assignment1.R

Joshua

2021-03-23

```
#load packages
library(ggplot2)
library(ggpubr)
par(mar=c(5.1, 4.1, 4.1, 2.1))
setwd("C:/Users/Joshua/Desktop/Joshua's Files/Australian National University/4th year/Semester 1/STAT303

#5.
lifetbl = read.csv("C:/Users/Joshua/Desktop/Joshua's Files/Australian National University/4th year/Semes

#6.
head(lifetbl)
```

```
##   Year Age      mx      qx   ax     lx   dx    Lx      Tx    ex
## 1 1921   0 0.05999 0.05750 0.28 100000 5750 95857 6317561 63.18
## 2 1921   1 0.01206 0.01199 0.50  94250 1130 93685 6221704 66.01
## 3 1921   2 0.00578 0.00576 0.50  93120  537 92851 6128020 65.81
## 4 1921   3 0.00289 0.00288 0.50  92583  267 92450 6035168 65.19
## 5 1921   4 0.00325 0.00325 0.50  92316  300 92166 5942719 64.37
## 6 1921   5 0.00252 0.00251 0.50  92016  231 91900 5850553 63.58
```

```
#7.
lifetbl = lifetbl[lifetbl$Year == 1943,]

#8.
class(lifetbl$Age) #we want age to be numeric
```

```
## [1] "character"
```

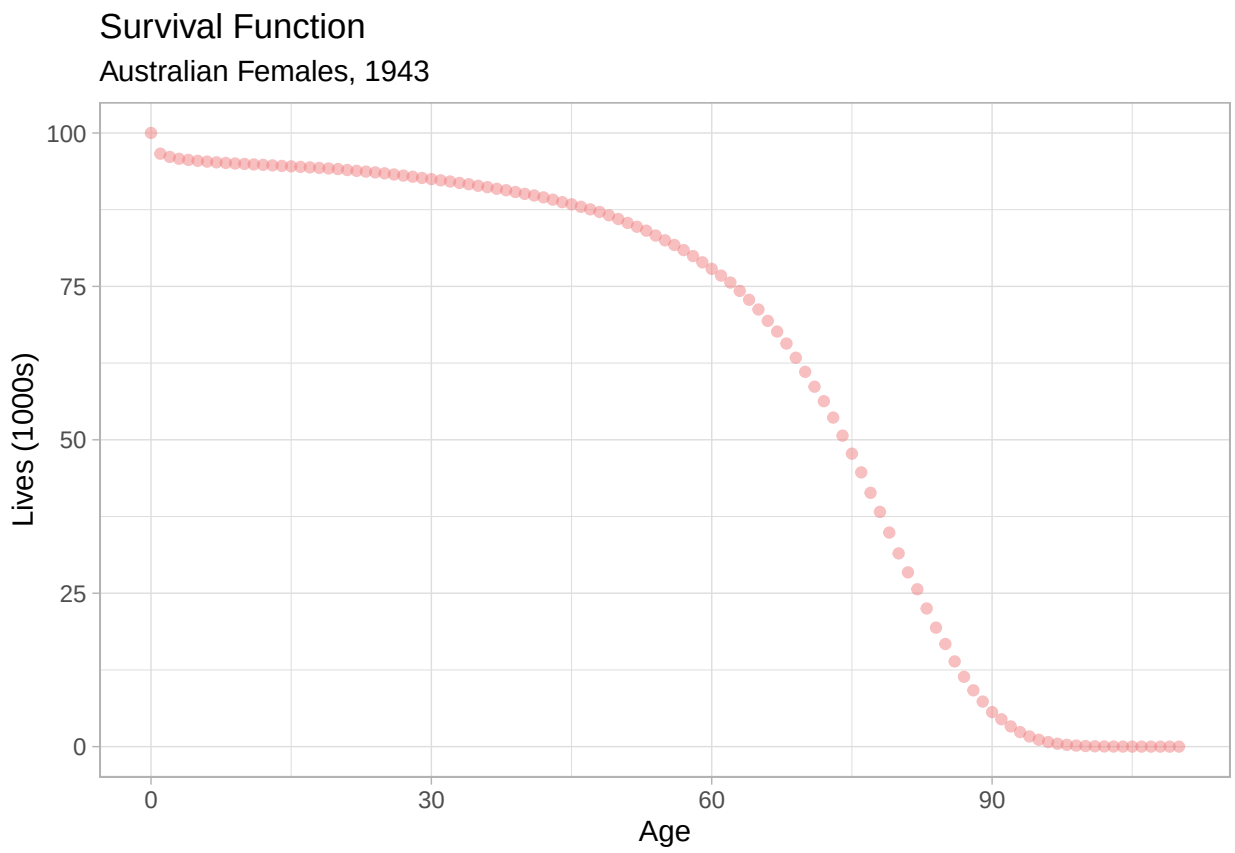
```
lifetbl$Age2 = lifetbl$Age #create new column duplicating age value
lifetbl$Age2[lifetbl$Age2 == "110+"] = 110
lifetbl$Age2 = as.numeric(lifetbl$Age2)
class((lifetbl$Age2)) #check that age is now numeric
```

```
## [1] "numeric"
```

```
#9.
attach(lifetbl)

plt = ggplot(lifetbl, aes(x = Age2, y = lx/1000))
plt = plt+geom_point(shape = 19, colour = "lightcoral", alpha = 0.5)
```

