ΖΗΤΗΜΑ 1 (Βαθμ. 2.5)
(a) Στο γενικό γραμμικό μοντέλο \( y = \beta_0 + \beta_1 x \), \( e \sim N(0, \sigma^2 I_n) \) ισχύει ότι η e.e.t. \( \hat{\beta}_1 = (X'X)^{-1}X'y - N_p(\beta_1, \sigma^2(X'X)^{-1}) \), και \( \text{SSE} = \sum \frac{(y_i - \hat{y}_i)^2}{\sigma^2} \), όπου \( X \) ο πίνακας σχεδιασμού και SSE το άδροσκομ τετραγώνων λάθους σφάλματος. Διδομένου ότι \( \hat{\beta}_1 \) και \( \text{SSE} \) είναι ανεξάρτητα, βρείτε ένα γ-διάστημα εμπιστοσύνης για τον αργιόστο συντελεστή \( \beta_1 \) μιας μεταβλητής \( x_1 \) του μοντέλου.

(β) Εστ το \( y = \beta_0 + \beta_1 x + \varepsilon \), χωρίς επεξεργαστικές μεταβλητές, τότε με βάση το δ. \( y_1, y_2, \ldots, y_n \), ισευριστών παρατηρήσεων από την κατανομή αυτή, δείξτε με τη μεθοδο μέγιστης πιθανολόγησης ότι \( \hat{\beta}_0 = \bar{y} \) και κατά συνέπεια SSE=SSE.

ΖΗΤΗΜΑ 2 (Βαθμ. 2.5)

Εστι το γενικό γραμμικό μοντέλο \( y = \beta_0 + \varepsilon \).

(i) Δείξτε ότι η άδροσκομ τετραγώνων λάθους παλινδρόμησης είναι \( \text{SSR}=y'(H-H_y)\bar{y} \), όπου \( H = X(X'X)^{-1}X' \) ο πίνακας προβελτίσης και \( H_y \) ο πίνακας του σημείου διά της στοιχεία είναι 1.

(ii) Δώστε τον σχεδιασμό του συντελεστή προοιμοδομού \( R^2 \). Τι εκφράζεται;

(iii) Γράψτε το διαγνωστικό συντελεστή συγκέντρωσης \( r_{y,2} \) μεταξύ των παρατηρήσεων \( y \) και \( \hat{y} \). Όταν η σταθερά \( \beta_0 \) περιλαμβάνεται στο μοντέλο παλινδρόμησης, ισχύει \( \sum_{i=1}^{n} \hat{y}_i = 0 \). Στην περίπτωση αυτή δείχτε ότι \( r_{y,2} = R^2 \).

(iv) Εστι κ ο αριθμός των επεξεργαστικών μεταβλητών. Δείχτε ότι η ελαχιστοποίηση για τον έλεγχο \( H_0: \beta_1=\beta_2=\ldots=\beta_k=0 \) είναι \( H_0: \eta \) η συντελεστής \( \hat{\beta}_j = 0 \), γράφεται ως \( F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \).

ΖΗΤΗΜΑ 3 (Βαθμ. 2.5)

(a) Εστι το γενικό γραμμικό μοντέλο \( E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \), το πρώτος όρος περιλαμβάνει τη σταθερά \( \beta_0 \) και οι \( \beta_1 \) μεταβλητές, το δεύτερος \( \beta_2 \) μεταβλητής. Δείχτε ότι η e.e.t. \( \hat{\beta}_2 \) είναι αμεροληπτική εκμήνωση της \( \hat{\beta}_1 \) στην περίπτωση που ο δεύτερος όρος \( x_2 \) παραλείπεται από το μοντέλο, ενώ χρειάζεται.

