4. GLIM for data with constant coefficient of variation

§4.1. Data with constant coefficient of variation and some simple models

- **Coefficient of variation**

  The coefficient of variation of a random variable $Y$ is defined as

  \[ c.v. = \frac{\sqrt{\text{Var}(Y)}}{|EY|}. \]

  If the coefficient of variation is a constant, then

  \[ \text{Var}(Y) = \sigma^2 (EY)^2, \text{ i.e. } V(\mu) = \mu^2. \]

- **Variance stabilization and Log-normal model**

  Note $\frac{Y-\mu}{\mu} \approx \sigma$. If $\sigma$ is small,

  \[ \ln Y = \ln \mu + \frac{1}{\mu} (Y - \mu) - \frac{1}{2\mu^2} (Y - \mu)^2 + \cdots \]

  Hence,

  \[ E \ln Y = \ln \mu - \frac{\sigma^2}{2}, \quad \text{Var}(\ln Y) \approx \sigma^2. \]

  If a multiplicative model can be assumed for $Y$ and the covariates $X_1, \ldots, X_p$, i.e., $\mu = \exp\{\beta_0 + \sum_{j=1}^p \beta_j X_j\} = \exp\{x^t \beta\}$, then $Y$ can be treated as if it has a log-normal distribution, and LS regression can be applied to $\ln Y$ and $x$. 
• **Weighted non-linear LS regression and Gamma model**

Assume the multiplicative model for $Y$ and $X$ and the variance function $V(\mu) = \mu^2$. The $\beta$ can be estimated by iteratively weighted non-linear LS regression: minimizing

$$\sum_{i=1}^{n} \frac{1}{\mu_i^2} (y_i - \exp\{x_i^t \beta\})^2,$$

while treating $\mu_i$ as fixed at each iteration, which is equivalent to solving

$$\frac{\partial}{\partial \beta} \left[ \sum_{i=1}^{n} \frac{1}{\mu_i^2} (y_i - \exp\{x_i^t \beta\})^2 \right] = \sum_{i=1}^{n} \frac{1}{\mu_i^2} (y_i - \exp\{x_i^t \beta\}) \frac{\partial \exp\{x_i^t \beta\}}{\partial \beta} = 0,$$

which, in turn, is equivalent to solving

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{n} \left( \frac{y_i}{\mu_i} - \ln \mu_i \right) = \sum_{i=1}^{n} \left( \frac{y_i}{\mu_i^2} - \frac{1}{\mu_i} \right) \frac{\partial \exp\{x_i^t \beta\}}{\partial \beta} = 0,$$

Weighted non-linear LS regression is equivalent to the GLIM assuming Gamma distribution and the log link function.
• Comparison of efficiency loss

<table>
<thead>
<tr>
<th>Assumed model</th>
<th>True model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-normal</td>
<td>0</td>
<td>G-L</td>
</tr>
<tr>
<td>Gamma</td>
<td>L-G</td>
<td>0</td>
</tr>
</tbody>
</table>

The efficiency loss G-L is larger than the efficiency loss L-G.

§4.2 Gamma distribution and its properties

• Two ways of parameterization:

  Conventional parameterization $G(\alpha, \beta)$:

  \[
  \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} y^{\alpha-1} e^{-y/\beta}.
  \]

  \[
  EY = \alpha \beta, \quad \text{Var}(Y) = \alpha \beta^2.
  \]

  GLIM parameterization in terms of mean and dispersion parameter $\mu$ and $\nu$:

  \[
  \frac{1}{\Gamma(\nu)} \left( \frac{\nu}{\mu} \right)^\nu y^{\nu-1} e^{-y \nu / \mu}.
  \]

  \[
  EY = \mu, \quad \text{Var}(Y) = \mu^2 / \nu.
  \]