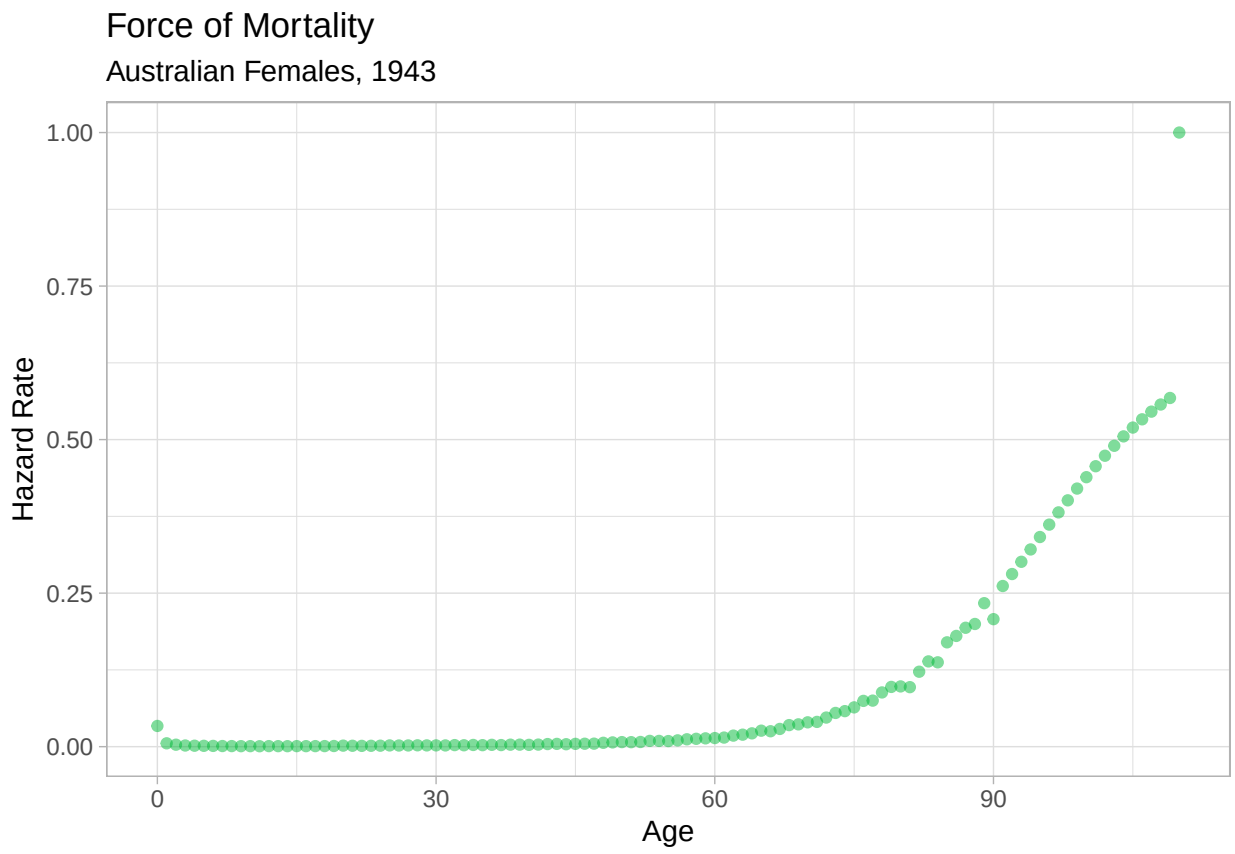
```
plt = plt+labs(x = "Age", y = "Lives (1000s)")
plt = plt+ggtitle("Survival Function", subtitle = "Australian Females, 1943")
plt = plt+theme_light()
plt
```




```
ggsave(filename = "survival.png", dpi = 1200)
```

```
## Saving 6.5 x 4.5 in image
```

```
#10. # mu approx. qx
plt = ggplot(lifetbl, aes(x = Age2, y = qx))
plt = plt+geom_point(shape = 19, colour = "#00BA38", alpha = 0.5)
plt = plt+labs(x = "Age", y = "Hazard Rate")
plt = plt+ggtitle("Force of Mortality", subtitle = "Australian Females, 1943")
plt = plt+theme_light()
plt
```




```
ggsave(filename = "hazard.png", dpi = 1200)
```

```
## Saving 6.5 x 4.5 in image
```

```
#11.
#Curtate expectations of life:
sum(lx[-1])/lx[1]
```

```
## [1] 67.75926
```

```
sum(lx[29:111])/lx[28]
```

```
## [1] 45.35909
```

```
#12. mu approx. qx
gompertz = lm(log(qx) ~ Age2)
summary(gompertz)
```

```
##
## Call:
## lm(formula = log(qx) ~ Age2)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -0.7137 -0.4279 -0.1079  0.2590  4.3501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.739154   0.123290  -62.77   <2e-16 ***
## Age2         0.065672   0.001937   33.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6539 on 109 degrees of freedom
## Multiple R-squared:  0.9134, Adjusted R-squared:  0.9126
## F-statistic:  1150 on 1 and 109 DF,  p-value: < 2.2e-16
```

