ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΧΤΑ
ΑΚΑΔΗΜΑΙΚΑ
ΜΑΘΗΜΑΤΑ

Στατιστική

10ο Μάθημα: Προσομοίωση Εξέτασης στο μάθημα της Στατιστικής

(Λυμένα και Άλυτα Θέματα)

Γεώργιος Μενεξές
Τμήμα Γεωπονίας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

10ο Μάθημα

Προσομοίωση Εξέτασης στο μάθημα της Στατιστικής

(Λυμένα και Άλυτα Θέματα)

Σκηνή Πρώτη

Ερωτήσεις Σωστού-Λάθους (μέρος Ι)

1. Ο μέσος όρος είναι δείκτης αξιοπιστίας των τιμών μιας κατανομής;
2. Το τυπικό σφάλμα εκφράζει την ακρίβεια στην εκτίμηση παραμέτρων ενός στατιστικού πληθυσμού;
3. Η τυπική απόκλιση είναι ίση με το τετράγωνο της διακύμανσης;
4. Ο συντελεστής παραλλακτικότητας *CV* εκφράζει τη διακύμανση μιας κατανομής ως ποσοστό του μέσου όρου της κατανομής;
5. Στο διάστημα τιμών $\bar{X} \pm 2s$ αναμένουμε να βρίσκεται το 99,9% των τιμών μιας Κανονικής Κατανομής;
6. Η κρίσιμη τιμή της *F* Κατανομής με 10 και 20 βαθμούς ελευθερίας σε επίπεδο σημαντικότητας *α*=0,05 είναι ίση με 2,35;

Ερωτήσεις Σωστού-Λάθους (μέρος ΙΙ)

1. Στο *t*-test, αν η μηδενική υπόθεση απορριφθεί σε *α*=0,05 τότε η παρατηρούμενη στάθμη σημαντικότητας του ελέγχου (*p*-value) θα πρέπει να είναι μεγαλύτερη από 0,05;
2. Εφαρμόσαμε την ANOVA σε ένα CRD που περιλαμβάνει 5 επεμβάσεις (π.χ. ποικιλίες) με 3 επαναλήψεις. Οι βαθμοί ελευθερίας που αντιστοιχούν στο πειραματικό σφάλμα είναι 10;
3. Αν θέλουμε να συγκρίνουμε ανά δύο 5 μέσους όρους τότε όλες οι δυνατές συγκρίσεις (στατιστικοί έλεγχοι) είναι σε πλήθος 25;
4. Αν πολλαπλασιάσουμε τις τιμές ενός δείγματος με τον ίδιο αριθμό, ο μέσος όρος πολλαπλασιάζεται επίσης με τον αριθμό αυτό;
5. Αν προσθέσουμε στις τιμές ενός δείγματος τον ίδιο αριθμό, η παραλλακτικότητα πολλαπλασιάζεται με τον αριθμό αυτό;

Ερωτήσεις Σωστού-Λάθους (μέρος ΙΙΙ)

1. Ο μέσος όρος είναι ο πιο έγκυρος δείκτης κεντρικής τάσης;
2. Η Κατανομή *F* προκύπτει από το λόγο δύο τυχαίων μεταβλητών που ακολουθούν Κανονική Κατανομή;
3. Για 10 β.ε. και ε.σ. α=0,05, ανάμεσα στις τιμές -2,228 και +2,228 το εμβαδό της επιφάνειας κάτω από την καμπύλη της *t*-Κατανομής είναι ίσο με 0,95;
4. Κατά την απογραφή των μονάδων ενός πληθυσμού, είναι σημαντικό να υπολογίζονται διαστήματα εμπιστοσύνης και να διενεργούνται έλεγχοι υποθέσεων;
5. Παράτυπες (*outliers*) είναι οι τιμές μιας κατανομής για τις οποίες τα αντίστοιχα *z-scores* είναι μεγαλύτερα από 3;

Απαντήσεις

- Μέρος I: 1)Λ, 2)Σ, 3)Λ, 4)Σ, 5)Λ, 6)Σ
- Μέρος II: 1)Λ, 2)Σ, 3)Λ, 4)Σ, 5)Λ
- Μέρος III: 1)Λ, 2)Λ, 3)Σ, 4)Λ, 5)Σ

Προσοχή:

Για κάθε λάθος απάντηση δεν λαμβάνεται υπόψη μία σωστή.

Σκηνή Δεύτερη

Ερωτήσεις Σύντομης Ανάπτυξης (Ι)

1. Τι εκφράζει ο συντελεστής παραλλακτικότητας *CV*;
2. Στην ANOVA, τι εκφράζει το Μέσο Τετράγωνο που αντιστοιχεί στο Σφάλμα;
3. Σε ποιο διάστημα παίρνει τιμές ο Συντελεστής Συνάφειας *V* του *Cramer*;
4. Τι εκφράζει το επίπεδο σημαντικότητας α σε ένα στατιστικό έλεγχο;
5. Τι εκφράζει ένα (1-α)% διάστημα εμπιστοσύνης για μια παράμετρο ενός πληθυσμού.
6. Πότε μια τυχαία μεταβλητή ονομάζεται διακριτή;
7. Σε ένα στατιστικό έλεγχο, πότε διαπράττουμε σφάλμα Τύπου Ι;
8. Αναφέρετε ένα παράδειγμα ζευγαρωτών παρατηρήσεων.

Ερωτήσεις Σύντομης Ανάπτυξης (ΙΙ)

1. Τι εκφράζει το Κεντρικό Οριακό Θεώρημα.
2. Δώστε ένα παράδειγμα Διωνυμικής Κατανομής.
3. Πότε χρησιμοποιούμε την Κατανομή *Poisson* ως προσέγγιση της Διωνυμικής;
4. Έστω ότι ρίχνουμε ένα ζάρι και ένα νόμισμα. Ποιος είναι ο Δειγματοχώρος; Να παρασταθεί και διαγραμματικά.
5. Πότε δύο ενδεχόμενα A και B είναι ανεξάρτητα;
6. Πότε δύο ενδεχόμενα A και B είναι ασυμβίβαστα;
7. Τι εκφράζει η πιθανότητα $P(A \cup B)$;
8. Δώστε ένα παράδειγμα όπου εμπλέκεται δεσμευμένη πιθανότητα.
9. Ποιες είναι οι προϋποθέσεις εφαρμογής του ελέγχου X^2;

Ερωτήσεις Σύντομης Ανάπτυξης (ΙΙΙ)

1. Αν Α, Β και Γ παριστάνουν τα σύνολα των φοιτητών που διαβάζουν τις εφημερίδες Μ1, Μ2 και Μ3 αντίστοιχα. Να εκφράσετε λεκτικά τα σύνολα: α) $A \cup B \cup Γ$, β) ΑΒΓ, γ) $(A \cup B)'$.
2. Οι παραλλακτικότητες δύο δειγμάτων είναι 25 και 15 (n_1=7 και n_2=31). Να ελεγχθεί σε α=0,05 ότι οι δύο παραλλακτικότητες (των αντίστοιχων πληθυσμών) είναι ίσες.

Σκηνή Τρίτη

Πρόβλημα 1_{36}

Μια γεωργική υπηρεσία για να εκτιμήσει την αποδοτικότητα δύο υβριδίων καλαμποκιού τα έδωσε σε 8 γεωργούς, που τα χωράφια τους βρίσκονταν σε διαφορετικές περιοχές, και εκτίμησε τις αποδόσεις κατά στρέμμα (βλέπε Πίνακα).

Γεωργός	Υβρίδιο Α	Υβρίδιο Β
1	344	320
2	348	316
3	224	232
4	372	364
5	336	308
6	372	328
7	292	296
8	316	264

Πρόβλημα 1 (συνέχεια)

1. Να υπολογιστούν οι μέσοι όροι της απόδοσης για τα δύο υβρίδια.
2. Να υπολογιστούν οι τυπικές αποκλίσεις της απόδοσης για τα δύο υβρίδια.
3. Να υπολογιστούν τα τυπικά σφάλματα στην εκτίμηση των μέσων όρων για τα δύο δείγματα.
4. Να υπολογιστούν οι συντελεστές παραλλακτικότητας *CV* για τα δύο δείγματα.
5. Ποιο από τα δύο δείγματα παρουσιάζει μεγαλύτερη ομοιογένεια;
6. Ποιο στατιστικό έλεγχο θα εφαρμόσετε για να διαπιστώσετε ποιο υβρίδιο είναι πιο αποδοτικό;
7. Διατυπώστε τη Μηδενική και την Εναλλακτική Υπόθεση του ελέγχου.
8. Έστω ότι ο έλεγχος που εφαρμόστηκε έδωσε p=0,032. Σε τι συμπέρασμα καταλήγετε;
9. Ποιο υβρίδιο είναι πιο αποτελεσματικό (σε α=0,05);
10. Αν το επίπεδο σημαντικότητας προκαθοριστεί σε α=0,01. Σε τι συμπέρασμα θα καταλήγατε με βάση την προηγούμενη τιμή p;
11. Έστω ότι ο έλεγχος που εφαρμόστηκε έδωσε p=0,344. Σε τι συμπέρασμα καταλήγετε;

Δίνονται: Άθροισμα Τετραγώνων των Διαφορών για το πρώτο Υβρίδιο=16798, Άθροισμα Τετραγώνων των Διαφορών για το δεύτερο Υβρίδιο=11738

Πρόβλημα 1: Υπολογισμοί

Γεωργός	Υβρίδιο Α	Υβρίδιο Β	ΑΤ1	ΑΤ2
1	344	320	342.25	272.25
2	348	316	506.25	156.25
3	224	232	10302.25	5112.25
4	372	364	2162.25	3660.25
5	336	308	110.25	20.25
6	372	328	2162.25	600.25
7	292	296	1122.25	56.25
8	316	264	90.25	1560.25
Αθροίσματα	**2604**	**2428**	**16798**	**11438**
ΜΟ	325.5	303.5		
ΤΑ	49.0	40.4		
CV	15.0%	13.3%		
Std Error	17.3	14.3		

Πρόβλημα 2

Σε ένα πείραμα μελετήθηκε η θερμοκρασία του εδάφους σε βάθος 5 εκ. σε τρεις διαφορετικές θέσεις και σε πέντε σημεία για κάθε θέση με τα εξής αποτελέσματα σε βαθμούς Κελσίου.

