέγονται τυχαία 5 φοιτητές. Ποια η κατανομή πιθανότητας των φοιτητών που αντιγράφουν; ΔΙΩΝΥΜΙΚΗ

Σύμφωνα με τα δεδομένα: $n = 5$ και $p = \frac{10}{50} = 0,20$

ΦΟΙΤΗΤΕΣ ΠΟΥ ΑΝΤΙΓΡΑΦΟΥΝ X	$P(X=x)$
0	0,3277
1	0,4096
2	0,2048
3	0,512
4	0,0064
5	0,0003

• Ποιος αριθμός φοιτητών είναι πιο πιθανό να αντιγράφει;

ΛΥΣΗ 1
Η μεγαλύτερη πιθανότητα του πίνακα (0,4096)
Δηλαδή $= 1$

ΛΥΣΗ 2
Η μαθηματική ελπίδα:
$E(x) = n \cdot p = 5 \cdot 0,20 = 1$

29. Από τις στατιστικές του Υπουργείου Συγκοινωνιών είναι γνωστό ότι η πιθανότητα να συμβεί ατύχημα σε ένα αυτοκίνητο που διέρχεται από συγκεκριμένη διασταύρωση μεταξύ 7:00 π.μ. και 8:00 π.μ. είναι 0,7%. Επίσης από τις ίδιες στατιστικές προκύπτει ότι από τη συγκεκριμένη διασταύρωση διέρχονται κάθε πρωί κατά μ.ο. 1000 αυτοκίνητα. Ποια η κατανομή πιθανοτήτων του ημερήσιου αριθμού των ατυχημάτων στη συγκεκριμένη διασταύρωση;

ΑΤΥΧΗΜΑΤΑ X	$P(X=x)$
0	0,4966
1	0,3476
2	0,1217
3	0,0284
4	0,0050
5	0,0007
6	0,0001

Είναι $\lambda = n \cdot p = 0,7$ ατυχήματα

$p = \frac{0,7}{1000} = 0,0007$

$n = 1000$

Τύπος Διωνυμικής Κατανομής:

$$P(X=x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

30. Για ένα δείγμα $n = 400$ ισχύει $\bar{x} = 50$ και $S = 15$. Υπολογίστε το διάστημα εμπιστοσύνης.

Υπολογίζουμε το τυπικό σφάλμα του $\bar{x}$: $S_{\bar{x}} = S/\sqrt{n} = \frac{15}{\sqrt{400}} = 0,75$

ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ: $\bar{x} - 2S_{\bar{x}} < \mu < \bar{x} + 2S_{\bar{x}}$ (95%)

$50 - 2 \cdot 0,75 < \mu < \bar{x} + 2 \cdot 0,75$

$48,5 < \mu < 51,5$

ΑΛΛΑ ΔΙΑΣΤΗΜΑΤΑ:

$\bar{x} - S_{\bar{x}} < \mu < \bar{x} + S_{\bar{x}}$ (68%)

$\bar{x} - 3S_{\bar{x}} < \mu < \bar{x} + 3S_{\bar{x}}$ (99,7%)

31. Μια εταιρεία κατασκευής μπαταριών για κινητά τηλέφωνα παράγει μπαταρία με μέσο χρόνο διάρκειας σε κατάσταση αναμονής 191 ώρες και τυπική απόκλιση περίπου 5 ώρες.

α) Αγοράζει κάποιος μια μπαταρία. Ποια η πιθανότητα να έχει διάρκεια ζωής κάτω από 180 h; Τυποποιημένη κανονική κατανομή.

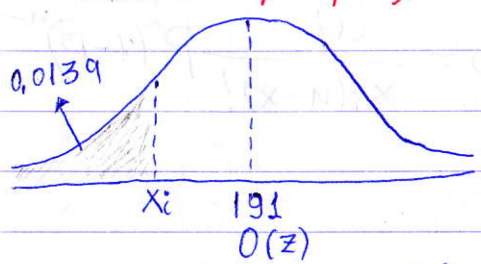
Ψάχνω το Z για να πάω στους πίνακες:

$$P(x<180) = P\left(Z < \frac{180-191}{5}\right) = P(Z<-2,2)$$

Πηγαίνω σε πίνακα εμβαδών τυποποιημένης κανονικής κατανομής για $Z=-2,2 \leftrightarrow P(x<180)=0,0139 \simeq 1,39\%$

Τύπος: $P(x<x_i) = P\left(Z<\frac{x_i-\mu}{\sigma}\right)$

β) Ποια είναι η μέγιστη διάρκεια ζωής κάτω της οποίας θα ζήσουν το 1,39% των μπαταριών; (αντίστροφο)



Στο X_i αντιστοιχεί ένα Z με δείκτη $0,0139 \rightarrow Z_{0,0139}$

$P(x<X_i) = 0,0139$

$P(Z<Z_{0,0139}) = 0,0139$

$-2,2 = \frac{X_i-191}{5} \leftrightarrow X_i = 180$ ώρες

γ) Αν κάποιος αγοράσει 1000 μπαταρίες, πόσες θα ζήσουν κάτω από 180h;

$1000 \times P(x<180) = 13,9 \simeq 14$ μπαταρίες.

Αφαιρώ από τη μονάδα γιατί οι πίνακες μετράνε τιμές από $-\infty$ μέχρι μια συγκεκριμένη τιμή. Άρα από το συνολικό εμβαδόν (=1) αφαιρώ όλες τις προηγούμενες P.

δ) Ποια η πιθανότητα μια μπαταρία να έχει διάρκεια ζωής πάνω από 195h;

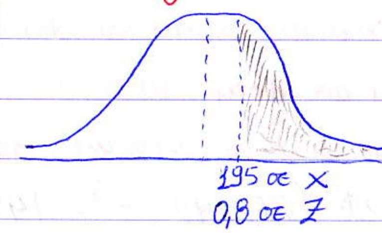
$P(x>195) = 1-P(x<195)$

το κάνουμε Z

$1-P(Z<0,8) = 1-0,78814$

$= 0,21186$

$\frac{195-191}{5} = 0,8$ Από πίνακα $0,8 \rightarrow 0,78814$



↑ αντίστροφο

ε) Ποια η ελάχιστη διάρκεια ζωής που ζουν το 21,186% των μπαταριών;

Πιθανότητα του X να είναι μεγαλύτερο της X_i που είναι σε έκφραση εμβαδού 21,186% $P(x>x_i) = 0,21186$

$P(Z>Z_i) = 0,21186 \leftrightarrow 1-P(Z<Z_i) = 0,21186$

$P(Z<Z_i) = 0,78814 \leftrightarrow Z_i = 0,8 = \frac{X_i-191}{5} \leftrightarrow X_i = 195h.$

στ) Ποια η P η διάρκεια ζωής της μπαταρίας να είναι μεταξύ 180-195h.

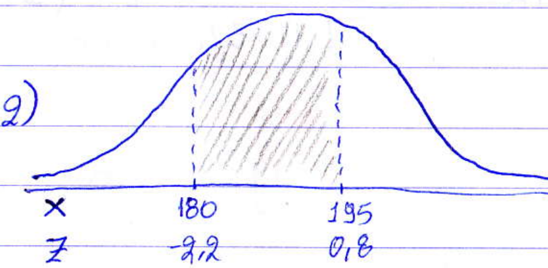
$$P(180<Z<195) = P\left(\frac{180-191}{5} < Z < \frac{195-191}{5}\right)$$

$= P(-2,2<Z<0,8)$

$= P(Z<0,8) - P(Z<-2,2)$

$= 0,78814 - 0,0139$

$= 0,77424$



32. Αν υποτεθεί ότι οι ετήσιες αποδοχές εργαζομένων στη Χ. Εταιρεία ακολουθούν την κανονική κατανομή με μέσο αριθμητικό 26542 € και τυπική απόκλιση 14.529 €, επιλέγουμε έναν υπάλληλο στην τύχη.

α) Ποια η P να έχει ετήσιες αποδοχές το πολύ 7944,88€;

$X \sim N(26.542, \sigma^2 = 14529^2)$

$P(X < 7.944,88) = P\left(Z < \frac{7944,88 - 26542}{14.529}\right)$

$P(Z < -1,28) = 0,10027$

β) Αν στην εταιρεία εργάζονται 190 υπάλληλοι, πόσοι έχουν ετήσιες αποδοχές το πολύ 7944,88 €;

$190 \times 0,10027 = 19$ υπάλληλοι

SOS

γ) Εάν επιλέξουμε 15 υπαλλήλους στην τύχη ποια η πιθανότητα 3 από αυτούς να έχουν ετήσιες αποδοχές 7944,88€.

Χρήση Poison ή Διωνυμικής.

$n = 15 \qquad p = 0,10027 \simeq 0,10$

$$P(X=3) = \frac{15!}{3!(15-3)!} \cdot 0,1^3 (1-0,1)^{15-3} = 0,1285$$

33. Μια μεγάλη κατασκευαστική εταιρεία πρόκειται να χρηματοδοτήσει έναν ποδηλατικό αγώνα που διοργανώνει ο δήμος της πόλης. Αναμένεται να δηλώσουν συμμετοχή 30.000 άτομα. Από προηγούμενες εκδηλώσεις είναι γνωστό ότι ο χρόνος κάλυψης της διαδρομής κατανέμεται κανονικά με μέσο 180 λεπτά και τυπική απόκλιση 20 λεπτά. $X \sim N(180, 20^2)$

α) Πόσα μετάλλια θα χρειαστούν εάν η εταιρεία δώσει από ένα σε όσους τερματίσουν σε χρόνο κάτω από 2 ώρες;

$P(X < 120) = P\left(Z < \frac{120-180}{20}\right) = P(Z < -3) = 0,00135$

Επομένως $0,00135 \times 30.000$ άτομα $= 41$ μετάλλια.

β) Πόσα φούτερ θα χρειαστούν εάν η εταιρεία δώσει δωρεάν από ένα σε όσους τερματίσουν σε χρόνο κάτω από 3,5 αλλά πάνω από 2 ώρες;

$P(120 < X < 210) = P\left(\frac{120-180}{20} < Z < \frac{210-180}{20}\right)$

$= P(-3 < Z < 1,5) = P(Z < 1,5) - P(Z < -3)$

$= 0,93319 - 0,00135 = 0,93184$ φουτερ.

Επομένως $30.000 \times 0,93184 = 27.955$ φουτερ.

γ) Πόσα μπλουζάκια θα χρειαστούν για όσους τερματίσουν σε χρόνο $> 3,5$ ώρες;

Α' τροπος.

