Using the Weibull accelerated failure time regression model to predict time to health events

Enwu Liu[1*,2], Karen Lim[1]

1 Musculoskeletal Health and Ageing Research Program, Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Victoria, Australia
2 College of Medicine and Public Health, Flinders University, Adelaide, South Australia, Australia

* enwu.liu@acu.edu.au

Abstract

We describe a statistical method protocol to use a Weibull accelerated failure time (AFT) model to predict time to a health-related event. This prediction method is quite common in engineering reliability research but rarely used for medical predictions such as survival time. A worked example for how to perform the prediction using a published dataset is included.

Keywords

Weibull regression, prediction, survivial time

Introduction

In medical literature, most prediction models are used to predict the probability of an event occurring or condition developing over a specified time period, such as the Framingham 10 Year Risk of General Cardiovascular Disease [1] and FRAX: a Tool for Estimating 10 year Fracture Risk. [2] In engineering reliability research, it is common to use Weibull accelerated failure time (AFT) models to predict 'time to failure', for example the lifespan of a machine; when a component will need replacement; and an optimal maintenance schedule that maximizes reliability of the entire system. [3] Weibull AFT models are also used to predict shelf-life of perishable foods and warranty period of goods. [4, 5] This statistical method estimates when the event will occur without being bound to a defined time period (i.e. absolute time; when a component will need replacement vs. relative time; 10-y risk of needing replacement). In addition to predicting engineering or mechanic events, this statistical method might also be useful for predicting medical events such as fracture, myocardial infarction and death.

In this paper, we have not intended to develop and present a prediction tool. Rather, we aim to illustrate how to use the Weibull AFT model and assess point accuracy in a medical context.

Weibull distribution

The Weibull distribution is also called the type III extreme value distribution. [6] The distribution has three parameters: the location parameter μ, the scale parameter ρ and

the shape parameter γ. The location parameter μ is predetermined as the minimum value in the distribution. In survival or failure analysis, μ of 0 usually selected to produce a two-parameter distribution.

The cumulative distribution function (CDF) for a two-parameter Weibull distributed random variable $T \sim W(\rho, \gamma)$ is given as

$$F_T(t; \rho, \gamma) = 1 - \exp\left[-(\frac{t}{\rho})^\gamma\right] \tag{1}$$

for $t \geq 0, \rho > 0, \gamma > 0$ and $F(t; \rho, \gamma) = 0$ for $t < 0$

The probability density function (PDF) of the Weibull distribution is given as

$$f_T(t; \rho, \gamma) = F^{'}(t; \rho, \gamma) = \frac{\gamma}{\rho}(\frac{t}{\rho})^{\gamma-1} \exp\left[-(\frac{t}{\rho})^\gamma\right] \tag{2}$$

The survival function of the Weibull distribution is given as

$$S_T(t) = 1 - F_T(t) = \exp\left[-(\frac{t}{\rho})^\gamma\right] \tag{3}$$

The mean survival time or mean time to failure (MTTF) is given as

$$\begin{aligned}
E(T) &= \int_0^\infty S(t)dt && \text{(Darth Vader rule)} \\
&= \int_0^\infty \exp\left[-(\frac{t}{\rho})^\gamma\right] dt && \text{let } (\frac{t}{\rho})^\gamma = u \Rightarrow t = \rho u^{\frac{1}{\gamma}} \\
&= \int_0^\infty e^{-u} \rho \frac{1}{\gamma} u^{\frac{1}{\gamma}-1} du \\
&= \rho \frac{1}{\gamma} \Gamma(\frac{1}{\gamma}) && \text{note } \frac{1}{\gamma}\Gamma(\frac{1}{\gamma}) = \Gamma(\frac{1}{\gamma}+1) \\
&= \rho \Gamma(\frac{1}{\gamma}+1)
\end{aligned} \tag{4}$$

Log-Weibull distribution

The log-Weibull distribution is also called Gumbel distribution, or type I extreme value distribution. [7] Let $T \sim W(\rho, \gamma)$ and $Y = g(T) = logT$ is a one-one transformation from support $\mathcal{T} = \{t | t > 0\}$ to $\mathcal{Y} = \{y | -\infty < y < \infty\}$. The inverse of Y is

$$T = g^{-1}(Y) = e^Y$$

and the Jacobian is given as

$$|J| = |\frac{dg^{-1}(Y)}{dY}| = e^Y$$

The probability density function of Y is then derived from (2)

$$
\begin{aligned}
f_Y(y) = f_T(g^{-1}(y))|J| &= \frac{\gamma}{\rho}(\frac{e^y}{\rho})^{\gamma-1}\exp\left[-(\frac{e^y}{\rho})^\gamma\right]e^y = \frac{\gamma}{\rho}\frac{e^{y(\gamma-1)}}{\rho^{(\gamma-1)}}e^y\exp\left[-\frac{e^{\gamma y}}{\rho^\gamma}\right] \\
&= \gamma\frac{e^{\gamma y}}{\rho^\gamma}\exp\left[-\frac{e^{\gamma y}}{\rho^\gamma}\right] \qquad\qquad \text{note } (\rho^\gamma = e^{\gamma\log\rho}) \\
&= \gamma\frac{e^{\gamma y}}{e^{\gamma log\rho}}\exp\left[-\frac{e^{\gamma y}}{e^{\gamma log\rho}}\right] = \gamma e^{\gamma y-\gamma\log\rho}\exp\left[-e^{\gamma y-\gamma log\rho}\right] \\
&= \gamma e^{\gamma(y-\log\rho)}\exp\left[-e^{\gamma(y-\log\rho)}\right] \qquad (\text{let } \gamma = \frac{1}{b} \text{ and } \log\rho = a) \\
&= \frac{1}{b}\exp(\frac{y-a}{b})\exp\left[-\exp(\frac{y-a}{b})\right] \text{ where } -\infty < y < \infty
\end{aligned} \tag{5}
$$