Θέση Α	Θέση Β	Θέση Γ
9	10	12
9	12	12
10	11	10
7	10	13
8	15	11

Πρόβλημα 2 (*συνέχεια*)

1. Διαφέρει στατιστικά σημαντικά, σε επίπεδο σημαντικότητας α=0,05, η θερμοκρασία στις τρεις θέσεις;
2. Να διατυπώσετε τη Μηδενική και Εναλλακτική Υπόθεση του ελέγχου που θα εφαρμόσετε.
3. Αν υπάρχουν διαφορές, ποιες θέσεις διαφέρουν μεταξύ τους με βάση το κριτήριο της ΕΣΔ (σε α=0,05);
4. Έχει το πείραμα ικανοποιητική ακρίβεια;
5. Τι τιμή έχει ο Διορθωτικός Όρος;

Δίνονται: Συνολικό Άθροισμα Τετραγώνων=57,6 και Άθροισμα Τετραγώνων Σφάλματος=27,6

Πρόβλημα 2: Υπολογισμοί

Descriptive Statistics

Dependent Variable: tem

loc	Mean	Std. Deviation	N
1	8.60	1.140	5
2	11.60	2.074	5
3	11.60	1.140	5
Total	10.60	2.028	15

Tests of Between-Subjects Effects

Dependent Variable: tem

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	30.000[a]	2	15.000	6.522	.012
Intercept	1685.400	1	1685.400	732.783	.000
loc	30.000	2	15.000	6.522	.012
Error	27.600	12	2.300		
Total	1743.000	15			
Corrected Total	57.600	14			

a. R Squared = .521 (Adjusted R Squared = .441)

Multiple Comparisons

Dependent Variable: tem

LSD

(I) loc	(J) loc	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-3.00*	.959	.009	-5.09	-.91
	3	-3.00*	.959	.009	-5.09	-.91
2	1	3.00*	.959	.009	.91	5.09
	3	.00	.959	1.000	-2.09	2.09
3	1	3.00*	.959	.009	.91	5.09
	2	.00	.959	1.000	-2.09	2.09

Based on observed means.

*. The mean difference is significant at the .05 level.

$ΕΣΔ_{0,05}$=2,01

***CV*=14,3%**

ΔΟ=1685,4

Πρόβλημα 3

- Από μια ποικιλία βαμβακιού πάρθηκε τυχαίο δείγμα 20 φυτών και προσδιορίστηκε το μέσο βάρος του σύσπορου βαμβακιού ενός καρυδιού από το κάθε φυτό. Τα αποτελέσματα σε γραμμάρια ήταν τα εξής:

Α/Α Φυτού	Μέσο βάρος	Α/Α Φυτού	Μέσο βάρος
1	8,4	11	8,0
2	5,8	12	7,7
3	7,8	13	7,0
4	6,4	14	7,7
5	7,9	15	7,3
6	7,2	16	6,7
7	7,3	17	6,3
8	8,4	18	7,3
9	5,1	19	6,0
10	7,6	20	4,2

Πρόβλημα 3 (*συνέχεια*)

1. Με βάση τα δεδομένα του δείγματος αυτού εκτιμήστε το μέσο βάρος καρυδιού σε ολόκληρη την ποικιλία με ένα 95% και με ένα 90% διάστημα εμπιστοσύνης.
2. Να υπολογίστε τα όρια εμπιστοσύνης της παραλλακτικότητας του πληθυσμού από τον οποίο προήλθε το δείγμα για επίπεδο εμπιστοσύνης 95%.
3. Να υπολογίσετε το μέγεθος δείγματος που απαιτείται για να έχει το διάστημα εμπιστοσύνης για το μέσο όρο σε α=0,05 εύρος ίσο με 0,5.

Απάντηση

- Η διακύμανση σ^2 του πληθυσμού άγνωστη και θα πρέπει να εκτιμηθεί από το δείγμα (δηλ. θα πρέπει να υπολογίσουμε την s^2)
- α=0,05 και α=0,10
- n=20<30.
- Υπολογίζουμε το μέσο όρο $\boxed{\bar{X} = 7,005}$
- Υπολογίζουμε την τυπική απόκλιση $\boxed{s = 1,097}$

Απάντηση (*συνέχεια*)

- Οι βαθμοί ελευθερίας είναι 20-1=19.
- Θα χρησιμοποιήσουμε τους πίνακες της *t*-Κατανομής.

$$\bar{X} - t_{a/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{a/2}\frac{s}{\sqrt{n}} \Rightarrow$$

$$7,005 - 2,093\frac{1,097}{\sqrt{20}} \leq \mu \leq 7,005 + 2,093\frac{1,097}{\sqrt{20}} \Rightarrow$$

$$(6,492, \quad 7,518)$$

Το 95% δ.ε.

Κρίσιμη τιμή της *t*-Κατανομής για 19 β.ε. και επίπεδο σημαντικότητας *α*/2=0,025

Απάντηση (*συνέχεια*)

$$\bar{X} - t_{a/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{X} + t_{a/2}\frac{s}{\sqrt{n}} \Rightarrow$$

$$7,005 - 1,729\frac{1,097}{\sqrt{20}} \le \mu \le 7,005 + 1,729\frac{1,097}{\sqrt{20}} \Rightarrow$$

$$(6,581, \quad 7,429)$$

Το 90% δ.ε.

Κρίσιμη τιμή της *t*-Κατανομής για 19 β.ε. και επίπεδο σημαντικότητας $\alpha/2$=0,05

Απάντηση (συνέχεια)

$$\frac{(n-1)s^2}{X^2_{a/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{X^2_{1-a/2}} \Rightarrow$$

$$\frac{19.(1,097)^2}{32,852} \leq \sigma^2 \leq \frac{19.(1,097)^2}{8,907} \Rightarrow$$

$$(0,696, \quad 2,567)$$

Το 95% δ.ε. για την παραλλακτικότητα

$$s_1 = 1,097 \Rightarrow s_1^2 = 1,203$$

Απάντηση (συνέχεια)

$$s_1 = 1,097 \Rightarrow s_1^2 = 1,203 \quad \mu\varepsilon\ 19\ \beta.\varepsilon.$$

$$n \geq \frac{4t_{a/2}^2 s^2}{d^2} \Rightarrow n \geq \frac{4.(2,093)^2.1,203}{0,5^2} = 84,32 \approx 85$$

1

Βρίσκουμε την κρίσιμη τιμή της *t*-Κατανομής για *α*=0,025 και 85-1=84 β.ε. και επαναλαμβάνουμε τον υπολογισμό

$$n \geq \frac{4t_{a/2}^2 s^2}{d^2} \Rightarrow n \geq \frac{4.(1,989)^2.1,203}{0,5^2} = 76,15 \approx 77$$

2

Πρόβλημα 4

- Σε μια μελέτη για το αποτέλεσμα της έκθεσης των ανθέων του φυτού *XXXR* σε διαφορετικές περιβαλλοντικές συνθήκες επιλέχθηκαν 10 υγιή φυτά με άνθη ελεύθερα εκτεθειμένα στην κορυφή και άνθη όσο το δυνατόν πιο καλά κρυμμένα στο κάτω μέρος.
- Στη συνέχεια καταμετρήθηκαν οι αριθμοί των σπόρων, που βρέθηκαν να είναι:

Πρόβλημα 4 (*συνέχεια*)

Α/Α Φυτού	1	2	3	4	5	6	7	8	9	10
Άνθη στην κορυφή	4	5	6	4	5	4	4	3	5	7
Άνθη στο κάτω μέρος	5	4	5	3	4	4	4	4	3	2

Με βάση τα παραπάνω δεδομένα να βρείτε ένα 90% διάστημα εμπιστοσύνης για την πραγματική διαφορά μ_1-μ_2 στον αριθμό των σπόρων του φυτού *XXXR* που προέρχεται από άνθη στο πάνω και στο κάτω μέρος του φυτού.

Πρόβλημα 4 (*συνέχεια*)

- Ο έλεγχος Kolmogorov-Smirnov για τον έλεγχο της Κανονικότητας των διαφορών έδωσε *p*-value>0,10. Σε τι συμπέρασμα καταλήγετε;
- Ποια είναι η Μηδενική Υπόθεση του ελέγχου Kolmogorov-Smirnov;

Απάντηση

- Επειδή οι μετρήσεις αναφέρονται στο ίδιο φυτό, είναι ζευγαρωτές και το 95% δ.ε. δίνεται:

$$\boxed{\bar{z} \pm \frac{s_z}{\sqrt{n}} t_{n-1;a/2}}$$

Οι διαφορές x_i-y_i

- Όπου z_i= -1, 1, 1, 1, 1, 0, 0, -1, 2 και 5

Απάντηση (συνέχεια)

$$\overline{z} = 0{,}900, \quad s_z = 1{,}729$$

$$\overline{z} \pm \frac{s_z}{\sqrt{n}} t_{n-1;a/2} \Rightarrow 0{,}900 \pm \frac{1{,}729}{\sqrt{10}} 1{,}833$$

$\mu\varepsilon$

$$t_{n-1;a/2} = t_{9;0,05} = 1{,}833$$

$$(-0{,}102, \quad 1{,}902)$$

Το 90% δ.ε.

Απάντηση (*συνέχεια*)

- Η υπόθεση ότι οι διαφορές προέρχονται από Κανονική Κατανομή δεν μπορεί να απορριφθεί σε ε.σ. α=0,05.
- H_0: Οι διαφορές ακολουθούν Κανονική Κατανομή.

Πρόβλημα 5

- Το επιστημονικό περιοδικό *Journal of Fish Biology* δημοσίευσε μια μελέτη που έκανε σύγκριση των παράσιτων που βρέθηκαν στα είδη ψαριών στη Μεσόγειο και στον Ατλαντικό.
- Στη Μεσόγειο από τα 588 ψάρια που πιάστηκαν και εξετάστηκαν βρέθηκαν μολυσμένα από παράσιτα τα 211. Στον Ατλαντικό ωκεανό, από 123 ψάρια που εξετάστηκαν βρέθηκαν μολυσμένα τα 26.

1. Συγκρίνετε την αναλογία των παράσιτων στις δύο θάλασσες χρησιμοποιώντας ένα 90% διάστημα εμπιστοσύνης. Ερμηνεύστε τα αποτελέσματα.
2. Ποιον έλεγχο υπόθεσης θα χρησιμοποιούσατε εναλλακτικά; Ποια θα είναι η Μηδενική Υπόθεση του ελέγχου αυτού;
3. Με βάση το διάστημα εμπιστοσύνης για τη διαφορά των δύο αναλογιών μπορείτε να πάρετε απόφαση σχετικά με την απόρριψη ή όχι της Μηδενικής Υπόθεσης στο προηγούμενο ερώτημα;

Απάντηση

$$\hat{p}_1 = \frac{211}{588} = 0,36 \quad \hat{p}_2 = \frac{26}{123} = 0,211$$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}},$$

$$n = 588, \quad m = 123, \quad z_{0,10/2} = z_{0,05} = 1,64$$

$$(0,36 - 0,211) \pm 1,64\sqrt{\frac{0,36\times 0,64}{588} + \frac{0,211\times 0,789}{123}} =$$

$$0,149 \pm 0,069 \Rightarrow$$

$$(0,08, \quad 0,218)$$

Το 90% δ.ε.

