Exercise 02

Exercise 2.1

(a)

In constructing the analytical sample for the analysis of first marriages in Italy based on cohort and education, a systematic approach was followed.

Italy was selected as the country of focus, and **first marriage**, reflecting a significant life event influenced by educational attainment, was chosen as the outcome. The sample was limited to **female** respondents to eliminate gender effects. Cohorts from **1940 to 1999**, with marriages occurring between 1955 and 2019, were included. Listwise deletion was applied to handle missing or invalid values, resulting in an analytical sample comprising **1193** observations.

(b)

The analysis involves 780 events (first marriages) and 413 non-events (individuals not marrying during the observation period).

(c)

Based on the definition of seven education levels within the sample size distribution, three EDUCATION categories were created: "Lower Secondary or below" (1-2), "Upper Secondary" (3-4), and "Vocational or tertiary education" (5-7), balancing sample sizes and providing meaningful classification.

The COHORT variable was categorized into two groups reflecting historical and social contexts: the 1940-1969 Cohort, experiencing post-war recovery, economic growth, and social changes, and the 1970-1999 Cohort, growing up during globalization, technological advancements, and evolving values.

(d)

The sample statistics for education level by birth cohort are presented in Table 1.

Table 1. The Sample Distribution for Education by Birth Cohort

DATA06$EDUCATION	DATA06$COHORT 1940-69	1970-99	Total
1-Lower	363	125	488
	55.4%	23.2%	
2-Upper	210	279	489

	32.1%	51.9%	
3-Advanced	82	134	216
	12.5%	24.9%	
Total	655	538	1193
	54.9%	45.1%	

The table indicates a decline in the proportion of lower education levels from 55.4% to 23.2% across the two birth cohorts. Simultaneously, the proportion of upper education levels increased from 32.1% to 51.9%, and advanced education rose from 12.5% to 24.9%. This suggests a general trend of educational expansion and upgrading among Italian women.

Exercise 2.2

(a)

I anticipate differences across birth cohorts, expecting that women born in later cohorts marry later due to social and economic shifts in Italy. Factors such as the rise of individualism, gender equality, and declining traditional values contribute to this hypothesis. Historical shifts, such as youth and student rebellions in the late 1960s and early 1970s, were associated with delayed marriage and childbirth (Brückner & Mayer, 2005). Meanwhile, Postponement may be stronger for events that are hard to reverse and have far-reaching consequences (Billari & Liefbroer, 2010). Therefore, my first hypothesis is:

H1: The more recent birth cohort, the later first marriage.

Additionally, I expect variations in the timing of the first marriage based on education levels. The postponement of marriage and fertility was strongly associated with high education levels for both sexes (Lesthaeghe, 2010), because higher education may increase women's opportunity costs of marriage, such as career prospects, income, and autonomy. Moreover, higher education may also affect women's preferences and expectations for marriage, such as spousal characteristics, division of labor, and fertility. Thus, my second hypothesis is:

H2: The higher the level of education, the later the first marriage.

Assuming that educational differences have narrowed across cohorts, such that the gap in the age at first marriage between women with different levels of education has decreased over time. This is because the diffusion of higher education among women has reduced the selectivity and heterogeneity of the highly educated group. My third hypothesis is:

H3: The gap in the age at first marriage between women with different levels of education has narrowed across cohorts.

(b)

Implementing the survival function for the EVENT of first marriage reveals a significant gap between the two cohorts, as illustrated in Figure 1.

