University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

01 University of Plymouth Research Outputs University of Plymouth Research Outputs

2017-08-27

A Bayesian Survival Analysis of a Historical Dataset: How Long Do Popes Live?

Stander, Julian

http://hdl.handle.net/10026.1/9194

10.1080/00031305.2017.1328374
The American statistician
Taylor & Francis

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

A Bayesian Survival Analysis of a Historical Dataset: How Long Do Popes Live?

Julian Stander, Luciana Dalla Valle, and Mario Cortina–Borja

Abstract

University courses in statistical modeling often place great emphasis on methodological theory, illustrating it only briefly by means of limited and repeatedly used standard examples. Unfortunately, this approach often fails to actively engage and motivate students in their learning process. The teaching of statistical topics such as Bayesian survival analysis can be enhanced by focusing on innovative applications. Here we discuss the visualization and modelling of a data set of historical events comprising the post–election survival times of popes. Inference, prediction and model checking are performed in the Bayesian framework, with comparisons being made with the frequentist approach. Further opportunities for similar statistical investigations are outlined.

1 Introduction

Often, data about historical events can provide an interesting and well-structured approach to teaching statistical methodology. This paper discusses such a historical data set that provides a teaching example of interest in three very popular areas of modern statistics: data visualization, biostatistics and Bayesian modeling. We have successfully presented this material in a module on data modeling, taught to final-stage undergraduates (that is, students in the third year of full-time university study) in Mathematics and Statistics at Plymouth University, UK. The students had a strong mathematical and statistical background that included some experience of frequentist inference and of using R (R Core Team, 2017) to perform a variety of analysis and modeling tasks. Students responded well to the material with 88% agreeing with the statement 'the teaching methods used helped me to learn' and 81% agreeing that 'the use of technology enhanced my learning'. One student stated that 'it was fun and interesting to see this application of a different way of doing statistics'. After exposure to this example, students were able to undertake a range of Bayesian modeling tasks with little additional support.

In Section 2 we present a data set based on post–election survival times of Roman Catholic popes from 1404 and show how to visualize the information that it contains using a Lexis diagram. Section 3 discusses a survival analysis model for these data, with inference about the model parameters being performed in the Bayesian framework using the JAGS program (Plummer, 2003), and makes comparisons between the Bayesian and frequentist inferential approaches. We also extensively discuss inference when some of the data are 'censored', that is, when we only know a lower bound for some of the post–election survival times. In Section 4 the predictive distribution is used to make statements about future survival times, while simple diagnostics for model checking are presented in Section 5. Finally, Section 6 discusses additional investigations that students may undertake, including those based on other historical data sets, before presenting brief conclusions. Basic JAGS code for Bayesian inference is presented in Appendix 1, while R code for frequentist inference is given in Appendix 2.

2 A Historical Data Set and a Lexis diagram

The Argentinian Jorge Mario Bergoglio was elected Pope of the Roman Catholic Church on 13 March 2013 and took the name Francis. In August 2014, he jokingly announced that he expected to live another two or three years, and that he may even retire within this period (Guardian, 2014). Although like other populations it may be expected that popes have experienced a generally increased longevity, a statistical question naturally springs to mind: how many years can a pope such as the present one expect to live after his election? A statistical answer to this question requires data, and these data can provide teachers with considerable scope for discussion about visualization, biostatistics and Bayesian modeling.

To answer the above question we analysed data on the post–election survival times of popes, starting with Pope Innocent VII, whose pontificate began in 1404. These data, which are supplied online, were obtained from Wikipedia (2016) and confirmed using Kelly and Walsh (2010). The beginning of the fifteenth century was chosen as the starting point, as dates of birth, election and death (or resignation) are accurately documented from then. The resulting data set is based on the 62 popes before Pope Francis. Apart from Pope Gregory XII (resigned 1415, died 1417) and Pope Emeritus Benedict XVI (Joseph Ratzinger, Francis's predecessor, who resigned on 28th February 2013), all the popes analysed died in office. Thus, except for these two popes, the post–election survival time is the same as the pontificate duration. The median age at election of these 62 popes is 63.5 years (interquartile range 12.5 years), meaning that generally they have achieved a certain maturity. The median post–election survival time of the 61 popes excluding Benedict XVI is 9 years (interquartile range 9 years); Benedict XVI's survival time was 11.7 years on 25th December 2016.

The post–election survival times of popes elected since 1404 are displayed in Figure 1 by a Lexis diagram, produced in `R` using the `Epi` package (Carstensen and Plummer, 2011). Benedict XVI and Francis were alive on 25th December 2016, and are indicated by grey triangles. Francis celebrated his 80th birthday on 17th December 2016, while Benedict XVI was 90 years old on 16th April 2017. A Lexis diagram provides an excellent visualization of survival time data that can stimulate discussion. For example, in Figure 1 the lines are longer and tend to finish on higher values in recent pontificates, confirming that popes have experienced a generally increased life expectancy. This is investigated further in Section 3 using survival analysis.

3 Bayesian Survival Analysis

3.1 A Data Model Based on the Weibull Distribution

The historical data set under consideration provides the teacher with an immediate application of survival analysis to understand how the post–election survival time T_i of the i-th pope, $i = 1, \ldots, 62$, depends on the age at and year of election. For illustration purposes, the Weibull distribution is used to model the positive times T_i. The Weibull probability density function of T_i takes the form

$$f_{T_i}(t_i \mid r, \mu_i) = r\,\mu_i\,t_i^{r-1}\exp\left(-\mu_i\,t_i^r\right), \;\; r, \mu_i > 0, \text{ for } t_i > 0, \tag{1}$$

in which t_i is the observed post–election survival time of the i-th pope. The associated cumulative distribution function takes the form

$$\Pr\left(T_i \leq t_i \mid r, \mu_i\right) = 1 - \exp\left(-\mu_i\,t_i^r\right). \tag{2}$$

The first parameter r of the Weibull distribution is a positive scale parameter. We model the log of the second parameter $\mu_i > 0$ as a linear function of the covariates age at election