• Moment and Cumulant generating functions

  \[
  M(t) = (1 - \mu t / \nu)^{-\nu},
  \]

  \[
  K(t) = -\nu \ln(1 - \mu t / \nu).
  \]
Cumulants:

\[ \kappa_1 = E(Y) = \mu, \]
\[ \kappa_2 = \text{Var}(Y) = \mu^2/\nu, \]
\[ \kappa_3 = E(Y - \mu)^3 = 2\mu^3/\nu^2, \]
\[ \kappa_4 = 6\mu^4/\nu^3, \]
\[ \cdots \]
\[ \kappa_r = (r - 1)!\mu^r/\nu^{r-1}. \]

- **Special cases**
  
  (i) \( \nu = 1 \) — Exponential distribution:
  \[
  \frac{1}{\mu} e^{-y/\mu}. \]
  
  (ii) \( \mu = n, \nu = n/2 \) — \( \chi^2 \)-distribution with d.f \( n \):
  \[
  \frac{1}{2^{n/2}\Gamma(n/2)} y^{n-2} e^{-y/2}. \]

- **Some other properties**
  
(i) If \( Y \sim \Gamma(\mu, \nu) \), then, for \( c > 0 \),
  \[ cY \sim \Gamma(c\mu, \nu). \]

(ii) If \( Y_1, \ldots, Y_n, \text{i.i.d.} \sim \Gamma(\mu, \nu) \), then
  \[
  \sum_{i=1}^n Y_i \sim \Gamma(n\mu, n\nu), \quad \bar{Y} \sim \Gamma(\mu, n\nu). \]
(iii) As $\nu \to \infty$,

$$\sqrt{\nu}(Y - \mu) \to N(0, \mu^2)$$.

- **Shapes of the p.d.f. of Gamma distributions**

Fig. 8.1. *The gamma distribution for $\nu = 0.5, 1.0, 2.0$ and $5.0$, $\mu = 1$.***
§4.3. Models for data with constant coefficient of variation

- **Assumptions**

  Data: \((Y_i, x_i), i = 1, \ldots, n.\)
  
  \[ Y_i \sim \Gamma(\mu_i, \nu \tau_i) \]
  
  \[ \eta_i = g(\mu_i), \]
  
  \[ \eta_i = x_i^t \beta. \]

  **Case of common coefficient of variation:** \(\tau_i \equiv 1.\)

  **Case of different coefficients of variation:** \(\tau_i\) are not all equal.

  **Remark:** The Gamma model assumed above is a prototype of the models for data with constant coefficient of variation. In application, the data needs not necessarily follow a Gamma distribution. The Gamma model can be applied if the constant coefficient of variation can be justified.

  The role the Gamma distribution with multiplicative model plays on data with constant coefficient is to the role the Normal distribution with additive model plays on data with constant variance.

- **Consideration on link function and linear predictor**

  (i) Canonical link, \(\eta(\mu) = 1/\mu,\) and inverse polynomial.

    The canonical link does not map the range of \(\mu\) onto the whole real line. Caution must be taken in the estimation to make sure that \(\hat{\mu} > 0.\)
A example of canonical link in needs:

$x$: plant density (number of plants in unit area).

Since yield per plant is inversely proportional to density, the mean yield per plant can be modeled as

$$\frac{1}{\beta_0 x + \beta_1}$$

Thus, yield per unit area is given by

$$\mu = \frac{x}{\beta_0 x + \beta_1}$$

This gives rise to

$$\eta = \mu^{-1} = \beta_1/x + \beta_0.$$ 

In practical problems, the mean response is subjected to certain restrictions. As a function of the covariates, it might be bounded, have maximum or minimum, or have asymptotes, etc.. The choice of the link function and linear predictor must reflect these natures. The inverse polynomials can be useful in this regards.
Inverse polynomial:

Inverse linear: $\beta_1/x + \beta_0$.

Inverse quadratic: $\beta_1/x + \beta_0 + \gamma_1 x$.

