Life Contingencies: The Mathematics, Statistics, and Economics of Life Insurance

An open text authored by the Actuarial Community

Contents

Preface

Date: 23 June 2021

Book Description

Like the companion book Loss Data Analytics, this book on life contingencies will be an interactive, online, freely available text.

- The online version will contain many interactive objects (quizzes, computer demonstrations, interactive graphs, video, and the like) to promote *deeper learning*. A subset of the book will be available for *offline* reading in pdf and EPUB formats.
- Will focus on data and statistical aspects of life contingent events.
- Will emphasize cash flow fundamentals, an approach that allows users to easily adapt approaches to handle complex products.
- This modular approach emphasizing data and cash flow fundamentals has additional advantages:
  - computational aspects become practically relevant through spreadsheet (e.g., `Microsoft Excel`) and numerical (`R`) examples, and
  - an emphasis on the foundations provides an easy entry point for learners who wish an introduction to the field.

How will the text be used?

This book will be useful in actuarial curricula worldwide. Our primary **target audience** is second or third year undergraduates with little to no experience in insurance. Learners may be international; although the book will be in English, we do not expect knowledge of native idiosyncrasies that might be used in the classroom. It will cover the learning objectives of the major actuarial organizations. Thus, it will be suitable for classroom use at universities as well as for use by independent learners seeking to pass professional actuarial examinations.

A secondary audience is the actuarial practitioner (perhaps international) who wishes to retool and learn about modern approaches in the risk management of life contingent events. Thus, the text will also be useful for the continuing professional development of actuaries and other professionals in insurance and related financial risk management industries.

Why is this good for the profession?

An online text is a type of open educational resource (OER). One important benefit of an OER is that it equalizes access to knowledge, thus permitting a broader community to learn about the actuarial profession. Moreover, it has the capacity to engage viewers through active learning that deepens the learning process, producing analysts more capable of solid actuarial work.

Why is this good for students and teachers and others involved in the learning process? Cost is often cited as an important factor for students and teachers in textbook selection (see a recent post on the $400 textbook). Students will also appreciate the ability to "carry the book around" on their mobile devices.

Life Contingent Calculations

Life contingences is a quantitative discipline, enjoying the rigor and discipline of mathematics. Like any mathematical discipline, one traditionally learns about it through the development of formulaic expressions, that is, their proofs, special cases, analysis of special features, and so on. Users of this text find that we do not shy away from presenting summaries of main conclusions using formulaic expressions. Nonetheless, rather than developing insights from mathematical proofs of the primary findings, we demonstrate their impact through short illustrative examples and links to practical applications.

As with other sources that introduce life contingencies, we utilize spreadsheets extensively. In our teaching, we find that spreadsheets are useful for communication and dynamically visualizing results as they evolve over time. However, unlike other sources, we supplement this with approaches that emphasize programming; in this text, we use `R`. Programming methods such as through `R` (and Python, another good candidate) easily accommodate more complex situations that require more computing and, moreover, are built to graphically portray results in an attractive fashion. Analytics, the process of using data to make decisions, is enjoying tremendous attention from many industries; this is certainly true of in data-driven fields that use life contingent methods. By working with data and using programming methods such as `R` in the study of life contingences, users see the connections within many fields that support the actuarial science discpline. Instruction may emphasize any one of the three approaches, traditional mathematical development, spreadsheets, or a computing approach. However, this text contains all three as we believe that future generations of actuaries need to be familiar with all of these different ways to analyze, and communicate, problems that can be solved using life contingent methods.

Project Goal

The project goal is to have the actuarial community author our textbooks in a collaborative fashion. To get involved, please visit our Open Actuarial Textbooks Project Site.

Acknowledgements

We acknowledge the Society of Actuaries for permission to use problems from their examinations.

We thank Rob Hyndman, Monash University, for allowing us to use his excellent style files to produce the online version of the book.

We thank Yihui Xie and his colleagues at Rstudio for the R bookdown package that allows us to produce this book.

We also wish to acknowledge the support and sponsorship of the International Association of Black Actuaries in our joint efforts to provide actuarial educational content to all.

IABA
INTERNATIONAL ASSOCIATION OF BLACK ACTUARIES

Contributors

The project goal is to have the actuarial community author our textbooks in a collaborative fashion. The following contributors have taken a leadership role in developing *Life Contingencies*.

- **Vali Asimit**
- **Dani Bauer**
- **Adam Butt**
- **Edward (Jed) Frees**
- **Emiliano Valdez**

For our Readers

Like any book, we have a set of notations and conventions. It will probably save you time if you regularly visit our Appendix Chapter **??** to get used to ours.

Freely available, interactive textbooks represent a new venture in actuarial education and we need your input. Although a lot of effort has gone into the development, we expect hiccoughs. Please let your instructor know about opportunities for improvement, write us through our project site, or contact chapter contributors directly with suggested improvements.

Chapter 1

Introduction

Placeholder

Chapter 2

Modeling Lifetimes

The analysis of life contingent exposures such as insurer's liability when selling a life insurance contract or a pension fund's obligations when offering a new pension scheme starts with modeling individual *lifetime* and *death*. These models, in turn, have to be calibrated in the context of relevant (mortality) data.

Therefore, in this chapter, we first present different types of data that describe the lifetimes and mortality of certain population. We deliberately introduce the data in Section **??** without assuming prior knowledge of life contingent modeling, and aim to develop an intuitive understanding of some of the associated challenges.

We then introduce the conventional framework for modeling lifetimes, particularly by introducing the concept of a future lifetime variable $T_{\mathbf{x}}$ and its properties, and we then explore various approaches of specifying and estimating its distribution. Here, we discuss the traditional actuarial models for lifetimes such as analytical mortality laws and life tables, but we also introduce conditional predictive models that characterize the lifetime distribution based on a set of covariates or *features*.

