**Άσκηση 1**

Προκειμένου να ελεγχθεί αν ο «τρέχον μισθός» των 474 υπαλλήλων μιας επιχείρησης εξαρτάται από τον «αρχικό μισθό», το «επίπεδο εκπαίδευσης», την «προϋπηρεσία στην επιχείρηση» και την «προηγούμενη εμπειρία» πραγματοποιήθηκε Ανάλυση Παλινδρόμησης (Regression Analysis). Τα αποτελέσματα παρουσιάζονται στους παρακάτω πίνακες:

|  |
| --- |
| **ANOVAb** |
| **Model** | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
| **1** | **Regression** | 1,118 | 4 | 0,2794 | 502,06648 | 0,000 |
| **Residual** | 0,261 | 469 | 0,0005565 |  |  |
| **Total** | 1,379 | 473 |  |  |  |
| **Model Summaryb** |
| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** | **Durbin-Watson** |
| **1** | 0,900 | 0,810 | 0,809 | 7,465.139 | 1,921 |
|  |
| **Coefficientsa** |
| **Model** | **Unstandardized Coefficients** | **Standardized Coefficients** | **t** | **Sig.** | **Collinearity Statistics** |
| **B** | **Std. Error** | **Beta** | **Tolerance** | **VIF** |
|  | (Constant) | -16149,671 | 3255,470 |  | -4,961 | 0,000 |  |  |
| Αρχικός Μισθός | 1,768 | 0,059 | 0,815 | 30,111 | 0,000 | 0,551 | 1,814 |
| Επίπεδο εκπαίδευσης (χρόνια) | 669,914 | 165,596 | 0,113 | 4,045 | 0,000 | 0,516 | 1,937 |
| Προϋπηρεσία στην επιχείρηση (μήνες) | 161,486 | 94,246 | 0,095 | 1,715 | 0,087 | 0,112 | 8,196 |
| Προηγούμενη εμπειρία (μήνες) | -17,303 | 3,528 | -0,106 | -4,904 | 0,000 | 0,865 | 1,156 |
|  |

1) Είναι τα δεδομένα κατάλληλα για ανάλυση παλινδρόμησης και γιατί;

2) Ποιες είναι οι ανεξάρτητες και ποια η εξαρτημένη μεταβλητή;

3) Ποιο είναι το ποσοστό της διακύμανσης (variance explained) της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τις ανεξάρτητες μεταβλητές;

4) Υπάρχει ένδειξη συγγραμμικότητας (collinearity) ή/και αυτοσυσχέτισης (autocorrelation); Σε περίπτωση θετικής απάντησης ποια ενέργεια απαιτείται;

5) Είναι όλοι οι συντελεστές παλινδρόμησης (regression coefficients) σημαντικοί; Σε περίπτωση αρνητικής απάντησης ποια ενέργεια απαιτείται;

6) Ποια από τις ανεξάρτητες μεταβλητές επιδρά εντονότερα στην εξαρτημένη μεταβλητή;

7) Κατασκευάστε το παλινδρομικό μοντέλο.

