BS1a Applied Statistics
Lectures 9-10

Dr Geoff Nicholls

Week 5 MT10

Box-Cox

Observations of $y, x_1, ..., x_p$ with $y_k \geq 0$.

y not linear with $x_1, ..., x_p$ try

$$y' = (y^\lambda - 1)/\lambda$$

treating λ as an(other) unknown parameter.

$(y^\lambda - 1)/\lambda$ gives powers of y and $\log(y)$.

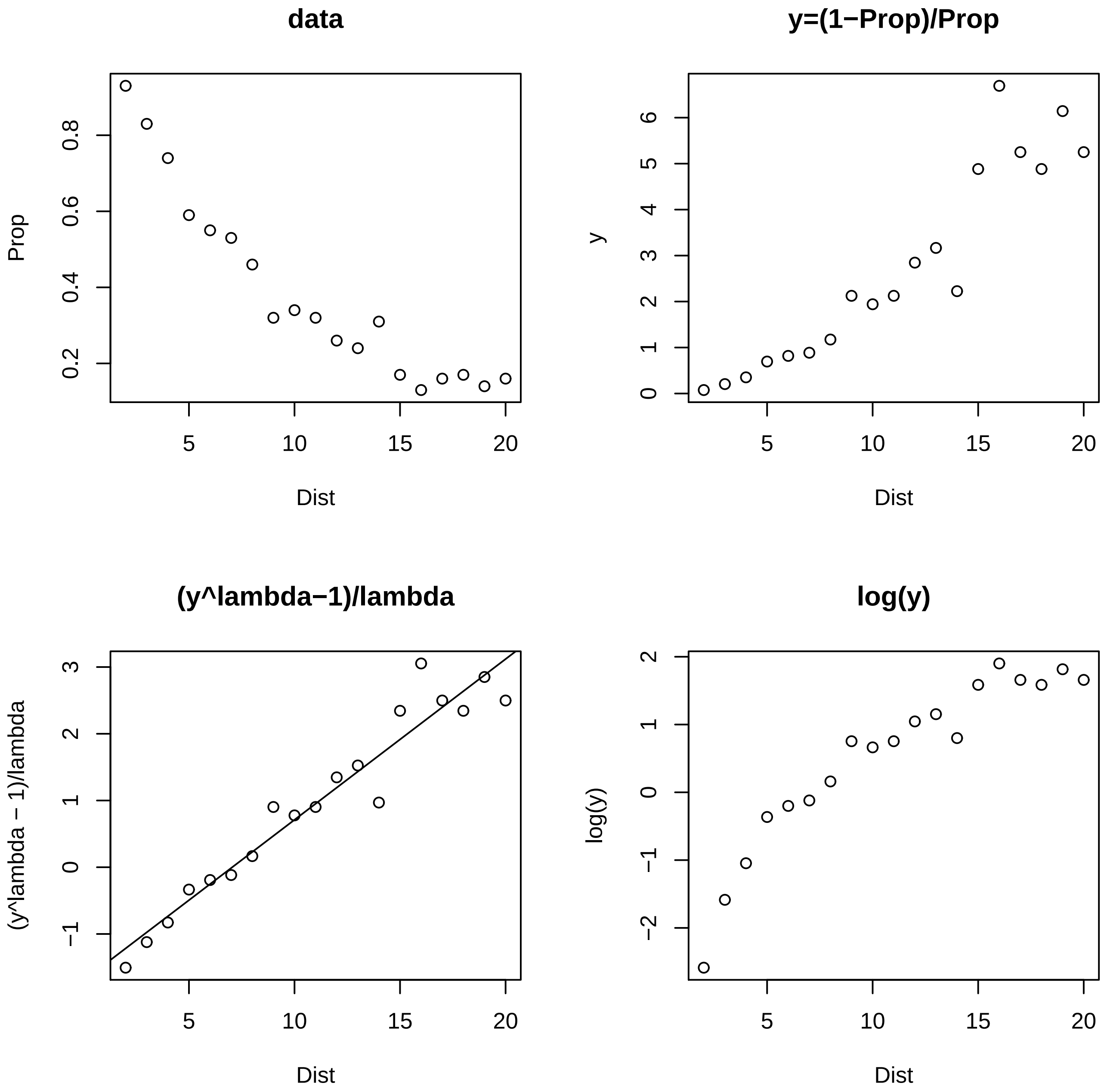
Likelihood is now

$$L(\beta, \sigma^2, \lambda; y') \propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_k (y'_k - \mathbf{x}_k^T \beta)^2 \right).$$

Exercise Compute MLE's for β, σ^2 and λ (transform to get $L(\beta, \sigma^2, \lambda; y)$ then maximise - this is the first exercise of PS4).