The inverse polynomials have certain desirable properties as illustrated by the following figure:

For more about inverse polynomials, see Nelder (1966), Inverse polynomials, a useful group of multi-factor response functions. *Biometrics*, **22**, 128-141.
(ii) Log-link, $\eta(\mu) = \ln \mu$, and certain particular scales of covariates.

Fig. 8.4. Plots of various logarithmic functions having asymptotes:

(a) $\log(\mu) = 1 + x + 1/x$,  
(b) $\log(\mu) = 1 - x - 1/x$,  
(c) $\log(\mu) = 1 + x - 1/x$,  
(d) $\log(\mu) = 1 - x + 1/x$. 
(iii) Identity link, $\eta(\mu) = \mu$, and the modeling of variance in robust design.

Scheme of Robust Design:

<table>
<thead>
<tr>
<th>Index induced by covariates</th>
<th>Noise levels</th>
<th>1 \cdots k</th>
<th>Mean $Y$</th>
<th>Variance $s^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$Y_{11} \cdots Y_{1k}$</td>
<td>$Y_1$</td>
<td>$s^2_1$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$Y_{21} \cdots Y_{2k}$</td>
<td>$Y_2$</td>
<td>$s^2_2$</td>
<td></td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>i</td>
<td>$Y_{i1} \cdots Y_{ik}$</td>
<td>$Y_i$</td>
<td>$s^2_i$</td>
<td></td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>n</td>
<td>$Y_{n1} \cdots Y_{nk}$</td>
<td>$Y_n$</td>
<td>$s^2_n$</td>
<td></td>
</tr>
</tbody>
</table>

Covariates affecting the mean: $X_1, \ldots, X_p$.

Covariates affecting the variance: $Z_1, \ldots, Z_q$.

The model:

$$E\bar{Y}_i = \sum_{j=1}^{p} \beta_j x_{ij},$$

$$E s^2_i = \sum_{l=1}^{q} \gamma_l z_{ij}.$$ 

Note:

$$\frac{(k-1)s^2_i}{\sigma^2_i} \sim \chi^2_{k-1} = \Gamma(k-1, (k-1)/2).$$

$$E s^2_i = \sigma^2_i = \mu_i, \quad \text{Var}(s^2_i) = \frac{\sigma^4_i}{(k-1)/2} = \frac{\mu_i^2}{\nu}.$$
• use of glm

For the case of common coefficient of variance:

\[
\text{glm}(\text{response} \sim \text{model formula}, \text{family} = \text{Gamma}(\text{link}), \text{data}=\text{data.set})
\]

For the case of different coefficient of variance:

\[
\text{glm}(\text{response} \sim \text{model formula}, \text{family} = \text{Gamma}(\text{link}), \text{weight} = \text{w.vector}, \text{data}=\text{data.set})
\]

where \( \text{w.vector}= (\tau_1, \ldots, \tau_n)^t \).

Available link functions:

\[ \log, \text{inverse}, \text{identity} \]

• Estimation of dispersion parameter \( \sigma^2 = 1/\nu \)

\[
\hat{\sigma}^2 = \frac{1}{n - q} \sum_{i=1}^{n} \tau_i \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^2.
\]

Remark: If the Gamma distribution can be assumed, \( \nu \) can be estimated by maximum likelihood estimation. The MLE is obtained by solving

\[
2n(\ln \hat{\nu} - \frac{\Gamma'(\hat{\nu})}{\Gamma(\hat{\nu})}) = D(y; \hat{\mu}),
\]

where \( D(y; \hat{\mu}) \) is the deviance. But (a) MLE is sensitive to rounding errors and (b) MLE is not robust if the assumption of Gamma distribution is violated.
### §4.4. Examples

