

EC306 Assignment 1: Child Care and Mothers' Labour Supply

Answer Key

Prof. Justin Smith

2026-10-21

Background

Since 1997, Quebec has offered regulated child care at a low fixed daily fee: originally \$5 a day, and still under \$10 a day in 2019. Until 2021, parents in the rest of Canada mostly paid market prices that were many times higher. In 2021 the federal government began signing agreements with the other provinces and territories to bring average fees down to \$10 a day by 2026. Because the rest of Canada has been catching up since then, this assignment uses data from 2019.

In the labour supply model, child care is a cost of working: a parent who takes a job has to pay for care during the hours they are away. Cheaper child care raises the net return to working, so we would expect it to draw parents of young children, and mothers especially, into the labour force. In this assignment you will look at whether the data are consistent with that prediction, and then think about whether the comparison you make can be read as causal.

The data

The data come from the Labour Force Survey public use microdata file for March 2019. The sample is everyone aged 25 to 54.

Variable	Description
id	Person identifier
province	Province of residence
gender	"Women+" or "Men+"
agegroup	Five-year age group
educ	Highest level of education

Variable	Description
<code>lfsstat</code>	Labour force status: “Employed, at work”, “Employed, absent from work”, “Unemployed”, or “Not in labour force”
<code>kidage</code>	Age of youngest child living with the person: “Under 6”, “6 to 12”, “13 to 17”, “18 to 24”, or “None under 25”
<code>hours</code>	Usual weekly hours at all jobs (missing if not employed)

Instructions

- Write your R code in the chunks provided. You can add more chunks if you need them.
- When a question tells you to store a result under a particular name, such as `q2` or `gap_u6`, use exactly that name. Those objects are marked by computer, and a different name counts as missing.
- Do not round the numbers you store. Round only when you display them.
- Tables and graphs must be presentable: use `modelsummary()`, `datasummary()` or `datasummary_df()` for tables, and give every graph readable axis labels. Do not show a table by just printing a data frame.
- The hints under each question point to sections of the R tutorial, the Introduction to R and Quarto page on the course website. Each section has a worked example of the step the hint describes, using different data.
- Render the document after each question, not only at the end.
- Every written answer goes under **Answer:**, and what it must cover is set out just above, under **Comment on the results**.
- Each question is marked out of 5 from your PDF, using the rubric on the assignment page. The objects you are told to create are also checked by computer, for another 5 marks, so the assignment is out of 30.

Setup

Do not change this chunk or the next one.

```
library(tidyverse)
library(modelsummary)
```

```
lfs <- read_csv("ec306assign1.csv")

# Put the categories of kidage in their natural order, so tables and
# graphs list them from youngest to oldest instead of alphabetically.
lfs <- lfs |>
  mutate(kidage = factor(kidage, levels = c("Under 6", "6 to 12", "13 to 17",
```

```
"18 to 24", "None under 25"))))
```

```
glimpse(lfs)
```

```
Rows: 46,808
```

```
Columns: 8
```

```
$ id      <dbl> 1, 2, 3, 4, 6, 9, 10, 11, 12, 14, 16, 20, 22, 23, 26, 27, 29, ~
$ province <chr> "Quebec", "Ontario", "Ontario", "Quebec", "Quebec", "Quebec", ~
$ gender   <chr> "Women+", "Men+", "Women+", "Women+", "Women+", "Women+", "Me~
$ agegroup <chr> "40 to 44 years", "30 to 34 years", "30 to 34 years", "45 to ~
$ educ     <chr> "High school graduate", "Bachelor's degree", "Postsecondary c~
$ lfsstat   <chr> "Employed, at work", "Employed, at work", "Employed, absent f~
$ kidage    <fct> None under 25, 6 to 12, Under 6, Under 6, 6 to 12, None under~
$ hours     <dbl> 40.0, 35.0, 40.0, 35.0, 32.0, 26.0, 38.0, 40.0, 50.0, 40.0, 4~
```

Question 1 (5 marks)

Create the following dummy variables and add them to `lfs`:

- `lfp`: 1 if the person is in the labour force (employed and at work, employed but absent from work, or unemployed), 0 otherwise
- `woman`: 1 if `gender` is “Women+”, 0 otherwise
- `quebec`: 1 if `province` is “Quebec”, 0 otherwise
- `kid_u6`: 1 if the person’s youngest child is under 6 (`kidage` is “Under 6”), 0 otherwise

Then make a table showing the mean, standard deviation, minimum, and maximum of the four dummy variables.