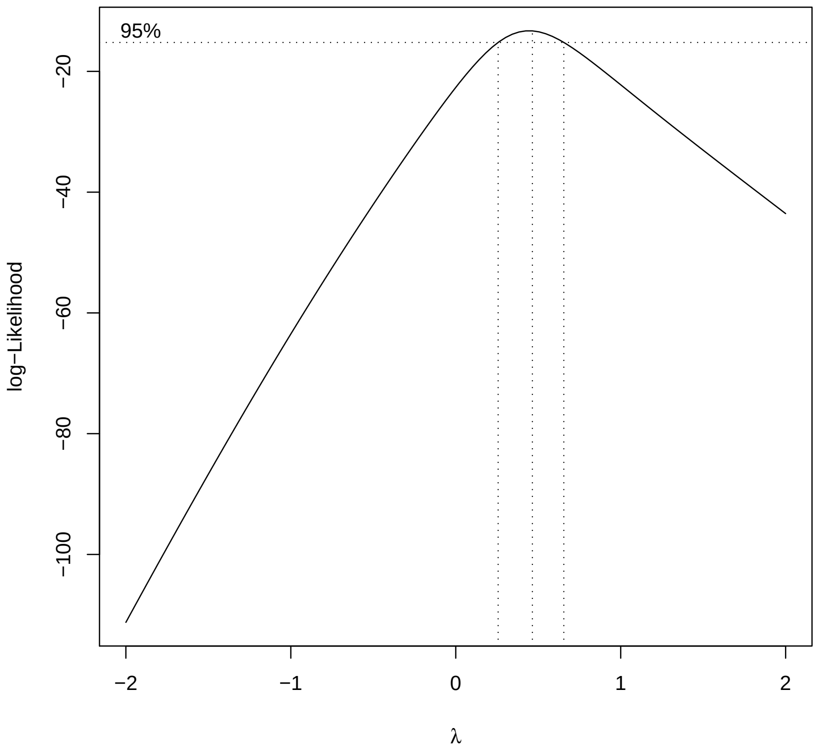
Example: fraction of successful putts as a function of distance in feet.

```
> putts<-data.frame(Dist=2:20, Prop=c(0.93,0.83,
    0.74,0.59,0.55,0.53,0.46,0.32,0.34,0.32,0.26,
    0.24,0.31,0.17,0.13,0.16,0.17,0.14,0.16))
> putts
   Dist Prop
1     2 0.93
2     3 0.83
3     4 0.74
...
17   18 0.17
18   19 0.14
19   20 0.16
> y<-(1-putts$Prop)/putts$Prop
> x<-putts$Dist
```



The λ value was estimated by maximising the likelihood, as above.

```
> putts.bc<-boxcox(y~x)
```




```
> putts.lm<-lm(sqrt(y)~x)
> summary(putts.lm)
...
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.14342    0.09818   1.461    0.162
x             0.12293    0.00799  15.386 2.07e-11
...
```

$$\sqrt{\frac{1-\text{Prop}}{\text{Prop}}} = \text{Dist} + \epsilon$$

Weighted Regression

$$Y_i = \mathbf{x}_i\beta + \epsilon \quad \text{for } i = 1, 2, ..., n$$

$\epsilon_i \sim \mathrm{N}(0, \sigma^2/w_i)$ non-constant variance.

Eg: $Y_i = n_i^{-1}\sum_j Y_{i,j}$ yielding var$(Y_i) \simeq \sigma^2/n_i$.