It shows log-Weibull distribution has a Gumbel distribution $G(a,b)$, where $a = \log\rho$ and $b = \frac{1}{\gamma}$. The CDF of the log-Weibull distribution can be derived from the above PDF or by the definition directly.

$$F_Y(y) = P(Y \le y) = P(log(T) \le y) = P(T \le e^y) = F_T(e^y)$$

By Eq(1) we get

$$
\begin{aligned}
F_Y(y) = F_T(e^y) &= 1-\exp\left[-(\frac{e^y}{a})^\gamma\right] = 1-\exp\left[-\frac{e^{\gamma y}}{a^\gamma}\right] \\
&= 1-\exp\left[-\frac{e^{\gamma y}}{e^{\gamma\log\rho}}\right] = 1-\exp\left[-e^{\gamma(y-\log\rho)}\right] \\
&= 1-\exp\left[-\exp(\frac{y-a}{b})\right] \text{ where } \gamma = \frac{1}{b}, \log\rho = a
\end{aligned} \tag{6}
$$

The survival function of $Y = \log(T)$ is given by

$$S_Y(y) = 1 - F_Y(y) = \exp\left[-\exp(\frac{y-a}{b})\right] \tag{7}$$

The hazard function is given by

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y)} = \frac{1}{b}\exp(\frac{y-a}{b})$$

Weibull AFT regression model

Let T be the survival time. Suppose we have a random sample of size n from a target population. For a subject $i(i = 1, 2, ..., n)$, we have observed values of covariates $x_{i1}, x_{i2}, ..., x_{ip}$ and possibly censored survival time t_i. We can express the Weibull AFT model as

$$\log(t_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \sigma\epsilon_i = \boldsymbol{x_i'\beta} + \sigma\epsilon_i, \qquad i = 1, 2, ..., n \tag{8}$$

where $\boldsymbol{\beta} = (\beta_0, ..., \beta_p)$ are the regression coefficients of interest, σ is a scale parameter and $\epsilon_1, ...\epsilon_n$ are independent and identically distributed according to a Gumbel distribution with PDF

$$f_\epsilon(x) = \exp(x)\exp\left[-\exp(x)\right] \tag{9}$$

and the CDF

$$F_\epsilon(x) = 1 - \exp\left[-\exp(x)\right] \tag{10}$$

Note Eq(9) and Eq(10) are equal to set $a = 0, b = 1$ in Eq(5) and Eq(6). We denote it as a $G(0, 1)$ distribution or a standard Gumbel distribution.

Now, let us find the PDF of T from Eq(8)

$$\log(T) = \boldsymbol{x'\beta} + \sigma\epsilon$$
$$\Rightarrow T = e^{\boldsymbol{x'\beta}+\sigma\epsilon}$$
$$\Rightarrow \epsilon = g^{-1}(T) = \frac{\log(T) - \boldsymbol{x'\beta}}{\sigma} \tag{11}$$
$$\Rightarrow |J| = |\frac{d(g^{-1}(T))}{dT}| = \frac{1}{\sigma T} \tag{12}$$

Puting Eq(11) and Eq(12) to Eq(9) we get

$$\begin{aligned} f_T(t) = f_\epsilon(g^{-1}(t))|J| &= \exp(\frac{\log(t) - \boldsymbol{x'\beta}}{\sigma}) \exp\left[-\exp(\frac{\log(t) - \boldsymbol{x'\beta}}{\sigma})\right] \frac{1}{\sigma t} \\ &= (\frac{t}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}} \exp\left[-(\frac{t}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] \frac{1}{\sigma t} \\ &= \frac{1/\sigma}{\exp(\boldsymbol{x'\beta})}(\frac{t}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}-1} \exp\left[-(\frac{t}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] \end{aligned} \tag{13}$$

Comparing Eq(13) with Eq (2) and let $\gamma = \frac{1}{\sigma}$ and $\rho = \exp(\boldsymbol{x'\beta})$, we can see T has a Weibull distribution $T \sim W(\exp(\boldsymbol{x'\beta}), \frac{1}{\sigma})$.

Referring to Eq(3), now the survival function of $T \sim W(\exp(\boldsymbol{x'\beta}), \frac{1}{\sigma})$ can be written as

$$S_T(t) = \exp\left[-(\frac{t}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] \tag{14}$$

Referring to Eq(4) the expected survival time of $W(\exp(\boldsymbol{x'\beta}), \frac{1}{\sigma})$ is given as

$$E(T) = \exp(\boldsymbol{x'\beta})\Gamma(\sigma + 1) \tag{15}$$

Since most statistical software use $\log(T)$ to calculate these parameters, let us show the distribution and characteristics of the $\log(T)$. Let

$$Y = \log(T) = \boldsymbol{x'\beta} + \sigma\epsilon$$
$$\Rightarrow \epsilon = g^{-1}(Y) = \frac{Y - \boldsymbol{x'\beta}}{\sigma} \tag{16}$$
$$\Rightarrow |J| = |\frac{d(g^{-1}(Y))}{dY}| = \frac{1}{\sigma} \tag{17}$$

Putting Eq(16) and Eq(17) to Eq(9), we get

$$f_Y(y) = f_\epsilon(g^{-1}(Y))|J| = \frac{1}{\sigma}\exp(\frac{y - \boldsymbol{x'\beta}}{\sigma}) \exp\left[-\exp(\frac{y - \boldsymbol{x'\beta}}{\sigma})\right] \tag{18}$$

Comparing Eq(18) to Eq(5) we can see that Y i.e. $\log(T)$ has a $G(\boldsymbol{x'\beta}, \sigma)$ distribution. We also can see the error term ϵ, which has a $G(0, 1)$ distribution. similar to the error term in a simple linear regression with a $N(0, \sigma^2)$ distribution.