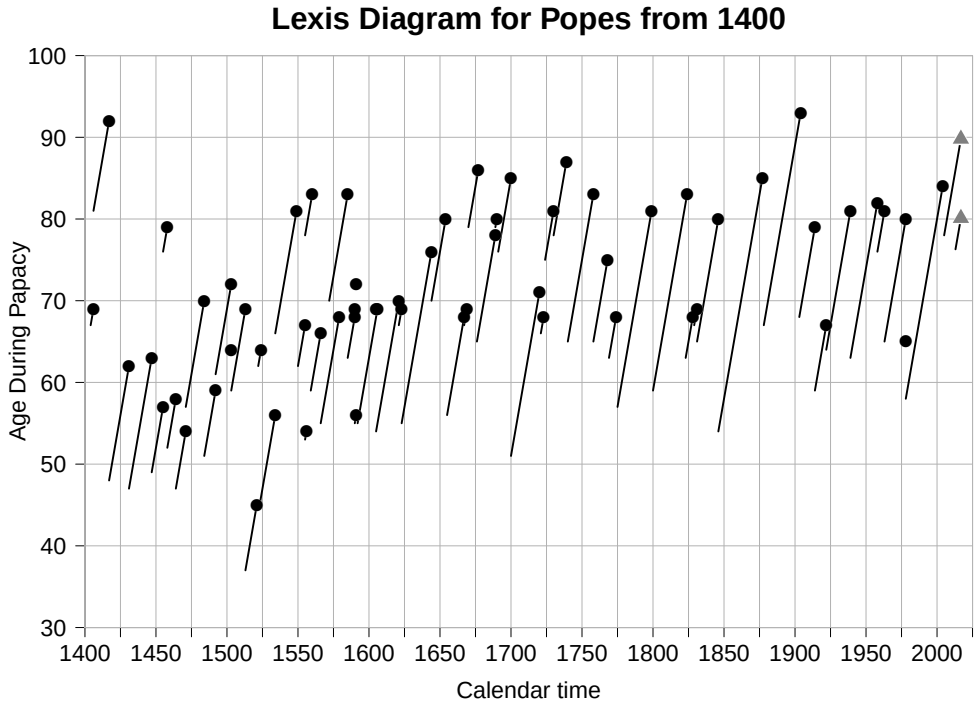

Figure 1: Post–election survival times of popes from 1404. The south–west end of each line indicates the age at and year of election. The north–east end reports the age at and year of death (black circle), or in the case of Pope Emeritus Benedict XVI and Pope Francis their ages on 25th December 2016 (grey triangles).

x_{1i} and year of election x_{2i}, both of which were centered by subtracting the corresponding sample means. Hence, our model takes the form:

$$\begin{aligned} T_i &\sim \text{Weibull}(r, \mu_i) \\ \log(\mu_i) &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \end{aligned} \tag{3}$$

where log is to base e. In this model the scale parameter r is the same for all popes. The Weibull parametrization adopted is the one used by BUGS (Lunn *et al*. (2013), page 346) or JAGS. With this parametrization

$$\text{median}[T_i] = \left\{ \frac{\log(2)}{\mu_i} \right\}^{1/r}, \tag{4}$$

and

$$\text{mean}[T_i] = \Gamma\left(1 + \frac{1}{r}\right) \left(\frac{1}{\mu_i}\right)^{1/r},$$

in which Γ is the gamma function. More generally it can be shown that the p-th quantile of T_i takes the form

$$q_p[T_i] = \left\{ \frac{-\log(1-p)}{\mu_i} \right\}^{1/r}, \ p \in (0,1). \tag{5}$$

The Weibull distribution can provide a starting point for an interesting discussion. In addition to properties of this distribution including its shape and use, the teacher could point out that, if $T \sim \text{Weibull}(r, \mu)$, then

$$T \stackrel{d}{=} \left(\frac{1}{\mu}\right)^{1/r} \epsilon^{1/r}, \tag{6}$$

in which $\stackrel{d}{=}$ means equal in distribution and $\epsilon \sim \text{Exp}(1)$, that is ϵ is an exponential random variable with rate parameter 1. It can immediately be seen from (6) that the exponential distribution is a special case of the Weibull distribution. Students can then be asked to derive that

$$\log(T) \stackrel{d}{=} -\frac{1}{r}\log(\mu) + \frac{1}{r}\log(\epsilon),$$

and hence, using (3), that

$$\log(T) \stackrel{d}{=} -\frac{1}{r}(\beta_0 + \beta_1 x_1 + \beta_2 x_2) + \frac{1}{r}\log(\epsilon). \tag{7}$$

The teacher can then help the students to understand that log survival time, which can take any real value, is being modeled using a mathematical form that is very similar to that adopted in the general linear model. In particular, $\log(T)$ is modeled as a linear function of the covariates plus a scaled random variable that can take any real value. This may enhance their understanding of the model.

Further insight may be gained by asking students to show that the model formulation implies that the log mean/median of T_i are also linear functions of x_{1i} and x_{2i} taking the form $\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i}$, where the coefficients $\alpha_j = -\beta_j/r, j = 0, 1, 2$. They can also show that, because of the logarithmic transformation, $\exp(\alpha_1)$ and $\exp(\alpha_2)$ are the multiplicative scale factors of the mean/median post–election survival time for a unit increase in age at and year of election. In fact, $100\{\exp(\alpha_1) - 1\}$ and $100\{\exp(\alpha_2) - 1\}$ represent the percentage increases in mean/median post election survival time for a unit increase in age at and year of election.

One feature of survival analysis emphasized in many biostatistics courses is 'censoring'. The data under consideration include one censored observation. This is because Benedict XVI was still alive on 25th December 2016 and so his data provide a lower bound of 11.7 years for his survival time. So, assuming that Benedict XVI is indexed by $i = 1$, we know only that $T_1 > 11.7$. This type of censoring is referred to as 'right censoring'; for a full discussion of the different types of censoring, see Cox and Oakes (1984), for example. Francis is still alive but the data set is based on the 62 popes before him, so that data on Francis do not feed into predictions related to him. The teacher could expand the discussion of censoring by explaining its importance in a medical context and showing how the associated data model or likelihood $L(\text{data} \mid r, \beta_0, \beta_1, \beta_2)$ can be modified to take into account censored observations; see Cox and Oakes (1984) and Venables and Ripley (2002) for more general discussions. In particular, when there is no censoring $L(\text{data} \mid r, \beta_0, \beta_1, \beta_2) = \prod_{i=1}^{62} f_{T_i}(t_i \mid r, \mu_i)$, where the probability density function f_{T_i} is defined in (1) and μ_i depends on the parameters β_0, β_1 and β_2 through (3). When this likelihood is modified to take into account the censoring of Benedict XVI's survival time, it becomes

$$L(\text{data} \mid r, \beta_0, \beta_1, \beta_2) = \Pr(T_1 > 11.7 \mid r, \mu_1) \prod_{i=2}^{62} f_{T_i}(t_i \mid r, \mu_i). \tag{8}$$

The 'survivor function' $\Pr(T_1 > 11.7 \mid r, \mu_1) = 1 - \Pr(T_1 \leq 11.7 \mid r, \mu_1) = \exp(-\mu_1 \times 11.7^r)$ from (1). In a recent contribution, Kundu and Mitra (2016) construct a likelihood in a similar way to (8) and perform Bayesian inference on the parameters of the Weibull distribution for left truncated (survival times are observed only if they are greater than a given value) and right censored data. Kundu and Mitra (2016) state that 'it will be of interest to consider the case when there are some covariates.... More work is needed along that direction.' This paper provides an example of such covariate modeling.

3.2 Bayesian Inference

There are four parameters r, β_0, β_1 and β_2 in our model about which to make inference. Many traditional statistics courses would discuss inference about these parameters in the frequentist framework based on the likelihood L, and this is briefly mentioned in Section 3.3. Recently, however, more emphasis has been placed on inference in the Bayesian framework, partly due to the availability of software to perform the associated computations; see Ntzoufras (2009), Brooks *et al*. (2011), Gelman *et al*. (2013), Lunn *et al*. (2013), Kruschke (2015), and Levy and Mislevy (2016) for example.