(β) Εστι \( y = y - \bar{y} \), \( X = X - (1-\mathbf{1}) \) ενός γενικού γραμμικού μοντέλου. Δώστε τον σχεδιασμό δύο περιπτώσεων τυποποιημένων υπολογίσμων. Πώς μιας χωρίσμενων;

ΖΗΤΗΜΑ 4 (Βαθμ. 2.5)

(a) Εφανείται η σχέση μεταξύ \( y \) και \( x_2 \) και έστω δείκτης μεταβλητής \( x_2 \), \( x_2 = 0 \) αν το δεδομένα είναι της κατανομής \( A \) και \( x_2 = 1 \), αν είναι της κατανομής \( B \). Περιγράψτε πώς μέσω αυτής της \( x_2 \) στο μοντέλο \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon \), μπορούμε να ελέγξουμε αν χρειάζεται να προσαρμοστούν \( (\theta) \) δύο διαφορετικές ευθείες \( (\theta) \) δύο παραλληλικές ευθείες ή \( (\theta) \) μια κοινή ευθεία και για τις δύο κατανομές, όπου \( x_2 = x_i x_2 \), η μεταβλητή που εκφράζει την αλληλεπίδραση μεταξύ των μεταβλητών \( x_1 \) και \( x_2 \).

(β) Να γίνουν αυτοί οι έλεγχοι αν \( y = \rho x_2 \) φιλοτεχνήθηκε, \( x_1 = \) πλακί ακτινοβολία και \( x_2 = \) διαδοχικότητα του νερού, \( x_2 = 0 \) αν χαμηλή , \( x_2 = 1 \) αν υψηλή , με βάση τα ακόλουθα αποτελέσματα:

Regression Analysis: y με x_1, x_2, x_3
The regression equation is
\[ y = 114 + 43.5 \times_1 - 25.9 \times_2 - 26.6 \times_3 \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>113.38</td>
<td>29.47</td>
<td>3.86</td>
<td>0.003</td>
</tr>
<tr>
<td>x_1</td>
<td>43.480</td>
<td>3.213</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_2</td>
<td>-25.94</td>
<td>44.620</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_3</td>
<td>4.188</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ s = 33.4748 \quad R^2 = 96.8\% \quad \text{R-Sq(adj)} = 95.08\% \]
Analysis of Variance
Source   DF   SS
Regression 328736
Residual Error 11206
Total      13 349942

Regression Analysis: y με x₁, x₂
The regression equation is
y = 214 + 31.3x₁ - 224x₂

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>214.60</td>
<td>37.49</td>
<td>5.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>x₁</td>
<td>3.636</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x₂</td>
<td>-224.39</td>
<td>33.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Sq = 89.0%  R-Sq(adj) = 87.0%

Analysis of Variance
Source   DF   SS
Regression 38358
Residual Error 11 38358
Total      13 349942

Regression Analysis: y με x₁
The regression equation is
y = 186 + 22.8x₁

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>186.48</td>
<td>79.98</td>
<td>2.33</td>
<td>0.038</td>
</tr>
<tr>
<td>x₁</td>
<td>7.308</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S = 126.769  R-Sq = 44.9%  R-Sq(adj) = 40.3%

Analysis of Variance
Source   DF   SS
Regression 157098
Residual Error 1 157098
Total      13 349942

ZHΣΗΜΑ 5 (ΒΑΘΜ. 2.5)
Κατασκευαστικής λάστιχων, αυτοκινήτου θέλει να εξετάσει αν υπάρχουν διαφορές μεταξύ των τεσσάρων θέσεων τροχου ενός σχήματος, ως κριτήριο ήλεκτρομετρικό. Για την ίδια απόσπαση, έχει κάθε θέση χρησιμοποιήθηκαν 5
tυχαία επιλεγμένα λάστιχα αυτοκινήτου από συνολικά 20 κομμάτια, επίσης τυχαία επιλεγμένα.

