In Class Activity – Linear & Logistic Regression (ICA 10)

Please enter your responses at https://tinyurl.com/AIF19-ICA10

1. Each row in the table below shows sales for a year, and the amount spent on advertising in that year. In this case, our *outcome* of interest is sales. It is the thing we want to predict.

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

A linear regression model using *Advertising* as the predictor, has an intercept of 168 and a slope of 23.

a. Write out the equation for the model.

Sales = 168 + 23 * advertising

b. What are the predicted sales in Year 3?

168 + 23 * 30 = 858 million euros

c. What is the error of the prediction from (b)?

Error = predicted - actual = 858 – 856 = 2 million euros

d. What are the expected sales without any advertising?

The intercept or 168 million euros

e. How much would sales increase if advertising *increased* by 1 euro?

The slope or 23 million euros.

We are now interested in including both *the year and advertising* to estimate sales. This gives us an intercept of 323, a slope of 14 for advertising, and a slope of 45 for year.

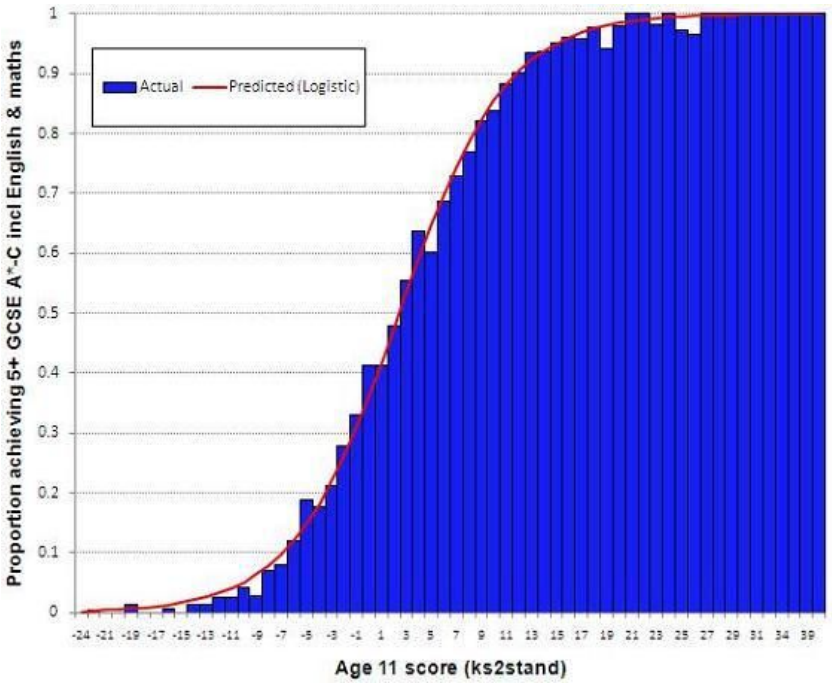
f. Write out the equation for the model.

sales = 323 + 14 * advertising + 45 * year

g. What are the estimated sales for year 5?

Sales = 323 + 14 * 43 + 45 * 5 = 1150 million euros

2. Consider the graph below, which depicts the proportion of age 16 students who pass five or more English and Math classes with A-C grades (Y-axis) based on their standardized test scores at age 11 (X-axis). We fit a logistic model to this data, the model is shown below in red.



a. From a visual examination, does the model appear to be a good fit for the data?

yes

b. What is the probability that a student who scores a -7 on the test at age 11 will pass at age 16?

About .08.

c. What is the minimum age 11 score that will result in a binary pass (vs. fail decision) at age 16 based on the logistic regression model?

About 3 as this is where the logistic function yields a predicted probability greater than 0.5.

d. If the probability of passing is 0.4, what are the odds of passing?

Odds = p/(1-p) = 0.4/0.6 = .667

e. If the log odds of passing is 0.85 for probability 0.7, what are the log odds of passing for probability 0.3?

-0.85 since 0.3 is the inverse of 0.7 (1-0.7) so the sign is flipped. Here is the math, odds = (0.3/0.7) = 0.43 and ln (0.43) = -0.85

f. The logistic curve above has a slope of .235 and an intercept of -.337. Which of the following describes the logistic regression function (p is the probability of passing)?
    i. p = -.337 + .235 * age 11 score.
    ii. log [p/(1-p)] = -.337 + .235 * age 11 score.
    iii. p/(1-p) = -.337 + .235 * age 11 score.
    iv. passing = -.337 + .235 * age 11 score.

Answer is ii.