**Απαντήσεις**

1. Από τον πίνακα ANOVA παρατηρούμε ότι το Sig. της F τιμής είναι <0,05 και συνεπώς τα δεδομένα είναι κατάλληλα για ανάλυση παλινδρόμησης καθώς υπάρχουν ικανοποιητικές συσχετίσεις μεταξύ των μεταβλητών.
2. Εξαρτημένη είναι η μεταβλητή «τρέχον μισθός» και ανεξάρτητες οι μεταβλητές ο «μισθός πρόσληψης», το «επίπεδο εκπαίδευσης», η «προϋπηρεσία στην επιχείρηση» και η «προηγούμενη εμπειρία».
3. Από τον πίνακα Model Summary η τιμή του R- Square είναι 0,810. Αυτό σημαίνει ότι το 81% της διακύμανσης των τιμών της εξαρτημένης μεταβλητής ερμηνεύεται από τις μεταβολές των ανεξάρτητων μεταβλητών.
4. Υπάρχει ένδειξη συγγραμμικότητας για την μεταβλητή «Προϋπηρεσία στην επιχείρηση» καθώς η τιμή του δείκτη V.I.F είναι μεγαλύτερη του ορίου 5 (Πίνακας Model Summary). Απαιτείται εξάλειψη της συγκεκριμένης μεταβλητής και δημιουργία εκ νέου του παλινδρομικού μοντέλου. Επίσης δεν έχουμε σοβαρή ένδειξη αυτοσυσχέτισης καθώς ο δείκτης Durbin Watson έχει τιμή περίπου 2 (1,921) (Πίνακας Coefficients).
5. Οι συντελεστές παλινδρόμησης τριών ανεξάρτητων μεταβλητών είναι στατιστικά σημαντικοί καθώς το Sig. των αντιστοίχων t- τιμών είναι <0,0. Μόνο ο συντελεστής παλινδρόμησης της μεταβλητής «Προϋπηρεσία στην επιχείρηση» είναι στατιστικά ασήμαντος. Απαιτείται εξάλειψη της συγκεκριμένης μεταβλητής και δημιουργία εκ νέου του παλινδρομικού μοντέλου (πίνακας Coefficients).
6. Η ανεξάρτητη μεταβλητή που επιδρά περισσότερο στην εξαρτημένη μεταβλητή είναι η

«Αρχικός Μισθός» καθώς ο Standardized Coefficient Beta (0,815) είναι κατά απόλυτη τιμή μεγαλύτερος από τους αντίστοιχους των άλλων μεταβλητών (πίνακας Coefficients).

1. Τρέχον μισθός = -16.149,671+1,768\* Αρχικός μισθός+ 669,914\*Επίπεδο εκπαίδευσης+ 161,486\* Προϋπηρεσία στην επιχείρηση -17,303\* Προηγούμενη εμπειρία (πίνακας Coefficients).

**Άσκηση 2**

Μια εταιρεία η οποία κατασκευάζει πέτρινες επενδύσεις τοίχων προσπαθεί να αναλύσει τις μεταβλητές που καθορίζουν τη ζήτηση του προϊόντος. Οι πέτρινες επενδύσεις χρησιμοποιούνται σε διαμερίσματα και γραφεία. Για τους 24 τελευταίους μήνες κατέγραψε τις μηνιαίες πωλήσεις, τον μηνιαίο αριθμό οικοδομικών αδειών, το επιτόκιο τραπεζικού δανεισμού, το ποσοστό των ανοίκιαστων διαμερισμάτων και το ποσοστό των ανοίκιαστων γραφείων.

Με βάση τα παραπάνω στοιχεία η εταιρεία πραγματοποίησε Ανάλυση Παλινδρόμησης τα αποτελέσματα της οποίας παρουσιάζονται στους παρακάτω πίνακες.

|  |
| --- |
| **Model Summaryb** |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
| 1 | ,945a | ,894 | ,871 | 40,132 | 1,955 |
| **ANOVAa** |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 256793,400 | 4 | 64198,350 | 39,860 | ,000b |
| Residual | 30601,558 | 19 | 1610,608 |  |  |
| Total | 287394,958 | 23 |  |  |  |

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | Collinearity Statistics |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | -111,828 | 134,343 |  | -,832 | ,000 |  |  |
| Οικοδομικές άδειες | 4,763 | ,395 | ,917 | 12,057 | ,000 | ,969 | 1,032 |
| Επιτόκιο δανεισμού | 16,989 | 15,159 | ,484 | 1,121 | ,027 | ,997 | 1,003 |
| Ανοίκιαστα διαμ. | -10,528 | 6,394 | -,225 | -1,646 | ,011 | ,973 | 1,027 |
| Ανοίκιαστα γραφεία | 1,308 | 2,791 | ,035 | ,469 | ,645 | ,186 | 5,376 |

