

La régression Bêta

Une alternative intéressante pour modéliser des proportions

Maxime Lacroix

28 septembre 2018

Mise en contexte

Contexte

- Régression sur une variable réponse tenue entre $(0,1)$
- Par exemple, un taux ou une proportion
- Régression linéaire “classique” à éviter

Première solution : transformation logit

Première transformation possible

$$\tilde{y} = \log\left(\frac{y}{1-y}\right) \quad (1)$$

- Avantage :
 - Les données ne sont plus bornées, la régression linéaire est envisageable
- Désavantages :
 - Interprétation différente
 - Fort potentiel d'hétéroscédasticité
 - Les données sont souvent asymétrique → problèmes pour les tests d'hypothèses et les intervalles de confiance.

```
##           speed           dist
##  Min.      : 4.0    Min.      :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean    :15.4    Mean     : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.     :25.0    Max.      :120.00
```

Going to sleep

- Get in bed
- Count sheep

Densité d'une loi bêta

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (2)$$

Hello Equation 2