Topics in Labor & Demo: Midterm Problem Set

Gabriel Tabak, Guilherme Cabrine & Gustavo Henrique

2025-05-30

Loading/Installing needed packages

Question 1

Importing databases

```
# Importing SDEM final dataset
sdem = read_csv(here("Data/Final", "sdem.csv"))
```

```
## Rows: 4416213 Columns: 23
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (23): cd_a, ent, con, v_sel, n_hog, h_mud, n_ren, n_pro_viv, year, par_c...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Cleaning memory
gc()
```

```
##             used  (Mb) gc trigger  (Mb)  max used  (Mb)
## Ncells   1071396  57.3    2142754 114.5   1422285  76.0
## Vcells 103430221 789.2  126209511 963.0 103688098 791.1
```

```
# Price index Data
mexcpi = read_excel(here("Data/Raw", "MEXCPIALLAINMEI.xlsx"),
                    sheet = "Annual") %>%
  mutate(year = as.numeric(format(observation_date, "%Y"))) %>%
  rename(cpi = MEXCPIALLAINMEI) %>%
  select(-1)
```

Item a

```r
# Get some household variables
sdem_agg = sdem %>% group_by(cd_a, ent, con, v_sel, n_hog, h_mud, year) %>%
  summarise(age_hhead = eda[which(par_c == 101)],
            age_hspouse = first(eda[par_c %in% c(201,202)]),
            chld_less16 = as.integer(sum(par_c %in% c(301,302,303) &
                                           eda < 16) > 0))
```

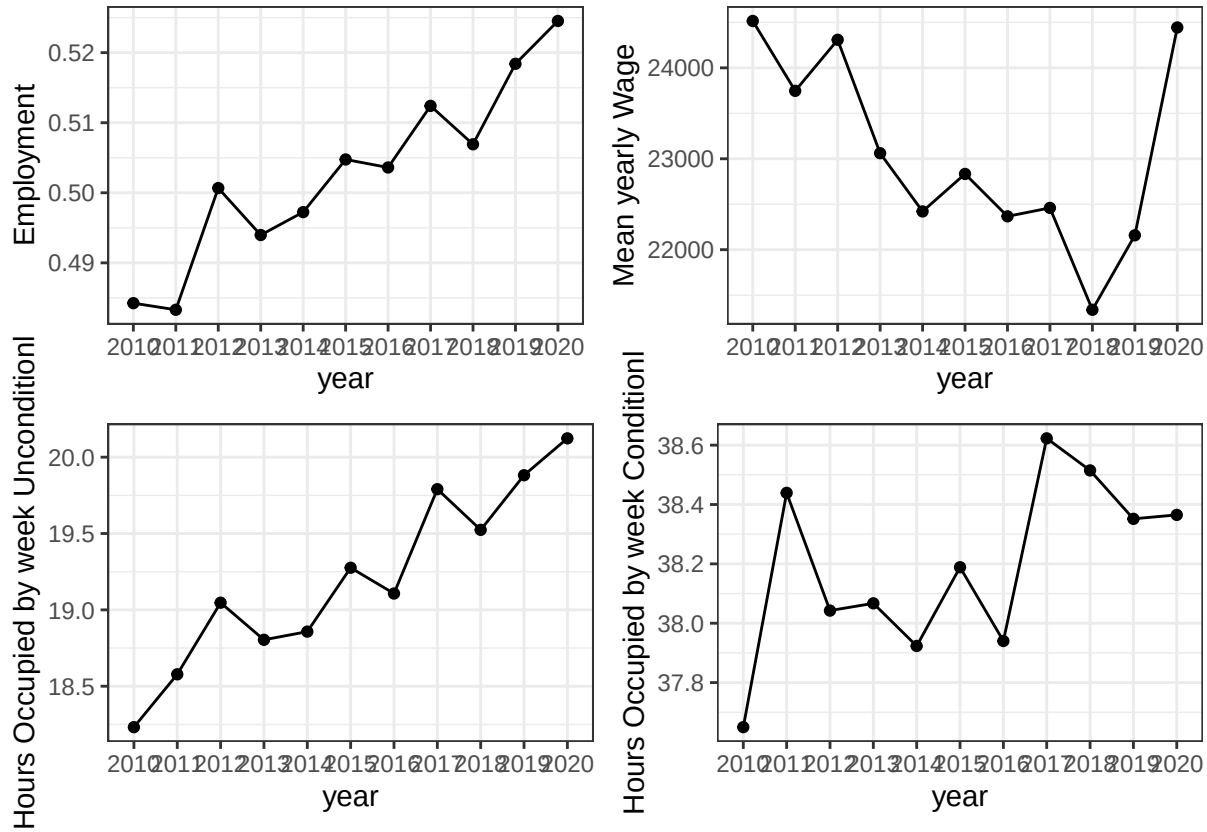
```
## `summarise()` has grouped output by 'cd_a', 'ent', 'con', 'v_sel', 'n_hog',
## 'h_mud'. You can override using the `.groups` argument.
```

```r
# Join with the data
sdem = sdem %>%
  left_join(sdem_agg)
```

```
## Joining with `by = join_by(cd_a, ent, con, v_sel, n_hog, h_mud, year)`
```

```r
# Removing unneeded df
rm(sdem_agg)

## Employed
# hrsocu > 0 and non empty, also, age_hhead > 13 and age_hhead <= 98
sdem = sdem %>%
  mutate(Employed = ifelse((hrsocup >0) & (age_hhead <= 98) & (age_hhead > 13),
                                                        1, 0))

# Get the conditional hours
sdem$HourC = ifelse(sdem$Employed == 1,
                          sdem$hrsocup, NA)

# Deflate the salary
sdem = sdem %>%
  left_join(mexcpi, by = c("year"))

sdem = sdem %>%
  mutate(Wage = 52*ing_x_hrs*hrsocup*(100/cpi))

df1 = sdem %>%
  filter(sex == 2,
         eda %in% 25:65) %>%
  group_by(year) %>%
  summarise(EmplR = mean(Employed),
            YWage = mean(Wage),
            Hrs = mean(hrsocup),
            HrsC = mean(HourC, na.rm = T)) %>%
  mutate(year = as.factor(year))

g1 = ggplot(df1, aes(year, EmplR)) + geom_point() +
  geom_line(aes(group = 1)) + theme_bw() +
  ylab("Employment")

g2 = ggplot(df1, aes(year, YWage)) + geom_point() +
  geom_line(aes(group = 1)) +theme_bw() +
```

```
  ylab("Mean yearly Wage")

g3 = ggplot(df1, aes(year, Hrs)) + geom_point() +
  geom_line(aes(group = 1)) + theme_bw() +
  ylab("Hours Occupied by week Unconditionl")

g4 = ggplot(df1, aes(year, HrsC)) + geom_point() +
  geom_line(aes(group = 1)) + theme_bw() +
  ylab("Hours Occupied by week Conditionl")

(g1 | g2) / (g3 | g4)
```




```
ggsave(here("Out", "plot1.png"), width = 12, height = 12, dpi=600)

rm(g1, g2, g3, g4, df1)
```

