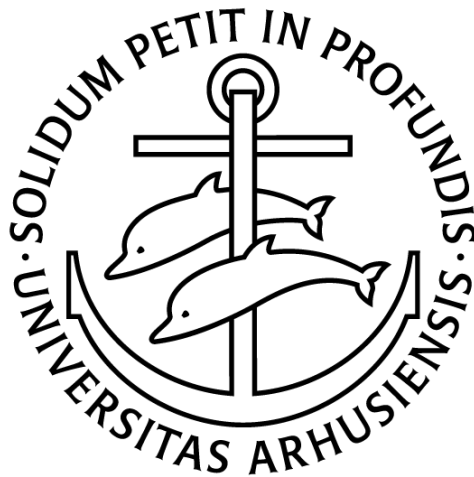


SURVIVAL ANALYSIS WITH SAS

HANDIN 1

Andreas Kracht Frandsen*

201506176



* Faculty of Mathematics, Aarhus University, andreas.kracht.frandsen@post.au.dk.

Contents

Contents	i
List of Figures	ii
List of Tables	iii
Preface	iv
1 Exercise 13	1
1.1 SAS Code	1
1.1.1 The Houses Data Set	1
1.1.2 Import of the Houses Data Set	1
1.1.3 Model Fitting	2
1.1.4 Creation of Plots	2
1.2 Plots	3
1.2.1 The First Figure	3
1.2.2 The Second Figure	5
2 Exercise 16	8
2.1 Models	8
2.1.1 Additive Model – base	8
2.1.2 Additive Model – drug	8
2.1.3 Additive Model – base+drug	9
2.1.4 Additive Model – base+drug (gamma)	9
2.1.5 Additive Model – base+drug (log link)	9
2.1.6 Interaction Model – base+drug+base*drug	9
3 Exercise 22	10
3.1 Question 1	10
3.1.1 Weibull Distributed Random Variable	10
3.1.2 Survival Function	10
3.1.3 Hazard Function	11
3.1.4 Cumulative Hazard Function	11
3.2 Question 2	11
3.3 Question 3	12
3.4 Question 4	12
A SAS Code of Exercise 16	14
A.1 Code A	14
A.1.1 Code AA	14
Bibliography	15

List of Figures

1.1	Output from the SGPlot Procedure.	4
1.2	Output from the SGPanel Procedure.	6
3.1	Density of a Exp(1) distribution.	13

List of Tables

1.1	The first 10 observations of houses.txt.	1
1.2	The first 10 observations of outdata.sas7bdat.	2
2.1	The first 10 observations of FEV.dat.	8

Preface

This document answers Exercise 13, 16 and 22 of Handin 1 in the course Survival Analysis with SAS.

To see an interactive HTML version of this document, with the possibility to edit out mistakes or make comments, go to this website afrandsen.rbind.io/bare/h1saws.html. It will be updated continuously if I find any mistakes myself through GitHub.

1 Exercise 13

In this exercise we use the SAS data set `houses` again. Execute the following code.

```
PROC GLM DATA = houses;  
CLASS new;  
MODEL price = new size;  
OUTPUT OUT = outdata PREDICTED = predvalues;  
RUN;  
QUIT;
```

Use the data set `outdata` to produce the same graphs as in Figure 5.1 and Figure 5.2 (Pedersen, 2019), noting in particular the ingenious choice of colors and symbols in the first figure. Save as pdf files.

1.1 SAS Code

1.1.1 The Houses Data Set

In this exercise we use the data set `houses` which is obtained from the course website¹. Table 1.1 shows the variables and the first 10 observations.

Table 1.1: The first 10 observations of `houses.txt`.

	taxes	beds	baths	new	price	size
1	3104	4	2	0	279.9	2048
2	1173	2	1	0	146.5	912
3	3076	4	2	0	237.7	1654
4	1608	3	2	0	200.0	2068
5	1454	3	3	0	159.9	1477
6	2997	3	2	1	499.9	3153
7	4054	3	2	0	265.5	1355
8	3002	3	2	1	289.9	2075
9	6627	5	4	0	587.0	3990
10	320	3	2	0	70.0	1160

1.1.2 Import of the Houses Data Set

First we start by importing our data set to our SAS 9.4 session.

```
DATA houses;  
  INFILE '~/Survival Analysis/Supplementary Notes/houses.txt'  
  FIRSTOBS = 2;  
  INPUT case taxes beds baths new price size;  
RUN;
```

Thus we take use of the `DATA` step. First we pass on the path to our data using the `INFILE` statement. Since our observations start in the second row, we must use the `FIRSTOBS` argument to tell

¹ Blackboard, Survival Analysis with SAS.