Let $W = \mathrm{diag}(w_1, ..., w_n)$ and $Y' = W^{1/2}Y$ and $X' = W^{1/2}X$ so

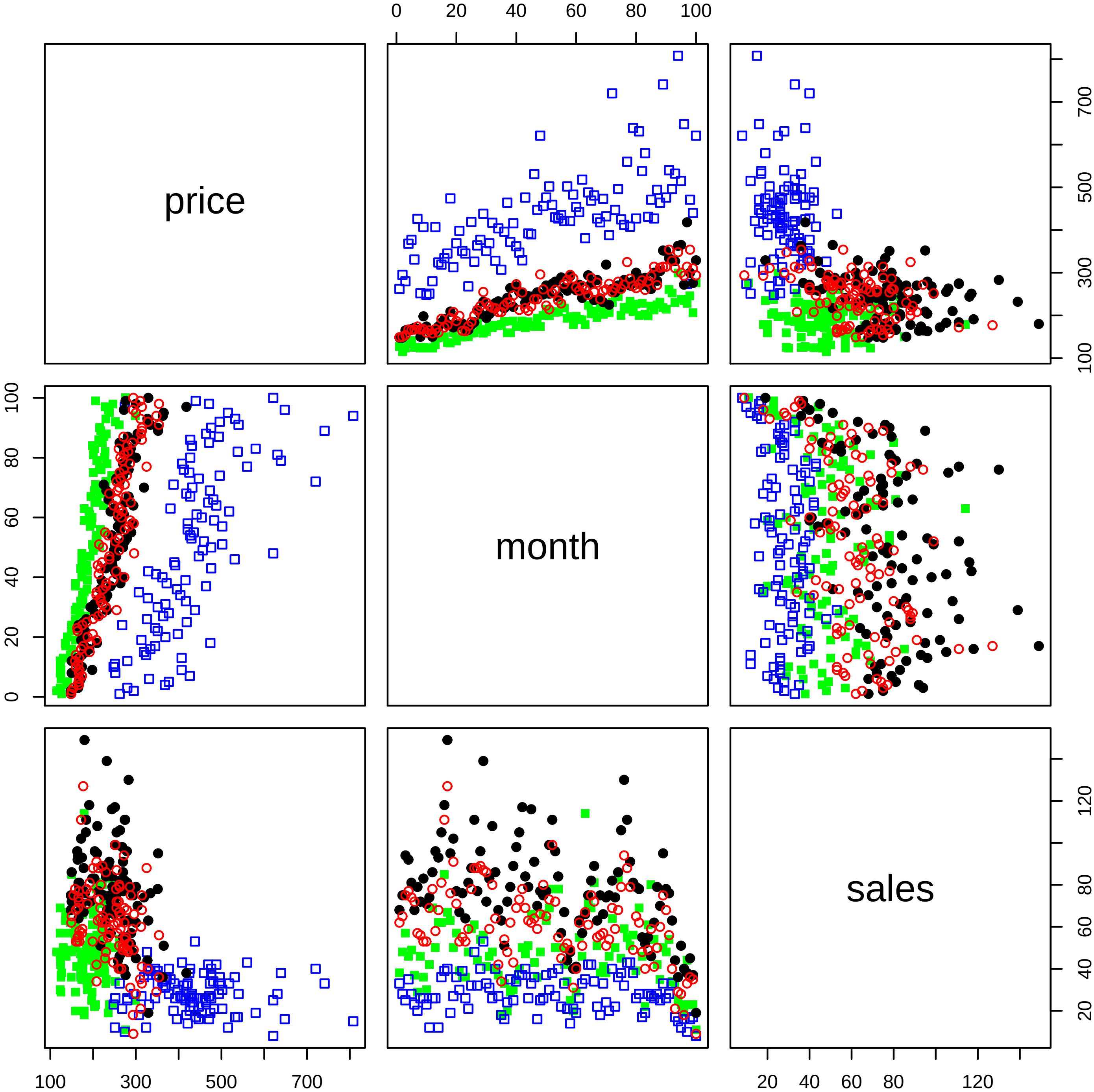
$$Y' = X'\beta + \epsilon'$$

$\epsilon' \sim \mathrm{N}(0, \sigma^2 I_n)$

$\hat{\beta} = (X^T W X)^{-1} X^T W Y$ estimates β, and

$$s^2 = (Y - X\hat{\beta})^T W (Y - X\hat{\beta})/(n - p)$$

is unbiased for σ^2. This is WLS.


price
month
sales
0
20
40
60
80
100
100
300
500
700
100
300
500
700
0
20
40
60
80
100
20
40
60
80
120
20
40
60
80
120

Example: OHP revisited

OHP data were monthly averages, `Sales` gives number of sales by month/House Type

Let $w_i = \texttt{sales}[i]$ in the ith month. Very simple NLM for flats alone.

$$y_k \sim \alpha + \gamma_M x_{k,M} + \epsilon_k, \qquad \epsilon_k \sim \mathrm{N}(0, \sigma^2/w_i),$$

Not enough (see below). Signs of $\sigma \propto \mathrm{E}(y)$. Would like to consider

$$w_i = 1/\mathrm{E}(Y_i)^2$$

but we dont have $\mathrm{E}(Y_i)$ (it is a result of the regression we want to do). Fit unweighted then fit $w_i = 1/\widehat{y}_i^2$ or $w_i = \texttt{sales}[i]/\widehat{y}_i^2$.

