TMA4315: Project 1

jototlan@stud.ntnu.no (10018), martigtu@stud.ntnu.no (10037)

Problem 1

a)

Since the response variables $y_i \sim \text{Bernoulli}(\pi_i)$, where $\pi_i = \Pr(y_i = 1 \mid \boldsymbol{x}_i)$. The conditional mean is given by $\text{E}(y_i) = \pi_i$, which is connected to the covariates via the relationship

$$\boldsymbol{x}_i^T \boldsymbol{\beta} = \eta_i = \Phi^{-1}(\pi_i),$$

or equivalently $\pi_i = \Phi(\eta_i)$. This results in the likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i} \\ &= \prod_{i=1}^{n} \Phi(\eta_i)^{y_i} (1-\Phi(\eta_i))^{1-y_i}. \end{aligned}$$

Thus, the log-likelihood becomes

$$l(\boldsymbol{\beta}) = \ln(L(\boldsymbol{\beta})) = \sum_{i=1}^{n} \underbrace{y_i \ln(\Phi(\eta_i)) + (1-y_i)\ln(1-\Phi(\eta_i))}_{=l_i(\boldsymbol{\beta})} = \sum_{i=1}^{n} l_i(\boldsymbol{\beta}).$$

To find the score function, we calculate

$$\begin{aligned} \frac{\partial l_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{y_i}{\Phi(\eta_i)} \frac{\partial \Phi(\eta_i)}{\partial \boldsymbol{\beta}} - \frac{1-y_i}{1-\Phi(\eta_i)} \frac{\partial \Phi(\eta_i)}{\partial \boldsymbol{\beta}} \\ &= \frac{y_i}{\Phi(\eta_i)} \phi(\eta_i) \boldsymbol{x}_i - \frac{1-y_i}{1-\Phi(\eta_i)} \phi(\eta_i) \boldsymbol{x}_i \\ &= \frac{y_i(1-\Phi(\eta_i)) - (1-y_i)\Phi(\eta_i)}{\Phi(\eta_i)(1-\Phi(\eta_i))} \phi(\eta_i) \boldsymbol{x}_i \\ &= \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1-\Phi(\eta_i))} \phi(\eta_i) \boldsymbol{x}_i. \end{aligned}$$

Consequently, the score function is given by

$$\boldsymbol{s}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\partial l_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1-\Phi(\eta_i))} \phi(\eta_i) \boldsymbol{x}_i.$$

The score function can be written on matrix form as

$$\boldsymbol{s}(\boldsymbol{\beta}) = X^T D \Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu}),$$

where X is the design matrix, $D = \text{diag}(\phi(\eta_i))$, $\Sigma = \text{diag}(\text{Var}(y_i)) = \text{diag}(\Phi(\eta_i)(1 - \Phi(\eta_i)))$ and $\boldsymbol{\mu} = [\Phi(\eta_1) \cdots \Phi(\eta_n)]^T$. Next, we find the expected Fisher information, $F(\boldsymbol{\beta})$. We find it by using the result

$$\begin{aligned} F(\boldsymbol{\beta}) &= \text{Var}(\boldsymbol{s}(\boldsymbol{\beta})) = \text{Var}\left(\sum_{i=1}^{n} \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))}\phi(\eta_i)\boldsymbol{x}_i\right) \\ &= \sum_{i=1}^{n}\left[\frac{\phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))}\right]^2 \text{Var}(y_i\boldsymbol{x}_i) = \sum_{i=1}^{n}\left[\frac{\phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))}\right]^2 \boldsymbol{x}_i \text{Var}(y_i)\boldsymbol{x}_i^T \\ &= \sum_{i=1}^{n}\left[\frac{\phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))}\right]^2 \pi_i(1 - \pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^T = \sum_{i=1}^{n}\frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1 - \Phi(\eta_i))}\boldsymbol{x}_i\boldsymbol{x}_i^T, \end{aligned}$$

where in the third equality we have used that the y_i's are independent. The expected Fisher information can also be verified to have this form by the general relationship