SAS to start reading observations from the second row, by setting it to 2. Next we tell SAS which columns and thereby variables in the houses data set it should read. We want every variable even though we aren't going to use all of them. Thus we use the INPUT statement which takes variables as arguments. (SAS Institute Inc., 2013)

1.1.3 Model Fitting

Next we run the GLM procedure as stated in the exercise.

```
PROC GLM DATA = houses;
CLASS new;
MODEL price = new size;
OUTPUT OUT = outdata PREDICTED = predvalues;
RUN;
QUIT;
```

We obtain the new data set outdata which is the same as the houses data set, but with an extra variable predvalues with the predicted prices. This variable can be spotted in Table 1.2 below.

Table 1.2: The first 10 observations of outdata.sas7bdat.

	taxes	beds	baths	new	price	size	predvalues
1	3104	4	2	0	279.9	2048	197.6
2	1173	2	1	0	146.5	912	65.7
3	3076	4	2	0	237.7	1654	151.9
4	1608	3	2	0	200.0	2068	199.9
5	1454	3	3	0	159.9	1477	131.3
6	2997	3	2	1	499.9	3153	383.7
7	4054	3	2	0	265.5	1355	117.1
8	3002	3	2	1	289.9	2075	258.5
9	6627	5	4	0	587.0	3990	423.1
10	320	3	2	0	70.0	1160	94.5

1.1.4 Creation of Plots

```
ODS PDF FILE = '~/Survival Analysis/Supplementary Notes/Graph1.pdf' NOTOC;
OPTIONS NODATE NONUMBER;

ODS GRAPHICS ON / NOBORDER
                ATTRPRIORITY=NONE
                WIDTH=9IN;

PROC SGPLOT DATA = outdata;
SCATTER X = size Y = price / GROUP = new;

STYLEATTRS DATASYMBOLS=(TRIANGLE STAR)
            DATACONTRASTCOLORS=(BLUE ORANGE);

REG X = size Y = predvalues / GROUP = new
                                NOMARKERS
                                LINEATTRS=(PATTERN=SOLID);

YAXIS LABEL = "Selling price";
XAXIS LABEL = "Size";
KEYLEGEND / NOBORDER DOWN = 2;
RUN;
```



```
ODS GRAPHICS / RESET = ALL;
ODS GRAPHICS OFF;

ODS PDF CLOSE;

ODS PDF FILE = '~/Survival Analysis/Supplementary Notes/Graph2.pdf' NOTOC;
OPTIONS NODATE NONUMBER;

ODS GRAPHICS ON / NOBORDER HEIGHT = 9IN;

PROC SGPanel DATA = outdata;
  PANELBY new / COLUMNS = 2
              ROWS = 1;
  SCATTER X = size Y = price;
  REG x = size y = predvalues / NOMARKERS;
  ROWAXIS LABEL = "Selling price";
  COLAXIS LABEL = "Size";
  KEYLEGEND / NOBORDER;
RUN;

ODS GRAPHICS / RESET = ALL;
ODS GRAPHICS OFF;

ODS PDF CLOSE;
```

1.2 Plots

1.2.1 The First Figure

Figure 1.1 below show the pdf output achieved from the first ODS PDF statement in the SAS code from Section 1.1.4.²

Figure 1.1 below show the svg output achieved from the first ODS GRAPHICS statement in the SAS code from Appendix .³

² Yes, the caption is correctly positioned. I want to explicitly show the 'A4' size from ODS PDF, and has thereby not scaled the pdf.

³ Notice the infinite scaling of svg output.