$30.000 - 41 - 27.955 = 2004$ μπλουζακια

Β' τροπος

$P(X > 210) = 1 - P(X < 210) = 1 - P\left(Z < \frac{210-180}{20}\right)$

$= 1 - 0,93184 = 0,06681 \rightarrow 30000 \times 0,06681 = 2004$

$X \sim N \rightarrow (\mu = n \cdot p \,,\, S^2)$

34. Η πιθανότητα να απαντήσει ένα τηλεφωνικό κέντρο είναι 0,6. Επιχειρούμε να καλέσουμε 20 φορές.

α) Ποια η πιθανότητα στις 9 φορές να είναι ελεύθερο το κέντρο;

$n = 20 \quad p = 0{,}6 \quad \mu = n \cdot p = 12 \quad \sigma^2 = n \cdot p(1-p) = 4{,}8$

ΔΙΩΝΥΜΙΚΗ

$$P(x=9) = \frac{20!}{9!(20-9)!} 0{,}6^9 (1-0{,}6)^{20-9} = 0{,}0710$$

ΚΑΝΟΝΙΚΗ

$$P(x=9) = P(8{,}5 < x < 9{,}5) = P\left(\frac{8{,}5-12}{\sqrt{4{,}8}} < Z < \frac{9{,}5-12}{\sqrt{4{,}8}}\right)$$

$$= P(-1{,}6 < Z < -1{,}14) = P(Z < 1{,}14) - P(Z < -1{,}6)$$

$$= 0{,}12714 - 0{,}05480 = 0{,}07234$$

35. Η πιθανότητα να απαντήσει το τηλεφωνικό κέντρο είναι 22%. Επιχειρώ 100 φορές. Ποια η πιθανότητα 8 φορές να είναι ελεύθερο;

Διωνυμική → Δε βγαίνει (μεγάλο n) ⟹ Κανονική

Poison → Δεν υπάρχει

$P = 0{,}22 \quad n = 100 \quad \lambda = n \cdot p = 22 \;(\rightarrow E(x) = \mu) \quad \sigma^2 = n \cdot p(1-p)$

$$P(x=8) = P(7{,}5 < X < 8{,}5) = P\left(\frac{7{,}5-22}{\sqrt{22}} < Z < \frac{8{,}5-22}{\sqrt{22}}\right)$$

Διόρθωση συνέχειας

$$= P(-3{,}09 < Z < -2{,}88) = 0{,}00099$$

36. Έστω ότι ο χρόνος σε λεπτά που χρειάζεται για τις απαντήσεις ενός γραπτού διαγωνίσματος είναι $X \sim N(70, 12^2)$. Πότε πρέπει να τελειώσει ο χρόνος της εξέτασης ώστε το 90% των εξεταζομένων να προλάβει να ολοκληρώσει το διαγώνισμα;

$$P(X < x_1) = P(Z < z_1) = 0{,}90$$

Ψάχνω στον πίνακα με τα εμβαδά της τυποποιημένης κανονικής για εμβαδόν πλησιέστερο στο 0,9

$z_1 = 1{,}28 \rightarrow 0{,}89973$

Επομένως $z_1 = \frac{x_1 - 70}{12} \Leftrightarrow 1{,}28 = \frac{x_1 - 70}{12}$

$x_1 = 0{,}8536$ λεπτά

37. Ο Διευθυντής του αεροδρομίου μελετά τις καθυστερήσεις των αναχωρήσεων. Από τα στατιστικά στοιχεία προκύπτει ότι σχεδόν το 84,1% των πτήσεων έχουν καθυστέρηση μικρότερη των 25 λεπτών ενώ περίπου το $\frac{1}{5}$ των πτήσεων καθυστερούν το πολύ 15,8 λεπτά. Με την υπόθεση ότι ο χρόνος καθυστέρησης κατανέμεται κανονικά ποια είναι η μέση διάρκεια (μ) και η τυπική απόκλιση (σ) των καθυστερήσεων;

Γνωρίζουμε: $\left.\begin{matrix} P(x<25) = 0{,}841 \\ P(x<15{,}8) = 0{,}20 \end{matrix}\right\} \begin{matrix} P\left(Z < \frac{25-\mu}{\sigma}\right) = 0{,}841 \\ P\left(Z < \frac{15{,}8-\mu}{\sigma}\right) = 0{,}20 \end{matrix}$

Από Πίνακες: $\left.\begin{matrix} \frac{25-\mu}{\sigma} = 1 \\ \frac{15{,}8-\mu}{\sigma} = -0{,}84 \end{matrix}\right\} \left.\begin{matrix} 25-\mu = \sigma \\ 15{,}8-\mu = -0{,}84 \end{matrix}\right\} \begin{matrix} \mu = 20 \text{ λεπτά} \\ \sigma = 5 \text{ λεπτά} \end{matrix}$

38. Από στατιστικές αεροπορικής που ειδικεύεται σε πτήσεις με αεροπλάνα μέχρι 100 θέσεων είναι γνωστό ότι η P να ακυρώσει ένας επιβάτης την κράτησή του είναι 14%.

α) Ποια η πιθανότητα στην επόμενη πτήση αεροπλάνου 15 θέσεων να ακυρώσουν κράτηση 2 επιβάτες;

$p=0{,}14 \quad n=15$ Διωνυμική

$$P(x=2)=\frac{15!}{2!(15-2)!}0{,}14^2(1-0{,}14)^{15-2}=0{,}2897$$

β) Ποια η πιθανότητα την επόμενη πτήση 100 θέσεων να ακυρώσουν 10 επιβάτες; Διωνυμική → Όχι γιατί $n>20$

Poisson → όχι γιατί $\lambda=np=14$ ΔΕΝ υπάρχει

Πάμε με κανονική διορθώνοντας την ασυνέχεια:

$$P(9{,}5<X<10{,}5)=P\left(\frac{9{,}5-14}{3{,}47}<Z<\frac{10{,}5-14}{3{,}47}\right)$$

$\mu=n\cdot p=14$

$s=\sqrt{np(1-p)}=3{,}47$

$$=P(-1{,}3<Z<-1)$$

$$=\underset{0{,}15866}{P(Z<-1)}-\underset{0{,}09680}{P(Z<-1{,}3)}=0{,}06186$$

ή 6,186%

γ) Αν κάθε ακύρωση σε πτήσεις των 15 θέσεων συνεπάγεται απώλεια εσόδων ύψους 100€ ποια είναι η μέση αναμενόμενη οικονομική απώλεια (μαθηματική ελπίδα) για κάθε πτήση;

Μαθηματική ελπίδα είναι ο μέσος σε κατανομή πιθανοτήτων.

$E(x)=\sum x P(x=x)$

x	$P(x=x)$	Απώλεια εσόδων Y
0	0,1041	0
1	0,2542	100
2	0,2897	200
3	0,2044	300
⋮	⋮	
8	0,0003	800

$E(Y)=\sum Y_i P(x=x_i)$

$=0\times0{,}1041+100\times0{,}2542\ldots$

$=209{,}91$€

αναμενόμενη οικονομική απώλεια.

39. Αναφέρετε τις βασικές ιδιότητες της Διωνυμικής/Poisson/Κανονικής.

ΔΙΩΝΥΜΙΚΗ

Μέσος αριθμητικός $\mu=E(x)=n\cdot p$

Διακύμανση $\sigma^2=n\cdot p(1-p)$

Τυπική απόκλιση $\sigma=\sqrt{np(1-p)}$

$$f(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

POISSON

Μέσος αριθμητικός $\lambda=E(x)=n\cdot p$

Διακύμανση $\sigma^2=\lambda$

Τυπική απόκλιση $\sigma=\sqrt{\lambda}$

$$f(x)=\frac{e^{-\lambda}\cdot\lambda^x}{x!}$$

ΚΑΝΟΝΙΚΗ

$\mu=T_0=M$

μέσος αριθμ. / μέγιστη συχνότητα / διάμεσος

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

40. Μια επιχείρηση παραγωγής ελαστικών διαθέτει 40 ελαστικά τα οποία έχουν διάρκεια ζωής κατά μέσο όρο 22000 km με μέση απόκλιση 1600 km. Να υπολογιστεί το διάστημα εμπιστοσύνης με πιθανότητα 95% για τη μέση διάρκεια ζωής ΟΛΩΝ των ελαστικών που αγοράζονται από την επιχείρηση.

$1-a = 0{,}95$ → Επίπεδο εμπιστοσύνης $n = 40$

95% → Διάστημα εμπιστοσύνης $\bar{x} = 22.000$

$a = 5\%$ → Περιθώριο σφάλματος $s = 1600$

Επειδή $n \geq 30$ επιλέγω κανονική κατανομή.

$$\bar{x} - |Z_{a/2}| S_{\bar{x}} < \mu < \bar{x} + |Z_{a/2}| S_{\bar{x}}$$

$$22000 - 1{,}96 \frac{1600}{\sqrt{40}} < \mu < 22.000 + 1{,}96 \frac{1600}{\sqrt{40}}$$

$Z_{\frac{0{,}05}{2}} = Z_{0{,}025} = -1{,}96$ $S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1600}{\sqrt{40}}$

Από πίνακα εμβαδών τυποποιημένης κανονικής

$$21500{,}15 < \mu < 22.495{,}58$$

Επομένως αν τραβήξω 100 δείγματα, των 95 οι μέσοι χρόνοι ζωής θα είναι σε αυτά τα όρια (km). Στα άλλα 5 ο μέσος θα έχει σφάλμα.

*Αν το $n < 30$ η κατανομή δειγματοληψίας του μέσου ακολουθεί t-student με μέσο μ και τυπική απόκλιση $S_{\bar{x}} = \frac{s}{\sqrt{n}}$

- ΑΝ σ_x^2 γνωστή → κανονική κατανομή με $\sigma^2 = \frac{\sigma_x^2}{n} \to \sigma = \frac{\sigma_x}{\sqrt{n}}$
- ΑΝ σ_x^2 άγνωστη → $n \geq 30$ κανονική με τυπική α. $S_{\bar{x}} = \frac{s}{\sqrt{n}}$
  → $n < 30$ t-student με $S_{\bar{x}} = \frac{s}{\sqrt{n}}$

41. Ο μέσος χρόνος αναμονής για την παραλαβή αποσκευών σε αεροδρόμιο 14 επιβατών είναι 18,57 λεπτά με τυπική απόκλιση 2,76 λεπτά. Να υπολογιστεί το 99% διάστημα εμπιστοσύνης του μέσου χρόνου αναμονής αποσκευών στο αεροδρόμιο.

$n = 14$ $\bar{x} = 18{,}57$ $s = 2{,}76$ $a = 1\%$

→ $n < 30$ → διακύμανση πληθυσμού άγνωστη → t-student

$$\bar{x} - t_{v,\frac{a}{2}} S_{\bar{x}} < \mu < \bar{x} + t_{v,\frac{a}{2}} S_{\bar{x}}$$

$S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2{,}76}{\sqrt{14}}$ $t_{13\,\frac{0{,}01}{2}} = 3{,}012$ Από πίνακα με τιμές κριτηρίου t-student για $a = 1\%$ και $v = 13$

$v = n - 1 = 13$

$$16{,}34 < \mu < 20{,}8$$

Σε αυτές τις τιμές κυμαίνεται ο μέσος χρόνος αναμονής.

