# Multiple logistic regression

April 19, 2020

Applied STAtistics group at AAU

Department of Mathematical Sciences

Aalborg University

AALBORG UNIVERSITY

DENMARK

# Introduction

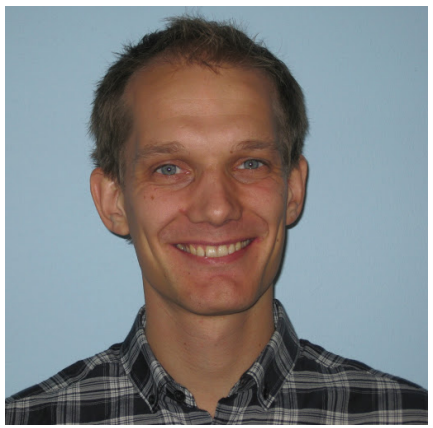


Outline of session:

- Data
- Model
- Inference
- Model selection

---

Lecturer for this session is Ege Rubak, Dept. of Math. Sciences, AAU



# Data



Wisconsin Breast Cancer Database covers 683 observations of 10 variables in relation to examining tumors in the breast.

- Nine clinical variables with a score between 0 and 10.
- The binary variable `Class` with levels `benign/malignant`.

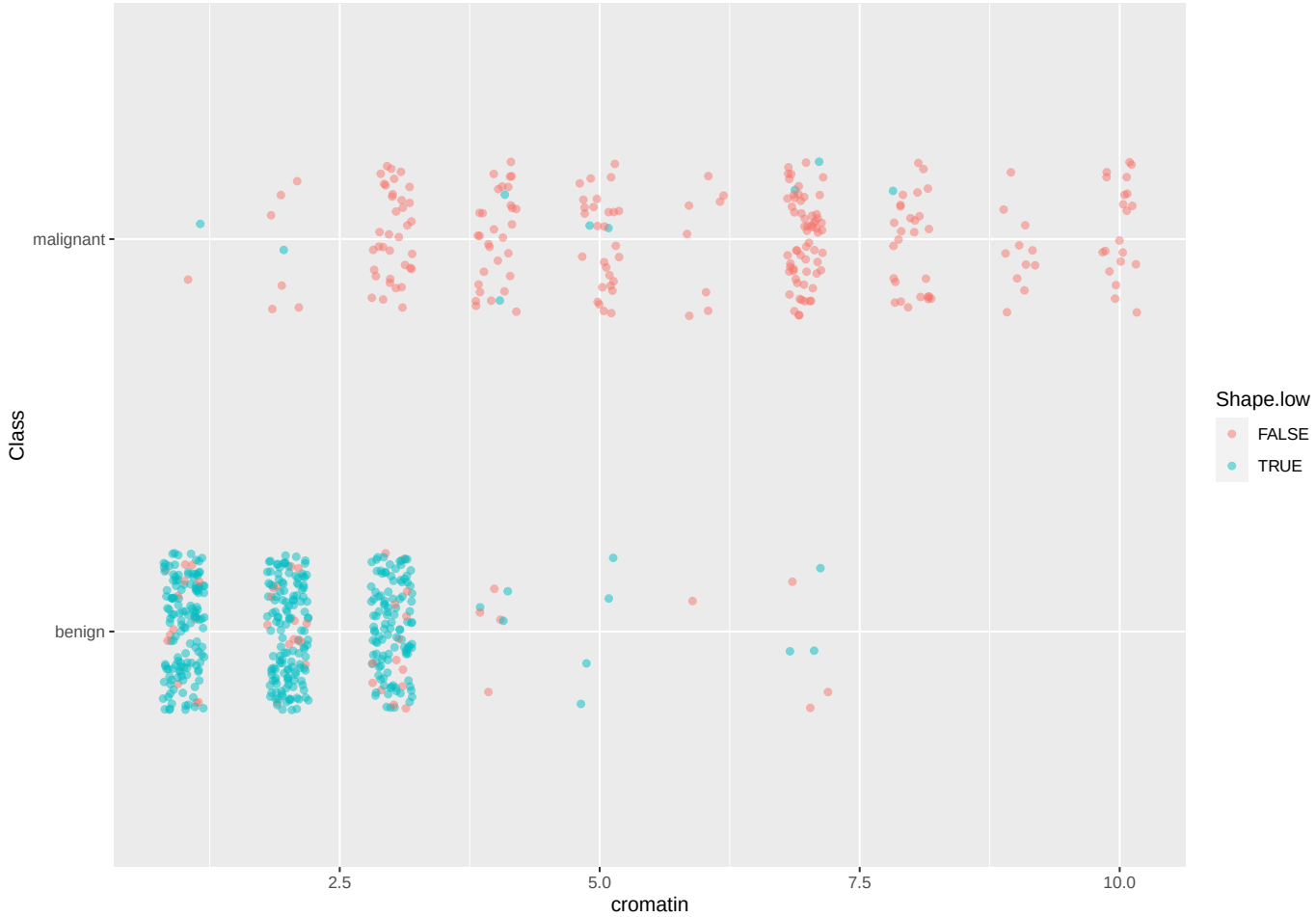
We will use 4 of the predictors, where 2 have been discretized.