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 συγγραμμικότητας (σε θετική περίπτωση τι ενέργεια απαιτείται;)
5. Να ελέγξετε αν οι συντελεστές παλινδρόμησης είναι στατιστικά σημαντικοί (σε αντίθετη περίπτωση τι ενέργεια απαιτείται;)
6. Να προσδιορίσετε την ανεξάρτητη μεταβλητή η οποία επηρεάζει περισσότερο την εξαρτημένη μεταβλητή
7. Να προβλέψετε την μηνιαία ζήτηση αν ο αριθμός των νέων αδειών είναι 50, το επιτόκιο 9%, το ποσοστό των ανοίκιαστων διαμερισμάτων 3,6% και των ανοίκιαστων γραφείων 14,3% (υποθέτοντας ότι δεν θα χρειαστεί να αφαιρεθεί καμία ανεξάρτητη μεταβλητή).

**Απαντήσεις:**

1. Εξαρτημένη μεταβλητή:μηνιαίες πωλήσεις

 Ανεξάρτητες μεταβλητές: μηνιαίος αριθμός οικοδομικών αδειών, επιτόκιο τραπεζικού δανεισμού, ποσοστό ανοίκιαστων διαμερισμάτων και ποσοστό ανοίκιαστων γραφείων.

1. Από τον πίνακα ANOVA παρατηρούμε ότι το στατιστικό F= 39,860 είναι στατιστικά σημαντικό καθώς το Sig. = 0,000<0,05. Αυτό δείχνει ότι υπάρχουν ικανοποιητικές συσχετίσεις μεταξύ των μεταβλητών και συνεπώς τα δεδομένα είναι κατάλληλα για ανάλυση παλινδρόμησης.
2. Το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τις μεταβολές των ανεξάρτητων μεταβλητών είναι 89,4%. Το υπόλοιπο 10,6% της διακύμανσης ερμηνεύεται από τυχαίους και ανερμήνευτους παράγοντες.
3. Ο δείκτης Durbin-Watson παίρνει τιμές στο διάστημα [0, 4] και τιμές κοντά στο 2 δείχνουν ότι δεν υπάρχει πρόβλημα αυτοσυσχέτισης. Στην συγκεκριμένη περίπτωση η τιμή είναι 1,955 κάτι που δείχνει ότι δεν υπάρχει πρόβλημα αυτοσυσχέτισης καθώς είναι κοντά στην τιμή 2.

Πρόβλημα συγγραμμικότητας υπάρχει όταν η τιμή του δείκτη V.I.F>5. Στην συγκεκριμένη περίπτωση μόνο η ανεξάρτητη μεταβλητή «Ανοίκιαστα γραφεία» με τιμή V.I.F = 5,376 έχει πρόβλημα συγγραμμικότητας. Απαιτείται η αφαίρεση της μεταβλητής αυτής και η επαναδημιουργία του μοντέλου χωρίς την συγκεκριμένη μεταβλητή.

1. Οι συντελεστές παλινδρόμησης είναι στατιστικά σημαντικοί όταν του Sig. των τιμών t είναι μικρότερο του 0,05. Από τις ανεξάρτητες μεταβλητές του μοντέλου η μεταβλητή «Ανοίκιαστα γραφεία» έχει πρόβλημα καθώς η τιμή t =0,469 είναι στατιστικά ασήμαντη καθώς το Sig. =0,645>0,05. Απαιτείται η αφαίρεση της μεταβλητής αυτής και η επαναδημιουργία του μοντέλου χωρίς την συγκεκριμένη μεταβλητή.
2. H ανεξάρτητη μεταβλητή η οποία επηρεάζει περισσότερο την εξαρτημένη μεταβλητή είναι αυτή με την μεγαλύτερη τιμή Βeta (κατ’ απόλυτο τιμή). Έτσι, η μεταβλητή «Οικοδομικές άδειες» με τιμή Βeta= 0,917 είναι αυτή που επηρεάζει περισσότερο την εξαρτημένη.
3. Η εκτίμηση των μηνιαίων πωλήσεων είναι:

Μηνιαίες Πωλήσεις = -111,828+4,763\*X1+16,986\*X2-10,528\*X3+1,308\*X4

**Άσκηση 3**

The following table gives the annual return (%), the safety rating (0= riskiest, 10=safest) the annual expense ratio (%) and the Country for 20 foreign companies.