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{n}\frac{h'(\eta_i)^2}{\text{Var}(y_i)}\boldsymbol{x}_i\boldsymbol{x}_i^T,$$

where $h'(\eta_i) = \Phi'(\eta_i) = \phi(\eta_i)$ and $\text{Var}(y_i) = \pi_i(1 - \pi_i) = \Phi(\eta_i)(1 - \Phi(\eta_i))$. We also note that the expected Fisher information can be written on matrix form as

$$F(\boldsymbol{\beta}) = X^T W X,$$

where $W = \text{diag}\left(\frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1-\Phi(\eta_i))}\right)$.

b)

The Fisher scoring algorithm is given by

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + F(\boldsymbol{\beta}^{(t)})^{-1}\boldsymbol{s}(\boldsymbol{\beta}^{(t)}).$$

The deviance is defined as

$$D = 2(l_{\text{saturated}} - l(\hat{\boldsymbol{\beta}})),$$

where the saturated model is a model with as many explanatory variables as there are observations. This implies that we can fit a parameter for each data point. For the Bernoulli distribution this parameter is $\hat{\pi}_i = y_i$. This means that the likelihood function of the saturated model is given by

$$L_{\text{saturated}} = \prod_{i=1}^{n}\hat{\pi}_i^{y_i}(1 - \hat{\pi}_i)^{1-y_i} = \prod_{i=1}^{n} y_i^{y_i}(1 + y_i)^{1-y_i} = 1,$$

Where we have used $0^0 = 1$. Consequently, the log-likelihood $l_{\text{saturated}} = \ln(1) = 0$ and the deviance becomes $D = -2l(\hat{\boldsymbol{\beta}})$. Next follows the implementation of `myglm` in R:

```r
Phi <- function(x) return (pnorm(x))
phi <- function(x) return (dnorm(x))

myglm <- function(formula, data, start = NULL){
  # response variable
  resp <- all.vars(formula)[1]
  y <- as.matrix( data[resp] )

  # model matrix
  X <- model.matrix(formula, data)
  n <- dim(X)[1]
  p <- dim(X)[2]

  # starting beta
  if (is.null(start)){
    beta = rep(0, p)
  }
  else {
    beta = start
  }

  # Fisher scoring algorithm
  max_iter <- 50
  tol <- 1e-10
  iter <- 0
  rel.err <- Inf

  F.inv = NULL
  eta = NULL

  while (rel.err > tol & iter < max_iter){
    # Calculate eta.
    eta <- X %*% beta

    # Calculate score.
    D <- diag(as.vector(phi(eta)), n, n)
    Sigma <- diag(as.vector(Phi(eta)*(1 - Phi(eta))), n, n)
    mu.vec <- as.vector(Phi(eta))
    score = t(X) %*% D %*% solve(Sigma) %*% (y - mu.vec)

    # Calculate Fisher information and its inverse.
    W <- diag(as.vector(phi(eta)^2 / (Phi(eta)*(1-Phi(eta)))), n, n)
    F <- t(X) %*% W %*% X
    F.inv <- solve(F)

    # Update beta.
    beta.new <- beta + F.inv %*% score

    iter <- iter + 1
    rel.err <- max(abs(beta.new - beta) / abs(beta.new))
    beta <- beta.new
  }
```

```
  # Calculating std.errors and deviance.
  SE <- sqrt(diag(F.inv))
  D = -2 * sum(y*log(pnorm(eta)) + (1 - y)*log(1 -pnorm(eta)))

  return (list("coefficients" = data.frame("Estimate" = c(beta), "Std.Error" = c(SE)),
               "deviance" = D,
               "vcov" = F.inv))
}
```

c)

To test that our implementation of `myglm` works, we will simulate 1000 Bernoulli draws, each with a random probability $\Phi(x_i)$, where we simulate the x_i's from a uniform distribution. Then the response can be written as $y_i \sim \text{Bernoulli}(\Phi(x_i))$. First we create the data set in R:

```
# simulate covariates.
x = runif(1000, 0, 1)

# draw n Bernoulli with prob Phi(x)
y <- rbinom(1000, 1, pnorm(x))
df <- data.frame(y, x)
```