Note that when we refit using the results of the first fit in the second, we replace one model error (non-constant variance) with another (correlation between observations). We do at least diagnose the issue.

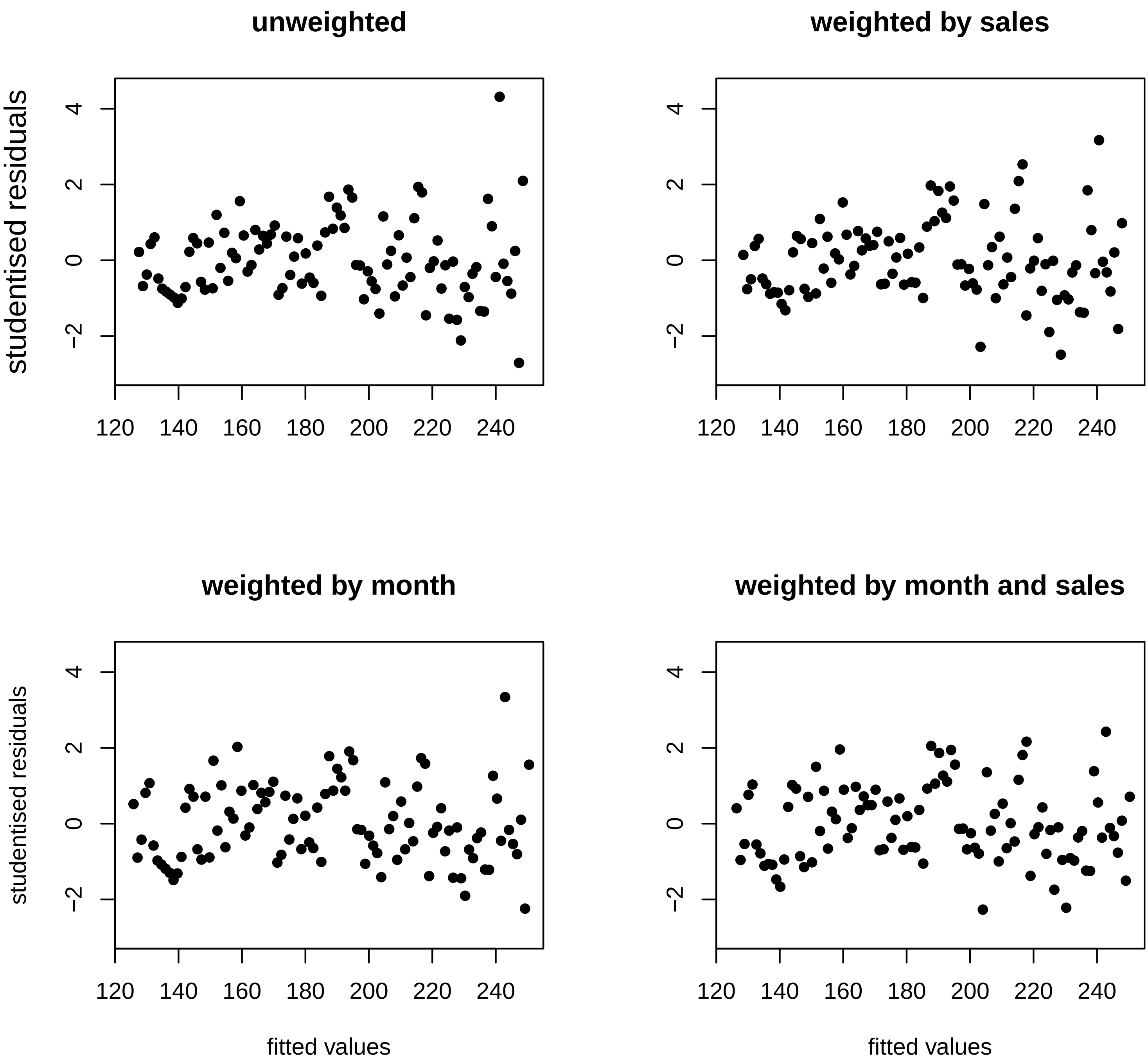
```
> #read the data
> fohp<-read.table('ohp.txt',header=TRUE)
> names(fohp)<-c('price','type','month','sales')
>
> #collect flats
> ohp<-fohp[ (fohp[,2]=='Flat'),]

#unweighted NLM
> ohp.lm<-lm(price~month,data=ohp)
> yhat<-fitted.values(ohp.lm)

> #weighted by number of sales
> ohp.wlm1<-lm(price~month,weights=ohp$sales,data=ohp)

> #weighted by fitted values (sigma ~ mean price)
> ohp.wlm2<-lm(price~month,weights=1/yhat^2,data=ohp)

> #weighted by sales and fitted values
> ohp.wlm3<-lm(price~month,weights=ohp$sales/yhat^2,data=ohp)
```