In the Bayesian framework, inference is based on the posterior distribution of the unknown parameters given the data. In our case this can be written as $\pi(r, \beta_0, \beta_1, \beta_2 \,|\, \text{data})$. By Bayes Theorem, this posterior distribution is proportional to the likelihood of the data multiplied by the prior distribution of the unknown parameters. This prior distribution summarises what may be known about the model parameters before seeing the data. Here prior parameter independence is assumed so that the posterior probability density function takes the form

$$\pi(r, \beta_0, \beta_1, \beta_2 \,|\, \text{data}) \propto L(\text{data} \,|\, r, \beta_0, \beta_1, \beta_2)\, \pi(r)\, \pi(\beta_0)\, \pi(\beta_1)\, \pi(\beta_2). \tag{9}$$

The assumption of prior parameter independence is often made to simplify the specification of the model and its implementation. It does not imply posterior parameter independence. The assumption could be relaxed, for example by adopting a multivariate normal distribution for $(\beta_0, \beta_1, \beta_2)$, if there were specific prior knowledge about parameter dependence.

The following prior distributions for r, β_0, β_1 and β_2 were adopted:

$$\begin{aligned} r &\sim \text{Exp}(\text{rate} = 0.001) \\ \beta_0,\ \beta_1,\ \beta_2 &\sim N(\text{mean} = 0, \text{variance} = 10,000). \end{aligned}$$

These priors support a very wide range of possible parameter values and so represent considerable uncertainty. In our case, it is reasonable to assume such uncertainty as it reflects our lack of previous knowledge about the parameters. Kundu and Mitra (2016) provide some discussion about the choice of the prior on r. Below we explore briefly the sensitivity of the results to prior assumptions. As is now standard (Brooks *et al*., 2011), inference proceeds by simulating values from the posterior probability density function (9) using a Markov chain Monte Carlo algorithm. Our simulation-based inference was implemented in JAGS (Plummer, 2003) using `R2jags` (Su and Yajima, 2015). Basic BUGS/JAGS code is given in Appendix 1.

There are two approaches to handle censored observations in the Bayesian framework, as discussed in Section 9.6 of Lunn *et al*. (2013). The first approach, described briefly in Plummer (2003), treats unknown survival times as model parameters about which to make inference. The second approach is similar in flavour to the one used by Kundu and Mitra (2016) and is based on specifying in the BUGS/JAGS code the contribution to the likelihood function (8) discussed in Section 3.1 from censored observations. We now explain in detail the relevant parts of the code for the first approach given in Appendix 1. The indicator variable `censored` takes the value 1 for Benedict XVI whose survival time is censored and 0 for the other popes. The survival times are stored in the variable `survival`, except in the case of censoring when `NA` ('not available') is assigned: `survival[1] = NA` . The variable `censoring_limits` is defined as `censoring_limits[1] = 11.7` and `censoring_limits[i] = 32`, `i` $= 2, \ldots, 62$, in which 11.7 is the censored survival time for Benedict XVI and 32 is an arbitrarily chosen survival time that must be greater than or equal to all the non-censored survival times. The function `dinterval(s,c)` takes the value 0 if `s` $\leq$ `c` and 1 if `s` $>$ `c`. When `i` $= 1$, the requirement `censored[1] ~ dinterval(survival[1], censoring_limits[1])` becomes `1 ~ dinterval(NA, 11.7)`, which forces the unknown `survival[1]` to be

simulated subject to the constraint that `survival[1]` > 11.7. When `i` $= 2, \ldots, 62$, the requirement `censored[i] ~ dinterval(survival[i], censoring_limits[i])` becomes `0 ~ dinterval(survival[i], 32)` which is automatically satisfied. In general, this requirement would be automatically satisfied for any value greater than or equal to all the non-censored survival times, so the actual value chosen is unimportant provided that it is sufficiently large.

To implement the second approach based on specifying the contribution to the likelihood function from censored observations, we use the 'zeros trick' of Section 9.5.1 of Lunn *et al*. (2013). In particular, we invent an observation $z = 0$ (referred to as `z_censored` in the code) which is assumed to be drawn from a Poisson(ϕ) distribution. Since $\Pr(z = 0) = \exp(-\phi)$, the likelihood contribution of z will be $\exp(-\phi)$. Setting this to the required contribution $\exp\left(-\mu \times 11.7^r\right)$, in which μ is the value of the Weibull parameter μ for the censored observation, means that $\phi = \mu \times 11.7^r > 0$. We found that the results from the two approaches were very largely the same. Lunn *et al*. (2013) point out that the second approach requires the survivor function to be known, so is less generally applicable, although computationally more efficient, than the first approach.

Traceplots of simulated r, β_0, β_1 and β_2 values, together with a variety of convergence diagnostics (Ntzoufras, 2009), suggested that 500,000 simulated values were sufficient. Posterior summaries are given in Table 1. The teacher can explain how the inference procedure is based on hypothetical popes simulated in the light of the observed data and the modeling assumptions. There is much scope for additional discussion including experimentation about the sensitivity of the results to prior assumptions. The prior probability density functions of β_0, β_1 and β_2 adopted above do not have a strong influence on the posterior as they are essentially flat. The effect of specifying stronger prior distributions, such as $\beta_0, \beta_1, \beta_2 \sim N(0, 1)$ and $r \sim \text{Exp}(1)$ that give more weight to smaller parameter values, can be investigated by easily modifying the code. The 2.5% and 97.5% quantiles reported in Table 1 generally change a little to reflect these modifications (r: $(0.840, 1.277)$; β_0: $(-2.950, -1.758)$; β_1: $(0.005, 0.057)$ and β_2: $(-0.003, 0.000)$.)

In the context of our historical data the posterior distribution of the median post–election survival time given by (4) is more interesting than inference about the parameters themselves. One of the beauties of the BUGS/JAGS simulation-based approach is that it is easy to transform values of r, β_0, β_1 and β_2 simulated from the posterior distribution into a sample from the posterior distribution of any function of those parameters simply by computing the values of that function. Here, the values of β_0, β_1 and β_2 are used to compute μ_i through (3), from which the median post-election survival time can be obtained through (4) using the values of r. Code for this is provided in Appendix 1. Figure 2 shows posterior densities, based on 500,000 draws, of the median post–election survival time for hypothetical popes elected in 1750 and 1950, aged 60 and 80 at the time of election. Figure 2 suggests that, as the years have passed, the median survival time has increased. Naturally, popes who are older when elected tend to have shorter pontificates. This can also

	Mean	Standard Deviation	Quantiles 2.5%	5%	50%	95%	97.5%
r	1.151	0.127	0.914	0.947	1.147	1.366	1.412
β_0	−2.623	0.347	−3.334	−3.209	−2.612	−2.071	−1.975
β_1	0.033	0.014	0.006	0.011	0.034	0.056	0.060
β_2	−0.002	0.001	−0.003	−0.003	−0.002	−0.001	0.000