Next, we fit a model using myglm and print the betas, standard error, variance covariance matrix and the deviance:

```
### fit using myglm
mymodel <- myglm(y ~ x, data = df)

# beta and SE
mymodel$coefficients
```

```
##                Estimate  Std.Error
## (Intercept) -0.08853247 0.08238277
## x            1.03110831 0.14995061
```

```
# vcov
mymodel$vcov
```

```
##              (Intercept)           x
## (Intercept)  0.006786921 -0.01066885
## x           -0.010668849  0.02248519
```

```
# deviance
mymodel$deviance
```

```
## [1] 1234.685
```

Now we can compare these results to the inbuilt `glm` function:

```
### fit using glm
model <- glm(y ~ x, family = binomial(link = "probit"), data = df)

# beta and SE
s <- summary(model)
s$coefficients[1:2,1:2]
```

```
##                Estimate Std. Error
## (Intercept) -0.08853243  0.0823828
```

```
## x            1.03110820  0.1499508
```

```
# vcov
vcov(model)
```

```
##               (Intercept)           x
## (Intercept)  0.006786926 -0.01066887
## x           -0.010668866  0.02248525
```

```
# deviance
model$deviance
```

```
## [1] 1234.685
```

We see that our implementation gives the same numerical values.

Problem 2

a)

First, we load the `juul` data set:

```
library(ISwR)
data(juul)
juul$menarche <- juul$menarche - 1
juul.girl <- subset(juul, age>8 & age<20 & complete.cases(menarche))
```

Next, we fit the desired model with the inbuilt `glm` function.

```
mod.probit <- glm(menarche ~ age, family=binomial(link="probit"), data= juul.girl)
```

We want to test

$$H_0 : \beta_{\texttt{age}} = 0 \quad \text{vs.} \quad H_1 : \beta_{\texttt{age}} \neq 0,$$

for which we use a Wald test, which utilizes that

$$Z = \frac{\hat{\beta}_{\texttt{age}}}{\text{SE}(\hat{\beta}_{\texttt{age}})} \sim \mathcal{N}(0, 1),$$

under the null hypothesis. We calculate the test statistic and p-value below.

```
beta_age <- mod.probit$coefficients[2]
z = beta_age/sqrt(vcov(mod.probit)[2,2])
p_val <- 2*pnorm(-z)
sprintf("%.30f", p_val)
```

```
## [1] "0.000000000000000000000000019869"
```

We see that the p-value is essentially 0, which is strong evidence that `age` has an effect on the response variable.

b)

Relating to the `juul` data set, we define for each observation/individual

$$y_i = \begin{cases} 1, & \text{if menarche has occured.} \\ 0, & \text{if menarche has not occured.} \end{cases}$$

and t_i as the age at the time of examination, which corresponds to `age` in the data set. Let $T_i \sim \mathcal{N}(\mu, \sigma^2)$, where T_i is the time until menarche occurs for the i'th individual. Furthermore, let

$$\begin{aligned} \pi_i &:= P(y_i = 1) = P(T_i \leq t_i) \\ &= P\left(\frac{T_i - \mu}{\sigma} \leq \frac{t_i - \mu}{\sigma}\right) = \Phi\left(\frac{t_i - \mu}{\sigma}\right). \end{aligned}$$

This, in turn, gives

$$\Phi^{-1}(\pi_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma} t_i = \beta_0 + \beta_1 t_i,$$

where $\beta_0 = -\mu/\sigma$ and $\beta_1 = 1/\sigma$. Or equivalently:

$$\sigma = \frac{1}{\beta_1}, \quad \mu = -\frac{\beta_0}{\beta_1}.$$

Due to the functional invariance of the maximum likelihood estimator, we can write the MLEs of σ and μ as

$$\hat{\sigma}(\hat{\beta}_1) = \frac{1}{\hat{\beta}_1}, \quad \hat{\mu}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\hat{\beta}_0}{\hat{\beta}_1},$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the MLEs of β_0 and β_1, respectively. Thus, the maximum likelihood estimates for this data set can be computed as in the code below.

```
mod.probit <- glm(menarche ~ age, family = binomial(link = 'probit'), data = juul.girl)
b <- mod.probit$coefficients
mle.mu <- -b[1]/b[2]
mle.sigma <- 1/b[2]
```