Item b

```
df = sdem %>%
  filter(sex == 2, eda %in% 25:65) %>%
  mutate(hj_peq = case_when(
    n_hij != 0 & chld_less16 != 0 & (par_c %in% c(101,201)) ~ 1,
    T ~ 0
```

```
  ))

Tables = function(df1, status){
  df1 = df1 %>%
    filter(e_con == status)

  final = matrix(0, 5, 8)
  final = data.frame(final)

  # All
  final[1,c(1,5)] = c(mean(df1$Employed), mean(df1$Wage/52))

  final[1,c(2,6)] = df1 %>%
    filter(cs_p13_1 < 4) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  final[1,c(3,7)] = df1 %>%
    filter(cs_p13_1 == 4) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  final[1,c(4,8)] = df1 %>%
    filter(cs_p13_1 %in% c(7,8,9)) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  # Less than 35 with young child
  final[2,c(1,5)] = df1 %>%
    filter(hj_peq == 1, eda < 35) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  final[2,c(2,6)] = df1 %>%
    filter(hj_peq == 1, eda < 35, cs_p13_1 < 4) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  final[2,c(3,7)] = df1 %>%
    filter(hj_peq == 1, eda < 35, cs_p13_1 == 4) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  final[2,c(4,8)] = df1 %>%
    filter(hj_peq == 1, eda < 35,cs_p13_1 %in% c(7,8,9)) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  # Less than 35 with no child
  final[3,c(1,5)] = df1 %>%
    filter(hj_peq == 0, eda < 35) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()
```

```r
final[3,c(2,6)] = df1 %>%
  filter(hj_peq == 0, eda < 35,cs_p13_1 < 4) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[3,c(3,7)] = df1 %>%
  filter(hj_peq == 0, eda < 35,cs_p13_1 == 4) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[3,c(4,8)] = df1 %>%
  filter(hj_peq == 0, eda < 35,cs_p13_1 %in% c(7,8,9)) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()


# Edad: 35-54
final[4,c(1,5)] = df1 %>%
  filter(eda >= 35, eda<=54) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[4,c(2,6)] = df1 %>%
  filter(eda >= 35, eda<=54,cs_p13_1 < 4) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[4,c(3,7)] = df1 %>%
  filter(eda >= 35, eda<=54,cs_p13_1 == 4) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[4,c(4,8)] = df1 %>%
  filter(eda >= 35, eda<=54,cs_p13_1 %in% c(7,8,9)) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

# Edad: 55+
final[5,c(1,5)] = df1 %>%
  filter(eda >= 55) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[5,c(2,6)] = df1 %>%
  filter(eda >= 55,cs_p13_1 < 4) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()

final[5,c(3,7)] = df1 %>%
  filter(eda >= 55,cs_p13_1 == 4) %>%
  summarise(mean(Employed), mean(Wage/52)) %>%
  as.numeric()
```

```
  final[5,c(4,8)] = df1 %>%
    filter(eda >= 55,cs_p13_1 %in% c(7,8,9)) %>%
    summarise(mean(Employed), mean(Wage/52)) %>%
    as.numeric()

  rownames(final) = c("All", "Less 35 - Young Chld",
                      "Less 35 - No Chld", "35-54",
                      "55+")

  colnames(final) = c("All", "Less HS", "HS", "College+",
                      "All - W", "Less HS - W", "HS - W", "College+ - W")
  return(final)
}


# Marital status: 5 = Married, 1 = Cohabiting & 6 = Single

Tables(df,5)
```

```
##                            All   Less HS        HS  College+  All - W
## All                  0.4212632 0.3578620 0.4437487 0.6010074 374.6771
## Less 35 - Young Chld 0.4013977 0.3249770 0.3977149 0.5912518 372.7436
## Less 35 - No Chld    0.4940442 0.3686817 0.4528258 0.6526341 460.7936
## 35-54                0.4643280 0.3987468 0.4846720 0.6388321 424.9992
## 55+                  0.2838355 0.2757894 0.2917112 0.3644491 194.4755
##                      Less HS - W   HS - W College+ - W
## All                     207.8507 372.1216     895.9341
## Less 35 - Young Chld    194.1302 329.7260     868.9485
## Less 35 - No Chld       221.8309 330.9636     805.1541
## 35-54                   239.0613 418.7880     989.6751
## 55+                     136.5604 230.3193     551.8663
```

```
Tables(df,1)
```

```
##                            All   Less HS        HS  College+  All - W
## All                  0.4613146 0.4133242 0.4965083 0.6689120 381.9526
## Less 35 - Young Chld 0.3968082 0.3451432 0.4366172 0.6158276 324.7059
## Less 35 - No Chld    0.4732909 0.3757503 0.4766594 0.6670692 428.8468
## 35-54                0.5116342 0.4664387 0.5783260 0.7251808 427.9895
## 55+                  0.3862294 0.3779557 0.4619772 0.4475874 245.4957
##                      Less HS - W   HS - W College+ - W
## All                     266.5754 429.3883     949.7548
## Less 35 - Young Chld    223.9082 366.4286     831.0939
## Less 35 - No Chld       251.3204 376.4126     853.3516
## 35-54                   305.7285 528.2982    1121.5445
## 55+                     203.9153 470.3692     635.6474
```

```
Tables(df,6)
```

```
##                            All   Less HS        HS  College+  All - W
## All                  0.6936368 0.6228403 0.7394974 0.7536346 623.8936
## Less 35 - Young Chld 0.7874754 0.7641909 0.7998629 0.8319198 815.4034
```

```
## Less 35 - No Chld    0.7145817 0.6268780 0.7370195 0.7488291 636.1009
## 35-54                0.7210457 0.6538024 0.7666159 0.8054700 655.4747
## 55+                  0.4711964 0.4773813 0.5041693 0.4925293 392.8549
##                      Less HS - W   HS - W College+ - W
## All                     413.9822 582.6194     879.6044
## Less 35 - Young Chld    620.6142 816.3482    1323.2247
## Less 35 - No Chld       406.8764 539.7035     798.7628
## 35-54                   438.8688 635.4535    1052.6012
## 55+                     291.7393 428.5993     723.9857
```