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **Safety Rating** | **Annual Expense Ratio %** | **Annual Return %** |
| 1 | 7,1 | 1,59 | 49 |
| 2 | 7,2 | 1,35 | 52 |
| 3 | 6,8 | 1,68 | 89 |
| 4 | 7,1 | 1,56 | 58 |
| 5 | 6,2 | 2,16 | 131 |
| 6 | 7,4 | 1,80 | 59 |
| 7 | 6,5 | 1,88 | 99 |
| 8 | 7,0 | 0,90 | 53 |
| 9 | 6,9 | 1,79 | 77 |
| 10 | 7,2 | 1,49 | 54 |
| 11 | 7,1 | 1,05 | 57 |
| 12 | 7,7 | 1,25 | 61 |
| 13 | 7,0 | 1,83 | 88 |
| 14 | 7,0 | 1,94 | 122 |
| 15 | 7,2 | 1,09 | 71 |
| 16 | 6,9 | 1,50 | 51 |
| 17 | 7,0 | 1,28 | 60 |
| 18 | 7,1 | 1,65 | 50 |
| 19 | 6,5 | 1,61 | 93 |
| 20 | 7,0 | 1,50 | 47 |

Using these data, an estimated regression equation relating the ***annual return*** to the ***safety rating*** and the ***annual expense ratio*** was developed.

**The results are presented in the tables 1, 2 and 3.**

|  |
| --- |
| **Table 1.** Model Summaryb |
| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** | **Durbin-Watson** |
| 1 | ,763a | ,582 | ,533 | 16,977 | 1,962 |
| a. Predictors: (Constant), Annual Expense Ratio %, Safety Rating |
| b. Dependent Variable: Annual Return % |

|  |
| --- |
| **Table 2.** ANOVAa |
| **Model** | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
| 1 | Regression | 13646,414 | 2 | 6823,207 | 25,762 | ,000b |
| Residual | 9799,486 | 37 | 264,851 |  |  |
| Total | 23445,900 | 39 |  |  |  |
| a. Dependent Variable: Annual Return % |
| b. Predictors: (Constant), Annual Expense Ratio %, Safety Rating |

|  |
| --- |
| **Table 3.** Coefficientsa |
| **Model** | **Unstandardized Coefficients** | **Standardized Coefficients** | **t** | **Sig.** | **Collinearity Statistics** |
| **B** | **Std. Error** | **Beta** | **Tolerance** | **VIF** |
| 1 | (Constant) | 247,358 | 74,864 |  | 3,304 | ,002 |  |  |
| Safety Rating | -32,845 | 9,456 | -,430 | -3,473 | ,001 | ,737 | 1,357 |
| Annual Expense Ratio % | 34,589 | 9,577 | ,447 | 3,611 | ,001 | ,737 | 1,357 |
| 1. Dependent Variable: Annual Return %
 |

1. **Does the regression model predicts the dependent variable significantly well?**

In order to test how well the regression equation fits the data (i.e., predicts the dependent variable) the table 2 (ANOVA) is needed. This table indicates if the regression model predicts the dependent variable significantly well. Since Sig.F = 0,000<0,05, overall the regression model statistically significantly predicts the outcome variable.

1. **What is the percentage of the variance of “Annual Return” that is explained by the variables “Annual Expense Ratio” and “Safety Rating”?** The percentage of the Variance explained by the independent variables is 58,2% (R-square from table 1).
2. **Is there a statistically significant relation between dependent and independent variables?**
3. If the regression coefficient is statistically significant, then we can conclude that there is a significant relation between dependent and independent variable. From table 3 we can see that the regression coefficients of the independent variables are statistically significant as the Sig. of t- statistic is less than 5%.
4. **Is there problem of Autocorrelation?**

Durbin –Watson statistic with a value of 1,962 (very close to 2) indicates that there is no problem of autocorrelation (table 1).