unweighted
weighted by sales
weighted by month
weighted by month and sales
studentised residuals
studentised residuals
fitted values
fitted values
120
140
160
180
200
220
240
−2
0
2
4

```
> summary(ohp.wlm3)
...
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 125.15076    2.47997   50.47   <2e-16
month         1.25198    0.05204   24.06   <2e-16
...
Residual standard error: 0.5502 on 98 degrees of freedom
F-statistic: 578.8 on 1 and 98 DF,  p-value: < 2.2e-16
```

Exercise Explain how to compute the RSE $s = 0.55$ and $F = 578.8$.

Exercise We fit the weighted response so $r'(e') \sim t(n-p-1)$ and $\widehat{Y}'$ are independent under NLM for Y'. However, R returns $\widehat{Y} = X\widehat{\beta}$ and $e = (Y - X\widehat{\beta})$ (*ie* in original unweighted coordinates). Verify $r'(e) \sim t(n-p-1)$ and $\widehat{Y}$ independent under NLM for Y'.

Generalised Linear Models

Restrict observation model to $Y \sim f(y|\theta)$

$$f(y|\theta) = \exp\left(\frac{y\theta - \kappa(\theta)}{\phi} + c(y;\phi)\right), \quad y \in \Omega$$

$\phi = 1$ natural exponential family

$\phi > 0$ natural exponential dispersion family.

$$\exp(\kappa/\phi) = \int_\Omega \exp\left(\frac{y\theta}{\phi} + c(y)\right)\, dy$$

$\theta \in \{\theta;\ \kappa(\theta) < \infty\}$

y natural observation

θ natural parameter

ϕ dispersion parameter

Normal: $Y \sim \mathrm{N}(\mu, \sigma^2)$ is NEDF

$$
\begin{aligned}
f(y|\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\
&= \exp\left(\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2} - \frac{y^2}{2\sigma^2}\right) \\
&= \exp\left(\frac{y\theta - \kappa(\theta)}{\phi} + c(y;\phi)\right)
\end{aligned}
$$

$\theta = \mu$, $y \in R$, $\phi = \sigma^2$,
$\kappa = \mu^2/2$
$c(y;\phi) = -(1/2)\log(2\pi\phi) - y^2/2\phi$

Student's-t: $Y \sim \mathrm{t}(\nu)$ is not in this class

$$
f(y|\nu) \propto (1 + y^2/\nu)^{-\left(\frac{\nu+1}{2}\right)}
$$

cant factorise y and ν in exponent.

$$e^{\kappa/\phi} = \int_\Omega \text{exp}\left(\frac{y\theta}{\phi} + c(y)\right)\, dy$$

$\kappa(\theta)$ cummulant GF (at $\phi = 1$).

$$\begin{aligned} \kappa' e^{\kappa/\phi} &= \int_\Omega \left(\frac{y}{\phi}\right) \text{exp}\left(\frac{y\theta}{\phi} + c(y)\right)\, dy \\ \kappa' &= \text{E}(Y) \end{aligned}$$

$$\begin{aligned} \left(\frac{(\kappa')^2}{\phi^2} + \frac{k''}{\phi}\right) e^{\kappa/\phi} &= \int_\Omega \left(\frac{y}{\phi}\right)^2 \text{exp}\left(\frac{y\theta}{\phi} + c(y)\right) dy \\ \left(\frac{(\kappa')^2}{\phi^2} + \frac{k''}{\phi}\right) &= \text{E}(Y^2)/\phi^2 \end{aligned}$$

$$\begin{aligned} \text{var}(Y) &= \text{E}(Y^2) - \text{E}(Y)^2 \\ &= (\kappa')^2 + \phi\kappa'' - (\kappa')^2 \\ &= \phi\kappa''. \end{aligned}$$

Let $\mu_i = \text{E}(Y_i)$, $\text{var}(Y_i) = \phi V(\mu_i)$.