2.1 Mortality data: Life Expectancies, Deaths, Counts, & Features

2.1.1 Life Expectancies

Perhaps the most relevant question to any individual when it comes to their future lifetime is: how long am I expected to live? Government agencies around the world publish vital statistics such as the *life expectancies* for their population, which are typically separated by age and gender – and potentially other attributes such as race.

The file `CDCLifeExp.csv` is provided with the supplemental information of this text and includes an excerpt of the U.S. national vital statistics that provides "Expectation of life, by age, race, [...] and sex: United States, 2017" in Table **??**.

```
library(knitr)
us_les <- read.csv("Data/CDCLifeExp.csv")
kable(us_les, caption="Life Expectancies From 2017 U.S. National Vital Statistics")
```

This excerpt provides the life expectancy for ages 0 (newborn), 20, 40, 60, and 80 observed within a population, for males and females with separate figures for the hispanic subpopulation. There are a few immediate observations.

First, females generally seem to have a longer life expectancy than males, whereas the aggregate "Total" life expectancy is in between the two figures. This is intuitive as the aggregate population is (largely) made up of male and female individuals, so that the "Total" life expectancy is a weighted average of the gender-specific life expectancies, relative to the composition of the population.

Table 2.1: Life Expectancies From 2017 U.S. National Vital Statistics

Age	Total	Male	Female	Hispanic..Total	Hispanic..Male	Hispanic..Female
0	78.6	76.1	81.1	81.8	79.1	84.3
20	59.4	57.0	61.8	62.5	59.9	64.9
40	40.7	38.7	42.6	43.5	41.2	45.5
60	23.3	21.7	24.7	25.5	23.6	27.0
80	9.2	8.4	9.8	10.5	9.4	11.1

Second, life expectancy is decreasing in age, which again is intuitive. It may be somewhat less obvious that the differences in life expectancies are less than the differences in age; subtracting the lines in Table **??** gives incremental life expectancies for 20-year age gaps.

```
kable(us_les[1:4,]-us_les[2:5,])
```

Age	Total	Male	Female	Hispanic..Total	Hispanic..Male	Hispanic..Female
-20	19.2	19.1	19.3	19.3	19.2	19.4
-20	18.7	18.3	19.2	19.0	18.7	19.4
-20	17.4	17.0	17.9	18.0	17.6	18.5
-20	14.1	13.3	14.9	15.0	14.2	15.9

Hence, while a 40-year-old male is twenty years older than a 20-year-old male, the 20-year-old's life expectancy is 18.3 years higher. The difference of 1.7 years is due to the possibility of the 20-year-old not surviving up to age 40. This effect is clearly more pronounced when comparing the 40-year-old with a 60-year-old. In other words, the average age at death increases with age.

Third, the life expectancies for the hispanic subpopulation exceed those of the total population, which suggests that other subpopulations must exhibit a lower life expectancy. There are many questions of potential reasons for this difference, although these fall more in the demographic or even sociological realm.

From an actuarial perspective, a relevant question may be how we could model the mortality data. In other words, is there a simple parametric model that may describe the progression of life expectancy across ages, at least in the context of one particular population? We will return to this question in the context of our mortality models, particularly in Section **??**.

As an early caveat to the question raised at the beginning of this section, it is not necessarily accurate to take these figures as estimates of a given individual's future lifetime or even its expectation. This is the case since the *life expectancy* is usually generated based on recent mortality *experience* rather than *forecasts*. As we will discuss in Section **??**, this is the difference between the so-called *period* and *cohort* life expectancies.

2.1.2 Population Mortality Counts

We now bring into consideration *mortality experience* for populations that had been observed over time, which is available at the Human Mortality Database (HMD) for a wide range of countries. The available data include *Exposures* by age, sex, and calender year period, i.e. how many people people of a given age and sex lived in the country's population during a given period of time, and corresponding *Deaths*, i.e. how many of these individuals had died.

In the supplemental information to this text, we provide exposures and deaths for the U.S. population, downloaded from the HMD as `HMD_Expo.csv` and `HMD_Deaths.csv`. We use the data over five year intervals starting at 1935 until 2015. Let us take a look at the exposures:

```
us_exp <- read.csv("Data/HMD_Expo.csv")
us_exp$Year_start <- as.character(us_exp$Year_start)
us_exp$Year_end <- as.character(us_exp$Year_end)
kable(head(us_exp), align = "cccrrr", digits = 2, format.args = list(big.mark = ","))
```

```
us_exp$Year_start <- as.numeric(us_exp$Year_start)
us_exp$Year_end <- as.numeric(us_exp$Year_end)
```

Year_start	Year_end	Age	Female	Male	Total
1935	1939	0	4,869,267	5,057,569	9,926,836
1935	1939	1	4,802,597	4,936,238	9,738,835
1935	1939	2	5,119,574	5,244,634	10,364,208
1935	1939	3	5,159,494	5,287,402	10,446,896
1935	1939	4	5,189,350	5,307,754	10,497,104
1935	1939	5	5,359,159	5,531,967	10,891,126

and deaths:

```
us_deaths <- read.csv("Data/HMD_Deaths.csv")
us_deaths$Year_start <- as.character(us_deaths$Year_start)
us_deaths$Year_end <- as.character(us_deaths$Year_end)
kable(head(us_deaths), align = "cccrrr", digits = 2, format.args = list(big.mark = ","))
us_deaths$Year_start <- as.numeric(us_deaths$Year_start)
us_deaths$Year_end <- as.numeric(us_deaths$Year_end)
```

Year_start	Year_end	Age	Female	Male	Total
1935	1939	0	253,145.89	335,492.00	588,637.89
1935	1939	1	36,010.02	42,169.20	78,179.22
1935	1939	2	17,718.83	21,208.22	38,927.05
1935	1939	3	12,450.36	14,852.17	27,302.53
1935	1939	4	10,154.85	11,771.25	21,926.10
1935	1939	5	8,678.43	10,339.57	19,018.00