Table 1: Summaries of the posterior density $\pi(r, \beta_0, \beta_1, \beta_2 \,|\, \text{data})$. The posterior mean, standard deviation and selected quantiles including the median are shown. 95% credible intervals can be obtained from the 2.5% and 97.5% quantiles.

be seen from Figure 3 in which we have added to the Lexis diagram shown in Figure 1 the posterior means of the 0.05, 0.5 and 0.9 quantiles of end of pontificate age corresponding to popes aged 55 and 65 years at election. These ages were chosen because there are several popes who began their pontificate at 55 or 65 years. The life trajectories for these popes have been highlighted in Figure 3. Apart from two popes who had very short pontificates, these quantile curves suggest good posterior support for the highlighted data. Popes with very short pontificates are discussed in more detail in Section 5. The fact that the upper quantiles shown in Figure 3 take rather high values is in part due to the assumptions about the right tail made by our parametric Weibull model.

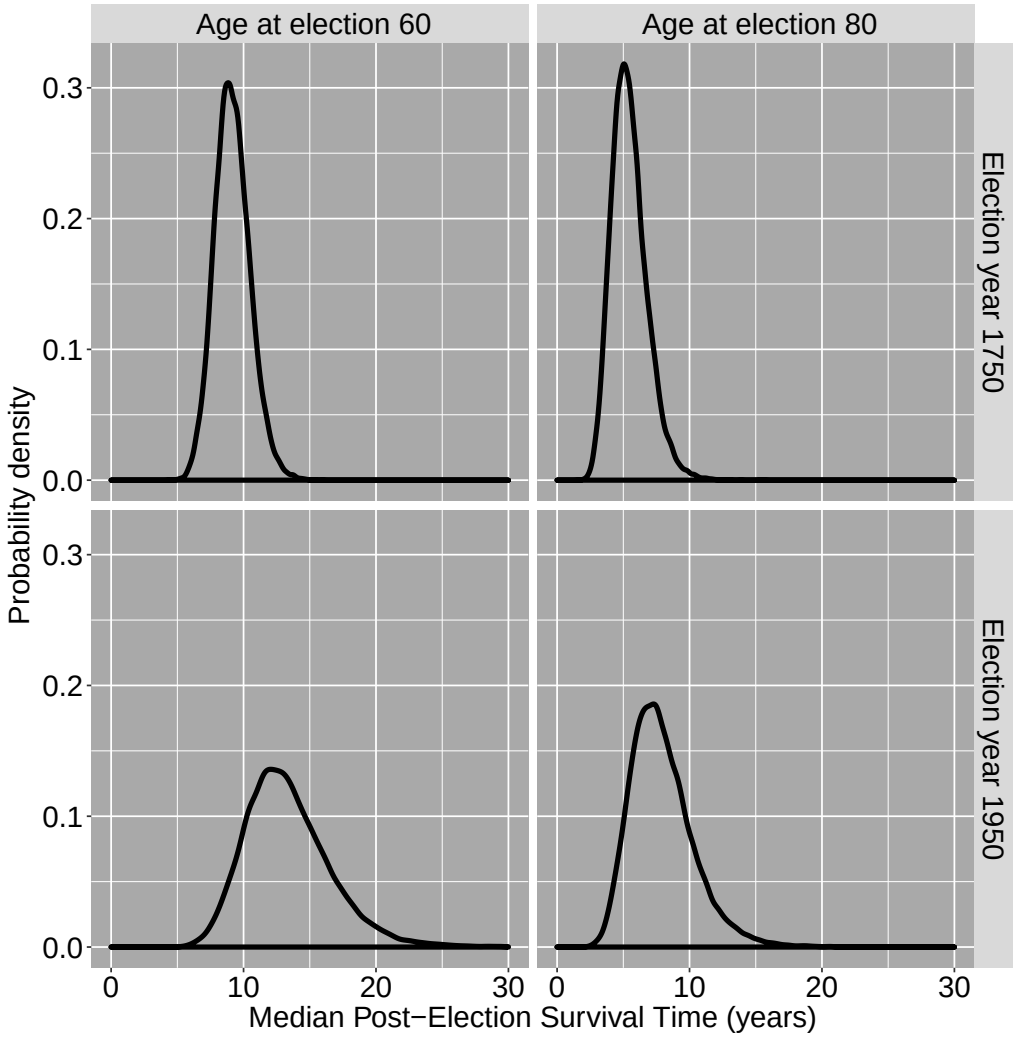

Figure 2: Posterior distributions of the median post–election survival time for popes elected in 1750 and 1950, aged 60 and 80 when elected.