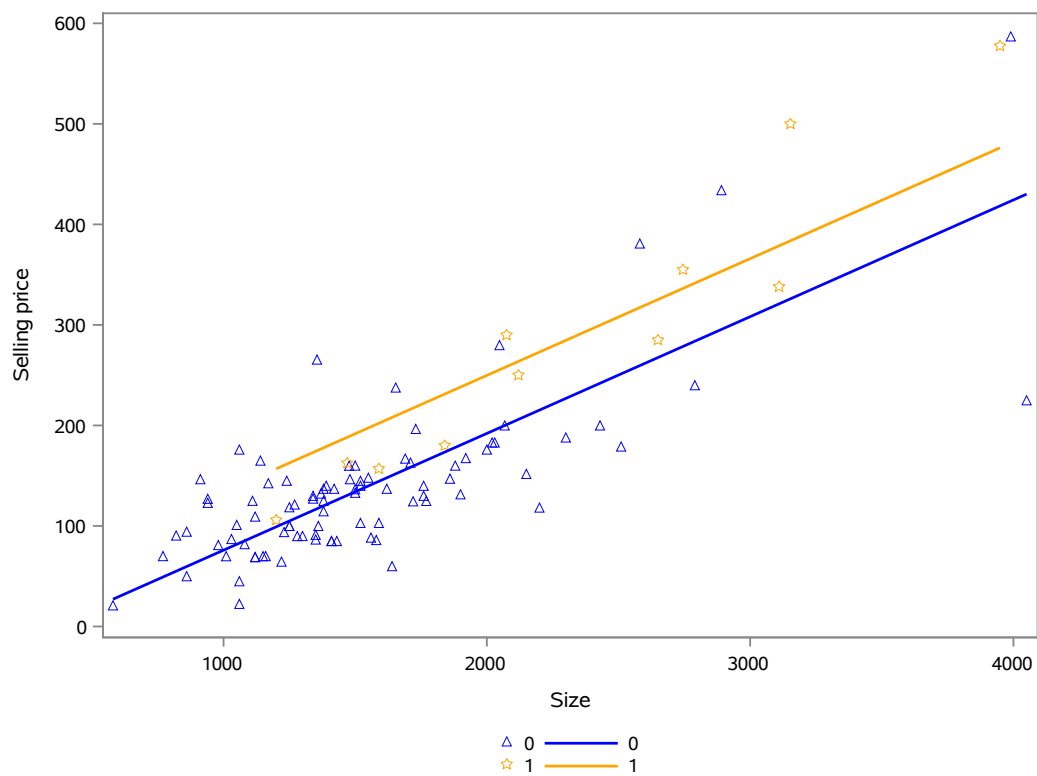


Figure 1.1: Output from the SGPlot Procedure.

1.2.2 The Second Figure

Figure 1.2 below show the pdf output achieved from the second ODS PDF statement in the SAS code from Section 1.1.4.

Figure 1.2 below show the svg output achieved from the second ODS GRAPHICS statement in the SAS code from Appendix .