That is, $\hat{\mu} = 13.1856339$ and $\hat{\sigma} = 1.1596528$. The standard errors (estimates of the standard deviation) of $\hat{\mu}$ and $\hat{\sigma}$ can then be computed using the delta method. A first order Taylor expansion of $\hat{\mu}$ gives

$$\hat{\mu} \approx \hat{\mu}(\boldsymbol{b}) + \nabla\hat{\mu}(\boldsymbol{b})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{b}),$$

where we have used the notation $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$ and expanded around $\hat{\boldsymbol{\beta}} = \boldsymbol{b} = (b_0, b_1)^T$. Next, we take the variance of this linear approximation, such that

$$\begin{aligned}\mathrm{Var}(\hat{\mu}) &\approx \mathrm{Var}\left(\nabla\hat{\mu}(\boldsymbol{b})^T\hat{\boldsymbol{\beta}}\right) \\ &= \nabla\hat{\mu}(\boldsymbol{b})^T\mathrm{Var}(\hat{\boldsymbol{\beta}})\nabla\hat{\mu}(\boldsymbol{b}).\end{aligned}$$

Using $\nabla\hat{\mu} = (-1/\hat{\beta}_1, \hat{\beta}_0/\hat{\beta}_1^{\,2})^T$, we can calculate the standard error as in the code below:

```
mod.probit <- glm(menarche ~ age, family = binomial(link = 'probit'), data = juul.girl)
b <- mod.probit$coefficients
grad.mu <- c(-1/b[2], b[1]/b[2]^2)
se.mu <- sqrt(t(grad.mu) %*% vcov(mod.probit) %*% grad.mu)
se.mu
```

```
##           [,1]
## [1,] 0.1185288
```

That is, $\mathrm{SE}(\hat{\mu}) = 0.1185288$. We follow the same procedure to estimate the standard error of $\hat{\sigma}$. First, we approximate it by a first order Taylor series expansion:

$$\hat{\sigma} \approx \hat{\sigma}(\boldsymbol{b}) + \nabla\hat{\sigma}(\boldsymbol{b})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{b}),$$

which implies that the variance can be approximated as

$$\begin{aligned}\mathrm{Var}(\hat{\sigma}) &\approx \mathrm{Var}\left(\nabla\hat{\sigma}(\boldsymbol{b})^T\hat{\boldsymbol{\beta}}\right) \\ &= \nabla\hat{\sigma}(\boldsymbol{b})^T\mathrm{Var}(\hat{\boldsymbol{\beta}})\nabla\hat{\sigma}(\boldsymbol{b}).\end{aligned}$$

Using that $\nabla\hat{\sigma} = (0, -1/\beta_1^2)^T$, we calculate the standard error as in the code below.

```
grad.sigma <- c(0, -1/b[2]^2)
se.sigma <- sqrt(t(grad.sigma) %*% vcov(mod.probit) %*% grad.sigma)
se.sigma
```

```
##             [,1]
## [1,] 0.1090121
```

That is, $\mathrm{SE}(\hat{\sigma}) = 0.1090121$.

c)

We fit the desired model in R:

```
mod.logit <- glm(menarche ~ age, family = binomial(link = 'logit'), data = juul.girl)
mod.logit$coefficients
```

```
## (Intercept)         age
##  -20.013212    1.517289
```

To find the distribution of the T_i's, we start with the cumulative distribution:

$$\Pr(T_i \leq t_i) = \Pr(y_i = 1 \mid t_i) = \pi_i = \frac{1}{1 + e^{-\eta_i}}.$$