1. **Is there problem of Collinearity?**

We have problem of collinearity if the V.I.F index is greater than 5. In table 3 are presented the V.I.F for the two independent variables and they are 1,357 for both. So, there is no proble of collinearity.

1. **Estimate the annual return for a company that has a safety rating of 7,5 and annual expense ratio of 2.**

Annual Return= 247,358 -32,845\*7,5 +34,589\*2=70,1985

**Άσκηση 4**

In order to test the impact of “Engine Displacement”, “Horsepower” and “Vehicle’s Weight” in “Acceleration” of a car, we have performed a Regression analysis. The tables below contain the results from the Regression Analysis.

| **Model Summaryb** |
| --- |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
| 1 | ,799a | ,639 | ,636 | 1,702 | 1,914 |
|  |
| **ANOVAb** |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 2030,521 | 3 | 676,840 | 233,521 | 0,000a |
| Residual | 1147,773 | 396 | 2,898 |  |  |
| Total | 3178,294 | 399 |  |  |  |

| **Coefficientsa** |
| --- |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | Collinearity Statistics |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | 17,233 | ,480 |  | 35,885 | ,000 |  |  |
| Engine Displacement | -,009 | ,003 | -,345 | -3,506 | ,001 | ,256 | 3,906 |
| Horsepower | -,088 | ,005 | -1,195 | -17,324 | ,000 | ,192 | 5,222 |
| Vehicle Weight  | ,003 | ,000 | ,935 | 10,994 | ,000 | ,349 | 2,865 |

**Questions:**

1. Are the data appropriate for Regression Analysis? why?
2. Which are the dependent and independent variables?
3. Which is the percentage of the variance explained by the independent variables?
4. Is there evidence for collinearity and autocorrelation?
5. Are regression coefficients significant?
6. Which of the independent variables affect most the dependent variable?
7. Can you construct the regression model?

**Άσκηση** **5**

Πραγματοποιήθηκε έρευνα σε 285 εργαζόμενους συγκεκριμένης επιχείρησης, για να διαπιστωθεί αν και κατά πόσο η «ικανοποίηση» των εργαζομένων εξαρτάται από το επίπεδο των «αποδοχών», τις «προοπτικές εξέλιξης» και το επίπεδο «συνεργασίας με τους προϊσταμένους». Από την ανάλυση παλινδρόμησης η οποία πραγματοποιήθηκε προέκυψαν οι επόμενοι πίνακες.

|  |
| --- |
| **Model Summaryb** |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
| 1 | 0,723a | 0,523 | 0,518 | ,51149 | 1,996 |

|  |
| --- |
| **ANOVAa** |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 80,749 | 3 | 26,916 | 102,883 | 0,000b |
| Residual | 73,515 | 281 | 0,262 |  |  |
| Total | 154,264 | 284 |  |  |  |

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | Collinearity Statistics |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | 1,115 | ,168 |  | 6,615 | ,000 |  |  |
| Αποδοχές | 0,475 | ,053 | 0,482 | 8,896 | ,000 | 0,577 | 1,734 |
| Προοπτικές | 0,212 | ,051 | 0,227 | 4,197 | ,000 | 0,582 | 1,719 |
| Συνεργασία | 0,117 | ,052 | 0,121 | 2,253 | ,025 | 0,587 | 1,704 |

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
5. Να ελέγξετε αν οι συντελεστές παλινδρόμησης είναι στατιστικά σημαντικοί
6. Να προσδιορίσετε την ανεξάρτητη μεταβλητή η οποία επηρεάζει περισσότερο την εξαρτημένη μεταβλητή
7. Να κατασκευάσετε το παλινδρομικό μοντέλο.