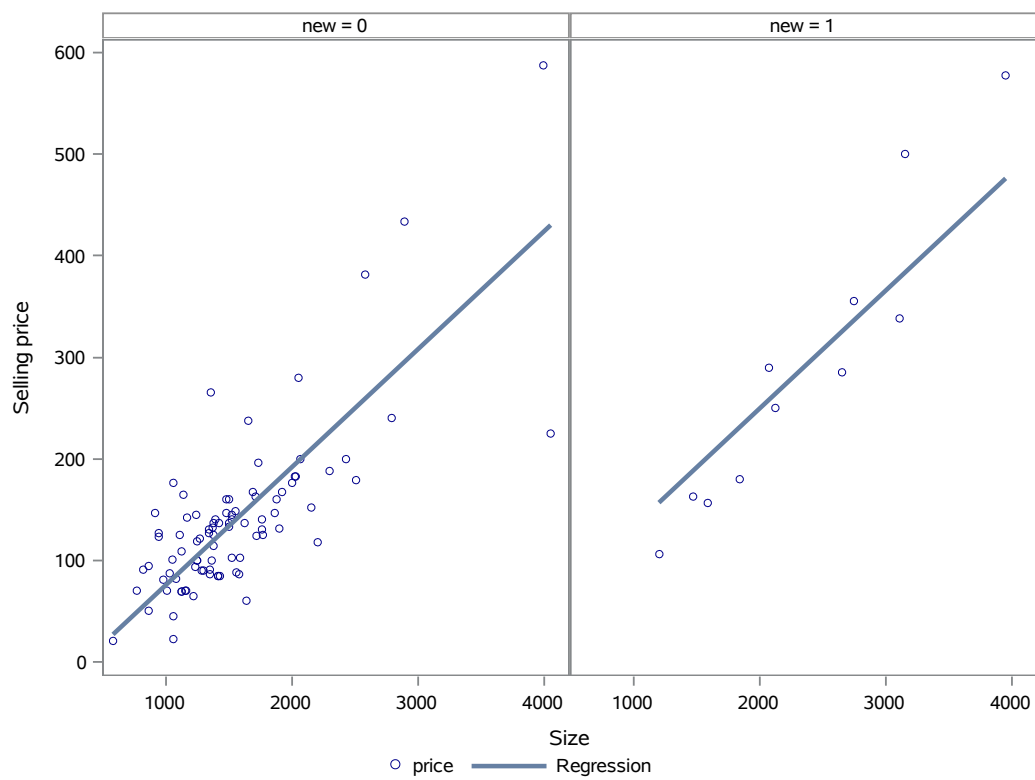


Figure 1.2: Output from the SGPanel Procedure.

We notice that Figure 1.1 and Figure 1.2 are indeed equal to Figure 5.1 and Figure 5.2 (Pedersen, 2019).

2 Exercise 16

Exercise 1.21 (Agresti, 2015) presented a study comparing forced expiratory volume after 1 hour of treatment for three drugs (a, b, and p = placebo), adjusting for a baseline measurement x_1 . Table 4.1 (Agresti, 2015) shows the results of fitting some normal GLMs (with identity link, except one with log link) and a GLM assuming a gamma response. Interpret results.

1. Do all analyses in SAS
2. Write up the models mathematically.
3. Include graphs to illustrate (lack of) fit.
4. Which models (or which model) seem to perform best?

The dataset FEV is obtained from (College of Liberal Arts and Sciences, 2000). Table 2.1 shows the variables and the first 10 observations.

Table 2.1: The first 10 observations of FEV.dat.

	base	fev1	fev2	fev3	fev4	fev5	fev6	fev7	fev8	drug
1	2.46	2.68	2.76	2.50	2.30	2.14	2.40	2.33	2.20	a
2	3.50	3.95	3.65	2.93	2.53	3.04	3.37	3.14	2.62	a
3	1.96	2.28	2.34	2.29	2.43	2.06	2.18	2.28	2.29	a
4	3.44	4.08	3.87	3.79	3.30	3.80	3.24	2.98	2.91	a
5	2.80	4.09	3.90	3.54	3.35	3.15	3.23	3.46	3.27	a
6	2.36	3.79	3.97	3.78	3.69	3.31	2.83	2.72	3.00	a
7	1.77	3.82	3.44	3.46	3.02	2.98	3.10	2.79	2.88	a
8	2.64	3.67	3.47	3.19	2.19	2.85	2.68	2.60	2.73	a
9	2.30	4.12	3.71	3.57	3.49	3.64	3.38	2.28	3.72	a
10	2.27	2.77	2.77	2.75	2.75	2.71	2.75	2.52	2.60	a

2.1 Models

We will study the following models

2.1.1 Additive Model – base

$$Y_{ij} \sim N(\alpha + \gamma b_{ij}, \sigma^2),$$

where $i = a, b, p, j = 1, \dots, 24$.

The above model is a NLM with $\beta = (\alpha, \gamma)^\top$.

2.1.2 Additive Model – drug

$$Y_{ij} \sim N(\alpha + \beta_i, \sigma^2),$$

where $i = a, b, p, j = 1, \dots, 24$.

The above model is a NLM with $\beta = (\alpha, \beta_a, \beta_b, \beta_p)^\top$.

2.1.3 Additive Model – base+drug

$$Y_{ij} \sim N\left(\alpha + \gamma b_{ij} + \beta_i, \sigma^2\right),$$

where $i = a, b, p, j = 1, \dots, 24$.

The above model is a NLM with $\beta = (\alpha, \gamma, \beta_a, \beta_b, \beta_p)^\top$.

2.1.4 Additive Model – base+drug (gamma)

$$Y_{ij} \sim \Gamma\left(\alpha + \gamma b_{ij} + \beta_i, k\right),$$

where $i = a, b, p, j = 1, \dots, 24$.

2.1.5 Additive Model – base+drug (log link)

$$Y_{ij} \sim N\left(\mu_{ij}, \sigma^2\right),$$

where $i = a, b, p, j = 1, \dots, 24$.

$$\mu_{ij} = \log(\mathbb{E}(Y_{ij})) = \alpha + \gamma b_{ij} + \beta_i.$$

The above model is a NLM with $\beta = (\alpha, \gamma, \beta_a, \beta_b, \beta_p)^\top$.