The pdf of T_i is then given as

$$\begin{aligned} f_{T_i}(t_i) &= \frac{\mathrm{d}}{\mathrm{d}t_i}\left(\frac{1}{1+e^{-\eta_i}}\right) = \frac{\beta_1 e^{-\beta_0-\beta_1 t_i}}{(1+e^{-\beta_0-\beta_1 t_i})^2} \\ &= \frac{e^{-(t_i-(-\beta_0/\beta_1))/(1/\beta_1)}}{1/\beta_1(1+e^{-(t_i-(-\beta_0/\beta_1))/(1/\beta_1)})^2} = \frac{e^{-(t_i-\mu)/s}}{s(1+e^{-(t_i-\mu)/s})^2}. \end{aligned}$$

This is the logistic distribution, with parameters $\mu = -\beta_0/\beta_1$ and $s = 1/\beta_1$, where we have used the parametrization from Wikipedia. We compute estimates of the mean and variance from the estimates of β_0 and β_1, which are given in the code output above. This gives $\mathrm{E}(T_i) = -\beta_0/\beta_1 \approx 13.1901147$, and $\sqrt{\mathrm{Var}(T_i)} = s\pi/\sqrt{3} = \pi/(\sqrt{3}\beta_1) \approx 1.1954214$.

d)

We now assume that the latent ages follow a log-normal distribution, i.e.

$$T_i \sim \mathrm{Lognormal}(\mu, \sigma^2).$$

This is equivalent to stating that $\ln T_i \sim \mathcal{N}(\mu, \sigma^2)$. Now we can follow the same approach as in 2b):

$$\begin{aligned} \pi_i &:= \Pr(y_i = 1) = \Pr(T_i \leq t_i) = \Pr(\ln T_i \leq \ln t_i) \\ &= \Pr\left(\frac{\ln T_i - \mu}{\sigma} \leq \frac{\ln t_i - \mu}{\sigma}\right) = \Phi\left(\frac{\ln t_i - \mu}{\sigma}\right). \end{aligned}$$

This, in turn, gives

$$\Phi^{-1}(\pi_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}\ln t_i = \beta_0 + \beta_1 \ln t_i,$$

where $\beta_0 = -\mu/\sigma$ and $\beta_1 = 1/\sigma$. Consequently, we fit GLM with a probit link-function on $\ln t_i$:

```
mod.lognorm <- glm(menarche ~log(age), family = binomial(link = "probit"), data = juul.girl)
mu.hat <- -mod.lognorm$coefficients[1]/mod.lognorm$coefficients[2]
sigma.hat <- 1/mod.lognorm$coefficients[2]
```

Exactly as in 2b), due to the functional invariance of MLEs, we can estimate the mean of T_i as

$$\exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right) = 13.1797035,$$

and we can estimate the standard deviation as

$$\sqrt{[\exp(\hat{\sigma}^2) - 1]\exp(2\hat{\mu} + \hat{\sigma}^2)} = 1.1700147.$$

The formulas for mean and standard deviation of the log-normal distribution are gathered from Wikipedia.

e)

The cloglog link function is given by (we drop the index i here)

$$g(\pi) = \text{cloglog}(\pi) = \ln(-\ln(1-\pi)).$$

Since our model is assumed to have the form `menarche ~ log(age)` with the cloglog link, η becomes

$$\eta = \beta_0 + \beta_1 \ln(t),$$

where t is age. Thus the probability that menarche has occurred is given as

$$\pi = g^{-1}(\eta) = 1 - e^{-e^{\eta}} = 1 - e^{-e^{\beta_0 + \beta_1 \ln t}} = 1 - e^{-(e^{\beta_0} t^{\beta_1})},$$

where it was used that $e^{\beta_1 \ln t} = t^{\beta_1}$. Using that π is the cumulative distribution function of T, i.e. $\Pr(T \leq t) = \pi$, it follows that the distribution of T is given by

$$T \sim \frac{\partial}{\partial t}\pi = \frac{\partial}{\partial t}\left(1 - e^{-(e^{\beta_0} t^{\beta_1})}\right) = \beta_1 e^{\beta_0}\ t^{\beta_1 - 1}\ e^{-(e^{\beta_0} t^{\beta_1})}.$$