Let us be more specific and plot the exposures and deaths for a 70-year-old U.S. females over time, which is given in Figure **??**:

```
options(digits = 9)
options(scipen = 9)
par(mfrow=c(1,2))
plot(us_exp[us_exp$Age == 70,2], us_exp[us_exp$Age == 70,4]/1000, type = "l", lwd = 5, col = "green", ma
plot(us_deaths[us_deaths$Age == 70,2], us_deaths[us_deaths$Age == 70,4]/1000, type = "l", lwd = 5, col =
```

The stark increase in exposures is due to two effects; on one hand, the U.S. population had been growing substantially since 1935, and on the other hand, individuals had been showing an enhanced life expectancy over time. As a consequence, the number of 70-year-old U.S. females had increased from less than two million to more than six million. In contrast, the number of deaths had been more steady. While with an increasing number of exposures the number of deaths have increased, the chance of dying for a 70-year-old had been decreasing. The latter becomes more evident by plotting the ratio of deaths and exposures as in Figure **??**

```
par(mfrow=c(1,1))
minlim <- min(us_deaths[us_deaths$Age == 70,4]/us_exp[us_exp$Age == 70,4])
maxlim <- max(us_deaths[us_deaths$Age == 70,4]/us_exp[us_exp$Age == 70,4])
plot(us_deaths[us_deaths$Age == 70,2],us_deaths[us_deaths$Age == 70,4]/us_exp[us_exp$Age == 70,4],type =
```

```
rm(minlim, maxlim)
```

Note that the chance of deceasing for a 70-year-old female was more than 4% around 1940, but it had decreased over time to a level lower than 2%. As we will see in Section **??**, the ratio of deaths and exposures is closely related to *death rates*.

Alternatively, we can focus on a specific period – say the most recent five years in our data, i.e. 2010-2014 – and plot the deaths over exposures across ages as in Figure **??**

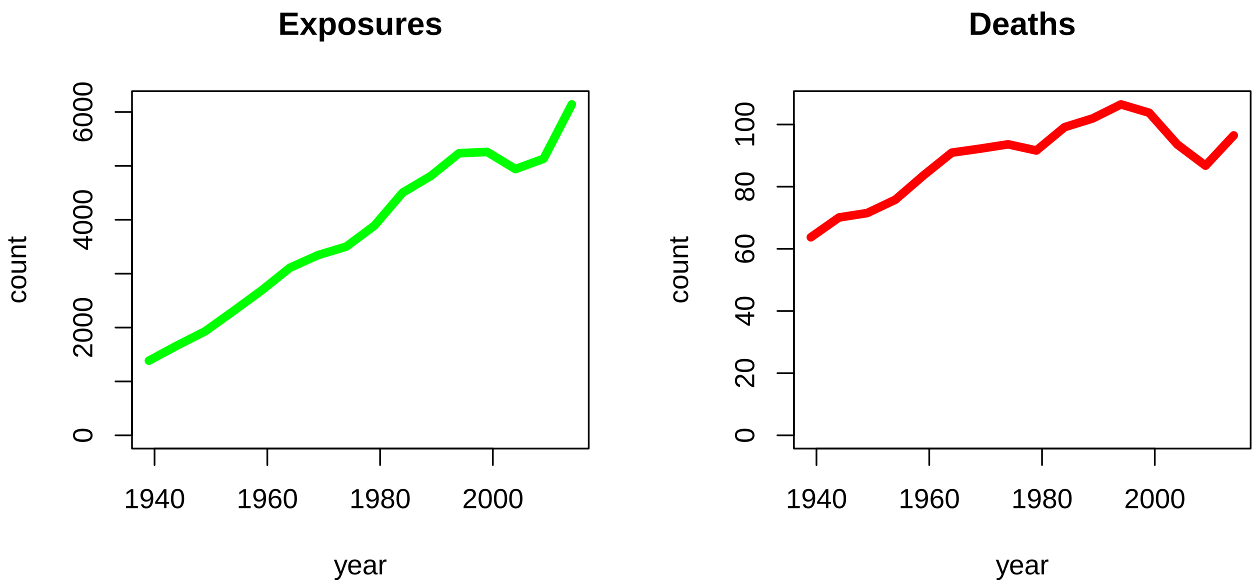

Figure 2.1: Exposures and Deaths (in thousands) for Females age 70 from the HMD U.S.

