

Notat 2:

**Approksimation af en logistisk regressionsmodel
med en lineær normalfordelingsmodel**

SÆTNING. *Lad X være binomialfordelt med antalsparameter n og sandsynlighedsparameter p . Hvis n er stor og p ikke er for tæt på 0 eller 1 gælder at*

$$Y = \text{logit} \left(\frac{X}{n} \right) = \log \left(\frac{\frac{X}{n}}{1 - \frac{X}{n}} \right) = \log \left(\frac{X}{n - X} \right)$$

er approksimativt normalfordelt med

$$EY = \text{logit}(p) = \log \left(\frac{p}{1 - p} \right) \quad \text{og} \quad \text{var}(Y) = \frac{1}{np(1 - p)}.$$

BEVIS. Ved rækkeudvikling til første orden af funktionen logit omkring p fås

$$\text{logit} \left(\frac{X}{n} \right) \approx \text{logit}(p) + \text{logit}'(p) \left(\frac{X}{n} - p \right)$$

hvor

$$\text{logit}'(p) = \frac{d}{dp} (\log p - \log(1 - p)) = \frac{1}{p} + \frac{1}{1 - p} = \frac{1}{p(1 - p)}.$$

Altså

$$\text{logit} \left(\frac{X}{n} \right) \approx \text{logit}(p) + \frac{1}{p(1 - p)} \left(\frac{X}{n} - p \right).$$

På grund af den sædvanlige approksimation af binomialfordelingen med en normalfordeling kan vi skrive

$$X \approx np + \sqrt{np(1 - p)}U,$$

hvor U er normeret normalfordelt. Indsættes dette får vi

$$\begin{aligned} \text{logit} \left(\frac{X}{n} \right) &\approx \text{logit}(p) + \frac{1}{p(1 - p)} \left(\frac{np + \sqrt{np(1 - p)}U}{n} - p \right) \\ &= \text{logit}(p) + \frac{U}{\sqrt{np(1 - p)}}. \end{aligned}$$

Heraf følger påstanden umiddelbart.

Dette resultat kan bruges til approksimation af en logistisk regressionsmodel med en vægtet linær regressionsmodel. Fordelen ved at gøre det på den måde er, at man herved får udvidet modellen med en skalaparameter, som kan tage højde for en eventuel overspredning.

EKSEMPEL. SAS programmet PRG2.SAS (se under SAS eksempler) skulle ifølge dette føre til approksimativt det samme som PRG1.SAS.