ΣΕΕ2
Γραμμικά και μη-Γραμμικά μοντέλα

Φώτης Σιάννης

Τμήμα Μαθηματικών, ΕΚΠΑ

Εαρινό Εξάμηνο 2020

Πλάνο Μαθήματος

1. Γραμμικά μοντέλα (Linear Models)
   - Γενικό γραμμικό μοντέλο (ΕΕΤ, ΕΜΠ, ιδιότητες εκτιμητών)
   - Στατιστική συμπερασματολογία για τους παραμέτρους
   - Προβλέψεις
   - Κριτήρια επιλογής μεταβλητών - Σύγκριση μοντέλων
   - Ετεροσκεδαστικότητα - Γενικευμένος εκτιμητής ΕΤ (GLS)
   - Ψευδομεταβλητές (Dummy Variables)
2. Γενικευμένα γραμμικά μοντέλα (Generalized Linear Models)
   - Συνιστώσες των GLM
   - Εκτίμηση Μέγιστης Πιθανοφάνειας για GLM (Αλγόριθμος Newton-Raphson, Μέθοδος Fisher scoring)
   - Επιλογή μοντέλου
   - Διωνυμικά Δεδομένα - Λογιστική Παλινδρόμηση (Logistic Regression)
   - Μοντέλα Ποισσον (Log-linear models)

Τετραγωνικές Μορφές (Quadratic Forms)

Εισαγωγικά:
Τετραγωνική μορφή είναι κάθε πολυώνυμο του οποίου όλοι οι όροι είναι τάξης δύο. Έστω λοιπόν ότι έχουμε το linear contrast

$$C_1^* = Y_1 + Y_2 - 2Y_3.$$

Το άθροισμα τετραγώνων (AT) είναι:

$$SS(C_1^*) = \frac{(C_1^*)^2}{6},$$

όπου το 6 στο παρανομαστή είναι το άθροισμα των τετραγώνων των συντελεστών. Αυτό επιλέγεται έτσι ώστε να κάνει το συντελεστή του σ^2 στην αναμενόμενη τιμή του AT ίση με 1.

Έστω ότι

$$C_1 = \frac{C_1^*}{\sqrt{6}},$$

τότε

$$\begin{aligned} C_1 &= a'Y \\ &= \frac{1}{\sqrt{6}} Y_1 + \frac{1}{\sqrt{6}} Y_2 - \frac{2}{\sqrt{6}} Y_3, \end{aligned}$$

όπου: $a = \left(\frac{1}{\sqrt{6}} \ \frac{1}{\sqrt{6}} \ \frac{-2}{\sqrt{6})} \right)'$

και: $Y = (Y_1 \ Y_2 \ Y_3)'$.

Τότε

$$\begin{aligned} SS(C_1) = C_1^2 &= (a'Y)'(a'Y) \\ &= Y'(aa')Y \\ &= Y'AY, \end{aligned}$$

όπου

$$\begin{aligned} A = aa' &= \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \end{pmatrix} \\ &= \begin{pmatrix} 1/6 & 1/6 & -2/6 \\ 1/6 & 1/6 & -2/6 \\ -2/6 & -2/6 & 4/6 \end{pmatrix}. \end{aligned}$$

Συνεπώς:

$$Y'AY = (Y_1, Y_2, Y_3) \begin{pmatrix} 1/6 & 1/6 & -2/6 \\ 1/6 & 1/6 & -2/6 \\ -2/6 & -2/6 & 4/6 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$
$$= \frac{1}{6}Y_1^2 + \frac{1}{6}Y_2^2 + \frac{4}{6}Y_3^2 + \frac{2}{6}Y_1Y_2 - \frac{4}{6}Y_1Y_3 - \frac{4}{6}Y_2Y_3.$$

Το πιο πάνω αποτέλεσμα

- Βγαίνει αναλυτικά αν πάρουμε το C_1^2
- Τα διαγώνια στοιχεία του A είναι οι συντελεστές των τετραγωνικών όρων και το άθροισμα των συμμετρικών όρων είναι οι συντελεστές των γινομένων.

ΟΡΙΣΜΟΣ: Το $Y'AY$ είναι μια τεραγωνική μορφή ως προς τις μεταβλητές Y_1, Y_2 & Y_3 με πίνακα συντελεστών (defining matrix) τον πίνακα A.

Έστω τώρα μια επιπλέον τετραγωνική μορφή του Y που είναι κάθετη (orthogonal) στο C_1. Έχουμε:

$$C_2 = \frac{Y_1 - Y_2}{\sqrt{2}} = d'Y,$$

όπου

$$d = \left(\frac{1}{\sqrt{2}} \quad \frac{-1}{\sqrt{2}} \quad 0 \right)'$$

Τότε

$$SS(C_2) = Y'DY,$$

όπου

$$D = dd' = \begin{pmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Βαθμοί Ελευθερίας

Καθένα από τα πιο πάνω AT (SS) έχει 1 BE, μιας και υπάρχει μια και μοναδική γραμμική σχέση (contrast).

- Στις TM, οι BE είναι ίσοι με την τάξη του πίνακα συντελεστών*
- Αν ο πίνακας είναι ταυτοδύναμος, τότε: BE $= tr($πίνακα$)$
- Οι πίνακες A και D είναι ταυτοδύναμοι, μιας και

$$A'A = A \quad \text{και} \quad D'D = D.$$

- Συνεπώς, οι BE είναι: $tr(A) = tr(D) = 1$.

- Αν τα δούμε συνδυαστικά, έχουμε:

$$K'Y = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

- Η ΤΜ είναι της μορφής

$$Y'KK'Y = Y'FY,$$

όπου

$$F = KK' = \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}.$$

- Ο πίνακας F είναι ταυτοδύναμος, οπότε: $\mathrm{BE} = tr(F) = 2$.
- Αυτό συμβαίνει επειδή τα αρχικά contrasts είναι κάθετα (δες σημείωση πιο κάτω)

- Δυο ΤΜ στο Y είναι κάθετες αν οι πίνακες συντελεστών έχουν γινόμενο μηδέν (συμμετρικοί πίνακες και οι δύο)

$$AD = DA = \begin{pmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/6 & 1/6 & -2/6 \\ 1/6 & 1/6 & -2/6 \\ -2/6 & -2/6 & 4/6 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

ΣΗΜΕΙΩΣΗ: Έχουμε

$$DA = dd'aa',$$

άρα για να γίνει μηδέν, αρκεί: $d'a = 0$. Συνεπώς, οι ΤΜ θα είναι κάθετες (orthogonal) αν τα διανύσματα των συντελεστών είναι κάθετα (εσωτερικό γινόμενο ίσο με μηδέν).

ΠΑΡΑΤΗΡΗΣΕΙΣ:

- Όταν οι γραμμικές συναρτήσεις είναι κάθετες, τότε το άθροισμα των AT (και ο BE) των δύο σχέσεων (contrasts) χωριστά ισούται με το AT (και BE) των σχέσεων όταν υπολογίζονται ταυτόχρονα.
- Αυτό ισχύει και για περισόττερα από 2 contrasts, αρκεί να είναι ανά δύο ανεξάρτητα.
- Συνεπώς:

$$\textit{Orthogonality} \Longrightarrow \text{Ανεξαρτησία Πληροφορίας}$$

Ανάλυση Διασποράς

Γνωρίζουμε ότι:

$$Y = \hat{Y} + \hat{\epsilon},$$

και

$$Y'Y = \sum_{i=1}^{n} y_i^2 = SS(Total_{Uncorrected}).$$

Συνεπώς, αυτή είναι μια ΤΜ της μορφής:

$$Y'Y = Y'\mathrm{I}_n Y,$$

όπου ο πίνακας συντελεστών είναι ο I_n.

- Ο I_n είναι συμμετρικός και ταυτοδύναμος
- Το $SS(Total_{Uncorrected})$ έχει ΒΕ ίσο με το $tr(\mathrm{I}_n) = n$, δηλαδή ίσο με το πλήθος των στοιχείων στο Y.
- * Ο μοναδιαίος πίνακας είναι ο μοναδικός ταυτοδύναμος πίνακας πλήρους τάξης (full rank).

Έχουμε λοιπόν:

$$\begin{aligned}
Y'Y =& (\hat{Y}+\hat{\epsilon})'(\hat{Y}+\hat{\epsilon}) \\
=& \hat{Y}'\hat{Y} + \hat{Y}'\hat{\epsilon} + \hat{\epsilon}'\hat{Y} + \hat{\epsilon}'\hat{\epsilon} \overset{\hat{Y}=PY \ \& \ \hat{\epsilon}=(\mathrm{I}_n-P)Y}{=} \\
=& (PY)'(PY) + (PY)'[(\mathrm{I}_n - P)Y] + [(\mathrm{I}_n - P)Y]'(PY) \\
& + [(\mathrm{I}_n - P)Y]'[(\mathrm{I}_n - P)Y] \\
=& Y'P'PY + Y'P'(\mathrm{I}_n - P)Y + Y'(\mathrm{I}_n - P)'PY \\
& + Y'(\mathrm{I}_n - P)'(\mathrm{I}_n - P)Y.
\end{aligned}$$

Όμως:

- $P'P = P$
- $(\mathrm{I}_n - P)'(\mathrm{I}_n - P) = (\mathrm{I}_n - P)$
- $P'(\mathrm{I}_n - P) = P - P = \mathbf{0}$ <u>και</u> $(\mathrm{I}_n - P)'P = P - P = \mathbf{0}$

Οπότε:

$$Y'Y = Y'PY + Y'(\mathrm{I}_n - P)Y = \hat{Y}'\hat{Y} + \hat{\epsilon}'\hat{\epsilon}, \tag{1}$$

όπου το συνολικό $SS(Total_{Uncorrected})$ έχει χωριστεί σε δύο ΤΜ με πίνακες συντελεστών P και $(\mathrm{I}_n - P)$ αντίστοιχα.

Συνεπώς:

- Το $SS(Model) = \hat{Y}'\hat{Y}$ είναι το μέρος του $Y'Y$ που αποδίδεται στο μοντέλο
- Το $SS(Res) = \hat{\epsilon}'\hat{\epsilon}$ είναι το μέρος του $Y'Y$ που δεν εξηγείται από το μοντέλο
- Η καθετότητα των δύο ΤΜ $(P'(\mathrm{I}_n - P) = (\mathrm{I}_n - P)'P = \mathbf{0})$ εγγυάται ότι τα δύο SS μπορούν να προστεθούν για να μας δώσουν το συνολικό SS.

Οι ΒΕ εξαρτώνται από τη τάξη των πινάκων. Οπότε:

- Η τάξη του $P = [X(X'X)^{-1}X']$ εξαρτάται από τη τάξη του πίνακα X. Επόμένως, για μοντέλα πλήρους τάξης, η τάξη του X είναι ίση με το πλήθος των στηλών του πίνακα που είναι ίσο με το πλήθος των $\beta's$. Συνεπώς:

$$tr(P) = tr[X(X'X)^{-1}X'] = tr[(X'X)^{-1}X'X] = tr(\mathrm{I}_{k'}) = k'.$$

- Οι ΒΕ του $SS(Res)$ είναι $(n - k')$, αφου

$$tr(\mathrm{I}_n - P) = tr(\mathrm{I}_n) - tr(P)n - k'.$$

- Τα αντίστοιχα μέσα τετράγωνα MS υπολογίζονται διαιρώντας τα SS με τους αντίστοιχους ΒΕ.
- Υπολογιστικά έχουμε:

$$\begin{aligned} SS(Model) = \hat{Y}'\hat{Y} &= (X\hat{\beta})'[X(X'X)^{-1}X'Y] \\ &= \hat{\beta}'X'X(X'X)^{-1}X'Y = \hat{\beta}'X'Y \end{aligned}$$

και

$$SS(Res) = Y'Y - SS(Model) = Y'Y - \hat{\beta}'X'Y$$

ANOVA TABLE

Source	*SS*	*DF*
Model	$\hat{Y}'\hat{Y} = \hat{\beta}'X'Y$	$r(P) = k'$
Residuals	$\hat{\epsilon}'\hat{\epsilon} = Y'Y - \hat{\beta}'X'Y$	$r(\mathtt{I}_n - P) = n - k'$
Total	$Y'Y$	$r(\mathtt{I}_n) = n$

ΣΗΜΕΙΩΣΕΙΣ:

- Το $SS(Total_{Uncorrected})$ αποτελεί το άθροισμα των τετραγωνικών αποκλίσεων από το μηδεν.
- Το ενδιαφέρον όμως είναι στις αποκλίσεις από το μέσο και πόση πληροφορία υπάρχει από τις ανεξάρτητες μεταβλητές για να εξηγήσει αυτή την απόκλιση.
- Διαφορετικά, η καλύτερη εκτίμηση για το Y θα είναι προφανώς ο δειγματικός μέσος $\bar{y}$, που αποτελεί τη καλύτερη εκτίμηση του πληθυσμιακού μέσου.

Το καλύτερο μέτρο της προσφοράς πληροφορίας των X είναι η διαφορά μεταξύ του μοντέλου με τα X και του μοντέλου χωρίς τα X, όπου η μόνη πληροφορία είναι ο ολικός μέσος, έστω $SS(\mu)$.

Οπότε:

$$SS(Regr) = SS(Model_{Uncorrected}) - SS(\mu),$$

όπου το $SS(\mu)$ συνήθως ονομάζεται διορθωτικός παράγοντας (correction factor).
Το μοντέλο μόνο με το μέσο είναι κι αυτό της μορφής:

$$Y = X\beta + \epsilon,$$

μόνο που ο X περιλαμβάνει μόνο μια στήλη με $'1'$ και $\beta = \mu$.

Τότε:

$$\hat{\beta} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'Y = \frac{1}{n}\mathbf{1}'Y = \bar{y},$$

όπου: $\mathbf{1} = (1\ 1\ 1 \ldots 1)'$ ένα $n \times 1$ διάνυσμα με $'1'$.
Τότε, σύμφωνα με τα παραπάνω έχουμε:

$$SS(\mu) = \hat{\beta}'\mathbf{1}'Y = \frac{1}{n}(\mathbf{1}'Y)'(\mathbf{1}'Y) = Y'\left(\frac{1}{n}\mathbf{1}\mathbf{1}'\right)Y.$$

Λαμβάνοντας υπόψη ότι: $\mathbf{1}'Y = \sum_{i=1}^{n} y_i$, τότε:

$$SS(\mu) = \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

Θέτουμε:

$$\mathbf{11}' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} = \mathrm{J}.$$

Ο πίνακας: $\frac{1}{n}\mathbf{11}' = \frac{\mathrm{J}}{n}$ είναι ταυτοδύναμος, τάξης

$$r\left(\frac{\mathrm{J}}{n}\right) = tr\left(\frac{\mathrm{J}}{n}\right) = 1.$$

Συνεπώς, ο διορθωτικός παράγοντας

$$SS(\mu) = Y'\left(\frac{\mathrm{J}}{n}\right)Y$$

έχει BE=1.

Το ΑΤ που αποδίδεται στο μοντέλο είναι:

$$\begin{aligned} SS(Regr) &= SS(Model_{Uncorrected}) - SS(\mu) \\ &= Y'PY - Y'\left(\frac{\mathrm{J}}{n}\right)Y \\ &= Y'\left(P - \frac{\mathrm{J}}{n}\right)Y. \end{aligned}$$

Ο πίνακας συντελεστών $\left(\frac{\mathrm{J}}{n}\right)$ είναι κάθετος στους $\left(P - \frac{\mathrm{J}}{n}\right)$ και $(\mathrm{I}_n - P)$, οπότε

$$\begin{aligned} Y'Y &= Y'\left(\frac{\mathrm{J}}{n}\right)Y + Y'\left(P - \frac{\mathrm{J}}{n}\right)Y + Y'(\mathrm{I}_n - P)Y \\ &= SS(\mu) + SS(Regr) + SS(Res), \end{aligned}$$

με 1, $k = k' - 1$ και $n - k'$ ΒΕ αντίστοιχα.

Όμως:

$$\begin{aligned} SS(Total_{Corrected}) &= SS(Total_{Uncorrected}) - SS(\mu) \\ &= Y'Y - Y'\left(\frac{\mathrm{J}}{n}\right)Y \\ &= \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} \\ &= \sum_{i=1}^{n} y_i^2 - n\bar{y} \\ &= \sum_{i=1}^{n} (y_i - \bar{y})^2, \end{aligned}$$

το οποίο ονομάζεται απλά: $SS(Total) = SST$.

Άρα:

$$SST = Y'\mathrm{I}_n Y - Y'\left(\frac{\mathrm{J}}{n}\right)Y = Y'\left(\mathrm{I}_n - \frac{\mathrm{J}}{n}\right)Y,$$

και έχει:

$$tr\left(\mathrm{I}_n - \frac{\mathrm{J}}{n}\right) = tr\left(\mathrm{I}_n\right) - tr\left(\frac{\mathrm{J}}{n}\right) = n - 1$$

βαθμούς ελευθερίας.

ANOVA TABLE

Source	*DF*	*SS*
$Model_{Corrected}$	k	$\hat{\beta}'X'Y - n\bar{y}^2$
Residuals	$n - k' = n - k - 1$	$Y'Y - \hat{\beta}'X'Y$
$Total_{Uncorrected}$	n	$Y'Y$
Correction Factor	1	$n\bar{y}^2$
$Total_{Corrected}$	$n - 1$	$Y'Y - n\bar{y}^2$

Μέσες Τιμές των Τετραγωνικών Μορφών

ΓΕΝΙΚΑ: Έστω $E(Y) = \mu$ Y το διάνυσμα με τις αναμενόμενες τιμές και $V(Y) = V_Y = V\sigma^2$ ο πίνακας διασπορών-συνδιασπορών.
Ισχύει:

$$\begin{aligned} E[Y'AY] &= tr(AV_y) + \mu' A\mu \\ &= \sigma^2 tr(AV) + \mu' A\mu. \end{aligned}$$

Έχοντας τις υποθέσεις : $E(Y) = X\beta$ και $V(Y) = \sigma^2 \mathtt{I}$, τότε:

$$E[Y'AY] = \sigma^2 tr(A) + \beta' X' AX\beta,$$

βάζοντας στη θέση του A τον κατάλληλο πίνακα.
Σημείωση: Αν ο A είναι ταυτοδύναμος, τότε $tr(A)$ είναι οι ΒΕ της ΤΜ.

Απόδειξη: Ξεκινάμε από το γεγονός ότι το $Y'AY$ είναι στοιχείο (1×1).

$$\begin{aligned} E[Y'AY] &= E\left[tr(Y'AY)\right] = E\left[tr(AYY')\right] \\ &= tr\left[A\ E\left(YY'\right)\right] = tr\left[A\ \left(V_Y + \mu\mu'\right)\right] \\ &= tr(AV_Y) + tr(A\mu\mu') = tr(AV_Y) + tr(\mu' A\mu) \\ &= tr(AV_Y) + \mu' A\mu. \end{aligned}$$

Οπότε, έχουμε:

$$\begin{aligned} E\left[SS(Model_{Un})\right] = E(Y'PY) &= \sigma^2 tr(P) + \beta' X'PX\beta \\ &= k'\sigma^2 + \beta' X'X\beta, \end{aligned}$$

αφού: $tr(P) = k'$ και $PX = X$.

Σημειώνουμε ότι η ΤΜ $\beta' X'PX\beta = \beta' X'X\beta$ αποτελεί μια ΤΜ του β, συμπεριλαμβανομένου του β_0.

Πιο συγκεκριμένα:

$$
\begin{aligned}
E[SSR] = E[SS(Regr)] &= E\left[Y'\left(P - \frac{\mathrm{J}}{n}\right)Y\right] \\
&= \sigma^2 tr\left(P - \frac{\mathrm{J}}{n}\right) + \beta' X'\left(P - \frac{\mathrm{J}}{n}\right)X\beta \\
&= k\sigma^2 + \beta' X'\left(\mathrm{I}_n - \frac{\mathrm{J}}{n}\right)X\beta,
\end{aligned}
$$

αφού: $X'P = X'$.

Σημειώνουμε ότι η ΤΜ $\beta' X'\left(\mathrm{I}_n - \frac{\mathrm{J}}{n}\right)X\beta$ αποτελεί μια ΤΜ του β η οποία **δεν** συμπεριλαμβάνει το β_0. Ο λόγος είναι ότι ο πίνακας $X'\left(\mathrm{I}_n - \frac{\mathrm{J}}{n}\right)X$ έχει τη πρώτη γραμμή και τη πρώτη στήλη με $0's$, γεγονός που εξαιρεί το β_0 από τη συγκεκριμένη ΤΜ. Συνεπώς, μόνο οι παράμετροι των $X's$ είναι στη συγκεκριμένη ΤΜ.