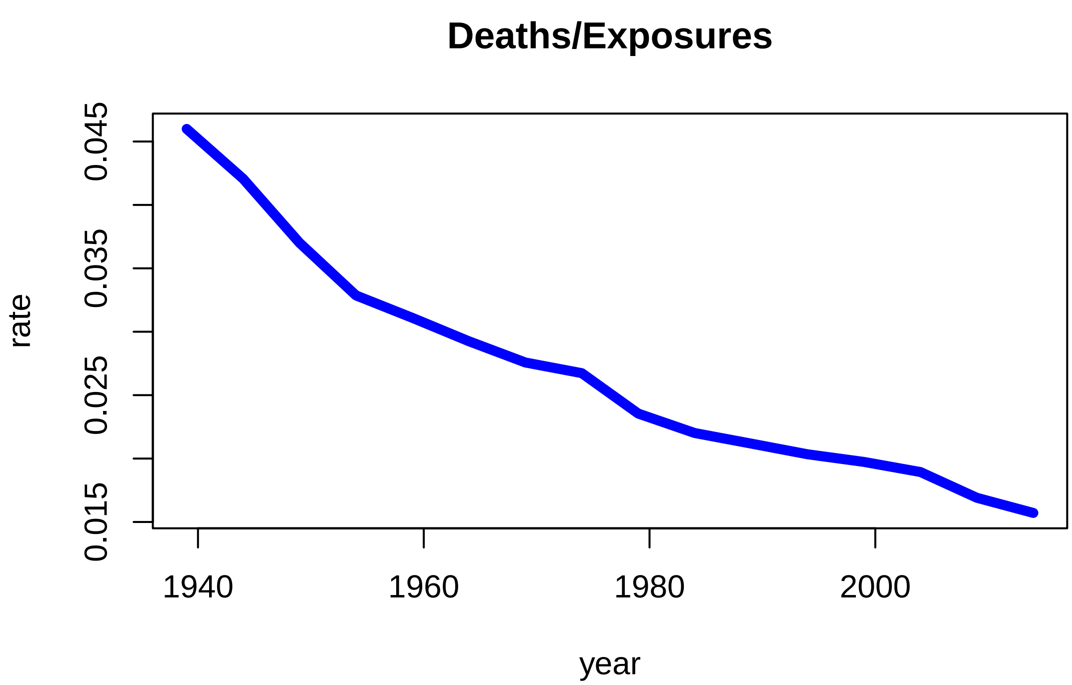

Figure 2.2: Death rates for Females age 70 from the HMD U.S.

```
us_mx_fem_2014 <- us_deaths[us_deaths$Year_start == 2010,4]/us_exp[us_exp$Year_start == 2010,4]
us_mx_mal_2014 <- us_deaths[us_deaths$Year_start == 2010,5]/us_exp[us_exp$Year_start == 2010,5]
plot(us_mx_fem_2014,type = "l", lwd = 5, col = "blue", main="Deaths/Exposures", xlab = "year", ylab = "
```

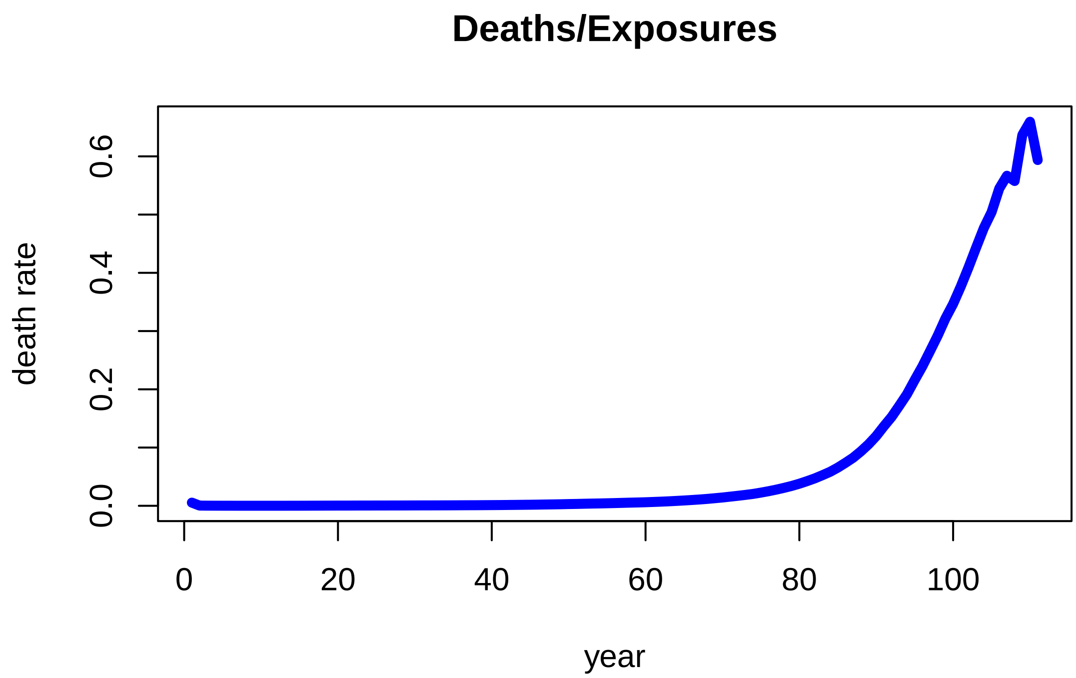

Figure 2.3: Death rates for Females across ages for 2010-2014 from the HMD U.S.

Not very surprisingly, the death rate increases fairly regularly with respect to age and its growth looks exponential. This observation is the foundation for the most famous and common analytic mortality model that is detailed in Section **??** known as the *Gompertz law*. It should be noted that the death rates at higher ages, specifically from 106 to 110, are showing a larger variability that could be explained by the fact very small sub-populations of high ages are observed and the estimates are more uncertain than those at lower ages.

We note some precaution when using the Human Mortality Database (HMD). It is intended to provide a wide range of interested parties with detailed and up-to-date mortality and population data covering about 41 countries. There are some countries that may exhibit different mortality patterns from these 41 countries, and these include for example, some of the most populous countries such as China, India, and Indonesia, and several countries in the African continents such as South Africa, Nigeria, and Egypt.

2.1.3 Individual Mortality Data

While mortality *counts* are a common way to organize mortality data across a large population, which had been our emphasis in the previous section, the individual survival time data provide valuable pieces of information. Indeed, the data available to an insurance company typically consists of records of individuals. The company records individual details for each insured person from when they purchased the contract; these include personal characteristics such as sex, age, date-of-birth, etc., but also medical records and other *underwriting information*. In addition, the company knows whether or not the person had died at the current time—and if so, the time of death.

This latter aspect is common for *survival* or *event history data*. For each observation i, a number of covariates or *features* x_i, as well as the *event time* T_i *if* the event has already happened. Otherwise, we only know that by the current *cutoff* time, the event has not happened yet. In the context of *survival analysis*, which is the branch of statistics that deals with *survival* or *event data*, this is known as *(right) censoring*, i.e. the data are (right-)censored.

Data protection policies are omnipresent and insurance companies are not any different with respect to policyholder data. On one hand, there are regulatory data protection that disallow disclosure of personal identifiable pieces of information. On the other hand, data are a key resource to a life insurer and sharing this valuable piece of information creates a competition disadvantage by revealing important information to the company's competitive position. Therefore, rather than relying on real survival data, we consider a synthetic dataset consisting of a hypothetical portfolio of policyholders.