| id | nuclei | cromatin | Size.low | Size.medium | Shape.low | Class |
|---|---|---|---|---|---|---|
| 1 | 1 | 3 | TRUE | FALSE | TRUE | benign |
| 2 | 10 | 3 | FALSE | TRUE | FALSE | benign |
| ... | ... | ... | ... | ... | ... | ... |
| 25 | 7 | 3 | TRUE | FALSE | FALSE | malignant |
| 26 | 1 | 2 | TRUE | FALSE | TRUE | benign |
| ... | ... | ... | ... | ... | ... | ... |
| 682 | 4 | 10 | FALSE | FALSE | FALSE | malignant |
| 683 | 5 | 10 | FALSE | FALSE | FALSE | malignant |

# Plot of data



## Plots of tumor class vs. cromatin coloured by shape score (low or not)



# Multiple logistic regression



We generalize the simple logistic regression model to the case, where we have $k$ predictors $x_1, x_2, \ldots, x_k$. Where some might be dummies for a factor.

$$\texttt{logit}(P(y = 1 \mid x_1, x_2, \ldots, x_k)) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

Interpretation of $\beta$-values is unaltered: If we fix $x_2, \ldots, x_k$ and increase $x_1$ by one unit, then the relative change in odds is given by $\exp(\beta_1) - 1$.

# Example without interaction



For Class modelled by cromatin score ($x_1$), and shape score ($x_2$ – dummy for low or not) without interaction:

$$\texttt{logit}(P(y = 1 \,|\, x_1, x_2)) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

For tumors with a low shape score ($x_2 = 1$) we get

$$\texttt{logit}(P(y = 1 \,|\, x_1, x_2)) = \alpha + \beta_2 + \beta_1 x_1$$

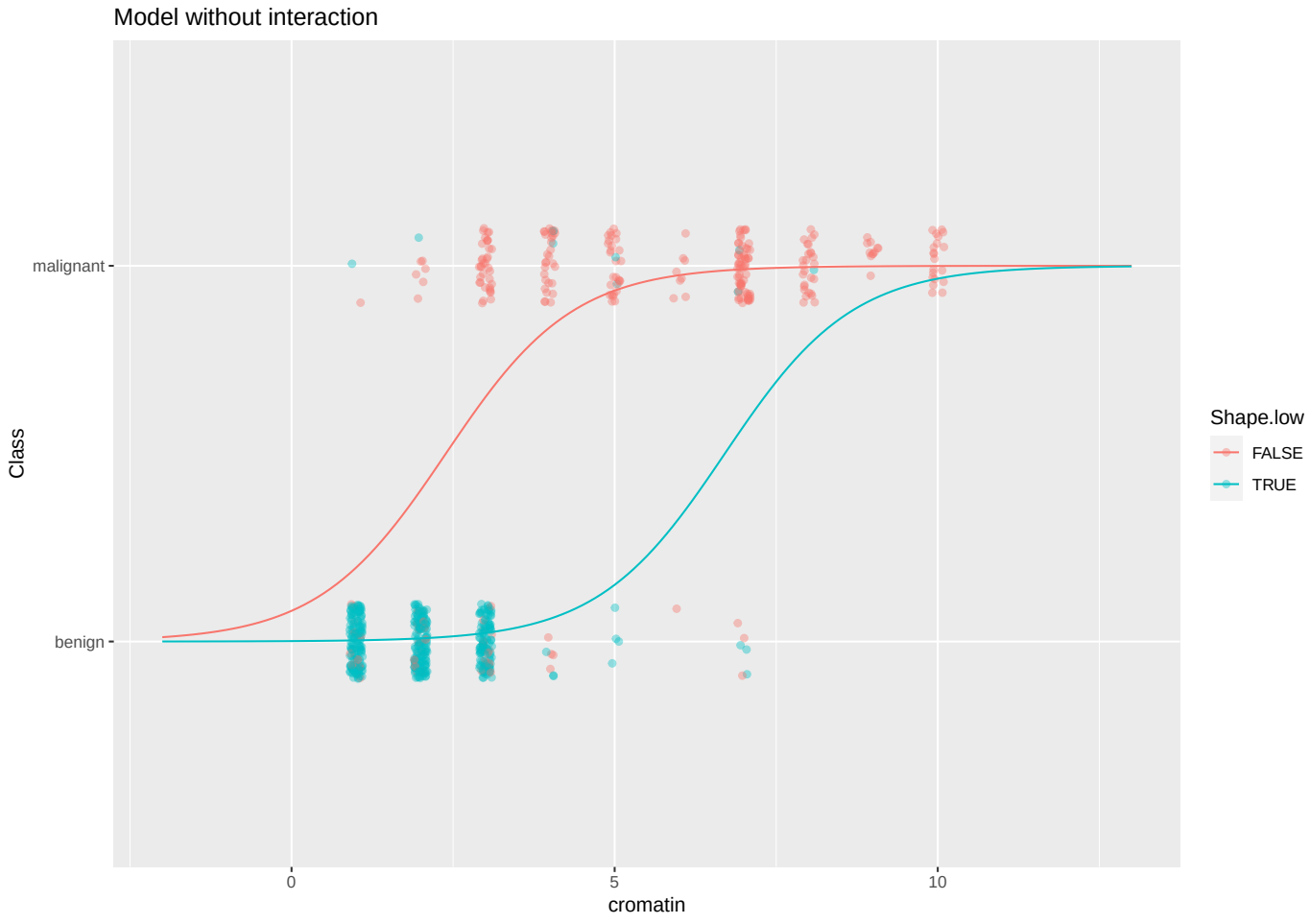
When the shape score is not low ($x_2 = 0$) we get