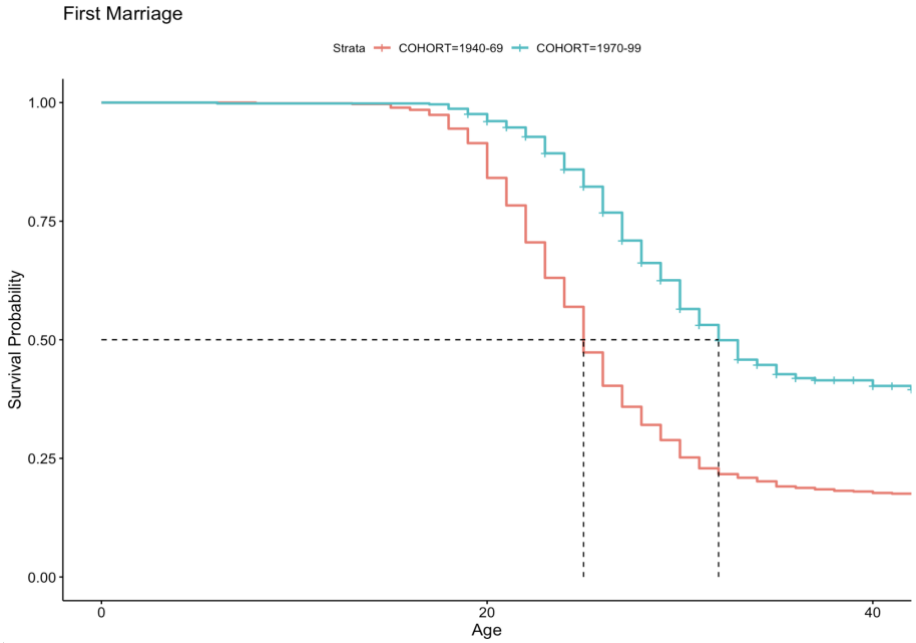

Figure 1. Survival Probability of the Timing of First Marriage (Italy Female)

The median timing of first marriage is 25 for the "1940-69" cohort and 32 for the "1970-99" cohort (see Table 2), supported by a Log-rank test p-value < 0.01, reinforcing Hypothesis 1.

Table 2. The Median Survival Time of Cohorts

strata	median	lower	upper
COHORT=1940-69	25	25	26
COHORT=1970-99	32	31	35

Hypothesis 2 focuses on educational attainment. Table 3 shows substantial differences between the three education categories, with median values of 25, 29 and 32, respectively.

Table 3. The Median Survival Time of Education Categories

strata	median	lower	upper
EDUCATION=1-Lower	25	24	25
EDUCATION=2-Upper	29	28	30
EDUCATION=3-Advanced	32	30	35

The same pattern is evident in the survival function plot (see **Figure 2**). The Log-Rank test also yields a p-value less than 0.01, supporting Hypothesis 2.

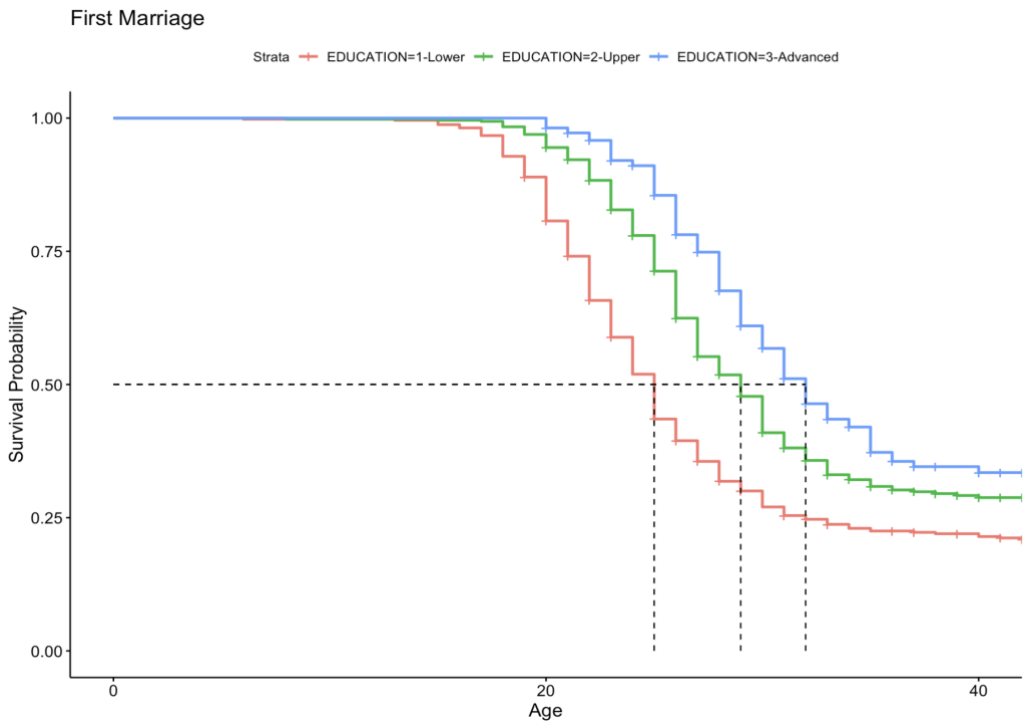

Figure 2. Survival Probability of the Timing of First Marriage by Education (Italy Female)

Examining birth cohort and education level (see **Figure 3**), the "1970-99" cohort experiences first marriage much later. Meanwhile, the gap in different education categories expands in the "1970-99" cohort compared to the "1940-69" cohort.