Ομοίως:

$$\begin{aligned} E[SSE] = E[SS(Res)] &= E\left[Y'\left(\mathtt{I}_n - P\right)Y\right] \\ &= \sigma^2 tr\left(\mathtt{I}_n - P\right) + \beta' X'\left(\mathtt{I}_n - P\right)X\beta \\ &= (n-k')\sigma^2 + \beta' X'\left(X - X\right)\beta \\ &= (n-k')\sigma^2, \end{aligned}$$

όπου ο συντελεστής του σ^2 είναι ακριβώς οι ΒΕ του $SS(Res)$, όπως πιο πάνω είναι οι ΒΕ του $SS(Regr)$.

Διαιρώντας με τους αντίστοιχους ΒΕ έχουμε:

$$\begin{aligned} E[MS(Regr)] &= \sigma^2 + \frac{\beta' X'\left(\mathtt{I}_n - \frac{\mathtt{J}}{n}\right)X\beta}{k} \\ E[MS(Res)] &= \sigma^2. \end{aligned}$$

- Αυτό δείχνει ότι το $MSE = MS(Res)$ αποτελεί αμερόληπτο εκτιμητή του σ^2.
- Αντίστοιχα, το $MSR = MS(Regr)$ αποτελεί έναν εκτιμητή του σ^2 συν μια ΤΜ των β_j εκτός του β_0.
- Συνεπώς, η σύγκριση του MSR με το MSE προσφέρει τη βάση για να κρίνουμε τη σημαντικότητα των $\beta_j,\ j = 1, \ldots, k$ και πιο συγκεκριμένα των ανεξάρτητων $X's$.
- Η αναμενόμενη τιμή του MSR θα είναι πάντα μεγαλύτερη αυτής του MSE, αφού η ΤΜ του β δε μπορεί να είναι αρνητική.
- Συνεπώς, ο λόγος:

$$\frac{MSR}{MSE}$$

προσφέρει τον έλεγχο

$$H_0 :\ \beta_j = 0, \quad \forall j \ \text{ εκτός του } \beta_0$$

Κατανομή των Τετραγωνικών Μορφών

Γενικά: Αν έχουμε $Y \sim N(\mu, V\sigma^2)$, όπου V ένας μη ιδιάζων πίνακας (non-singular or regular matrix), δηλ. έχει αντίστροφο. Το μ μπορεί να είναι $X\beta$ και $V = \mathtt{I}_n$. Αν ο AV είναι πίνακας προβολής, τότε

- η ΤΜ $Y'AY/\sigma^2 = Y'(A/\sigma^2)Y$ ακολουθεί χ^2 κατανομή (non-central χ^2 distribution) με
  - α) BE $= tr(AV)$ και
  - β) non-centrality parameter: $\Omega = \mu' A\mu/\sigma^2$
    (υπολογίζεται από την αναμενόμενη τιμή της ΤΜ).
- Οι ΤΜς $Y'AY$ και $Y'BY$ είναι ανεξάρτητες αν: $AVB = 0$. Αν $V = \mathtt{I}_n$ τότε $AB = 0$, που σημαίνει ότι είναι κάθετες (orthogonal) η μια στην άλλη.
- Η ΤΜ $Y'AY$ είναι ανεξάρτητη της γραμμικής συνάρτησης BY αν: $BVA = 0$ ή $BA = 0$ εφόσον $V = \mathtt{I}_n$.

Για το πολλαπλό γραμμικό μοντέλο ισχύουν:

- Τα SS του μοντέλου, μέσου, παλινδρόμησης και υπολοίπων έχουν πίνακες που είναι ταυτοδύναμοι. Για παράδειγμα:

$$\frac{SS(Model_{Un})}{\sigma^2} = \frac{Y'PY}{\sigma^2},$$

αφου $P^2 = P$. Τότε, το $\frac{SS(Model_{Un})}{\sigma^2}$ ακολουθεί μια χ^2 κατανομή με

  - BE=$r(P) = tr(P) = k'$ και
  - $\Omega = \beta' X' P X \beta / \sigma^2$.

- Ομοίως, το

$$\frac{SS(\mu)}{\sigma^2} = \frac{Y'\left(\mathrm{J}/n\right)Y}{\sigma^2}$$

ακολουθεί μια χ^2 κατανομή με

  - BE=$r(\mathrm{J}/n) = tr(\mathrm{J}/n) = 1$ και
  - $\Omega = \beta' X' (\mathrm{J}/n) X \beta / \sigma^2$.

- Επίσης, το

$$\frac{SSR}{\sigma^2} = \frac{SS(Regr)}{\sigma^2} = \frac{Y'\left(P - \mathtt{J}/n\right)Y}{\sigma^2}$$

ακολουθεί μια χ^2 κατανομή με

  - > BE$=r(P - \mathtt{J}/n) = tr(P - \mathtt{J}/n) = k$ και
  - > $\Omega = \beta' X'(P - \mathtt{J}/n)X\beta/\sigma^2 = \beta' X'(\mathtt{I} - \mathtt{J}/n)X\beta/\sigma^2$.

- Τέλος, το

$$\frac{SSE}{\sigma^2} = \frac{SS(Res)}{\sigma^2} = \frac{Y'\left(\mathtt{I}_n - P\right)Y}{\sigma^2}$$

ακολουθεί μια χ^2 κατανομή με

  - > BE$=r(\mathtt{I}_n - P) = tr(\mathtt{I}_n - P) = n - k'$ και
  - > $\Omega = \beta' X'(\mathtt{I}_n - P)X\beta/\sigma^2 = 0$.

ΠΑΡΑΤΗΡΗΣΕΙΣ:

- Αφού έχουμε: $(\mathtt{I}_n - P)(P - \mathtt{J}/n) = \mathbf{0}$, τότε τα $SSE = Y'(\mathtt{I}_n - P)Y$ και $SSR = Y'(P - \mathtt{J}/n)Y$ είναι ανεξάρτητα.
- Ομοίως, αφού:
  - $P(\mathtt{I}_n - P) = \mathbf{0}$
  - $\mathtt{J}/n(P - \mathtt{J}/n) = \mathbf{0}$
  - $\mathtt{J}/n(\mathtt{I}_n - P) = \mathbf{0}$

  τότε τα:
  - $SS(Model_{Un})$ και SSE
  - $SS(\mu)$ και SSR
  - $SS(\mu)$ και SSE

  είναι ανεξάρτητα.
- Τέλος, αφού $X'(\mathtt{I}_n - P) = \mathbf{0}$, τότε κάθε γραμμική συνάρτηση

$$K'\hat{\beta} = K'(X'X)^{-1}X'Y = BY$$

  είναι ανεξάρτητη του: $SSE = SS(Res) = Y'(\mathtt{I}_n - P)Y$.

* Συμπέρασμα: Η υπόθεση κανονικότητας για τα ϵ οδηγεί όλα τα SS διαιρούμενα με το σ^2 να ακολουθούν (non-central) χ^2 κατανομές.

$F-$ *test* για το Γραμμικό Μοντέλο

Ο λόγος:

$$\mathbf{F} = \frac{MSR}{MSE}$$

αποτελεί έναν έλεγχο για την υπόθεση

$$H_0 : \ \beta_1 = \beta_2 = \ \ldots \ = \ \beta_k \ = \ 0.$$

Δείξαμε ότι το MSE αποτελεί μια α.ε. του σ^2 ενώ το MSR υπερεκτιμά το σ^2 κατά έναν όρο που αποτελεί μια ΤΜ του β, εκτός του β_0, τον

$$\frac{\beta' X' \left(\mathrm{I} - \frac{\mathrm{J}}{n}\right) X\beta}{k}.$$

Η ΤΜ μπορεί να γίνει ίση με το μηδέν αν

$$\left(\mathrm{I} - \frac{\mathrm{J}}{n}\right) X\beta = 0.$$

Εφόσον ο X είναι πλήρους τάξης, τότε μπορεί να δειχθεί ότι παραπάνω όρος γίνεται μηδέν όταν

$$\beta_1 = \beta_2 = \ \ldots \ = \ \beta_k \ = \ 0.$$

Άν λοιπόν οι παρατηρούμενες τιμές του F είναι σημαντικά μεγαλύτερες του 1, αυτό δείχνει ότι η πιο πάνω ποσότητα δεν είναι μηδέν.

Η ποσότητα **F** λοιπόν, ως λόγος δύο ανεξάρτητων χ^2 τυχαίων ποσοτήτων διαιρεμένων με τους αντίστοιχους BE τους αποτελεί μια τυχαία ποσότητα που ακολουθεί τη κατανομή $\mathcal{F}$, άρα

$$\mathbf{F} \sim \mathcal{F}_{k,n-k'}.$$

Ο έλεγχος αυτός, που αποτελεί τη τελευταία στήλη στο πίνακα *ANOVA*, αποτελεί έναν έλεγχο για το μοντέλο συνολικά.

Στατιστική Συμπερασματολογία

Έστω το γραμμικό μοντέλο

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \ldots, n.$$

Τότε:

- Έχουμε
$$\hat{\beta}_0 \sim N\left(\beta_0, V(\hat{\beta}_0)\right)$$
- και
$$\hat{\beta}_1 \sim N\left(\beta_1, V(\hat{\beta}_1)\right).$$
- Γνωρίζουμε ότι
$$V(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

- Τυποποιούμαι και παίρνουμε

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{V(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim N(0,1)$$

- Από τη στιγμή που το σ^2 είναι άγνωστο, τότε παίρνουμε

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n - k'} = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n - 2}.$$

- Οπότε, το

$$\widehat{V(\hat{\beta}_1)} = S^2(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

η οποία είναι α.ε. του $V(\hat{\beta}_1)$.

Συνεπώς, αφού:

α) Έχουμε

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim N(0,1),$$

β) και

$$\frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{\sigma^2} = \frac{\hat{\epsilon}'\hat{\epsilon}}{\sigma^2} \sim \chi^2_{n-2}$$

γ) και ανεξάρτητα

τότε παίρνουμε

$$\mathbf{t} = \frac{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}}}{\sqrt{\frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{\sigma^2}/(n-2)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{V(\hat{\beta}_1)}}} \sim t_{n-2}$$

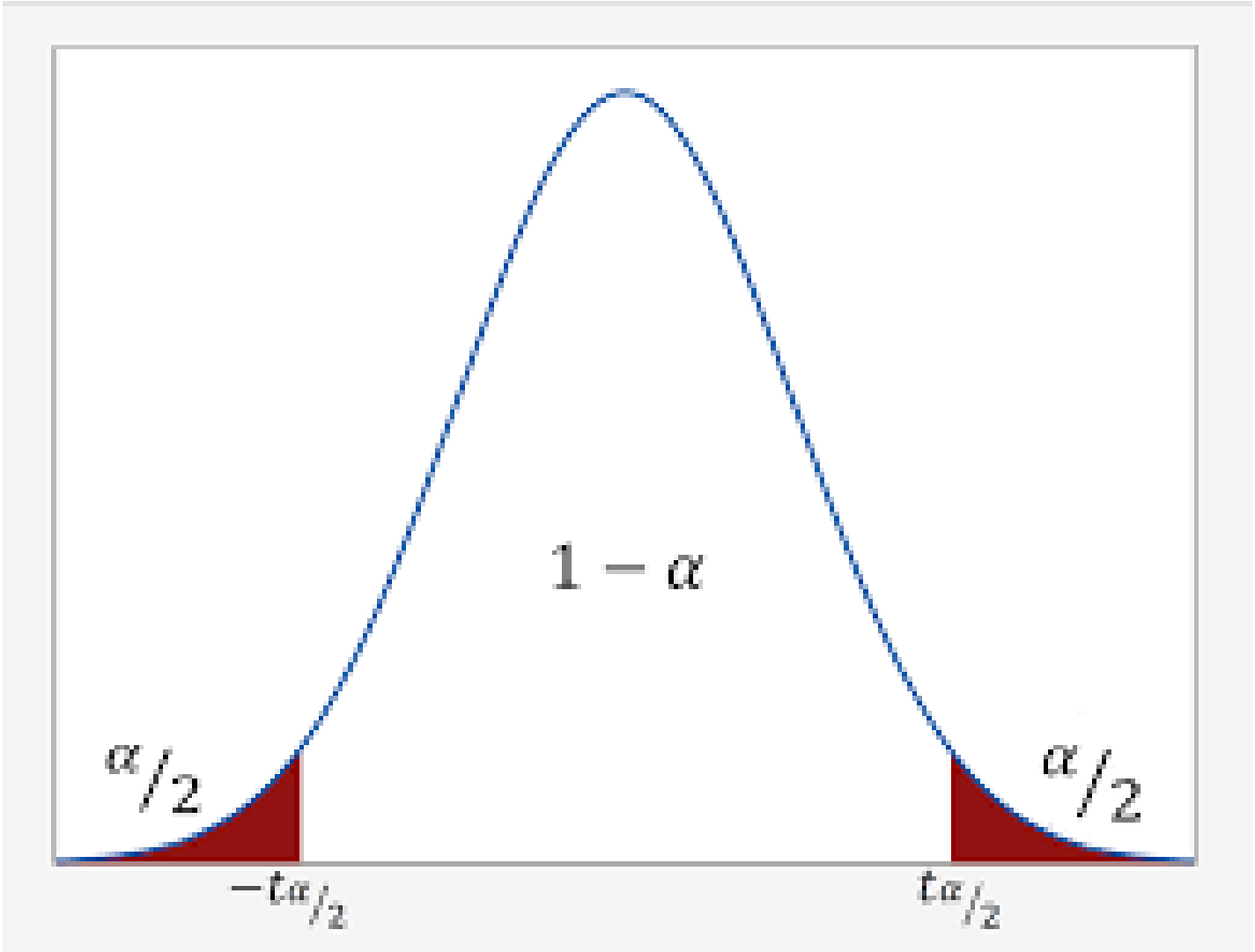
Διάστημα Εμπιστοσύνης για το β_1

Γνωρίζοντας τη κατανομή του $\hat{\beta}_1$ μπορούμε να κατασκευάσουμε ένα (1-α)100% διάστημα εμπιστοσύνης για το β_1. Έχουμε:

$$P(-t_{n-2,\alpha/2} \leq \frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \leq t_{n-2,\alpha/2}) = 1-\alpha \Rightarrow$$

$$P(-t_{n-2,\alpha/2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \leq \hat{\beta}_1-\beta_1 \leq t_{n-2,\alpha/2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}) = 1-\alpha \Rightarrow$$

$$P(\hat{\beta}_1-t_{n-2,\alpha/2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \leq \beta_1 \leq \hat{\beta}_1+t_{n-2,\alpha/2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}) = 1-\alpha$$


$1-\alpha$
$\alpha/2$
$\alpha/2$
$-t\alpha/2$
$t\alpha/2$

Γενίκευση για Πολλαπλή Παλινδρόμηση

Έχουμε το μοντέλο

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_1 x_{ik} + \epsilon_i,$$

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \ldots, n.$$

Έχουμε:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

και

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1},$$

του οποίου μια α.ε. είναι

$$S^2(\hat{\beta}) = \widehat{V(\hat{\beta})} = \hat{\sigma}^2(X'X)^{-1}$$

με

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k-1} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k'}.$$

Έστω

- β_j, το j-στο στοιχείο του $\hat{\beta}$ και
- $S^2(\hat{\beta}_j)$ το jj-διαγώνιο στοιχείο του πίνακα $S^2(\hat{\beta})$.

Τότε, το ΔΕ για το β_j είναι

$$\beta_j \pm t_{n-k-1,\alpha/2}\sqrt{S^2(\hat{\beta}_j)}.$$

Έλεγχοι Υποθέσεων για το β_1

Στο απλό γραμμικό μοντέλο έχουμε τους εξής πιθανούς ελέγχους. Η αρχική υπόθεση είναι

$$H_0 : \beta_1 = \beta_1^0$$

έναντι

$$(\alpha)\ H_1 : \beta_1 \neq \beta_1^0 \quad \text{ή}\ (\beta)\ H_1 : \beta_1 > \beta_1^0 \quad \text{ή}\ (\gamma)\ H_1 : \beta_1 < \beta_1^0.$$

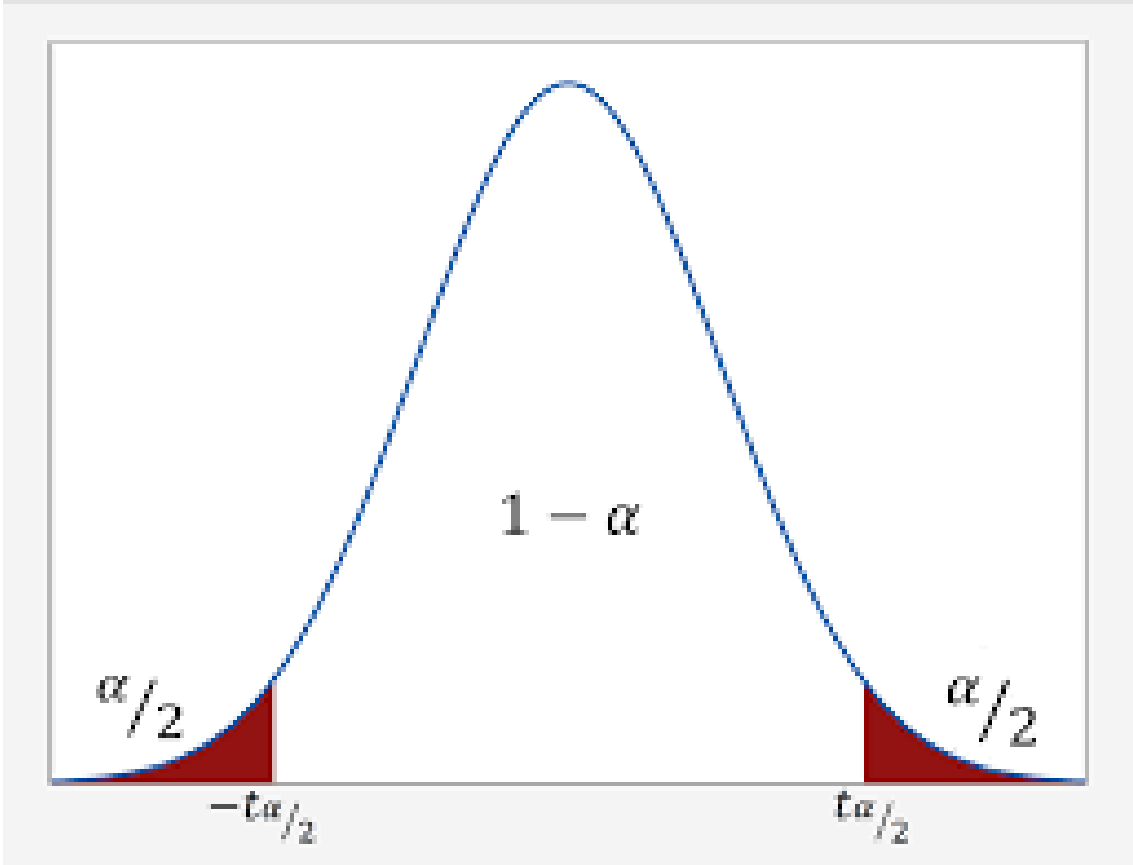
Όταν $\beta_1^0 = 0$ τότε έχουμε έλεγχο στατιστικής σημαντικότητας.