Απάντηση (*συνέχεια*)

- Η διαφορά των δύο ποσοστών είναι θετική.
- Η αναλογία p_1 είναι μεγαλύτερη από την αναλογία p_2 (σε ε.σ. α=0,10).
- Η Μεσόγειος έχει μεγαλύτερη αναλογία μολυσμένων ψαριών απ’ ό,τι ο Ατλαντικός Ωκεανός.

Απάντηση (*συνέχεια*)

- Έλεγχο για τη σύγκριση δύο ποσοστών (βλέπε Τυπολόγιο)
- Αφού το διάστημα εμπιστοσύνης δεν περιέχει το μηδέν οι δύο αναλογίες διαφέρουν στατιστικά σημαντικά σε ε.σ. α=0,90.

Πρόβλημα 6

- Η απόδοση σε σύσπορο βαμβάκι (g/πειρ. τεμ.) τυχαίων γραμμών βαμβακιού ενός πληθυσμού μάρτυρα κι ενός από σπόρο ακτινοβολημένο με ακτίνες ***γ*** δίνεται στον πίνακα που ακολουθεί.

1. Διαφέρουν τα δύο δείγματα ως προς τη μέση απόδοση σε ε.σ. α=0,05;
2. Ποιο είναι το τυπικό σφάλμα της διαφοράς των δύο μέσων όρων;

Πρόβλημα 6 (*συνέχεια*)

Μάρτυρας	Ακτινοβολημένο
980	1130
1310	1320
1090	490
1270	980
1240	760
1440	1020
1230	1290
1180	1250
1050	1120
1060	910
	530

Απάντηση

- Ο Στατιστικός Έλεγχος είναι δίπλευρος:

$$H_0: \mu_1=\mu_2$$

$$H_1: \mu_1\neq\mu_2$$

Σε ε.σ. α=0,05

Απάντηση (*συνέχεια*)

- Υπολογίζουμε:

$$n_1 = 10 \qquad \bar{Y}_1 = 1.185,0 \qquad s_1^2 = 19.761,1$$

$$n_2 = 11 \qquad \bar{Y}_2 = 981,8 \qquad s_2^2 = 82.416,4$$

- Ελέγχουμε την ισότητα (**ομοιογένεια**) των παραλλακτικοτήτων των δύο πληθυσμών:

$$F = \frac{82.416,4}{19.761,1} = 4,17 > F_{10,9;0,025} = 3,96 \Rightarrow$$

Οι δύο παραλλακτικότητες διαφέρουν στατιστικά σημαντικά σε ε.σ. *α*=0,05

Απάντηση (*συνέχεια*)

- Απορριπτική Περιοχή του Ελέγχου:

$$\boxed{R = \left\{ |t| > t_{\nu;a/2} \right\}}$$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{\bar{Y}_1 - \bar{Y}_2}{s_{\bar{Y}_1 - \bar{Y}_2}}$$

$$\text{Α}\nu \quad n_1 = n_2 = n \quad \tau\acute{o}\tau\varepsilon \quad \nu = 2(n-1)$$

$$\text{Α}\nu \quad n_1 \neq n_2 \quad \tau\acute{o}\tau\varepsilon$$

$$\nu = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2} \right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}}$$

Απάντηση (*συνέχεια*)

- Κάνουμε υπολογισμούς:

Τυπικό Σφάλμα Διαφοράς των δύο μέσων όρων

$$s_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}} \Rightarrow s_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{19.761,1}{10}+\frac{82.416,4}{11}} = \sqrt{9.468,5} = 97,3$$

$$t = \frac{1.185,0-981,8}{97,3} = 2,088$$

Βαθμοί Ελευθερίας

$$\nu = \frac{(19.761,1/10+82.416,4/11)^2}{\left[(19.761,1/10)^2/9\right]+\left[(82.416,4/11)^2/10\right]} = 14,8 \approx 15$$

Απάντηση (*συνέχεια*)

- Η κρίσιμη τιμή της $$\boxed{t_{15;0,025} = 2,131}$$

- Η τιμή του *t* που υπολογίσαμε από το δείγμα 2,088 είναι μικρότερη από την κρίσιμη-θεωρητική της *t* Κατανομής, δηλαδή πέφτει στην Περιοχή Αποδοχής του ελέγχου.

- Συνεπώς με βάση τα διαθέσιμα δεδομένα η μηδενική υπόθεση παραμένει (δεν απορρίπτεται) σε ε.σ. α=0,05.

- Τα δύο δείγματα δεν διαφέρουν στατιστικά σημαντικά ως προς τη μέση απόδοση σε ε.σ. α=0,05.

Πρόβλημα 7

- Ένας ερευνητής θέλει να ελέγξει αν μια νέα σειρά πειραματόζωων είναι πιο **ομοιόμορφη** από άποψη σωματικού βάρους σε σύγκριση με την εν χρήσει.

1. Από ένα δείγμα 25 ζώων της νέας σειράς υπολογίζει την παραλλακτικότητα σε 750 (s^2) ενώ η εν χρήσει σειρά έχει παραλλακτικότητα 625 (σ^2). Ποιον στατιστικό έλεγχο θα πραγματοποιήσει ο ερευνητής;

Απάντηση

- Το πρόβλημα ανάγεται στον Έλεγχο Υπόθεσης ότι η παραλλακτικότητα ενός πληθυσμού έχει μια συγκεκριμένη τιμή.
- Στη συγκεκριμένη περίπτωση είναι πιο λογικό να υποθέσουμε ότι ο έλεγχος θα αφορά στην ερευνητική υπόθεση ότι η νέα σειρά πειραματόζωων έχει παραλλακτικότητα μεγαλύτερη από 625.

Απάντηση (*συνέχεια*)

Ο στατιστικός έλεγχος που πρέπει να πραγματοποιηθεί είναι:

H_0: $\sigma^2=625$ (ή $\sigma^2 \leq 625$)

H_1: $\sigma^2>625$

Σε ε.σ. $\alpha=0{,}05$.

Απάντηση (*συνέχεια*)

H_0	H_1		
$\sigma^2=\sigma_0^2$	$\sigma^2>\sigma_0^2$	$R=\{X^2>X^2_{n-1;\,a}\}$	όπου $X^2=\dfrac{(n-1)s^2}{\sigma_0^2}$
	$\sigma^2<\sigma_0^2$	$R=\{X^2<\mathscr{X}'^2_{n-1;\,1-a}\}$	
	$\sigma^2\neq\sigma_0^2$	$R=\{X^2>X^2_{n-1;\,a/2}$ ή $X^2<\mathscr{X}'^2_{n-1;\,1-a/2}\}$	

Υπολογίζουμε το στατιστικό:

$$X^2=\frac{(n-1)s^2}{\sigma_0^2}\Rightarrow$$

$$X^2=\frac{(25-1)750}{625}=28{,}8$$

Απάντηση (*συνέχεια*)

- Η απορριπτική περιοχή είναι:

$$\boxed{\begin{aligned} &R=\left\{X^{2}>X_{n-1;a}^{2}\right\} \Rightarrow \\ &R=\left\{X^{2}>X_{24;0,05}^{2}\right\}=\left\{X^{2}>36,42\right\} \end{aligned}}$$

- Περιλαμβάνει το δεξί άκρο της X^2 Κατανομής, δηλ. τις τιμές που είναι μεγαλύτερες από 36,42. Η τιμή που υπολογίσαμε από το δείγμα 28,8 βρίσκεται μέσα στην περιοχή αποδοχής και επομένως η μηδενική υπόθεση δεν μπορεί να απορριφθεί σε ε.σ. α=0,05.

Πρόβλημα 8

- Το ύψος των δένδρων ενός δάσους μετριέται συνήθως από το έδαφος. Η μέθοδος αυτή είναι ακριβής αλλά δαπανηρή.
- Σύμφωνα με μια άλλη μέθοδο το ύψος μπορεί να υπολογιστεί με βάση αεροφωτογραφίες.
- Για να μελετήσουμε τη δυνατότητα χρησιμοποίησης αεροφωτογραφιών, μετρήσαμε το ύψος 10 δένδρων και με τους δύο τρόπους και πήραμε τα παρακάτω αποτελέσματα:

Πρόβλημα 8 (*συνέχεια*)

	Ύψος δένδρων σε πόδια	
Α/Α Δένδρου	**Από το έδαφος**	**Από αεροφωτογραφία**
1	**43**	**37**
2	**39**	**35**
3	**39**	**34**
4	**42**	**41**
5	**46**	**39**
6	**43**	**37**
7	**38**	**35**
8	**44**	**40**
9	**51**	**48**
10	**43**	**36**

Μπορούμε να ισχυριστούμε ότι οι δύο μέθοδοι είναι ισοδύναμες σε ε.σ. α=0,05;

Απάντηση

- Πρόκειται για ζευγαρωτές παρατηρήσεις. Οι μετρήσεις και με τις δύο μεθόδους αφορούν στην ίδια πειραματική ή δειγματοληπτική μονάδα (το ίδιο δένδρο μετρήθηκε δύο φορές).

Απάντηση (*συνέχεια*)

- Ο Στατιστικός Έλεγχος είναι δίπλευρος:

H_0: $\mu_1=\mu_2$

H_1: $\mu_1\neq\mu_2$

Σε ε.σ. α=0,05

Απάντηση (*συνέχεια*)

- Υπολογίζουμε τις διαφορές x_i-y_i

Α/Α Δένδρου	Από το έδαφος x_i	Από αεροφωτογραφία y_i	Διαφορές $x_i - y_i$
1	43	37	6
2	39	35	4
3	39	34	5
4	42	41	1
5	46	39	7
6	43	37	6
7	38	35	3
8	44	40	4
9	51	48	3
10	43	36	7

Απάντηση (*συνέχεια*)

- Υπολογίζουμε:

$$\bar{z} = 4,6$$
$$s_z^2 = 3,82$$
$$s_z = 1,96$$

- Απορριπτική Περιοχή του Ελέγχου:

$$R = \left\{ \left| \frac{\bar{z}\sqrt{n}}{s_z} \right| > t_{n-1;a/2} \right\} \Rightarrow$$

$$R = \left\{ \left| \frac{4,6\sqrt{10}}{1,96} \right| > t_{9;0,025} \right\}$$

$$\left| \frac{4,6\sqrt{10}}{1,96} \right| = 7,42 > t_{9;0,025} = 2,262$$

Απάντηση (*συνέχεια*)

- Η τιμή του στατιστικού ***t*** που υπολογίσαμε από το δείγμα **7,42** είναι μεγαλύτερη από την κρίσιμη (θεωρητική) τιμή **2,262** και συνεπώς η μηδενική υπόθεση απορρίπτεται σε ε.σ. α=0,05.
- Οι δύο μέθοδοι δεν είναι ισοδύναμες σε ε.σ. α=0,05.
- Οι δύο μέθοδοι διαφέρουν στατιστικά σημαντικά σε ε.σ. α=0,05.