- **Example 1. Car insurance claims**

<table>
<thead>
<tr>
<th>Policy-holder’s age</th>
<th>Car group</th>
<th>Vehicle age</th>
<th>0–3</th>
<th>4–7</th>
<th>8–9</th>
<th>10+</th>
</tr>
</thead>
<tbody>
<tr>
<td>17–20</td>
<td>A</td>
<td>£ 289</td>
<td>No. 8</td>
<td>£ 282</td>
<td>No. 8</td>
<td>£ 133</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 372</td>
<td>No. 10</td>
<td>£ 249</td>
<td>No. 28</td>
<td>£ 288</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 189</td>
<td>No. 9</td>
<td>£ 288</td>
<td>No. 13</td>
<td>£ 179</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 763</td>
<td>No. 3</td>
<td>£ 850</td>
<td>No. 2</td>
<td>—</td>
</tr>
<tr>
<td>21–24</td>
<td>A</td>
<td>£ 302</td>
<td>No. 18</td>
<td>£ 194</td>
<td>No. 31</td>
<td>£ 135</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 420</td>
<td>No. 59</td>
<td>£ 243</td>
<td>No. 96</td>
<td>£ 196</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 268</td>
<td>No. 44</td>
<td>£ 343</td>
<td>No. 39</td>
<td>£ 293</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 407</td>
<td>No. 24</td>
<td>£ 320</td>
<td>No. 18</td>
<td>£ 205</td>
</tr>
<tr>
<td>25–29</td>
<td>A</td>
<td>£ 268</td>
<td>No. 56</td>
<td>£ 285</td>
<td>No. 55</td>
<td>£ 181</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 275</td>
<td>No. 125</td>
<td>£ 243</td>
<td>No. 172</td>
<td>£ 179</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 334</td>
<td>No. 163</td>
<td>£ 274</td>
<td>No. 129</td>
<td>£ 208</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 383</td>
<td>No. 72</td>
<td>£ 305</td>
<td>No. 50</td>
<td>£ 116</td>
</tr>
<tr>
<td>30–34</td>
<td>A</td>
<td>£ 236</td>
<td>No. 43</td>
<td>£ 270</td>
<td>No. 53</td>
<td>£ 160</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 259</td>
<td>No. 179</td>
<td>£ 226</td>
<td>No. 211</td>
<td>£ 161</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 340</td>
<td>No. 197</td>
<td>£ 260</td>
<td>No. 125</td>
<td>£ 189</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 400</td>
<td>No. 104</td>
<td>£ 349</td>
<td>No. 55</td>
<td>£ 147</td>
</tr>
<tr>
<td>35–39</td>
<td>A</td>
<td>£ 207</td>
<td>No. 43</td>
<td>£ 129</td>
<td>No. 73</td>
<td>£ 157</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 208</td>
<td>No. 191</td>
<td>£ 214</td>
<td>No. 219</td>
<td>£ 149</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 251</td>
<td>No. 210</td>
<td>£ 232</td>
<td>No. 131</td>
<td>£ 204</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 233</td>
<td>No. 119</td>
<td>£ 325</td>
<td>No. 43</td>
<td>£ 207</td>
</tr>
<tr>
<td>40–49</td>
<td>A</td>
<td>£ 254</td>
<td>No. 90</td>
<td>£ 213</td>
<td>No. 98</td>
<td>£ 149</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 218</td>
<td>No. 380</td>
<td>£ 209</td>
<td>No. 434</td>
<td>£ 172</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 239</td>
<td>No. 401</td>
<td>£ 250</td>
<td>No. 253</td>
<td>£ 174</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 387</td>
<td>No. 199</td>
<td>£ 299</td>
<td>No. 88</td>
<td>£ 325</td>
</tr>
<tr>
<td>50–59</td>
<td>A</td>
<td>£ 251</td>
<td>No. 69</td>
<td>£ 227</td>
<td>No. 120</td>
<td>£ 172</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 196</td>
<td>No. 366</td>
<td>£ 229</td>
<td>No. 353</td>
<td>£ 164</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 268</td>
<td>No. 310</td>
<td>£ 250</td>
<td>No. 148</td>
<td>£ 175</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 391</td>
<td>No. 105</td>
<td>£ 228</td>
<td>No. 46</td>
<td>£ 346</td>
</tr>
<tr>
<td>60+</td>
<td>A</td>
<td>£ 264</td>
<td>No. 64</td>
<td>£ 198</td>
<td>No. 100</td>
<td>£ 167</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>£ 224</td>
<td>No. 228</td>
<td>£ 193</td>
<td>No. 233</td>
<td>£ 178</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>£ 269</td>
<td>No. 183</td>
<td>£ 258</td>
<td>No. 103</td>
<td>£ 227</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>£ 385</td>
<td>No. 62</td>
<td>£ 324</td>
<td>No. 22</td>
<td>£ 192</td>
</tr>
</tbody>
</table>
The Variables:
Response Y: average claims in pounds sterling,
Covariates:

Policyholder age (PA): 17-20, 21-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60+;
Car group (CG): A, B, C, D;
Vehicle age (VA): 0-3, 4-7, 8-9, 10+

Determination of Variance function

From the original raw data, for each average, the corresponding sample variance can be computed. The relationship of the variance and the mean can be assessed, which can aid at the determination of the variance function. This can be done by checking the plot:

\[ s_{ijk}^2 \quad vs. \quad y_{ijk}. \]

Since \( \hat{\text{Var}}(y_{ijk}) = s_{ijk}^2/m_{ijk} \), \( m_{ijk} \) must be used as the weights in glm. Thus

\[ V(\mu_{ijk}) = \sigma^2 \mu_{ijk}^2/m_{ijk}. \]

Determination of Link function

An intuitive consideration: if the insurance company wants to determine the policy price for a given period length, it would like to know the rate at which instalments of 1 pound must be paid to service an average claim over the given period. This rate is given by
the inverse of the average claim. Hence the inverse link $\eta_{ijk} = \mu_{ijk}^{-1}$ can be considered.

A more rigorous verification is provided below:

\[ \eta_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_k. \]

The linear predictor should also be subjected to checking.
R codes

PA_factor(rep(rep(c(17,21,25,30,35,40,50,60),
  c(4,4,4,4,4,4,4,4)),4))
CG_factor(rep(rep(c("A","B","C","D"),8),4))
VA_factor(rep(c(0,4,8,10),c(32,32,32,32)))
claim.amount <-
c(289,372,189,763,302,420,268,407,268,275,334,383,
  236,259,340,400,207,208,251,233,254,218,239,387,
  251,196,268,391,264,224,269,385,282,249,288,850,
  194,243,343,320,285,243,274,305,270,226,260,349,
  129,214,232,325,213,209,250,299,227,229,250,228,
  198,193,258,324,133,288,179, 0,135,196,293,205,
  181,179,208,116,160,161,189,147,157,149,204,207,
  149,172,174,325,172,164,175,346,167,178,227,192,
  160, 11, 0, 0,166,135,104, 0,110,264,150,636,
  110,107,104, 65,113,137,141, 0, 98,110,129,137,
  98,132,152,167,114,101,119,123)
claim.num <-
c(8,10,9,3,18,59,44,24,56,125,163,72,43,179,197,104,
  43,191,210,119,90,380,401,199,69,366,310,105,64,228,
  183,62,8,28,13,2,31,96,39,18,55,172,129,50,53,211,
  125,55,73,219,131,43,98,434,253,88,120,353,148,46,
  100,233,103,22,4,1,1,0,10,13,7,2,17,36,18,6,15,39,
  30,8,21,46,32,4,35,97,50,8,42,95,33,10,43,73,20,6,
  1,1,0,0,4,3,2,0,12,10,8,1,12,19,9,2, 14,23,8,0,22,
  59,15,9,35,45,13,1,53,44,6,6)
car.dat<-data.frame(claim.amount,claim.num,PA,CG,VA)
options(contrasts=c("contr.treatment","contr.poly"))
car.glm<-glm(claim.amount~PA+CG+VA,family=Gamma(inverse),
  weight=claim.num,data=car.dat)
summary(car.glm)
Estimated Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.003411</td>
<td>0.000418</td>
<td>8.160993</td>
</tr>
<tr>
<td>PA21</td>
<td>0.000101</td>
<td>0.000436</td>
<td>0.232406</td>
</tr>
<tr>
<td>PA25</td>
<td>0.000350</td>
<td>0.000412</td>
<td>0.848662</td>
</tr>
<tr>
<td>PA30</td>
<td>0.000462</td>
<td>0.000411</td>
<td>1.125999</td>
</tr>
<tr>
<td>PA35</td>
<td>0.001370</td>
<td>0.000419</td>
<td>3.268478</td>
</tr>
<tr>
<td>PA40</td>
<td>0.000969</td>
<td>0.000405</td>
<td>2.395944</td>
</tr>
<tr>
<td>PA50</td>
<td>0.000916</td>
<td>0.000408</td>
<td>2.246508</td>
</tr>
<tr>
<td>PA60</td>
<td>0.000920</td>
<td>0.000416</td>
<td>2.213389</td>
</tr>
<tr>
<td>CGB</td>
<td>0.000038</td>
<td>0.000169</td>
<td>0.223233</td>
</tr>
<tr>
<td>CGC</td>
<td>-0.000614</td>
<td>0.000170</td>
<td>-3.611006</td>
</tr>
<tr>
<td>CGD</td>
<td>-0.001421</td>
<td>0.000181</td>
<td>-7.866716</td>
</tr>
<tr>
<td>VA4</td>
<td>0.000366</td>
<td>0.000101</td>
<td>3.631972</td>
</tr>
<tr>
<td>VA8</td>
<td>0.001651</td>
<td>0.000227</td>
<td>7.281134</td>
</tr>
<tr>
<td>VA10</td>
<td>0.004154</td>
<td>0.000442</td>
<td>9.390544</td>
</tr>
</tbody>
</table>