42. Ο διευθυντής Marketing μιας εταιρείας κινητής τηλεφωνίας θέλει να εκτιμήσει τη μέση διάρκεια συνδιαλέξεων συγκεκριμένης ομάδας συνδρομητών με σκοπό την προώθηση ενός νέου είδους συμβολαίου κινητής τηλεφωνίας. Η επιθυμητή ακρίβεια της εκτίμησης του διαστήματος εμπιστοσύνης του μέσου είναι $e = \pm 2$ δευτερόλεπτα σε επίπεδο εμπιστοσύνης 95%. Παλαιότερη έρευνα με παρόμοιο αντικείμενο αλλά για άλλη κατηγορία συνδρομητών είχε δώσει εκτίμηση της τυπικής απόκλισης $s = 30$ δευτ. Τι μέγεθος δείγματος χρειάζεται για να έχουμε την επιθυμητή ακρίβεια;

$e = \pm 2$ 95% εμπιστοσύνης $\rightarrow a = 5\%$ $s = 30$

$$n = \frac{Z^2_{a/2} \cdot \sigma^2}{e^2} \Leftrightarrow \underset{\substack{\downarrow \\ \text{μέγεθος δείγματος}}}{n} = \frac{1{,}96^2 \cdot 30^2}{2^2} = 864{,}36 \simeq 865$$

Σφάλμα εκτίμησης $S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{30}{\sqrt{865}}$

43. Οι ετήσιες αποδοχές των εργαζομένων σε μια εταιρεία ακολουθούν κανονική κατανομή με μέσο μ και σ = 15.000€. Από τυχαίο δείγμα 100 εργαζομένων προκύπτει άθροισμα αμοιβών $\Sigma x = 2.700.000$€ (σύνολο ετήσιων αποδοχών 100 εργαζομένων του δείγματος).

α) Ποιο το διάστημα εμπιστοσύνης μέσου με 95% πιθανότητα;

$$\bar{x} - |Z_{a/2}|\sigma_{\bar{x}} < \mu < \bar{x} + |Z_{\frac{a}{2}}|\sigma_{\bar{x}}$$

$\bar{x} = \frac{\Sigma x}{n} = \frac{2700}{100} = 27$

$$24{,}66 < \mu < 29{,}94$$

$a = 5\% \leftrightarrow \left|Z_{\frac{0{,}05}{2}}\right| = 1{,}96$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1{,}5$

β) Ποια η ακρίβεια εκτίμησης (σε χιλιάδες ευρώ) με πιθανότητα 95%;

$e = |Z_{a/2}|\sigma_{\bar{x}} = 1{,}96 \times 1{,}5 = 2{,}94$

ΑΚΡΙΒΕΙΑ ΕΚΤΙΜΗΣΗΣ (e) ή αλλιώς, δειγματοληπτικό σφάλμα

- $e = \bar{x} - \mu$
- $e = |Z_{a/2}|\sigma_{\bar{x}}$
- $e = (P - \pi)$
- $e = |Z_{a/2}|\frac{\sigma}{\sqrt{n}}$ γνωστή διακύμανση
- $e = |Z_{a/2}|\frac{s}{\sqrt{n}}$ άγνωστη διακύμανση

γ) Ποιο μέγεθος δείγματος απαιτείται προκειμένου το σφάλμα της εκτίμησης του μέσου με πιθανότητα 95% να μειωθεί στο μισό;

$n = \frac{Z^2_{a/2} \cdot \sigma^2}{e^2}$ για e όμως χρησιμοποιώ το:

$e = \frac{2{,}94}{2} = 1{,}47$ για να μειώσω στο μισό το σφάλμα

Επομένως: $n = \frac{1{,}96^2 \cdot 15^2}{1{,}47^2} = 400$

Επομένως χρειάζεται n=400 ή αλλιώς, τετραπλασιασμός δείγματος.

δ.) Ποιο το διάστημα εμπιστοσύνης μέσα στα όρια του οποίου αναμένεται με πιθανότητα 95% της δαπάνης της εταιρίας για τις ετήσιες αποδοχές των υπαλλήλων της που είναι 190.

$\mu = \frac{\Sigma x}{N} \leftrightarrow \Sigma x = N \cdot \mu$ με N=190

Επομένως, έχοντας το 95% διάστημα εμπιστοσύνης:

$$190(24{,}66) < \underbrace{N \cdot \mu}_{\text{Συνολική δαπάνη πληθυσμού}} < 190(29{,}94)$$

44. Από 2000 καταστήματα πώλησης ηλεκτρικών ειδών μιας περιοχής παίρνουμε τυχαίο δείγμα 400 καταστημάτων από τα οποία τα 136 πωλούσαν σε τιμή μικρότερη από εκείνη που συνιστούσε ο κατασκευαστής.

α) Υπολογίστε το διάστημα εμπιστοσύνης μέσα στα όρια του οποίου αναμένεται με πιθανότητα 95% το ποσοστό όλων των καταστημάτων που πωλούν σε τιμή μικρότερη του τιμοκαταλόγου.

$N = 2000$

$x = 136$

$n = 400$

$p = \Pi = \frac{x}{n}$ → μονάδες που έχουν ένα συγκεκριμένο χαρακτηριστικό

p ↓ ποσοστό στο δείγμα, Π ↓ ποσοστό στον πληθυσμό

Τότε χρησιμοποιούμε διάστημα εμπιστοσύνης ποσοστού!

$$95\% : p - |Z_{\alpha/2}| S_p < \Pi < p + |Z_{\alpha/2}| S_p$$

$* \ p = \frac{x}{n} = \frac{136}{400} = 0{,}34$

$* \ |Z_{\frac{\alpha}{2}}| = |Z_{\frac{0{,}05}{2}}| = |-1{,}96| = 1{,}96$

$* \ S_p = \sqrt{\frac{p(1-p)}{n}} = 0{,}0237$

$$0{,}2906 < \Pi < 0{,}3864$$

$$29\% < \Pi < 38\%$$

β) Υπολογίστε το διάστημα εμπιστοσύνης στα όρια του οποίου αναμένεται με πιθανότητα 95% ο συνολικός αριθμός καταστημάτων που πωλούν σε τιμή μικρότερη του καταλόγου.

$$2000 \times 0{,}2906 < \underset{(N\cdot\Pi)}{X} < 2000 \times 0{,}3864$$

$$581 < X < 773$$

45. Ο υπεύθυνος Marketing εταιρείας κινητών τηλεφώνων θέλει να εκτιμήσει το ποσοστό των συνδιαλέξεων που αφορούν επαγγελματικές κλήσεις με επιθυμητή ακρίβεια $e = 1\%$ με επίπεδο εμπιστοσύνης 95%.

α) Τι μέγεθος δείγματος χρειάζεται για να επιτευχθεί η ακρίβεια αυτή;

$e = 1\%$ Υποθέτουμε προσωρινά ότι $p = 0{,}5$

$$n = \frac{Z^2_{\alpha/2}\, p(1-p)}{e^2} \Leftrightarrow n = \frac{1{,}96^2 \cdot 0{,}5(1-0{,}5)}{0{,}01^2} = 9604$$

Επομένως βάση της τιμής $p = 0{,}5$ απαιτείται έρευνα με μέγεθος δείγματος $n = 9604$ για να εξασφαλίσουμε ακρίβεια $e = 1\%$.

β) Μετά την ολοκλήρωση της έρευνας προέκυψε ποσοστό επαγγελματικών κλήσεων $p = 0{,}8$. Ποια η ακρίβεια του διαστήματος εμπιστοσύνης;

$$e = |Z_{\alpha/2}| S_p = |Z_{\alpha/2}| \sqrt{\frac{p(1-p)}{n}}$$ → ΑΚΡΙΒΕΙΑ ΕΚΤΙΜΗΣΗΣ

$$e = 1{,}96 \sqrt{\frac{0{,}8(1-0{,}8)}{9.604}} = 0{,}008 < 0{,}01$$

Το σφάλμα είναι μικρότερο από το επιθυμητό. Υπολογίζουμε εκ νέου το μέγεθος του δείγματος που απαιτείται για την ακρίβεια που θέλουμε ώστε να μειώσουμε το κόστος της έρευνας:

$$n = \frac{1{,}96^2 \cdot 0{,}8(1-0{,}8)}{0{,}01^2} = 6146{,}56$$

ΣΗΜΕΙΩΣΗ $C = C_0 + n \cdot c$ → κόστος δειγματοληπτικής μονάδας

C ↓ Κόστος έρευνας, C_0 ↓ Πάγια Έξοδα, n ↓ μέγεθος δείγματος

46. Το τμήμα Marketing της Εταιρίας Φυσικού Αερίου θέλει να εκτιμήσει το ποσοστό των οικειών που χρησιμοποιεί κεντρικό σύστημα θέρμανσης με αέριο (γκάζι). Η έρευνα γίνεται για πρώτη φορά και καμιά πληροφορία δεν υπάρχει για το πιθανό ποσοστό χρήσης γκαζιού. Τι μέγεθος δείγματος απαιτείται για πιθανό μέγεθος σφάλματος ± 4% με επίπεδο εμπιστοσύνης 95%.

