

Εκφώνηση:

Στην άσκηση αυτή θα εξετάσετε κατά πόσο από τον βαθμό της προόδου σε ένα μάθημα μπορεί να προβλεφθεί η επιτυχία (= τελικός βαθμός τουλάχιστον 5) στο μάθημα αυτό.

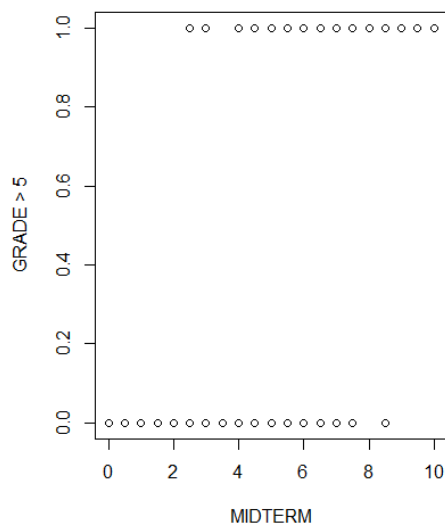
- Φτιάξτε ένα υπόδειγμα για τη σχέση μεταξύ βαθμού προόδου και επιτυχίας χρησιμοποιώντας λογιστική παλινδρόμηση στα δεδομένα που προέκυψαν από κάποιο μάθημα κατά το ακαδ. έτος 2014-15. Θεωρείτε ότι η λογιστική παλινδρόμηση είναι κατάλληλη ως υπόδειγμα;
- Βάσει του υποδείγματος αυτού, πόσο εκτιμάτε ότι είναι το ποσοστό επιτυχίας των φοιτητών όταν παίρνουν βαθμό 5 στην πρόοδο;
- Απαντήστε εάν σχετίζεται ο βαθμός προόδου με την επιτυχία, χρησιμοποιώντας έναν έλεγχο σημαντικότητας.
- Χρησιμοποιώντας το παραπάνω υπόδειγμα, μπορείτε να προβλέψετε εάν θα περάσει το μάθημα ένας φοιτητής που πήρε 5 στην πρόοδο;

A)

Θα χρησιμοποιήσουμε ως μεταβλητή επεξήγησης την MIDTERM και ως απόκρισης την GRADE

Έτσι έχουμε:

```
plot(MIDTERM, GRADE>5)
```



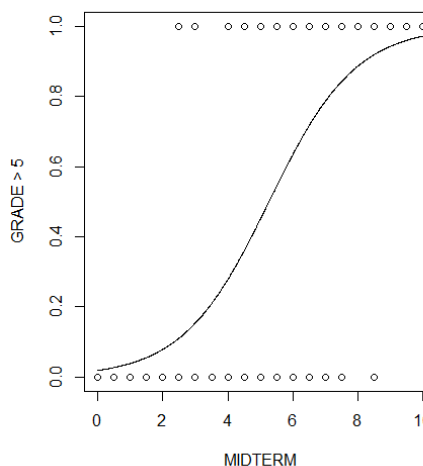
Παρατηρούμε ότι καθώς αυξάνεται ο βαθμός προόδου η επιτυχία είναι πιο συχνή, έτσι θεωρούμε πως το υπόδειγμα είναι κατάλληλο για την χρήση του λογιστικής παλινδρόμησης.

```
> m<-glm(GRADE>5~MIDTERM,family = binomial("logit"))
> m

Call:  glm(formula = GRADE > 5 ~ MIDTERM, family = binomial("logit"))

Coefficients:
(Intercept)      MIDTERM 
   -3.9689         0.7541 

Degrees of Freedom: 110 Total (i.e. Null);  109 Residual
(16 observations deleted due to missingness)
Null Deviance:      153.7 
Residual Deviance:  90.57    AIC: 94.57 
> x <- seq(from=0, to=10, by=0.01)
> y<-predict(m, newdata=data.frame(MIDTERM=x), type="response")
> lines(x,y)
```



B)

Από το παραπάνω μοντέλο παίρνουμε την πρόβλεψη:

```
> predict(m, data.frame(MIDTERM=5), type="response")-> y1
> y1
      1
0.4505533
```

Γ)

Θεωρούμε την υπόθεση H_0 πως δεν υπάρχει συσχέτιση και H_a πως υπάρχει, έτσι βάση του:

```
> summary(m)
```

Call:

```
glm(formula = GRADE > 5 ~ MIDTERM, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2470	-0.5775	0.2354	0.5828	2.0981

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.9689	0.7856	-5.052	4.37e-07	***
MIDTERM	0.7541	0.1353	5.575	2.47e-08	***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 153.65 on 110 degrees of freedom
Residual deviance: 90.57 on 109 degrees of freedom
(16 observations deleted due to missingness)
AIC: 94.57

Number of Fisher Scoring iterations: 5

Έχουμε λοιπόν $p\text{-value} = 2.47 \times 10^{-8}$ άρα απορρίπτουμε την μηδενική υπόθεση, με αποτέλεσμα να έχουμε συσχέτιση μεταξύ του βαθμού προόδου και του τελικού.

Δ)

Από το ερώτημα Β, έχουμε:

```
> predict(m, data.frame(MIDTERM=5), type="response")-> y1
> y1
      1
0.4505533
```

Άρα θεωρούμε πως δεν θα περάσει ο φοιτητής το μάθημα.