Estimation of dispersion parameter

rp_resid(car.glm,type="pearson")
dispersion_sum(rp^2)/109
> dispersion
[1] 1.209015
Conclusion

The largest average claims are made by policyholders in the youngest four age groups, i.e., up to age 34, the smallest claims by those aged 35-39, and intermediate claims by those aged 40 and over.

The value of claims decreases with car age.

There are marked differences between the car groups, group D being the most expensive and group C intermediate. No significance difference between groups A and B.

- **Example 2. Developmental rate of Drosophila melanogaster**

<table>
<thead>
<tr>
<th>Temp. °C</th>
<th>Exp. No.</th>
<th>Duration (hours)</th>
<th>Batch size</th>
<th>Std. dev.</th>
<th>Temp. °C</th>
<th>Duration (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.95</td>
<td>25</td>
<td>67.5 ± 0.33</td>
<td>54</td>
<td>2.41</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>16.16</td>
<td>44</td>
<td>57.1 ± 0.12</td>
<td>182</td>
<td>2.28</td>
<td>25.0</td>
<td>0.50</td>
</tr>
<tr>
<td>16.19</td>
<td>26</td>
<td>56.0 ± 0.12</td>
<td>153</td>
<td>1.46</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>17.15</td>
<td>28</td>
<td>48.4 ± 0.12</td>
<td>129</td>
<td>1.40</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>18.20</td>
<td>25</td>
<td>41.2 ± 0.16</td>
<td>64</td>
<td>1.30</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>19.08</td>
<td>33</td>
<td>37.80 ± 0.059</td>
<td>94</td>
<td>0.57</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>20.07</td>
<td>28</td>
<td>33.33 ± 0.080</td>
<td>82</td>
<td>0.73</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>22.14</td>
<td>25</td>
<td>26.50 ± 0.083</td>
<td>57</td>
<td>0.63</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>23.27</td>
<td>28</td>
<td>24.24 ± 0.038</td>
<td>135</td>
<td>0.44</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>21.09</td>
<td>33</td>
<td>22.44 ± 0.029</td>
<td>188</td>
<td>0.40</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>24.81</td>
<td>42</td>
<td>21.13 ± 0.017</td>
<td>217</td>
<td>0.36</td>
<td>25.0</td>
<td>0.50</td>
</tr>
<tr>
<td>24.84</td>
<td>40</td>
<td>21.05 ± 0.027</td>
<td>141</td>
<td>0.46</td>
<td>25.0</td>
<td>0.50</td>
</tr>
<tr>
<td>25.06</td>
<td>27</td>
<td>20.39 ± 0.064</td>
<td>37</td>
<td>0.38</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>25.06</td>
<td>27</td>
<td>20.41 ± 0.037</td>
<td>84</td>
<td>0.34</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>25.80</td>