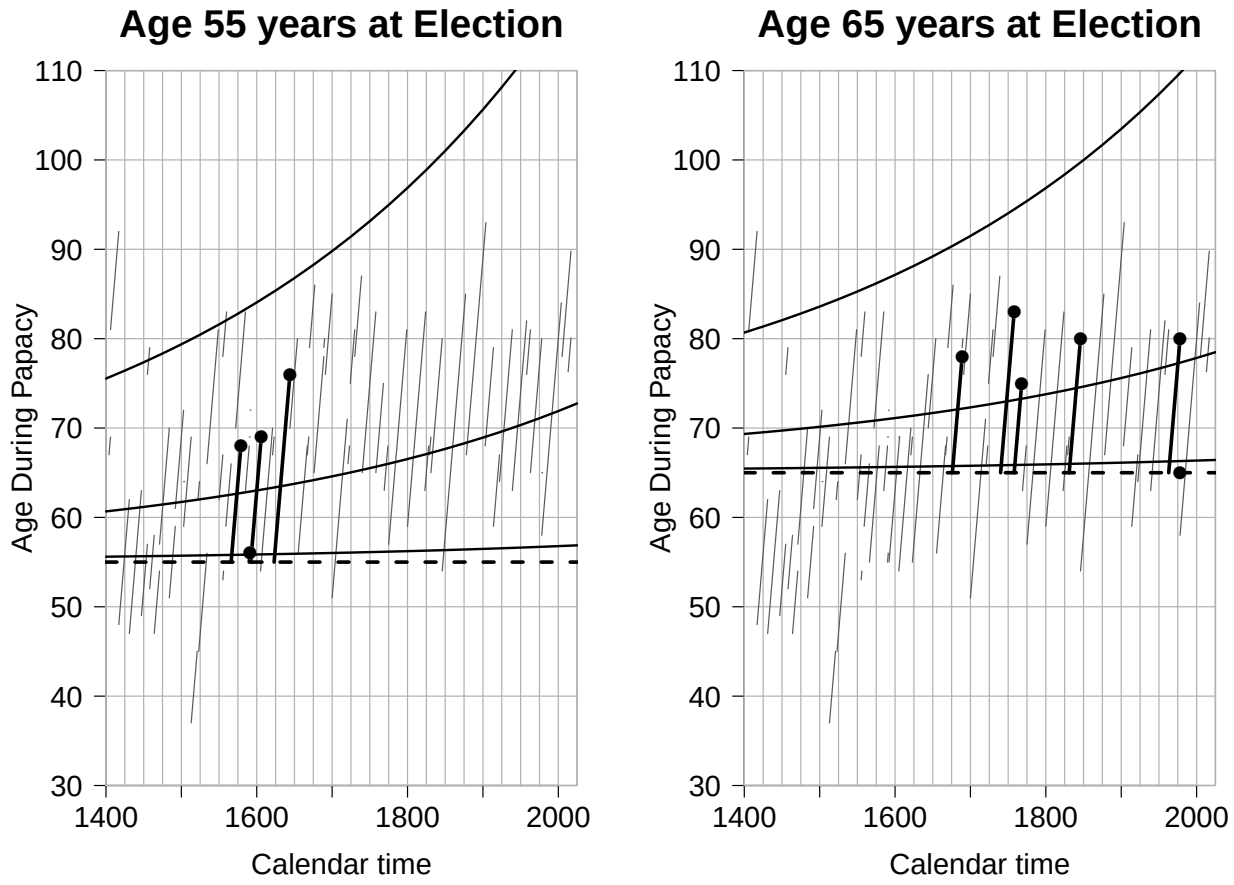

Figure 3: The posterior means of the 0.05 (lower curve), 0.5 (middle curve) and 0.9 (upper curve) quantiles of end of pontificate age corresponding to popes of 55 (left panel) and 65 (right panel) years at election added to the Lexis diagram of Figure 1. The dashed horizontal lines are drawn at 55 and 65 years and popes of this age when elected are highlighted.

We can further quantify Figure 3 by considering the multiplicative scale factors $\exp(\alpha_1)$ and $\exp(\alpha_2)$ of the mean/median post–election survival time for a unit increase in age at and year of election. Posterior means and 95% highest posterior density credible intervals for these quantities and the associated percentage increases $100\{\exp(\alpha_1) - 1\}$ and $100\{\exp(\alpha_2) - 1\}$ are shown in Table 2. The highest posterior density credible intervals were calculated using the function `HPDinterval` available from the `coda` package (Plummer *et al.*, 2006). The posterior means for $\exp(\alpha_1)$ and $\exp(\alpha_2)$ are 0.971 (less than 1) and 1.002 (greater than 1), meaning that a higher age at election leads to a reduced mean/median post election survival time, while a later year of election leads to an increase. Similarly, the posterior means of $100\{\exp(\alpha_1) - 1\}$ and $100\{\exp(\alpha_2) - 1\}$ are -2.860% and 0.163%, meaning that an additional year of age leads to a considerable percentage decrease in these survival times, while the passing of a further year leads to a smaller percentage increase. These quantities can be linked to Figure 3. For example, the posterior mean of the median survival time (the gap between the middle curve and the dashed horizontal line in Figure 3) is approximately 16/12 for popes aged 55/65 years when elected in 2000, so that the ratio between these two approximate survival times is $16/12 = 0.75$ for ages that differ by 10 years. This corresponds well to the posterior mean of $\exp(10\alpha_1)$ which is 0.753. Similarly, the changes in the posterior mean of the median survival times over the 600 year period between 1400 to 2000 are 6.2 to 16/4.4 to 12 years for popes elected at 55/65 years. The ratios $16/6.2 \approx 2.6$ and $12/4.4 \approx 2.7$ correspond well to the posterior mean of $\exp(600\alpha_2)$ of around 2.8.

3.3 The Frequentist Alternative

As our emphasis in this paper is on inference in the Bayesian framework, we only briefly discuss the frequentist alternative. If students wish to make comparisons with frequentist-based maximum likelihood estimation results, they could be directed towards the function `survreg` of the `survival` package (Therneau, 2015). It should, however, be pointed out that the model fitted by `survreg` takes the form

$$\log(T) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \sigma \log(\epsilon),$$

which is parametrized in a different way from (7); see Section 13.2 of Venables and Ripley (2002). In particular, $\alpha_j = -\beta_j/r$, $j = 0, 1, 2$, and a simple comparison with (7) tells us that $\sigma = 1/r$. We provide `R` code in Appendix 2 for performing inference about these parameters, for making predictions and for extracting residuals. Note that standard errors are available for $\log(\sigma)$ rather than for σ. Censoring is handled by creating a

Quantity	Interpretation	Posterior Mean	95% Credible Interval
$\exp(\alpha_1)$	Multiplicative scale factor for age	0.971	$(0.950, 0.994)$
$\exp(\alpha_2)$	Multiplicative scale factor for year	1.002	$(1.0003, 1.0030)$
$100\{\exp(\alpha_1) - 1\}$	Percentage increase for age	-2.860%	$(-5.039, -0.589)\%$
$100\{\exp(\alpha_2) - 1\}$	Percentage increase for year	0.163%	$(0.030, 0.298)\%$

Table 2: Posterior means and 95% highest posterior density credible intervals for the multiplicative scale factors and percentage increases for the mean/median post–election survival time for a unit increase in age at and year of election.

survival object using the `Surv` function that contains information about censoring. Table 3 presents a comparison between Bayesian and frequentist inference for these parameters, and differences are seen to be small. An advantage of using a simulation approach in the Bayesian framework is the ease with which the distributions of parameter transformations such as α_0, α_1, α_2 and $\log(\sigma)$ can be obtained. For a recent, excellent discussion of the Bayesian and frequentist inference approaches see Efron and Hastie (2016), for example.

4 Predictive Distribution

Francis was 76 years old when elected in 2013. His pontificate has lasted 3.79 years as of 25th December 2016. The future survival time T^{new} of a pope such as Francis can be understood by considering the predictive distribution and in particular the probability density function:

$$\pi\left(T^{\text{new}} \,|\, \text{data},\, T^{\text{new}} > 3.79\right), \tag{10}$$

in which 3.79 years is the assumed current pontificate length. The sample from the posterior density $\pi(r, \beta_0, \beta_1, \beta_2 \,|\, \text{data})$, together with simulations from an appropriate Weibull distribution, can be used to generate values from this predictive distribution. All this can be done in a straightforward way in BUGS/JAGS and code is supplied in Appendix 1. Kundu and Mitra (2016) perform inference about probability density functions such as (10), although not in the presence of covariates.