Πρόβλημα 9

- Έχουμε 5 δείγματα (μηδικής) F_1 απογόνων και θέλουμε να ελέγξουμε αν οι αναλογίες των κλάσεων γύρης είναι οι ίδιες στους αντίστοιχους πληθυσμούς (α=0,05)

	Κλάσεις Παραγωγής Γύρης				
F1	**1η**	**2η**	**3η**	**4η**	Σύνολο
1ο	**11**	**6**	**9**	**14**	**40**
2ο	**7**	**6**	**7**	**9**	**29**
3ο	**14**	**5**	**7**	**11**	**37**
4ο	**11**	**4**	**7**	**20**	**42**
5ο	**22**	**2**	**12**	**16**	**52**
Σύνολο	**65**	**23**	**42**	**70**	**200**

Πρόβλημα 9 (*συνέχεια*)

- Ποιον στατιστικό έλεγχο θα χρησιμοποιήσετε;
- Να διατυπώσετε τη Μηδενική και Εναλλακτική Υπόθεση του ελέγχου.
- Να κατασκευάσετε τον πίνακα με τα προφίλ γραμμών.
- Να κατασκευάσετε τον πίνακα με τις αναμενόμενες συχνότητες.
- Να εφαρμόσετε τον έλεγχο και να καταλήξετε σε συμπεράσματα.

Απάντηση

- **H_0:** Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης **είναι ίδιες** και για τις πέντε ομάδες απογόνων (F_1 γενεές)
- **H1:** Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης **δεν είναι ίδιες** και για τις πέντε ομάδες απογόνων (F_1 γενεές)

σε επίπεδο σημαντικότητας $\boldsymbol{\alpha}$**=0,05**

Απάντηση: Προφίλ Γραμμών

- Κατασκευάζουμε τον πίνακα που περιέχει τα “προφίλ” των γραμμών:

Δείγματα F1 απογόνων $\dfrac{\textit{συχνότητα κελιού}}{\textit{σύνολο γραμμής}} \times 100$ **Crosstabulation**

% within Δείγματα F1 απογόνων

		Κλάσεις Παραγωγής Γύρης				
		1	2	3	4	Total
Δείγματα F1 απογόνων	1	27.5%	15.0%	22.5%	35.0%	100.0%
	2	24.1%	20.7%	24.1%	31.0%	100.0%
	3	37.8%	13.5%	18.9%	29.7%	100.0%
	4	26.2%	9.5%	16.7%	47.6%	100.0%
	5	42.3%	3.8%	23.1%	30.8%	100.0%
Total		32.5%	11.5%	21.0%	35.0%	100.0%

$$\frac{11}{40} \times 100 = 27.5$$

$$\frac{16}{52} \times 100 = 30.8$$

Απάντηση: Αναμενόμενες Συχνότητες

- Υπολογίζουμε τις αναμενόμενες συχνότητες σε κάθε κελί του πίνακα:

$$\text{αναμενόμενη συχνότητα} = \frac{\text{σύνολο γραμμής} \times \text{σύνολο στήλης}}{\text{γενικό σύνολο}}$$

		Κλάσεις Παραγωγής Γύρης				
		1	2	3	4	Total
Δείγματα F1 απογόνων	1	13.0	4.6	8.4	14.0	40.0
	2	9.4	3.3	6.1	10.2	29.0
	3	12.0	4.3	7.8	13.0	37.0
	4	13.7	4.8	8.8	14.7	42.0
	5	16.9	6.0	10.9	18.2	52.0
Total		65.0	23.0	42.0	70.0	200.0

$$\frac{40 \times 65}{200} = 13.0$$

$$\frac{50 \times 72}{200} = 18.2$$

Απάντηση: Βοηθητικός Πίνακας

Δείγματα F1 απογόνων * Κλάσεις Παραγωγής Γύρης Crosstabulation

			Κλάσεις Παραγωγής Γύρης				
			1	2	3	4	Total
Δείγματα F1 απογόνων	1	Count	11	6	9	14	40
		Expected Count	13.0	4.6	8.4	14.0	40.0
	2	Count	7	6	7	9	29
		Expected Count	9.4	3.3	6.1	10.2	29.0
	3	Count	14	5	7	11	37
		Expected Count	12.0	4.3	7.8	13.0	37.0
	4	Count	11	4	7	20	42
		Expected Count	13.7	4.8	8.8	14.7	42.0
	5	Count	22	2	12	16	52
		Expected Count	16.9	6.0	10.9	18.2	52.0
Total		Count	65	23	42	70	200
		Expected Count	65.0	23.0	42.0	70.0	200.0

Count: Συχνότητα

Expected Count: Αναμενόμενη Συχνότητα

Απάντηση: Το Στατιστικό X^2

- Υπολογίζουμε το στατιστικό X^2:

$$\boxed{X^2 = \sum \frac{\left(\pi\alpha\rho\alpha\tau\eta\rho o\acute{\upsilon}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha - \alpha\nu\alpha\mu\varepsilon\nu\acute{o}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha\right)^2}{\alpha\nu\alpha\mu\varepsilon\nu\acute{o}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha}}$$

$$\boxed{X^2 = \frac{\left(11-13{,}0\right)^2}{13{,}0} + \frac{\left(6-4{,}6\right)^2}{4{,}6} + \cdots + \frac{\left(16-18{,}2\right)^2}{18{,}2} = 12{,}125}$$

Απάντηση: Χρήση Πίνακα της X^2 Κατανομής

- Συγκρίνουμε το στατιστικό X^2 με την κρίσιμη τιμή της X^2 Κατανομής, με (5-1)(4-1)=12 β.ε., σε επίπεδο σημαντικότητας α=0,05:

Ανατρέχουμε στους Πίνακες

της X^2 Κατανομής

Απάντηση: Σύγκριση

- $X^2(12)_{0,05}=21{,}03$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=12{,}125$ (δειγματικό)
- $X^2=12{,}125<21{,}03=X^2(12)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

Απάντηση: Απόφαση-Συμπέρασμα

- Η μηδενική υπόθεση δεν απορρίπτεται σε ε.σ. α=0,05.
- Με βάση τα διαθέσιμα δεδομένα δεν μπορεί να απορριφθεί η μηδενική υπόθεση σε ε.σ. α=0,05.
- Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης και για τις πέντε ομάδες απογόνων (F_1 γενεές) δεν διαφέρουν στατιστικά σημαντικά σε ε.σ. α=0,05.
- Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης "είναι ίδιες" και για τις πέντε ομάδες απογόνων (F_1 γενεές) σε ε.σ. α=0,05.

Πρόβλημα 10

Φυλλοφόρα μοσχεύματα δύο ποικιλιών ελιάς που ριζοβόλησαν ή όχι μετά από 84 ημέρες κάτω από υδρονέφωση

	Μοσχεύματα		
Ποικιλία	**Με ρίζες**	**Χωρίς ρίζες**	**Σύνολο**
Χαλκιδικής	**100**	**60**	**160**
Αμφίσσης	**109**	**51**	**160**
Σύνολο	**209**	**111**	**320**

Πρόβλημα 10 (*συνέχεια*)

- Να ελεγχθεί με δύο τρόπους αν τα ποσοστά ριζοβολίας των 2 φυλλοφόρων μοσχευμάτων ελιάς είναι ίσα (δεν διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας α=0,05).

Απάντηση

Στατιστικός Έλεγχος

$$\mathrm{H}_0 : p_1 = p_2$$

$$\mathrm{H}_1 : \textit{τα δύο ποσοστά διαφέρουν } (p_1 \neq p_2)$$

Απάντηση (*συνέχεια*)

- Υπολογίζουμε τις αναμενόμενες συχνότητες

Στο κελί (1,1) η E_{11}= $\dfrac{160\times 209}{320}=104,5$

Στο κελί (2,1) η E_{21}= $\dfrac{160\times 209}{320}=104,5$

Στο κελί (1,2) η E_{12}= $\dfrac{160\times 111}{320}=55,5$

Στο κελί (2,2) η E_{22}= $\dfrac{160\times 111}{320}=55,5$

Απάντηση (*συνέχεια*)

Διόρθωση Συνέχειας του *Yates*

$$X^2 = \frac{\left(|100-104,5|-0,5\right)^2}{104,5} + \frac{\left(|60-55,5|-0,5\right)^2}{55,5} + \frac{\left(|109-104,5|-0,5\right)^2}{104,5} + \frac{\left(|51-55,5|-0,5\right)^2}{55,5} =$$

$$= 0,88$$

Γενική Σχέση

$$X^2 = \sum \frac{\left(|O-E| - \frac{1}{2}\right)^2}{E}$$

***O*: Παρατηρούμενη Συχνότητα**

***E*: Αναμενόμενη-Θεωρητική Συχνότητα**

Απάντηση (*συνέχεια*)

- $X^2(1)_{0,05}=3,84$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=0,88$ (δειγματικό)
- $X^2=0,88<3,84=X^2(1)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

Απάντηση (*συνέχεια*)

- Τα δύο ποσοστά ριζοβολίας μπορούν να συγκριθούν και με το z-τεστ.

$$\boxed{z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}} \Longleftrightarrow \boxed{X^2 = \frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Απάντηση (*συνέχεια*)

- Ο Στατιστικός Έλεγχος z (z-test)

  H_0: p_1-p_2=0

  H_1: p_1-$p_2 \neq 0$

  ή

  H_0: p_1=p_2

  H_1: $p_1 \neq p_2$

  Σε ε.σ. α

Απάντηση (*συνέχεια*)

- **Απορριπτική Περιοχή:**

$$R = \left\{ |z| > z_{a/2} \right\}$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{s}$$

$$s \approx \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Σύμφωνα με τη Μηδενική Υπόθεση τα δύο ποσοστά είναι ίσα και επομένως μπορούμε να συγχωνεύσουμε τα δύο δείγματα σε ένα και να υπολογίσουμε ένα κοινό *p* (και *q*=1-*p*)

$$\hat{p} = \frac{209}{320} = 0,653$$

Τυπικό Σφάλμα της Διαφοράς των Δύο Ποσοστών

Απάντηση (*συνέχεια*)