Referring to Eq(13) and Eq(18), we can see in the Weibull AFT model, T has a Weibull $W(\exp(\boldsymbol{x'\beta}, \frac{1}{\sigma}))$ distribution, and $\log(T)$ has a Gumbel $G(\boldsymbol{x'\beta}, \sigma)$ distribution.

From Eq (7) the survival function of Y i.e. $\log(T)$ is given as

$$S_Y(y) = \exp\left[-\exp(\frac{y - \mathbf{x}'\beta}{\sigma})\right] \tag{19}$$

And the expectation of Y i.e $\log(T)$ is calculated as

$$
\begin{aligned}
E(Y) = E(\log(T)) &= \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sigma} \exp(\frac{y - \boldsymbol{x'\beta}}{\sigma}) \exp\left[-\exp(\frac{y - \boldsymbol{x'\beta}}{\sigma})\right] dy \qquad (\text{let } z = \exp(\frac{y - \boldsymbol{x'\beta}}{\sigma})) \\
&= \frac{1}{\sigma} \int_0^{\infty} (\sigma \log z + \boldsymbol{x'\beta}) z \exp(-z) d(\sigma \log z + \boldsymbol{x'\beta}) \qquad (y = \sigma \log z + \boldsymbol{x'\beta}) \\
&= \int_0^{\infty} (\sigma \log z + \boldsymbol{x'\beta}) \exp(-z) dz = \sigma \int_0^{\infty} \log z \exp(-z) dz + \boldsymbol{x'\beta} \int_0^{\infty} \exp(-z) dz \\
&= \sigma \int_0^{\infty} \frac{\partial}{\partial \alpha} \left[z^{\alpha} e^{-z}\right]_{\alpha=0} dz + \boldsymbol{x'\beta} \\
&= \sigma \frac{\partial}{\partial \alpha} \left[\int_0^{\infty} z^{\alpha} e^{-z} dz\right]_{\alpha=0} (\text{differentiating under the integral, note } \frac{dz^{\alpha}}{d\alpha} = \log(z) z^{\alpha}) \\
&\quad + \boldsymbol{x'\beta} \\
&= \sigma \Gamma(1)' + \boldsymbol{x'\beta} \\
&= \boldsymbol{x'\beta} - \sigma\xi \quad (\text{where } \Gamma(1)' = -\xi \approx -0{\cdot}57721, \ \ \xi \text{ is the Euler-Mascheroni Constant })
\end{aligned}
$$

Note, since $\log(x)$ is a concave down function by Jensen's inequality, $E(\log(T)) \leq \log(E(T))$, Eq (15) is the correct formula to calculate the expected survival time rather than $\exp(\boldsymbol{x'\beta} - \sigma\xi)$.

Parameter estimation for the Weibull AFT model

The parameters of Weibull AFT model can be estimated by the maximum likelihood method. The likelihood function of the n observed $\log(t)$ time, $y_1, y_2, ..., y_n$ is given by

$$
\begin{aligned}
L(\beta, \sigma; y_i) &= \prod_{i=1}^{n} [f_Y(y_i)]^{\delta_i} [S_Y(y_i)]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left\{\frac{1}{\sigma} \exp(\frac{y_i - \boldsymbol{x'\beta}}{\sigma}) \exp\left[-\exp(\frac{y_i - \boldsymbol{x'\beta}}{\sigma})\right]\right\} \left\{\exp\left[-\exp(\frac{y_i - \boldsymbol{x'\beta}}{\sigma})\right]\right\} \quad (20)
\end{aligned}
$$

where δ_i is the event indicator for ith subject with $\delta_i = 1$ if an event has occurred and $\delta_i = 0$ if the event has not occurred. The maximum likelihood estimates $p + 1$ parameters $\sigma, \beta_1 ...., \beta_p$. We can take the log of the likelihood function and use the Newton-Raphson method to calculate these parameters. Most statistical software packages can perform these calculations.

Calculate expected survival time using the Weibull AFT model

In reliability research, expected survival time is called mean time to failure (MTTF) or mean time between failures (MTBF). [8]

Suppose we want to predict a person i's mean survival time t_i using the Weibull AFT model. First, we use the maximum likelihood estimation (MLE) method of Eq(20) to calculate the $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ then by the invariance property of MLE, we use Eq(15) to calculate the predicted MTTF directly.

$$
t_i = E(\hat{t_i}) = \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}})\Gamma(\hat{\sigma} + 1) \qquad (21)
$$

After we calculate the MTTF, we can use the Delta method to calculate the confidence interval for the MTTF. We will treat the predicted MTTF as a function of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$. The standard error of the MTTF can be calculated as

$$SE = \left\{ \begin{pmatrix} \frac{\partial E(\hat{t}_i)}{\partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial E(\hat{t}_i)}{\partial \hat{\sigma}} \end{pmatrix}^t \Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}} \begin{pmatrix} \frac{\partial E(\hat{t}_i)}{\partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial E(\hat{t}_i)}{\partial \hat{\sigma}} \end{pmatrix} \right\}^{\frac{1}{2}} \tag{22}$$

where $\Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}}$ is the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$. The variance-covariance matrix can be estimated by the observed Fisher information of the Weibull AFT model. The (1-α)% confidence interval is given as

$$\hat{t}_i - z_{1-\frac{\alpha}{2}} SE < t_i < \hat{t}_i + z_{1-\frac{\alpha}{2}} SE \tag{23}$$

where α is the type I error and z is the quantile of the standard normal distribution.