Recall that the probability density function of a Weibull distribution is given by

$$\text{Weibull}(x; \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$$

where k is the shape parameter and λ is the scale parameter. If $k = \beta_1$ and $\lambda = e^{-\frac{\beta_0}{\beta_1}}$, then

$$\begin{aligned}\text{Weibull}\left(t; e^{-\frac{\beta_0}{\beta_1}}, \beta_1\right) &= \frac{\beta_1}{e^{-\frac{\beta_0}{\beta_1}}}\left(\frac{x}{e^{-\frac{\beta_0}{\beta_1}}}\right)^{\beta_1 - 1} e^{-\left(t\ e^{\frac{\beta_0}{\beta_1}}\right)^{\beta_1}} \\ &= \beta_1 e^{\beta_0}\ t^{\beta_1 - 1}\ e^{-(e^{\beta_0} t^{\beta_1})},\end{aligned}$$

which shows that $T \sim \text{Weibull}\left(t; e^{-\frac{\beta_0}{\beta_1}}, \beta_1\right)$. The shape and scale parameters are given as

$$k = \beta_1, \qquad \lambda = e^{-\beta_0/\beta_1}.$$

Due to the functional invariance of the maximum likelihood estimator, we can write the MLEs of k and λ as

$$\hat{k} = \hat{\beta}_1, \qquad \hat{\lambda} = e^{-\hat{\beta}_0/\hat{\beta}_1}.$$

Fitting the appropriate model in R and finding the estimate for the shape and scale parameter:

```r
mod.cloglog <- glm(menarche ~ log(age), family = binomial(link = 'cloglog'), data = juul.girl)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
beta0 <- mod.cloglog$coefficients[1]
beta1 <- mod.cloglog$coefficients[2]

k <- beta1
lam <- exp(-beta0/beta1)
```

Thus, we find the maximum likelihood estimates $\hat{k} = 13.5719271$ and $\hat{\lambda} = 13.7168936$.

If $W \sim \text{Weibull}(x; \lambda, k)$, the mean and variance of W are given as

$$E[\text{W}] = \lambda\, \Gamma\left(1 + \frac{1}{k}\right), \quad \text{Var}[\text{W}] = \lambda^2 \left(\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right),$$

where Γ is the gamma function. Again, using the functional invariance of MLEs, we find estimates $\hat{\mu}$ and $\hat{\sigma}$ of the mean and standard deviation, respectively.

```r
mu.hat <- lam * gamma(1+1/k)
sigma.hat <- lam * sqrt( gamma(1+2/k) - gamma(1+1/k)^2 )
```

Thus, an estimate of the mean is $\hat{\mu} = 13.2025564$ and an estimate of the standard deviation is $\hat{\sigma} = 1.1882943$.

f)

First, a short summary of the different models:

- (a) probit link, $T \sim \mathcal{N}(\mu, \sigma^2)$, $D = 197.3901002$
- (b) logit link, $T \sim \text{Logistic}(\mu, s)$, $D = 200.6642831$
- (d) probit link, $T \sim \text{Lognormal}(\mu, \sigma^2)$, $D = 198.0463696$
- (e) cloglog link, $T \sim \text{Weibull}(\lambda, k)$, $D = 198.5931916$

The first model (a) is perhaps the most intuitive, as we know that the normal distribution seemingly arises in countless situations in nature. It may also be the simplest model to implement and interpret/explain to people from other disciplines. Model (b) is also a natural choice, as it constitutes a standard logistic regression, which is perhaps the most popular model to do binary regression with. However, both model (a) and (b) allows a small probability that the age, T, is negative, which could be disconcerting. Models (d) and (e) remedy this problem, since the log-normal and Weibull distribution only have a non-zero pdf for positive values. These models also allow for a skewed distribution which could more reasonable depending on the distribution of the sample. Looking at the deviance, D, model (a) has the lowest, which suggests that this model fits the data the best.