**ΑΣΚΗΣΗ 5η**

**χ2 Έλεγχος Ανεξαρτησίας**

Ένα εργοστάσιο λειτουργεί σε 24ωρη βάση, με τρείς βάρδιες. Ο Υπεύθυνος ποιότητας του εργοστασίου παρατήρησε ότι το τελευταίο διάστημα παράγεται σημαντικός αριθμός ελαττωματικών προϊόντων. Για να ελέγξει ως πιθανή αιτία τους εργαζόμενους σε διαφορετικές βάρδιες, πήρε δείγμα από κάθε μια και κατέγραψε το πλήθος των ελαττωματικών προϊόντων. Τα αποτελέσματα ανά βάρδια παρουσιάζονται στον επόμενο πίνακα.

|  |
| --- |
| **Βάρδια \* Χαρακτηρισμός Crosstabulation** |
|  | **Χαρακτηρισμός** | **Total** |
| **Καλό** | **Ελαττωματικό** |
| **Βάρδια** | **1η βάρδια** | Count | 230 | 20 | **250** |
| Expected Count | 237,5 | 12,5 | **250,0** |
| % within Βάρδια |  | α= | **100,0%** |
| % within Χαρακτηρισμός | 42,1% | 33,3% | **41,7%** |
| % of Total | 40,0% | 1,7% | **41,7%** |
| **2η βάρδια** | Count | 191 | 9 | **200** |
| Expected Count | β= |  | **200,0** |
| % within Βάρδια | 95,5% | 4,5% | **100,0%** |
| % within Χαρακτηρισμός |  | γ= | **33,3%** |
| % of Total | 31,8% | 1,5% | **33,3%** |
| **3η βάρδια** | Count | 139 | 11 | **150** |
| Expected Count | 142,5 | 7,5 | **150,0** |
| % within Βάρδια | 92,7% | 7,3% | **100,0%** |
| % within Χαρακτηρισμός | 24,4% | 36,7% | **25,0%** |
| % of Total | δ= | 1,8% | **25,0%** |
|  **Total** | **Count** | **560** | **40** | **600** |
| **Expected Count** | **570,0** | **30,0** | **600,0** |
| **% within Βάρδια** | **95,0%** | **5,0%** | **100,0%** |
| **% within Χαρακτηρισμός** | **100,0%** | **100,0%** | **100,0%** |
| **% of Total** | **95,0%** | **5,0%** | **100,0%** |

1. Ποια μέθοδος χρησιμοποιήθηκε και γιατί;
2. Ποιες είναι οι υποθέσεις της μεθόδου;
3. Να συμπληρωθούν στα γραμμοσκιασμένα κελιά οι τιμές που λείπουν

(κελιά α, β, γ, δ).

1. Από το σύνολο των ***παραχθέντων προϊόντων της πρώτης βάρδιας***, ποιο ποσοστό είναι ελλαττωματικό*;*
2. Από το σύνολο των ***καλών προϊόντων*** (μη ελλαττωματικών) ποιο ποσοστό το παρήγαγε η τρίτη βάρδια;
3. Από το ***γενικό σύνολο των παραχθέντων προϊόντων*** ποιο ποσοστό παρήχθη από την δεύτερη βάρδια και είναι ελλαττωματικό;
4. Σε επίπεδο σημαντικότητας 5% μπορούμε να ισχυριστούμε ότι οι μεταβλητές «βάρδια» και «χαρακτηρισμός» είναι ανεξάρτητες (εμπεριστατωμένη αιτιολόγηση);
5. Είναι αξιόπιστα τα αποτελέσματα;

|  |
| --- |
| **Chi-Square Tests** |
|  | Value | df | Asymp. Sig. (2-sided) |
| Pearson Chi-Square | 2,351a | 2 | 0,039 |
| Likelihood Ratio | 2,185 | 2 | 0,033 |
| N of Valid Cases | 600 |  |  |
| a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,0. |