Item c

```
## household chores W
df = sdem %>%
  filter(e_con %in% c(1,5),
         sex == 2,
         eda %in% 25:65,
         par_c %in% c(101,201))

## household chores man
dfM = sdem %>%
  filter(e_con %in% c(1,5),
         sex == 1,
         par_c %in% c(101,201))

# Removing unneeded df
rm(sdem)

# Cleaning memory
gc()
```

```
##            used  (Mb) gc trigger   (Mb)  max used   (Mb)
## Ncells  1306540  69.8    5414706  289.2  16524367  882.5
## Vcells 46586630 355.5  251715981 1920.5 314620458 2400.4
```

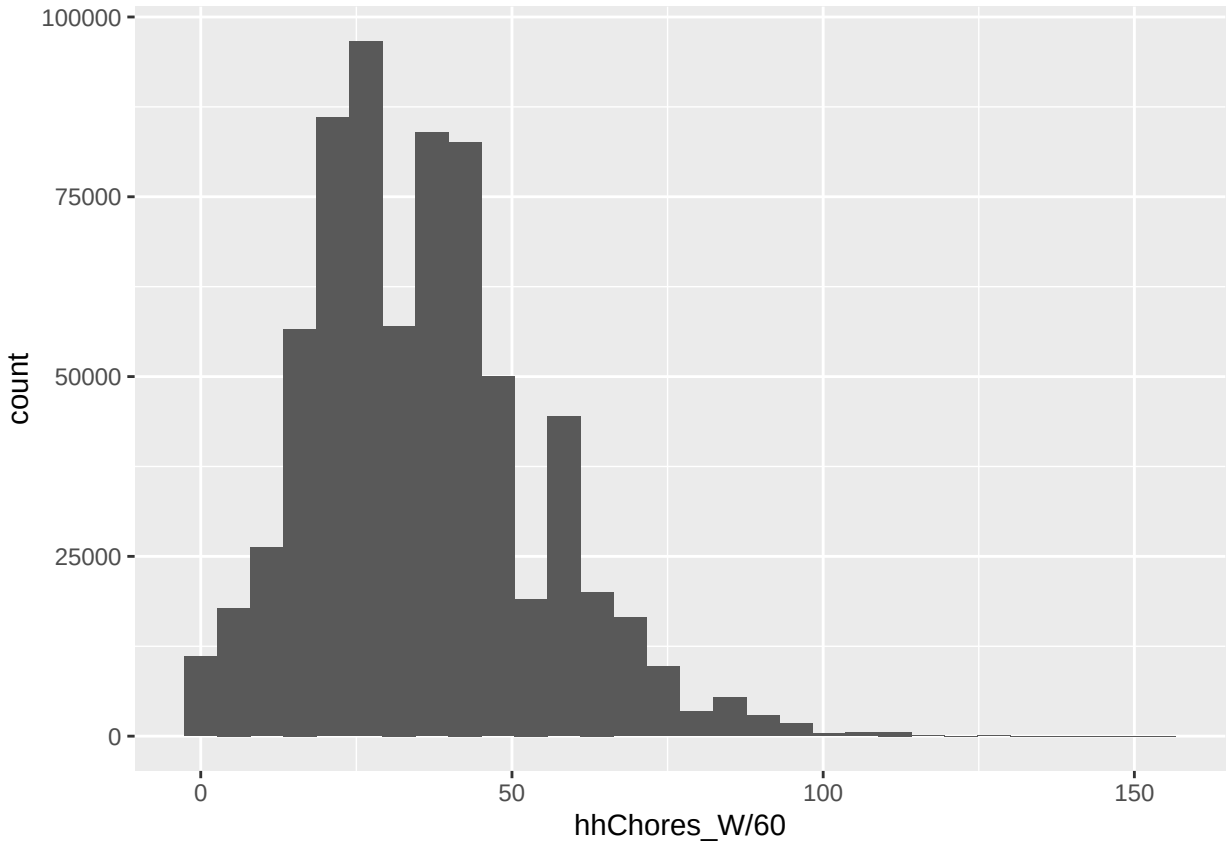
```
# Remove pairs with 2 W
Aux = df %>%
  group_by(cd_a, ent, con, v_sel, n_hog, h_mud, year) %>%
  summarise(n = n()) %>%
  filter(n>1)
```

```
## `summarise()` has grouped output by 'cd_a', 'ent', 'con', 'v_sel', 'n_hog',
## 'h_mud'. You can override using the `.groups` argument.
```

```
X = paste(df$cd_a, df$ent, df$con, df$v_sel,
          df$n_hog, df$h_mud, df$year)
X2 = paste(Aux$cd_a, Aux$ent, Aux$con, Aux$v_sel,
           Aux$n_hog, Aux$h_mud, Aux$year)

df = df[!X%in%X2,]