<table>
<thead>
<tr>
<th>θέση 1</th>
<th>θέση 2</th>
<th>θέση 3</th>
<th>θέση 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.94</td>
<td>18.28</td>
<td>28.54</td>
<td>20.18</td>
</tr>
<tr>
<td>19.01</td>
<td>21.20</td>
<td>27.99</td>
<td>18.79</td>
</tr>
<tr>
<td>20.33</td>
<td>19.39</td>
<td>30.07</td>
<td>19.20</td>
</tr>
<tr>
<td>17.12</td>
<td>14.82</td>
<td>37.23</td>
<td>34.34</td>
</tr>
<tr>
<td>15.92</td>
<td>11.28</td>
<td>38.85</td>
<td>34.71</td>
</tr>
</tbody>
</table>

Υποθέτουμε ότι η κατασκευή

\[
x_1 = \begin{cases} 
1, \text{ αν θέση 2} \\
0, \text{ αλλιώς}
\end{cases}
\]

\[
x_2 = \begin{cases} 
1, \text{ αν θέση 3} \\
0, \text{ αλλιώς}
\end{cases}
\]

\[
x_3 = \begin{cases} 
1, \text{ αν θέση 4} \\
0, \text{ αλλιώς}
\end{cases}
\]

\[
\text{προσφυγόμετα εύ κατασκευή του μοντέλου παλινδρόμησης, Εγ}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 .
\]
(ii) Να γίνει ο ελέγχος $H_0: \beta_1 = \beta_2 = 0$ με ενσωματωτή $H_1$: τουλάχιστον ένα $\beta_j \neq 0$.

\[ \text{Δίνονται } \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij}^2 = 12176.7, \ S = 5.368, \ S = \left( \frac{\text{SSE}}{n-k-1} \right)^{1/2}. \]

(ii) Να συμπληρωθεί και να ερμηνευθεί ο παρακάτω πίνακας

<table>
<thead>
<tr>
<th>Μεταβλητές</th>
<th>$\hat{\beta}$</th>
<th>se($\hat{\beta}$)</th>
<th>t</th>
<th>p-τιμή</th>
</tr>
</thead>
<tbody>
<tr>
<td>Σταθερά</td>
<td>25.444</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>-6.78</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>-8.45</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_4$</td>
<td>7.09</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ΣΗΜΕΙΩΜΑ 6 (Βαθμ. 2.5)**

Για τη λειτουργία μιας μονάδας παραγωγής επί 21 ημέρες, εξετάζεται η γραμμική εξάρτηση της διαρροής αμμοσιάς $\nu$ (σε log), από τις μεταβλητές $X_1$ (ταχύτητα λειτουργίας της μονάδας) και $X_2$ (θερμοκρασία νερού, °C).

(i) Συμπληρώστε τα παρακάτω πίνακα. Συμπέραντα αποτελέσματα σας.

[Δίνεται: $S = 0.172$, $r_{X_1, X_2} = 0.782$, SST=5.482, AIC=-2.25, $d=2d-n \ln(2) = n + 2 + n \ln(n) + 2p \] \]

<table>
<thead>
<tr>
<th>Μεταβλητές</th>
<th>$\hat{\beta}$</th>
<th>se($\hat{\beta}$)</th>
<th>t</th>
<th>p-τιμή</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Σταθερά</td>
<td>-0.752</td>
<td>0.273</td>
<td>-2.75</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.035</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.063</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2 = \ldots \%$, $C_p = \frac{\text{SSE}(p)}{\text{SSE}(\text{μπροστά})} + 2p - n = \ldots$, AIC=\ldots

(ii) Για το παρακάτω μοντέλο δίνεται ότι $h_{21} = 0.276$. Αποτελεί η παρατήρηση 21 σημείο επιρροής του μοντέλου;

(iii) Δεδομένου ότι στο μοντέλο υπάρχουν οι μεταβλητές $X_1$ και $X_2$, θεωρείται ότι το μοντέλο βελτιώνεται με την προσθήκη της $X_2$; $\text{SSE}_\text{μπροστά} = 0.3858$, $R^2 = \ldots \%$, $C_p = \ldots$, AIC=\ldots

(iv) Εξετάστε εκ νέου αν η παρατήρηση 21 αποτελεί σημείο επιρροής για το μοντέλο (iii) ($h_{21} = 0.288$).
ΖΗΤΗΜΑ 7 (Βαθμ. 2.5)

Έστω μοντέλο παλινδρόμησης Poisson \( R(y; \beta) = \frac{\exp(-\beta) \beta^y}{y!} \), \( y=0,1,2, \ldots \), με συνάρτηση σύνδεσης \( g(u_i)=\ln(u_i)=-\beta x_i \) και με

ελεγχοσυνάρτηση Deviance\(=\sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i}\right) - y_i - \hat{\mu}_i\right]\). 