In the supplemental information to this text, we provide survival information for a hypothetical insurance company in `SyntheticInsurerData.csv`. The company has sold whole life insurance policies for since 1955. Policyholders have to go through an underwriting examination, and in addition to policyholders' age, sex (0 for female, 1 for male), smoking status (0 for non-smoker, 1 for smoker), and the month of sale, the company records the applicants body-mass index (BMI) and the systolic blood pressure at the time of underwriting. Finally, for those policyholders with a claim, i.e., for the policyholders that have died, the company records the time of death (relative to the month of underwriting). The data is organized in the order of sales, so that the oldest entries are at the top of the data and the newest entries are at the bottom:

```
library(data.table)
ins_data <- fread("Data/SyntheticInsurerData.csv",data.table=FALSE)
kable(head(ins_data), align = "ccccrrrr")
#displaying the very end of the table similar to the very top of the table requires a few more steps
tmp_tail_ins_data <- tail(ins_data); rownames(tmp_tail_ins_data) <- NULL
kable(tmp_tail_ins_data, align = "ccccrrrr")
```

Month_of_Sale	Age	Sex	Smoking	BMI	BloodPressure	Claim	Time_of_death
1	27	0	0	25.8	117	YES	55.6332063
1	51	1	0	17.6	109	YES	18.5339297
1	59	1	0	22.5	132	YES	15.8815835
1	37	1	0	22.9	109	YES	57.4026087
1	62	0	0	30.9	147	YES	27.8345380
1	31	0	0	17.3	91	YES	64.0640696

Month_of_Sale	Age	Sex	Smoking	BMI	BloodPressure	Claim	Time_of_death
780	43	0	0	17.1000000	110	NO	NA
780	57	1	1	19.3000000	118	NO	NA
780	40	0	0	20.1000000	117	NO	NA
780	27	1	0	20.6000000	90	NO	NA
780	55	1	1	20.1000000	118	NO	NA
780	23	1	0	19.0301596	82	NO	NA

Evidently, most of the policies sold in the first months—back in 1955—have matured, and most recently underwritten policyholders are still alive. Let us further investigate the mortality dataframe and we start by summarizing the attributes of our data:

```
suppressWarnings(library(psych))
tmp_describe_ins_data<-psych::describe(ins_data)
kable(tmp_describe_ins_data, digits = 2, format.args = list(big.mark = ","))
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
Month_of_Sale	1	160,781	394.58	216.72	382.00	393.08	277.25	1.0	780.0	779.0	0.08	-1.17
Age	2	160,781	39.99	11.61	39.00	39.65	11.86	19.0	65.0	46.0	0.22	-0.73
Sex	3	160,781	0.70	0.46	1.00	0.75	0.00	0.0	1.0	1.0	-0.87	-1.25
Smoking	4	160,781	0.30	0.46	0.00	0.25	0.00	0.0	1.0	1.0	0.88	-1.22
BMI	5	160,781	22.79	4.55	21.70	22.20	3.85	16.1	69.6	53.5	1.46	3.26
BloodPressure	6	160,781	114.74	15.75	114.00	114.36	16.31	57.0	208.0	151.0	0.26	0.09
Claim*	7	160,781	1.37	0.48	1.00	1.34	0.00	1.0	2.0	1.0	0.54	-1.71
Time_of_death	8	59,382	28.27	13.69	28.36	28.24	15.21	0.0	64.2	64.2	0.02	-0.73

In summary, there are 160,781 insureds, out of which 57,529, i.e. 35.78%, have died; the average age at purchase is about 40, and our portfolio has a high percentage of men (70%) and non-smokers (70%). The average BMI is 22.8 and the average blood pressure is 114.8, which are close to (but somewhat lower than) U.S. national averages; for details, see e.g. the Withings Health Observatory that provides real-time information for US Americans.

To illustrate the sales history, we plot the monthly sales of the company, which is given as Figure **??**.

```
monthly_sales <- rep(0,780)
for (i in 1:780){
  monthly_sales[i] <- sum(ins_data$Month_of_Sale == i)
}
plot(monthly_sales,type = "l", lwd = 2, col = "green", main="Monthly Sales", xlab = "month", ylab = "Sa
abline(h = mean(monthly_sales), col = "red", lty = 2)
```

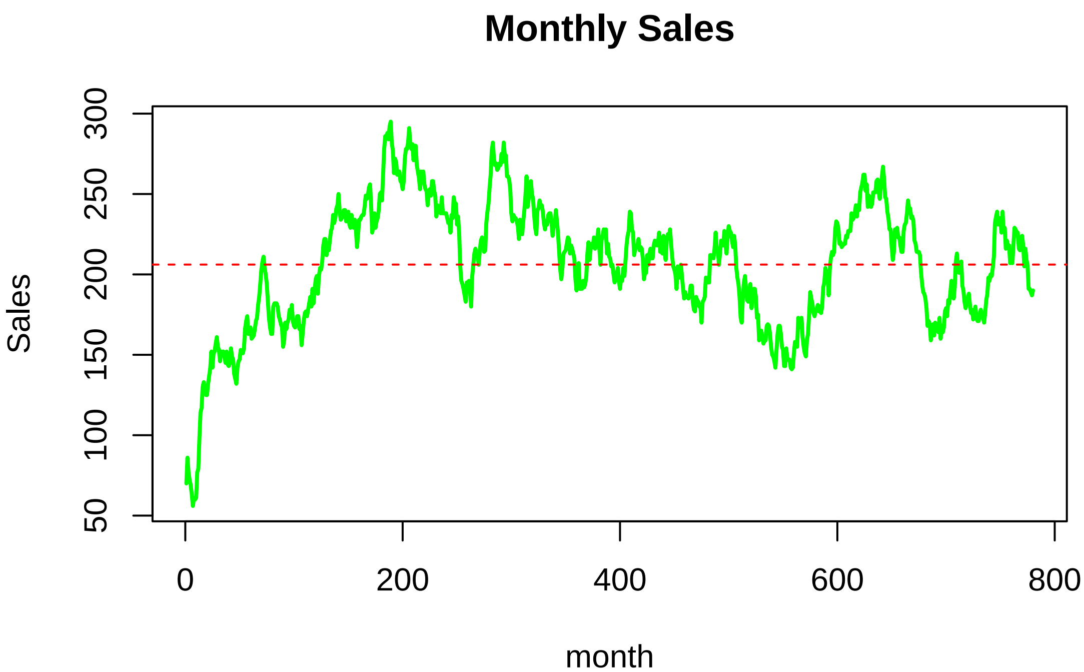

Figure 2.4: Monthy sales for the synthetic mortality data

In summary, the company sells about 200 insurance contracts each month, although there is some variation over time; while sales initially increased, the company experience some ebbs and flows, potentially due to marketing efforts and/or the effectiveness or attractiveness of the products.

We now investigate the traits of the insureds for which various histograms are given as Figure **??**.