```

```r
rm(X, X2, Aux)

df$p11_h2 = ifelse(is.na(df$p11_h2), 0, df$p11_h2)
df$p11_h5 = ifelse(is.na(df$p11_h5), 0, df$p11_h5)
df$p11_h7 = ifelse(is.na(df$p11_h7), 0, df$p11_h7)
df$p11_m2 = ifelse(is.na(df$p11_m2), 0, df$p11_m2)
df$p11_m5 = ifelse(is.na(df$p11_m5), 0, df$p11_m5)
df$p11_m7 = ifelse(is.na(df$p11_m7), 0, df$p11_m7)

df = df %>%
  filter(p11_h2 != 98, p11_h5 != 98, p11_h7 != 98)

df$p11_h2 = ifelse(df$p11_h2 == 99 | is.na(df$p11_h2), 0, df$p11_h2)
df$p11_h5 = ifelse(df$p11_h5 == 99 | is.na(df$p11_h5), 0, df$p11_h5)
df$p11_h7 = ifelse(df$p11_h7 == 99 | is.na(df$p11_h7), 0, df$p11_h7)


df$hhChores_W = ifelse(df$year < 2013,  df$p11_h2*60 + df$p11_m2 +
                                         df$p11_h5*60 + df$p11_m5,0)
df$hhChores_W = ifelse(df$year >= 2013, df$p11_h2*60 + df$p11_m2 +
                                         df$p11_h7*60 + df$p11_m7, df$hhChores_W)

df %>%
  ggplot() +
  geom_histogram(mapping = aes(hhChores_W/60))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



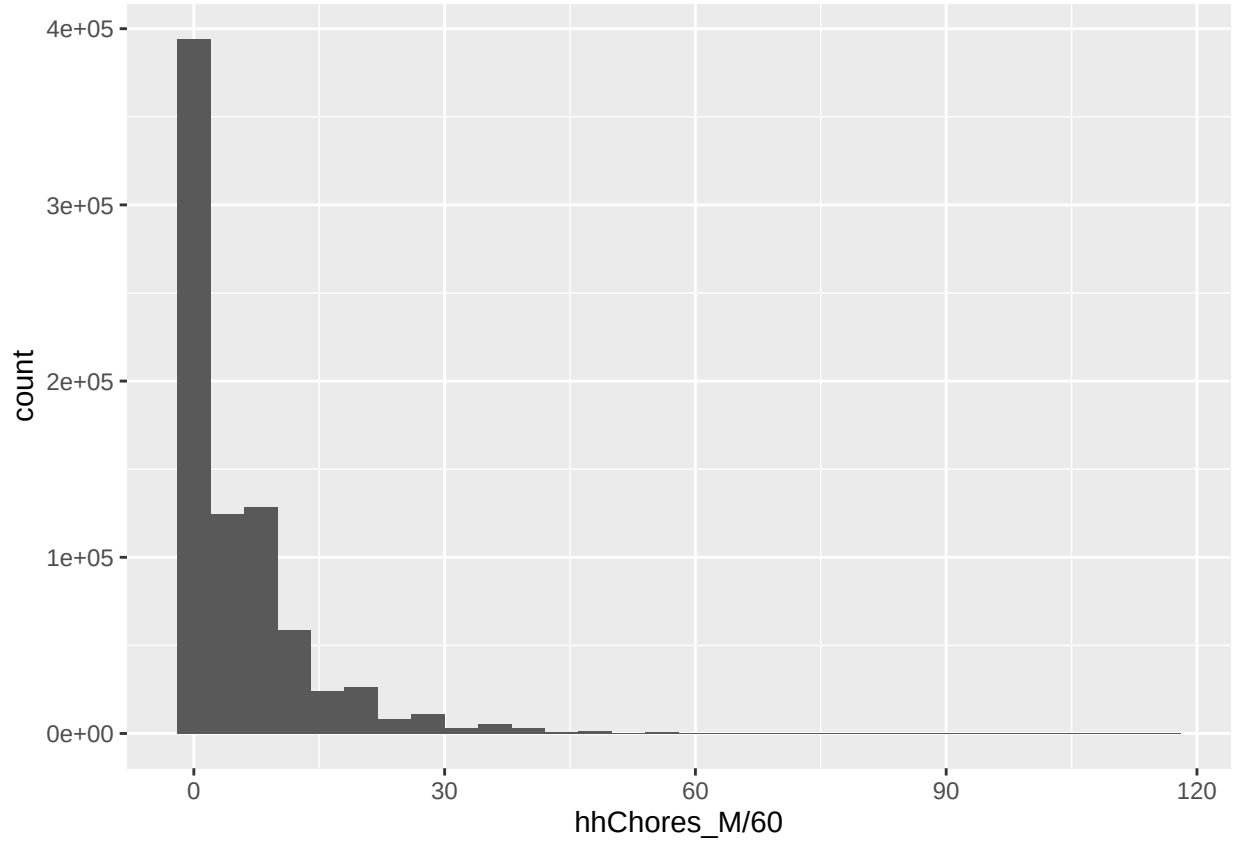