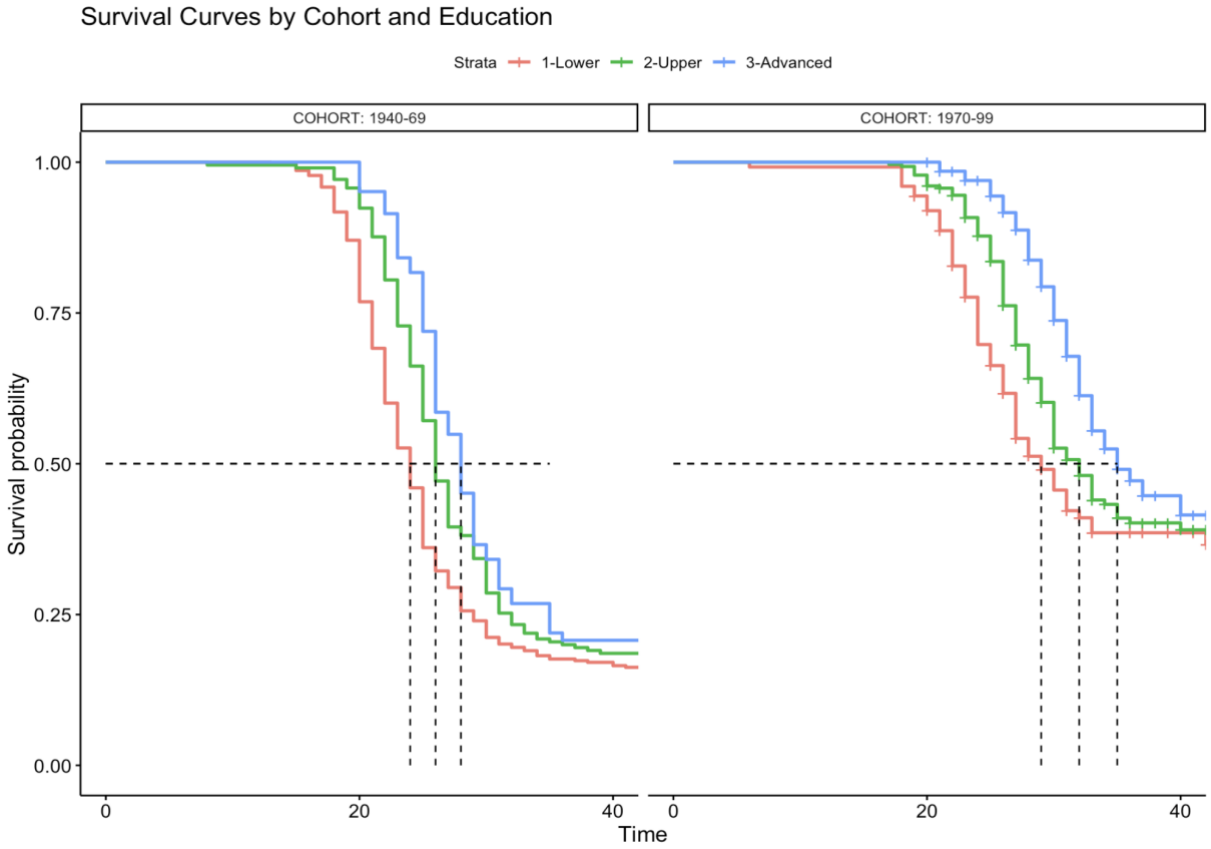

Figure 3. Survival Probability of the Timing of First Marriage by Cohort and Education (Italy Female)

Median values of different educational attainments are 24/26/28 and 29/32/35, respectively (see Table 4). Although the Log-Rank test yields a p-value less than 0.01, the gap expands rather than narrows, rejecting Hypothesis 3.

Table 4. The Median Survival Time of Cohorts and Education Categories

strata	median	lower	upper
COHORT=1940-69, EDUCATION=1-Lower	24	23	25
COHORT=1940-69, EDUCATION=2-Upper	26	26	27
COHORT=1940-69, EDUCATION=3-Advanced	28	26	29
COHORT=1970-99, EDUCATION=1-Lower	29	27	33
COHORT=1970-99, EDUCATION=2-Upper	32	30	35
COHORT=1970-99, EDUCATION=3-Advanced	35	33	NA

In conclusion, more recent birth cohorts and higher education levels correlate with later first marriages. However, the gap between education categories has not narrowed in recent cohorts but has widened.

Exercise 2.3

The patterns observed in the data reveal a significant link between education and the timing of first marriage, with variations across birth cohorts. Higher levels of education are associated with later marriages, potentially influenced by extended periods of schooling, elevated career aspirations, and distinct preferences for partners. Notably, the widening gap in the age at first marriage among education categories in the 1970-1999 cohort compared to the 1940-1969 cohort suggests an escalating impact of education on marriage timing, mirroring societal changes such as expanded higher education, increased female labor force participation, and transformed gender roles.