Κάτω από την H_0 η στατιστική συνάρτηση ελέγχου είναι

$$\mathbf{t} = \frac{\hat{\beta}_1 - \beta_1^0}{S(\hat{\beta}_1)} \sim t_{n-2}.$$

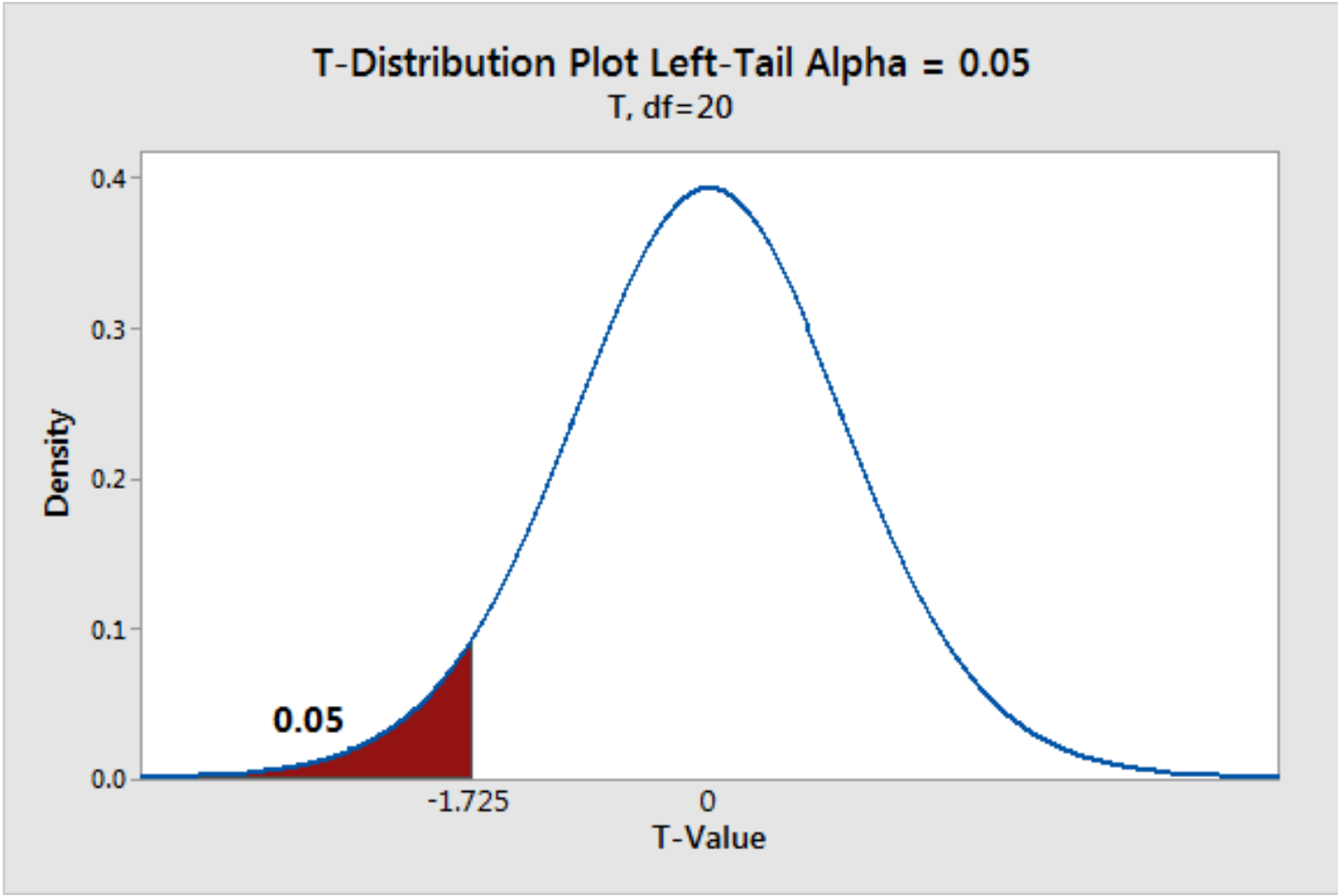
Συνεπώς απορρίπτουμε την H_0 έναντι της H_1 (α) για ακραίες θετικές ή αρνητικές τιμές της ελεγχοσυνάρτησης.

Δηλαδή, απορρίπτουμε την H_0 σε ε.σ.σ. α αν: $|\mathbf{t}| > t_{n-2,a/2}$.



Αντιστοίχως,

- για την H_1 (β) απορρίπτουμε όταν: $\mathbf{t} > t_{n-2,a}$, ενώ
- για την H_1 (γ) απορρίπτουμε όταν: $\mathbf{t} < t_{n-2,a}$.



Γενικά στο μοντέλο πολλαπλής παλινδρόμησης έχουμε

$$\mathbf{t} = \frac{\hat{\beta}_j - \beta_j}{S(\hat{\beta}_j)} \sim t_{n-k-1},$$

και

- Το ΔΕ είναι: $\beta_j \pm t_{n-k-1,\alpha/2} S(\hat{\beta}_j)$
- Ο ΕΥ με σ.σ.ε.: $\frac{\hat{\beta}_j - \beta_j}{S(\hat{\beta}_j)}$.

Για το πολλαπλό μοντέλο έχουμε

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}).$$

Άρα

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 (X'X)^{-1}_{jj}),$$

και

$$\frac{\hat{\epsilon}'\hat{\epsilon}}{\sigma^2} = \frac{\hat{\epsilon}'\hat{\epsilon}\ \frac{n-k-1}{n-k-1}}{\sigma^2} = \frac{\hat{\sigma}^2(n-k-1)}{\sigma^2} \sim \chi^2_{n-k-1}.$$

Συνεπώς:

$$\mathbf{t} = \frac{(\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2 (X'X)^{-1}_{jj}}}{\sqrt{\frac{\hat{\sigma}^2(n-k-1)}{\sigma^2}/(n-k-1)}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 (X'X)^{-1}_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{S(\hat{\beta}_j)} \sim t_{n-k-1}.$$

Άσκηση

Έστω το μοντέλο

$$y_i = \beta_0 + \beta_1 x_{i1}^2 + \frac{\beta_3}{x_{i2}} + \epsilon_i,$$

όπου:

$$\epsilon_i \sim N(0, \sigma^2) \text{ ανεξάρτητα για } i = 1, 2, \ldots, n.$$

α) Να βρεθεί ο ΕΕΤ του $\beta = (\beta_0, \beta_1, \beta_2)$ καθώς και η αναμενόμενη τιμή και ο πίνακας διακύμανσης του $\hat{\beta}$.

β) Θεωρείστε το μοντέλο σε μορφή πινάκων, με

$$Y = X\beta + \epsilon.$$

Ένα δείγμα με $n = 30$ μας έδωσε:

$$(X'X)^{-1} = \begin{pmatrix} 1.70 & -0.06 & -0.08 \\ -0.06 & 0.92 & 0.55 \\ -0.08 & 0.55 & 0.54 \end{pmatrix}, \quad X'Y = \begin{pmatrix} 1.60 \\ -30 \\ 47 \end{pmatrix}$$

και $\hat{\epsilon}'\hat{\epsilon} = \sum_{i=1}^{30} \hat{\epsilon}_i^2 = 6.$

Να γίνουν οι παρακάτω έλεγχοι:

(*i*) $H_0 : \beta_1 = 0$ έναντι $H_1 : \beta_1 \neq 0$, και

(*ii*) $H_0 : \beta_2 = 7$ έναντι $H_1 : \beta_2 > 7$.

Λύση:

α) Έχουμε: $Y = X\beta + \epsilon$ και $\epsilon \sim N(\mathbf{0}, \mathtt{I}_n)$, με

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11}^2 & \frac{1}{x_{12}} \\ 1 & x_{21}^2 & \frac{1}{x_{22}} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1}^2 & \frac{1}{x_{n2}} \end{pmatrix} \text{ και } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

Κατά τα γνωστά ελαχιστοποιώ: $\hat{\epsilon}'\hat{\epsilon} = \sum_{i=1}^{30} \hat{\epsilon}_i^2$, όπου

$$\hat{\epsilon} = Y - X\hat{\beta}.$$

Οπότε:

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

και:

$$\frac{\partial}{\partial \hat{\beta}}(Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}) = \mathbf{0} \Rightarrow \ldots \Rightarrow \hat{\beta} = (X'X)^{-1}X'Y,$$

με $(X'X)$ αντιστρέψιμο.

(β) Έχουμε:

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 1.70 & -0.06 & -0.08 \\ -0.06 & 0.92 & 0.55 \\ -0.08 & 0.55 & 0.54 \end{pmatrix} \begin{pmatrix} 1.60 \\ -30 \\ 47 \end{pmatrix}$$

$$= \begin{pmatrix} 0,76 \\ -1,85 \\ 8,75 \end{pmatrix}$$

Επίσης:

$$V(\hat{\beta}) = \hat{\sigma^2}(X'X)^{-1} = \frac{\sum_{i=1}^{30} \hat{\epsilon}_i^2}{n-3}(X'X)^{-1}$$

$$= \frac{6}{30-3}\begin{pmatrix} 1.70 & -0.06 & -0.08 \\ -0.06 & 0.92 & 0.55 \\ -0.08 & 0.55 & 0.54 \end{pmatrix}$$

$$= \begin{pmatrix} 0.374 & -0.013 & -0.018 \\ & 0.202 & 0.121 \\ & & 0.119 \end{pmatrix}$$

(i) Έλεγχος: $H_0 : \beta_1 = 0$ έναντι $H_1 : \beta_1 \neq 0$.
Ελεγχοσυνάρτηση (για $\alpha = 0.05$):

$$\mathbf{t} = \frac{\hat{\beta}_1 - 0}{\sqrt{\widehat{V(\hat{\beta}_1)}}} = \frac{-1.85}{\sqrt{0.202}} = -4.116.$$

Γνωρίζουμε ότι $\mathbf{t} \sim t_{27}$, καθώς επίσης ότι $t_{27,0.025} = 2.056..$
Αφού λοιπόν $|\mathbf{t}| > 2,056 = t_{27,0.025}$ τότε απορρίπτουμε την H_0.

ii) Έλεγχος: $H_0 : \beta_2 = 7$ έναντι $H_1 : \beta_2 > 7$.
Ελεγχοσυνάρτηση (για $\alpha = 0.05$):

$$\mathbf{t} = \frac{\hat{\beta}_2 - 7}{\sqrt{\widehat{V(\hat{\beta}_2)}}} = \frac{8,75 - 7}{\sqrt{0.119}} = 5,073.$$

Γνωρίζουμε ότι $\mathbf{t} \sim t_{27}$, καθώς επίσης ότι $t_{27,0.05} = 1,703..$
Αφού λοιπόν $|\mathbf{t}| > 1,703 = t_{27,0.05}$ τότε απορρίπτουμε την H_0.

Γενική Μορφή Ελέγχων

- Γενικά έχουμε την υπόθεση

$$H_0 : \ K'\beta = m \ \text{ έναντι } \ H_1 : K'\beta \neq m,$$

όπου K' ένας $(\ell \times k')$ πίνακας συντελεστών που ορίζουν $\ell-$γραμμικές συναρτήσεις των $\beta' s$ που είναι προς έλεγχο και m ένα $(\ell \times 1)$ διάνυσμα σταθερών τιμών, που συχνά είναι μηδέν.

- Οι $\ell-$εξισώσεις στην H_0 πρέπει να είναι γραμμικώς ανεξάρτητες (όχι απαρραίτητα κάθετες).
- Αυτό σημαίνει ότι ο K' πίνακας είναι πλήρους τάξης με $r(K') = \ell$.
- Οι γραμμικές εξισώσεις δεν μπορούν να είναι περισσότερες από τα β, αλλιώς ο K' δεν θα είναι πλήρους τάξης.

Παράδειγμα

Έστω

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

και θέλουμε να ελέγξουμε την υπόθεση ότι

$$\beta_1 = \beta_2, \quad \beta_1 + \beta_2 = 2\beta_3 \quad \text{και} \quad \beta_0 = 20.$$

Επόμενως,

$$\begin{aligned} H_0 : \beta_1 - \beta_2 &= 0 \\ \beta_1 + \beta_2 - 2\beta_3 &= 0 \\ \beta_0 &= 20. \end{aligned}$$

Αυτό μπορεί να γραφεί στη μορφή: $K'\beta = m$, όπου:

$$K = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \text{και} \quad m = \begin{pmatrix} 0 \\ 0 \\ 20 \end{pmatrix}.$$

Η εναλλακτική υπόθεση είναι: $H_1 : K'\beta \neq m$.

Έχουμε:

- Η ΕΕΤ του $K'\beta - m$ είναι $K'\hat{\beta} - m$, χρησιμοποιώντας την εκτίμηση του β.
- Υπό την υπόθεση της κανονικότητας, το $K'\hat{\beta} - m$ ακολουθεί κανονική κατανομή με

$$E(K'\hat{\beta} - m) = K'\beta - m$$

και

$$V(K'\hat{\beta} - m) = K'(X'X)^{-1}K\sigma^2 = V\sigma^2.$$

- Υπό την H_0, η μέση τιμή η μέση τιμή είναι μηδέν

$$E(K'\hat{\beta} - m) = \mathbf{0}$$

.

- Το ΑΤ της γραμμικής υπόθεσης H_0 είναι

$$Q = (K'\hat{\beta} - m)'V^{-1}(K'\hat{\beta} - m),$$

όπου $V^{-1} = [K'(X'X)^{-1}K]^{-1} = A$ είναι ο πίνακας συντελεστών της ΤΜ του $K'\hat{\beta} - m$.

- Οπότε: $tr(AV) = tr(\mathtt{I}_\ell) = \ell$.
- Η αναμενόμενη τιμή του Q είναι

$$E(Q) = \ell\sigma^2 + (K'\hat{\beta} - m)'[K'(X'X)^{-1}K]^{-1}(K'\hat{\beta} - m).$$

- Συνεπώς, η ποσότητα Q/σ^2 ακολουθεί μια non-central χ^2_ℓ (αφού $AV = \mathtt{I}_\ell$ ταυτοδύναμος).
- Οι ΒΕ είναι: $r(A) = r(K) = \ell$.
- Έχουμε: $\Omega = \frac{(K'\hat{\beta}-m)'V^{-1}(K'\hat{\beta}-m)}{\sigma^2}$ τον non-centrality parameter. Υπό την H_0 έχουμε: $\Omega = 0$.
- Συνεπώς, η ποσότητα Q/ℓ μπορεί να χρησιμοποιηθεί (ως αριθμητής) σε ένα $\mathcal{F} - test$ που θα ελέγχει την H_0.

- Ως παρανομαστής στον συγκεκριμένο έλεγχο μπορούμε να έχουμε την όποια α.ε. του σ^2, συνήθως το MSE. Συνεπώς:

$$\mathcal{F} = \frac{Q/r(K)}{S^2} = \frac{Q/\ell}{S^2}.$$

- Αυτή η δομή είναι κατάλληλη για τον έλεγχο της όποιας H_0.

Ειδικές Περιπτώσεις

Α) Απλή Υπόθεση (απλό γρ. μοντέλο)

Σε αυτή τη περίπτωση ο πίνακας K' είναι διάνυσμα γραμμή, έτσι ώστε η ποσότητα $K'(X'X)^{-1}K$ να είναι ένας αριθμός (scalar). Συνεπώς, ο αντίστροφος είναι $\frac{1}{K'(X'X)^{-1}K}$ και το ΑΤ της υπόθεσης παίρνει τη μορφή

$$Q = \frac{(K'\hat{\beta} - m)^2}{K'(X'X)^{-1}K},$$

με μόλις έναν ΒΕ. Οπότε:

$$\mathbf{F} = \frac{(K'\hat{\beta} - m)^2}{[K'(X'X)^{-1}K]\hat{\sigma}^2}.$$

Στη περίπτωση αυτή το $F - test$ είναι το τετράγωνο του $t - test$, αφού:

$$\mathbf{t} = \frac{K'\hat{\beta} - m}{\sqrt{[K'(X'X)^{-1}K]\hat{\sigma}^2}},$$

όπου ο παρανομαστής είναι η τυπική απόκλιση του αριθμητή.

B) Τα ℓ από τα β_s να είναι μηδεν

- Η αρχική υπόθεση είναι ότι τα ℓ από τα β_s να είναι μηδεν.
- Σε αυτή τη περίπτωση ο πίνακας K θα αποτελείται από μηδενικά εκτός από έναν '1' σε κάθε γραμμή που θα υποδεικνύει ποιό από τα β_s θα ελέγχεται αν είναι μηδέν.
- Συνεπώς, ο πολλαπλασιασμός $K'(X'X)^{-1}K$ παίρνει τον $(\ell \times \ell)$ υποπίνακα από τον $(X'X)^{-1}$ που είναι οι συντελεστές για τις διασπορές/συνδιασπορές των β_s που είναι προς έλεγχο.
- Συνεπώς, αν θέλουμε να ελέγξουμε τα $\beta_1,\ \beta_3,\ \beta_5$ παίρνει τη μορφή

$$Q = (\hat{\beta}_1\ \hat{\beta}_3\ \hat{\beta}_5) \begin{pmatrix} c_{11} & c_{13} & c_{15} \\ c_{31} & c_{33} & c_{35} \\ c_{51} & c_{53} & c_{55} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_3 \\ \hat{\beta}_5 \end{pmatrix},$$

όπου c_{ij} είναι το $(i+1) \times (j+1)$ στοιχείο του $(X'X)^{-1}$ (λόγω β_0).

- Το συγκεκριμένο ΑΤ μετρά τη συμβολή των (x_1, x_3, x_5) σε ένα μοντέλο που ήδη έχει τις υπόλοιπες μεταβλητές.

Γ) Ένα απο τα β_s να είναι μηδεν

- Η υπόθεση είναι $H_0 : \beta_j = 0$.
- Σε αυτή τη περίπτωση ο K' είναι διάνυσμα γραμμή, με μηδενικά και ένα '1' στη θέση του β που είναι για έλεγχο.
- Ο πίνακας $K'(X'X)^{-1}K$ στην Q παίρνει το $(j+1)$ διαγώνιο στοιχείο του $(X'X)^{-1}$, το c_{jj}, το οποίο είναι ο συντελεστής της διασποράς του $\hat{\beta}_j$ (πολλαπλασιάζει το $\hat{\sigma}^2$).
- Το ΑΤ παίρνει τη μορφή

$$Q = \frac{\hat{\beta}_j^2}{c_{jj}}.$$

- Το $F - test$ παίρνει τη μορφή

$$\mathbf{F} = \frac{\hat{\beta}_j^2}{c_{jj}\hat{\sigma}^2}.$$

- Το αντίστοιχο $t - test$ παίρνει τη μορφή

$$\mathbf{t} = \frac{\hat{\beta}_j}{\sqrt{c_{jj}\hat{\sigma}^2}}.$$

Ψευδομεταβλητές – Dummy (Binary) Variables

- Μια ψευδομεταβλητή d είναι μια μεταβλητή που παίρνει τιμή 0 ή 1
- Ουσιαστικά μια ψευδομεταβλητή εκφράζει την ύπαρξη ή όχι κάποιου (ποιοτικού) χαρακτηριστικού
- Για πχ το φύλο του ατόμου που απάντησε ένα ερωτηματολόγιο είναι άντρας ($d = 1$) ή γυναίκα ($d = 0$)
- Άλλο πχ έχει να κάνει με κλινικές δοκιμές, όπου κάποιοι ασθενείς λαμβάνουν τη νέα θεραπεία ($d = 1$) ενώ οι υπόλοιποι λαμβάνουν την υπάρχουσα/standard θεραπεία ($d = 0$)
- Σημειώνεται ότι η κατηγοριοποίηση δεν είναι μοναδική και ότι μια ψευδομεταβλητή μπορεί να ορισθεί με δυο τρόπους, για πχ
  - $d = 1$ για άντρα και $d = 0$ για γυναίκα, και
  - $d = 1$ για γυναίκα και $d = 0$ για άντρα

Παλινδρόμηση με Χρήση Ψευδομεταβλητών

Έστω το ακόλουθο μοντέλο που περιλαμβάνει μια συνεχή και μια δίτιμη (binary) μεταλητή (ψευδομεταβλητή)

$$y_i = \beta_0 + \beta_1 d_i + \beta_2 x_i + \epsilon_i.$$

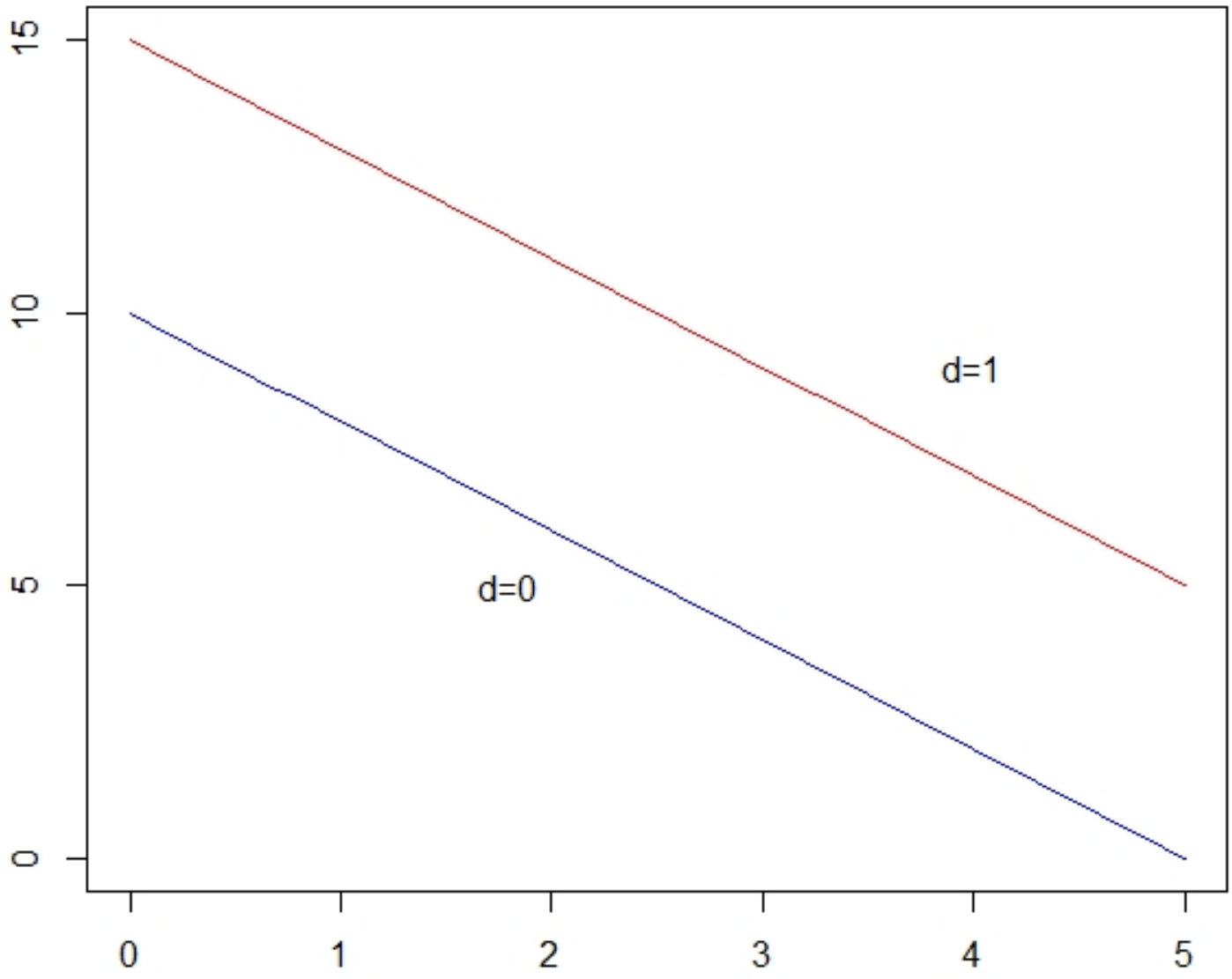
Όσον αφορά τη ψευδομεταβλητή, αυτή ορίζει τις εξής δύο περιπτώσεις:

- Αν $d = 0$, έχουμε
$$y_i = \beta_0 + \beta_2 x_i + \epsilon_i$$
- Ενώ αν $d = 1$, έχουμε
$$y_i = \beta_0 + \beta_1 + \beta_2 x_i + \epsilon_i$$

Αν δηλαδή είχαμε το μοντέλο

$$y = 10 + 5d - 2x,$$

τότε


15
10
5
0
0
1
2
3
4
5
d=1
d=0

Ερμηνεία

- Οι παρατηρήσεις χωρίζονται σε δύο ομάδες (group), σύμφωνα με την τιμή της μεταβλητής d (ανά φύλο, θεραπεία, κλπ)
- Η ομάδα με $d = 0$ ονομάζεται ομάδας αναφοράς (baseline group)
- Η παράμετρος β_1 της d εκφράζει (ποσοτικοποιεί) την αναμενόμενη επίδραση της ομάδας με $d = 1$ έναντι της ομάδας με $d = 0$, την ώρα που διατηρούμε όλες τις υπόλοιπες μεταβλητές (πχ. το x) σταθερές
- Συνεπώς, ο έλεγχος της μορφής

$$H_0 : \beta_1 = 0$$

ουσιατικά αντιπροσωπεύει την υπόθεση ότι η αναμενόμενη τιμή της y και στις δύο ομάδες είναι ίδια

Κατηγορικές Μεταβλητές (Categorical Variables)

- Μπορούμε να χρησιμοποιήσουμε ψευδομεταβλητές για να εισάγουμε στο μοντέλο μας κατηγορικές μεταβλητές με m-κατηγορίες (επίπεδα), όπου $m > 2$.
- Για κατηγορική μεταβλητή με m-κατηγορίες θα χρειαστούμε $m - 1$ ψευδομεταβλητές
- Έστω για παράδειγμα ότι η κατηγορική μεταβλητή είναι 'Εκπαίδευση' (*edu*), με επίπεδα

	edu
1	Γυμνάσιο
2	Λύκειο
3	Πανεπιστήμιο

και το ερώτημα είναι ποιά η επίδραση του επιπέδου εκπαίδευσης στο μισθό ενός εργαζομένου (Y).

- Εισάγοντας τη μεταβλητή *edu* απευθείας στο μοντέλο (ως συνεχής μεταβλητή)

$$y = \beta_0 + \beta_1 x_{edu} + \epsilon$$

σημαίνει ότι η επίδραση που θα έχει το Απολυτήριο Λυκείου σε σύγκριση με το Απολυτηρίου Γυμνασίου θα είναι ακριβώς η ίδια με αυτή που θα έχει το Πτυχίο Πανεπιστημίου σε σχέση με το Απολυτήριο Λυκείου.

- Για να παραστήσουμε τη μεταβλητή *edu* χρειαζόμαστε 2 ψευδομεταβλητές

	edu	d_1	d_2
1	Γυμνάσιο	0	0
2	Λύκειο	1	0
3	Πανεπιστήμιο	0	1

- Το επίπεδο αναφοράς (baseline) είναι αυτό που έχει όλες τις ψευδομεταβλητές ίσες με μηδέν (Γυμνάσιο)
- Αν $d_1 = 1$ τότε έχουμε απόφοιτο Λυκείου, αλλιώς 0
- Αν $d_2 = 1$ τότε έχουμε απόφοιτο Πανεπιστημίου, αλλιώς 0
- Είναι προφανές ότι δε μπορούμε να έχουμε $d_1 = 1$ και $d_2 = 1$ ταυτόχρονα
- Το μοντέλο παίρνει τη μορφή

$$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \epsilon.$$

- Εφόσον έχουμε *edu*=Γυμνάσιο (άρα $d_1 = d_2 = 0$), τότε

$$y = \beta_0 + \epsilon,$$

όπου το β_0 αντιστοιχεί στο επίπεδο αναφοράς

- Το β_1 είναι η επίδραση (effect) του Λυκείου σε σχέση με το Γυμνάσιο
$$y = \beta_0 + \beta_1 + \epsilon,$$
- Το β_2 είναι η επίδραση (effect) του Πανεπιστημίου σε σχέση με το Γυμνάσιο
$$y = \beta_0 + \beta_2 + \epsilon,$$
- Διάφοροι έλεγχοι ως προς τα β_1 και β_2 μπορούν να ελέγξουν διάφορες υποθέσεις.
- Ίδια είναι η ερμηνεία και στη παρουσία άλλων επεξηγηματικών μεταβλητών

Ψευδομεταβλητές Χρονικής Επίδρασης

- Είναι αρκετά συνηθισμένο η εξαρτημένη μεταβλητή να εξαρτάται από το χρόνο
- Ενδέχεται στη διάρκεια της μελέτης να συμβεί ένα 'γεγονός' (πχ. μια κρίση, αλλαγή πολιτικής ή θεραπείας, ένα διάλλειμα, κλπ)
- Η επίδραση του γεγονότος μπορεί να εκφραστεί στο μοντέλο με την εισαγωγή μιας ψευδομεταβλητής.
- Για πχ. έστω η μελέτη του επιπέδου ζακχάρου στο χρόνο y_t στο αίμα ασθενούς με ζαχαρώδη διαβήτη σε συνάρτηση με το σωματικό του βάρος x_t

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

- Μας ενδιαφέρει να μελετήσουμε την επίδραση που μπορεί να έχει η λήψη φαρμακευτικής αγωγής στο επίπεδο ζακχάρου στο αίμα. Έστω η ψευδομεταβλητή

$$d_t = \begin{cases} 0, & \text{όχι θεραπεία τη στιγμη } t \\ 1, & \text{θεραπεία τη στιγμη } t \end{cases}$$

περιγράφει το πότε ο ασθενής λαμβάνει θεραπεία

- Έχουμε λοιπόν το μοντέλο

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 d_t + \epsilon_t$$

το οποίο

  - για $d_t = 0$ παίρνει τη μορφή

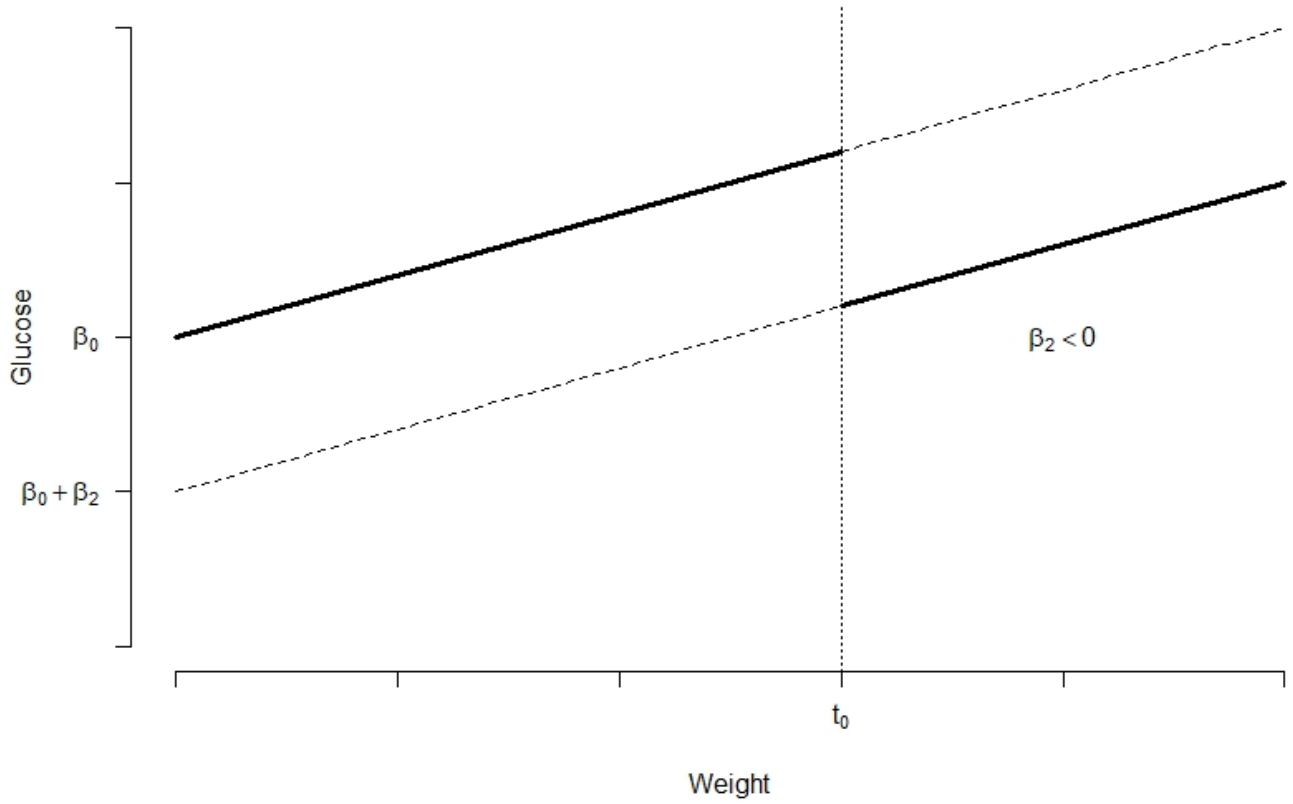
$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

  - για $d_t = 1$ παίρνει τη μορφή

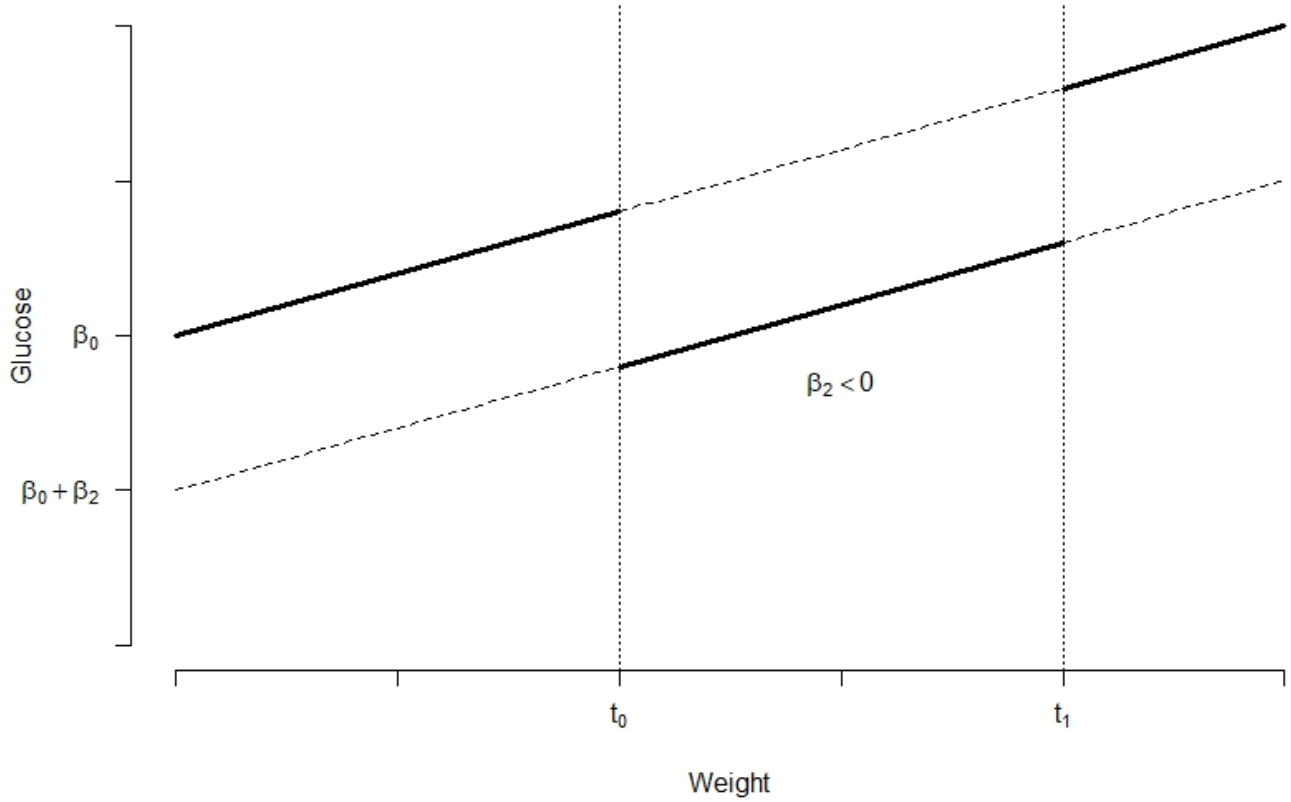
$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \beta_2 + \epsilon_t \\ &= (\beta_0 + \beta_2) + \beta_1 x_t + \epsilon_t \\ &= \beta_0^* + \beta_1 x_t + \epsilon_t \end{aligned}$$

όπου: $\beta_0^* = \beta_0 + \beta_2$.

- Είναι ξεκάθαρο ότι το πιο πάνω μοντέλο προκαλεί παράλληλη μετατόπιση του επιπέδου ζακχάρου του ασθενούς.
- Αν $d_t = 0$ έως τη χρονική στιγμή έναρξη θεραπείας t_0, έχουμε



- Αν $d_t = 1$ για διάστημα (t_0, t_1), έχουμε



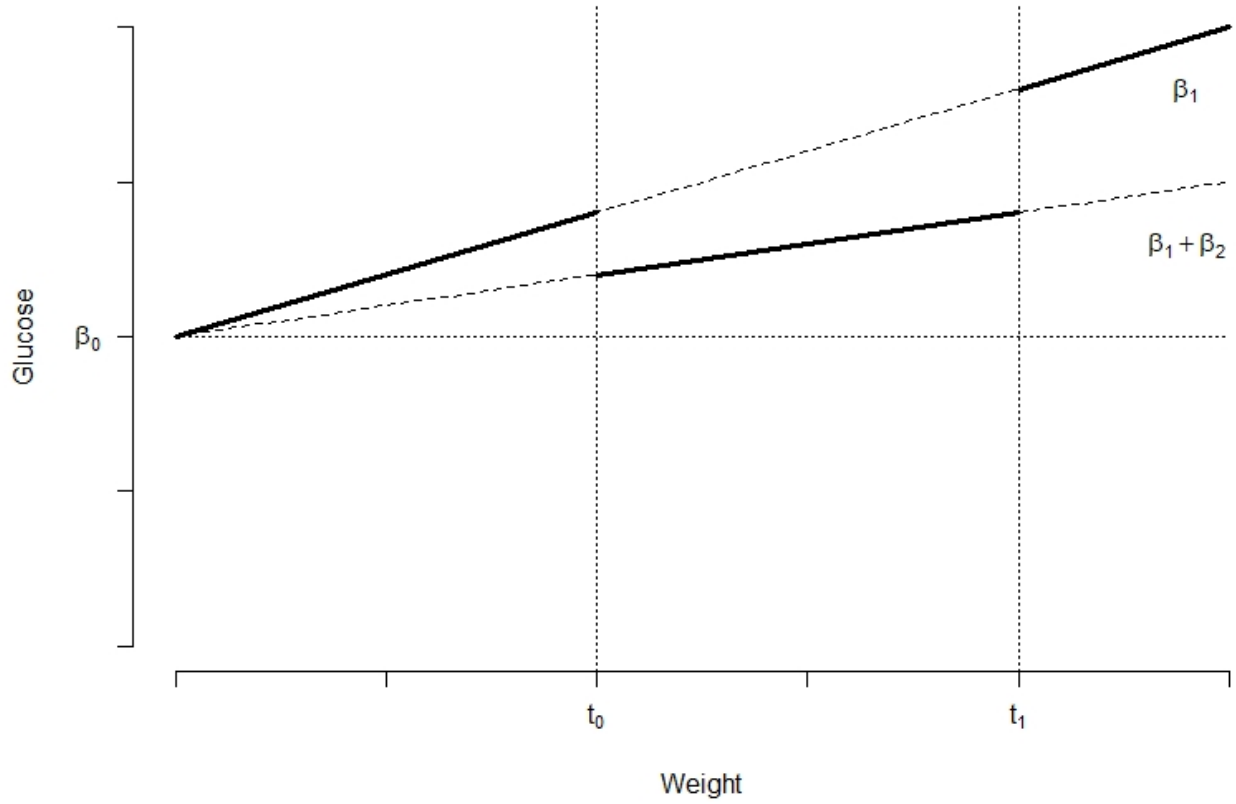
Μοντέλο με Αλληλεπίδραση (Interaction Term)

- Σε ένα γραμμικό μοντέλο, αλληλεπίδραση εμφανίζεται όταν μια ανεξάρτητη (επεξηγηματική) μεταβλητή έχει διαφορετική επίδραση στη μεταβλητή ενδιαφέροντος Y ανάλογα με τα επίπεδα/τιμές μιας άλλης ανεξάρτητης μεταβλητής
- Έστω το προηγούμενο πχ και έστω ότι η λήψη της θεραπείας δεν επιφέρει αλλαγή στο σταθερό όρο αλλά στο συντελεστή κλίσης.
- Με λίγα λόγια, έστω ότι έχουμε το μοντέλο

$$\begin{aligned} y_t &= \beta_0 + (\beta_1 + \beta_2 d_t)x_t + \epsilon_t \\ &= \beta_0 + \beta_1 x_t + \beta_2 d_t x_t + \epsilon_t \end{aligned}$$

- Ο συντελεστής κλίσης από β_1 αλλάζει σε $\beta_1 + \beta_2$ όταν $d_t = 1$

- Το $d_t x_t$ είναι ο όρος αλληλεπίδρασης (interaction term)
- Οπότε



Σύνθετο Μοντέλο

Έστω το μοντέλο που προκαλεί παράλληλη μετατόπιση και αλλαγή στη κλίση

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \beta_2 d_t + \beta_3 d_t x_t + \epsilon_t \\ &= \beta_0 + (\beta_1 + \beta_3 d_t) x_t + \beta_2 d_t + \epsilon_t \end{aligned}$$

Οπότε:

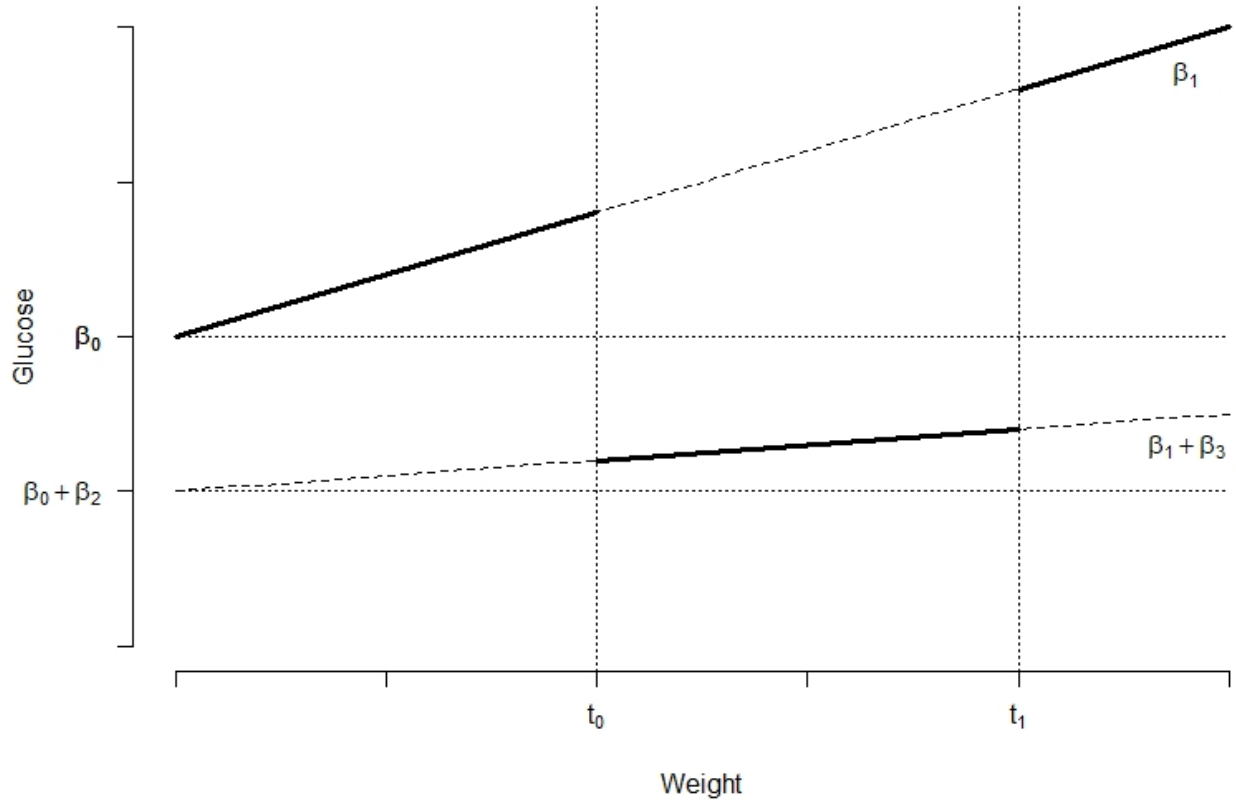
- Όταν $d_t = 0$ έχουμε

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

- και όταν $d_t = 1$ έχουμε

$$y_t = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_t + \epsilon_t$$

Οπότε



Επίδραση Παραγόντων

Έστω ότι θέλουμε να μελετήσουμε την επίδραση παραγόντων όπως το φύλο, προηγούμενη θεραπεία, μορφωτικό επίπεδο κ.α. Αυτό μπορεί να συμβεί εισάγοντας στο μοντέλο (ποιοτικές) ψευδομεταβλητές που εκφράζουν τα συγκεκριμένα χαρακτηριστικά.