Υποθέτουμε ότι $P=0,5$

$$n = \frac{Z^2_{\alpha/2} \cdot P(1-P)}{e^2} = \frac{1,96^2 \cdot 0,5(1-0,5)}{0,04^2} = 600$$

παρατηρήσεις

β) Η έρευνα διεξήχθηκε και βρέθηκε ότι το ποσοστό χρήσης αερίου ανέρχεται σε 20% των οικιών. Ποιο είναι το διάστημα εμπιστοσύνης του ποσοστού Π του πληθυσμού σε επίπεδο εμπιστοσύνης 95%;

95% $\quad p - |Z_{\alpha/2}| S_p < \Pi < p + |Z_{\alpha/2}| S_p$

* $p = 0,20$
* $|Z_{\alpha/2}| = 1,96$

* $S_p = \sqrt{\frac{P(1-P)}{n}} = 0,0163$

Επομένως: $0,168 < \Pi < 0,232$

$16,8\% < \Pi < 23,2\%$

Ποιο μικρό από αυτό που βρήκαμε! Για να βρω το δείγμα που χρειάζεται το υπολογίζω από την αρχή βλ. προηγούμενη

γ) Με τι σφάλμα έγινε η εκτίμηση;

$$e = |Z_{\alpha/2}| \sqrt{\frac{P(1-P)}{n}} = 1,96 \sqrt{\frac{0,20(1-0,20)}{600}} = 0,032$$

Αλλος τροπος (Προσεγγιστικος)

$$e = \frac{\text{ΠΑΝΩ ΟΡΙΟ} - \text{ΚΑΤΩ ΟΡΙΟ}}{2} = \frac{0,232-0,168}{2} = 0,032$$

47. Για το πρόβλημα της προηγούμενης άσκησης ο Εμπορικός Διευθυντής της εταιρίας ισχυρίζεται ότι η ιδανική ακρίβεια της εκτίμησης είναι ± 2%. Τι μέγεθος δείγματος απαιτείται για πιθανό μέγεθος σφάλματος ± 2% με επίπεδο εμπιστοσύνης 95%; $e = \pm 2\%$

$$n = \frac{1,96^2 \cdot 0,5(1-0,5)}{0,02^2} = 2401 \rightarrow$$

Συντηρητική εκτίμηση με υποθετικό $p=0,5$

β) Ποια η προβλεπόμενη δαπάνη διεξαγωγής της έρευνας δεδομένου ότι το κόστος ανά ερωτηματολόγιο ανέρχεται σε 5 €;

$$C = n \cdot c = 2401 \cdot 5 = 12.005\ €$$

γ) Αν μπορούν να δαπανηθούν 8.000 € πόσα άτομα μπορούν να ερωτηθούν και τι σφάλμα θα προκύψει;

Έχουμε $n \cdot 5 = 8000 \Leftrightarrow n = 1600$

$$e = |Z_{\frac{a}{2}}| \sqrt{\frac{P(1-P)}{n}}$$

ΣΗΜΕΙΩΣΗ ΘΕΩΡΙΑΣ → Διαστήματα εμπιστοσύνης 95%

- ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΜΕΣΟΥ (με Z)

$$\bar{X} - |Z_{\alpha/2}| \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + |Z_{\alpha/2}| \frac{\sigma}{\sqrt{n}}$$

- ΠΟΣΟΣΤΟΥ Π ΤΟΥ ΠΛΗΘΥΣΜΟΥ

$$P - |Z_{\alpha/2}| S_P < \Pi < P + |Z_{\alpha/2}| S_P \quad \text{με } S_p = \sqrt{\frac{P(1-P)}{n}}$$

- t-student ΜΕΣΟΥ

$$\bar{X} - t_{v,\frac{a}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{v,\alpha/2} \frac{S}{\sqrt{n}} \quad \text{με } v = n-1$$

48. Το τμήμα Marketing της Εταιρία Φυσικού Αερίου θέλει επίσης να εκτιμήσει το μέσο ύψος της μηνιαίας δαπάνης των οικειών που χρησιμοποιούν κεντρικό σύστημα θέρμανσης με αέριο. Επειδή η έρευνα γίνεται για πρώτη φορά και καμία πληροφορία δεν υπάρχει για το πιθανό ύψος των λογαριασμών της συγκεκριμένης κατηγορίας καταναλωτών, σαν εκτίμηση της τυπικής απόκλισης θα χρησιμοποιηθεί η αντίστοιχη εκτίμηση από παλαιότερη έρευνα μεταξύ των καταναλωτών για τη μηνιαία δαπάνη ηλεκτρικού ρεύματος που βρέθηκε ίση με 40 ευρώ.

α) Τι μέγεθος δείγματος απαιτείται για πιθανό μέγεθος σφάλματος ±3 ευρώ με επίπεδο εμπιστοσύνης 95%;

$s = 40 \qquad e = \pm 3$

Βρίσκω το μέγεθος του δείγματος:

$$n = \frac{Z_{\alpha/2}^2 \, s^2}{e^2} = \frac{1{,}96^2 \, 40^2}{3^2} = 683$$

β) Από την έρευνα προέκυψαν οι εξής εκτιμήσεις: μέση μηνιαία δαπάνη 200€ και τυπική απόκλιση 30. Ποιο είναι το διάστημα εμπιστοσύνης της μέσης μηνιαίας δαπάνης μ του πληθυσμού σε επίπεδο εμπ. 95%;

$$\bar{x} - |Z_{\alpha/2}| \frac{s}{\sqrt{n}} < \mu < \bar{x} + |Z_{\alpha/2}| \frac{s}{\sqrt{n}}$$

$$200 - 1{,}96 \frac{30}{\sqrt{683}} < \mu < 200 + 1{,}96 \frac{30}{\sqrt{683}}$$

$$197{,}5 < \mu < 202{,}25$$

Υπολογίζουμε και το σφάλμα της εκτίμησης

$$e = \frac{202{,}25 - 197{,}5}{2} = 2{,}37 \text{ ευρώ} \rightarrow \text{Προσεγγιστική αξία}$$

$$e = |Z_{\alpha/2}| \frac{s}{\sqrt{n}} = 1{,}96 \frac{30}{\sqrt{683}} = 2{,}25 \text{ ευρώ} \rightarrow \text{Ακριβές σφάλμα}$$

49. Αναφέρετε τα πιο συνηθισμένα διαστήματα εμπιστοσύνης: μέσου.

$$68\% \rightarrow \bar{x} - \sigma_{\bar{x}} < \mu < \bar{x} + \sigma_{\bar{x}}$$

$$95\% \rightarrow \bar{x} - \overset{\text{ή } 2}{1{,}96}\,\sigma_{\bar{x}} < \mu < \bar{x} + \overset{\text{ή } 2}{1{,}96}\,\sigma_{\bar{x}}$$

$$99{,}7\% \rightarrow \bar{x} - 3\sigma_{\bar{x}} < \mu < \bar{x} + 3\sigma_{\bar{x}}$$

↳ με $\sigma_{\bar{x}} = \sigma / \sqrt{n}$

ΣΗΜΕΙΩΣΗ ΘΕΩΡΙΑΣ

- ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΤΟΥ ΣΥΝΟΛΟΥ (Τ) ΜΕΣΟΥ + ΠΟΣΟΣΤΟΥ

$$N\left(\bar{x} - Z_{\alpha/2} \frac{s}{\sqrt{n}}\right) < N\mu < N\left(\bar{x} + Z_{\alpha/2} \frac{s}{\sqrt{n}}\right)$$

$$N\left(p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) < N \cdot \pi < N\left(p + Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

50. Τι ονομάζουμε επίπεδο στατιστικής σημαντικότητας α;

Ο έλεγχος των υποθέσεων εμπεριέχει κινδύνους και η αξιολόγηση των συνεπειών από μια λανθασμένη απόφαση είναι το βασικό κριτήριο της επιλογής του ορίου μεταξύ πιθανού και απίθανου. Το όριο αυτό ονομάζεται επίπεδο σημαντικότητας και συμβολίζεται με α.

51. Μια εταιρεία παρασκευής δέρματος επιθυμεί το μέσο πάχος του παραγόμενου δέρματος να είναι 4 mm με τυπική απόκλιση 0,1 mm. Μας απασχολεί αν η παραγωγή εξελίσσεται κανονικά, σύμφωνα με τις προδιαγραφές, δεδομένου ότι δείγμα 50 μετρήσεων έχει μέσο πάχος 4,02 mm.

$n = 50 \quad \bar{x} = 4{,}02 \quad \mu = 4\,mm \quad \sigma = 0{,}1\,mm$

1. Ορίζουμε έλεγχο υπόθεσης: $H_0: \mu = 4$ και $H_1: \mu \neq 4$
2. Εξετάζουμε ένα δείγμα $n = 50$ με $\bar{x} = 4{,}02$
3. Η διακύμανση γνωστή, επομένως:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad \text{με} \quad \sigma_{\bar{x}} = \sigma/\sqrt{n} = \frac{0{,}1}{\sqrt{50}} = 0{,}014$$

Αν $|Z| > |Z_{\alpha/2}|$ τότε απορρίπτεται η H_0

4. Εφόσον δε δίνεται, επιλέγω ένα επίπεδο σημαντικότητας, πχ $\alpha = 5\%$.

Επομένως:

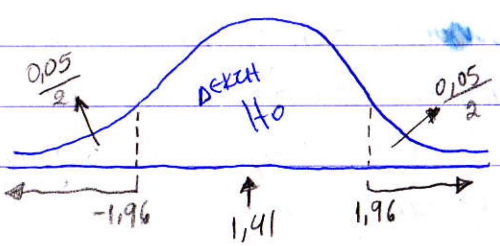
$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{4{,}02 - 4}{0{,}014} = 1{,}41$$

$$|Z_{\alpha/2}| = \left|Z_{\frac{0{,}05}{2}}\right| = 1{,}96$$

$1{,}41 \not> 1{,}96$ ΔΕΚΤΗ Η H_0

Επομένως δουλεύει σύμφωνα με τις προδιαγραφές

ΔΙΑΓΡΑΜΜΑΤΙΚΑ



• να ελεγχθεί η υπόθεση με $\alpha = 5\%$ αν: $H_0: \mu = 4 \quad H_1: \mu > 4$

Έχουμε μονοκατάληκτο κριτήριο ελέγχου (μόνο > 4)

Απορρίπτω τη H_0 αν $Z > Z_{1-\alpha}$

$Z = 1{,}41$

$Z_{1-\alpha} = Z_{1-0{,}05} = Z_{0{,}95} = 1{,}645$ ← Από πίνακα εμβαδών τυποποιημένης

$1{,}41 \not> 1{,}645$ ΔΕΚΤΗ Η H_0

• να ελεγχθεί η υπόθεση με $\alpha = 5\%$ αν $H_0: \mu = 4 \quad H_1: \mu < 4$

Απορρίπτουμε τη H_0 αν $Z < Z_\alpha$

$Z = 1{,}41$

$Z_\alpha = Z_{0{,}05} = -1{,}645$

$1{,}41 \not< -1{,}645$

Επομένως ΔΕΚΤΗ Η H_0

Συνοπτική Θεωρία

• Διακύμανση (σ^2): γνωστή
- ▷ Δικατάληκτο → απορρίπτω H_0 αν $|Z| > |Z_{\alpha/2}|$
- ▷ Μονοκατάληκτο → απορρίπτω H_0 αν $Z < Z_\alpha$ ή $Z > Z_{1-\alpha}$

ΔΕ: $\bar{x} - |Z_{\alpha/2}|\sigma_{\bar{x}} < \mu < \bar{x} + |Z_{\alpha/2}|\sigma_{\bar{x}}$ με: $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

• Διακύμανση (σ^2): άγνωστη
- ▷ Δικατάληκτο → $n \geq 30$ $\bar{x} - |Z_{\alpha/2}|S_{\bar{x}} < \mu < \bar{x} + |Z_{\alpha/2}|S_{\bar{x}}$ με $S_{\bar{x}} = \frac{s}{\sqrt{n}}$
  - → $n < 30$ $\bar{x} - |t_{v,\frac{\alpha}{2}}| < \mu < \bar{x} + |t_{v,\frac{\alpha}{2}}|$

    Απορρίπτω H_0 αν $|t| > |t_{v,\frac{\alpha}{2}}|$
- ▷ Μονοκατάληκτο → Απορρίπτω H_0 αν $t < -t_{v,\frac{\alpha}{2}}$ ή $t > t_{v,\alpha/2}$

52. Η βιομηχανία κάνει δοκιμαστική παραγωγή για νέο υλικό που θα χρησιμοποιηθεί για κάλυμα μικρών σκαφών. Το πάχος του νέου υλικού καθορίζεται σε 5 mm. Ο υπεύθυνος για να κάνει τις τελικές ρυθμίσεις στη μηχανή παραγωγής θέλει να ελέγξει αν ο μέσος της δοκιμαστικής παραγωγής ισούται με 5 mm.