- Η μηδενική υπόθεση **δεν απορρίπτεται** σε ε.σ. α=0,05.
- Με βάση τα διαθέσιμα δεδομένα **δεν μπορεί να απορριφθεί** η μηδενική υπόθεση σε ε.σ. α=0,05.
- Τα ποσοστά ριζοβολίας των 2 φυλλοφόρων μοσχευμάτων ελιάς **"είναι ίσα"** μεταξύ τους. Στις δύο ποικιλίες τα αντίστοιχα ποσοστά **δεν διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας α=0,05.**

Πρόβλημα 11

- Σε πρόσφατη έρευνα ρωτήθηκαν κάποια άτομα σχετικά με το ποιο θεωρούν ως το σημαντικότερο πρόβλημα σήμερα.
- Οι απαντήσεις στην ερώτηση αυτή διασταυρώθηκαν με τις διάφορες ηλικιακές κατηγορίες των ατόμων που συμμετείχαν στην έρευνα.
- Τα αποτελέσματα δίνονται στον πίνακα που ακολουθεί:

Πρόβλημα 11 (*συνέχεια*)

Κλάσεις Ηλικιών	Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία
18-24	24	4	23	82	12
25-34	42	2	20	146	13
35-44	54	0	12	136	18
45-54	40	1	6	121	9
55-64	37	1	7	68	15
65+	20	5	3	61	13

Πρόβλημα 11 (*συνέχεια*)

- Υπάρχει σχέση (συσχέτιση, συνάφεια) μεταξύ της ηλικίας των ερωτώμενων και της άποψή τους σχετικά με το ποιο είναι το σημαντικότερο πρόβλημα (α=0,05);
- Να κατασκευάσετε τον πίνακα με τα προφίλ γραμμών και στηλών.
- Να κατασκευάσετε τον πίνακα αντιστοιχιών.
- Να κατασκευάσετε τον πίνακα με τις αναμενόμενες συχνότητες.
- Ποιος είναι ο βαθμός συνάφειας των δύο εξεταζόμενων χαρακτηριστικών;
- Που οφείλεται κυρίως η συνάφεια;
- Ισχύουν οι προϋποθέσεις του ελέγχου που εφαρμόσατε;

Απάντηση: Προφίλ Γραμμών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

% within Κλάσεις Ηλικιών

		Πρόβλημα					Total
		Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία	
Κλάσεις Ηλικιών	18-24 ετών	16.6%	2.8%	15.9%	56.6%	8.3%	100.0%
	25-34	18.8%	.9%	9.0%	65.5%	5.8%	100.0%
	35-44	24.5%		5.5%	61.8%	8.2%	100.0%
	45-54	22.6%	.6%	3.4%	68.4%	5.1%	100.0%
	55-64	28.9%	.8%	5.5%	53.1%	11.7%	100.0%
	65+	19.6%	4.9%	2.9%	59.8%	12.7%	100.0%
Total		21.8%	1.3%	7.1%	61.7%	8.0%	100.0%

Απάντηση: Προφίλ Στηλών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

% within Πρόβλημα

		Πρόβλημα					Total
		Τζόγος	Στεγαστικό	Σπουδές	Οικον ομικό	Υγεία	
Κλάσεις Ηλικιών	18-24 ετών	11.1%	30.8%	32.4%	13.4%	15.0%	14.6%
	25-34	19.4%	15.4%	28.2%	23.8%	16.3%	22.4%
	35-44	24.9%		16.9%	22.1%	22.5%	22.1%
	45-54	18.4%	7.7%	8.5%	19.7%	11.3%	17.8%
	55-64	17.1%	7.7%	9.9%	11.1%	18.8%	12.9%
	65+	9.2%	38.5%	4.2%	9.9%	16.3%	10.3%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Απάντηση: Πίνακας Αντιστοιχιών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

% of Total

		Πρόβλημα					Total
		Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία	
Κλάσεις Ηλικιών	18-24 ετών	2.4%	.4%	2.3%	8.2%	1.2%	14.6%
	25-34	4.2%	.2%	2.0%	14.7%	1.3%	22.4%
	35-44	5.4%		1.2%	13.7%	1.8%	22.1%
	45-54	4.0%	.1%	.6%	12.2%	.9%	17.8%
	55-64	3.7%	.1%	.7%	6.8%	1.5%	12.9%
	65+	2.0%	.5%	.3%	6.1%	1.3%	10.3%
Total		21.8%	1.3%	7.1%	61.7%	8.0%	100.0%

Απάντηση: Αναμενόμενες Συχνότητες

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

Expected Count

		Πρόβλημα					Total
		Τζόγος	Στεγαστικό	Σπουδές	Οικον ομικό	Υ γεία	
Κλάσεις Ηλικιών	18-24 ετών	31.6	1.9	10.3	89.5	11.7	145.0
	25-34	48.6	2.9	15.9	137.6	17.9	223.0
	35-44	48.0	2.9	15.7	135.8	17.7	220.0
	45-54	38.6	2.3	12.6	109.2	14.2	177.0
	55-64	27.9	1.7	9.1	79.0	10.3	128.0
	65+	22.2	1.3	7.3	62.9	8.2	102.0
Total		217.0	13.0	71.0	614.0	80.0	995.0

Απάντηση: Αποτελέσματα Ελέγχου

Sig.➔*p*-value

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	Monte Carlo Sig. (2-sided) 99% Confidence Interval Lower Bound	Monte Carlo Sig. (2-sided) 99% Confidence Interval Upper Bound	Monte Carlo Sig. (1-sided) Sig.	Monte Carlo Sig. (1-sided) 99% Confidence Interval Lower Bound	Monte Carlo Sig. (1-sided) 99% Confidence Interval Upper Bound
Pearson Chi-Square	59.480[a]	20	.000	.000[b]	.000	.000			
Likelihood Ratio	55.136	20	.000	.000[b]	.000	.000			
Fisher's Exact Test	52.065			.000[b]	.000	.000			
Linear-by-Linear Association	.362[c]	1	.548	.563[b]	.550	.575	.291[b]	.279	.303
N of Valid Cases	995								

a. 6 cells (20.0%) have expected count less than 5. The minimum expected count is 1.33.

b. Based on 10000 sampled tables with starting seed 272886377.

c. The standardized statistic is -.601.

Αφού *p*<0,05 η Μηδενική Υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας *α*=0,05

Απάντηση: Συμπέρασμα

- Υπάρχει στατιστικά σημαντική συσχέτιση, σε επίπεδο σημαντικότητας α=0,05, μεταξύ της ηλικίας των ερωτώμενων και της άποψής τους σχετικά με το ποιο είναι το σημαντικότερο πρόβλημα.

Ποια είναι η ένταση-βαθμός της συνάφειας;

Symmetric Measures

		Value	Approx. Sig.	Monte Carlo Sig.		
				Sig.	99% Confidence Interval	
					Lower Bound	Upper Bound
Nominal by Nominal	Phi	.244	.000	.000[c]	.000	.000
	Cramer's V	.122	.000	.000[c]	.000	.000
N of Valid Cases		995				

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on 10000 sampled tables with starting seed 1090229469.

Η τιμή του δείκτη συνάφειας *V* του Cramer μαρτυρά ασθενούς εντάσεως συσχέτιση μεταξύ των δύο χαρακτηριστικών που διασταυρώνονται

Απάντηση: Ο δείκτης *V* του *Cramer*

$$V = \sqrt{\frac{X^2}{Np}}$$

όπου $p = min(k-1, l-1)$.

k **είναι ο αριθμός των γραμμών και *l* ο αριθμός των στηλών του πίνακα συμπτώσεων**

Παίρνει τιμές στο διάστημα [0, 1]

Απάντηση: Που οφείλεται κυρίως η συνάφεια των δύο χαρακτηριστικών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

			Πρόβλημα					
			Τζόγος	Στεγαστικό	Σπουδές	Οικον ομικό	Υ γεία	Total
Κλάσεις Ηλικιών	18-24 ετών	% within Κλάσεις Ηλικιών	16.6%	2.8%	15.9%	56.6%	8.3%	100.0%
		Adjusted Residual	-1.7	1.7	4.4	-1.4	.1	
	25-34	% within Κλάσεις Ηλικιών	18.8%	.9%	9.0%	65.5%	5.8%	100.0%
		Adjusted Residual	-1.2	.6	1.2	1.3	-1.4	
	35-44	% within Κλάσεις Ηλικιών	24.5%	.0%	5.5%	61.8%	8.2%	100.0%
		Adjusted Residual	1.1	-1.9	-1.1	.0	.1	
	45-54	% within Κλάσεις Ηλικιών	22.6%	.6%	3.4%	68.4%	5.1%	100.0%
		Adjusted Residual	.3	-1.0	-2.1	2.0	-1.6	
	55-64	% within Κλάσεις Ηλικιών	28.9%	.8%	5.5%	53.1%	11.7%	100.0%
		Adjusted Residual	2.1	-.6	-.8	-2.1	1.6	
	65+	% within Κλάσεις Ηλικιών	19.6%	4.9%	2.9%	59.8%	12.7%	100.0%
		Adjusted Residual	-.6	3.4	-1.7	-.4	1.8	
Total		% within Κλάσεις Ηλικιών	21.8%	1.3%	7.1%	61.7%	8.0%	100.0%

Adjusted Residual:

Διορθωμένο Τυποποιημένο Υπόλοιπο

Απάντηση (*συνέχεια*)

- Ισχύουν οι περιορισμοί του *Cochran*.

Πρόβλημα 12

Κατανομή των Αποδόσεων σε χλγ. Χλωρού Βάρους 168 Μεμονωμένων Φυτών Μηδικής

Πραγματικά Όρια Κλάσεων	Αριθμός Φυτών
<0,55	0
0,55-1,05	5
1,05-1,55	8
1,55-2,05	20
2,05-2,55	20
2,55-3,05	27
3,05-3,55	30
3,55-4,05	20
4,05-4,55	13
4,55-5,05	11
5,05-5,55	7
5,55-6,05	7
>6,05	0
Σύνολο	168

Πρόβλημα 12 (*συνέχεια*)

Να ελεγχθεί η υπόθεση ότι η απόδοση των φυτών κατανέμεται κανονικά (σε ε.σ. α=0,05).