```
# Remove pairs with 2 M
Aux = dfM %>%
  group_by(cd_a, ent, con, v_sel, n_hog, h_mud, year) %>%
  summarise(n = n()) %>%
  filter(n>1)
```

```
## `summarise()` has grouped output by 'cd_a', 'ent', 'con', 'v_sel', 'n_hog',
## 'h_mud'. You can override using the `.groups` argument.
```

```
X = paste(dfM$cd_a, dfM$ent, dfM$con, dfM$v_sel,
          dfM$n_hog, dfM$h_mud, dfM$year)
X2 = paste(Aux$cd_a, Aux$ent, Aux$con, Aux$v_sel,
           Aux$n_hog, Aux$h_mud, Aux$year)

dfM = dfM[!X%in%X2,]

rm(X, X2, Aux)

dfM$p11_h2 = ifelse(is.na(dfM$p11_h2), 0, dfM$p11_h2)
dfM$p11_h5 = ifelse(is.na(dfM$p11_h5), 0, dfM$p11_h5)
dfM$p11_h7 = ifelse(is.na(dfM$p11_h7), 0, dfM$p11_h7)
dfM$p11_m2 = ifelse(is.na(dfM$p11_m2), 0, dfM$p11_m2)
dfM$p11_m5 = ifelse(is.na(dfM$p11_m5), 0, dfM$p11_m5)
dfM$p11_m7 = ifelse(is.na(dfM$p11_m7), 0, dfM$p11_m7)
```

```r
dfM = dfM %>%
  filter(p11_h2 != 98, p11_h5 != 98, p11_h7 != 98)

dfM$p11_h2 = ifelse(dfM$p11_h2 == 99 | is.na(dfM$p11_h2), 0, dfM$p11_h2)
dfM$p11_h5 = ifelse(dfM$p11_h5 == 99 | is.na(dfM$p11_h5), 0, dfM$p11_h5)
dfM$p11_h7 = ifelse(dfM$p11_h7 == 99 | is.na(dfM$p11_h7), 0, dfM$p11_h7)

dfM$hhChores_M = ifelse(dfM$year < 2013,  dfM$p11_h2*60 + dfM$p11_m2 +
                          dfM$p11_h5*60 + dfM$p11_m5,0)

dfM$hhChores_M = ifelse(dfM$year >= 2013, dfM$p11_h2*60 + dfM$p11_m2 +
                          dfM$p11_h7*60 + dfM$p11_m7, dfM$hhChores_M)

dfM %>%
  ggplot() +
  geom_histogram(mapping = aes(hhChores_M/60))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
dfM = dfM %>%
  select(cd_a, ent, con, v_sel, n_hog,
                      h_mud, n_ren, n_pro_viv, year, hhChores_M)

#### Join- W + M
```

```r
df = df %>%
  left_join(dfM, by = c("cd_a", "ent", "con", "v_sel", "n_hog", "h_mud", "year"))

df = df[!is.na(df$hhChores_M),]

df %>% filter(e_con == 1) %>%
  group_by(year) %>%
  summarise(Ratio = mean(hhChores_W/(hhChores_M + hhChores_W), na.rm = T))
```