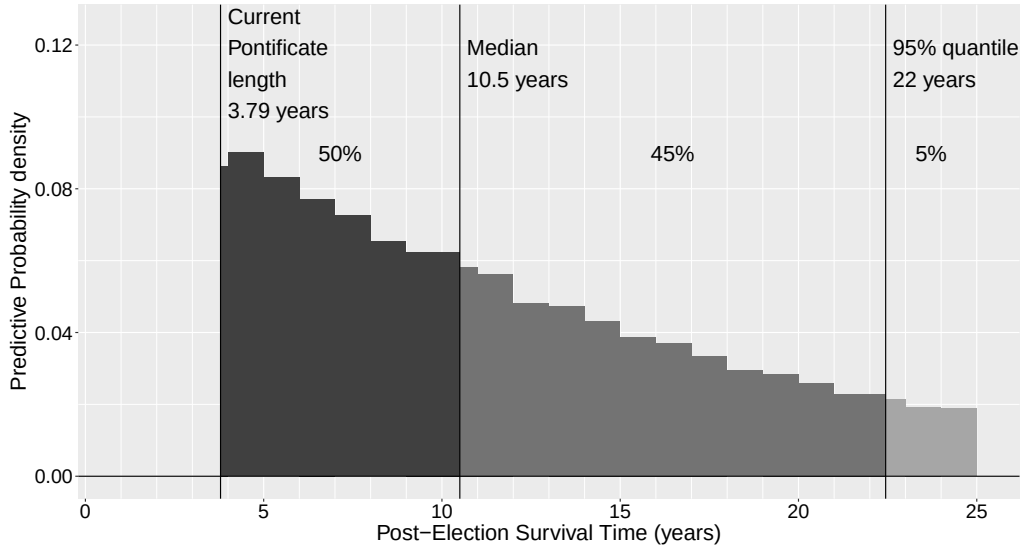

Figure 4: The predictive probability density function of the post–election survival time for a hypothetical pope elected in 2013, aged 76, who, like Francis, has already served 3.79 years, presented as a histogram of simulated future survival times. 50% of simulated times lie below the median of 10.5 years, while 95% of them are below (and 5% are above) the 95% quantile of 22 years.

Figure 4 shows the predictive distribution of the post–election survival time for a pope such as Francis. The median survival time is 10.5 years, with the 95% quantile being approximately 22 years. The corresponding ages at death are approximately 86 and 98 years, leading to the conclusion that the forecast made by Francis in 2014 about his own survival was rather pessimistic: based on the post–election survival of past popes, he can expect to live around 10 years with probability 50%, and 22 years with probability 5%.

This example should provide students with a basic understand of the use of the predictive distribution to make statements about future survival times. Students can then use this methodology to say something about the future survival of hypothetical popes with different

	Bayesian			Frequentist		
	Posterior Mean	Posterior Standard Deviation	Posterior 95% Credible Interval	Parameter Estimate	Standard Error	Approximate 95% Confidence Interval
$\log(\sigma)$	-0.133	0.111	$(-0.343, 0.092)$	-0.133	0.112	$(-0.357, 0.090)$
σ	0.881	0.099	$(0.709, 1.097)$	0.875		$(0.700, 1.094)$
α_0	2.276	0.118	$(2.044, 2.508)$	2.271	0.116	$(2.045, 2.498)$
α_1	-0.029	0.012	$(-0.053, -0.006)$	-0.027	0.011	$(-0.050, -0.005)$
α_2	0.002	0.001	$(0.000, 0.003)$	0.002	0.001	$(0.001, 0.003)$

Table 3: A comparison between Bayesian and frequentist inference. The posterior means, standard deviations and 95% credible intervals for the parameters $\log(\sigma)$, σ, α_0, α_1 and α_2 are presented. Frequentist estimates, standard errors and approximate 95% confidence intervals are also given. `R` does not provide a standard error for the estimate of σ. The approximate 95% confidence interval for $\log(\sigma)$ was calculated using parameter estimate $\pm$ 2 standard errors. The confidence interval for σ was obtained by applying the exponential function to the resulting values.

covariate values. They can, for example, investigate how year of election, age and present pontificate length affect the distribution of future survival times.

5 Diagnostics for Model Checking

A routine part of any statistical analysis should be to use some kind of diagnostics to check the reasonableness of the model. An extensive discussion of Bayesian model checking is given in Chapter 10 of Ntzoufras (2009). Many of the diagnostic techniques described there are based on the predictive distribution, discussed in Section 4. Here a simple approach, based on the posterior distribution of specially defined residuals and motivated by Section 10.3.5 of Ntzoufras (2009), is used to check model (3) . From equation (6) it is easy to show that $\mu T^r = \epsilon \sim \text{Exp}(1)$. Students can be asked to establish that the cumulative distribution function of ϵ takes the form $1 - e^{-\epsilon}$, that $1 - e^{-\epsilon} \sim U[0,1]$ and that $\Phi^{-1}\left(1 - e^{-\epsilon}\right) \sim N(0,1)$, in which $U[0,1]$ is the uniform distribution on $(0,1)$, $N(0,1)$ is the standard normal distribution, Φ is the cumulative distribution function of a standard normal random variable and Φ^{-1} is the associated quantile function. They can then modify the basic BUGS/JAGS code in Appendix 1 to sample from the posterior distribution of the 'transformed residuals' $\Phi^{-1}\left(1 - e^{-\epsilon_i}\right)$, in which $\epsilon_i = \mu_i T_i^r$. It should be noted that the JAGS function `qnorm` needs to be supplied with all its arguments: `qnorm(1 - exp(-epsilon[i]), 0, 1)`. A plot illustrating the posterior distribution of these transformed residuals is shown in Figure 5, together with guide limits at -2 and 2. Five mainly 16-th Century and one 20-th Century pope have been highlighted as they have posterior distributions that give considerable support to very low residual values. These popes had very short pontificates ranging from around 0.03 to 0.17 years. Students could be asked to remove these popes from the data set, to repeat the above analyses and to comment on the differences in results. They may find, for example, that there are some noticeable changes to the results presented in Table 1, that the median post–election survival time for a pope such as Francis increases from around 10 years to considerably more than 11 years, but that the 95% quantile of this survival time remains at around 22 years. As well as these transformed residuals, we also worked with standardized residuals defined as $(T_i - \text{E}[T_i])/\text{sd}[T_i]$. These standardized residuals were less sensitive to the data from popes with short pontificates than the transformed residuals.