Calculate median survival time using Weibull AFT model

Another important statistic in survival analysis is the median survival time or percentile survival time. The pth percentile of survival time is calculated from the survival function. For the individual i the pth percentile of survival time is calculated as

$$S_T(t_i(p)) = \frac{100-p}{100}$$

For the Weibull AFT model, we use Eq(14) to calculate pth survival time of an individual i.

$$\begin{aligned} S_T(t_i) &= \exp\left[-(\frac{t_i}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] = \frac{100-p}{100} \\ &\Rightarrow (\frac{t_i}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}} = -\log(\frac{100-p}{100}) \\ &\Rightarrow t_i = \left[-\log(\frac{100-p}{100})\right]^{\sigma} \exp(\boldsymbol{x_i'\beta}) \end{aligned}$$

After we obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ from Eq(20) and use the the invariance property of MLE, the median survival time is estimated by

$$t_i(50) = (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}}) \tag{24}$$

We can treat the estimated survival time percentile as a function of $\hat{\sigma}$ and $\hat{\boldsymbol{\beta}}$ when p is fixed by using the Delta method to calculate the standard error of predicted pth survival time. The method is the same as in Eq(22) and Eq(23)

Minimum prediction error survival time (MPET)

Both mean and median survival time estimates are biased when the sample size is small and the model includes censoring. To attenuate bias, Henderson et al. purposed a method to find the optimum prediction time with the minimum prediction error. [9] They specified that if an observed survival time t falls in the interval $\frac{p}{k} < t < kp$ where

p is the predicted survival time and $k > 1$ then the prediction is accurate. The probability of prediction error E_k condition on the predicted time p is given as

$$\begin{aligned} P(E_k|p) &= P(T < p/k) + P(T > kp) \qquad \text{(observed time falls outside the bounds)} \\ &= 1 - S_T(p/k) + S_T(kp) \end{aligned}$$

The minimum probability of prediction error $P(E_k|p)$ is achieved when

$$f_T(p/k) = k^2 f(kp) \tag{25}$$

Now let us calculate the minimum prediction error for the Weibull AFT model. From Eq(13) we get

$$\begin{aligned} f_T(p/k) &= \frac{1/\sigma}{\exp(\boldsymbol{x'\beta})}(\frac{p/k}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}-1} \exp\left[-(\frac{p/k}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] \\ k^2 f_T(kp) &= k^2 \frac{1/\sigma}{\exp(\boldsymbol{x'\beta})}(\frac{kp}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}-1} \exp\left[-(\frac{kp}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] \end{aligned}$$

Applying Eq(25) and canceling the common parts, we get

$$k^{1-\frac{1}{\sigma}} \exp\left[-(\frac{p/k}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right] = k^{1+\frac{1}{\sigma}} \exp\left[-(\frac{kp}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}\right]$$

Taking the log of both sides we get

$$(1 - \frac{1}{\sigma})\log(k) - (\frac{p/k}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}} = (1 + \frac{1}{\sigma})\log(k) - (\frac{kp}{\exp(\boldsymbol{x'\beta})})^{\frac{1}{\sigma}}$$

Finally,when rearranging these terms we get

$$p = \left[\frac{\frac{2}{\sigma}\log(k)}{k^{\frac{1}{\sigma}} - k^{-\frac{1}{\sigma}}}\right]^{\sigma} \exp(\boldsymbol{x'\beta}) \tag{26}$$

Here p is the minimum prediction error survival time. We may use the Delta method to get the standard error of the minimum prediction error survival time. Bootstrap methods also could be used to get a confidence interval.

An example to predict the survival time

We use a published larynx cancer dataset of 90 male participants [10] to demonstrate the calculations of the predictions. The dataset includes five variables: stage of cancer ('stage'; 1=stage 1, 2=stage 2, 3=stage 3, 4=stage 4); time to death or on-study time in months ('time'); age at diagnosis of larynx cancer ('age'); year of diagnosis ('diagyr'); and death indicator ('delta'; 0=alive, 1=dead). The authors added a unique particiapnt identifier variable (ID) to the dataset and renamed the variable 'delta' to 'death'. The dataset can be downloaded from *here*. The resulting dataset has the following structure

```
ID stage time age diagyr death
  1     1  0.6  77     76     1
  2     1  1.3  53     71     1
  3     1  2.4  45     71     1
  4     1  2.5  57     78     0
  5     1  3.2  58     74     1
  ... ... ... ...
 46     2  6.2  74     72     1
```

```
47     2  7.0  62     73     1
48     2  7.5  50     73     0
49     2  7.6  53     73     0
50     2  9.3  61     71     0
51     3  0.3  49     72     1
52     3  0.3  71     76     1
... ... ... ...
89     4  3.8  84     74     1
90     4  4.3  48     76     0
```

We will use two predictors to predict the survival time: 1) stage of cancer and 2) age at cancer diagnosis. Since stage of cancer is a categorical variable, we will create three dummy variables for 'stage' and use stage 1 as the default reference group. The Weibull AFT model can be written as

$$\log(T) = \beta_0 + \beta_1 * stage2 + \beta_2 * stage3 + \beta_3 * stage4 + \beta_4 * age + \sigma\epsilon \qquad \epsilon \sim G(0,1)$$

Most statistical software can run the Weibull AFT regression model. Here we display R syntax

```
library(survival)
larynx<-read.csv("D:/larynx.csv")
wr <- survreg(Surv(time, death) ~ factor(stage) + age,
data = larynx,dist="w")
summary(wr)
```

The above syntax yields the following results

```
Call:
survreg(formula = Surv(time, death) ~ factor(stage) + age, data = larynx,
    dist = "w")
                 Value Std. Error      z        p
(Intercept)     3.5288     0.9041  3.903 9.50e-05
factor(stage)2 -0.1477     0.4076 -0.362 7.17e-01
factor(stage)3 -0.5866     0.3199 -1.833 6.68e-02
factor(stage)4 -1.5441     0.3633 -4.251 2.13e-05
age            -0.0175     0.0128 -1.367 1.72e-01
Log(scale)     -0.1223     0.1225 -0.999 3.18e-01
Scale= 0.885
Weibull distribution
Loglik(model)= -141.4   Loglik(intercept only)= -151.1
Chisq= 19.37 on 4 degrees of freedom, p= 0.00066
Number of Newton-Raphson Iterations: 5
n= 90
```

Suppose we wanted to apply all three methods (MTTF, median and MPET) to predict the survival time of patient ID=46, who had stage 2 cancer and was aged 74 years.