```
anova(gompertz)
```

```
## Analysis of Variance Table
##
## Response: log(qx)
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Age2        1 491.49  491.49  1149.6 < 2.2e-16 ***
## Residuals 109  46.60    0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#log linear scale
plt1 = ggplot(lifetbl, aes(x = Age2, y = qx))
plt1 = plt1+ scale_y_continuous(trans = "log", labels=scales::number_format(accuracy = 0.01))
plt1 = plt1+geom_smooth(method=lm , color="#C77CFF", fill="#C77CFF", se=TRUE, size = 0.5)
plt1 = plt1+geom_point(shape = 19, colour = "#E7861B", alpha = 0.3, size = 1)
plt1 = plt1+labs(x = "Age", y = "Log Hazard Rate")
plt1 = plt1+theme_light(base_size = 10)

#transformed back to original scale #include legend or caption explaining this
plt2 = ggplot(lifetbl, aes(x = Age2, y = qx))
plt2 = plt2+labs(x = "Age", y = "Hazard Rate")
plt2 = plt2+theme_light(base_size = 10)

x0 = seq(from = 0, to = 110)

CI = predict(gompertz, newdata = data.frame(Age2 = x0), interval = "confidence") #takes in gompertz mod

plt2 = plt2+geom_ribbon(aes(ymin = exp(CI[,2]), ymax = exp(CI[,3]), fill = "95% Confidence Intervals"),
plt2 = plt2+geom_line(aes(y = exp(CI[,1]), size = "Gompertz Model"), colour = "#C77CFF")
plt2 = plt2+scale_fill_manual(values = c("#C77CFF","#00BFC4"))