$$\texttt{logit}(P(y = 1 \,|\, x_1, x_2)) = \alpha + \beta_1 x_1$$

Slopes are the same – curves have just shifted horizontally.

# Graphically





# Test of significance of the model



- The null model with no effect of the predictors corresponds to

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

- I.e. there is a constant risk of malignant tumor regardless of predictor values.
- The test statistic is the Deviance which should be small under $H_0$.
- It is evaluated in $\chi^2$-distribution with $k$ degrees of freedom (number of parameters set to zero).

| Resid. Df | Resid. Dev | Df | Deviance | Pr(>Chi) |
|---|---|---|---|---|
| 682 | 884.4 | NA | NA | NA |
| 680 | 218.7 | 2 | 665.6 | 2.899e-145 |

- The 95%-critical value for the $\chi^2(2)$ distribution is 6 and the p-value is in practice zero.

# Example with interaction



For Class modelled by cromatin score ($x_1$), and shape score ($x_2$ – dummy for low or not) with interaction:

$$\texttt{logit}(P(y = 1 \,|\, x_1, x_2)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

For tumors with a low shape score ($x_2 = 1$) we get

$$\texttt{logit}(P(y = 1 \,|\, x_1, x_2)) = \alpha + \beta_2 + (\beta_1 + \beta_3) x_1$$

When the shape score is not low ($x_2 = 0$) we get

$$\texttt{logit}(P(y = 1 \,|\, x_1, x_2)) = \alpha + \beta_1 x_1$$

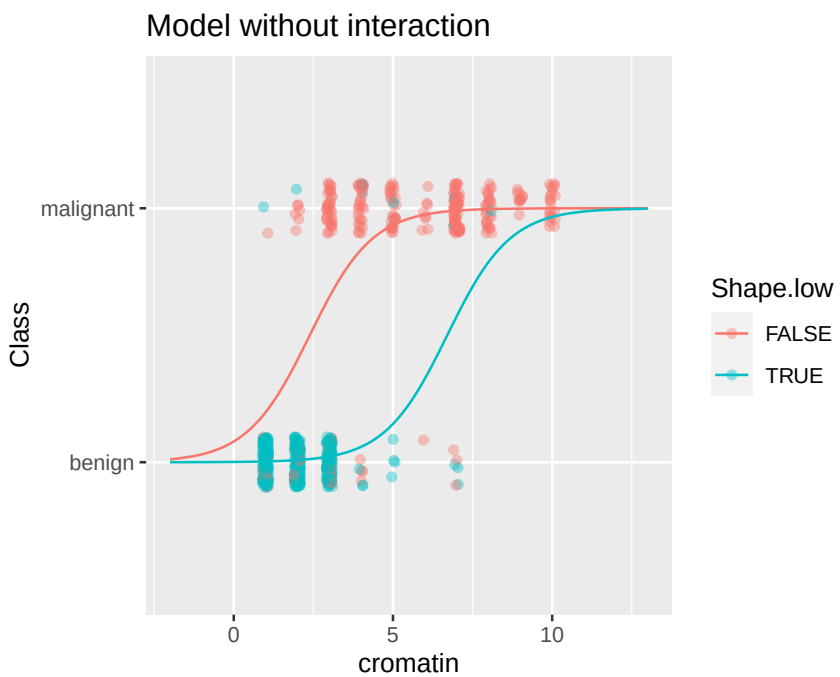
Again the Deviance is used to test $H_0 : \beta_3 = 0$ for insignificant interaction (or we could use the z-test on next page):