```
## # A tibble: 11 x 2
##     year Ratio
##    <dbl> <dbl>
##  1  2010 0.885
##  2  2011 0.881
##  3  2012 0.877
##  4  2013 0.869
##  5  2014 0.871
##  6  2015 0.868
##  7  2016 0.866
##  8  2017 0.864
##  9  2018 0.864
## 10  2019 0.851
## 11  2020 0.844
```

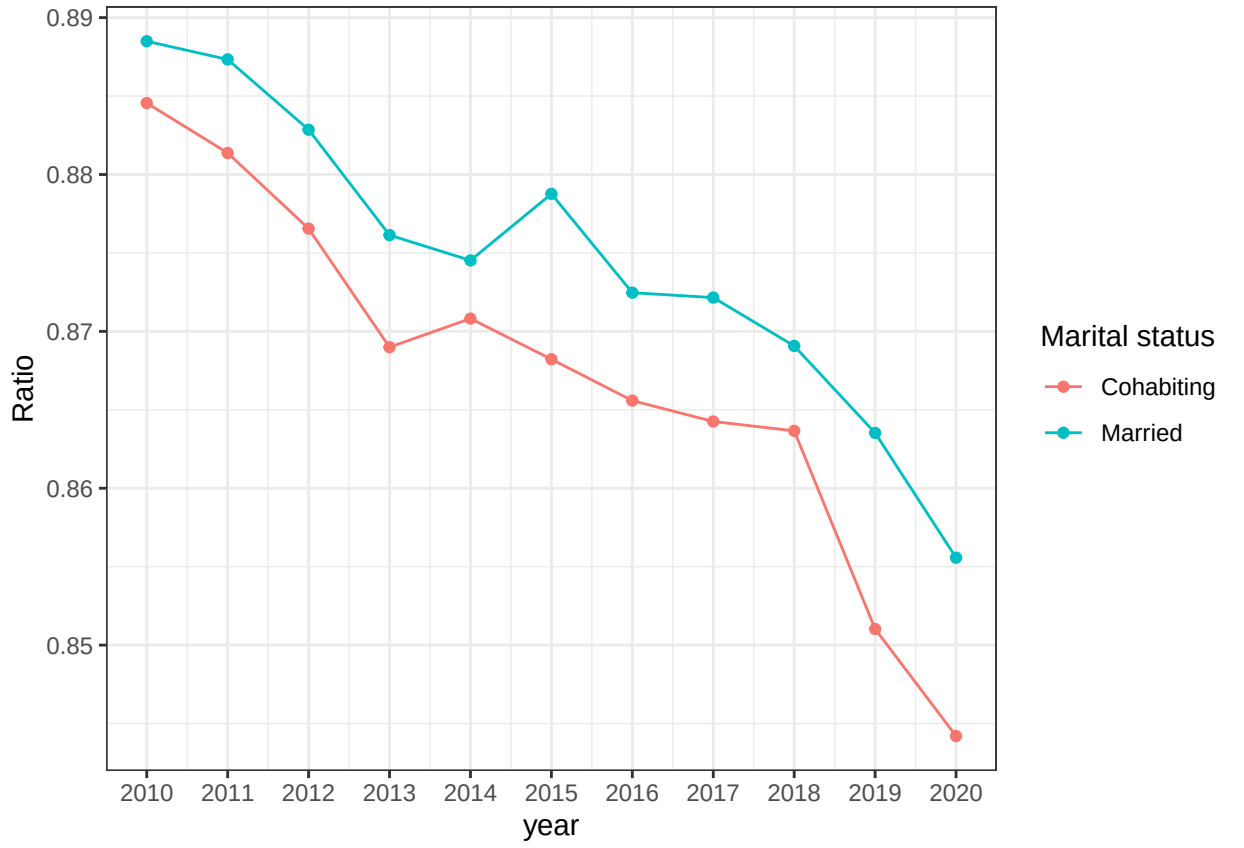
```r
df %>% filter(e_con == 5) %>%
  group_by(year) %>%
  summarise(Ratio = mean(hhChores_W/(hhChores_M + hhChores_W), na.rm = T))
```

```
## # A tibble: 11 x 2
##     year Ratio
##    <dbl> <dbl>
##  1  2010 0.888
##  2  2011 0.887
##  3  2012 0.883
##  4  2013 0.876
##  5  2014 0.875
##  6  2015 0.879
##  7  2016 0.872
##  8  2017 0.872
##  9  2018 0.869
## 10  2019 0.864
## 11  2020 0.856
```

```r
df %>%
  group_by(year, e_con) %>%
  summarise(Ratio = mean(hhChores_W/(hhChores_M + hhChores_W), na.rm = TRUE)) %>%
  mutate(status = case_when(
    e_con == 1 ~ "Cohabiting",
    T ~ "Married"
  )) %>%
  ggplot(aes(x = year, y = Ratio, color = as.factor(status))) +
  geom_point() +
```

```
  geom_line() +
  theme_bw() +
  labs(color = "Marital status") +
  scale_x_continuous(breaks = unique(df$year))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```




```
rm(df, dfM)
```