```
par(mfrow=c(1,3))
hist(ins_data$Age,main="Insured Age", xlab="age", border="red", col="green")
hist(ins_data$BMI,main="Body Mass Index", xlab="bmi", border="red", col="green")
hist(ins_data$BloodPressure,main="Systolic Blood Pressure", xlab="bp", border="red", col="green")
```

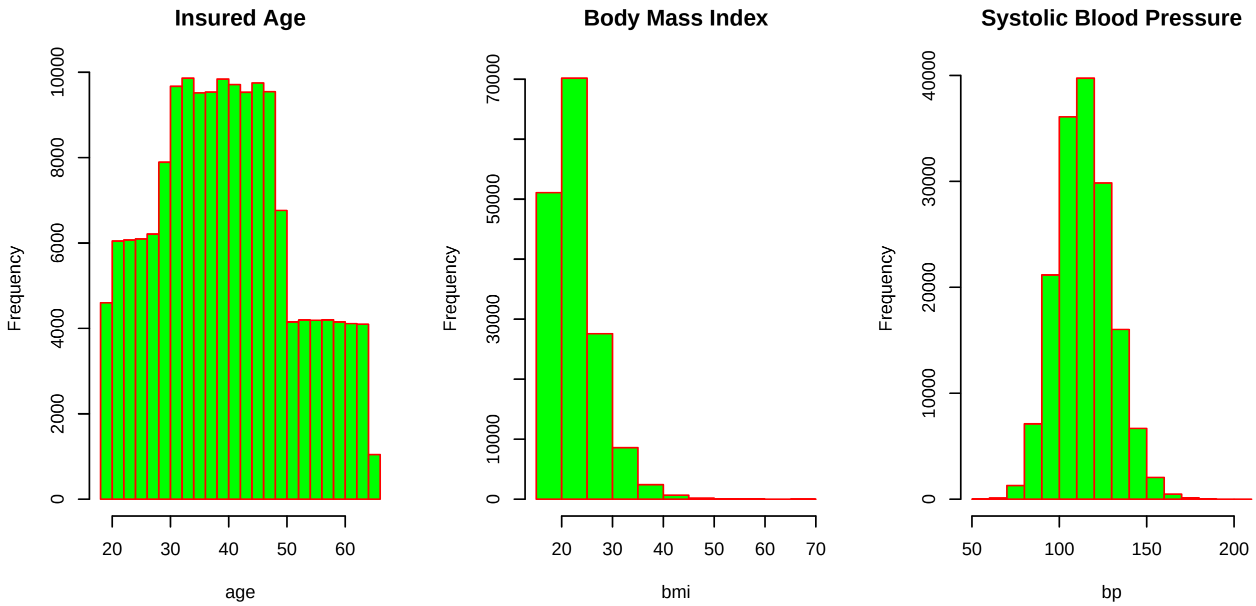

Figure 2.5: Histograms for the synthetic mortality data

Most individuals are in between thirty and fifty years of age when purchasing the coverage, but some sales to younger and some older individuals are observed. The *Body Mass Index (BMI)* is concentrated between 20 and 30, although there are some outliers with relatively large values. The systolic blood pressure is roughly bell shaped, with most applicants exhibiting blood pressure measurements at normal levels (below 120) or slightly elevated levels (120-130). We could further visualize the relationship between these three attributes by looking at the correlation correlation matrix corresponding to these three characteristics, which is given as Figure **??**.

```
suppressWarnings(library(corrplot))
tmp_corr_ins_data <- cor(ins_data[,c(2,5,6)])
colnames(tmp_corr_ins_data) <- c("Age", "BMI", "BP")
rownames(tmp_corr_ins_data) <- c("Age", "BMI", "BP")
corrplot(tmp_corr_ins_data, method="circle", order="hclust", addCoef.col = "red", tl.col="black",tl.srt=
```

Clearly, high blood pressure and elevated BMI are positively associated, and an elevated blood pressure is more common for elderly insureds, which are common observations in many populations.

One of our main focus is the realized lifetime, which could only be observed for those individuals where a claim had been paid. We could thus plot the distribution of the age at death, which is illustrated in Figure **??**.