We would use Eq(21) to calculate his expected survival time (mean or MTTF)

$$\begin{aligned} MTTF_{46} &= E(\hat{t}_{46}) = \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}})\Gamma(\hat{\sigma}+1) \\ &= exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) * \Gamma(1.885) \\ &= 7.7(months) \end{aligned}$$

We would use Eq(24) to predict his median survival time

$$\begin{aligned} Median_{46} &= (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}}) \\ &= \log(2)^{0.885} exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) \\ &= 5.8(months) \end{aligned}$$

We would use Eq(26) to calculate the MPET and we fix k=2.

$$
\begin{aligned}
MPET_{46} &= \left[\frac{\frac{2}{\sigma}\log(k)}{k^{\frac{1}{\sigma}} - k^{-\frac{1}{\sigma}}}\right]^{\sigma} \exp(\boldsymbol{x'\beta}) \\
&= \left[\frac{\frac{2}{\sigma}\log(k)}{2^{\frac{1}{0.885}} - k^{-\frac{1}{0.885}}}\right]^{0.885} * \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) \\
&= 7.4(months)
\end{aligned}
$$

It seems the three prediction methods MTTF, median and MPET produce estimates close to the actual survival time of patient ID=46, which was 6.2 months.

Note in R build in *predict* function for Weibull AFT model type= "response" calculates $\exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}})$ without the $\Gamma(1+\hat{\sigma})$ and type="lp" calculates $\boldsymbol{x'}\hat{\boldsymbol{\beta}}$ only. We should not use them to predict MTTF and there are no software to calculate the minimum prediction error survival time

We use the Delta method to calculate the 95% confidence interval of the median survival time. The standard error of the predicted median time is calculated using Eq(22)

$$
\begin{aligned}
SE &= \left\{ \begin{pmatrix} \frac{\partial E(\hat{t}_i)}{\partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial E(\hat{t}_i)}{\partial \hat{\sigma}} \end{pmatrix}^t \Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}} \begin{pmatrix} \frac{\partial E(\hat{t}_i)}{\partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial E(\hat{t}_i)}{\partial \hat{\sigma}} \end{pmatrix} \right\}^{\frac{1}{2}} \\
&= \left\{ \begin{pmatrix} \frac{\partial (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}})}{\partial \hat{\sigma}} \end{pmatrix}^t \Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}} \begin{pmatrix} \frac{\partial (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}})}{\partial \hat{\sigma}} \end{pmatrix} \right\}^{\frac{1}{2}} \\
&= \left\{ \begin{pmatrix} (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * stage2 \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * stage3 \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * stage4 \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * age \\ (\log 2)^{\hat{\sigma}} \log(\log 2) \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) \end{pmatrix}^t \Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}} \begin{pmatrix} (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * stage2 \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * stage3 \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * stage4 \\ (\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) * age \\ (\log 2)^{\hat{\sigma}} \log(\log 2) \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}) \end{pmatrix} \right\}^{\frac{1}{2}}
\end{aligned} \tag{27}
$$

First, we need to find the variance-covariance matrix $\Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}}$, which can be calculated by the observed Fisher information of the Weibull AFT model (calculated by most software). Again in R:

```
wr$var
```

We get

```
            (Intercept)   stage2   stage3  stage4      age Log(scale)
(Intercept)       0.817 -0.09049 -0.08479 -0.0444 -0.01114    0.02591
stage2           -0.090  0.16611  0.05319  0.0507  0.00057    0.00016
stage3           -0.085  0.05319  0.10237  0.0567  0.00042   -0.00731
stage4           -0.044  0.05068  0.05668  0.1320 -0.00020   -0.01070
age              -0.011  0.00057  0.00042 -0.0002  0.00016   -0.00026
Log(scale)        0.026  0.00016 -0.00731 -0.0107 -0.00026    0.01501
```

Note, R generates the variance-covariance of coefficients as log(scale), so we need to change the log(scale) to scale by performing extra calculations. Using formulas on page 401 of the book by John Klein [11], set $g_1(\sigma, \boldsymbol{\beta}) = \boldsymbol{\beta}$ and $g_2(\sigma, \boldsymbol{\beta}) = \sigma = e^{\log(\sigma)}$, i.e

$\sigma = e^{\theta}$ is a function of θ and $\theta = \log(\sigma)$, we get

$$
\begin{aligned}
Cov(\beta_0, \sigma,) &= Cov(\beta_0, e^{\log(\sigma)}) = Cov(\beta_0, \log(\sigma)) * \sigma = 0.02292735 \\
Cov(\beta_1, \sigma,) &= Cov(\beta_1, e^{\log(\sigma)}) = Cov(\beta_1, \log(\sigma)) * \sigma = 0.0001403178 \\
Cov(\beta_2, \sigma,) &= Cov(\beta_2, e^{\log(\sigma)}) = Cov(\beta_2, \log(\sigma)) * \sigma - 0.006469443 \\
Cov(\beta_3, \sigma,) &= Cov(\beta_3, e^{\log(\sigma)}) = Cov(\beta_3, \log(\sigma)) * \sigma = -0.009470604 \\
Cov(\beta_4, \sigma,) &= Cov(\beta_4, e^{\log(\sigma)}) = Cov(\beta_4, \log(\sigma)) * \sigma = -0.0002297781 \\
Cov(\sigma) &= Cov(e^{\log(\sigma)}) = (e^{\log(\sigma)})^2 Cov(\log(\sigma)) = \sigma^2 Var(log(\sigma)) = 0.0117501
\end{aligned}
$$

Now we will use these six values to replace the last row and column of the variance-covariance matrix of the coefficients and log(scale). We get

```
      [,1]     [,2]     [,3]    [,4]     [,5]     [,6]
[1,]  0.817 -0.09049 -0.08479 -0.0444 -0.01114  0.02293
[2,] -0.090  0.16611  0.05319  0.0507  0.00057  0.00014
[3,] -0.085  0.05319  0.10237  0.0567  0.00042 -0.00647
[4,] -0.044  0.05068  0.05668  0.1320 -0.00020 -0.00947
[5,] -0.011  0.00057  0.00042 -0.0002  0.00016 -0.00023
[6,]  0.023  0.00014 -0.00647 -0.0095 -0.00023  0.01175
```

This is the $\Sigma_{\hat{\sigma}\hat{\boldsymbol{\beta}}}$ matrix we needed.