Έστω για παράδειγμα ότι έχουμε τα ακόλουθα ποιοτικά χαρακτηριστικά

$$d_1 = \begin{cases} 1, & \text{άντρας} \\ 0, & \text{γυναίκα} \end{cases},$$

$$d_2 = \begin{cases} 1, & \text{μόρφωση επιπέδου τουλ. πανεπιστημίου} \\ 0, & \text{διαφορετικά} \end{cases},$$

$$d_3 = \begin{cases} 1, & \text{λήψη παλαιότερης θεραπείας} \\ 0, & \text{διαφορετικά} \end{cases}.$$

Το μοντέλο παίρνει τη μορφή

$$y_i = \beta_0 + \beta_1 d_{i1} + \beta_2 d_{i2} + \beta_3 d_{i3} + \epsilon_i, \quad i = 1, \ldots, n.$$

- Η εκτίμηση των παραμέτρων του μοντέλου γίνεται με ΕΕΤ (κατά τα γνωστά)
- Το $F - test$ ελέγχει την υπόθεση

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad vs \quad H_1 : \text{τουλάχιστον ένα} \neq 0$$

- Ένα χαρακτηριστικό με m-επίπεδα εκφράζεται στο μοντέλο με $(m-1)$-ψευδομεταβλητές
- Ο έλεγχος της στατιστικής σημαντικότητας μιας τέτοιας μεταβλητής γίνεται ελέγχοντας αν οι παράμετροι των ψευδομεταβλητών που εκφράζουν τη συγκεκριμένη κατηγορική μεταβλητή είναι ταυτόχρονα μηδέν ή τουλάχιστον μια είναι διάφορη του μηδενός (δες ειδική περίπτωση ελέγχων (Β) πιο πάνω)

Προβλέψεις

Έστω το μοντέλο

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_k x_{ik} + \epsilon_i, \quad i = 1, \ldots, n.$$

και έστω $X_0 = (1, x_{10}, x_{20}, \ldots, x_{k0})'$ το διάνυσμα τιμών των ανεξάρτητων μεταβλητών για τις οποίες θέλουμε να προβλέψουμε τη τιμή της εξαρτημένης y_0.

- Αληθινή τιμή του y_0: $Y_0 = X_0'\beta + \epsilon_0$
- Πρόβλεψη του y_0: $\hat{Y}_0 = X_0'\hat{\beta}$

Το σφάλμα πρόβλεψης είναι:

$$\hat{Y}_0 - Y_0 = X_0'(\hat{\beta} - \beta) - \epsilon_0.$$

Οπότε

$$\begin{aligned} E(\hat{Y}_0 - Y_0) &= E\left(X_0'\left(\hat{\beta} - \beta\right) - \epsilon_0\right) \\ &= X_0'\left(E\left(\hat{\beta}\right) - \beta\right) - E\left(\epsilon_0\right) \\ &= X_0'\left(\beta - \beta\right) \\ &= 0 \end{aligned}$$

Άρα, η $\hat{Y}_0$ είναι αμερόληπτη εκτιμήτρια της Y_0. Επίσης,

$$\begin{aligned} V(\hat{Y}_0 - Y_0) &= E\left[\left(\hat{Y}_0 - Y_0\right)^2\right] \\ &= E\left[\left(X_0'\left(\hat{\beta} - \beta\right) - \epsilon_0\right)^2\right] \\ &= E\left\{\left[X_0'\left(\hat{\beta} - \beta\right)\right]^2\right\} + E\left(\epsilon_0^2\right) - 2E\left[X_0'\left(\hat{\beta} - \beta\right)\epsilon_0\right] \\ &= \sigma^2 X_0'\left(X'X\right)^{-1} X_0 + \sigma^2 \\ &= \sigma^2\left[X_0'\left(X'X\right)^{-1} X_0 + 1\right] \end{aligned}$$

αφού

$$E\left\{\left[X_0'\left(\hat{\beta}-\beta\right)\right]^2\right\} = E\left[X_0'\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'X_0\right]$$
$$= X_0'V(\hat{\beta})X_0$$
$$= \sigma^2 X_0'\left(X'X\right)^{-1}X_0$$

και

$$E\left[X_0'\left(\hat{\beta}-\beta\right)\epsilon_0\right] = E\left[X_0'\left(X'X\right)^{-1}X'\epsilon\epsilon_0\right]$$
$$= X_0'\left(X'X\right)^{-1}X'E\left[\epsilon\epsilon_0\right]$$
$$= 0$$

αφού

$$\hat{\beta}-\beta = \left(X'X\right)^{-1}X'Y-\beta$$
$$= \left(X'X\right)^{-1}X'(X\beta+\epsilon)-\beta$$
$$= \beta+\left(X'X\right)^{-1}X'\epsilon-\beta$$
$$= \left(X'X\right)^{-1}X'\epsilon.$$

ΔΕ της Πρόβλεψης

Υποθέτουμε: $\epsilon_i \sim N(0, \sigma^2)$, ανεξάρτητα για $i = 1, 2, \ldots, n$.

Άρα το σφάλμα πρόβλεψης είναι: $\hat{Y}_0 - Y_0 = X_0'(\hat{\beta} - \beta) - \epsilon_0$
και ακολουθεί

$$\hat{Y}_0 - Y_0 \sim N(0, V(\hat{Y}_0 - Y_0)).$$

κανονική κατανομή ως γραμμικός συνδυασμός ανεξάρτητων κανονικών τυχαίων μεταβλητών. Άρα

$$\frac{\hat{Y}_0 - Y_0}{\sqrt{V(\hat{Y}_0 - Y_0)}} \sim N(0, 1).$$

Αν σ^2 άγνωστο:

$$\frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{\sigma}^2[X_0'(X'X)^{-1}X_0 + 1]}} \sim t_{n-k-1}.$$

Σημείωση: Το πιο στενό ΔΕ είναι για $X_0 = \bar{X}$, καθώς έτσι η $V(\hat{Y}_0 - Y_0)$ ελαχιστοποιείται.

Συντελεστής Προσδιορισμού (Coef. of Determination)

- Ο συντελεστής προσδιορισμού είναι

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

- Εκφράζει το ποσοστό της συνολικής μεταβλητότητας των y_i που ερμηνεύεται από τη παλινδρόμηση.
- Είναι ένα μέτρο ερμηνευτικής ικανότητας του μοντέλου.

Σημείωση: Σε μικρό δείγμα, αν ο αριθμός των ερμηνευτικών μεταβλητών αυξηθεί, τότε μπορεί η τιμή του R^2 να αυξηθεί πλασματικά (τείνει στη μονάδα), χωρίς αυτό να αντανακλά στη πραγματική ερμηνευτική αξία του μοντέλου. Αυτό οφείλεται στο μικρό αριθμό BE που είναι διαθέσιμοι για την εκτίμηση των παραμέτρων.

Προσαρμοσμένος Συντελεστής Προσδιορισμού (Adjusted Coefficient of Determination)

- Ο προσαρμοσμένος συντελεστής προσδιορισμού είναι

$$\begin{aligned} R^2_{Adj} &= 1 - \frac{n-1}{n-k-1}\frac{SSE}{SST} \\ &= 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} \\ &= 1 - \frac{MSE}{MST} < R^2 \end{aligned}$$

- Ο R^2_{Adj} λαμβάνει υπόψη του το πλήθος των παρατηρήσεων καθώς και το πλήθος των ανεξάρτητων μεταβλητών (δηλ. το πλήθος των παραμέτρων προς εκτίμηση). Είναι

$$R^2_{Adj} = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

- Αν n-μεγάλο και k-σταθερό, τότε $\frac{n-1}{n-k-1} \simeq 1$ και έτσι: $R^2_{Adj} \simeq R^2$.

- Για πολύ μικρές τιμές του R^2, ο R^2_{Adj} μπορεί να πάρει και αρνητικές τιμές
- Αν προσθέσουμε ανεξάρτητες μεταβλητές στο μοντέλο το R^2 πάντα αυξάνει ενώ το R^2_{Adj} όχι απαραίτητα
- Το $R^2(R^2_{Adj})$ μπορεί να χρησιμοποιηθεί ως κριτήριο για το ποιά ανεξάρτητη μεταβλητή μπορεί να εισαχθεί στο μοντέλο
- Προσθέτουμε μια μεταβλητή στο μοντέλο αν το R^2_{Adj} αυξάνει σημαντικά (κλιμακωτή παλινδρόμηση - stepwise regression)

Άσκηση

Έστω το μοντέλο

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n.$$

Έγινε εκτίμηση με τη μέθοδο ΕΤ με δείγμα $n = 20$ και έδωσε

$$\hat{\beta} = \begin{pmatrix} 0.97 \\ 0.70 \\ 1.78 \end{pmatrix}, \quad \widehat{V(\hat{\beta})} = \begin{pmatrix} 0.22 & 0.02 & -0.05 \\ & 0.05 & -0.03 \\ & & 0.04 \end{pmatrix} \quad \text{και} \quad \hat{\sigma}^2 = 2.52.$$

Προβλέψτε τη τιμή της Y στο σημείο $X_0 = (1, 0, 2)'$, βρείτε τη διακύμανση του σφάλματος πρόβλεψης και υπολογίστε το 95% ΔΕ (πρόβλεψης) για το Y_0.

Λύση:
Έχουμε

$$\hat{Y}_0 = X_0'\hat{\beta} = (1\ 0\ 2)\begin{pmatrix} 0.97 \\ 0.70 \\ 1.78 \end{pmatrix} = 4.53.$$

Επίσης έχουμε

$$\widehat{V(\hat{\beta})} = \hat{\sigma}^2(X'X)^{-1} \Rightarrow (X'X)^{-1} = \frac{\widehat{V(\hat{\beta})}}{\hat{\sigma}^2} \Rightarrow (X'X)^{-1} = \frac{1}{2.52}\widehat{V(\hat{\beta})}$$

Άρα

$$\widehat{V(\hat{Y}_0 - Y_0)} = \hat{\sigma}^2\left[X_0'(X'X)^{-1}X_0 + 1\right] = X_0'\widehat{V(\hat{\beta})}X_0 + \hat{\sigma}^2 = \ldots = 2.70.$$

Άρα, το 95 % ΔΕ είναι:

$$\hat{Y}_0 \pm t_{17,0.025}\sqrt{\widehat{V(\hat{Y}_0 - Y_0)}} = 4.53 \pm 2.11\sqrt{2.70} = [1.0633, 7.9967].$$

Εκτίμηση Μέγιστης Πιθανοφάνειας

Έχουμε το πολλαπλό γραμμικό μοντέλο

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(\mathbf{0}, \sigma^2 \mathtt{I}_n).$$

Θέτουμε: $\theta = (\beta_0, \beta_1, \beta_2, ..., \beta_k, \sigma^2)'$ το διάνυσμα των παραμέτρων.
Πιθανοφάνεια:

$$L(\theta) = (2\pi)^{-n/2} \left(\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\epsilon'\epsilon\right\}.$$

Λογαριθμίζοντας παίρνουμε

$$\begin{aligned}\ell(\theta) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\epsilon'\epsilon \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta).\end{aligned}$$

Παραγωγίζουμε ως προς θ και παίρνουμε το διάνυσμα των πρώτων μερικών παραγώγων της λογαριθμικής πιθανοφάνειας (log-likelihood)

$$\frac{\partial \ell(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell(\theta)}{\partial \beta} \\ \frac{\partial \ell(\theta)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial \ell(\theta)}{\partial \beta_0} \\ \frac{\partial \ell(\theta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial \ell(\theta)}{\partial \beta_k} \\ \frac{\partial \ell(\theta)}{\partial \sigma^2} \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{\sigma^2} X'(Y - X\beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}$$

το οποίο ονομάζεται διάνυσμα σκορ (score function): $u(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}$

Η λύση των πιο πάνω εξισώσεων δίνει

$$\begin{cases} \hat{\beta}_{ML} = (X'X)^{-1}X'Y \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{(ίδιος με ΕΕΤ)} \\ \hat{\sigma}^2_{ML} = \frac{1}{n}(Y - X\beta)'(Y - X\beta) = \frac{1}{n}\hat{\epsilon}'\hat{\epsilon} \quad \ldots\ldots \text{(όχι α.ε.)} \end{cases}$$

άρα: $\hat{\theta} = (\hat{\beta}'_{ML}; \hat{\sigma}^2_{ML})'$.

Για να είναι εκτιμητές μέγιστης πιθανοφάνειας θα πρέπει να ικανοποιούνται οι συνθήκες δευτέρας τάξης για μέγιστο. Θα πρέπει λοιπόν ο πίνακας δευτέρων παραγώγων της λογαριθμικής πιθανοφάνειας (Hessian) να είναι αρνητικά ορισμένος. Έχουμε:

$$H(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta' \partial \theta} = \begin{pmatrix} \frac{\partial^2 \ell(\theta)}{\partial \beta' \partial \beta} & \frac{\partial^2 \ell(\theta)}{\partial \sigma^2 \partial \beta} \\ \frac{\partial^2 \ell(\theta)}{\partial \beta' \partial \sigma^2} & \frac{\partial^2 \ell(\theta)}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix}$$

Τα στοιχεία του πίνακα $H(\theta)$ βρίσονται παίρνοντας παραγώγους του διανύσματος σκορ

$$\frac{\partial^2 \ell(\theta)}{\partial \beta' \partial \beta} = \frac{\partial}{\partial \beta}\left(\frac{1}{\sigma^2} X'(Y - X\beta)\right) = -\frac{1}{\sigma^2} X'X$$

$$\frac{\partial^2 \ell(\theta)}{\partial \sigma^2 \partial \beta} = \frac{\partial^2 \ell(\theta)}{\partial \beta' \partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma^2} X'(Y - X\beta)\right) = -\frac{1}{\sigma^4} X'\epsilon$$

$$\frac{\partial^2 \ell(\theta)}{\partial \sigma^2 \partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left(-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta)\right)$$

$$= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\epsilon'\epsilon.$$

Επομένως, ο Hessian πίνακας παίρνει τη μορφή

$$H(\theta) = \begin{pmatrix} -\frac{1}{\sigma^2}X'X & -\frac{1}{\sigma^4}X'\epsilon \\ -\frac{1}{\sigma^4}X'\epsilon & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\epsilon'\epsilon \end{pmatrix},$$

και ο οποίος είναι αρνητικά ορισμένος.

Αυτό σημαίνει ότι για κάθε $t \in \mathbb{R}^{k+1}$ με $t \neq 0$ έχουμε

$$t'H(\theta)t < 0$$

(παραλείπεται).

Πίνακας Πληροφορίας του Fisher

Ο πίνακας παίρνει τη μορφή

$$J(\theta) = E[-H(\theta)] = E\left[-\begin{pmatrix} -\frac{1}{\sigma^2}X'X & -\frac{1}{\sigma^4}X'\epsilon \\ -\frac{1}{\sigma^4}X'\epsilon & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\epsilon'\epsilon \end{pmatrix}\right]$$
$$= \begin{pmatrix} \frac{1}{\sigma^2}X'X & \mathbf{0} \\ \mathbf{0} & \frac{n}{2\sigma^4} \end{pmatrix}$$

καθώς $E(X'\epsilon) = \mathbf{0}$ και $E(\epsilon'\epsilon) = n\sigma^2$.

Ο αντίστροφος παίρνει τη μορφή

$$J(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\sigma^4}{n} \end{pmatrix}$$

όπου

$$V(\hat{\beta}_{ML}) = \sigma^2(X'X)^{-1}$$

και $\frac{2\sigma^4}{n}$ αποτελεί την ασυμπτωτική διακύμανση του $\hat{\sigma}^2_{ML}$.

Συμπεράσματα:

- Ο πίνακας $J(\theta)^{-1}$ αποτελεί το κατώτερο όριο της ανισότητας Cramer-Rao
- Ο εκτιμητής $\hat{\beta}_{ML}$ αποτελεί α.ε. του β και έχει διακύμανση $V(\hat{\beta}_{ML}) = \sigma^2(X'X)^{-1}$ που είναι το κάτω όριο Cramer-Rao. Συνεπώς είναι ένας αποτελεσματικός εκτιμητής!
- Ο εκτιμητής $\hat{\sigma}^2_{ML}$ δεν είναι α.ε. του σ^2.
- Το κάτω όριο Cramer-Rao για τον αμερόληπτο εκτιμητή του σ^2 είναι $\frac{2\sigma^4}{n}$. Αυτό δεν επιτυγχάνεται ούτε για τον α.ε. $\hat{\sigma}^2 = \frac{1}{n-k-1}\sum \hat{\epsilon}_i^2$ με

$$V(\hat{\sigma}^2) = \frac{2\sigma^4}{n-k-1} > \frac{2\sigma^4}{n}$$

- Ισχύει

$$J(\theta) = E\left[-H(\theta)\right] = E\left[-\frac{\partial^2 \ell(\theta)}{\partial\theta'\partial\theta}\right]$$
$$= E\left[-\left(\frac{\partial \ell(\theta)}{\partial\theta}\right)\left(\frac{\partial \ell(\theta)}{\partial\theta}\right)'\right] = E\left[u(\theta)u(\theta)'\right].$$

Ασυμπτωτικές Ιδιώτητες του ΕΜΠ

- Συνέπεια (ασυμπτωτική διακύμανση ισούται με το *LB CR* για α.ε.)
- Ασυμπτωτική αποτελεσματικότητα
- Ασυμπτωτική κανονικότητα

Ασυμπτωτική κατανομή του $\hat{\theta}_{ML}$

Έχουμε:

$$\hat{\theta}_{ML} \sim N(\theta, AsymV(\hat{\theta}_{ML}))$$

όπου:

$$AsymV(\hat{\theta}_{ML}) = J(\theta)^{-1} = \{E[-H(\theta)]\}^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\sigma^4}{n} \end{pmatrix}.$$

Επομένως, εκτιμητής του $AsymV(\hat{\theta}_{ML})$ είναι

$$[-H(\theta)]^{-1} = \begin{pmatrix} \hat{\sigma}^2_{ML}(X'X)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\hat{\sigma}^4_{ML}}{n} \end{pmatrix},$$

όπου $\hat{\sigma}^2_{ML}$ είναι συνεπής εκτιμητής του σ^2.

Σημείωση: Οι ασυμπτωτικές ιδιότητες του ΕΜΠ $\hat{\theta}_{ML}$ είναι χρήσιμες για να βρεθούν οι κατανομές στατιστικών κριτηρίων για τον έλεγχο υποθέσεων γραμμικών περιορισμών της μορφής

$$H_0 : R\beta = r$$

και κατ' επέκταση για τη σύγκριση μοντέλων.