Exercise $\frac{d\mu_i}{d\theta_i} = V(\mu_i)$ so μ_i increases with θ_i.

Modeling with GLM's
three decisions fix the model.

Distribution: $Y_i \sim f(y_i|\theta)$, iid for $i = 1, 2, ..., n$.
given the EV, how is the response distributed?

Linear predictor: $\eta = x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p$
what variables $x_1, x_2, ..., x_p$ should we use?

Link function $g(\mu_i) = \eta_i$
(increasing, continuous, differentiable function)
How does mean response increase with LP?

NLM $y_i = \mathbf{x}_i\beta + \epsilon_i,\ \ \epsilon_i \sim N(0, \sigma^2)$
Stochastic, $Y_i \sim$ N(μ_i, σ^2) jointly indpendent
Deterministic, $\eta_i = \mathbf{x}_i^T\beta$, $\eta = X\beta$
Link, choice $g(\mu_i) = \mu_i$ gives NLM E$(Y_i) = \mathbf{x}_i\beta$.

Log-Likelihood (GLM)

$$\ell(\beta; y) = \sum_{i=1}^{n} \frac{y_i\theta_i - \kappa(\theta_i)}{\phi} + c(y_i; \phi),$$

with

$$\theta_i = \theta(\mathbf{x}_i\beta)$$

since $\mu_i = \kappa'(\theta_i)$ and $g(\mu_i) = \eta_i$ both invertible

Canonical link function (important special case)

If we choose link function

$$g(\mu_i) = \kappa'^{-1}(\mu_i)$$

then $\theta_i = \eta_i$ since $g(\mu_i) = \theta_i$.

Log-Likelihood (canonical link, $\phi = 1$ or known)

$$\ell(\beta; y) = \sum_{i=1}^{n} \frac{y_i\eta_i - \kappa(\eta_i)}{\phi} + c(y, \phi), \qquad \eta_i = \mathbf{x}_i\beta.$$

Example GLM for a Poisson response

$Y_i \sim$ Poisson(λ_i), independent E$(Y_i) = \lambda_i$.

$$
\begin{aligned}
\exp(-\lambda)\frac{\lambda^y}{y!} &= \exp\left(y\log(\lambda) - \lambda - \log(y!)\right) \\
&= \exp\left(\frac{y\theta - \kappa(\theta)}{\phi} + c(y;\phi)\right)
\end{aligned}
$$

$\theta = \log(\lambda)$, $\phi = 1$, $\kappa(\theta) = \exp(\theta)$.

Check E(Y): $\mu = \kappa'(\theta) = \exp(\theta) = \lambda$.

Check var(Y): $\phi V(\mu) = \kappa''(\theta) = \lambda = \mu$.