If we use SAS software, we can directly get the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ by using the following statements

```
proc lifereg data=larynx order=data COVOUT outest=est;
class stage;
model time*death(0)=stage age/dist=weibull;
run;
proc print data=est;
run;
```

The right side vector of the Eq(27) is calculated as

$$
\begin{pmatrix}
(\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}) \\
(\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}) * stage2 \\
(\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}) * stage3 \\
(\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}) * stage4 \\
(\log 2)^{\hat{\sigma}} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}) * age \\
(\log 2)^{\hat{\sigma}} \log(\log 2) \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}})
\end{pmatrix}
$$

$$
= \begin{pmatrix}
(\log 2)^{0.885} \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) \\
(\log 2)^{0.885} \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) * 1 \\
(\log 2)^{0.885} \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) * 0 \\
(\log 2)^{0.885} a \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) * 0 \\
(\log 2)^{0.885} \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74) * 74 \\
(\log 2)^{0.885} \log\log(\log 2) \exp(3.5288 - 0.1477 * 1 - 0.5866 * 0 - 1.5441 * 0 - 0.0175 * 74)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
5.8383 \\
5.8383 \\
0 \\
0 \\
432.03 \\
-7.7915
\end{pmatrix}
$$

Now we have everything to calculate the standard error of the median survival time.

$$SE = \left\{ \begin{pmatrix} 5.8383 \\ 5.8383 \\ 0 \\ 0 \\ 432.03 \\ -7.7915 \end{pmatrix}^t \begin{pmatrix} 0.817 & -0.0905 & -0.0848 & -0.0444 & -0.0111 & 0.0229 \\ -0.090 & 0.1661 & 0.0532 & 0.0507 & 0.00057 & 0.00014 \\ -0.0859 & 0.0532 & 0.1024 & 0.0567 & 0.00042 & -0.0065 \\ -0.044 & 0.0507 & 0.0567 & 0.1320 & -0.00020 & -0.0095 \\ -0.011 & 0.00057 & 0.0004 & -0.00020 & 0.00016 & -0.00023 \\ 0.023 & 0.00014 & -0.0065 & -0.0095 & -0.00023 & 0.01175 \end{pmatrix} \begin{pmatrix} 5.8383 \\ 5.8383 \\ 0 \\ 0 \\ 432.03 \\ -7.7915 \end{pmatrix} \right\}^{\frac{1}{2}}$$
$$= 2.156133$$

Therefore, the 95% confidence interval is given as

$$95\%CI : (5.83 - 1.96 * 2.16 < Median_{46} < 5.83 + 1.96 * 2.16) = (1.60 \text{ to } 10.01) \text{ months.}$$

This result indicates we are 95% confident that the survival time will be within 1.60 to 10.01 months. We can also use the R build in function *predict* to predict the median survival time.

```
Median46<-predict(wr, newdata=data.frame(stage=2,age=74),type="quantile",
p=0.5,se.fit=TRUE)
Median46
```

We get

```
$fit
 5.838288
$se.fit
2.095133
```

This standard error differs slightly to our calculations as R uses Greedwood's formula to calculate the standard error of the survival function. [11]

Assess the point prediction accuracy

Two methods of assessing accuracy of predicted survival time have been proposed by Parkes and Christakis and Lamont. [12,13] Parkes suggests let t be the observed survival time and p be the predicted time. If $p/k \leq t \leq kp$ then the point prediction p is defined as "accurate" and any value outside the interval is "inaccurate". Parkes proposes $k = 2$ as suitable. As a more strict assessment, Christakis and Lamont purposed a 33% rule to measure accuracy, where the observed time is divided by the predicted survival time and defined as "accurate" if this quotient lies between 0.67 and 1.33. Values less than 0.67 or greater than 1.33 are defined as "errors". We compared the accuracy of our example using Parkes method ($k = 2$) and Christakis' method. The accuracy rate (i.e. proportion of "accurate" prediction over the total sample size) is presented in Table 1.

Discussion

In this paper, we introduced how to use a Weibull AFT model to predict when a health-related event will happen. Expected survival time, median survival time and minimum prediction error survival time from baseline to event were estimated and prediction accuracy assessed using Parkes' and Christakis and Lamont's method. When we fixed $k = 2$ the accuracy was 55.6% for median, 50% for MTTF and 51.1% for MPET. When used the method as suggested by Christakis and Lamont, the accuracy rate decreased to 30.0%,37.8 % and 33.3% (data not shown), respectively. Our example is limited as the sample is small and we only used two predictors. However, with a larger sample size and more predictors, accuracy may improve. In this example, we did not observe minimum prediction error time to be more accurate than median survival time.

Parametric survival models are advantageous in predicting survival time compared to semi-parametric Cox regression models. The Cox regression model which can be specified as $S_i(t|x_i) = S_0(t)^{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}$ cannot predict time directly but calculates the probability of an event occurring within a specified timeframe. However, one disadvantage of parametric survival models such as Weibull AFT is the need to to make stronger assumptions than semi-parametric models. [14]

Currently, most clinical prediction models describe a patient's likelihood of having or developing a certain disease as a traditional probability value or a risk score that is based on calculated probability. [15] However, probabilities are not intuitive to the general population and probability itself can be defined in different ways. [16] In practice, the time axis remains the most natural measure for both clinicians and patients. It is much easier to understand a survival time rather than probability of survival in a certain timepoint. Predicting when an event will happen provides a practical and interpretable guide for clinicians, health care providers and patients and and can help with decision making over remaining lifespan.