plt2 = plt2+scale_size_manual(values = 0.5)
plt2 = plt2+geom_point(shape = 19, colour = "#E7861B", alpha = 0.3, size = 1)
plt2 = plt2+theme(legend.title = element_blank(), plot.title = element_text(hjust = 0.5), legend.positi

figure = ggarrange(plt1, plt2, ncol=2, nrow=1, common.legend = TRUE, legend="bottom")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

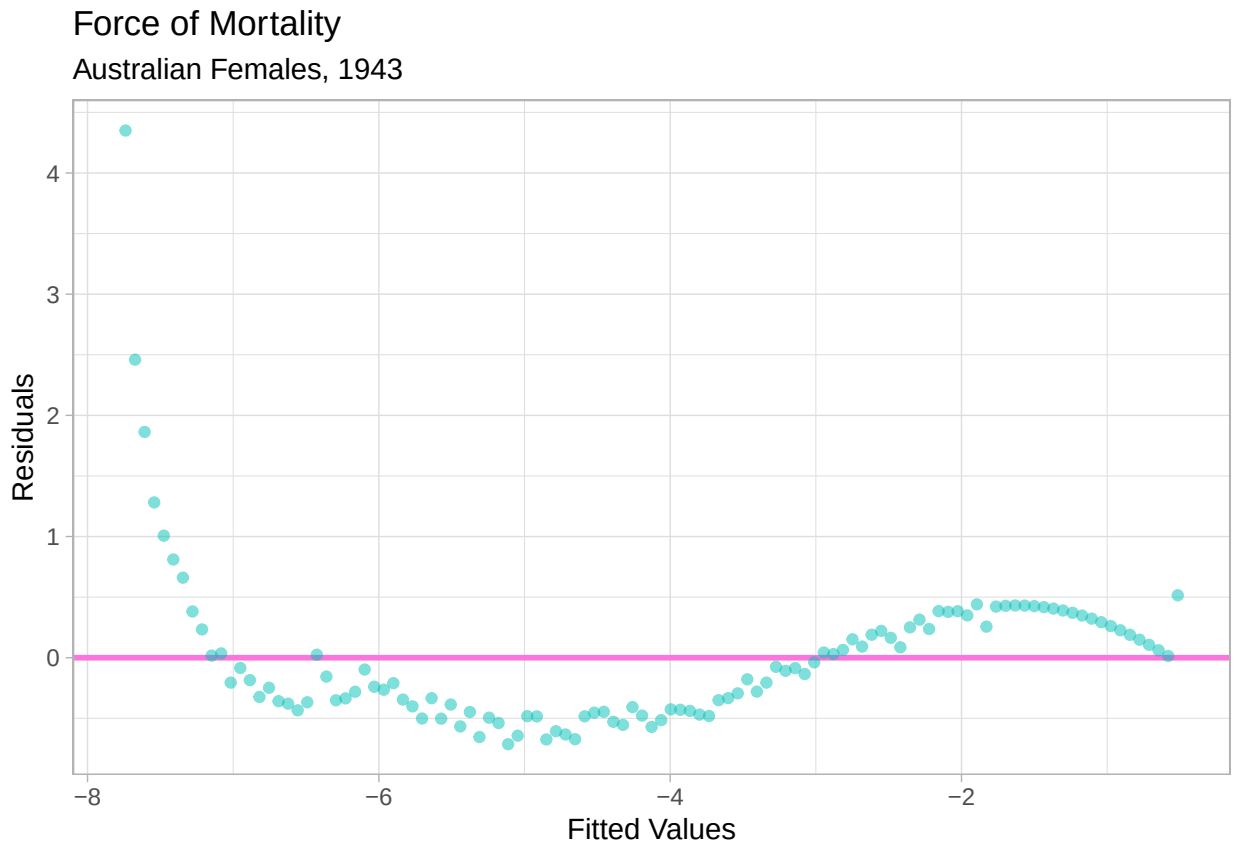
```
## `geom_smooth()` using formula 'y ~ x'
```

```
title = expression(atop("Force of Mortality", scriptstyle("Australian Females, 1943")))
annotate_figure(figure, top = text_grob(title, size = 12))
```


```
ggsave(filename = "gompertz.png", dpi = 1200)
```

```
## Saving 6.5 x 4.5 in image
```

```
#13.
plt = ggplot(gompertz, aes(x = gompertz$fitted.values, y = gompertz$residuals))
plt = plt+geom_hline(yintercept = 0, color = "#F876DD" , size = 1)
plt = plt+geom_point(shape = 19, colour = "#00C0B8", alpha = 0.5)
plt = plt+labs(x = "Fitted Values", y = "Residuals")
plt = plt+ggtitle("Force of Mortality", subtitle = "Australian Females, 1943")
plt = plt+theme_light()
plt
```




```
ggsave(filename = "residuals.png", dpi = 1200)
```

```
## Saving 6.5 x 4.5 in image
```

```
#looking at CI, there is not a 95% of true value staying within interval. cant be correct. quadratic or
```