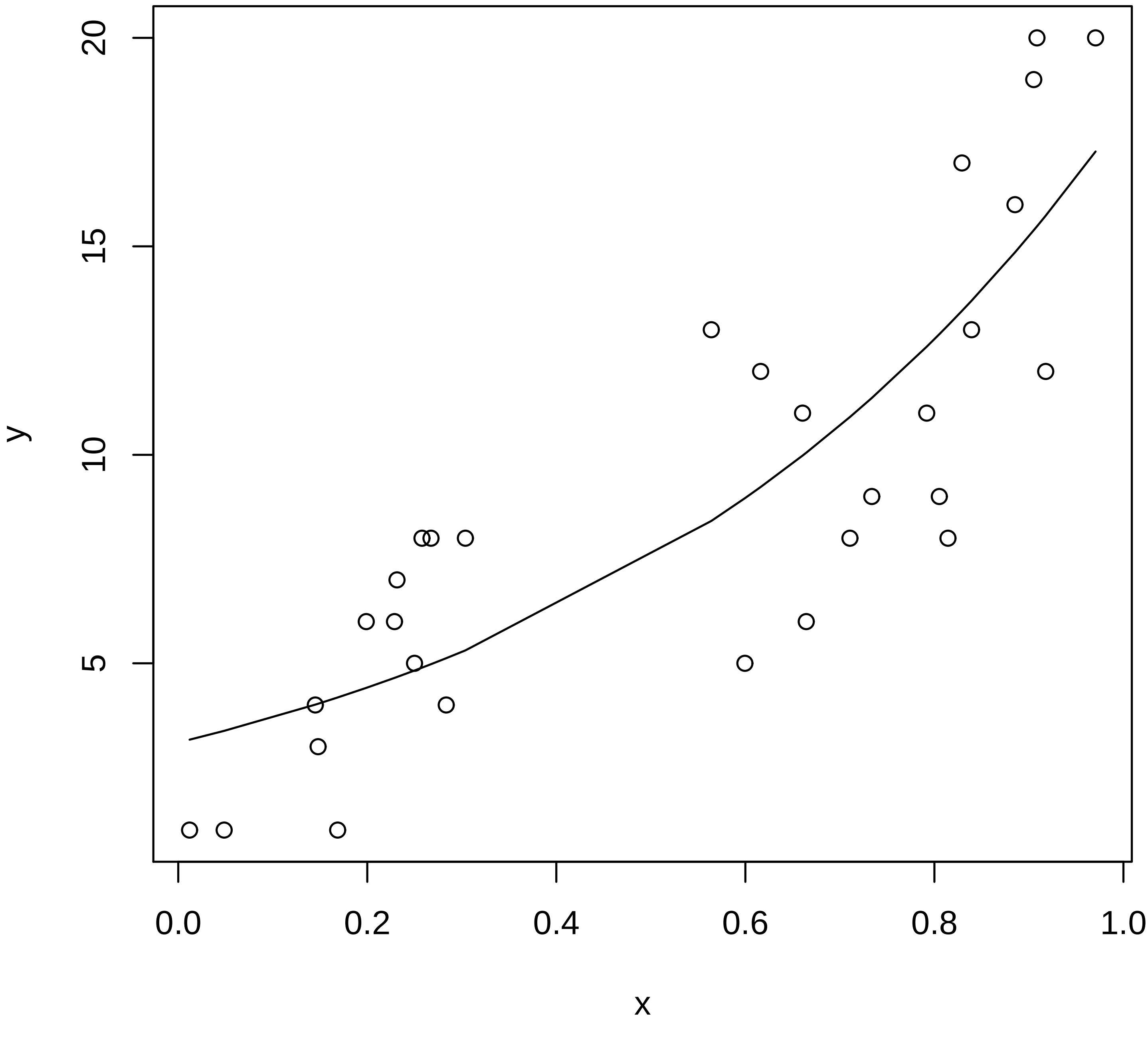
Link function $g(\mu) = \eta$ where $\eta = \mathbf{x}\beta$:

Canonical link $g(\mu) = \kappa'^{-1}(\mu) = \log(\mu)$

Log-likelihood

$$
\ell(\beta; y) = \sum_{i=1}^{n} y_i \mathbf{x}_i \beta - e^{\mathbf{x}_i \beta}
$$

y~Poisson(mu)   mu=exp(1+2x)



Example GLM for $Y_i \sim \text{Binomial}(m, p_i)$

$$\begin{aligned} f(y|\theta) &= C_y^m p^y (1-p)^{m-y} \\ &= \exp\left(y \log\left(\frac{p}{1-p}\right) + m\log(1-p) + \log(C_y^m)\right) \end{aligned}$$

$\theta = \log\left(\frac{p}{1-p}\right)$ (log odds), $\phi = 1$

$\kappa(\theta) = -m\log(1-p) = m\log(1+\exp(\theta))$

Check $\mathrm{E}(Y) = mp$:

$$\kappa'(\theta) = m\exp(\theta)/(1+\exp(\theta)).$$

Exercise Check variance $\phi V = mp(1-p)$.