2.1.6 Interaction Model – base+drug+base*drug

$$Y_{ij} \sim N\left(\alpha + \gamma b_{ij} + \beta_i + \delta_i b_{ij}, \sigma^2\right),$$

where $i = a, b, p, j = 1, \dots, 24$.

The above model is a NLM with $\beta = (\alpha, \gamma, \beta_a, \beta_b, \beta_p, \delta_a, \delta_b, \delta_p)^\top$.

3 Exercise 22

Let X have a Weibull distribution with parameters $\alpha, \lambda > 0$, that is, X has density f as indicated in Table 2.2 in [Klein and Moeschberger \(2003\)](#).

3.1 Question 1

Show that h and S are as indicated in [Klein and Moeschberger \(2003\)](#), Table 2.2. Determine H as well.

3.1.1 Weibull Distributed Random Variable

The random variable $X \sim \text{Weibull}(\alpha, \lambda)$ with $\alpha, \lambda > 0$, has the following probability density function

$$f(x) = \alpha \lambda x^{\alpha-1} \exp(-\lambda x^\alpha) \quad x \geq 0. \quad (3.1)$$

3.1.2 Survival Function

For a continuous random variable the survival function is defined as the complement of the cumulative distribution function

$$S(x) = 1 - F(x) = P(X > x) = \int_x^\infty f(t) dt. \quad (3.2)$$

Thus for the continuous random variable X we use Equation (3.2).

We insert Equation (3.1) in Equation (3.2) and perform the calculation for $x \geq 0$.

$$\begin{aligned} S(x) &= \int_x^\infty \alpha \lambda t^{\alpha-1} \exp(-\lambda t^\alpha) dt \\ &= \alpha \lambda \int_x^\infty t^{\alpha-1} \exp(-\lambda t^\alpha) dt \\ &= \alpha \lambda \int_{x^\alpha}^\infty t^{\alpha-1} \exp(-\lambda u) \frac{1}{\alpha t^{\alpha-1}} du \\ &= \lambda \int_{x^\alpha}^\infty \exp(-\lambda u) du \\ &= [-\exp(-\lambda u)]_{x^\alpha}^\infty \\ &= (-\exp(-\lambda \cdot \infty) - (-\exp(-\lambda \cdot x^\alpha))) \\ &= (0 - (-\exp(-\lambda \cdot x^\alpha))) \\ &= \exp(-\lambda x^\alpha). \end{aligned} \quad (3.3)$$

Where we in the third and fifth equality perform integration by substitution. We see that the above function is indeed equal to the survival function in Table 2.2 ([Klein and Moeschberger, 2003](#)).

3.1.3 Hazard Function

For a continuous random variable the hazard function (rate) is defined as

$$h(x) = \frac{f(x)}{S(x)}. \quad (3.4)$$

We insert Equation (3.3) and Equation (3.1) in Equation (3.4) and perform the calculation for $x \geq 0$.

$$\begin{aligned} h(x) &= \frac{f(x)}{S(x)} \\ &= \frac{\alpha \lambda x^{\alpha-1} \exp(-\lambda x^\alpha)}{\exp(-\lambda x^\alpha)} \\ &= \alpha \lambda x^{\alpha-1}. \end{aligned} \quad (3.5)$$

We see that the above function is indeed equal to the hazard function in Table 2.2 (Klein and Moeschberger, 2003).

3.1.4 Cumulative Hazard Function

For a continuous random variable the cumulative hazard function is defined as

$$H(x) = \int_0^x h(u) du = -\ln(S(x)). \quad (3.6)$$

We insert Equation (3.3) in Equation (3.6) and perform the calculation for $x \geq 0$.

$$H(x) = -\ln(\exp(-\lambda x^\alpha)) = \lambda x^\alpha. \quad (3.7)$$

We see that the above function is indeed equal to the cumulative hazard function in (Klein and Moeschberger (2003), page 32).

Thus we showed that the functions in Table 2.2 (Klein and Moeschberger, 2003) are correct and we are done.

3.2 Question 2

Determine the distribution of X^γ for $\gamma > 0$.