Δίνονται:

$$\bar{X} = 3,18$$

$$s = 1,22$$

Στην πράξη υπολογίζονται από το δείγμα

Απάντηση (*συνέχεια*)

Στατιστικός Έλεγχος:

- H_0: Η απόδοση των φυτών κατανέμεται κανονικά (ακολουθεί την Κανονική Κατανομή)
- H_1: Η απόδοση των φυτών δεν κατανέμεται κανονικά (δεν ακολουθεί την Κανονική Κατανομή)

σε ε.σ. α=0,05

Απάντηση (*συνέχεια*)

- Θεωρητικές Πιθανότητες:

  Υπολογίζονται με τη βοήθεια της Τυποποιημένης Κανονικής Κατανομής.

  Για παράδειγμα:

$$P(Y \leq 0,55) = P\left(\frac{Y-3,18}{1,22} \leq \frac{0,55-3,18}{1,22}\right) = P(z \leq -2,16) =$$

$$= 0,5000 - 0,4846 = 0,0154$$

$$P(0,55 \leq Y \leq 1,55) = P\left(\frac{0,55-3,18}{1,22} \leq \frac{Y-3,18}{1,22} \leq \frac{1,55-3,18}{1,22}\right) =$$

$$= P(-2,16 \leq z \leq -1,75) = 0,4846 - 0,4599 = 0,0247$$

Απάντηση (*συνέχεια*)

- Αναμενόμενες Συχνότητες:

  Πολλαπλασιάζοντας τις θεωρητικές συχνότητες (πιθανότητες) με το σύνολο των φυτών του δείγματος (168 συνολικά φυτά).

  Για παράδειγμα:

  $2{,}59=168\times0{,}0154$

  $4{,}15=168\times0{,}0247$

Απάντηση (*συνέχεια*)

Κατανομή των Αποδόσεων σε χλγ. Χλωρού Βάρους 168 Μεμονωμένων Φυτών Μηδικής

Πραγματικά Όρια Κλάσεων	Αριθμός Φυτών	Θεωρητική Πιθανότητα	Αναμενόμενη Συχνότητα
<0,55	0	0,0154	2,59
0,55-1,05	5	0,0247	4,15
1,05-1,55	8	0,0500	8,40
1,55-2,05	20	0,0861	14,46
2,05-2,55	20	0,1253	21,05
2,55-3,05	27	0,1547	25,99
3,05-3,55	30	0,1617	27,17
3,55-4,05	20	0,1432	24,06
4,05-4,55	13	0,1075	18,06
4,55-5,05	11	0,0684	11,49
5,05-5,55	7	0,0368	6,18
5,55-6,05	7	0,0168	2,82
>6,05	0	0,0094	1,58
Σύνολο	168	1,0000	168,00

Απάντηση (*συνέχεια*)

- Υπολογίζουμε το στατιστικό X^2:

$$\boxed{X^2 = \frac{(0-2,59)^2}{2,59} + \frac{(5-4,15)^2}{4,15} + \cdots + \frac{(0-1,58)^2}{1,58} = 15,30}$$

- Συγκρίνουμε το στατιστικό X^2=15,30 με την κρίσιμη της X^2 Κατανομής με 10 β.ε. σε επίπεδο σημαντικότητας *α*=0,05.

Απάντηση (*συνέχεια*)

- $X^2(10)_{0,05}=18,31$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=15,30$ (δειγματικό)
- $X^2=15,30<18,31=X^2(10)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

Η μηδενική Υπόθεση παραμένει

Απάντηση (*συνέχεια*)

Απόφαση-Συμπέρασμα:

- Με βάση τα διαθέσιμα δεδομένα η μηδενική υπόθεση δεν μπορεί να απορριφθεί σε ε.σ. α=0,05.
- Δεν ανιχνεύθηκαν στατιστικά σημαντικές διαφορές μεταξύ θεωρητικών και δειγματικών τιμών.
- Δεν έχουμε λόγους να αμφιβάλλουμε ότι η κατανομή του χλωρού βάρους είναι η Κανονική (σε ε.σ. α=0,05).

Σκηνή 4η

Λυμένες Ασκήσεις στον Πίνακα

1

① Το ημερομίσθιο ν εργ. διαφέρει από το μ.ημ. κατά μέσο όρο 133 δρχ.

Διόρθωση Sheppard

$\sigma^{*2} = \sigma^2 - \frac{c^2}{12}$, c = πλάτος τάξεων.

κέντρο

$\bar{x} = 477$

Κλάσεις Ημερομισθίων	x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$	$f_i x_i^2$
150-250	200	5	1000	-277	76.729	383.645	200.000
250-350	300	13	3900	-177	31.329	407.277	1.170.000
350-450	400	20	8000	-77	5.929	118.580	3.200.000
450-550	500	35	17.500	23	529	18.515	8.750.000
550-650	600	18	10.800	123	15.129	272.322	6.480.000
650-750	700	7	4.900	223	49.729	348.103	3.430.000
750-850	800	2	1.600	323	104.329	208.658	1.280.000
Σύνολο		100	47.700			1.757.100	24.510.000

① $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} \Rightarrow \bar{x} = \frac{47.700}{100} = 477$

② $\sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i} \Rightarrow \sigma^2 = \frac{1.757.100}{100} = 17.571 \Rightarrow \sigma = \sqrt{17.571} = 132,56 \simeq 133$

$\sigma^2 = \frac{\Sigma f_i x_i^2}{\Sigma f_i} - \Big(\underbrace{\frac{\Sigma f_i x_i}{\Sigma f_i}}_{\bar{x}}\Big)^2 \Rightarrow \sigma^2 = \frac{24.510.000}{100} - \left(\frac{47.700}{100}\right)^2 = 245.100 - 227.529 =$

$= 17.571 \Rightarrow \sigma = 132,56 \simeq 133$

③ Διόρθωση Sheppard

$\sigma^{*2} = 17.571 - \frac{(100)^2}{12} = 17.571 - 833,33 = 16.737,67 \Rightarrow \sigma^* = 129,37 \simeq 129$

④ $C.V. = \frac{133}{477} \cdot 100 = 27,88\%$ ⑤ $C.V.^* = \frac{129}{477} \cdot 100 = 27,04\%$

2

Untitled

$1/12 = 8,3\%$.

	var00002	
1	8,00	50
2	9,00	15
3	9,00	16
4	15,00	9
5	16,00	8
6	16,00	20
7	20,00	32
8	25,00	9
9	32,00	25
10	44,00	16
11	48,00	48
12	50,00	44

Median

min = 8

max = 50

range = max − min = 50 − 8 = 42.

	ΑΣ	ΣΣ	ΑΣ ΑΣΣ
8	1	1/12	1
9	2	2/12	3
15	1	1/12	4
16	2	2/12	6
20	1	1/12	7
25	1	1/12	8
32	1	1/12	9
44	1	1/12	10
48	1	1/12	11
50	1	1/12	12

$M.O = (8+9+\dots+50)/12 = \frac{292}{12} = 24,333$

= άθροισμα τιμών / πλήθος παρατηρήσεων.

Mode = 16 (2 φορές) κ 9 (2 φορές).

Median $\longrightarrow \frac{n+1}{2}$ θέση : $\frac{12+1}{2} = \frac{13}{2} = 6,5$ θέση

δηλ. μεταξύ 6ης κ 7ης θέσης.

$6^{o\iota} \longrightarrow 16$, $7^{os} \longrightarrow 20$

median $= \frac{16+20}{2} = \frac{36}{2} = 18$.

$Q_{25} \longrightarrow \frac{n+1}{4} : \frac{12+1}{4} = \frac{13}{4} = 3\,1/4$ θέση

$3^{os} \longrightarrow 9$

$4^{os} \longrightarrow 15$

$15 - 9 = 6$.

άρα $Q_{25} = 9 + \frac{1}{4} \cdot 6 =$

$= 9 + 1,5 = 10,5 \longrightarrow Q_{25}$

$Q_{75} \longrightarrow \frac{3(n+1)}{4} = \frac{39}{4} = 9\frac{3}{4}$.

$9^{os}: \longrightarrow 32$

$44 - 32 = 12$

$10^{os}: \longrightarrow 44$

$Q_{75} = 32 + \frac{3}{4} \cdot 12 = 32 + 9 = 41 \longrightarrow Q_{75}$.

3

Παράδειγμα: Ρίχνουμε 2 ζάρια. Έστω X η τ.μ. που μετρά το άθροισμα των δύο ενδείξεων.

α) Η κατανομή πιθανότητας της X

β) Η αθροιστική συνάρτηση κατανομής

γ) Οι γρ. παραστάσεις.

δ) EX

ε) VAR X.

α)

1ος κύβος

Δειγματοχώρος

2ος κύβος	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Οι δυνατές τιμές της X είναι 2, 3, ... 12.

3

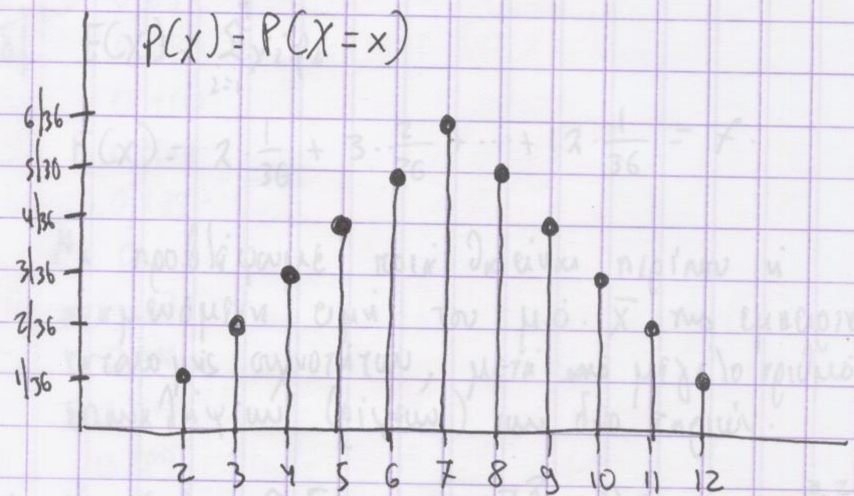
α) Κατανομή Πιθανότητας

X	f_i	P_i	F(X)
2	1	1/36	1/36
3	2	2/36	3/36
4	3	3/36	6/36
5	4	4/36	10/36
6	5	5/36	15/36
7	6	6/36	21/36
8	5	5/36	26/36
9	4	4/36	30/36
10	3	3/36	33/36
11	2	2/36	35/36
12	1	1/36	36/36