These patterns carry profound implications for policy interventions, influencing individual and familial well-being, as well as demographic trends. The delay in first marriages may impact fertility rates, pose risks of infertility or pregnancy complications, and shape marital relationships. Policymakers could consider incentives, like financial subsidies, tax benefits, or flexible work arrangements, to support those wishing to marry earlier. Alternatively, fostering acceptance of diverse family structures, including cohabitation, single parenthood, or childlessness, and ensuring equitable access to resources can be a focus.

Recognizing the evolving societal landscape in Italy, policy interventions should address the multifaceted challenges posed by delayed first marriages. Adaptations in educational curricula, initiatives supporting work-life balance, affordable housing, and family-friendly workplaces, alongside efforts to reshape societal perceptions through public awareness campaigns, can collectively contribute to a supportive environment. Understanding and responding to these patterns with thoughtful policies will empower individuals to make informed decisions, fostering a more inclusive and adaptable society.

Exercise 2.4

One variable that may introduce omitted variable bias in the analysis of first marriage timing based on cohort and education is the cultural shift towards individualism. Individualism, characterized by an emphasis on personal autonomy, self-expression, and independence, can significantly influence marriage decisions and the timing of such life events. In the context of Italy, where traditional family values have historically played a pivotal role, a failure to account for the evolving influence of individualism could impact the interpretation of results.

The importance of individualism lies in its potential to reshape societal norms and expectations surrounding marriage. As individualism gains prominence, people may prioritize personal goals, career aspirations, and self-discovery over traditional familial expectations. This shift in values could contribute to both the delay in first marriages and the widening gap observed across education categories and cohorts. Failure to include individualism as a variable may lead to an underestimation of its impact on marriage timing, potentially attributing observed trends solely to educational factors or cohort effects.

The results are affected as the absence of the individualism variable hinders a comprehensive understanding of the drivers behind delayed first marriages in Italy. The influence of cultural shifts towards individualism may confound the observed associations between education, cohort, and marriage timing. Policymakers relying on these findings for decision-making may overlook the nuanced role of evolving cultural values in shaping marriage patterns. Therefore, recognizing and incorporating variables related to cultural shifts, such as individualism, is crucial for a more accurate and nuanced analysis of first marriage timing in Italy.

References:

ESS Round 9: European Social Survey Round 9 Data (2018). Data file edition 3.2. Sikt - Norwegian Agency for Shared Services in Education and Research, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS9-2018.

Billari, F. C., & Liefbroer, A. C. (2010). Towards a new pattern of transition to adulthood? *Advances in Life Course Research*, *15*(2–3), 59–75. https://doi.org/10.1016/j.alcr.2010.10.003

Brückner, H., & Mayer, K. U. (2005). De-Standardization of the Life Course: What it Might Mean? And if it Means Anything, Whether it Actually Took Place? *Advances in Life Course Research*, *9*, 27–53. https://doi.org/10.1016/S1040-2608(04)09002-1

Lesthaeghe, R. (2010). The Unfolding Story of the Second Demographic Transition. *Population and Development Review*, *36*(2), 211–251. https://doi.org/10.1111/j.1728-4457.2010.00328.x

Wickham H, Miller E, Smith D (2023). _haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. *R package version 2.5.4*, <https://CRAN.R-project.org/package=haven>.

Therneau T (2023). _A Package for Survival Analysis in R_. *R package version 3.5-7*, <https://CRAN.R-project.org/package=survival>.