Από δείγμα 30 μετρήσεων προέκυψαν τα εξής αποτελέσματα: $\Sigma x^2 = 760{,}2$ και $\Sigma x = 150{,}6$. Μπορεί ο υπεύθυνος παραγωγής να προχωρήσει στην κύρια παραγωγή με βάση τις ρυθμίσεις που έγιναν στη δοκιμαστική παραγωγή; $(a = 5\%)$

Υπόθεση: $H_0: \mu = 5$ και $H_1: \mu \neq 5$ (δικατάληκτο)

Διακύμανση άγνωστη: t-student

Απορρίπτω την H_0 αν $|t| > \left|t_{v,\frac{a}{2}}\right|$

$$*\ t = \frac{\bar{x} - \mu_0}{S_{\bar{x}}} = \frac{5{,}02 - 5}{0{,}069} = 0{,}29$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{150{,}6}{30} = 5{,}02$$

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{0{,}38}{\sqrt{30}} = 0{,}069$$

$$S = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}} = 0{,}38$$

$$*\ t_{v,\frac{a}{2}} = t_{29,\frac{0{,}05}{2}} = 2{,}045$$

↳ $v = n - 1 = 30 - 1 = 29$

$0{,}29 \ngtr 2{,}045$

Δεκτή η H_0

Όλα είναι εντάξει!

53. Ο διευθυντής πωλήσεων εταιρείας καλλυντικών ισχυρίζεται ότι το μερίδιο που κατέχουν τα προϊόντα της εταιρίας στην αγορά είναι μεγαλύτερο από 10%. Από δείγμα 200 γυναικών προέκυψε ότι 32 χρησιμοποιούν προϊόντα της εταιρίας. Με πιθανότητα λάθους 1%, είναι ο ισχυρισμός του διευθυντή σωστός;

μονάδες που έχουν ένα χαρακτηριστικό

$n = 200 \quad x = 32 \quad P = \frac{32}{200} = 0{,}16$ ← ποσοστό στο δείγμα

μέγεθος δείγματος

Υπόθεση: $H_0: \Pi = 0{,}1$, $H_1: \Pi > 0{,}1$ μονοκατάληκτο

Επειδή έχω ποιοτικό χαρακτηριστικό κάνω έλεγχο ποσοστού.

Βρίσκω το P. Απορρίπτω την H_0 όταν: $|Z| > |Z_{a/2}|$

$Z < Z_a$

$Z > Z_{1-a}$

$$Z = \frac{P - \Pi_0}{S_P} = \frac{0{,}16 - 0{,}1}{0{,}021} = 2{,}86$$

$$S_P = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0{,}1(1-0{,}1)}{200}} = 0{,}021$$

$$Z_{1-a} = Z_{1-0{,}01} = 2{,}33$$

$2{,}86 > 2{,}33$

Απορρίπτεται η $H_0: \Pi = 0{,}1$

Επομένως $H_1: \Pi > 0{,}1$

Ευσταθεί ο ισχυρισμός του διευθυντή!

54. Από δείγμα 200 γυναικών, οι 32 χρησιμοποιούν προϊόντα της εταιρίας. α) Να υπολογιστούν τα όρια του διαστήματος εμπιστοσύνης, μέσα στα οποία αναμένεται με $P=95\%$ το ποσοστό της εταιρίας στην αγορά καλλυντικών.

$$p-|Z_{a/2}|S_p < \Pi < p+|Z_{a/2}|S_p$$

$$p=\frac{32}{200}=0{,}16 \qquad S_p=\sqrt{\frac{p(1-p)}{n}}=0{,}026$$

$$|Z_{\frac{0{,}05}{2}}|=1{,}96$$

Επομένως $0{,}109 < \Pi < 0{,}211 \rightarrow 10{,}9\% < \Pi < 21{,}1\%$

β) Μπορεί να υποστηριχθεί με πιθανότητα διαπράξεως σφάλματος τύπου Ι, 5% ότι το ποσοστό της εταιρίας στην αγορά καλλυντικών είναι 8%;

$H_0: \Pi=0{,}08$
$H_1: \Pi\neq 0{,}08$
Όμως δε χρειάζεται αφού έχω το διάστημα εμπιστοσύνης και βλέπω ότι σίγουρα δεν κατέχει στην αγορά το 8% αλλά το πάνω από 10,9%.

Το διάστημα εμπιστοσύνης για έλεγχο μιας υπόθεσης χρησιμοποιείται καλύτερα σε δικατάληκτο κριτήριο ελέγχου.

55. Ο προϊστάμενος ελέγχου ποιότητας μιας βιομηχανίας ενδιαφέρεται να διερευνήσει μία νέα μέθοδο συσκευασίας. Συγκεκριμένα ενδιαφέρεται να εξετάσει αν το ποσοστό ελαττωματικών συσκευασιών, με τη νέα μέθοδο μειώνεται κάτω από 10% που ήταν το ποσοστό ελαττωματικών συσκευασιών στο παρελθόν. Για να ελέγξει την υπόθεση επιλέγει τυχαίο δείγμα από 200 πακέτα προϊόντος που έχουν συσκευαστεί με τη νέα μέθοδο και παρατηρεί ότι 11 από αυτά είναι ελαττωματικά.
Τι συμπέρασμα βγαίνει βάσει αυτού του δείγματος;

Υπόθεση: $H_0: \Pi=0{,}10$, $H_1: \Pi\neq 0{,}10$ δικατάληκτο. Απορρίπτω H_0 αν: $Z<Z_a$

$$Z=\frac{p-\Pi_0}{S_p}=\frac{\frac{11}{200}-0{,}10}{\sqrt{\frac{0{,}10(1-0{,}10)}{200}}}=-2{,}12$$

Για $a=5\%$ $\longrightarrow$ $-2{,}12<-1{,}645$
$Z_{0{,}05}=-1{,}645$ Απορρίπτεται η H_0, H_1 ΔΕΚΤΗ

Για $a=1\%$ $\longrightarrow$ $-2{,}12<-2{,}32$ Δεκτή η H_0
$Z_{0{,}01}=-2{,}32$

Όταν η τιμή που ελέγχω είναι κοντά στην κριτική τιμή τότε καλό είναι να ελέγχουμε ξανά με μικρότερο ποσοστό σφάλματος την υπόθεση για την ορθότητά της.

56. Ένας χρηματιστηριακός αναλυτής μελετά την απόδοση των μετοχών του χρηματιστηρίου Αθηνών κατά το περασμένο έτος. Το πρόβλημα που τον απασχολεί είναι εάν ο κλάδος του εμπορίου απέδωσε όσο και ο κλάδος της βιομηχανίας (μεταποίησης). Ο αναλυτής συλλέγει ένα τυχαίο δείγμα 18 μετοχών εμπορικών εταιριών και ένα τυχαίο δείγμα 20 μετοχών βιομηχανικών εταιριών και καταγράφει τις ποσοστιαίες αποδόσεις (%) των μετοχών τους.

	ΜΕΤΟΧΕΣ ΕΜΠΟΡΙΚΩΝ ΕΤΑΙΡΙΩΝ	ΜΕΤΟΧΕΣ ΒΙΟΜΗΧΑΝΙΚΩΝ ΕΤΑΙΡΙΩΝ
Μέγεθος δείγματος	$n_1 = 18$	$n_2 = 20$
Μέση απόδοση (%)	$\bar{X}_1 = 3,5$	$\bar{X}_2 = 2,7$
Τυπική απόκλιση	$S_1 = 1,3$	$S_2 = 1,5$

Υπόθεση: $H_0: \mu_1 - \mu_2 = 0 \quad (\mu_1 = \mu_2)$ } Δικατάληκτο

$H_1: \mu_1 - \mu_2 \neq 0$ } Δε γνωρίζουμε s^2

$n_1, n_2 < 30 \rightarrow$ t-student

$$t_v = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}} \Leftrightarrow t_v = \frac{3,5 - 2,7}{0,4578} = 1,747$$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{1,9856\left(\frac{1}{18} + \frac{1}{20}\right)} = 0,4578$$

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 1,9856$$

Βαθμός ελευθερίας: $V = n_1 + n_2 - 2 = 18 + 20 - 2 = 36$

$$t_{v,\frac{a}{2}} = t_{36,\frac{0,05}{2}} = 2,025$$

$$t_v < t_{v,\frac{a}{2}} \Leftrightarrow 1,747 < 2,025$$

Δεκτή η H_0: Δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των κλάδων.

Συνοπτική θεωρία

- Διακυμάνσεις (σ_1^2, σ_2^2): γνωστές

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}} \quad \text{με} \quad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ΑΓΝΩΣΤΗ ΔΙΑΚΥΜΑΝΣΗ: $v = n_1 + n_2 - 2$

$H_0: \mu_1 - \mu_2 = 0$, $H_1: \mu_1 - \mu_2 \neq 0$ — Απορρίπτω H_0 αν $|Z| > |Z_{\frac{a}{2}}|$ — $|t_v| > |t_{v,\frac{a}{2}}|$

$H_0: \mu_1 - \mu_2 = 0$, $H_1: \mu_1 - \mu_2 < 0$ — Απορρίπτω H_0 αν $Z < Z_a$ — $t_v < -t_{v,a}$

$H_0: \mu_1 - \mu_2 = 0$, $H_1: \mu_1 - \mu_2 > 0$ — Απορρίπτω H_0 αν $Z > Z_{1-a}$ — $t_v > t_{v,a}$

Δ.Ε. $\bar{X}_1 - \bar{X}_2 - |Z_{a/2}|\sigma_{\bar{x}_1 \bar{x}_2} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + |Z_{a/2}|\sigma_{\bar{x}_1 \bar{x}_2}$

- Διακυμάνσεις (σ_1^2, σ_2^2): άγνωστες

$\rightarrow n_1, n_2 \geq 30$: $Z = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}}$ με $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$\rightarrow n_1, n_2 < 30$: $t_v = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}}$ με $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

57. Οι υπεύθυνοι μιας γαλακτοβιομηχανίας πιστεύουν ότι τα παιδιά των αστικών κέντρων δαπανούν μεγαλύτερα χρηματικά ποσά για παγωτά σε σχέση με τα παιδιά των επαρχιακών πόλεων. Δείγμα 100 παιδιών από Αθήνα και Θεσσαλονίκη έδωσε μέση μηνιαία δαπάνη για παγωτά 36,5€ με τυπική απόκλιση 2,4€. Αντίστοιχο δείγμα 80 παιδιών από το Λουτράκι, το Άργος και το Ναύπλιο έδωσε μέση μηνιαία δαπάνη 35,8€ με τυπική απόκλιση 3,5 €.