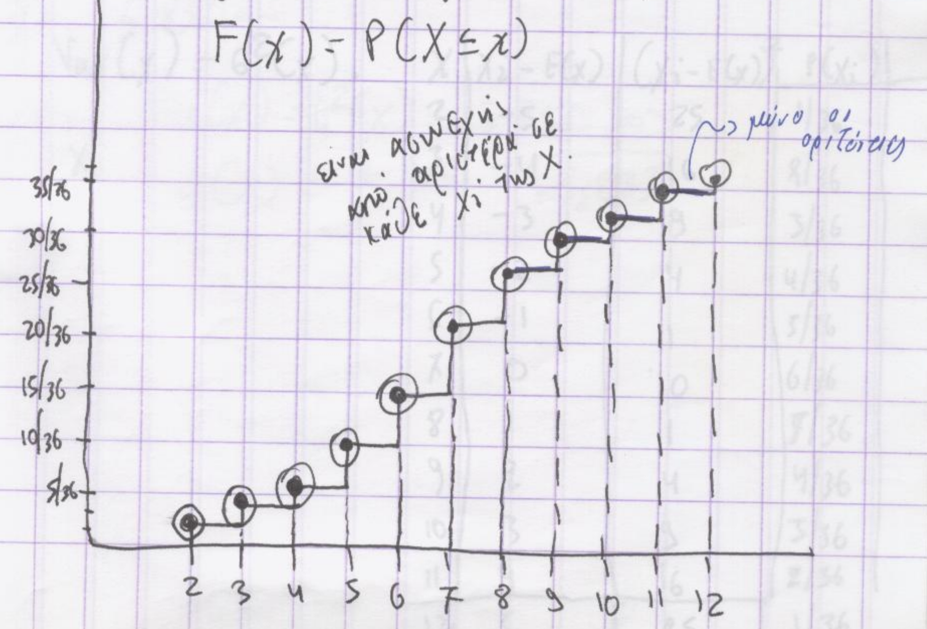
β) Αθροιστική Κατανομή

X	$F(x_i) = P(X \leq x)$
2	1/36
3	3/36
4	6/36
5	10/36
6	15/36
7	21/36
8	26/36
9	30/36
10	33/36
11	35/36
12	36/36.

γ). Κατανομή Πιθανότητας



Αθροιστική Συνάρτηση Κατανομής



Μαθηματική Ελπίδα: Μέση Τιμή - Αναμενόμενη Τιμή

δ) $E(X) = \sum_{i=1}^{\nu} x_i p_i$.

$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \cdots + 12 \cdot \frac{1}{36} = 7$.

Να προβλέψουμε ποιά θα είναι περίπου η αναμενόμενη τιμή του μ.ό. $\bar{X}$ της εμπειρικής κατανομής συχνοτήτων, μετά από μεγάλο αριθμό επαναλήψεων (ρίψεων) των δύο ζαριών.

ε) $Var(X) = E[X - E(X)]^2 = \sum (x_i - E(x))^2 p(x_i)$.

$Var(X) = \sigma^2(X)$.

X

X	$x_i - E(x)$	$(x_i - E(x))^2$	$p(x_i)$
2	−5	25	1/36
3	−4	16	2/36
4	−3	9	3/36
5	−2	4	4/36
6	−1	1	5/36
7	0	0	6/36
8	1	1	5/36
9	2	4	4/36
10	3	9	3/36
11	4	16	2/36
12	5	25	1/36

3

x_i	$(x_i - E(x))^2 \cdot P(x_i)$
2	0,694
3	0,889
4	0,750
5	0,444
6	0,139
7	0
8	0,139
9	0,444
10	0,750
11	0,889
12	0,694
Σ	5,833

άρα $Var(x) = \sigma^2(x) = 5,833 \Rightarrow$

$\sigma(x) = 2,415 \approx \sqrt{5,833}$.

[redacted]

4

[redacted]

Διωνυμική Κατανομή.

α) Να υπολογιστεί η πιθανότητα ότι 3 ή περισσότερα άτομα, σε τυχαίο δείγμα 10 ατόμων, προτιμούν τα ΕΛ.ΤΑ για τις ταχυμεταφορές τους, αν υποθέσουμε ότι το % των ατόμων στον πληθυσμό που προτιμούν τα ΕΛ.ΤΑ είναι $p = 0,35$

ΛΥΣΗ

$n=10,\ p=0,35,\ 1-p=0,65$

Η ζητούμενη πιθανότητα $P(X \geq 3) =$

$= P(X=3$ ή $X=4$ ή $X=5$ ή ... ή $X=10) =$

$= P(X=3)+P(X=4)+\dots+P(X=10) =$

$= 1 - P(X<3) = 1 - P(X=0$ ή $X=1$ ή $X=2)$

$= 1 - (P(X=0)+P(X=1)+P(X=2)) = 1 - 0,2617$

[redacted]

$P(X=0) =$ [redacted] $\rightarrow 0,0135 \quad = 0,7383$

$P(X=1) = \rightarrow 0,0725$

$P(X=2) = \rightarrow 0,1757$

$0,2617$

5

β) Ρίχνουμε ζάρι 6 φορές, 2 φορές το 5 (p?).

Λύση:

[redacted] Ε το ενδ. να έρθει 5

$p(E) = 1/6$. $p(\text{Αποτ.}) = 1 - 1/6 = 5/6$.

$n = 6$.

$x = 2$ (επιτυχίες)

$p = 1/6$, $q = 5/6$.

$P(X = x) = \binom{n}{x} p^x q^{n-x} \Rightarrow$

$P(X = 2) = \binom{6}{2} (1/6)^2 (5/6)^4 \Rightarrow$

$P(X = 2) = \frac{6!}{2!\,4!} \cdot \frac{1}{36} \cdot \frac{625}{1296} = 0,20$ ή 20%.

Διωνυμική Κατανομή:

- όταν $n < 50$
- η $p = ct$ η μεγαλύτερη του 10%.

$\binom{4}{0} 0,2^0\, 0,8^4 = (0,8)^4 = 0,4096$

$\binom{4}{1} = 0,2^1\, 0,8^3 = \frac{4!}{3!}\, 0,2 \cdot 0,8^3 = 0,8 \cdot 0,8^3 = 0,8^4 =$
$= 0,4096$.

[redacted]

Άσκηση.

20% ελ. λίδες, παίρνουμε δείγμα 4

P(μία λίδα)

P(0)

P (λιγότερες από δύο)

P (περισ. από δύο λίδες)

P (τουλ. δύο λίδες)

P (το πολύ δύο λίδες).

Ποιος είναι ο μέσος αριθμός κ ποια var των ελλατ. λιδών. σε δείγμα 400 λίδες.

Λύση:

Για $n=4$, $p=0,20$ κ $X=0,1,2,3,4$.

X	P(X)
0	0,4096
1	0,4096
2	0,1536
3	0,0256
4	0,0016

$P(X=1)=0,4096$

$P(X=0)=0,4096$

$P(X<2)$

$P(X>2)$

$P(X\geq 2)$

$P(X\leq 2)$

$\bar{X}=n.p=80$

$\sigma=\sqrt{npq}$
$=8$

6

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Poisson

7

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

- 3% φαρ. δηλώσεων δεν είναι ειλικρινείς

P; σε 150 φορ. δηλ.

α) 0, 1, 2, 3, 4, 5.

β) περισ. από 5.

γ) από 1 έως 3

δ) το πολύ 2.

Λύση:

Σπάνιο γεγονός $p = 0{,}03$

$\nu = 150$ μεγάλο

$\nu p = 0{,}03 \times 150 = 4{,}5 < 10$.

α) $P(X=0) = 0{,}0111$ $P(X=4) = 0{,}1898$

$P(X=1) = 0{,}0500$ $P(X=5) = 0{,}1708$

$P(X=2) = 0{,}1125$

$P(X=3) = 0{,}1687$

β) $P(X>5) = 1 - P(X \le 5) = 1 - 0{,}7029 = 0{,}2971$

γ) $P(1 \le X \le 3) = P(X=1) + P(X=2) + P(X=3)$

$= 0{,}3312$

δ) $P(X \le 2) = P(X=0) + P(X=1) + P(X=2) =$

$= 0{,}1736$.

Κανονική

Στρ. Μονάδα πρόκειται να δεχθεί 10.000 στρ.
Θέλουμε στολές. ⟶ 3 μεγέθη
α) Για στρ. με ύψος 165 μικρότερο. No1
β) -//- 165–182 No2
γ). 182 μεγαλύτερο. No3

Από άλλες έρευνες $\mu = 175$, $\sigma = 4$ cm
Πόσες στολές από κάθε μέγεθος

Λύση. ($N = 10.000$, $\mu = 175$, $\sigma = 4$ cm)

α) $P(x_1 \leq 165) = P\left(\frac{x_1-\mu}{\sigma} \leq \frac{165-175}{4}\right) =$

$= P(z_1 \leq -2,5) = \Pi(-2,5) = 0,0062$

$N \cdot \Pi(z_1) = N \cdot 0,0062 = 10.000 \cdot 0,0062 =$
$= 62$. στρ.

β) $P(\cancel{165} \leq x \leq 182) = P\left(\frac{165-175}{4} \leq \frac{x-\mu}{\sigma} \leq \frac{182-175}{4}\right) =$

$= P\left(-2,5 \leq \frac{x-\mu}{\sigma} \leq 1,75\right) = \Pi(1,75) - \Pi(-2,5)$
$= 0,9599 - 0,0062 = 0,9537$

$10.000 \times 0,9537 = 9.537$ στρ.

γ) $P(X \geq 182) = P\left(\frac{X-\mu}{\sigma} \geq \frac{182-175}{4}\right) =$

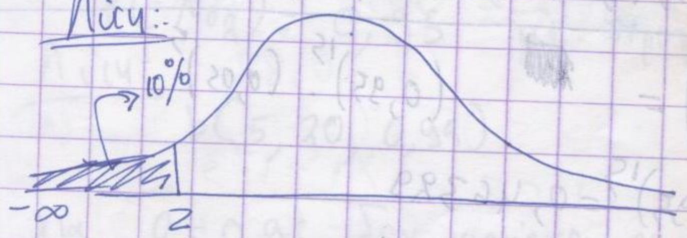
$= P(Z \geq 1,75) = 1 - P(Z < 1,75) =$

$= 1 - \Pi(1,75) = 1 - 0,9599 = 0,0401$

$10.000 \cdot 0,0401 = 401$ στρ.

- Κάτω ποιου αναστήματος βρίσκονται το 10% των στρατιωτών
- Άνω ποιου αναστήματος βρίσκονται το 15% των στρατιωτών.

Λύση:



$P = 10\% \rightsquigarrow z = -1,28.$

$-1,28 = \frac{x-175}{4} \Rightarrow$

$x = 169,88 \simeq 170$ cm.

$\binom{20}{5}(0,95)^5(0,05)^{15}$

8

$1-\Pi(z) = 15\% \Rightarrow$

$\Pi(z) = 85\% = 0,85$

$\Rightarrow z = 1,04$

$1,04 = \dfrac{x-175}{4} \Rightarrow x = 179$ cm

+++ • Ποια η P σε τυχ. δείγμα 20 στρατιωτών οι 5 να φορούν στολή No 2

$P(No_2) = 0,95$ από προηγούμενα. με στρογγυλοποίηση.