Importing databases

```
VivEni = read_csv(here("Data/Raw", "viviendas.csv"))
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 73405 Columns: 64
## -- Column specification --------------------------------------------------------
```

```
## Delimiter: ","
## chr  (9): folioviv, tipo_viv, mat_techos, tipo_finan, num_dueno1, num_dueno2...
## dbl (55): mat_pared, mat_pisos, antiguedad, antigua_ne, cocina, cocina_dor, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ConcEni = read_csv(here("Data/Raw", "concentradohogar.csv"))
```

```
## Rows: 74647 Columns: 126
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (5): folioviv, ubica_geo, est_dis, upm, educa_jefe
## dbl (121): foliohog, tam_loc, est_socio, factor, clase_hog, sexo_jefe, edad_...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
PobEni = read_csv(here("Data/Raw", "poblacion.csv"))
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 269206 Columns: 178
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (30): folioviv, numren, madre_id, padre_id, disc1, causa1, lenguaind, n...
## dbl (120): foliohog, parentesco, sexo, edad, madre_hog, padre_hog, disc2, di...
## lgl  (28): disc6, disc7, causa6, causa7, inscr_8, noatenc_1, noatenc_4, noat...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
GasEni = read_csv(here("Data/Raw", "gastoshogar.csv"))
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 4405250 Columns: 27
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (12): folioviv, clave, tipo_gasto, mes_dia, forma_pag1, forma_pag2, form...
## dbl (14): foliohog, frecuencia, cantidad, gasto, pago_mp, costo, inmujer, nu...
## lgl  (1): inst_2
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Item d

```
# Get Home Ownership
VivEni = VivEni %>%
  filter(tenencia != 6) %>%
  mutate(homeownership = ifelse(tenencia %in% c(1,2,5), 0, 1))

# Get daycar, insurances, child care
GasEni = GasEni %>% mutate(health_ins = ifelse(clave %in% c("J070", "J071", "J072"),
                                               gasto_tri, 0),
                           other_ins  = ifelse(clave %in% c("N008", "N009"),
                                              gasto_tri, 0),
                           child_care = ifelse(clave %in% c("E012"),
                                               gasto_tri, 0),
                           child_exp  = ifelse(!is.na(clave),
                                               as.integer(clave == "E012"), NA),
                           day_care = ifelse(clave %in% c("E008"),
                                               gasto_tri, 0),
                           day_exp  = ifelse(!is.na(clave),
                                               as.integer(clave == "E008"), NA)
                           )

GasH = GasEni %>% group_by(folioviv, foliohog) %>%
  summarise(health_ins = sum(health_ins, na.rm = T),
            other_ins  = sum(other_ins, na.rm = T),
            child_care = sum(child_care, na.rm = T),
            child_exp  = sum(child_exp, na.rm = T),
            day_care   = sum(day_care, na.rm = T),
            day_exp    = sum(day_exp, na.rm = T))
```

```
## `summarise()` has grouped output by 'folioviv'. You can override using the
## `.groups` argument.
```

```
PobH = PobEni %>% group_by(folioviv, foliohog) %>%
  summarise(age_hhead = edad[which(parentesco == 101)],
            age_hspouse = first(edad[parentesco %in% c(201,202)]),
            married = edo_conyug[which(parentesco == 101)])
```

```
## `summarise()` has grouped output by 'folioviv'. You can override using the
## `.groups` argument.
```

```
ConcEni = ConcEni %>% left_join(VivEni, by = c("folioviv"))

ConcEni = ConcEni %>% left_join(GasH, by = c("folioviv",
                                             "foliohog"))

ConcEni = ConcEni %>% left_join(PobH, by = c("folioviv",
                                             "foliohog"))

ConcEni = ConcEni %>%
  filter(clase_hog == 2)
```

```r
ConcEni = ConcEni %>% mutate(Income = ing_cor - ingtrab,
                             Food =  ali_dentro + ali_fuera,
                             Transportation = transporte,
                             HealthService = salud,
                             Utilities = pred_cons + agua + energia,
                             Education = educacion,
                             HouseKeeping = cuidados,
                             Rent = ifelse(homeownership == 0, alquiler, estim_alqu),
                             Insurance = health_ins + other_ins,
                             Health_ins = health_ins,
                             Home_ins = other_ins,
                             Childc = day_care + child_care)

vars_to_summarise = c("Income", "Food", "Transportation", "HealthService", "Utilities",
                      "Education", "HouseKeeping", "Rent", "Insurance", "Health_ins",
                      "Home_ins", "Childc")

tabela_resumo = ConcEni %>%
  filter(married %in% c(1, 2)) %>%
  group_by(married) %>%
  summarise(
    across(all_of(vars_to_summarise),
           list(Mean = ~mean(.x, na.rm = TRUE),
                Median = ~median(.x, na.rm = TRUE)))
  )