We calculate the survival function of X^γ . Since the distribution of a random variable is fully described by it's survival function.

$$\begin{aligned}
S_{X^\gamma}(x) &= P(X^\gamma > x) \\
&= P(\ln(X^\gamma) > \ln(x)) \\
&= P(\gamma \ln(X) > \ln(x)) \\
&= P\left(\ln(X) > \frac{\ln(x)}{\gamma}\right) \\
&= P\left(\exp(\ln(X)) > \exp\left(\frac{\ln(x)}{\gamma}\right)\right) \\
&= P\left(X > \exp\left(\frac{\ln(x)}{\gamma}\right)\right) \\
&= S_X\left(\exp\left(\frac{\ln(x)}{\gamma}\right)\right) \\
&= \exp\left(-\lambda \left(\exp\left(\frac{\ln(x)}{\gamma}\right)\right)^\alpha\right) \\
&= \exp\left(-\lambda x^{\frac{\alpha}{\gamma}}\right).
\end{aligned} \tag{3.8}$$

Thus from Table 2.2 (Klein and Moeschberger, 2003) we have $X^\gamma \sim \text{Weibull}\left(\frac{\alpha}{\gamma}, \lambda\right)$.

3.3 Question 3

Determine the distribution of λX^α .

We calculate the survival function of λX^α . Since the distribution of a random variable is fully described by its survival function.

$$\begin{aligned}
S_{\lambda X^\alpha}(x) &= P(\lambda X^\alpha > x) \\
&= P(\ln(\lambda X^\alpha) > \ln(x)) \\
&= P(\ln(\lambda) + \alpha \ln(X) > \ln(x)) \\
&= P\left(\ln(X) > \frac{\ln(x) - \ln(\lambda)}{\alpha}\right) \\
&= P\left(\exp(\ln(X)) > \exp\left(\frac{\ln(x) - \ln(\lambda)}{\alpha}\right)\right) \\
&= P\left(X > \exp\left(\frac{\ln(x) - \ln(\lambda)}{\alpha}\right)\right) \\
&= S_X\left(\exp\left(\frac{\ln(x) - \ln(\lambda)}{\alpha}\right)\right) \\
&= \exp\left(-\lambda \left(\exp\left(\frac{\ln(x) - \ln(\lambda)}{\alpha}\right)\right)^\alpha\right) \\
&= \exp\left(-\lambda x^{\frac{\alpha}{\gamma}}\right) \\
&= \exp(-x).
\end{aligned} \tag{3.9}$$

Thus from Table 2.2 (Klein and Moeschberger, 2003) we have $\lambda X^\alpha \sim \text{Weibull}(1, 1) \stackrel{d}{=} \text{Exp}(1)$.

Figure 3.1 below show the distribution.

3.4 Question 4

Let $n \in \mathbb{N}$ and X_1, \dots, X_n be i.i.d. and have a Weibull distribution with parameters $\alpha, \lambda > 0$ as common distribution. Determine the distribution of $\min(X_1, \dots, X_n)$.