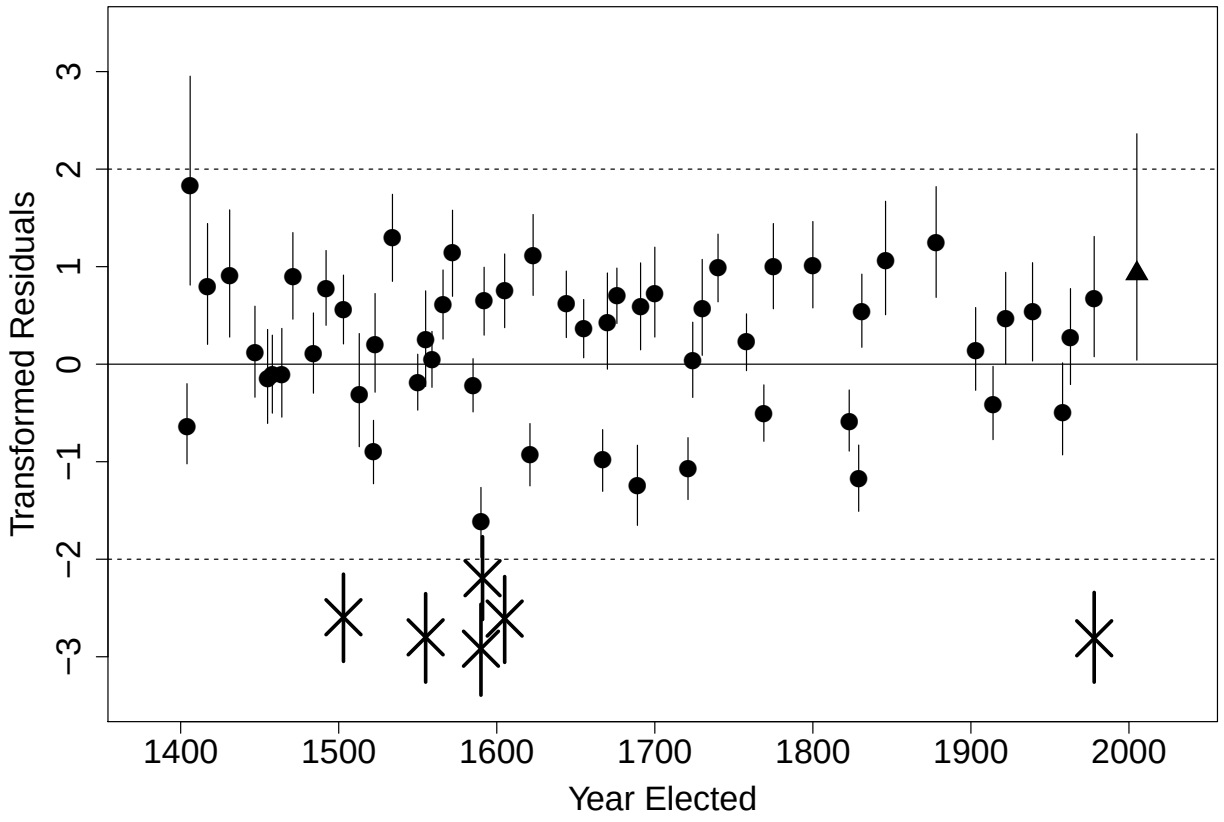

Figure 5: The posterior median and 95% credible interval of all the transformed residuals against year of election. A cross and thicker line is used when the posterior median of the residual is less than -2. The triangle indicates that the corresponding survival time for Benedict XVI is censored.

6 Further Investigations, Different Data and Conclusions

The historical data set used and the problem of quantifying Francis's statement about his own longevity can open up various additional avenues for discussion. For example, our analysis does not take any account of Francis's particular state of health. His life expectancy may be somewhat reduced as he is missing part of a lung (BBC, 2014) and has walking difficulties that have led to falls (BBC, 2016, for example), though one should balance this by his having access to excellent medical care.

There is considerable scope for discussion of the modeling that has been performed. For example, model (3) could be extended to include a multiplicative term $\beta_3 x_{3i}$, in which x_{3i} is the product of the age at and year of election of the i-th pope centered to have mean 0 and scaled to have standard deviation 1 to prevent numerical problems. The posterior support for β_3 could be investigated by students, who would find that there is some support around zero. 95% credible intervals for the other model parameters r, β_0, β_1 and β_2 and their widths are given in Table 4, together with the widths of the corresponding intervals for model (3) given in Table 1. It can be seen that the credible interval widths for the extended model are larger than those for model (3), indicating that the additional term has increased posterior uncertainty. In addition to this investigation, alternative models could also be considered.

Actuarial type tables such as WorldLifeExpectancy (2016) suggest that an 80 year old Argentinian man has a life expectancy of 87.2 years (the corresponding figures for an Argentinian woman, and a US man and woman are 89.5, and 88.7 and 90 years, respectively). Hence our prediction for Francis of 86 years is slightly lower than that given in the published life table, possibly because popes have particular life courses. However, our Bayesian approach offers the full predictive distribution and our modeling provides some historical insights. Such a comparison can introduce students to the use of actuarial tables and can lead to some fun investigations.

Extrapolating conclusions from data based on just popes to the whole male population may be dangerous because of the particular life courses of popes and the fact that their median election age of over 63 years means that they have generally achieved a certain maturity. However, Hanley, Carrieri and Serraino (2006) made a very careful comparative study of the longevity of popes and artists between the 13-th and 19-th century. For each pope they considered artists born in the same year who were still alive when the pope was elected. They found that from the 14-th until the 18-th century artists tended to outlive popes. Again, the class could be engaged in a discussion about how these comparisons could be made, perhaps referring again to Lexis diagrams such as Figure 1; see Hanley, Carrieri and Serraino (2006) for a very interesting discussion.

A similar analysis could be conducted for individuals comprising special populations such as presidents, monarchs and other heads of state. These data are readily available. When analysing populations that include females, a covariate indicating gender should be

	Extended Model		Model (3)
	95% Credible Intervals	Width	Width
r	$(0.937, 1.464)$	0.527	0.498
β_0	$(-3.434, -2.017)$	1.417	1.359
β_1	$(-0.494, 0.108)$	0.602	0.054
β_2	$(-0.023, 0.001)$	0.024	0.003

Table 4: 95% credible intervals and their widths for the parameters r, β_0, β_1 and β_2 when model (3) is extended by the addition of $\beta_3 x_{3i}$. The width of the corresponding intervals for model (3) given in Table 1 are also reported.

included. Discussion about this could be motivated by consideration of the actuarial tables mentioned above. Similar analyses could be conducted for subpopulations with particular mortality schemes, for example people in specific professions. Britton and Shipley (2010) found that among British civil servants 'those reporting being bored are more likely to die younger than those who are not bored'. The class could be asked to discuss how they could investigate whether job interest is a protective factor for people in stressful occupations.

In conclusion, we believe that using historical data provides ample scope for enhancing the teaching and discussion of a range of issues such as data visualization, survival analysis, and Bayesian statistical modeling. It therefore has considerable potential for improving the student learning experience.

Acknowledgements

We would like to thank the Editor, Dr Nicole Lazar, the Associate Editor and the editorial team for very constructive suggestions that have improved the content and presentation of this contribution.

References

BBC (2013), http://www.bbc.co.uk/news/world-europe-21777141.

BBC (2016), http://www.bbc.co.uk/news/36915870.

Britton A., and Shipley M. J. (2010), "Bored to death?" *Int. J. Epidemiol*, **19**, 370–1.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011), *Handbook of Markov Chain Monte Carlo*, Chapman and Hall/CRC.

Carstensen, B., and Plummer, M. (2011), "Using Lexis Objects for Multi-State Models in R," *Journal of Statistical Software*, **38 (6)**, 1–18. http://www.jstatsoft.org/v38/i06/.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall/CRC.

Efron, B., and Hastie, T. (2016), *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian Data Analysis*, second edition, Chapman & Hall.

Guardian (2014),
https://www.theguardian.com/world/2014/aug/19/
pope-francis-expects-live-two-three-years-may-retire

Hanley, J. A., Carrieri, M. P., and Serraino, D. (2006), "Statistical fallibility and the longevity of popes: William Farr meets Wilhelm Lexis, letter to the editor," *Int. J. Epidemiol*, **35** (3), 802–5.