Canonical link $g(\mu) = \kappa'^{-1}(\mu)$

$$g(\mu) = \log\left(\frac{\mu/m}{1-\mu/m}\right)$$

Log-likelihood

$$\ell(\beta; y) = \sum_{i=1}^{n} y_i \mathbf{x}_i\beta - m\log(1+e^{\mathbf{x}_i\beta})$$

GLM MLEs

The MLEs $\beta = \widehat{\beta}$ for the GLM satisfy score equations

$$\frac{\partial \ell(\beta; y)}{\partial \beta_i} = 0, \quad i = 1, 2, ..., p.$$

$\beta \in R^p, \ell(\beta; y) : R^p \to R.$

In terms of the gradient operator

$$\frac{\partial \ell}{\partial \beta} = \left(\frac{\partial \ell}{\partial \beta_1}, \frac{\partial \ell}{\partial \beta_2}, ..., \frac{\partial \ell}{\partial \beta_p} \right)^T$$

$$\frac{\partial \ell}{\partial \beta} = 0 \quad \text{at } \beta = \widehat{\beta}.$$

Hessian operator is

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta_1^2} & \frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_p} \\ \frac{\partial^2 \ell}{\partial \beta_2 \partial \beta_1} & & & \frac{\partial^2 \ell}{\partial \beta_2 \partial \beta_p} \\ \vdots & & & \vdots \\ \frac{\partial^2 \ell}{\partial \beta_p \partial \beta_1} & \frac{\partial^2 \ell}{\partial \beta_p \partial \beta_2} & \cdots & \frac{\partial^2 \ell}{\partial \beta_p^2} \end{pmatrix}$$

Observed information

$$J(y) = -\partial^2 \ell / \partial\beta\partial\beta^T$$

Expected information

$$I = -\mathsf{E}(\partial^2 \ell / \partial\beta\partial\beta^T)$$

$$\widehat{\beta} \xrightarrow{D} \mathsf{N}(\beta, I^{-1}) \quad \text{with } n.$$

Compute I

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta_i} &= \sum_{k=1}^{n} \frac{\partial \ell}{\partial \eta_k} \frac{\partial \eta_k}{\partial \beta_i} \\
&= \frac{\partial \ell}{\partial \eta^T} \frac{\partial \eta}{\partial \beta_i} \\
\frac{\partial \ell}{\partial \beta^T} &= \frac{\partial \ell}{\partial \eta^T} \frac{\partial \eta}{\partial \beta^T} \\
&= \frac{\partial \ell}{\partial \eta^T} X \\
\frac{\partial^2 \ell}{\partial \beta \beta^T} &= \frac{\partial \eta^T}{\partial \beta} \frac{\partial}{\partial \eta} \frac{\partial \ell}{\partial \beta^T} \\
&= X^T \frac{\partial^2 \ell}{\partial \eta \partial \eta^T} X
\end{aligned}$$

Let

$$W = -\mathsf{E}\left(\frac{\partial^2 \ell}{\partial\eta\partial\eta^T}\right)$$

so that $I = X^T W X$ and

$$\hat\beta \xrightarrow{D} \mathsf{N}(\beta, (X^T W X)^{-1}).$$

W is diagonal since $\ell = \sum_k \ell_k$ with $\ell_k = \ell_k(\eta_k)$.

$$\begin{aligned}
W_{kk} &= -\mathsf{E}\left(\frac{\partial^2 \ell}{\partial\eta_k^2}\right) \\
&= -\mathsf{E}\left(\left[\frac{d\theta_k}{d\eta_k}\right]^2 \frac{\partial^2 \ell}{\partial\theta_k^2}\right) \\
\eta_k &= g(\kappa'(\theta_k)) \\
1 &= \frac{d\theta_k}{d\eta_k} g'(\mu_k)\kappa''(\theta_k) \\
\frac{d\theta_k}{d\eta_k} &= \frac{1}{g'(\mu_k)V(\mu_k)} \\
\frac{\partial^2 \ell}{\partial\theta_k^2} &= -\kappa''/\phi \\
W_{kk} &= \frac{1}{g'(\mu_k)^2 \phi V(\mu_k)}
\end{aligned}$$

Example: Poisson, canonical link

$\phi = 1$, $\kappa(\theta) = \exp(\theta)$, $V(\mu) = \mu$.
Link $g(\mu) = \eta$, canonical link $\log(\mu) = \eta$.

Log-likelihood

$$\ell(\beta; y) = \sum_{i=1}^{n} y_i \eta_i - e^{\eta_i}$$

$$\begin{aligned} W_{kk} &= -\mathsf{E}\left(\frac{\partial^2 \ell}{\partial \eta_k^2}\right) \\ &= e^{\eta_k} \\ &= \mu_k \\ &= \frac{1}{g'(\mu_k)^2 \phi V(\mu_k)} \end{aligned}$$

$$\begin{aligned} I &= J \\ &= X^T \mathrm{diag}(\mu_1, ..., \mu_n) X \end{aligned}$$