Λύση:

$b(5, 20, 0,95)$.

για $p = 0,95$ δεν υπάρχει σε πίνακες

Όμως: $b(\kappa, \nu, p) = b(\nu-\kappa, \nu, 1-p)$

$\Rightarrow b(5, 20, 0,95) = b(20-5, 20, 1-0,95)$

$= b(15, 20, 0,05)$

$= 0$

Σκηνή Πέμπτη
Bonus Θέματα

Πρόβλημα Επιδημιολογίας

Από 100.000 ανθρώπους, οι 51.500 είναι γυναίκες και οι 48.500 είναι άνδρες. Από τις γυναίκες 9.000 έχουν το πρόβλημα υγείας «*φ*» και από τους άνδρες 30.200 έχουν το ίδιο πρόβλημα. Έστω ότι επιλέγουμε ένα άτομο τυχαία.

Πρόβλημα Επιδημιολογίας (*συνέχεια*)

- Έχουμε Ω={*γφ, γμ, αφ, αμ*} ως δειγματικό χώρο με το ***γφ*** να σημαίνει γυναίκα με πρόβλημα υγείας, το ***γμ*** γυναίκα χωρίς πρόβλημα, το ***αφ*** άνδρας με πρόβλημα υγείας και το ***αμ*** άνδρας χωρίς πρόβλημα υγείας.

$\Rightarrow$ P***(γφ)*** = 0,090

P***(γμ)*** = 0,425

P***(αφ)*** = 0,302

P***(αμ)*** = 0,183

Πρόβλημα Επιδημιολογίας (*συνέχεια*)

Έστω A το ενδεχόμενο της επιλογής ενός ανθρώπου με πρόβλημα υγείας και έστω B το ενδεχόμενο επιλογής γυναίκας.

- το $A \cap B$ είναι το ενδεχόμενο επιλογής μιας γυναίκας με πρόβλημα υγείας,
- το $A \cup B$ είναι το ενδεχόμενο επιλογής ενός ανθρώπου με πρόβλημα υγείας ή γυναίκας,
- το B-A είναι το ενδεχόμενο επιλογής μια γυναίκας χωρίς πρόβλημα υγείας

Πρόβλημα Επιδημιολογίας (*συνέχεια*)

- $P(A) = 0{,}090 + 0{,}302 = 0{,}392$
- $P(B) = 0{,}090 + 0{,}425 = 0{,}515$
- $P(A \cap B) = 0{,}090$
- $P(A \cup B) = 0{,}090 + 0{,}425 + 0{,}302 = 0{,}817$
- $P(B\text{-}A) = 0{,}425$

Πρόβλημα Διωνυμικής Κατανομής

- Έστω ότι ρίχνουμε ένα ζάρι 5 φορές και ενδιαφερόμαστε αν το αποτέλεσμα κάθε ρίψης ήταν «1» ή «όχι 1».

  1. Ποια η πιθανότητα να μην έρθει ούτε μια φορά στις 5 προσπάθειες το «1»?

  2. Ποια η πιθανότητα να έρθει ακριβώς τρεις φορές στις 5 προσπάθειες το «1»?

  3. Ποια η πιθανότητα να έρθει τουλάχιστο δύο φορές στις 5 προσπάθειες το «1»?

Απάντηση

- Ας θεωρήσουμε ως επιτυχία (Ε) το αποτέλεσμα της ρίψης να είναι το “1” και ως αποτυχία (Α) το αποτέλεσμα να είναι «όχι 1» (κοινώς το «όχι 1» σημαίνει ότι το αποτέλεσμα μπορεί να είναι 2,3,4,5 ή 6).
- Η πιθανότητα να έρθει «1» όταν ρίχνουμε ένα ζάρι είναι 1/6. Άρα η πιθανότητα επιτυχίας είναι $p=1/6$ και επομένως η πιθανότητα αποτυχίας θα είναι $q=1-p=5/6$.
- Αν με X συμβολίσουμε το συνολικό αριθμό επιτυχιών στις 5 επαναλήψεις του πειράματος, τότε $X \sim \mathrm{B}(5,1/6)$. Έχουμε λοιπόν:

Απάντηση (*συνέχεια*)

1. $P(X=0)=\dfrac{5!}{0!(5-0)!}\left(\dfrac{1}{6}\right)^0\left(\dfrac{5}{6}\right)^{5-0}\cong 0,4019$

2. $P(X=3)=\dfrac{5!}{3!(5-3)!}\left(\dfrac{1}{6}\right)^3\left(\dfrac{5}{6}\right)^{5-3}=\dfrac{5\times4\times3!}{3!2!}\left(\dfrac{1}{6}\right)^3\left(\dfrac{5}{6}\right)^2=10\times\dfrac{5^2}{6^5}\cong 0,03215$

3. $P(X\geq 2)=1-P(X<1)=1-\left[P(X=0)+P(X=1)\right]$ (1)

όμως η πιθανότητα $P(X=0)$ υπολογίστηκε στο πρώτο ερώτημα ενώ για την $P(X=1)$ έχουμε:

$$P(X=1)=\frac{5!}{1!(5-1)!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{5-1}=5\times\frac{5^4}{6^5}\cong 0,4019$$

άρα η σχέση (1) γίνεται:

$$P(X\geq 2)\cong 1-[0,4019+0,4019]=0,1962$$

Πρόβλημα *Poisson* Κατανομής

- Μια υπάλληλος η οποία εισάγει δεδομένα στον Η/Υ κάνει κατά μέσο όρο τρία λάθη ανά σελίδα. Να υπολογιστεί η πιθανότητα σε τυχαία επιλεγμένη σελίδα να βρεθούν δύο λάθη.
- Εστω X ο αριθμός των λαθών ανά σελίδα. Σύμφωνα με τα δεδομένα $X \sim \mathrm{P}(\lambda=3)$.
- Ζητάμε την πιθανότητα $\mathrm{P}(X=2)$

$$P(X=2)=\frac{e^{-3}3^2}{2!}=\frac{\frac{1}{e^3}*3^2}{1*2}=\frac{(0{,}050)*9}{2}=0{,}225$$

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Ασκήσεις Πιθανοτήτων

1. Δίνονται $P(A')=0,3$, $P(B)=0,4$ και $P(AB')=0,5$. Να υπολογίσετε τις πιθανότητες $P(A)$, $P(AB)$, $P(A \cup B)$.
2. Ένα παλιό τρακτέρ χαλάει 65% από βλάβη μηχανής, 20% από αμέλεια του οδηγού, 5% από βλάβη μηχανής και αμέλεια οδηγού και από άλλες αιτίες. Ποια είναι η πιθανότητα να χαλάσει το τρακτέρ «μόνο από βλάβη μηχανής ή μόνο από αμέλεια οδηγού»;
3. Ρίχνουμε δύο ζάρια μια φορά. Ποιος είναι ο Δειγματοχώρος; Ποια είναι τα γεγονότα: α) Το άθροισμα των ενδείξεων είναι διαιρετό διά 4, β) Οι ενδείξεις των ζαριών είναι ίδιες, γ) οι ενδείξεις των ζαριών διαφέρουν το πολύ κατά 3. Ποιες είναι οι πιθανότητες των ενδεχομένων που ορίστηκαν στα α), β) και γ)

Άσκηση Συμπλήρωσης (1)

Ο αριθμός των απασχολούμενων υπαλλήλων μηχανογραφημένου λογιστηρίου ενός δείγματος 40 μικρομεσαίων επιχειρήσεων δίνεται στον παρακάτω πίνακα.

Αριθμός Υπαλλήλων	Αριθμός Επιχειρήσεων (Απόλυτες Συχνότητες, f_i)	Σχετικές Συχνότητες, P_i	Αθροιστικές Συχνότητες, $Φ_i$	Αθροιστικές Σχετικές Συχνότητες, F_i
1	3			
2	3			
3	7			
4	12			
5	8			
6	6			
7	1			
	Σύνολο=40			

1. Να συμπληρωθούν τα στοιχεία που λείπουν.
2. Πόσες επιχειρήσεις απασχολούν έως και 4 υπαλλήλους;
3. Τι ποσοστό των επιχειρήσεων απασχολεί έως και 3 υπαλλήλους;
4. Τι ποσοστό των επιχειρήσεων απασχολεί ακριβώς 4 υπαλλήλους;
5. Πόσες επιχειρήσεις απασχολούν έως και 6 υπαλλήλους;
6. Να υπολογιστεί ο μέσος αριθμός εργαζομένων στις επιχειρήσεις.

Άσκηση Συμπλήρωσης (2)

Ένα τυχαίο δείγμα 200 υπαλλήλων ταξινομήθηκε ανάλογα με τη θέση του υπαλλήλου μέσα στην επιχείρηση και τη συμπεριφορά του σχετικά με την έκπτωση που ζητούν οι πελάτες. Τα αποτελέσματα δίνονται στον παρακάτω πίνακα συμπτώσεων.

	Κάνει έκπτωση	Δεν κάνει έκπτωση	Είναι αναποφάσιστος	Σύνολα
Πωλητής	30	15	15	60
Τμηματάρχης	?	50	10	100
Διαχειριστής	10	?	5	?
Σύνολα	80	90	?	?

1. Να συμπληρωθούν τα στοιχεία που λείπουν (δηλώνονται με ?).
2. Ποιο είναι το μέγεθος του δείγματος;
3. Ποιο είναι το ποσοστό % των Πωλητών που κάνει έκπτωση;
4. Ποιο είναι το ποσοστό % των Αναποφάσιστων που είναι Διαχειριστές;
5. Ποιο είναι το ποσοστό % των Υπαλλήλων που δεν κάνει έκπτωση;
6. Ποιο είναι το ποσοστό % των Πωλητών που δεν κάνουν έκπτωση στο σύνολο του δείγματος;

Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Γεώργιος Μενεξές. «Στατιστική. Προσομοίωση Εξέτασης στο μάθημα της Στατιστικής (Λυμένα και Άλυτα Θέματα)». Έκδοση: 1.0. Θεσσαλονίκη 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: https://opencourses.auth.gr/courses/OCRS484/

Σημείωμα Αδειοδότησης

ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΧΤΑ
ΑΚΑΔΗΜΑΙΚΑ
ΜΑΘΗΜΑΤΑ

Τέλος Ενότητας

Επεξεργασία: Μαρία Αλεμπάκη

Θεσσαλονίκη, Εαρινό Εξάμηνο 2014-2015

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης