Exercise 04

Exercise 4.4

a)

Migration background can significantly impact unemployment duration due to human capital factors such as language barriers, differences in educational qualifications, and discrimination in the labor market (De Jong & Madamba, 2001). Individuals with a migration background often face challenges in finding employment due to cultural and institutional barriers.

Additionally, gender can also play a role in unemployment duration. In general, migration contributes to labor force non-participation and unemployment among married women (Morrison & Lichter, 1988). Meanwhile, men and women who have contributed little to the family income before becoming unemployed are less likely to look for work than men and women who have provided a higher share of income. Accounting for the actual income distribution of couples and the structural gender inequality present within the large majority of partnerships, women are affected much more often by this mechanism than men (Jacob & Kleinert, 2014).

Furthermore, the interaction between migration background and gender can amplify the challenges faced by migrant women in the labor market, potentially prolonging their unemployment duration compared to their male counterparts. Therefore, the testable hypothesis that deals with the impact of migration background and gender (in interaction) on unemployment duration is formulated as follows:

H1: Individuals with a migration background and of the female gender will experience longer unemployment durations compared to individuals without a migration background and of the male gender.

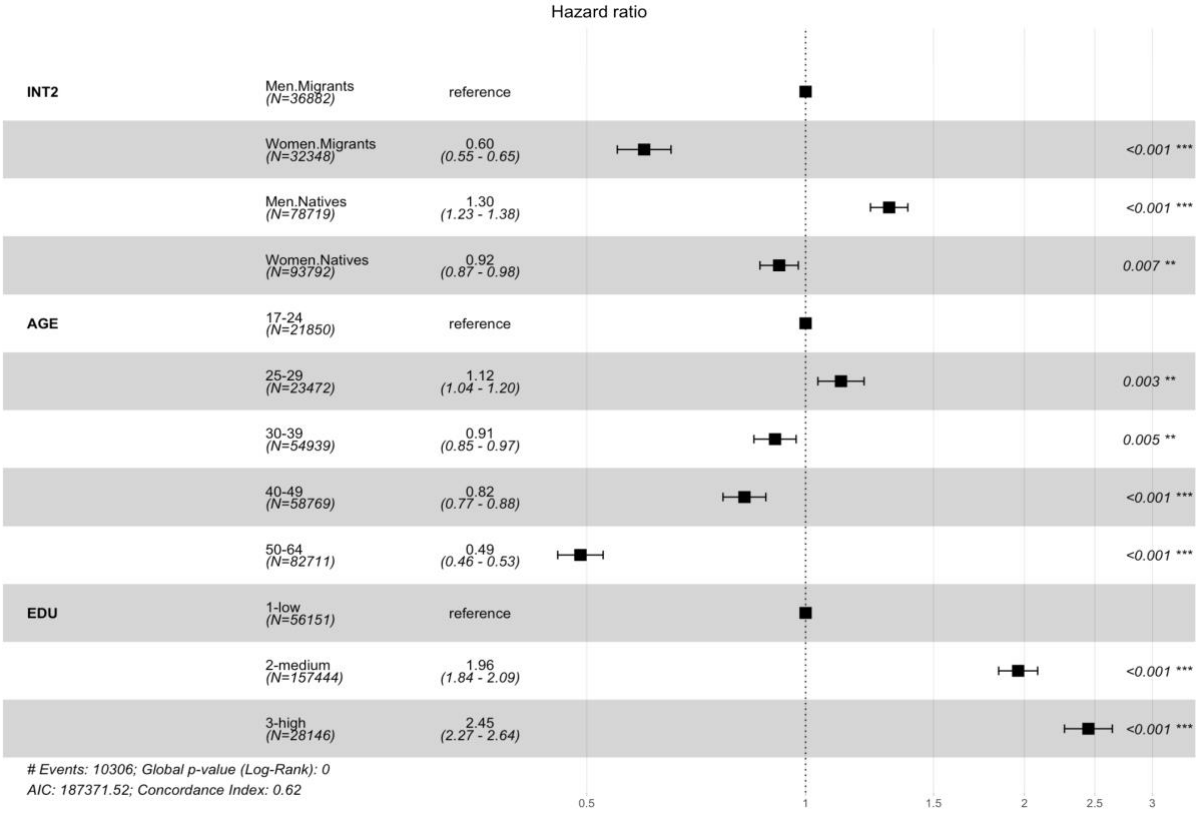
b)

The forest plot (see **Figure 1**) displays the hazard ratios for the variables included in the model: age, education, and the interaction between gender and migration background. The results from the forest plot indicate that the interaction between gender and migration background has a significant effect on unemployment duration.

The hazard ratio for Women Migrants is 0.60 (95% CI: 0.55 - 0.65), indicating that women with a migration background have a significantly lower hazard compared to men with a migration background. On the other hand, the hazard ratio for Men Natives is 1.30 (95% CI: 1.23 - 1.38), suggesting that Men Natives have a higher hazard compared to the reference group of men with a migration background. Similarly, the hazard ratio for Women Natives is 0.92 (95% CI: 0.87 - 0.98), indicating a lower hazard compared to men without a migration background. Furthermore, the risk is lower for individuals with medium to high levels of education compared to those with low education. The risk also increases with age, particularly in people aged 50 and above.

These results support hypothesis **H1** that the impact of migration on unemployment duration differs by gender. Women migrants have a significantly lower hazard compared to men migrants. This suggests that women with a migration background may face additional challenges in the labor market, leading to obviously longer unemployment durations compared to the other three groups: men migrants, women natives, and men natives.

Figure 1. Forest Plot for a Cox-model with migration in interaction with gender



Exercise 4.5

a)

The gendered effect of family status on unemployment duration can be hypothesized to be influenced by traditional gender roles and societal expectations. This gender asymmetry should be particularly pronounced for married couples because marriage is a stronger prerequisite for resource sharing and joint labor market decisions, upholding more traditional gender norms (Jacob & Kleinert, 2014). Moreover, there is a considerable “marriage premium” for men; that is, married men re-enter work more quickly than single men, while married women are less likely to enter re-employment than single women (Jacob & Kleinert, 2014). Furthermore, we expect that unmarried women will have shorter unemployment durations compared to unmarried men, as they may have more flexibility and fewer familial responsibilities. Therefore, here is the hypothesis:

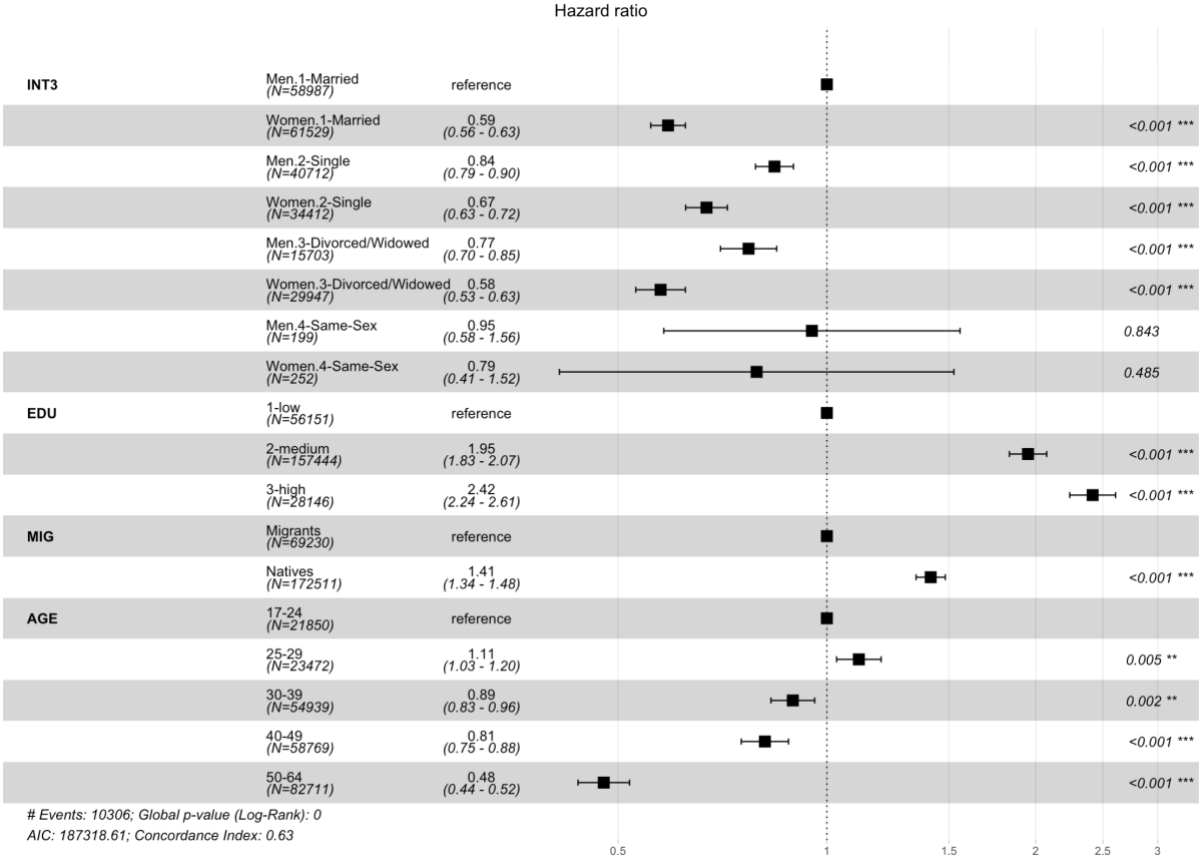
H2: Family status has a gendered effect on unemployment duration, with married men experiencing shorter durations compared to unmarried men, while married women experience longer durations compared to unmarried women.

To test this hypothesis, an analysis of the interaction between gender and family status in relation to unemployment duration would be necessary. This could provide insights into the differential impact of family status on unemployment duration for men and women, considering societal expectations and traditional gender roles.

b)

The forest plot (see **Figure 2**) illustrates the hazard ratios for the variables included in the model: education, migration status, age, and the interaction between gender and family status. The results indicate a significant impact of family status and gender on unemployment duration. The hazard ratio for Married Women is 0.59, suggesting that married women have a 41% lower hazard to re-enter work compared to married men. On the other hand, the hazard ratio for Single Women is 0.67, indicating a lower hazard compared to married men but higher than married women. Meanwhile, the hazard ratio for Unmarried Men is 0.84, suggesting a lower hazard compared to the reference group of married men. These results support hypothesis H2, indicating that the gendered effect of family status on unemployment duration is significant.

Figure 2. Forest Plot for a Cox-model with family status in interaction with gender



In addition, the divorced and widowed individuals (women: HR = 0.58, men: HR = 0.77) also display a reduced risk of occurrence in comparison to the married group. Notably, divorced or widowed women exhibit the lowest hazard ratio among all groups. Same-sex migrants do not show a statistically significant difference, neither for women nor for men.

Education, migrant status, and age significantly influence unemployment duration. Individuals aged 50 and above face an increased risk of longer unemployment periods. Meanwhile, those with medium to high levels of education have a lower hazard compared to

those with low education. Migrant status is linked to higher unemployment durations compared to non-migrants.

These findings highlight the multifaceted nature of unemployment duration and underscore the need for targeted policies and interventions to address the specific challenges faced by different demographic groups in the labor market.

Exercise 4.6

One crucial omitted variable that could bias the results is "language proficiency" in Germany. Language proficiency is a significant factor for labor market outcomes among migrants. Individuals with limited proficiency in the local language may face barriers in finding employment opportunities that require strong communication skills. This could lead to longer unemployment durations for migrants compared to non-migrants.

Similarly, differences in cultural and social capital, such as professional networks and knowledge of local labor market norms, could also play a role in influencing unemployment duration. These factors are often not captured in standard datasets but could significantly impact the results. Furthermore, discriminatory practices in the labor market, which are not directly observable in the data, could also contribute to longer unemployment durations for migrants.

Therefore, while the model indicates a strong effect of migration status on unemployment duration, it is important to acknowledge that these results may be influenced by omitted variables that could bias the causal interpretation of the effect. Further research that accounts for these potential omitted variables would provide a more comprehensive understanding of the causal relationship between migration status and unemployment duration.

References:

De Jong, G. F., & Madamba, A. B. (2001). A Double Disadvantage? Minority Group, Immigrant Status, and Underemployment in the United States. *Social Science Quarterly*, *82*(1), 117–130. https://doi.org/10.1111/0038-4941.00011

Goebel, Jan, Markus M. Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp. 2019. The German Socio-Economic Panel (SOEP). Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 239 (2), 345-360. (https://doi.org/10.1515/jbnst-2018-0022)

Jacob, M., & Kleinert, C. (2014). Marriage, Gender, and Class: The Effects of Partner Resources on Unemployment Exit in Germany. *Social Forces*, *92*(3), 839–871. https://doi.org/10.1093/sf/sot130

Morrison, D. R., & Lichter, D. T. (1988). Family Migration and Female Employment: The Problem of Underemployment among Migrant Married Women. *Journal of Marriage and Family*, *50*(1), 161–172. https://doi.org/10.2307/352436

R Packages:

Enzmann JAIRscadwbD, Schwartz M, Jain N, Kraft S (2023). _descr: Descriptive Statistics_. R package version 1.1.8, <https://CRAN.R-project.org/package=descr>.

Göran Broström (2022). Event History Analysis with R, Second Edition. Chapman and Hall/CRC, Boca Raton.

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>.

Rich B (2023). _table1: Tables of Descriptive Statistics in HTML_. R package version 1.4.3, <https://CRAN.R-project.org/package=table1>.

Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible summary tables with the gtsummary package. The R Journal 2021;13:570–80. https://doi.org/10.32614/RJ-2021-053.

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

```
library(haven)
library(descr)
library(survival)
library(survminer)
library(lmtest)
library(table1)
library(gtsummary)


#### Step 1: Load data ####
rm(list=ls())

# All data from SOEP (http://companion.soep.de/)
# Meta Data (constant)
FIX <- read_dta("assets/FIX.dta") # 20862 obs.
FIX <- FIX[order(FIX$ID),]

# Personal Data (yearly TV)
VARYING <- read_dta("assets/VARYING.dta")  # 36833 obs.
VARYING <- VARYING[order(VARYING$ID,VARYING$SYEAR),]

# merge the two datasets by ID
SUBSET01 <- merge(VARYING,FIX, by=c("ID"))
SUBSET01 <- SUBSET01[order(SUBSET01$ID,SUBSET01$SYEAR),]

# Employment Biographies (monthly TV)
CALEN <- read_dta("assets/CALEN.dta")  # 241741 obs.
CALEN <- CALEN[order(CALEN$ID,CALEN$SYEAR,CALEN$BEGIN),]

# merge CALEN with SUBSET01 by ID and SYEAR
SUBSET02 <- merge(CALEN, SUBSET01, by=c("ID","SYEAR"))
SUBSET02 <-
SUBSET02[order(SUBSET02$ID,SUBSET02$SYEAR,SUBSET02$BEGIN),]


#### Step 2: Variable construction ####

#VAR: SEX
SUBSET02$SEX<-"NA"
SUBSET02$SEX[SUBSET02$SEX01==1]<-"Men"
SUBSET02$SEX[SUBSET02$SEX01==2]<-"Women"
```

```
SUBSET02$SEX<-as.factor(SUBSET02$SEX)

#VAR: AGECAT
SUBSET02$AGE<-"4-Other"
SUBSET02$AGE[SUBSET02$AGE01>=17]<-"17-24"
SUBSET02$AGE[SUBSET02$AGE01>=25]<-"25-29"
SUBSET02$AGE[SUBSET02$AGE01>=30]<-"30-39"
SUBSET02$AGE[SUBSET02$AGE01>=40]<-"40-49"
SUBSET02$AGE[SUBSET02$AGE01>=50]<-"50-64"
SUBSET02$AGE<-as.factor(SUBSET02$AGE)

#VAR: MIG
SUBSET02$MIG<-"NA"
SUBSET02$MIG[SUBSET02$MIGBACK==1]<-"Natives"
SUBSET02$MIG[SUBSET02$MIGBACK==2]<-"Migrants"
SUBSET02$MIG<-as.factor(SUBSET02$MIG)

#VAR: WORRY HEALTH
SUBSET02$HEALTH<-"NA"
SUBSET02$HEALTH[SUBSET02$HEALTH01==1]<-"1-Poor"
SUBSET02$HEALTH[SUBSET02$HEALTH01==2]<-"2-Medium"
SUBSET02$HEALTH[SUBSET02$HEALTH01==3]<-"3-Good"
SUBSET02$HEALTH<-as.factor(SUBSET02$HEALTH)

#VAR: EDU
SUBSET02$EDU<-"NA"
SUBSET02$EDU[SUBSET02$PGISCED97==1]<-"1-low"
SUBSET02$EDU[SUBSET02$PGISCED97==2]<-"1-low"
SUBSET02$EDU[SUBSET02$PGISCED97==3]<-"2-medium"
SUBSET02$EDU[SUBSET02$PGISCED97==4]<-"2-medium"
SUBSET02$EDU[SUBSET02$PGISCED97==5]<-"2-medium"
SUBSET02$EDU[SUBSET02$PGISCED97==6]<-"3-high"
SUBSET02$EDU<-as.factor(SUBSET02$EDU)

#VAR: FAMILY STATUS
SUBSET02$FAM<-"NA"
SUBSET02$FAM[SUBSET02$PGFAMSTD==1]<-"1-Married"
SUBSET02$FAM[SUBSET02$PGFAMSTD==2]<-"1-Married"
SUBSET02$FAM[SUBSET02$PGFAMSTD==3]<-"2-Single"
SUBSET02$FAM[SUBSET02$PGFAMSTD==4]<-"3-Divorced/Widowed"
SUBSET02$FAM[SUBSET02$PGFAMSTD==5]<-"3-Divorced/Widowed"
SUBSET02$FAM[SUBSET02$PGFAMSTD==6]<-"1-Married"
SUBSET02$FAM[SUBSET02$PGFAMSTD==7]<-"4-Same-Sex"
SUBSET02$FAM[SUBSET02$PGFAMSTD==8]<-"4-Same-Sex"
```

```r
SUBSET02$FAM <- as.factor(SUBSET02$FAM)

# construct the event variable (employment)
SUBSET02$Surv <- with(SUBSET02, Surv(BEGIN, END, EVENT==1))

#Summary Statistics
table1(~ HEALTH + AGE + SEX + MIG | MIG, data = SUBSET02)



#### Step 3: Analysis ####

# fit the Weibull model with migration status
SURVIVAL01 <- survfit(SUBSET02$Surv ~ SUBSET02$MIGBACK)
plot(SURVIVAL01, col = c("blue", "red"), xlab="unemployment
duration (in months)")
legend("topright", legend = c("Natives", "Migrants"), col =
c("blue", "red"), lwd=2, cex = 1,)

# fit the Cox model with gender, migration status, age
MODEL01 <- coxph(SUBSET02$Surv ~ SEX+MIG+AGE, data=SUBSET02)
OUTPUT01 <- tbl_regression(MODEL01, exponentiate = TRUE)
tbl_merge(
  tbls = list(OUTPUT01),
  tab_spanner = c("Model01"))

# fit the Cox model with gender, migration status, age, education
MODEL02 <- coxph(SUBSET02$Surv ~ SEX+MIG+AGE+EDU, data=SUBSET02)
MODEL03 <- coxph(SUBSET02$Surv ~ SEX+MIG+AGE+EDU+HEALTH,
data=SUBSET02)
OUTPUT02 <- tbl_regression(MODEL02, exponentiate = TRUE)
OUTPUT03 <- tbl_regression(MODEL03, exponentiate = TRUE)
tbl_merge( tbls = list(OUTPUT01, OUTPUT02, OUTPUT03),
           tab_spanner = c("Model01", "Model02", "Model03"))

# Conduct the likelihood ratio test
lrtest(MODEL01, MODEL02)
lrtest(MODEL02, MODEL03)

# Exercise 4.4

# Interaction effect between health and migration status
SUBSET02$INT <- interaction(SUBSET02$HEALTH, SUBSET02$MIG)
MODEL03_2 <- coxph(SUBSET02$Surv ~ INT+AGE+SEX+EDU,
data=SUBSET02)
```

```r
OUTPUT03_2 <- tbl_regression(MODEL03_2, exponentiate = TRUE)

# Use the survminer package to visualize the results
ggforest(MODEL03_2, data = NULL,fontsize = 1)

# Interaction effect betwwen gender and migration status
SUBSET02$INT2 <- interaction(SUBSET02$SEX, SUBSET02$MIG)
MODEL04 <- coxph(SUBSET02$Surv ~ INT2+AGE+EDU, data=SUBSET02)
OUTPUT04 <- tbl_regression(MODEL04, exponentiate = TRUE)
# Visualize the results
ggforest(MODEL04, data = NULL,fontsize = 1)

# Exercise 4.5

# Interaction effect between gender and family status
SUBSET02$INT3 <- interaction(SUBSET02$SEX, SUBSET02$FAM)
MODEL05 <- coxph(SUBSET02$Surv ~ INT3+EDU+MIG+AGE, data=SUBSET02)
OUTPUT05 <- tbl_regression(MODEL05, exponentiate = TRUE)

# Visualize the results
ggforest(MODEL05, data = NULL,fontsize = 1)
```