α) Να διατυπώσετε την υπόθεση μηδέν (H_0) και την εναλλακτική (H_1).

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 > 0$

β) Είναι ο παραπάνω συλλογισμός σωστός σε επίπεδο σημαντικότητας 5%;

Διακυμάνσεις γνωστές $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}}$ με $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Συνάρτηση ελέγχου Z:

$$Z = \frac{36,5 - 35,8}{\sqrt{\frac{2,4^2}{100} + \frac{3,5^2}{80}}} = 1,52$$

$$|Z_{1-a}| = |Z_{1-0,05}| = 1,645$$

Απορρίπτω H_0 αν: $Z > Z_{1-a}$

$1,52 \not> 1,96$

Επομένως δεκτή η H_0

Δηλαδή αυτό που πιστεύουν δεν είναι σωστό!

58. Το τμήμα έκδοσης πιστωτικών καρτών μιας τράπεζας, θέλει να διερευνήσει εάν η οικογενειακή κατάσταση των κατόχων πιστωτικής κάρτας επιδρά στο ποσοστό των κατόχων που έχουν ανεξόφλητες ληξιπρόθεσμες υποχρεώσεις. Τα δύο τυχαία δείγματα 150 παντρεμένων και 200 ανύπαντρων κατόχων πιστωτικής κάρτας έδωσαν τα εξής:

	Παντρεμένοι	Ανύπαντροι
Μέγεθος δείγματος	$n_1 = 150$	$n_2 = 200$
Κάτοχοι με υποχρεώσεις	$X_1 = 4$	$X_2 = 9$
Εκτίμηση ποσοστού	$P_1 = \frac{4}{150} = 0,027$	$P_2 = \frac{9}{200} = 0,045$

Δύο δείγματα, άρα έλεγχος δύο παραμέτρων.

Έλεγχος δύο ποσοστών αφού είναι ποιοτικό (εξοφλημένες ή ανεξόφλητες)

Υπόθεση: $H_0: \Pi_1 - \Pi_2 = 0$, $H_1: \Pi_1 - \Pi_2 \neq 0$ — Επιλέγουμε αν δε δίνεται: $a = 5\%$

$$Z = \frac{P_1 - P_2}{S_{P_1 - P_2}} = \frac{0,027 - 0,045}{0,02} = -0,90$$

$$S_{P_1-P_2} = \sqrt{0,037\left(\frac{1}{150} + \frac{1}{200}\right)} = 0,02$$

$$P_c = \frac{X_1 + X_2}{n_1 + n_2} = \frac{4+9}{150+200} = 0,037$$

$$Z_{\frac{a}{2}} = Z_{\frac{0,05}{2}} = -1,96$$

Απορρίπτω H_0 αν: $|Z| > |Z_{a/2}|$

$|-0,90| \not> |-1,96|$

Δεκτή η H_0!

Επομένως δεν επιδρά η οικογενειακή κατάσταση!

59. Από τα αρχεία ενός νοσοκομείου προκύπτει ότι από 500 άνδρες που εισήχθηκαν για θεραπεία, οι 70 είχαν υψηλή πίεση αίματος, ενώ από 500 γυναίκες οι 40 είχαν το ίδιο πρόβλημα. Να υπολογιστεί με πιθανότητα 95%, το διάστημα εμπιστοσύνης μέσα στο οποίο βρίσκεται η διαφορά μεταξύ των ποσοστών ανδρών και γυναικών που έχουν υψηλή πίεση αίματος. Υπάρχει σημαντική διαφορά μεταξύ των δύο ποσοστών, σε $a=5\%$;

Ποιοτικό χαρακτηριστικό → Άρα διαφορά ποσοστών

ΔΕ: $P_1-P_2-|Z_{a/2}|S_{P_1-P_2} < \Pi_1-\Pi_2 < P_1-P_2+|Z_{a/2}|S_{P_1-P_2}$

$P_1=\frac{70}{500}=0,14$ $\quad P_2=\frac{40}{500}=0,08$

$Z_{a/2}=Z_{\frac{0,05}{2}}=1,96$

$P_c=\frac{x_1+x_2}{n_1+n_2}=\frac{70+40}{500+500}=0,11 \rightarrow S_{P_1-P_2}=\sqrt{0,11\left(\frac{1}{500}+\frac{1}{500}\right)}=0,0198$

$(0,14-0,08)-1,96\cdot 0,0198<\Pi_1-\Pi_2<(0,14-0,08)+1,96\cdot 0,0198$

ΕΡΜΗΝΕΙΑ $\quad 0,0212<\Pi_1-\Pi_2<0,0988$

- Το μηδέν δεν εμπεριέχεται ανάμεσα στο όριο, άρα υπάρχει σημαντική διαφορά μεταξύ ποσοστό γυναικών και ανδρών που έχουν υψηλή πίεση.
- Επειδή και τα δύο ποσοστά είναι άνω του μηδενός, τότε $\Pi_1>\Pi_2$

Άρα οι άντρες έχουν μεγαλύτερο ποσοστό υψηλής πίεσης.

ΣΥΝΟΠΤΙΚΗ ΘΕΩΡΙΑ

ΕΛΕΓΧΟΣ ΔΙΑΦΟΡΑΣ ΔΥΟ ΠΟΣΟΣΤΩΝ

Χρησιμοποιώ μόνο κανονική κατανομή

$Z=\frac{P_1-P_2}{S_{P_1-P_2}}$ με $S_{P_1-P_2}=\sqrt{P_c\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ όπου $P_c=\frac{x_1+x_2}{n_1+n_2}$

$P_1=\frac{x_1}{n_1}$ και $P_2=\frac{x_2}{n_2}$

Απορρίπτω H_0 όταν
- $|Z|>|Z_{a/2}|$ (δικατάληκτο) $\Pi_1-\Pi_2\neq 0$
- $Z<Z_a$ (μονοκατάληκτο) $\Pi_1-\Pi_2<0$
- $Z>Z_{1-a}$ (μονοκατάληκτο) $\Pi_1-\Pi_2>0$

Διάστημα Εμπιστοσύνης:

$$P_1-P_2-|Z_{a/2}|S_{P_1-P_2}<\Pi_1-\Pi_2<P_1-P_2+|Z_{a/2}|S_{P_1-P_2}$$

- Αν το μηδέν περιλαμβάνεται στο διάστημα εμπιστοσύνης τότε δεχόμαστε την H_0, αν όχι δεχόμαστε την H_1 σε δικατάληκτο.
- Σε μονοκατάληκτο $(\Pi_1-\Pi_2<0)$ δεχόμαστε την H_1 όταν και οι δύο τιμές των ορίων είναι αρνητικές.
- Σε μονοκατάληκτο $(\Pi_1-\Pi_2>0)$ δεχόμαστε την H_1 όταν και οι δύο τιμές των ορίων είναι θετικές.

60. Το τμήμα χορηγήσεων μιας τράπεζας διερευνά αν η πιστοληπτική ικανότητα των επιχειρήσεων σχετίζεται με τον οικονομικό κλάδο που ανήκουν. Στον πίνακα δίνεται η κατανομή ανά κλάδο δραστηριότητας και πιστοληπτικής ικανότητας τυχαίου δείγματος 300 επιχειρήσεων που υπέβαλαν αίτηση για δανειοδότηση. Να ελεγχθεί σε 1% αν η πιστοληπτική ικανότητα των επιχειρήσεων σχετίζεται με τον κλάδο που αυτές ανήκουν:

ΠΙΣΤΟΛΗΠΤΙΚΗ ΙΚΑΝΟΤΗΤΑ / ΚΛΑΔΟΣ	ΒΙΟΜΗΧΑΝΙΑ	ΕΜΠΟΡΙΟ	ΥΠΗΡΕΣΙΕΣ	ΣΥΝΟΛΟ
ΥΨΗΛΗ	20	25	45	90
ΜΕΣΑΙΑ	25	40	30	95
ΧΑΜΗΛΗ	30	65	20	115
ΣΥΝΟΛΟ	75	130	95	300

Υποθέσεις:

H_0: Η πιστοληπτική ικανότητα είναι ανεξάρτητη του κλάδου

H_1: Η πιστοληπτική ικανότητα σχετίζεται με τον κλάδο.

Φτιάχνουμε πίνακα Fe (αναμενόμενες συχνότητες) $f_{e_{ij}} = \frac{\text{ΑΘΡΟΙΣΜΑ ΓΡΑΜΜΗΣ} \times \text{ΑΘΡΟΙΣΜΑ ΣΤΗΛΗΣ}}{\text{ΓΕΝΙΚΟ ΑΘΡΟΙΣΜΑ}}$

23	39	29	90
24	41	30	95
29	50	36	115
75	130	95	300

Το άθροισμα θα πρέπει να είναι ίδιο με τα f_o και σε γραμμές και σε στήλες.

$f_{e_{11}} = \frac{90\times75}{300} = 22,5 \simeq 23$

$f_{e_{12}} = \frac{90\times130}{300} = 39$

$f_{e_{13}} = \frac{90\times95}{300} = 29$

$f_{e_{21}} = \frac{95\times75}{300} = 24$

$f_{e_{22}} = \frac{95\times130}{300} = 41,16 \simeq 41$

$f_{e_{23}} = \frac{95\times95}{300} = 30$

$f_{e_{31}} = \frac{115\times75}{300} = 29$

$f_{e_{32}} = \frac{115\times130}{300} = 50$

$f_{e_{33}} = \frac{115\times95}{300} = 36$

ΣΤΡΟΓΓΥΛΟΠΟΙΟΥΜΕ !