Likelihood Ratio Test

Το κριτήριο του λόγου πιθανοφανειών ελέγχει τη μέγιστη τιμή της λογαριθμικής πιθανοφάνειας για το μοντέλο χωρίς περιορισμούς (unrestricted) με εκείνη του μοντέλου με περιορισμούς (restricted)

$$H_0 : R\beta = r.$$

Έχουμε:

$$\begin{aligned}
\ell(\hat{\theta}_{ML}) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\hat{\sigma}^2_{ML}) - \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{2\hat{\sigma}^2_{ML}} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n}\right) - \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{2\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n}} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sum_{i=1}^n \hat{\epsilon}_i^2\right) + \frac{n}{2}\ln(n) - \frac{n}{2} \\
&= -\frac{n}{2}\left[\ln(2\pi) + 1 - \ln(n)\right] - \frac{n}{2}\ln\left(\sum_{i=1}^n \hat{\epsilon}_i^2\right).
\end{aligned}$$

Έστω τώρα $\hat{\theta}^*_{ML}$ ο ΕΜΠ του μοντέλου με περιορισμούς. Είναι:

$$\ell_{RE}(\hat{\theta}^*_{ML}) = -\frac{n}{2}\big[\ln(2\pi) + 1 - \ln(n)\big] - \frac{n}{2}\ln\left(\sum_{i=1}^{n}\hat{\epsilon}_i^{*2}\right).$$

Συνεπώς, το κριτήριο λόγου πιθανοφανειών παίρνει τη μορφή

$$LR = -2\left[\ell_{RE}(\hat{\theta}^*_{ML}) - \ell_{UN}(\hat{\theta}_{ML})\right],$$

το οποίο παίρνει πάντα θετικές τιμές καθώς: $\ell_{RE}(\hat{\theta}^*_{ML}) < \ell_{UN}(\hat{\theta}_{ML})$.
Οπότε:

$$LR = n\left[\ln\left(\sum_{i=1}^{n}\hat{\epsilon}_i^{*2}\right) - \ln\left(\sum_{i=1}^{n}\hat{\epsilon}_i^{2}\right)\right] = n\ln\left(\frac{\hat{\epsilon}^{*\prime}\hat{\epsilon}^*}{\hat{\epsilon}'\hat{\epsilon}}\right),$$

όπου κάτω από την H_0 έχουμε: $LR \sim \chi^2_q$, καθώς $n \to \infty$.
Οι ΒΕ q καθορίζονται από τους περιορισμούς που προκύπτουν από την H_0.

Wald Test

Έστω:

$$H_0 : R\beta = r.$$

Από τα πιο πάνω σχετικά με τη κατανομή του $\hat{\beta}_{ML}$ προκύπτει ότι

$$R\hat{\beta}_{ML} - r \sim N\left(R\beta - r, RAsymV(\hat{\beta}_{ML})R'\right).$$

Κάτω από την H_0 έχουμε

$$R\hat{\beta}_{ML} - r \sim N\left(\mathbf{0}, RAsymV(\hat{\beta}_{ML})R'\right),$$

όπου

$$(R\hat{\beta}_{ML} - r)' \left[RAsymV(\hat{\beta}_{ML})R'\right]^{-1} (R\hat{\beta}_{ML} - r) \sim \chi^2_q.$$

Το κριτήριο Wald παίρνει τη μορφή:

$$W = \frac{(R\hat{\beta}_{ML} - r)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta}_{ML} - r)}{\hat{\sigma}^2_{ML}} \sim \chi^2_q, \quad \text{για} \quad n \to \infty.$$

Είναι:

$$R\hat{\beta}_{ML} - r)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta}_{ML} - r) = \hat{\epsilon}^{*'}\hat{\epsilon}^* - \hat{\epsilon}'\hat{\epsilon}$$

και

$$\hat{\sigma}^2_{ML} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n}.$$

Άρα:

$$W = \frac{\hat{\epsilon}^{*'}\hat{\epsilon}^* - \hat{\epsilon}'\hat{\epsilon}}{\hat{\epsilon}'\hat{\epsilon}/n}.$$

Akaike Information Criterion (AIC)

- Το AIC (Akaike,1973) γενικά θεωρείτε το πρώτο κριτήριο για επιλογή μοντέλου (model selection)
- Ακόμα και σήμερα αποτελεί ένα από τα πιο γνωστά κριτήρια
- Ένα μοντέλο ποτέ δεν πρόκειται να παραστήσει (εκτιμήσει) τον ακριβή τρόπο (διαδικασία) με τον οποίο γεννήθηκαν τα παρατηρούμενα δεδομένα
- Η χρήση λοιπόν του μοντέλου έχει ως αποτέλεσμα την απώλεια πληροφορίας
- Το AIC εκτιμά τη σχετική απώλεια πληροφορίας του υποψήφιου μοντέλου
- Όσο μικρότερη η απώλεια πληροφορίας ενός μοντέλου τόσο καλύτερη η ποιότητα του

Έστω ότι έχουμε τα μοντέλα $M_1, M_2, \ldots, M_r$, όπου κάθε μοντέλο είναι μια κατανομή ή μίξη κατανομών

$$M_j = \{p(y; \theta_j) : \theta_j \in \Theta_j\}.$$

- Έστω τα δεδομένα $y_1, y_2, \ldots, y_n$ από μια κατανομή $f()$ η οποία δεν είναι μια από τις πιο πάνω κατανομές
- Έστω $\hat{\theta}_j$ η ΕΜΠ του μοντέλου j
- Μια εκτίμηση της $p()$ βασισμένη στο μοντέλο j είναι:

$$\hat{p}_j(y) = p(y; \hat{\theta}_j)$$

- Η ποιότητα της $p(y; \hat{\theta}_j)$ ως εκτιμήτρια της $f()$ μπορεί να μετρηθεί μέσω της Kullback-Leibler distance

$$\begin{aligned} K(p, \hat{p}_j) &= \int p(y) \log \frac{p(y)}{\hat{p}_j(y)} dy \\ &= \int p(y) \log p(y) dy - \int p(y) \log \hat{p}_j(y) dy \end{aligned}$$

- Ο πρώτος όρος δεν εξαρτάται από το j
- Για να κάνουμε το $K(p, \hat{p}_j)$ ελάχιστο ως προς j αρκεί να κάνουμε μέγιστο το

$$K_j = \int p(y) \log p(y; \hat{\theta}_j) dy$$

- Θα πρέπει λοιπόν να εκτιμήσουμε το K_j και διαισθητικά ένας εκτιμητής θα μπορούσε να είναι

$$\overline{K}_j = \frac{1}{n} \sum_{i=1}^{n} \log p(y; \hat{\theta}_j) = \frac{\ell_j(\hat{\theta}_j)}{n}$$

όπου $\ell_j(\hat{\theta}_j)$ είναι η συνάρτηση λογαριθμικής πιθανοφάνειας του μοντέλου j

- Παρόλα αυτά, η πιο πάνω εκτίμηση θεωρείται εξαιρετικά μεροληπτική (biased) επειδή τα δεδομένα χρησιμοποιούνται δύο φορές
  - μια για τον υπολογισμό των ΕΜΠ και
  - μια για την εκτίμηση του ολοκληρώματος
- Ο Akaike έδειξε ότι η μεροληψία είναι περίπου

$$\frac{d_j}{n} \quad \text{όπου} \quad d_j = dimension(\Theta_j)$$

είναι η διάσταση του μοντέλου j

- Οπότε

$$\hat{K}_j = \frac{\ell_j(\hat{\theta}_j)}{n} - \frac{d_j}{n} = \overline{K}_j - \frac{d_j}{n}$$

- Ορίζουμε λοιπόν

$$AIC(j) = -2n\hat{K}_j = -2\ell_j(\hat{\theta}_j) + 2d_j$$

- Το να μεγιστοποιήσουμε το $\hat{K}_j$ είναι το ίδιο με το να ελαχιστοποιήσουμε το $AIC(j)$
- Ο πολλαπλασιασμός με μια σταθερά $(-2n)$ δεν αλλάζει κάτι στο τρόπο σκέψης και λειτουργίας (γίνεται για ιστορικούς λογους (;))
- Το AIC χρησιμοποιείται ευρέως μιας και απαιτεί μόνον τις ασυμπτωτικές ιδιότητες της ΕΜΠ
- Μπορεί να χρησιμοποιηθεί για τον έλεγχο non-nested μοντέλων
- Μπορεί να χρησιμοποιηθεί για τον έλεγχο μοντέλων που προέρχονται από διαφορετικές κατανομές
- Το βέλτιστο μοντέλο χαρακτηρίζεται από τη μικρότερη τιμή του AIC

Bayesian Information Criterion (BIC)

- Το *BIC* (Stone, 1979) είναι ένα κριτήριο επιλογής μοντέλου αντίστοιχο του *AIC*
- Ορίζεται ως

$$BIC = -2\ell_j(\hat{\theta}_j) + 2\log(n)d_j$$

όπου η ποινή (penalty) είναι πιο αυστηρή από αυτή του *AIC* μιας και εξαρτάται και από το n

- Η λογική βασίζεται στο Bayesian τρόπο σκέψη, όπου τόσο το θ_j χαρακτηρίζεται από μια prior κατανομή όσο και το μοντέλο M_j έχει μια prior πιθανότητα να είναι στο σωστό μοντέλο
- Μετά από πράξεις και μερικά αναπτύγματα Taylor (...) προκύπτει ο πιο πάνω τύπος για το *BIC*
- Η χρήση του είναι επίσης αντίστοιχη με αυτή του *AIC*, όπου μικρότερες τιμές είναι επιθυμητές και χαρακτηρίζουν το βέλτιστο μοντέλο

Mallow's C_p Criterion

Το Mallow's C_p κριτήριο ελέγχει τη προσαρμογή μοντέλων με διαφορετικό πλήθος παραμέτρων. Έχει τη μορφή:

$$C_p = \frac{SSE(p)}{\sigma^2} - N + 2p$$

Άν το μοντέλο(p) είναι σωστό τότε το C_p θα είναι κοντά ή μικρότερο από p. Αυτό ισχύει γιατί όταν το πραγματικό μοντέλο έχει p παραμέτρους, τότε μπορεί να δειχθεί ότι ισχύει

$$E(C_p) = p$$

Συνεπώς ένα απλό γράφημα των μοντέλων έναντι του p μπορεί να μας οδηγήσει στο βέλτιστο μοντέλο.
Ένα πρόβλημα με το C_p είναι να βρούμε εκτίμηση του σ^2 για όλα τα p. Ως εκτιμήτρια χρησιμοποιούμε το S^2, το MSE από το μοντέλο με όλες τις k διαθέσιμες επεξηγηματικές μεταβλητές (regressors).

References

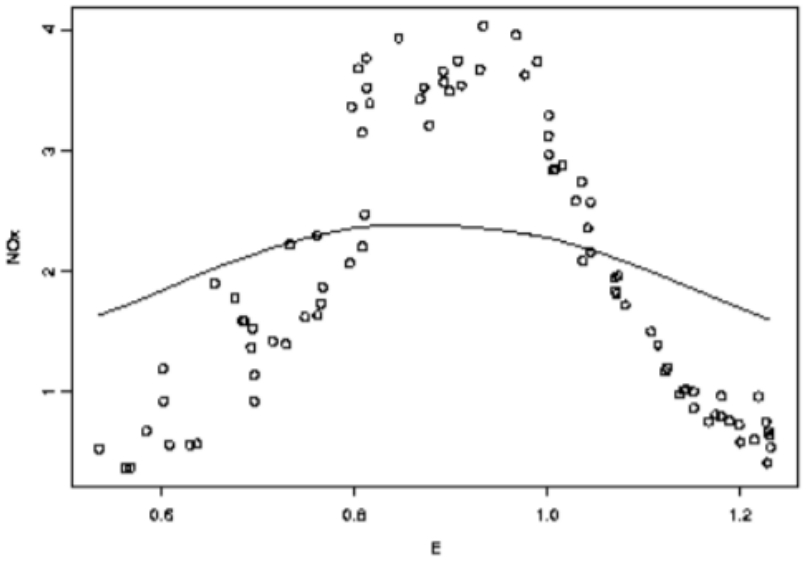
- Akaike Hirotugu (1973). "Information Theory and an Extension of the Maximum Likelihood Principle" (in: B. N. Petrov and F. Csaki, eds., 2nd International Symposium on Information Theory, Akademia Kiado, Budapest, pp. 267–281)
- Stone, M. (1979). "Comments on Model Selection Criteria of Akaike and Schwartz." Journal of the Royal Statistical Society B 41:276–278.

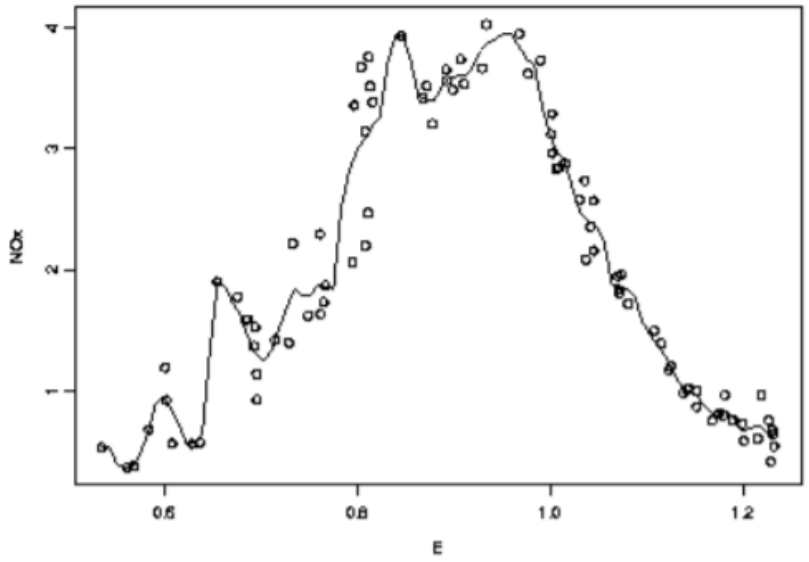
Model Selection in Linear Models

Το πρόβλημα που αντιμετωπίζουμε έχει να κάνει με την επιλογή ενός μοντέλου ανάμεσα σε όλα τα πιθανά (διαθέσιμα) μοντέλα παλινδρόμησης.
Αν λοιπόν:

- το μοντέλο είναι πολύ *μικρό*, τότε θα έχει φτωχή προγνωστική αξία, μεγάλη μεροληψία, μικρή διακύμανση και γενικά θα είναι ακατάλληλο για τα δεδομένα
- το μοντέλο είναι πολύ *μεγάλο*, τότε θα κάνει 'overfit' τα δεδομένα, θα έχει φτωχή προγνωστική αξία, μικρή μεροληψία αλλά μεγάλη διακύμανση
- το μοντέλο είναι *σωστό*, τότε θα έχει ισοροπία μεταξύ μεροληψίας και διακύμανσης και θα έχει καλή προγνωστική αξία

Μεροληψία *vs* Μεταβλητότητα





Διαφορά μεταξύ

- ενός *μικρού* μοντέλου με μεγάλη μεροληψία και μικρή διακύμανση (αριστερά) και
- ενός *μεγάλου* μοντέλου με μιρκή μεροληψία αλλά μεγάλη διακύμανση (δεξιά).

- Έστω ότι έχουμε το μοντέλο M_0 χωρίς κανέναν προγνωστικό παράγοντα (null model)
- Αν στο μοντέλο μας συμπεριλάβουμε $w = 1, 2, \ldots, k$ από τα διαθέσιμα X, τότε έχουμε $\binom{k}{w}$ πιθανά μοντέλα
- Σκοπός είναι να επιλέξουμε το καλύτερο βασιζόμενοι σε κάποια τακτική και φυσικά σε κάποια κριτήρια
  (R^2, R^2_{Adj} AIC, BIC, $p - value$, κλπ)

Στρατηγικές Επιλογής Μοντέλου

- Best Subset: αναζητούμε όλα τα πιθανά μοντέλα και επιλέγουμε αυτό με το μεγαλύτερο R^2 (ή το μικρότερο AIC/BIC)
- Stepwise (forward, backward or both): χρήσιμη στρατηγική όταν το πλήθος των επεξηγηματικών μεταβλητών είναι μεγάλο.
  Μπορούμε να επιλέξουμε ένα αρχικό μοντέλο και να είμαστε greedy
- Greedy: διαλέγουμε το μοντέλο που προσφέρει το μεγαλύτερο άλμα ως προς το κριτήριο μας (πχ R^2)

Forward Selection

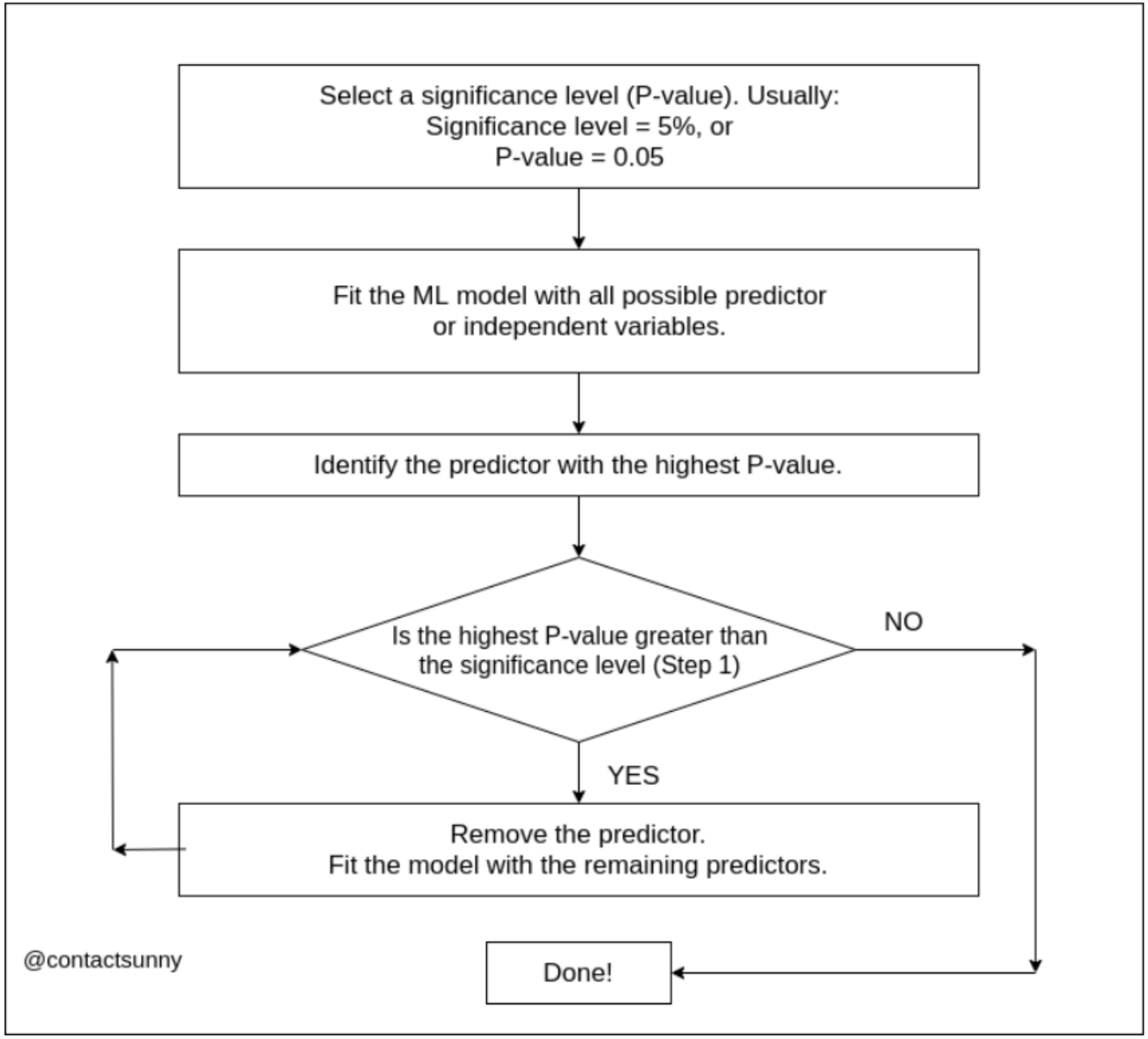
Είναι μια καλή στρατηγική μιας και είναι εύκολα εφαρμόσιμη και δίνει μια καλή αλληλουχία μοντέλων.
Έστω λοιπόν ότι έχουμε k επεξηγηματικές μεταβλητές (X)

- Ξεκινάμε από το null model (το μοντέλο μόνο με το σταθερό όρο)
- Τρέχουμε k απλά γραμμικά μοντέλα, ένα για κάθε μια από τις ανεξάρτητες μεταβλητές με το σταθερό όρο
- Διαλέγουμε αυτή τη μεταβλητή που προκαλεί τη μεγαλύτερη αλλαγή στο κριτήριο επιλογής και την εισάγουμε στο μοντέλο
- Από τις υπόλοιπες $k-1$ μεταβλητές, διαλέγουμε αυτή με την δεύτερη μεγαλύτερη αλλαγή στο κριτήριο επιλογής και την εισάγουμε στο μοντέλο
- Συνεχίζουμε με τις υπόλοιπες μεταβλητές μέχρι να ικανοποιηθεί κάποιος κανόνας για να σταματήσουμε, πχ η μεταβλητή να έχει $p-value$ μεγαλύτερο από κάποια συγκεκριμένη τιμή

Backward Elimination

Ακολουθούμε τα εξής βήματα:

1) Επιλέγουμε ένα κριτήριο (πχ. p-value) καθώς και το επίπεδο του κριτηρίου πάνω στο οποίο θα επιλέγουμε τις μεταβλητές που θα αφαιρεθούν από το μοντέλο (συνήθως το 5%)
2) Προσαρμόζουμε το μοντέλο με όλες τις διαθέσιμες μεταβλητές
3) Εντοπίζουμε τη μεταβλητή με το μεγαλύτερο p-value
4) Αν $p > 0.05$ τότε αφαιρούμε τη συγκεκριμένη μεταβλητή από το μοντέλο. Αν όμως $p < 0.05$, τότε καμία μεταβλητή δε βγαίνει από το μοντέλο και η διαδικασία ολοκληρώνεται (Βήμα 6)
5) Αφού αφαιρεθεί η μεταβλητή από το Βήμα 4, προσαρμόζουμε το μοντέλο με τις υπόλοιπες μεταβλητές και πάμε στο Βήμα 3
6) Όταν δεν υπάρχει πια άλλη μεταβλητή να βγει από το μοντέλο, η διαδικασία ολοκληρώνεται και έχουμε καταλήξει στο βέλτιστο μοντέλο (optimal model)



(Source: https://towardsdatascience.com)

Μια χρήσιμη και πολύ συνηθισμένη παραλλαγή του πιο πάνω Βήματος 2 είναι η ακόλουθη:

- Προσαρμόζουμε τα k απλά γραμμικά μοντέλα (μια μεταβλητή τη φορά)
- Επιλέγουμε αυτές που είναι στατιστικά σημαντικές, συχνά αυτές με $p < 0.10$
- Στη συνέχεια προσαρμόζουμε το μοντέλο μόνο με τις στατιστικά σημαντικές μεταβλητές
- Συνεχίζουμε στο Βήμα 3 της πιο πάνω διαδικασίας

Εκθετική Οικογένεια Κατανομών (ΕΟΚ)

- Οι παρατηρήσεις μας μπορούν να ακολουθούν κατανομή που ανήκει στην ΕΟΚ (Exponential Family of Distributions), και όχι απαραίτητα τη κανονική (μπορεί να είναι και διακριτή)
- Συνεπώς η τ.μ. Y ακολουθεί:

$$Y \sim f(y; \theta, \phi)$$

- Η $f(y; \theta, \phi)$ θα ανήκει στην ΕΟΚ με κανονική παράμετρο (natural parameter) τη θ και παράμετρο όχλησης (nuisance parameter) το ϕ αν
  - μπορεί να γραφεί ως

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{\alpha(\phi)} + c(y, \phi)\right\}$$

    για πραγματικές συναρτήσεις $\alpha(\phi), b(\theta)$ και $c(y, \phi)$, και
  - το στήριγμα (support) της Y δεν εξαρτάται από τα θ και ϕ.