<td>26</td>
<td>19.45 ± 0.026</td>
<td>196</td>
<td>0.36</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>26.92</td>
<td>33</td>
<td>18.77 ± 0.029</td>
<td>104</td>
<td>0.30</td>
<td>25.1</td>
<td>0.50</td>
</tr>
<tr>
<td>27.68</td>
<td>26</td>
<td>17.79 ± 0.041</td>
<td>148</td>
<td>0.49</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>28.89</td>
<td>29</td>
<td>17.38 ± 0.043</td>
<td>83</td>
<td>0.39</td>
<td>25.1</td>
<td>0.25</td>
</tr>
<tr>
<td>28.96</td>
<td>40</td>
<td>17.26 ± 0.031</td>
<td>95</td>
<td>0.43</td>
<td>25.0</td>
<td>0.50</td>
</tr>
<tr>
<td>29.00</td>
<td>44</td>
<td>17.18 ± 0.023</td>
<td>232</td>
<td>0.50</td>
<td>25.0</td>
<td>0.50</td>
</tr>
<tr>
<td>30.05</td>
<td>26</td>
<td>16.81 ± 0.032</td>
<td>148</td>
<td>0.39</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>30.80</td>
<td>26</td>
<td>16.97 ± 0.028</td>
<td>195</td>
<td>0.39</td>
<td>25.1</td>
<td>0.33</td>
</tr>
<tr>
<td>32.00</td>
<td>33</td>
<td>18.20 ± 0.290</td>
<td>58</td>
<td>2.23</td>
<td>25.1</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Determination of variance function

Plot average duration time against standard deviation:

![Plot showing relationship between standard deviation and duration time]

The plot seems linear, which implies

$$\text{Var(average duration time)} = \frac{\sigma^2 \mu^2}{BS},$$

BS: batch size.
Determination of link function and linear predictor

Plot mean duration time against temperature:

The right panel of the Figure above shows the Arrhenius’s law:

\[ \ln \mu = \frac{c}{T}. \]

The left panel can be well fitted by

\[ \ln \mu = \beta_0 + \beta_1 T + \beta_{-1} / (T - \delta). \]

If \( \beta_{-1} < 0 \), then, as \( T \to \delta \), \( \mu \to \infty \).
R codes and results

melanogaster<-read.table("melanogaster.txt")
attach(melanogaster)
plot(D2,Std)
x<-seq(15,35,by=0.5)
y<-1/x
par(mfrow=c(1,2))
plot(Temp,D2)
plot(x,y*840)
mela.glm<-glm(D2~Temp+I(1/(Temp-58.644)),
family=Gamma(log),weight=BS,data=melanogaster)
summary(mela.glm)
Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.2007</td>
<td>0.0439</td>
<td>72.82830</td>
</tr>
<tr>
<td>Temp</td>
<td>-0.2648</td>
<td>0.0036</td>
<td>-72.35518</td>
</tr>
<tr>
<td>I(1/(Temp - 58.644))</td>
<td>-217.0802</td>
<td>4.3935</td>
<td>-49.40890</td>
</tr>
</tbody>
</table>

(Dispersion Parameter for Gamma family taken to be 0.0159168 )

Minimum duration occurs at

\[ T = \delta - \left(\beta_{-1}/\beta_1\right)^{1/2} = 30.01. \]