Πρέπει να υπολογίσω το $X_v^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

Κάνω πίνακα:

f_o	f_e	$(f_o-f_e)^2/f_e$
20	23	$\frac{(20-23)^2}{23} = 0,3913$
25	39	0,0416
45	29	0,0345
25	24	⋮
40	41	⋮
30	30	⋮
30	24	⋮
65	50	
20	36	
		27,007

Υπολογίζω το v:

$v = (\mu-1)(c-1) = (\text{γραμμές}-1)(\text{στήλες}-1)$

$v = (3-1)(3-1) = 4$

Επομένως $X_4^2 = 27,007$

Χρειάζομαι και το $X_{v,a}^2$

$X^2_{4\ 0,01} = 13,277 \rightarrow$ Από πίνακες X^2

Απορρίπτω H_o αν: $X_v^2 > X_{va}^2$

$27,007 > 13,277$ Ισχύει

Απορρίπτω τη H_o: Οι ιδιότητες σχετίζονται.

ΣΥΝΟΠΤΙΚΗ ΘΕΩΡΙΑ

ΕΛΕΓΧΟΣ ΑΝΕΞΑΡΤΗΣΙΑΣ ΙΔΙΟΤΗΤΩΝ - ΚΡΙΤΗΡΙΟ X^2

H_o: Οι ιδιότητες δεν σχετίζονται (ανεξάρτητα τα δύο χαρακτηριστικά)

H_1: Οι ιδιότητες σχετίζονται (μη ανεξάρτητα)

Χρειαζόμαστε τα: $X_v^2 = \sum\frac{(f_o-f_e)^2}{f_e}$ και $X_{v,a}^2$ (πίνακες)

με v = βαθμός ελευθερίας $= (\mu-1)(c-1)$

Απορρίπτω H_o αν $X_v^2 > X_{v,a}^2$

ΣΗΜΕΙΩΣΗ ΓΙΑ ΠΙΝΑΚΑ fe

$f_{e_{ij}} = \frac{\text{ΑΘΡΟΙΣΜΑ ΓΡΑΜΜΗΣ} \times \text{ΑΘΡΟΙΣΜΑ ΣΤΗΛΗΣ}}{\text{ΓΕΝΙΚΟ ΑΘΡΟΙΣΜΑ}}$

61. Δίνεται X με $\rho_{yx}=0{,}95$ και Z με $\rho_{yz}=-0{,}95$

Ποιο υπόδειγμα είναι καταλληλότερο για να κάνω προβλέψεις;

Και τα δύο υποδείγματα είναι το ίδιο αξιόπιστα για προβλέψεις με διαφορετικές συσχετίσεις.

62. Έστω Y η εβδομαδιαία κατανάλωση μαρουλιού και X η τιμή πωλήσεως (€ ανά κεφαλάκι). Παίρνουμε ένα δείγμα με τις εξής τιμές:

Y	X	Y·X	X^2
20	0,95		
18	1		
12	1,10		
19	0,90		
10	1,10		
14	1,05		
8	1,20		
13	1		
$\Sigma y=114$	$\Sigma x=8{,}3$	$\Sigma xy=115{,}10$	$\Sigma x^2=8{,}75$

α) Ποια η εξίσωση παλινδρόμησης;

Είναι της μορφής: $Y_i = b_0 + b_1 X_i$

με $b_1 = \dfrac{n\Sigma yx-(\Sigma Y)(\Sigma X)}{n\Sigma x^2-(\Sigma x)^2}$

$b_0 = \dfrac{\Sigma Y}{n} - b_1\dfrac{\Sigma X}{n}$

Για $n=8$ (μέγεθος δείγματος):

$b_1=-41{,}96$ και $b_0=57{,}78$

$\hat{Y}_i = 57{,}78 - 41{,}96\,X_i$

β) Να ερμηνευτούν οι συντελεστές b_0 και b_1.

b_0: αν μηδενιστεί η τιμή πωλήσεως θα ζητούνται 57,78 κεφαλάκια

b_1: αν η τιμή πωλήσεως αυξηθεί (μειωθεί) κατά 1€ τότε η μέση εβδομαδιαία ζήτηση θα μειωθεί (αυξηθεί) κατά 41,96 κεφαλάκια.

γ) Να εκτιμηθεί η μέση εβδομαδιαία ζήτηση όταν η τιμή πώλησης είναι 1,15 ευρώ ανά κεφαλάκι. $\to X_i$

$\hat{Y} = 57{,}78 - 41{,}96(1{,}15)$

$Y = 9{,}526$ κεφαλάκια

ΣΥΝΟΠΤΙΚΗ ΘΕΩΡΙΑ

- ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

$r = \dfrac{\Sigma(x-\bar{x})(Y-\bar{Y})}{\sqrt{\Sigma[(x-\bar{x})^2][\Sigma(Y-\bar{Y})^2]}}$

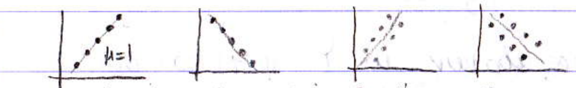
r: για δείγμα

ρ: για πληθυσμό

Μετράει ένταση συσχέτισης

$r \in [-1,1]$

1. Μπορεί να έχει σφάλμα
2. Μπορεί $\rho=0$ αλλά στο δείγμα $r=0{,}10$



- ΣΥΝΑΡΤΗΣΗ ΕΥΘΕΙΑΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

$y = b_0 + b_1 x$ (είναι μαθηματικό υπόδειγμα όταν έχω πλήρης συσχέτιση)

$b_1 = \dfrac{n\Sigma Yx-(\Sigma x)(\Sigma Y)}{n\Sigma x^2-(\Sigma x)^2}$ και $b_0 = \dfrac{\Sigma Y}{n} - b_1\dfrac{\Sigma x}{n} = \bar{Y} - b_1\bar{x}$

δ) Να υπολογιστεί η μέση ελαστικότητα και να ερμηνευθεί.

$\bar{\eta}_{yx} = b_1\dfrac{\bar{x}}{\bar{y}} = b_1\dfrac{\Sigma x/n}{\Sigma y/n} \leftrightarrow \bar{\eta}_{xy} = -41{,}96\dfrac{8{,}3}{114} = -3{,}055$

Ερμηνεία: Όταν η τιμή πωλήσεως του μαρουλιού αυξηθεί (μειωθεί) κατά 1% τότε η μέση τιμή πωλήσεων θα μειωθεί (αυξηθεί) κατά 3,055% (ΠΡΟΣΟΧΗ, δε μεταφέρεται η υποδιαστολή)

ε) Κατά πόσο θα πρέπει να μεταβληθεί η μέση τιμή πωλήσεων για να αυξηθεί η τιμή κατά 4%; $3{,}055 \times 4 = 12{,}22$

στ) Να υπολογιστεί ο συντελεστής προσδιορισμού και να ερμηνευτεί η τιμή του. $R^2 = SSR/SST$

• $SST = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 1758 - \frac{114^2}{8} = 133,5$

• $SSR = \Sigma Y^2 - b_0 \Sigma Y - b_1 \Sigma YX$

$= 1758 - 57,78 \cdot 114 + 41,96 \cdot 115,6 = 21,255$

Επομένως: $R^2 = \frac{SSR}{SST} = \frac{21,255}{133,5} = 0,84 \rightarrow 84\%$

Ερμηνεία: Το 84% των μεταβολών της Y οφείλεται στην X και το υπόλοιπο 16% σε άλλους παράγοντες.

ζ) Αν σας είναι γνωστό ότι $R^2 = 0,84$ και $b_1 = -41,96$ υπολογίστε τον συντελεστή συσχέτισης (r).

Αν R^2 γνωστός, τότε $r = \sqrt{R^2}$ και δίνουμε το πρόσημο του b_1.

$r = \sqrt{0,84} = -0,9168$

ΣΥΝΟΠΤΙΚΗ ΘΕΩΡΙΑ

Συντελεστής Προσδιορισμού

$R^2 = \frac{SSR}{SST}$ → Εκφράζει το ποσοστό των μεταβολών την Y που ερμηνεύονται από τη X. Αν R=0 τότε το X ερμηνεύει πλήρως την Y.

$SST = SSR + SSE$

Εκφώνηση: Αξιολόγηση προγνωστικής ικανότητας δείγματος → Βρίσκουμε R^2

η) Να γίνει έλεγχος του συντελεστή προσδιορισμού.

ΣΥΝΟΠΤΙΚΗ ΘΕΩΡΙΑ

H_0: η εξίσωση παλινδρόμησης δεν εξηγεί καθόλου τις μεταβολές της Y

H_1: η εξίσωση παλινδρόμησης εξηγεί μέρος των μεταβολών της Y.

Απορρίπταμε H_0 αν: $F_{1,n-2} > F_{(1,n-2),a}$

Υπόθεση: $H_0: \rho^2 = 0$ $\qquad n-2 = 8-2 = 6$

$H_1: \rho^2 \neq 0$

$F_{1,6} = \frac{SSR}{SSE/(n-2)} = \frac{21,255}{112,245/6} = 3$

$SSE = SST - SSR = 133,5 - 21,255 = 112,245$

$F_{1,6,0,05} = 5,99$ (από πίνακες)

$31,685 > 5,99$ → Απορρίπτεται η H_0.

63. Έρευνα αγοράς που έγινε σε 8 πόλεις και αφορούσε τη μέση ανά κάτοικο ετήσια κατανάλωση προϊόντος σε τεμάχια (Y) σε σχέση με τη μέση τιμή πώλησης σε ευρώ (X) έδωσε αποτελέσματα:

$\Sigma Y_i = 648 \quad \Sigma x_i y_i = 21325 \quad \Sigma x_i^2 = 8934 \quad S_e = 1{,}927$

$\Sigma X_i = 266 \quad \Sigma (x_i - \bar{x})^2 = 89{,}5 \quad \Sigma y_i^2 = 53056$

α) Να εκτιμηθεί η γραμμική εξίσωση παλινδρόμησης της ανά κάτοικο ετήσιας ζήτησης προϊόντος ως προς την τιμή και να ελεγχθεί η στατιστική σημαντικότητα του υποδείγματος σε $\alpha = 5\%$

$$\hat{Y} = b_0 + b_1 X_i$$

$$b_1 = \frac{n\Sigma YX - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} = -2{,}469$$

$$b_0 = \frac{\Sigma Y}{n} - b_1 \frac{\Sigma X}{n} = 163{,}103$$

Επομένως $\hat{Y} = 163{,}103 - 2{,}469 X_i$

Έλεγχος Στατιστικής Σημαντικότητας:

Υπόθεση: $H_0: b_1 = 0$, $H_1: b_1 \neq 0$

$$S_{b_1} = \frac{S_e}{\sqrt{\Sigma (x_i - \bar{x})^2}} = \frac{1{,}927}{\sqrt{89{,}5}} = 0{,}204$$

$$t_v = \frac{b_1}{S_{b_1}} = \frac{-2{,}469}{0{,}204} = -12{,}119$$

Απορρίπτω H_0 αν $|t_v| > |t_{v,\frac{\alpha}{2}}|$

$$t_{v,\frac{\alpha}{2}} = t_{6,\frac{0{,}05}{2}} = 2{,}447$$

$$\underbrace{|-12{,}119|}_{t_v} > \underbrace{|2{,}447|}_{t_{v,\frac{\alpha}{2}}}$$

Απορρίπτω H_0, άρα υπόδειγμα στατιστικά σημαντικό!