- Αρχικά το ϕ το θεωρούμε γνωστό!
- Έχουμε λοιπόν μια μονοπαραμετρική ΕΟΚ (με παράμετρο θ)
- Αναλύεται σε ένα γινόμενο παραγόντων:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta}{\alpha(\phi)}\right\}\exp\left\{\frac{-b(\theta)}{\alpha(\phi)}\right\}\exp\{c(y,\phi)\}$$

και μπορεί να θεωρήθεί ως πιθανοφάνεια από μια και μοναδική παρατήρηση

- Έχουμε:

$$\log f(y;\theta,\phi) = \frac{y\theta - b(\theta)}{\alpha(\phi)} + c(y,\phi),$$

οπότε η συνάρτηση σκορ (score function) γίνεται

$$u(\theta) = \frac{\partial}{\partial\theta}\log f(y;\theta,\phi) = \frac{y - b'(\theta)}{\alpha(\phi)}$$

- Συνεπώς

$$E\left[u(\theta)\right] = 0 \;\Rightarrow\; E\left[\frac{y - b'(\theta)}{\alpha(\phi)}\right] = 0 \;\Rightarrow \underline{\mu = E[y] = b'(\theta)}$$

- Ο πίνακας των δευτέρων παραγώγων (Hessian) είναι

$$H(\theta) = \frac{\partial}{\partial\theta} u(\theta) = -\frac{b''(\theta)}{\alpha(\phi)}$$

- και η πληροφορία του Fisher

$$I(\theta) = E\left[-H(\theta)\right] = \frac{b''(\theta)}{\alpha(\phi)}$$

- Έχουμε λοιπόν

$$V\left[u(\theta)\right] = E\left[u^2(\theta)\right] - E\left[u(\theta)\right]^2 = E\left[u^2(\theta)\right]$$
$$= -E\left[H(\theta)\right] = I(\theta) = \frac{b''(\theta)}{\alpha(\phi)} \tag{2}$$

- Επίσης

$$V\left[u(\theta)\right] = V\left[\frac{y - b'(\theta)}{\alpha(\phi)}\right] = \frac{V(y)}{\alpha(\phi)^2} \tag{3}$$

- Συνεπώς, από (2) και (3) έχουμε

$$\frac{V(y)}{\alpha(\phi)^2} = I(\theta) = \frac{b''(\theta)}{\alpha(\phi)} \;\Rightarrow\; \underline{V(y) = \alpha(\phi)b''(\theta)}$$

- Αφού λοιπόν

$$\mu = b'(\theta) \;\Rightarrow\; \theta = b'^{(-1)}(\mu)$$

(αντίστροφη συνάρτηση), τότε

$$V(y) = \alpha(\phi) b''(b'^{(-1)}(\mu)),$$

όπου $b''(b'^{(-1)}(\mu))$ είναι συνάρτηση του μέσου

- Η διακύμανση λοιπόν είναι συνάρτηση του μέσου (μ) (και του ϕ)
- Αν το ϕ είναι γνωστό, τότε η οικογένεια κατανομών είναι γνωστή ως EOK (Exponential Family of Distributions)
- Αν το ϕ είναι άγνωστο, τότε η οικογένεια κατανομών είναι γνωστή ως Exponential-Dispersion Family of Distributions και η παράμετρος ϕ είναι γνωστή ως dispersion parameter. Συνεπώς

  θ : Canonical parameter

  ϕ : Dispersion Parameter

Συναρτήσεις Σύνδεσμος (Link Functions)

Στο γραμμικό μοντέλο έχουμε

$$\mu(x;\beta) = E(Y|X = x) = x'\beta$$

όπου συνδέουμε την αναμενόμενη τιμή της Y με το

$$\eta = x'\beta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i,$$

το οποίο ονομάζουμε συστηματική συνιστώσα (linear predictor).

- Στο γραμμικό μοντέλο δεν είχαμε την ανάγκη να θέτουμε περιορισμούς μιας και η $\mu(x;\beta)$ παίρνει τιμές στο $\mathbb{R}$
- Για κάποιους τύπους δεδομένων όμως θα χρειαστεί να εισάγουμε κάποιους περιορισμούς
- Για πχ, αν $Y \sim Bernoulli(p)$ με $E(Y) = p$ και $0 \leq p \leq 1$, τότε θα πρέπει να περιορίσουμε και το $\mu(x;\beta)$ στο $[0, 1]$
- Όμως το $\eta = x'\beta$ παίρνει τιμές στο $\mathbb{R}$
- Απαιτείται λοιπόν ένας "τρόπος" (πχ μια συνάρτηση...ένας μετασχηματισμός) για να συνδέσει το $\mu(x;\beta)$ με το $\eta = x'\beta$
- Αναζητούμε $g()$ τέτοια ώστε

$$g(\mu) = \eta$$

Γενικευμένο Γραμμικό Μοντέλο (Generalized Linear Model)

Στο Γενικευμένο Γραμμικό Μοντέλο (GLM) έχουμε

- Τυχαία συνιστώσα, η οποία δίνει πληροφορία για τη μορφή της κατανομής της Y, και
- Συστηματική συνιστώσα, η οποία δίνει τη σχέση των χαρακτηριστικών που μελετάμε και της κατανομής αυτής (σχέση μεταξύ $X_{1i}, ..., X_{ki}$ και των Y_i, κυρίως των $E(Y_i)$)

Τυχαία Συνιστώσα

Οι $Y_1, Y_2, ..., Y_n$ είναι ανεξάρτητες και ανήκουν στην ίδια ΕΟΚ

$$f(y_i; \theta_i, \phi_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\alpha(\phi_i)} + c(y_i, \phi_i)\right\}$$

Επειδή είναι ανεξάρτητες, η από κοινού ισούται με το γινόμενο

$$f(\mathbf{y}; \theta, \phi) = \prod_{i=1}^{n} f(y_i; \theta_i, \phi_i) = \exp\left\{\sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\alpha(\phi_i)} + \sum_{i=1}^{n} c(y_i, \phi_i)\right\}$$

όπου

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix} \quad \text{και} \quad \phi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_n \end{pmatrix}$$

Συστηματική Συνιστώσα (linear predictor)

Στο απλό γραμμικό μοντέλο έχουμε

$$\underbrace{E(Y_i)}_{\mu_i} = \underbrace{\beta_0 + \beta_1 x_i}_{\eta_i}$$

Στο GLM υπάρχει συνάρτηση σύνδεσμος $g()$ για την οποία ισχύει

$$g(\mu_i) = \eta_i$$

- Υποθέτουμε ότι η $g()$ είναι αμφιμονοσήμαντη (΄1-1΄) και παραγωγίσιμη
- Είναι

$$g : \underbrace{\ldots}_{\mu_i} \to \underbrace{\mathbb{R}}_{\eta_i}$$

- Το πεδίο ορισμού της $g()$ εξαρτάται από την υπόθεση για τη κατανομή των Y_i
- Το πεδίο ορισμού της $g()$ πρέπει να είναι τέτοιο που να ισούται ή τουλάχιστον να περιέχει το σύνολο των $E(Y_i)$, για Y_i που ανήκει στην οικογένεια την οποία υποθέτουμε
- ΠΧ1: $Y_i \sim Poisson(\lambda) \Rightarrow \mu_i = E(Y_i) = \lambda > 0$
- ΠΧ2: $Y_i \sim Bin(n, p) \Rightarrow \mu_i = E(Y_i) = np, \quad p \in (0, 1)$
- Το $g(\mu_i)$ είναι πάντα μια γραμμική συνάρτηση των X_i

Γενικά, έχουμε παρατηρήσεις για n άτομα

$$\begin{array}{cccc} y_1 & x_{11} & \dots & x_{1k} \\ y_2 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ y_n & x_{n1} & \dots & x_{nk} \end{array}$$

ΠΧ: Έστω δεδομένα από $n = 100$ άτομα, και

$$Y_i = \begin{cases} 0, & \text{εργαζόμενος} \\ 1, & \text{άνεργος} \end{cases}$$

μια μεταβλητή που εκφράζει την εργασιακή κατάσταση του ατόμου i. Έστω π_i η πιθανότητα $P(y_i = 1)$ και

$$g(\pi_i) = \alpha + \beta x_i.$$

Αντ΄ αυτού, έστω μια κρυφή μεταβλητή (δε τη παρατηρούμε)

$$Y^* : \text{Απασχολισημότητα}$$

με

$$y_i^* = \alpha + \beta x_i + \epsilon_i,$$

όπου τα ϵ_i ακολουθούν μια τυχαία κατανομή που χαρακτηρίζεται από αθροιστική συνάρτηση κατανομής $F()$.

Υπόθεση: Αν $y_i^* < c$ τότε $y_i = 1$. Οπότε:

$$\begin{aligned} \pi_i = P(y_i = 1) = P(y_i^* < c) &= P(\alpha + \beta x_i + \epsilon_i < c) \\ &= P(\epsilon_i < c - \alpha - \beta x_i) \\ &= F(c - \alpha - \beta x_i) \end{aligned}$$

Συνεπώς, αφού Y_i ακολουθεί *Bernoulli* έχουμε:

$$\pi_i = E(Y_i) = \mu_i = F(\eta_i)$$

Άρα

$$\eta_i = F^{-1}(\mu_i),$$

όπου $F^{-1}()$ είναι η συνάρτηση σύνδεσμος $(g())$ η οποία εξαρτάται από την υπόθεση για την κατανομή των ϵ_i.

Κανονική Συνδετική Συνάρτηση (Canonical Link Function)

Έστω τ.δ. $y = (y_1, y_2, ..., y_n)$ και

$$g(\mu_i) = \eta_i = x_i'\beta$$

- Μας ενδιαφέρει η σχέση των y_i και x_i την οποία περιγράφει το β
- Έχουμε την από κοινού

$$f(\mathbf{y}; \theta, \phi) = \prod_{i=1}^{n} f(y_i; \theta_i, \phi_i) = \exp\left\{ \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\alpha(\phi_i)} + \sum_{i=1}^{n} c(y_i, \phi_i) \right\}$$

- Κάνουμε αναπαραμετροποίηση έτσι ώστε να μπορέσουμε να εκφράσουμε το θ ως συνάρτηση του β.
- Έχουμε για το θ_i ότι

$$\mu_i = b'(\theta_i)$$

καθώς επίσης

$$\eta_i = g(\mu_i) = x_i'\beta$$

- Οπότε

$$g\left(b'(\theta_i)\right) = x_i'\beta \quad \Rightarrow \quad \theta_i = b'^{-1}\left(g^{-1}(x_i'\beta)\right) = G(x_i'\beta)$$

- Αν λοιπόν διαλέξουμε

$$g = b'^{-1} \quad \text{δηλ.} \quad g(\mu_i) = b'^{-1}(\mu_i)$$

τότε

$$G(x_i'\beta) = b'^{-1}\left(b'(x_i'\beta)\right) = x_i'\beta$$

- Αυτή ονομάζεται κανονική συνδετική συνάρτηση (Canonical Link)
- Γενικά

$$\log f(\mathbf{y}; \beta, \phi) = \sum_{i=1}^{n} \left[\frac{y_i x_i'\beta - b(x_i'\beta)}{\alpha(\phi_i)} + c(y_i, \phi_i) \right]$$

ΚΑΝΟΝΙΚΕΣ ΣΥΝΔΕΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ

	Normal	*Poisson*(λ)	*Bernoulli*(p)
θ	$\theta = \mu$	$\theta = \log(\lambda)$	$\theta = \log\left(\frac{p}{1-p}\right)$
$b(\theta)$	$\frac{1}{2}\mu^2$	$\exp(\theta) = \lambda$	$\log(1 + e^{\theta})$
$\mu = b'(\theta)$	$\theta = \mu$	$\exp(\theta) = \lambda$	$\frac{e^{\theta}}{1+e^{\theta}} = p$
$g = b'^{-1}(\mu)$	μ	$log(\mu)$	$\log \frac{\mu}{1-\mu} = \log \frac{p}{1-p} = logit(p)$
	Ταυτοτική	log	*logistic* ή *logit*

Εκτίμηση Μέγιστης Πιθανοφάνειας

Έχουμε τη log-likelihood για canonical link

$$\log f(\mathbf{y}; \beta, \phi) = \sum_{i=1}^{n} \left[\frac{y_i x_i' \beta}{\alpha(\phi_i)} - \frac{b(x_i' \beta)}{\alpha(\phi_i)} + c(y_i, \phi_i) \right]$$

Παίρνουμε

$$\frac{\partial}{\partial \beta} \log f(\mathbf{y}; \beta, \phi) = 0 \quad \Leftrightarrow \quad u(\hat{\beta}) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} \left[\frac{y_i x_i'}{\alpha(\phi_i)} - \frac{b'(x_i' \beta) x_i'}{\alpha(\phi_i)} \right] = 0$$

- Το σύστημα των εξισώσεων που προκύπτει δεν έχει αναλυτική λύση
- Λύνεται μόνον αριθμητικά, με τη χρήση του αλγορίθμου Newton-Raphson ή με τη Μέθοδο Fisher Scoring (ταυτίζονται για κανονικό link)
- Hessian:

$$H(\beta) = \frac{\partial^2}{\partial\beta\partial\beta'} \log f(\mathbf{y}; \beta, \phi) = \frac{\partial u(\beta)}{\partial\beta} = -\sum_{i=1}^{n} \left[\frac{x_i b''(x_i'\beta) x_i'}{\alpha(\phi_i)} \right]$$

- Fisher Information:

$$I(\beta) = E[-H(\beta)] = \sum_{i=1}^{n} \left[\frac{x_i b''(x_i'\beta) x_i'}{\alpha(\phi_i)} \right] = X'WX$$

όπου $X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$, $W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & w_n \end{pmatrix}$ και $w_i = \frac{b''(x_i'\beta)}{a(\phi_i)}$

Αλγόριθμος Newton-Raphson/Fisher Scoring

(δες γραφήματα)

- Επαναληπτική μέθοδος που επιτυγχάνει αριθμητική επίλυση του συστήματος

$$u(\beta) = 0,$$

ανανεώνοντας επαναληπτικά τη τρέχουσα εκτίμηση β^j σε β^{j+1} με

$$\beta^{j+1} = \beta^j - H(\beta^j)^{-1} u(\beta^j)$$

- Στη πράξη αντί για $H(\beta)$ χρησιμοποιούμε $E[H(\beta)] = -I(\beta)$
- Για κανονικό link:

$$E[H(\beta)] = H(\beta)$$

άρα Newton-Raphson και Fisher Scoring ταυτίζονται.

- Έχουμε

$$\beta^{j+1} = \beta^j + I(\beta^j)^{-1} u(\beta^j)$$

όπου $I(\beta^j) = X' W^j X$ και $u(\beta^j) = \sum_{i=1}^{n} \left[\frac{(y_i - b'(x_i' \beta^j)) x_i'}{\alpha(\phi_i)} \right]$

Αλγόριθμος

B. 1: Ξεκινάμε με μια αρχική τιμή β^0 για το $\hat{\beta}$

B. 2: Στην επανάληψη j του αλγορίθμου ανανεώνουμε το β^j σε

$$\beta^{j+1} = \beta^j + I(\beta^j)^{-1} u(\beta^j)$$

B. 3: Συνεχίζουμε έως ότου

$$||\beta^{j+1} - \beta^j|| < \epsilon$$

όπου ϵ μια προκαθορισμένη πολύ μικρή τιμή

Συμπερασματολογία για το β

- Για μεγάλο n έχουμε

$$\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta} \sim N(\mathbf{0}, I(\beta)^{-1}) \;\Rightarrow\; \hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta} \sim N(\mathbf{0}, (X' W_{\underset{\sim}{\beta}} X)^{-1})$$

  όπου το $W_{\underset{\sim}{\beta}}$ εξαρτάται από το πραγματικό $\underset{\sim}{\beta}$.

- Το Δ.Ε. για το β_i βασίζεται στο ότι $\frac{\hat{\beta}_i - \beta_i}{\sqrt{\left[X' W_{\underset{\sim}{\beta}} X\right]_{ii}}} \sim N(0,1)$, οπότε

$$\beta_i \in \hat{\beta}_i \pm 1.96 \sqrt{\left[X' W_{\underset{\sim}{\beta}} X\right]_{ii}}$$

- Επειδή δε γνωρίζουμε το πραγματικό $\underset{\sim}{\beta}$ προσεγγίζουμε το $W_{\underset{\sim}{\beta}}$ με μια εκτιμήτρια

$$W_{\underset{\sim}{\hat{\beta}}} \rightarrow \hat{W}$$

- Συνεπώς

$$\beta_i \in \hat{\beta}_i \pm 1.96\sqrt{\left[X'\hat{W}X\right]_{ii}}$$

Σύγκριση Μοντέλων: Γενικευμένος έλεγχος λόγου πιθανοφανειών

- Σύγκριση μοντέλων όπου το ένα (υπό την H_0) είναι εμφωλευμένο (nested) στο εναλλακτικό (υπό την H_1). Η ασυμπτωτική κατανομή της ελεγχοσυνάρτησης του ελέγχου είναι χ^2
- Για τα GLM, εμφωλευμένα μοντέλα σημαίνει ότι οι H_0 και H_1 είναι
  - είναι μοντέλα βασισμένα στην ίδια ΕΟΚ
  - έχουν την ίδια συνδετική συνάρτηση (link), και
  - οι επεξηγηματικές μεταβλητές στο H_0 είναι υποσύνολο αυτών του μοντέλου H_1

Έστω ότι το μοντέλο H_1 περιέχει p επεξηγηματικές μεταβλητές και το H_0 ένα υποσύνολο με $q < p$ από αυτές. Χωρίς βλάβη της γενικότητας, έστω ότι έχουμε

$$H_0 : \text{μοντέλο με } \eta_i = \sum_{j=1}^{p} x_{ij}\beta_j, \qquad i = 1, 2, \ldots, n$$

$$H_1 : \beta_{q+1} = \beta_{q+2} = \ldots = \beta_p = 0$$

Το σύνολο τιμών των κανονικών παραμέτρων θ κάτω από την H_0 $\left(\Theta^{(0)}\right)$ είναι υποσύνολο αυτού κάτω από την H_1 $\left(\Theta^{(1)}\right)$

Η στατιστική συνάρτηση λόγου πιθανοφανειών για τον έλεγχο της H_0 έναντι της H_1 είναι:

$$L_{01} = 2\log\frac{\max_{\theta\in\Theta^{(1)}} f_Y(y;\theta)}{\max_{\theta\in\Theta^{(0)}} f_Y(y;\theta)} = 2\log f_Y(y;\hat{\theta}^{(1)}) - 2\log f_Y(y;\hat{\theta}^{(0)})$$

όπου $\hat{\theta}^{(0)}$ και $\hat{\theta}^{(1)}$ τα διανύσματα κανονικών παραμέτρων που προκύπτουν από τις σχέσεις $b'(\hat{\theta}_i) = \hat{\mu}_i$ και $g(\hat{\mu}_i) = \hat{\eta}_i = \sum x_{ij}\hat{\beta}_j$.