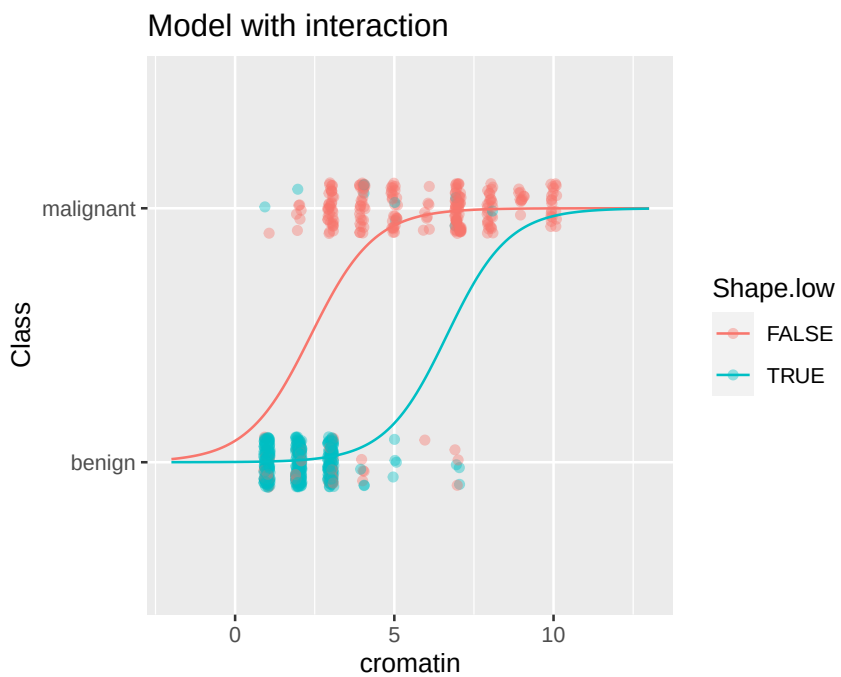
| Resid. Df | Resid. Dev | Df | Deviance | Pr(>Chi) |
|---|---|---|---|---|
| 680 | 218.7 | NA | NA | NA |
| 679 | 218.7 | 1 | 0.01448 | 0.9042 |

# Example with interaction



| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **(Intercept)** | -2.38 | 0.525 | -4.54 | 5.52e-06 |
| **cromatin** | 1 | 0.157 | 6.4 | 1.57e-10 |
| **Shape.lowTRUE** | -4.49 | 1.02 | -4.4 | 1.08e-05 |
| **cromatin:Shape.lowTRUE** | 0.0307 | 0.255 | 0.12 | 0.904 |



# Model selection by stepwise selection



- We propose a starting model with all predictors and their pairwise interactions, which is a 15 parameter model.
- We use the so-called Bayesian Information Criterion (BIC) to leave out terms one by one.
- We end up with a model with 7 parameters:

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **(Intercept)** | 0.0337 | 0.903 | 0.0373 | 0.97 |
| **nuclei** | 0.302 | 0.0837 | 3.6 | 0.000314 |
| **cromatin** | 0.446 | 0.144 | 3.09 | 0.00198 |
| **Size.lowTRUE** | -5.42 | 1.14 | -4.77 | 1.82e-06 |
| **Size.mediumTRUE** | -2.29 | 0.69 | -3.33 | 0.000874 |
| **Shape.lowTRUE** | -2.25 | 0.649 | -3.47 | 0.000525 |
| **nuclei:Size.lowTRUE** | 0.569 | 0.236 | 2.41 | 0.0157 |

# Prediction and classification



- We save the final model's estimate of the probability of `malignant`.

| Class | P(y=1) |
|---|---|
| benign | 0.004 |
| benign | 0.89 |
| benign | 0.01 |
| benign | 0.929 |
| benign | 0.004 |
| malignant | 0.999 |

- We may classify by rounding the probability:

| | 0 | 1 |
|---|---|---|
| **benign** | 432 | 12 |
| **malignant** | 11 | 228 |

23 patients are misclassified.

# Prediction and classification



- If we assign everyone with predicted probability of malignancy above 5% to further investigation, then we catch all malignant.

| | 0 | 1 |
|---|---|---|
| **benign** | 405 | 39 |
| **malignant** | 0 | 239 |

The expense is that the number of false positive increases from 12 to 39.

- If we instead set the alarm to 10%, then the number of false positives decreases from 39 to 26.

| | 0 | 1 |
|---|---|---|
| **benign** | 418 | 26 |
| **malignant** | 2 | 237 |

But at the expense of 2 false negative.