Upon recognizing that the Weibull AFT can be adapted from an engineering reliability framework to a medical framework, the next step involves developing a real prediction tool for medical predictive purposes, applying models to larger datasets and performing rigorous internal and external validation procedures. Such steps are outlined in the book Clinical prediction models: a practical approach to development, validation, and updating for the steps to develop a real prediction tool. [17]

References

1. Sullivan LM, Massaro JM, D'Agostino Sr RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Statistics in medicine 2004;23:1631-60.

2. Kanis J, McCloskey E, Johansson H, et al. Development and use of FRAX® in osteoporosis. Osteoporosis international 2010;21:407-13.

3. Ali JB, Chebel-Morello B, Saidi L, et al. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mechanical Systems and Signal Processing 2015;56:150-72.

4. Fu B, Labuza TP. Shelf-life prediction: theory and application. Food Control 1993;4:125-33.

5. Li X, Lu WF, Zhai L, et al. Remaining life prediction of cores based on data-driven and physical modeling methods. Handbook of Manufacturing Engineering and Technology. Springer; 2015. p. 3239-64.

6. Gorgoso-Varela JJ, Rojo-Alboreca A. Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands. Annals of forest science. 2014;71(7):741-50.

7. Lai C-D. Generalized Weibull Distributions. Generalized Weibull Distributions: Springer; 2014. p. 23-75.

8. Ho L, Silva A. Unbiased estimators for mean time to failure and percentiles in a Weibull regression model. International Journal of Quality & Reliability Management.2006;23(3):323-39.

9. Henderson R, Jones M, Stare J. Accuracy of point predictions in survival analysis. Statistics in medicine. 2001;20(20):3083-96.

10. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media; 2005.

11. Collett D. Modelling survival data in medical research Chapman and Hall/CRC; 2015.

12. Parkes CM. Accuracy of predictions of survival in later stages of cancer. British medical journal. 1972;2(5804):29.

13. Christakis NA, Smith JL, Parkes CM, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort studyCommentary: Why do doctors overestimate? Commentary: Prognoses should be based on proved indices not intuition. Bmj. 2000;320(7233):469-73.

14. Nardi A, Schemper M. Comparing Cox and parametric models in clinical studies. Statistics in medicine. 2003;22(23):3597-610.

15. Lee Y-h, Bang H, Kim DJ. How to establish clinical prediction models. Endocrinology and Metabolism. 2016;31(1):38-44.

16. Saunders S. What is probability? Quo vadis quantum mechanics? Springer; 2005. p. 209-38.

17. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer Science & Business Media; 2008.

Table 1. Prediction results and accuracy (last digit in predicted time: 0 inaccurate, 1, accurate)

ID	stage	age	death	time	Median(95% CI)	MTTF	MPET
1	1	77	1	0.6	6.42(3.16,9.68),0	8.47,0	8.11,0
2	1	53	1	1.3	9.77(4.07,15.46),0	12.9,0	12.34,0
3	1	45	1	2.4	11.23(3.07,19.39),0	14.84,0	14.19,0
4	1	57	0	2.5	9.11(4.32,13.89),0	12.03,0	11.5,0
5	1	58	1	3.2	8.95(4.36,13.54),0	11.82,0	11.3,0
6	1	51	0	3.2	10.11(3.88,16.34),0	13.36,0	12.78,0
7	1	76	1	3.3	6.54(3.29,9.78),1	8.62,0	8.25,0
8	1	63	0	3.3	8.2(4.37,12.03),0	10.83,0	10.36,0
9	1	43	1	3.5	11.63(2.72,20.54),0	15.36,0	14.7,0
10	1	60	1	3.5	8.64(4.4,12.89),0	11.41,0	10.91,0
11	1	52	1	4	9.94(3.98,15.89),0	13.13,0	12.55,0
12	1	63	1	4	8.2(4.37,12.03),0	10.83,0	10.36,0
13	1	86	1	4.3	5.49(1.96,9.01),1	7.24,1	6.92,1
14	1	48	0	4.5	10.66(3.52,17.79),0	14.08,0	13.47,0
15	1	68	0	4.5	7.52(4.12,10.91),1	9.92,0	9.49,0
16	1	81	1	5.3	5.99(2.63,9.34),1	7.9,1	7.56,1
17	1	70	0	5.5	7.26(3.96,10.56),1	9.58,1	9.16,1
18	1	58	0	5.9	8.95(4.36,13.54),1	11.82,0	11.3,1
19	1	47	0	5.9	10.84(3.38,18.31),1	14.33,0	13.7,0
20	1	75	1	6	6.65(3.41,9.89),1	8.78,1	8.39,1
21	1	77	0	6.1	6.42(3.16,9.68),1	8.47,1	8.11,1
22	1	64	0	6.2	8.06(4.34,11.77),1	10.64,1	10.18,1
23	1	77	1	6.4	6.42(3.16,9.68),1	8.47,1	8.11,1
24	1	67	1	6.5	7.65(4.19,11.1),1	10.1,1	9.66,1
25	1	79	0	6.5	6.2(2.9,9.5),1	8.18,1	7.83,1
26	1	61	0	6.7	8.49(4.4,12.58),1	11.21,1	10.73,1
27	1	66	0	7	7.78(4.25,11.31),1	10.27,1	9.83,1
28	1	68	1	7.4	7.52(4.12,10.91),1	9.92,1	9.49,1
29	1	73	0	7.4	6.89(3.65,10.12),1	9.09,1	8.69,1
30	1	56	0	8.1	9.27(4.28,14.26),1	12.24,1	11.71,1
31	1	73	0	8.1	6.89(3.65,10.12),1	9.09,1	8.69,1
32	1	58	0	9.6	8.95(4.36,13.54),1	11.82,1	11.3,1
33	1	68	0	10.7	7.52(4.12,10.91),1	9.92,1	9.49,1
34	2	86	1	0.2	4.73(0.68,8.78),0	6.25,0	5.97,0
35	2	64	1	1.8	6.95(2.37,11.54),0	9.18,0	8.78,0
36	2	63	1	2	7.07(2.4,11.75),0	9.34,0	8.93,0
37	2	71	0	2.2	6.15(1.96,10.34),0	8.12,0	7.77,0
38	2	67	0	2.6	6.6(2.22,10.97),0	8.71,0	8.33,0
39	2	51	0	3.3	8.72(2.28,15.17),0	11.52,0	11.02,0
40	2	70	1	3.6	6.26(2.03,10.49),1	8.26,0	7.9,0
41	2	72	0	3.6	6.05(1.89,10.2),1	7.98,0	7.63,0
42	2	81	1	4	5.17(1.13,9.21),1	6.82,1	6.52,1
43	2	47	0	4.3	9.36(1.98,16.73),0	12.36,0	11.82,0
44	2	64	0	4.3	6.95(2.37,11.54),1	9.18,0	8.78,0
45	2	66	0	5	6.71(2.28,11.15),1	8.86,1	8.48,1

