# Handout 3 – Parameter estimation

## Least Squares Estimate

Method used in linear regression to optimise model parameters/coefficients.

Residual Sum of Squares (RSS)



The lower the RSS, the better a set of parameters is. We find the parameter set that give us the lowest possible RSS score.

Why we can’t use RSS for Poisson/Logistic regression?

1. Least squares estimation looks to minimise the sum of squared differences between observed and estimates values. Negative and positive differences are treated the same (due to the squared).

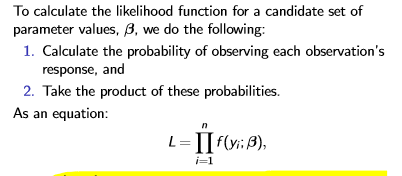
* Due to this least square is only appropriate when the response distribution is symmetric.
* Thus, for Poisson and Logistic regression we cannot use least squares.

1. Least squares weigh the residuals the same regardless of the amount of variance. This is not valid for response variables with non-constant variance.

* Thus cannot use it for Poisson and Logistic

## Maximum Likelihood Estimate

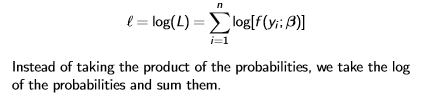
Method used in Poisson/Logistic regression (GLM’s) to estimate parameters is maximum likelihood.



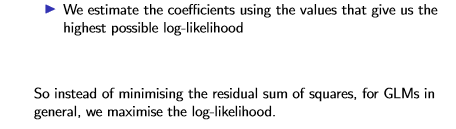
Typically:

* Points that lie close to the model line have high likelihood
* Point that are further from the model line have low likelihood

Maximum likelihood estimation involves finding the parameter set that maximise the likelihood estimate. We deal with log-likelihood instead of likelihood.



Because of the log transformation we can add instead of multiplying the probabilities.



# Handout 4 – Deviance

Deviance – goodness of fit statistic for GLM’s

## Model complexity

Increasing a model’s complexity will always:

* Decrease the RSS (linear regression)
* Increase the log-likelihood (GLM’s)

## Saturated Model

A model with as many parameters as there is observations is called a saturated model.

* Most complex model you can fit
* Regression line will go through each point
* RSS will be equal to 0
* Highest log-likelihood possible for the data
* Cannot further add any explanatory terms

## Null Model

A model with no explanatory terms

* Most simple model you can fit
* RSS will be maximum possible
* Log-likelihood will be minimum possible

## Handout 4 - Deviance

Deviance is twice is difference between the maximised log-likelihood (saturated model) and log-likelihood of the fitted model.



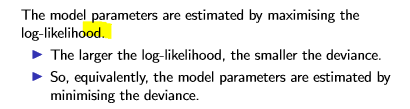
Deviance of the saturated model = 0

As the deviance gets larger:

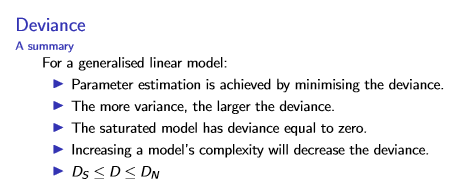
* the fitted model gets worse
* Less complex model
* Smaller log-likelihood

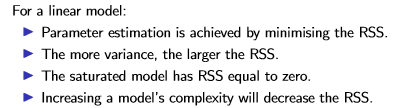
The null deviance is the deviance of the null model. It is the maximum deviance possible.





Minimise deviance to get optimal set of model parameters.





## Chapter 5 – Deviance as a goodness of fit statistic

We can use deviance to assess mode fit.

Models that have poor fit will have a large deviance. Poor model fits occur for two core reasons:

1. Model is too simple and does not have enough/the right explanatory terms
2. Response variable has more variance than assumed under the distribution

Use hypothesis test and assess whether the deviance comes from a chi-squared distribution of n-k degrees of freedom.

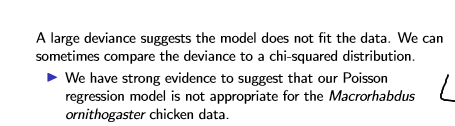
* N = number of observations
* K = number of parameters estimated

**High deviance = low p -value** and have strong evidence against the hypothesis that the model fit is good.

**Low deviance = high p-value** and no evidence against null hypothesis.

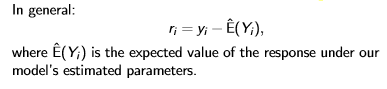
Need to satisfy chi-squared distribution assumptions for the test to hold:

* Holds true for Poisson where mean is greater than 5
* Holds true for Logistic where number of trials is large enough and
  + P near 0.5 n>=5
  + P close to 0 or 1, n must be large



## Handout 6 – Residuals

### Raw residuals



Raw residuals are difference between observed response and its expected value.

Linear regression raw residuals will be:

* Zero mean
* Constant variance

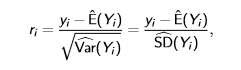
For Poisson/Logistic regression raw residuals will be:

* Zero mean
* Non-constant variance

Thus, we use Pearson residuals.

### Pearson residuals

Pearson residuals are standardised to have constant variance. If the model is correct, then the Pearson residual will have approximately constant variance.



To calculate Pearson residual: Divide residual by the standard deviation of response under the model

If model is appropriate the Pearson residuals will show:

* 0 Mean
* Constant variance
* Patternless band around 0
* Residuals will have come from normal distribution

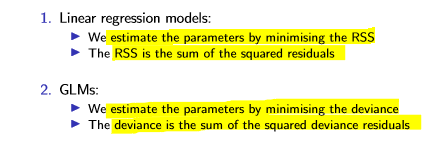
If observations are not sparse:

* Residuals will show banding (apparent patterns)

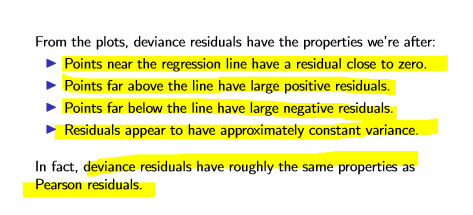
### Deviance residuals

Calculate residuals using deviance

*Compare Linear regression to GLM’s*



Deviance residual properties



If the model is appropriate:

If model is appropriate the Pearson residuals will show:

* 0 Mean
* Constant variance
* Patternless band around 0
* Residuals will have come from normal distribution

## Sparse Data - Banding

Note any banding or patterns seen in the residual plot is due to sparse data. The banding is because observations in the certain region can only take certain values and thus are very discrete.

Ungrouped data in logistic regression is the extreme case of sparisty; each observation can be one of two possible outcomes.

## Randomised Quantile Residuals

Pearson and Deviance residuals is problematic when the data is sparse. The data will show an apparent pattern even if the model is correct.

Randomised Quantile residuals randomly jitter the residuals to break up banding. This jittering works even if the data is sparse or not.