α) Γίνετε τον ορισμό των υπολοίπων Pearson και Deviance για το μοντέλο αυτό. Πώς το χρησιμοποιούμε;

β) Προσαρμόζοντας μοντέλα της παλινδρόμησης Poisson σε δεδομένα 44 ορυχείων μείον περιοχής, εξετάζεται η σχέση του αριθμού ρομπότ σε όραφη ορυχείου (Y), με τις συμμεταβλητές \( x_1 \) και \( x_2 \), (χαρακτηριστικά των ατμομέτων του ορυχείου) καθώς και με τη \( x_3 \), (έτη λειτουργίας του ορυχείου).

Αφού συμπληρωθούν οι παρακάτω πίνακες θα ερμηνευτούν οι εκτιμημένες ποσότητες \( \exp(\hat{\beta}_i) \) και η γραφική παράσταση των υπολοίπων Deviance του τελικού μοντέλου. Συμπληρώνουν οι Ελέγχοι Wald, Deviance και τα κριτήρια AIC;

<table>
<thead>
<tr>
<th>ΜΟΝΤΕΛΟ: 3</th>
<th>Μεταβλητές</th>
<th>( \hat{\beta}_i )</th>
<th>se(( \hat{\beta}_i ))</th>
<th>( z_i )</th>
<th>p-τιμή</th>
<th>( \exp(\hat{\beta}_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Σταθερά</td>
<td>-3.39</td>
<td>0.9842</td>
<td>-3.445</td>
<td>&lt;0.001</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( x_1 )</td>
<td>0.05860</td>
<td>0.0117</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_2 )</td>
<td>-0.00375</td>
<td>0.0049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_3 )</td>
<td>-0.03408</td>
<td>0.0147</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ελεγχοσυνάρτηση Deviance δίνεται ως \( D_1=41.329 \) και η τιμή του κριτηρίου AIC_3=145.6

<table>
<thead>
<tr>
<th>ΜΟΝΤΕΛΟ: 2</th>
<th>Μεταβλητές</th>
<th>( \hat{\beta}_i )</th>
<th>se(( \hat{\beta}_i ))</th>
<th>( z_i )</th>
<th>p-τιμή</th>
<th>( \exp(\hat{\beta}_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Σταθερά</td>
<td>-3.593</td>
<td>0.9440</td>
<td>-3.813</td>
<td>&lt;0.001</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( x_1 )</td>
<td>0.05874</td>
<td>0.0117</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_2 )</td>
<td>-0.03563</td>
<td>0.0148</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ελεγχοσυνάρτηση Deviance δίνεται ως \( D_2=41.952 \) και η τιμή του κριτηρίου AIC_2=144.22

<table>
<thead>
<tr>
<th>ΜΟΝΤΕΛΟ: 1</th>
<th>Μεταβλητές</th>
<th>( \hat{\beta}_i )</th>
<th>se(( \hat{\beta}_i ))</th>
<th>( z_i )</th>
<th>p-τιμή</th>
<th>( \exp(\hat{\beta}_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Σταθερά</td>
<td>-3.32859</td>
<td>0.90886</td>
<td>-3.662</td>
<td>&lt;0.001</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( x_1 )</td>
<td>0.05234</td>
<td>0.01109</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ελεγχοσυνάρτηση Deviance δίνεται ως \( D_1=48.620 \) και η τιμή του κριτηρίου AIC_1=148.89

![Diagram](image)