**Λύση**

1. Χρησιμοποιήθηκε ο χ2 έλεγχος ανεξαρτησίας διότι οι μεταβλητές είναι ποιοτικές.
2. Μηδενική υπόθεση: Οι μεταβλητές είναι ανεξάρτητες Εναλλακτική υπόθεση: Οι μεταβλητές δεν είναι ανεξάρτητες (εξαρτημένες)
3. α=(20/250)\*100=8% β=(200\*560)/600=186,67 γ=(9/40)\*100=22,5% δ=(139/600)\*100=23,17%
4. (20/250)\*100=8%
5. (139/560)\*100=24,82%
6. (9/600)\*100=1,5%
7. Επειδή η τιμή 2,351 του Pearson Chi-Square είναι στατιστικά σημαντική (Sig.=0,039<0,05) απορρίπτουμε την μηδενική υπόθεση και αποδεχόμαστε την εναλλακτική. Δηλαδή οι μεταβλητές είναι εξαρτημένες. Άρα πιθανή αιτία παραγωγής πολλών ελλαττωματικών είναι η βάρδια.
8. Επειδή κανένα κελί δεν έχει αναμενόμενη τιμή μικρότερη του 5 τα αποτελέσματα είναι αξιόπιστα. (αν τουλάχιστον το 20% των κελιών έχουν αναμενόμενη τιμή μικρότερη του 5 τα αποτελέσματα δεν είναι αξιόπιστα).

**ΑΣΚΗΣΗ 6η**

**ANOVA**

Διαφέρουν οι ειδικότητες των ιατρών ως προς το χρόνο που αφιερώνουν στη φροντίδα των ασθενών τους; Μια έρευνα επέλεξε ένα τυχαίο δείγμα ιατρών διαφόρων ειδικοτήτων και κατέγραψε τον αριθμό των ωρών που αφιέρωσε ο καθένας στη φροντίδα των ασθενών του στη διάρκεια μιας εβδομάδας. Εφαρμόστηκε η κατάλληλη μέθοδος και τα αποτελέσματα εμφανίζονται στον επόμενο πίνακα.

**Πίνακας ANOVA**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Πηγή Μεταβλητότητας** | **Άθροισμα Τετραγώνων-SS** | **Βαθμοί Ελευθερίας-df** | **Μέσο Τετράγωνο-MS** | ***f*** | ***Sig.*** |
| Παράγοντας (TR) | 1406,390 | 4  | 351,598 | 5,169 | 0,035 |
| Σφάλματα (E) | 11903,578 | 175  | 68,020 |  |  |
| **Ολική** | 13309,968 | 179 |  |  |  |

1. Ποια είναι η κατάλληλη μέθοδος η οποία εφαρμόστηκε και γιατί;
2. Ποιες είναι οι υποθέσεις της μεθόδου;
3. Να προσδιορίσετε το πλήθος των ειδικοτήτων
4. Το πλήθος των ιατρών και
5. Να ελέγξετε, σε επίπεδο 2% αν υπάρχει διαφορά στις ειδικότητες ως προς το χρόνο που αφιερώνουν στη φροντίδα των ασθενών τους.

**Λύση**

1. Εφαρμόστηκε η μέθοδος της Ανάλυσης Διακύμανσης γιατί η μια μεταβλητή είναι ποιοτική (ειδικότητα) και η άλλη ποσοτική (χρόνος)
2. Μηδενική υπόθεση: Οι χρόνοι είναι ίσοι ανεξαρτήτως ειδικότητας Εναλλακτική υπόθεση: Οι χρόνοι διαφέρουν
3. Κ-1=4 άρα κ=5 ειδικότητες
4. Ν-1=179 άρα Ν=180 ιατροί
5. Επειδή η τιμή 5,169 της κατανομής F είναι στατιστικά μη σημαντική (Sig.=0,035>0,02) η μηδενική υπόθεση γίνεται αποδεκτή και συνεπώς ο χρόνος φροντίδας είναι ίδιος ανεξαρτήτως ειδικότητα.

**ΑΣΚΗΣΗ 7η**

**Παλινδρόμηση**

Πραγματοποιήθηκε ανάλυση παλινδρόμησης προκειμένου να προσδιοριστεί αν και κατά πόσο οι παράγοντες «καθηγητές», «αξιολόγηση και ανατροφοδότηση», «πρόγραμμα σπουδών», «οργάνωση και διοίκηση», «πόροι και τεχνολογία» επηρεάζουν την ικανοποίηση των φοιτητών από τις σπουδές τους. Τα αποτελέσματα παρουσιάζονται στους επόμενους πίνακες.