```
hist(ins_data$Age+ins_data$Time_of_death,main="Age at Death", xlab="age", border="red", col="green")
```

Not surprisingly, the majority of deceased policyholders have died at higher ages, since the bulk of deaths are concentrated between seventy and ninety years old. There are a few individuals that died relatively young, and similarly a few that were close to achieving the centenarian status, though it should be noted that some policyholders at higher ages may still be alive. Indeed, the ages of five oldest individuals that are still alive are

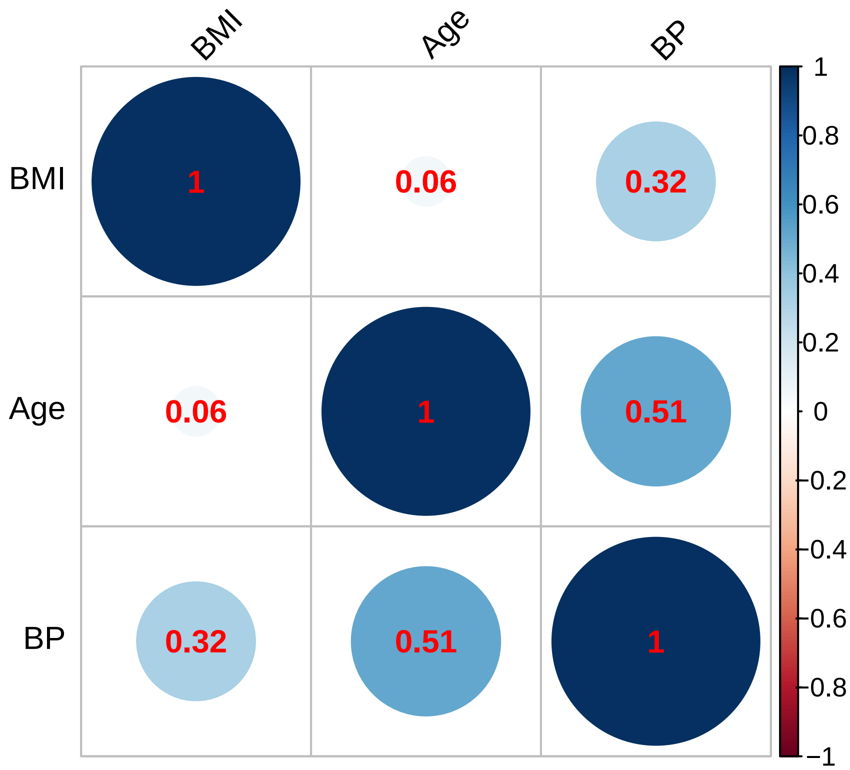

Figure 2.6: Correlation matrix for the synthetic mortality data corresponding to age, BMI and BP (blood pressure)

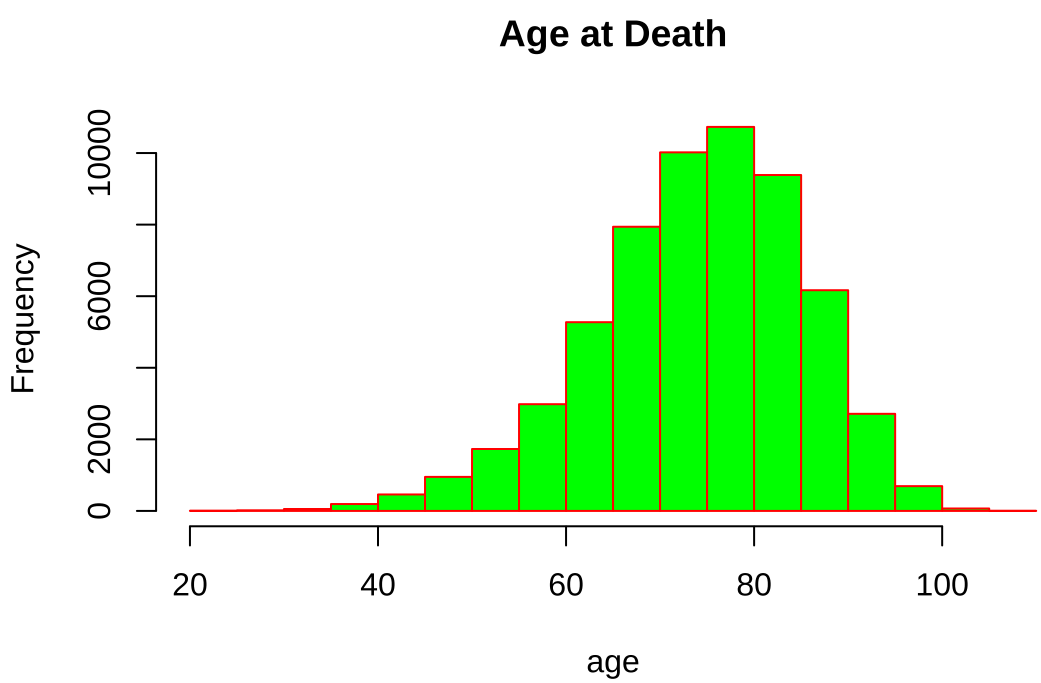

Figure 2.7: Age at Death for the synthetic mortality data

```
round(tail(sort(ins_data[ins_data$Claim == "NO",2]+ (780 - ins_data[ins_data$Claim == "NO",1])/12),5), 
```

```
[1] 102.50 102.67 103.50 104.17 105.58
```

and thus, it is clear that we observe not many centenarians.

It is intuitive that the number of deaths are associated with sales: The maximal number of deaths is the number of individuals that purchased insurance. This is evident when plotting the number of deaths by the year of sale, which is displayed in Figure **??**, particularly over the early years.