Kassambara A, Kosinski M, Biecek P (2021). *_survminer: Drawing Survival Curves using 'ggplot2'_. R package version 0.4.9*, <https://CRAN.R-project.org/package=survminer>.

```r
library(haven)
library(descr)
library(questionr)
library(survival)
library(survminer)

#### Step 1: Load data ####
rm(list=ls())
DATA01 <- read_dta("assets/ESS9e03_2.dta")


#### Step 2: Select cases & variables ####
# create a subset of Italy with the variables we need
DATA02 <- subset(DATA01, select=c(gndr, yrbrn, inwyys, cntry, eisced,
evmar, maryr), DATA01$cntry == "IT")

# filter the female as the target group
DATA03 <- subset(DATA02, DATA02$gndr == 2)

# filter the people who birth year is between 1940 and 1999
DATA04 <- subset(DATA03, yrbrn >= 1940 & yrbrn <= 1999)
# filter the people who have valid 'eisced' values
DATA05 <- subset(DATA04, DATA04$eisced > 0 & DATA04$eisced < 8)
# filter the people who either live with parents or married between 1955
and 2019
DATA06 <- subset(DATA05, DATA05$evmar == 2 | (DATA05$maryr >= 1955 &
DATA05$maryr < 2019))

# 1193 observations left


#### Step 3: Variable construction ####
# VAR: Event
DATA06$EVENT <- NA
DATA06$EVENT[DATA06$evmar == 2] <- 0
DATA06$EVENT[DATA06$evmar == 1] <- 1
DATA06$EVENT<-as.numeric(DATA06$EVENT)

# VAR: TIME
# if the oberservation already married, then 'TIME' equals the 'maryr' -
'yrbrn'
# if not, then 'the'TIME' equals 'inwyys' - 'yrbrn'
DATA06$TIME <- NA
DATA06$TIME[DATA06$evmar == 1] <- DATA06$maryr[DATA06$evmar == 1] -
DATA06$yrbrn[DATA06$evmar == 1]
DATA06$TIME[DATA06$evmar == 2] <- DATA06$inwyys[DATA06$evmar == 2] -
DATA06$yrbrn[DATA06$evmar == 2]
DATA06$TIME <- as.numeric(DATA06$TIME)

# VAR: COHORT classified as 2 categories
# 1040-1069: 1
# 1970-1999: 2
DATA06$COHORT <- NA
DATA06$COHORT[DATA06$yrbrn>=1940 & DATA06$yrbrn<=1969] <- "1940-69"
```

```r
DATA06$COHORT[DATA06$yrbrn>=1970 & DATA06$yrbrn<=1999] <- "1970-99"
DATA06$COHORT<-factor(DATA06$COHORT)

# VAR: EDUCATION classified as 3 categories based on education level and
sample size
DATA06$EDUCATION <- NA
DATA06$EDUCATION[DATA06$eisced==1 | DATA06$eisced==2] <- "1-Lower"
DATA06$EDUCATION[DATA06$eisced==3 | DATA06$eisced==4] <- "2-Upper"
DATA06$EDUCATION[DATA06$eisced==5 | DATA06$eisced==6 | DATA06$eisced==7]
<- "3-Advanced"
DATA06$EDUCATION <- factor(DATA06$EDUCATION)

# Sample statistics for EDUCATION by COHORT
crosstab(DATA06$EDUCATION, DATA06$COHORT, prop.c = T)

#### Step 4: Analysis ####
# Analysis: Survival Function
# survival function of first marriage group by 'COHORT'
SURVIVAL01 <- survfit(Surv(DATA06$TIME, DATA06$EVENT) ~ COHORT, data =
DATA06)

ggsurvplot(SURVIVAL01, data = DATA06, surv.median.line = "hv",
           # limited x axis to 0-40
           xlim = c(0, 40),
           xlab = "Age", ylab = "Survival Probability",
           title = "First Marriage")

# median values of first marriage group by 'COHORT'
surv_median(SURVIVAL01)
# Log-rank test
survdiff(Surv(DATA06$TIME, DATA06$EVENT) ~ COHORT, data = DATA06)

# survival function of first marriage group by 'EDUCATION'
SURVIVAL02 <- survfit(Surv(DATA06$TIME, DATA06$EVENT) ~ EDUCATION, data =
DATA06)

ggsurvplot(SURVIVAL02, data = DATA06, surv.median.line = "hv",
           # limited x axis to 0-40
           xlim = c(0, 40),
           xlab = "Age", ylab = "Survival Probability",
           title = "First Marriage")

# median values of first marriage group by 'COHORT'
surv_median(SURVIVAL02)
# Log-rank test
survdiff(Surv(DATA06$TIME, DATA06$EVENT) ~ EDUCATION, data = DATA06)

# survival function of first marriage group by 'COHORT' and 'EDUCATION'
ggsurvplot(
  survfit(Surv(DATA06$TIME, DATA06$EVENT) ~ EDUCATION, data = DATA06),
  data = DATA06,
  surv.median.line = "hv",
  facet.by = "COHORT",
  # limited x axis to 0-40
  xlim = c(0, 40),
  title = "Survival Curves by Cohort and Education"
)
```

```r
# median values of first marriage group by 'COHORT' and 'EDUCATION'
SURVIVAL03 <- survfit(Surv(DATA06$TIME, DATA06$EVENT) ~ COHORT +
EDUCATION, data = DATA06)
surv_median(SURVIVAL03)
# Log-rank test
survdiff(Surv(DATA06$TIME, DATA06$EVENT) ~ COHORT + EDUCATION, data =
DATA06)
```