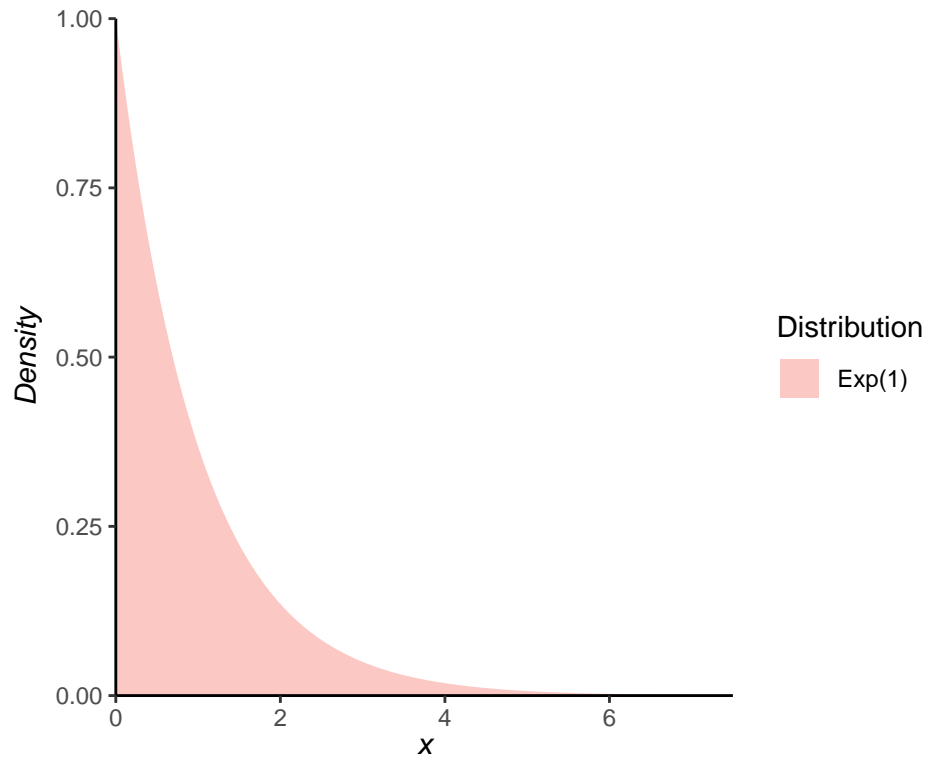


Figure 3.1: Density of a Exp(1) distribution.

The random variable X_i has the survival function

$$S_{X_i}(x) = \exp(-\lambda x^\alpha) \quad x \geq 0,$$

for $i = 1, 2, \dots, n$. As proved in Section 3.1.2, Equation (3.3).

Let $Y = \min(X_1, \dots, X_n)$, then the survival function of Y is

$$\begin{aligned}
 S_Y(y) &= P(Y > y) \\
 &= P(\min(X_1, \dots, X_n) > y) \\
 &= P(X_1 > y, X_2 > y, \dots, X_n > y) \\
 &= P(X_1 > y) \cdot P(X_2 > y) \cdots P(X_n > y) \\
 &= S_{X_1}(y) \cdot S_{X_2}(y) \cdots S_{X_n}(y) \\
 &= (\exp(-\lambda y^\alpha)) \cdot (\exp(-\lambda y^\alpha)) \cdots \exp(-\lambda y^\alpha) \\
 &= \exp(-n \cdot \lambda y^\alpha).
 \end{aligned} \tag{3.10}$$

Where we in the fourth equality take use of the independency of our random variables. Thus $Y = \min(X_1, \dots, X_n) \sim \text{Weibull}(\alpha, n\lambda)$.

A SAS Code of Exercise 16

A.1 Code A

A.1.1 Code AA

```
data fev;
  infile '~/Survival Analysis/Supplementary Notes/FEV.dat'
    firstobs=2;
  input patient base fev1 fev2 fev3 fev4 fev5 fev6 fev7 fev8 drug $1.;
run;

* AIC = 134.4;
proc genmod data=fev plots=all;
class drug;
model fev1 = base / dist=normal link=identity;
run;

* AIC = 152.4;
proc genmod data=fev plots=all;
class drug(ref='a');
model fev1 = drug / dist=normal link=identity;
run;

* AIC = 103.4;
proc genmod data=fev plots=all;
class drug(ref='a');
model fev1 = base drug / dist=normal link=identity;
run;

* AIC = 106.2;
proc genmod data=fev plots=all;
class drug(ref='a');
model fev1 = base drug / dist=gamma link=identity;
run;

* AIC = 106.8;
proc genmod data=fev plots=all;
class drug(ref='a');
model fev1 = base drug / dist=normal link=log;
run;

* AIC = 107.1;
proc genmod data=fev plots=all;
class drug(ref='a');
model fev1 = base drug base*drug / dist=normal link=identity;
run;
```

Bibliography

Agresti, A. (2015). *Foundations of linear and generalized linear models*. Wiley series in probability and statistics. John Wiley & Sons Inc, 1 edition.

College of Liberal Arts and Sciences (2000). <http://users.stat.ufl.edu/~aa/glm/data/FEV.dat>, Last accessed on 2019-10-08.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data*. Springer, 2nd edition.

Pedersen, J. (2019). *Survival Analysis (Supplementary Notes)*. Department of Mathematics, Aarhus University.

SAS Institute Inc. (2013). SAS 9.4 ® DATA Step Statements: Reference. <https://documentation.sas.com/?docsetId=lestmtsref&docsetTarget=titlepage.htm&docsetVersion=9.4&locale=en>, Last accessed on 2019-10-08.