Kelly, J. N. D., and Walsh, M. J. (2010), *Oxford Dictionary of Popes*, Oxford University Press.

Kruschke, J. K. (2015), *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, second edition, Elsevier.

Kundu, D., and Mitra, D. (2016), "Bayesian inference of Weibull distribution based on left truncated and right censored data," *Computational Statistics and Data Analysis*, **99**, 38–50.

Levy, R., and Mislevy, R. J. (2016), *Bayesian Psychometric Modeling*, CRC Press.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2013), *The BUGS Book: A Practical Introduction to Bayesian Analysis*, CRC Press.

Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, Wiley.

Plummer, M. (2003), "JAGS Version 3.4.0 user manual," https://sourceforge.net/
projects/mcmc-jags/files/Manuals/4.x/.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," R News, **6**, 7–11.

R Core Team (2017), "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Su, Y.-S., and Yajima, M. (2015), "R2jags: Using R to run JAGS," R package version 0.5-6. http://CRAN.R-project.org/package=R2jags.

Therneau, T. (2015), "A package for survival analysis in S version 2.38," http://CRAN.
R-project.org/package=survival.

Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edition, Springer.

Wikipedia (2016), https://en.wikipedia.org/wiki/List_of_popes.

WorldLifeExpectancy (2016),
http://www.worldlifeexpectancy.com/your-life-expectancy-by-age.

Appendix 1: BUGS/JAGS Code

First Approach: treat the unknown survival time as a model parameter

```
model{
 for(i in 1:n){
#
# Likelihood part
#
   censored[i] ~ dinterval(survival[i], censoring_limits[i])
     # To deal with censoring:
     # censored equals 1 for Benedict XVI as of 25/12/2016,
     # 0 for the other popes.
     # survival equals NA for Benedict XVI,
     # and the survival times of the other popes.
     # censoring_limits equals 11.7 for Benedict XVI,
     # and values (e.g. 32) greater than or equal to
     # the survival times of the other popes.
   survival[i] ~ dweib(r, mu[i]) # Basic Weibull assumption
   mu[i] <- exp(beta[i])  # Defining beta as log(mu)
   beta[i] <- beta_0 + beta_1*x_1[i] + beta_2*x_2[i]
     # beta = log(mu) is a linear function of the covariates
     }
#
#####################################################
#
# Priors
#
beta_0 ~ dnorm(0.0, 1.0E-4) # Prior on beta_0 is normal with low precision
beta_1 ~ dnorm(0.0, 1.0E-4) # Prior on beta_1 is normal with low precision
beta_2 ~ dnorm(0.0, 1.0E-4) # Prior on beta_2 is normal with low precision
#
r ~ dexp(0.001) # Prior on r
#
#####################################################
#
# Define the alphas
#
alpha_0 <- - beta_0 / r
alpha_1 <- - beta_1 / r
alpha_2 <- - beta_2 / r
#
# Percentage increases
#
percentage_increase_age <- 100*(exp(alpha_1) - 1)
percentage_increase_year <- 100*(exp(alpha_2) - 1)
#
#####################################################
```

```
#
# Posterior median at various covariate values
#
beta_med <- beta_0 + beta_1*x_1_new + beta_2*x_2_new
     # New values need to be supplied
t_median <- pow(log(2) * exp(-beta_med), 1 / r)
#
#####################################################
#
# Predictive distribution of age at the new values
#
beta_Francis <- beta_0 + beta_1*age_Francis + beta_2*year_Francis
     # Values of age_Francis and year_Francis need to be provided
mu_Francis <- exp(beta_Francis)
survival_Francis~ dweib(r, mu_Francis)T(present_length, upper_length)
     # Take into account the current pontificate length
     # Also specify a sensible upper bound
age_Francis_predictive <- age_at_election + survival_Francis
     # Work also with age
}
```

Second Approach: specify the contribution to the likelihood function from the censored observation

```
model{
#
#####################################################
#
# Likelihood contribution from censored survival time
#
# z_censored takes the value 0
# z_censored is assumed to follow a Poisson distribution
# with parameter phi_censored
#
z_censored ~ dpois(phi_censored)
#
# Required form of phi_censored, in which
# t_censored takes the value 11.7, censored survival time for Benedict XVI
#
phi_censored <- mu_censored * pow(t_censored, r)
#
# Usual form of mu for mu_censored
# x_1_censored and x_2_censored
# are the corresponding covariate values
#
mu_censored <- exp(beta_censored)
beta_censored <- beta_0 + beta_1*x_1_censored + beta_2*x_2_censored
#
#####################################################
#
# Likelihood contributions from non-censored observations, as before
#
# We loop over the 61 non-censored observations
# Previously indexed i in 2:62
#
for(j in 1:n_non_censored){ # n_non_censored is 61
 survival_non_censored[j] ~ dweib(r, mu[j])
 mu[j] <- exp(beta[j])
 beta[j] <- beta_0 + beta_1*x_1_non_censored[j] + beta_2*x_2_non_censored[j]
    }
#
#############################################
#
# Priors
#
beta_0 ~ dnorm(0.0, 1.0E-4)
beta_1 ~ dnorm(0.0, 1.0E-4)
beta_2 ~ dnorm(0.0, 1.0E-4)
#
r ~ dexp(0.001)
#
}
```

Appendix 2: R code for Frequentist Inference

```
require(survival)
#
# Censoring is handled by creating a survival object:
Surv(survival_raw, censored == 0)
# survival_raw contains all the survival times.
# censored equals 1 for Benedict XVI as of 25/12/2016,
# 0 for the other popes.
# The condition is TRUE for popes who are no longer alive or not censored
#
# Fit the model
#
s <- survreg(Surv(survival_raw, censored == 0) ~ x_1 +x_2)
summary(s) # Estimates and standard errors
confint(s) # Approximate 95% confidence intervals
#
# Predictions at the original data points
#
predict(s, se.fit = TRUE) # See ?predict.survreg
#
# Predictions at a combination of age and year values
#
age_new <- c(60, 80, 60, 80)
x_1_new <-  age_new - mean(popes$Age.Election) # Centered
year_new <- c(1750, 1750, 1950, 1950)
x_2_new <- year_new - mean(popes$Year.Elected) # Centered
#
predict(s, newdata = data.frame(x_1 = x_1_new, x_2 = x_2_new))
#
# Residuals
#
residuals(s, type = "response") # See ?residuals.survreg
```

Biographies

Julian Stander is Associate Professor (Reader) in Mathematics and Statistics in the School of Computing, Electronics and Mathematics, Plymouth University.

Luciana Dalla Valle is a Lecturer in Statistics in the School of Computing, Electronics and Mathematics, Plymouth University.

Mario Cortina–Borja is Professor of Biostatistics in the Population, Policy and Practice Programme, Great Ormond Street Institute of Child Health, University College London.