```
annual_death <- rep(0,65)
for (i in 1:65){
  for (j in 1:12){
    annual_death[i] <- annual_death[i] + sum(ins_data[ins_data$Month_of_Sale == (i-1)*12 +j,7] == "YES"
  }
}
plot(annual_death,type = "l", lwd = 4, col = "blue", main="Annual Deaths", xlab = "Year", ylab = "Death
```

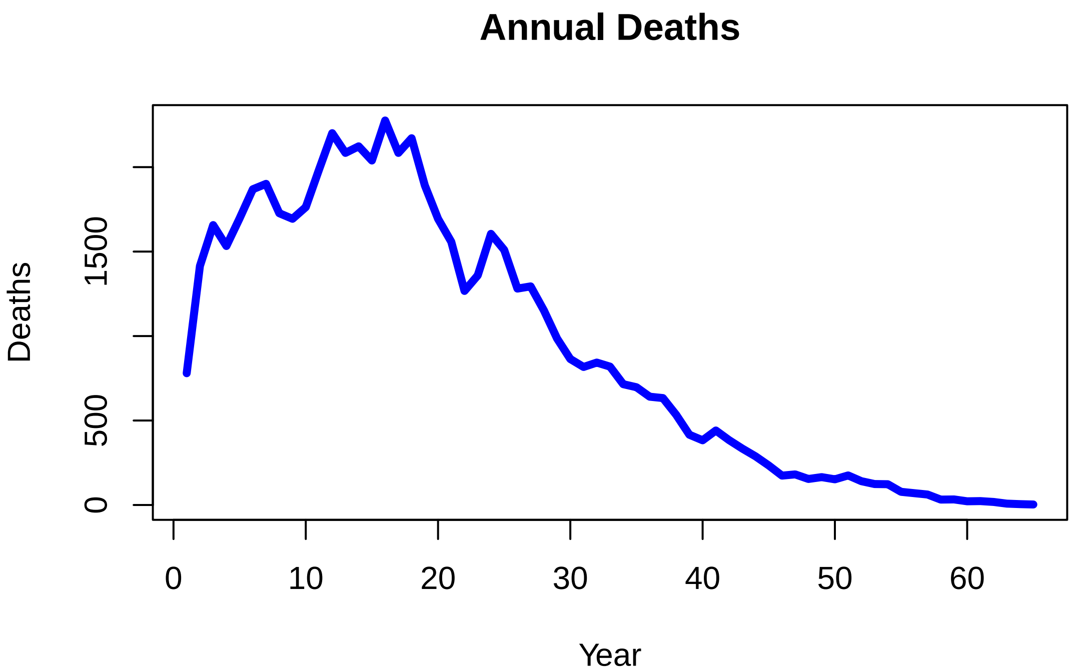

Figure 2.8: Annual deaths for the synthetic mortality data

We could investigate the relationship of the time of death to policyholder characteristics by creating a correlation plot amongst those that had already died. The correlation matrix appears as Figure **??**.

```
#suppressWarnings(library(corrplot))
tmp_dead_corr_ins_data <- cor(ins_data[ins_data$Claim == "YES",c(2,5,6,8)])
colnames(tmp_dead_corr_ins_data) <- c("Age", "BMI", "BP", "TD")
rownames(tmp_dead_corr_ins_data) <- c("Age", "BMI", "BP", "TD")
corrplot(tmp_dead_corr_ins_data, method="circle", order="hclust", addCoef.col = "red", tl.col="black", 
```

Hence, the time of death is (strongly) negatively associated with age, which is not surprising as elderly individuals are more likely to die. Similar behaviour is also observed for the BMI and blood pressure attributes that are negatively associated with the the time of death, but we should clarify that the BMI and blood pressure measurements are observed at policy inception though these negative associations explain why such attributes are good predictors of the insured's health. The pairwise correlation between age, BMI

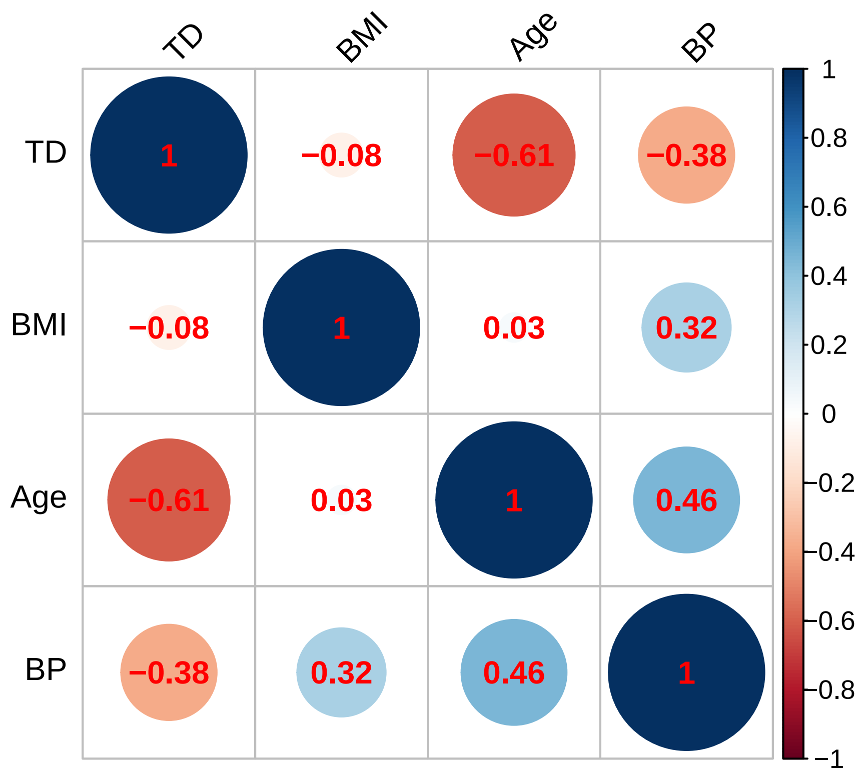

Figure 2.9: Correlation matrix for the synthetic mortality data corresponding to age, BMI, BP (blood pressure) and TD (time of death)

and blood pressure are in line with our findings from Figure **??** that includes all insureds irrespective of their life status.

Even this simple analysis could help in understanding why the life office should evaluate the individual risk per policy by taking into consideration the available information, and more importantly, to have a fair evaluation of the covariates with strong influence over the policy final payout.

2.2 Modeling Death

The previous section has introduced various examples of mortality data. Clearly, the most relevant question to actuaries is how to use the available piece of information in order to devise lifetime models that could become the basis for analyzing the life contingent exposures, which we require a theoretical underpinning. This section provides with the minimal theoretical foundation related to the conventional framework for life contingent modeling. We commence by introducing the most important concept of a lifetime random variable and the key actuarial quantities that allow for a formal interpretation of the data previously presented.

2.2.1 Lifetime random variable and its distribution

It should be first recognized that we have focused on understanding the lifetime uncertainty for a given individual or (sub)population, where the life status has been assumed to be dichotomous, i.e. *alive* and *dead* are the only life status under observation. Clearly, this simplified assumption helps us to create a parsimonious presentation that is fit for its purpose at this very moment, though one could consider *multi-state models* – with multiple life status such as alive, death, temporary disability, permanent disability and any other long term health condition, etc. – or bespoke models that relate to specific life insurance products where the modeler could consider non-health related factors such as early termination of a contract that