β) Να εκτιμήσετε τη συνολική ζήτηση σε τεμάχια και αξία αυτού του προϊόντος, σε μια πόλη με πληθυσμό 100.000 κατοίκους εάν η τιμή του προϊόντος διαμορφωθεί στο επίπεδο των 33€/τεμάχιο.

$$\hat{Y} = 163{,}103 - 2{,}469 X_i$$

$X = 33 \rightarrow Y = 81{,}626$ τμχ / κάτοικο (ΜΟ ζήτησης ανά κάτοικο)

o Συνολική ζήτηση: $100.000 \times 81{,}626 = 8.162.600$ τεμάχια

o Συνολική αξία: $8.162.600 \times 33 = 269.365.800$ €

64. Να συμπληρωθούν τα κενά του πίνακα Anova.

ΠΗΓΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ	ΑΘΡΟΙΣΜΑΤΑ ΤΕΤΡΑΓΩΝΩΝ	ΒΑΘΜΟΙ ΕΛΕΥΘΕΡΙΑΣ	ΜΕΣΑ ΤΕΤΡΑΓΩΝΑ	ΛΟΓΟΣ $F_{1,n-2}$
Παλινδρόμηση	SSR	1	SSR/1	$\frac{SSR/1}{SSE/(n-2)}$
Κατάλοιπος	SSE	n−2	SSE/(n−2)	
Σύνολο	SST	n−1		

α) Υπολογίστε τον συντελεστή προσδιορισμού R^2 $\quad R^2 = \frac{SSR}{SST}$

β) Να ελέγξετε τον συντελεστή προσδιορισμού R^2.

$H_0: P^2 = 0$, $H_1: P^2 \neq 0$

Απορρίπτουμε H_0 αν: $F_{1,n-2} > F_{(1,n-2),\alpha}$ με $F_{1,n-2} = \frac{SSR}{SSE/(n-2)}$

γ) Να υπολογίσετε τον συντελεστή συσχετίσεως.

$$r = \sqrt{R^2} \text{ (με πρόσημο } b_1\text{)} = \frac{\Sigma (x-\bar{x})(Y-\bar{Y})}{\sqrt{\Sigma [(x-\bar{x})^2][\Sigma (Y-\bar{Y})^2]}}$$

ΣΗΜΕΙΩΣΗ $r^2 = R^2$

{ ΕΡΜΗΝΕΙΕΣ - ΟΡΙΣΜΟΙ }

1. **Δειγματοληπτικό σφάλμα:** είναι η διαφορά μεταξύ του αποτελέσματος της απογραφής και του αποτελέσματος της δειγματοληψίας

2. **Συχνότητα f_i της τιμής x_i:** λέγεται ο φυσικός αριθμός που μας δείχνει πόσες φορές εμφανίζεται η τιμή x_i μέσα στο δείγμα.

3. **Κεντρική τάση:** μετράει την τάση συγκέντρωσης των τιμών γύρω από μια κεντρική τιμή (μέσος αριθμητικός, σταθμικός μέσος, διάμεσος, επικρατούσα τιμή)

4. **Διασπορά:** είναι το μέγεθος της ανομοιογένειας μεταξύ των τιμών, δηλαδή πόσο διαφέρουν μεταξύ τους ή πόσο "διεσπαρμένες" είναι.
(εύρος, τεταρτημοριακή απόκλιση, διακύμανση, τυπική απόκλιση, συντελεστής μεταβλητότητας)

5. **Τεταρτημοριακή απόκλιση:** δείχνει το εύρος των τιμών που συγκεντρώνεται στο μεσαίο (κεντρικό) 50% των παρατηρήσεων

6. **Διακύμανση / Τυπική απόκλιση:** όσο μικρότερη είναι η τιμή της διακύμανσης και ως εκ τούτου της τυπικής απόκλισης, τόσο μικρότερη η διασπορά των τιμών. Μετρούν τη μέση σπορά των τιμών γύρω από τον μέσο αριθμητικό.

7. Συντελεστής μεταβλητότητας εκφράζει τη διασπορά γύρω από τον μέσο ως ποσοστό του μέσου αριθμητικού.

8. Σχήμα: μια από τις τρεις ιδιότητες των αριθμητικών δεδομένων είναι το σχήμα που ακολουθεί η κατανομή τους (αρνητική, μηδενική, θετική ασυμμετρία)

9. Μέσος αριθμητικός: εάν το σύνολο (άθροισμα) των παρατηρήσεων μοιραζόταν σε n ποσά, το κάθε ποσό θα ισούται με $\bar{X}$

10. Διάμεσος είναι η μεσαία τιμή μιας ομάδας τιμών ιεραρχημένων σε αύξουσα τάξη μεγέθους

11. Επικρατούσα τιμή: ή σημείο μέγιστης συχνότητας είναι η τιμή με τη μεγαλύτερη συχνότητα σε ένα σύνολο μετρήσεων.

12. Συντελεστής ασυμμετρίας είναι θετικός όταν η κατανομή έχει θετική ασυμμετρία, μηδέν αν είναι μηδενική η ασυμμετρία και αρνητικός εάν έχει αρνητική ασυμμετρία $S_K \in [-1, 1]$

13. Πιθανότητα εκφράζει το πόσο "μέσα" πέφτουμε στις προβλέψεις μας

14. Ανεξάρτητα ενδεχόμενα: όταν η εμφάνιση του ενός δεν επηρεάζει την εμφάνιση του άλλου. Το αντίθετο για τα εξαρτημένα.

15. Ασυμβίβαστα ενδεχόμενα: όταν η εμφάνιση του ενός αποκλείει την εμφάνιση του άλλου $A \cap B = \emptyset$

16. Στατιστική ανεξαρτησία: όταν δύο ενδεχόμενα είναι ανεξάρτητα. Ισχύει $P(A/B) = P(A)$

17. Υπό-συνθήκη πιθανότητα είναι η πιθανότητα εμφάνισης του A δεδομένου ότι γνωρίζουμε την εμφάνιση του B $(P(A/B))$

18. Διωνυμική κατανομή: κατανομή κατά την οποία οι παρατηρήσεις χωρίζονται σε δύο αμοιβαίως αποκλειόμενες κατηγορίες (επιτυχία, αποτυχία), η πιθανότητα μια παρατήρηση να ανήκει στην κατηγορία "επιτυχία" ισούται με p και είναι πάντα σταθερή όταν οι παρατηρήσεις προέρχονται είτε από άπειρους πληθυσμούς χωρίς επανατοποθέτηση είτε από δείγμα $(n \leq 20)$ με επανατοποθέτηση.

19. Κατανομή Poison: Ίδια με τη διωνυμική όμως η πιθανότητα επιτυχίας είναι πολύ μικρή και ο αριθμός των παρατηρήσεων μεγάλος $(n \geq 20)$ προς άπειρος → μικρότερη από 0,01

20. Διακριτές μεταβλητές ή ασυνεχείς είναι αυτές που παίρναν μεμονωμένες τιμές (π.χ. ύψος, βάρος κτλ). Συνεχείς είναι αυτές που παίρνουν οποιαδήποτε πραγματική τιμή.

21. Τυπικό σφάλμα μέσου: η τυπική απόκλιση της κατανομής δειγματοληψίας του μέσου $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ ← Τυπική απόκλιση πληθυσμού

22. Επίπεδο εμπιστοσύνης: ισούται με 1−α ενώ με α συμβολίζουμε την πιθανότητα σφάλματος

23. Τυπικό σφάλμα ποσοστού η τυπική απόκλιση της κατανομής δειγματοληψίας του ποσοστού $S_p = \sqrt{\frac{P(1-P)}{n}}$ με $p = x/n$

24. Δειγματοληπτικό σφάλμα $e = \bar{x} - \mu = Z_{a/2} \frac{\sigma}{\sqrt{n}}$

25. Σφάλμα τύπου Ι: αναφέρεται στην περίπτωση που η H_0 απορρίπτεται ενώ στην πραγματικότητα είναι αληθής

26. Σφάλμα τύπου ΙΙ: αναφέρεται στην περίπτωση που η H_0 δεν απορρίπτεται (γίνεται δεκτή) ενώ στην πραγματικότητα είναι λανθασμένη.

27. Θετική συσχέτιση όσο αυξάνει η Χ, αυξάνει και η Ψ.

28. Συντελεστής συσχέτισης μετράει την ένταση της γραμμικής σχέσης μεταξύ δύο μεταβλητών $r \in [-1,1]$
−1= τέλεια αρνητική, 1= τέλεια θετική, 0= μη συσχέτιση

29. Νόθος συσχέτιση: δύο μεταβλητές αντανακλούν την επίδραση ενός τρίτου παράγοντα

30. Απλή γραμμική παλινδρόμηση περιγράφει τη σχέση μεταξύ των Χ και Υ με ένα υπόδειγμα της μορφής $Y_i = b_0 + b_1 X_i + e_i$

31. Συντελεστής προσδιορισμού: μετράει το ποσοστό των μεταβολών της Υ που οφείλονται στη Χ $R^2 \in [-1,1]$

32. Συντελεστής b_0: αν μηδενιστεί η Χ θα φτάναι Υ

33. Συντελεστής b_1: για κάθε αύξηση της Χ κατά 1 το Υ αυξάνεται ή μειώνεται κατά b_1

34. Μέση ελαστικότητα Όταν η X αυξηθεί κατά 1% τότε η μέση τιμή της Y θα μειωθεί κατά $\bar{n}_{yx}$

35. Προγνωστική ικανότητα: βρίσκουμε συντελεστή προσδιορισμού
$$R^2 = SSR/SST$$

Σημειώσεις

$Z_{a/2} = 1,96$ $\quad$ $Z_{0,01} = -2,32$

$Z_{0,05} = -1,645$ $\quad$ $Z_{0,99} = 2,33$

$Z_{0,95} = 1,645$

Κανονική Κατανομή	Κατανομή Poisson
$\mu = np$ $S^2 = np(1-p)$	$\mu = E(x) = np = S^2$

$$S_p = \sqrt{\frac{P(1-P)}{n}}$$ (για Δ.Ε ποσοστού)

$$Z = \frac{P - \Pi_0}{S_p}$$ (για έλεγχο υπόθεσης ποσοστού)

$$Z = \frac{P_1 - P_2}{S_{P_1 - P_2}}$$ (για έλεγχο διαφοράς ποσοστού)