Table 1 continued Prediction results and accuracy (last digit in predicted time: 0 inaccurate, 1, accurate)

ID	stage	age	death	time	Median(95% CI)	MTTF	MPET
46	2	74	1	6.2	5.84(1.73,9.94),1	7.7,1	7.37,1
47	2	62	1	7	7.2(2.43,11.97),1	9.51,1	9.09,1
48	2	50	0	7.5	8.88(2.22,15.54),1	11.73,1	11.22,1
49	2	53	0	7.6	8.42(2.38,14.47),1	11.13,1	10.64,1
50	2	61	0	9.3	7.33(2.46,12.2),1	9.67,1	9.25,1
51	3	49	1	0.3	5.83(2.41,9.24),0	7.69,0	7.36,0
52	3	71	1	0.3	3.97(2.19,5.75),0	5.24,0	5.01,0
53	3	57	1	0.5	5.07(2.68,7.45),0	6.69,0	6.4,0
54	3	79	1	0.7	3.45(1.56,5.34),0	4.55,0	4.35,0
55	3	82	1	0.8	3.27(1.32,5.23),0	4.32,0	4.13,0
56	3	49	1	1	5.83(2.41,9.24),0	7.69,0	7.36,0
57	3	60	1	1.3	4.81(2.68,6.94),0	6.35,0	6.07,0
58	3	64	1	1.6	4.48(2.58,6.39),0	5.92,0	5.66,0
59	3	74	1	1.8	3.76(1.96,5.56),0	4.97,0	4.75,0
60	3	72	1	1.9	3.9(2.12,5.68),0	5.14,0	4.92,0
61	3	53	1	1.9	5.43(2.6,8.27),0	7.17,0	6.86,0
62	3	54	1	3.2	5.34(2.63,8.05),1	7.05,0	6.74,0
63	3	81	1	3.5	3.33(1.4,5.27),1	4.39,1	4.2,1
64	3	52	0	3.7	5.53(2.56,8.49),1	7.3,1	6.98,1
65	3	66	0	4.5	4.33(2.49,6.17),1	5.71,1	5.47,1
66	3	54	0	4.8	5.34(2.63,8.05),1	7.05,1	6.74,1
67	3	63	0	4.8	4.56(2.61,6.51),1	6.02,1	5.76,1
68	3	59	1	5	4.89(2.69,7.1),1	6.46,1	6.18,1
69	3	49	0	5	5.83(2.41,9.24),1	7.69,1	7.36,1
70	3	69	0	5.1	4.11(2.32,5.89),1	5.42,1	5.19,1
71	3	70	1	6.3	4.04(2.26,5.82),1	5.33,1	5.1,1
72	3	65	1	6.4	4.41(2.54,6.27),1	5.82,1	5.56,1
73	3	65	0	6.5	4.41(2.54,6.27),1	5.82,1	5.56,1
74	3	68	1	7.8	4.18(2.38,5.98),1	5.52,1	5.28,1
75	3	78	0	8	3.51(1.64,5.38),0	4.63,1	4.43,1
76	3	69	0	9.3	4.11(2.32,5.89),0	5.42,1	5.19,1
77	3	51	0	10.1	5.63(2.52,8.73),1	7.43,1	7.11,1
78	4	65	1	0.1	1.69(0.77,2.61),0	2.23,0	2.14,0
79	4	71	1	0.3	1.52(0.7,2.34),0	2.01,0	1.92,0
80	4	76	1	0.4	1.4(0.61,2.18),0	1.84,0	1.76,0
81	4	65	1	0.8	1.69(0.77,2.61),0	2.23,0	2.14,0
82	4	78	1	0.8	1.35(0.56,2.13),1	1.78,0	1.7,0
83	4	41	1	1	2.57(0.3,4.84),0	3.4,0	3.25,0
84	4	68	1	1.5	1.6(0.74,2.47),1	2.12,1	2.03,1
85	4	69	1	2	1.58(0.73,2.42),1	2.08,1	1.99,1
86	4	62	1	2.3	1.78(0.78,2.79),1	2.35,1	2.25,1
87	4	74	0	2.9	1.44(0.65,2.24),0	1.91,1	1.82,1
88	4	71	1	3.6	1.52(0.7,2.34),0	2.01,1	1.92,1
89	4	84	1	3.8	1.21(0.42,2.01),0	1.6,0	1.53,0
90	4	48	0	4.3	2.28(0.57,3.99),1	3.01,1	2.87,1
Accuracy rate(%)					55.6%(50/90)	50%(45/90)	51.1%(46/90)