|  |
| --- |
| **Model Summaryb** |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
| 1 | ,810a | ,656 | ,651 | ,52318 | 1,991 |
| **ANOVAa** |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 186,025 | 5 | 37,205 | 135,927 | ,000b |
| Residual | 97,442 | 356 | ,274 |  |  |
| Total | 283,467 | 361 |  |  |  |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | Collinearity Statistics |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | -,746 | ,154 |  | -4,834 | ,000 |  |  |
| καθηγητές | ,421 | ,050 | ,331 | 8,339 | ,000 | ,612 | 1,634 |
| αξιολόγηση και ανατροφοδότηση | ,050 | ,035 | ,053 | 1,427 | ,154 | ,693 | 1,444 |
| πρόγραμμα σπουδών | ,494 | ,042 | ,440 | 11,618 | ,000 | ,674 | 1,484 |
| οργάνωση και διοίκηση | ,296 | ,041 | ,280 | 7,215 | ,000 | ,643 | 1,555 |
| πόροι και τεχνολογία | -,118 | ,046 | -,106 | -2,534 | ,012 | ,551 | 1,816 |

**Ζητούμενα (**εμπεριστατωμένη αιτιολόγηση σε κάθε απάντηση)

1. Να προσδιορίσετε την εξαρτημένη και τις ανεξάρτητες μεταβλητές
2. Να ελέγξετε αν τα δεδομένα είναι κατάλληλα για ανάλυση παλινδρόμησης
3. Να προσδιορίσετε το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τις μεταβολές των ανεξάρτητων μεταβλητών
4. Να ελέγξετε αν υπάρχει πρόβλημα αυτοσυσχέτισης και συγραμμικότητας. Σε περίπτωση θετικής απάντησης ποια ενέργεια απαιτείται;
5. Να ελέγξετε αν οι συντελεστές παλινδρόμησης είναι στατιστικά σημαντικοί. Σε περίπτωση αρνητικής απάντησης ποια ενέργεια απαιτείται;
6. Να προσδιορίσετε την ανεξάρτητη μεταβλητή η οποία επηρεάζει περισσότερο την εξαρτημένη μεταβλητή.
7. Να κατασκευάσετε το παλινδρομικό μοντέλο.

**Λύση**

1. Ανεξάρτητες: «καθηγητές», «αξιολόγηση και ανατροφοδότηση», «πρόγραμμα σπουδών», «οργάνωση και διοίκηση», «πόροι και τεχνολογία» Εξαρτημένη: «ικανοποίηση»
2. Είναι καθώς η τιμή 135,927 της κατανομής F είναι στατιστικά σημαντική (Sig.=0,000<0,05)
3. Το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τις μεταβολές των ανεξάρτητων μεταβλητών είναι 65,6%. Το υπόλοιπο 34,4% οφείλεται σε τυχαίους και ανερμήνευτους παράγοντες.
4. Πρόβλημα αυτοσυσχέτισης δεν υπάρχει καθώς ο δείκτης Durbin-Watson με τιμή 1,991 βρίσκεται πού κοντά στο 2. Επίσης δεν υπάρχει πρόβλημα συγραμμικότητας για καμία μεταβλητή καθώς όλοι οι δείκτες VIF είναι μικρότεροι του 5. Αν υπήρχε πρόβλημα θα έπρεπε η μεταβλητή να αποβληθεί από το μοντέλο και να δημιουργηθεί ένα νέο με μια μεταβλητή λιγότερη.
5. Οι συντελεστές παλινδρόμησης των μεταβλητών «καθηγητές», «πρόγραμμα σπουδών», «οργάνωση και διοίκηση», «πόροι και τεχνολογία» είναι στατιστικά σημαντικοί καθώς σε κάθε περίπτωση οι t τιμές είναι στατιστικά σημαντικές (sig.<0,05). Αντίθετα ο συντελεστής της μεταβλητής «αξιολόγηση και ανατροφοδότηση» δεν είναι στατιστικά σημαντικός (sig.=0,154>0,05). Θα πρέπει λοιπόν να αποβληθεί και να δημιουργηθεί ένα νέο μοντέλο χωρίς την συγκεκριμένη μεταβλητή.
6. Η ανεξάρτητη μεταβλητή που επηρεάζει περισσότερο την εξαρτημένη μεταβλητή είναι η «πρόγραμμα σπουδών» γιατί ο συντελεστής Β (0,440) είναι μεγαλύτερος κατ’ απόλυτο τιμή από τους υπόλοιπους.
7. Ικανοποίηση=-0,746+0,421\*καθηγητές+0,050\*αξιολόγηση και ανατροφοδότηση+0,496\*πρόγραμμα σπουδών+0,296\*οργάνωση και διοίκηση-0,118\*πόροι και τεχνολογία