ΣΗΜΕΙΩΣΗ: Έχουμε υποθέσει ότι τα $\alpha(\phi_i), \quad i = 1, 2, \ldots, n$ είναι γνωστά.

Ασυμπτωτικά λοιπόν ισχύει: $L_{01} \sim \chi^2_{(p-q)}$, οπότε απορίπτουμε την H_0 σε ε.σ.σ. α αν

$$L_{01} > \chi^2_{(p-q),\alpha}$$

Απόκλιση μοντέλου (Scaled deviance)

Έστω ένα μοντέλο όπου το β είναι n-διάστατο. Δηλ. έστω X ένας $(n \times n)$ αντιστρέψιμος πίνακας σχεδιασμού και $\eta = X\beta$. Το μοντέλο αυτό έχει τόσες παραμέτρους όσες παρατηρήσεις και μπορεί να παραμετροποιηθεί ισοδύναμα ως προς $\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$ και $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$.

Για link function $g()$ έχουμε:

$$\theta = b'^{-1}\left(g^{-1}(X\beta)\right).$$

Οπότε:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = X^{-1} \begin{pmatrix} g(b'(\theta_1)) \\ \vdots \\ g(b'(\theta_n)) \end{pmatrix}.$$

Για κανονικό link έχουμε:

$$\theta = X\beta \Rightarrow \beta = X^{-1}\theta$$

.

Το μοντέλο αυτό ονομάζεται κορεσμένο (saturated).

Στο κορεσμένο μοντέλο μπορούμε να υπολογίσουμε τις ΕΜΠ $\hat{\theta}$ κατευθείας από τη συνάρτηση πιθανοφάνειας, χωρίς πρώτα να βρούμε τα β. Έχουμε:

$$\log f_Y(y;\theta) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + \sum_{i=1}^{n} c(y_i,\phi_i)$$

όπου

$$\frac{\partial}{\partial\theta_k}\log f_Y(y;\theta) = \frac{y_k - b'(\theta_k)}{a(\phi_k)} = 0, \quad k = 1,2,\ldots,n$$

$$\Rightarrow b'(\hat{\theta}_k) = y_k \;\Rightarrow \hat{\mu}_k = y_k, \quad k = 1,2,\ldots,n \;\Rightarrow\; \hat{\theta}_k = b'^{-1}(y).$$

Από τα πιο πάνω

- Προκύπτει ότι οι παράμετροι είναι οι ίδιες οι παρατηρήσεις
- Το κορεσμένο μοντέλο είναι υπέρ-παραμετροποιημένο και δε χρησιμοποιείται στη πράξη
  (αρ. παραμέτρων = αρ. παρατηρήσεων)
- Στο $N(\theta_i, \sigma^2)$ μοντέλο το SST είναι μηδέν
- Είναι όμως πολύ χρήσιμο για συγκρίσεις καθώς κάθε μοντέλο είναι εμφωλευμένο στο κορεσμένο μοντέλο
- Η σύγκριση ενός μοντέλου με το κορεσμένο χρησιμεύει ως έλεγχος καλής προσαρμογής. Έχουμε την ελεγχοσυνάρτηση

$$L_{0S} = 2 \log f_Y(y; \hat{\theta}^{(S)}) - 2 \log f_Y(y; \hat{\theta}^{(0)})$$

όπου το $\hat{\theta}^{(S)}$ προκύπτει από τη σχέση $b'(\hat{\theta}) = \hat{\mu} = y$ και το $\hat{\theta}^{(0)}$ είναι συνάρτηση του ΕΜΠ του β.

- Είναι

$$L_{0S} = 2\sum_{i=1}^{n} \frac{y_i\left[\hat{\theta}_i^{(S)} - \hat{\theta}_i^{(0)}\right] - \left[b(\hat{\theta}_i^{(S)}) - b(\hat{\theta}_i^{(0)})\right]}{a(\phi_i)}$$

- Η $D^* = L_{0S}$ καλείται απόκλιση του μοντέλου H_0 από το κορεσμένο (scaled deviance), ενώ για κοινό ϕ έχουμε την (unscaled) deviance

$$D = \alpha(\phi)D^* = 2\sum_{i=1}^{n}\left(y_i\left[\hat{\theta}_i^{(S)} - \hat{\theta}_i^{(0)}\right] - \left[b(\hat{\theta}_i^{(S)}) - b(\hat{\theta}_i^{(0)})\right]\right)$$

- Κάτω από την H_0 η scaled deviance (όπως και η unscaled deviance όταν πχ στα κανονικά δεδομένα έχουμε $\alpha(\phi) = \phi = 1$) ασυμπτωτικά ακολουθεί

$$D^* = L_{0S} \sim \chi^2_{(n-q)}$$

για $(n-q)$ περιορισμούς για τα β (συνήθως παραμετροι ίσοι με μηδέν)

- Για γνωστά $a(\phi_i)$ η L_{0S} μπορεί να υπολογιστεί από τα δεδομένα.

Παρατηρήσεις:

- Ένα μοντέλο με μεγάλη Deviance δεν προσαρμόζει καλά τα δεδομένα! Όταν η χ^2 προσέγγιση είναι ακριβής, τότε τη χρησιμοποιήσουμε για έλεγχο καλής προσαρμογής (goodness-of-fit)
- Για τη σύγκριση δύο μοντέλων H_0 και H_1 έχουμε

$$L_{01} = 2\log f_Y(y;\hat{\theta}^{(1)}) - 2\log f_Y(y;\hat{\theta}^{(0)}) = L_{0S} - L_{1S} \ \sim \ \chi^2_{(p-q)}$$

- Εναλλακτικός έλεγχος καλής προσαρμογής ενός μοντέλου είναι ο χ^2 του Pearson

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i^{(0)})^2}{\widehat{Var}(Y_i)}.$$

Κάτω από την H_0 έχουμε

$$X^2 \sim \chi^2_{(n-q)}$$

δηλ. κάτω από την υπόθεση ότι έχουμε $(n-q)$ περιορισμούς για τα β (συνήθως παραμετροι ίσοι με μηδέν)

ΠΧ: Γραμμικό Μοντέλο

Έχουμε

$$y_i = x_i'\beta + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_i^2) \text{ και } \sigma_i^2 \text{ γνωστή.}$$

Επίσης,

- $g(\mu_i) = \mu_i$ (ταυτοτική συνδετική συνάρτηση),
- $b(\theta_i) = \frac{1}{2}\theta_i^2$,
- $\alpha(\phi_i) = \sigma_i^2$ και
- $\mu_i = \theta_i = \eta_i = x_i'\beta$.

Τέλος, για το κορεσμένο μοντέλο έχουμε: $\mu_i = \theta_i = y_i$.
Οπότε:

$$\log f_Y(y; \theta, \phi) = \log \prod_{i=1}^{n} f_Y(y; \theta_i, \phi_i) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\alpha(\phi_i)} + \sum_{i=1}^{n} c(y_i, \phi_i)$$

Άρα

$$L_{0S} = 2\sum_{i=1}^{n} \frac{y_i\left(\hat{\theta}_i^{(S)} - \hat{\theta}_i^{(0)}\right) - \left(b(\hat{\theta}_i)^{(S)} - b(\hat{\theta}_i)^{(0)}\right)}{\alpha(\phi_i)}$$

$$= 2\sum_{i=1}^{n} \frac{y_i\left(y_i - \hat{\theta}_i^{(0)}\right) - \left(\frac{1}{2}y_i^2 - \frac{1}{2}\hat{\theta}_i^{(0)^2}\right)}{\sigma_i^2}$$

$$= 2\sum_{i=1}^{n} \frac{y_i\left(y_i - x_i'\hat{\beta}_0\right) - \left(\frac{1}{2}y_i^2 - \frac{1}{2}(x_i'\hat{\beta}_0^2)^2\right)}{\sigma_i^2} = \ldots =$$

$$= \sum_{i=1}^{n} \frac{\left(y_i - x_i'\hat{\beta}_0\right)^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}_i^{(0)}\right)^2}{Var(y_i)}$$

- Στο γραμμικό μοντέλο, το L_{0S} ταυτίζεται με τη στατιστική συνάρτηση X^2 του Pearson
- Αν: $\sigma_i^2 = \sigma^2$ τότε:

$$L_{0S} = \frac{SSE}{\sigma^2}$$

ΠΧ: $y_i \sim Poisson(\lambda_i)$

Έχουμε,

- $\theta_i = \log(\lambda_i)$,
- $b(\theta_i) = \exp(\theta_i) = \exp\{\log(\mu_i)\}$,
- $\mu_i = b'(\theta_i) = \exp(\theta_i)$ και
- $Var(y_i) = \alpha(\phi_i)\mu_i = \mu_i$, αφού $\alpha(\phi_i) = 1$ γνώστό

Τέλος, για το κορεσμένο μοντέλο έχουμε: $\mu_i = \theta_i = y_i$.
Οπότε:

$$L_{0S} = 2\sum_{i=1}^{n}\left[y_i\left(\log\hat{\mu}_i^{(S)} - \log\hat{\mu}_i^{(0)}\right) - \left(\hat{\mu}_i^{(S)} - \hat{\mu}_i^{(0)}\right)\right]$$

$$= 2\sum_{i=1}^{n}\left[y_i \log\frac{y_i}{\hat{\mu}_i^{(0)}} - \left(y_i - \hat{\mu}_i^{(0)}\right)\right]$$

όπου: $\hat{\mu}_i^{(0)} = g^{-1}(x_i'\hat{\beta}_0) = \exp\{x_i'\hat{\beta}_0\}$ το canonical link

- Το X^2 του Pearson (μέτρο απόκλισης μεταξύ του μοντέλου μας και του κορεσμένου) παίρνει τη μορφή

$$X^2 = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}_i^{(0)}\right)^2}{\hat{Var}(y_i)} = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}_i^{(0)}\right)^2}{\hat{\mu}_i^{(0)}}$$

- Ένα πρόβλημα που μπορεί να προκύψει είναι το εξής: Να έχουμε μια παρατήρηση ίση με μηδέν, στη οποία περίπτωση ο λογάριθμος $\log \frac{y_i}{\hat{\mu}_i^{(0)}}$ δεν ορίζεται. Σε αυτή τη περίπτωση μια λύση θα είναι να εφαρμόσουμε τον έλεγχο X^2 του Pearson

Μοντέλα με άγνωστο ϕ

- Κάνουμε τη ρεαλιστική υπόθεση: $\alpha(\phi_i) = \frac{\sigma^2}{m_i}$,
  όπου σ^2 είναι άγνωστο και m_i γνωστά
- Με λίγα λόγια, δε γνωρίζουμε τη διακύμανση αλλά οι διακυμάνσεις θα είναι ίσες ($m_i = 1$) ή τουλάχιστον θα διαφέρουν κατά γνωστή ποσότητα (m_i)
- Τότε

$$L_{0S} = \frac{2}{\sigma^2} \sum_{i=1}^{n} \left(m_i y_i \left[\hat{\theta}_i^{(S)} - \hat{\theta}_i^{(0)} \right] - m_i \left[b(\hat{\theta}_i^{(S)}) - b(\hat{\theta}_i^{(0)}) \right] \right) = \frac{D_{0S}}{\sigma^2}$$

  όπου D_{0S} είναι η deviance και L_{0S} είναι η scaled deviance
- Το L_{0S} γράφεται ως μια ποσότητα που μπορεί να υπολογιστεί (D_{0S}) προς την άγνωστη ποσότητα σ^2

Σύγκριση Μοντέλων

Έχουμε

$$H_0 : \text{μοντέλο με } p \text{ παραμέτρους}$$
$$H_1 : \text{μοντέλο με } q \text{ παραμέτρους}$$

Η ελεγχοσυνάρτηση παίρνει τη μορφή

$$F = \frac{\frac{1}{q-p}(D_{0S} - D_{1S})}{\frac{1}{n-q}D_{1S}} = \frac{\frac{1}{q-p}\frac{D_{0S}-D_{1S}}{\sigma^2}}{\frac{1}{n-q}\frac{D_{1S}}{\sigma^2}}$$
$$= \frac{\frac{1}{q-p}(L_{0S} - L_{1S})}{\frac{1}{n-q}L_{1S}} \sim \frac{\frac{1}{q-p}\chi^2_{q-p}}{\frac{1}{n-q}\chi^2_{n-q}} = \mathcal{F}_{q-p,n-q}$$

Σημείωση: Η ελεγχοσυνάρτηση δεν εξαρτάται από την άγνωστη παράμετρο σ^2 και έχει γνωστή κατανομή

Εκτίμηση του σ^2

Γνωρίζουμε ότι

$$L_{0S} = \frac{D_{0S}}{\sigma^2} \sim \chi^2_{n-p}$$

Οπότε:

$$E(L_{0S}) = n - p \;\Rightarrow\; E\left(\frac{D_{0S}}{\sigma^2}\right) = n - p \;\Rightarrow\; \hat{\sigma}^2 = \frac{D_{0S}}{n-p}$$

η οποία είναι α.ε. του σ^2

Κατάλοιπα - Deviance residuals

- Για το γραμμικό μοντέλο έχουμε $SSE = \sum \hat{\epsilon}_i^2$, όπου $\epsilon_i = y_i - \hat{\mu}_i^{(0)}$ και

$$X^2 = \sum_i \frac{\left(y_i - \hat{\mu}_i^{(0)}\right)^2}{\sigma^2} = \frac{SSE}{\sigma^2}$$

- Στα GLM έχουμε

$$D_{0S} = 2\sum_{i=1}^{n} \left(y_i \left[\hat{\theta}_i^{(S)} - \hat{\theta}_i^{(0)} \right] - \left[b(\hat{\theta}_i^{(S)}) - b(\hat{\theta}_i^{(0)}) \right] \right)$$

οπότε τα Deviance Residuals είναι

$$r_i = \sqrt{\left| y_i \left[\hat{\theta}_i^{(S)} - \hat{\theta}_i^{(0)} \right] - \left[b(\hat{\theta}_i^{(S)}) - b(\hat{\theta}_i^{(0)}) \right] \right|} \times \underbrace{(\text{πρόσημο καταλοίπων})}_{y_i - \hat{\mu}_i^{(0)}}$$

Ερμηνεία των παραμέτρων β

Διωνυμικά Δεδομένα (Binomial Data)

Άλλα Link Functions

Poisson Regression

Παρατηρούμε $y_i \in \mathbb{N}$ (counts ή αρ. εμφανίσεων γεγονότων στη μονάδα) και έστω D_1 και D_2 έστω κατηγορικές μεταβλητές με $D_1 \in \{1, 2, \ldots, I\}$ και $D_2 \in \{1, 2, \ldots J\}$, όπου I και J δηλώνουν τα επίπεδα των D_1 και D_2 αντίστοιχα.

Υπόθεση:

$$y_i \sim Poisson(\mu_i)$$

όπου

$$f(y_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad \mu_i > 0$$

και

$$E(y_i) \;=\; V(y_i) \;=\; \mu_i.$$

Από την ΕΟΚ έχουμε: $\theta_i = \log(\mu_i)$, $b(\theta_i) = \exp(\theta_i)$ και $\alpha(\phi_i) = 1$.

Κανονικό link:

$$g(\mu_i) = \log(\mu_i) = \eta_i = x_i'\beta$$

που μας δίνει ένα log-linear μοντέλο

Ερμηνεία παραμέτρων: Αν D κατηγορική μεταλητή με J επίπεδα και έστω $\beta = (\beta_1, \beta_2, \ldots, \beta_{J-1})$ οι παράμετροι που αντιστοιχούν στις ψευδομεταβλητές που εκφράζουν την D, τότε

β_j: η αλλαγή του $\log(\mu_i)$ όταν η D πάρει τη τιμή j συγκρινόμενο με $D = 1$ (αθροιστική)

$\exp(\beta_j)$: η πολλαπλασιαστική αλλαγή του μ_i όταν η D πάρει τη τιμή j συγκρινόμενο με $D = 1$

Γενικά, είναι η αλλαγή που επιφέρεται στο $\log(\mu_i)$ (αθροιστική) ή στο μ_i (πολλαπλασιαστική) στη μοναδιαία αλλάγη της επεξηγηματικής μεταβλητής, με τις υπόλοιπες μεταβλητές να παραμένουν σταθερές

Πιθανοφάνεια

Έχουμε:

$$f(y_i; \mu_i) = \exp\{y_i \log \mu_i - \mu_i - \log(y_i!)\}.$$

Οπότε

$$\log f(\mathbf{y}; \mu) = \sum_{i=1}^{n} [y_i \log \mu_i - \mu_i - \log(y_i!)]$$

όπου

$$\theta_i = \log \mu_i = \eta_i = x_i' \beta$$
$$b(\theta_i) = \exp(\theta_i) \;\Rightarrow\; b'(\theta_i) = \mu_i = \exp(\theta_i)$$

ΣΗΜΕΙΩΣΗ: Η ΕΜΠ και συμπερασματολογία για το β με Fisher's Scoring

Deviance:

$$L_{0S} = 2\sum_{i=1}^{n}\left[y_i \log \frac{y_i}{\hat{\mu}_i^{(0)}} - \left(y_i - \hat{\mu}_i^{(0)}\right)\right] \sim \chi^2_{n-p}$$

Pearsons χ^2:

$$Q = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}_i^{(0)}\right)^2}{\hat{\mu}_i^{(0)}} \sim \chi^2_{n-p}$$

Offset

Συχνά, σε πρακτικές εφαρμογές τα counts (αρ. εμφανίσεων γεγονότων) δεν είναι στην ίδια μονάδα

ΠΧ: Έστω y_i ο αρ. περιπτώσεων (πλήθος) μιας ασθένειας σε ένα γεωγραφικό διαμέρισμα της χώρας (i-οστή περιοχή)
Θεωρούμε το μοντέλο

$$y_i \sim Poisson(\mu_i = e_i \lambda_i)$$

όπου e_i είναι ο πληθυσμός της συγκεκριμένης περιοχής ή η έκταση της περιοχής ή ο αναμενόμενος αρ. περιπτώσεων στη περιοχή (γνωστό)
Μετά υποθέτουμε:

$$\log(\lambda_i) = x_i'\beta \Rightarrow \ \lambda_i = \exp(x_i'\beta),$$

συνεπώς

$$\mu_i = e_i\lambda_i \ \Rightarrow \ \log(\mu_i) = \log(e_i) + \log(\lambda_i) = \log(e_i) + x_i'\beta$$

Άρα:

$$\log(\mu_i) = \eta_i = \underbrace{\log(e_i)}_{Offset} + x_i'\beta$$

και η εκτίμηση του β γίνεται με μικρές αλλαγές στη συνάρτηση πιθανοφάνειας

Με λίγα λόγια, το log-linear μοντέλο διαπραγματεύεται τον αριθμό εμφάνισης γεγονότων (counts). Όταν όμως ενδιαφερόμαστε για το **ρυθμό** εμφανισης γεγονότων, τότε στο μοντέλο μας εμφανίζεται το offset

ΠΧ (συνέχεια): Έστω

y_i: οι αρ. περιπτώσεων ασθενών με καρκίνο των χειλιών στα 56 γεωγραφικά διαμερίσματα της Αγγλίας

e_i: οι αναμενόμενοι αριθμοί περιπτώσεων

x_i: το ποσοστό των ανθρώπων με 'υπαίθρια απασχόληση'

Παίρνουμε:

		$s.e.$	$p-value$
$\hat{\beta}_0$	-0.203	0.0662	0.00215
$\hat{\beta}_1$	0.026	0.0060	0.0

με: Deviance=362.54 $>>$ $72.15 = \chi^2_{54,0.05}$ (;)

Συνεπώς, αύξηση του ποσοστού των ανθρώπων με 'υπαίθρια απασχόληση' κατά 1% οδηγεί σε πολλαπλασιαστική αύξηση του μέσου αριθμού εμφανίσεων καρκίνου των χειλιών κατά 0,026
($\exp(0.026) = 1.026341 \simeq 1 + 0.026$)