tabela_resumo
```

```
## # A tibble: 2 x 25
##   married Income_Mean Income_Median Food_Mean Food_Median Transportation_Mean
##     <dbl>       <dbl>         <dbl>     <dbl>       <dbl>               <dbl>
## 1       1       9064.         4972.    10018.       8653.               5011.
## 2       2      16970.         7415.    11048.       9347.               6912.
## # i 19 more variables: Transportation_Median <dbl>, HealthService_Mean <dbl>,
## #   HealthService_Median <dbl>, Utilities_Mean <dbl>, Utilities_Median <dbl>,
## #   Education_Mean <dbl>, Education_Median <dbl>, HouseKeeping_Mean <dbl>,
## #   HouseKeeping_Median <dbl>, Rent_Mean <dbl>, Rent_Median <dbl>,
## #   Insurance_Mean <dbl>, Insurance_Median <dbl>, Health_ins_Mean <dbl>,
## #   Health_ins_Median <dbl>, Home_ins_Mean <dbl>, Home_ins_Median <dbl>,
## #   Childc_Mean <dbl>, Childc_Median <dbl>
```

Question 2

Question 3

Question 4

Item a

Interior solution

Considering positive work in both period 1 and 2, our Lagrangian is

$$\mathcal{L} = \sum_{t=1}^{3} \beta^{t-1}[\alpha \ln C_t + (1-\alpha)\ln L_t] - \lambda \left[\sum_{t=1}^{3} \frac{C_t}{(1+r)^{t-1}} - A_1 - \sum_{t=1}^{2} \frac{w_t(L_0 - L_t)}{(1+r)^{t-1}}\right]$$

with CPO's:

$$[C_t]: \frac{\alpha\beta^{t-1}}{C_t} - \frac{\lambda}{(1+r)^{t-1}} = 0 \Longrightarrow C_t^* = \frac{\alpha[\beta(1+r)]^{t-1}}{\lambda}$$
$$[L_t]: \frac{(1-\alpha)\beta^{t-1}}{L_t} - \frac{\lambda w_t}{(1+r)^{t-1}} = 0 \Longrightarrow L_t^* = \frac{(1-\alpha)[\beta(1+r)]^{t-1}}{\lambda w_t}$$

Then, the Frischian labor supply function is simple

$$h_t^F(w_t, \lambda, t) = L_0 - L_t^* = L_0 - \frac{(1-\alpha)[\beta(1+r)]^{t-1}}{\lambda w_t}$$

Recall that the Marshallian elasticity accounts for variations in labor supply due to permanent shocks in the path of wage. So, to find the Marshallian labor supply function, we need to substitute C_t^* and L_t^* into the budget constraint and isolate λ to find a close expression for the Lagrangean multiplier, i. e.,

$$\sum_{t=1}^{3} \frac{C_t^*}{(1+r)^{t-1}} - A_1 - \sum_{t=1}^{2} \frac{w_t(L_0 - L_t^*)}{(1+r)^{t-1}} = 0$$
$$\sum_{t=1}^{3} \frac{1}{(1+r)^{t-1}} \cdot \frac{\alpha[\beta(1+r)]^{t-1}}{\lambda} - A_1 - \sum_{t=1}^{2} \left[\frac{w_t L_0}{(1+r)^{t-1}} - \frac{w_t}{(1+r)^{t-1}} \cdot \frac{(1-\alpha)[\beta(1+r)]^{t-1}}{\lambda w_t}\right] = 0$$
$$\sum_{t=1}^{3} \frac{\alpha\beta^{t-1}}{\lambda} - A_1 - \sum_{t=1}^{2} \left[\frac{w_t L_0}{(1+r)^{t-1}} - \frac{(1-\alpha)\beta^{t-1}}{\lambda}\right] = 0$$
$$\frac{\alpha}{\lambda}\sum_{t=1}^{3} \beta^{t-1} - A_1 - L_0 \sum_{t=1}^{2} \frac{w_t}{(1+r)^{t-1}} + \frac{(1-\alpha)}{\lambda}\sum_{t=1}^{2} \beta^{t-1} = 0$$

As we know

$$\sum_{t=1}^{T} \beta^{t-1} = \frac{1-\beta^T}{1-\beta}$$

So, we have

$$A_1 + L_0 \sum_{t=1}^{2} \frac{w_t}{(1+r)^{t-1}} = \frac{1}{\lambda(1-\beta)}[\alpha(1-\beta^3) + (1-\alpha)(1-\beta^2)]$$
$$\lambda \left[A_1 + L_0 \left(w_1 + \frac{w_2}{(1+r)} \right) \right] = \frac{1-\alpha\beta^3 + (1-\alpha)\beta^2}{(1-\beta)}$$
$$\lambda \left[\frac{(1+r)(A_1 + L_0 w_1) + w_2}{(1+r)} \right] = \frac{1-\alpha\beta^3 + (1-\alpha)\beta^2}{(1-\beta)}$$
$$\lambda^* = \frac{(1+r)[1-\alpha\beta^3 + (1-\alpha)\beta^2]}{(1-\beta)[(1+r)(A_1 + L_0 w_1) + w_2]}$$

Therefore, the Marshallian labor supply function is:

$$h_t^M(w_t, \lambda^*, t) = L_0 - \frac{(1-\alpha)[\beta(1+r)]^{t-1}}{\lambda^* w_t}$$
$$h_t^M(w_t, \lambda^*, t) = L_0 - \frac{\gamma(1-\alpha)[\beta(1+r)]^{t-1}}{w_t}$$

where

$$\gamma = \frac{(1-\beta)[(1+r)(A_1 + L_0 w_1) + w_2]}{(1+r)[1-\alpha\beta^3 + (1-\alpha)\beta^2]}$$

Corner solution

We have two cases to consider here:

1. **Work only in period 1 ($L_2 = L_0$)**

In this case, the budget constraint will be

$$

$$