Hints.

- Create the dummies with `mutate()`, and save the result back into `lfs`; otherwise the new variables are not kept. A comparison such as `province == "Ontario"` gives `TRUE` or `FALSE` for each person, and wrapping it in `as.numeric()` turns that into 1 or 0. See “Dummy variables” in the R tutorial.
- Three of the four categories of `lfsstat` count as being in the labour force. You can join comparisons with `|` (or), or use `!=` (is not equal to) on the one category that does not count. See “Dummy variables” and “`filter()` — keep some rows” in the R tutorial.
- Text has to match exactly, capitals and punctuation included. Run `count(lfs, lfsstat)` to see the exact spellings.
- For the table, use `datasummary()` with the statistics `Mean`, `SD`, `Min` and `Max`. See “Tables” in the R tutorial.

```
lfs <- lfs |>
  mutate(
    lfp = as.numeric(lfsstat != "Not in labour force"),
    woman = as.numeric(gender == "Women+"),
    quebec = as.numeric(province == "Quebec"),
    kid_u6 = as.numeric(kidage == "Under 6")
  )
```

```
datasummary(
  (`In labour force` = lfp) + (`Woman` = woman) + (`Lives in Quebec` = quebec) +
  (`Youngest child under 6` = kid_u6) ~ Mean + SD + Min + Max,
  data = lfs,
  fmt = 3
)
```

	Mean	SD	Min	Max
In labour force	0.868	0.338	0.000	1.000
Woman	0.510	0.500	0.000	1.000
Lives in Quebec	0.169	0.374	0.000	1.000
Youngest child under 6	0.204	0.403	0.000	1.000

Comment on the results. What does the table tell you about the people in the sample? Before you write, review how to interpret the mean of a dummy variable.

Answer:

The mean of a dummy variable is the share of the sample for whom it equals 1. Among Canadians aged 25 to 54, 86.8% are in the labour force. The sample is almost evenly split by gender (51.0% women). 16.9% of the sample lives in Quebec, somewhat below Quebec's share of the population, because the LFS oversamples the smaller provinces and these figures are unweighted. 20.4% have a youngest child under 6. The minimum of 0 and maximum of 1 confirm that each variable is a dummy, and for a dummy the standard deviation is determined by the mean, $\sqrt{p(1-p)}$, so it carries no extra information.

Question 2 (5 marks)

Keep only women. For each combination of `kidage` and `quebec`, calculate the labour force participation rate. That gives 10 rates: five child age groups, in Quebec and in the rest of Canada. Store the result in a data frame called `q2`, with the rates in a column called `lfp_rate`.

Then make a bar chart of the participation rates, with the age of the youngest child on the horizontal axis and the Quebec and rest-of-Canada bars side by side. You do not need a table for this question: the graph presents the rates.

Hints.

- Do not overwrite `lfs` with the women-only data, because later questions need everyone. Use `filter()` to save the women in a new data frame called `women`; you will use it again in Questions 3 and 4. See “`filter()` — keep some rows” in the R tutorial.
- The rates come from `group_by()` followed by `summarise()`, with both variables in `group_by()`. See “Grouping by two variables” in the R tutorial.
- Give `q2` to `ggplot()`, not `women`. `geom_col()` draws the bars, `position = "dodge"` puts them side by side, and mapping `factor(quebec)` to `fill` colours them by province. Replace the 0 and 1 in the legend with words. See “Bar charts of results you have calculated” in the R tutorial.

```
women <- lfs |> filter(woman == 1)

q2 <- women |>
  group_by(kidage, quebec) |>
  summarise(lfp_rate = mean(lfp), n = n())
```

```
ggplot(q2, aes(x = kidage, y = lfp_rate, fill = factor(quebec))) +
  geom_col(position = "dodge") +
  scale_fill_discrete(labels = c("Rest of Canada", "Quebec")) +
  scale_y_continuous(labels = scales::percent) +
  coord_cartesian(ylim = c(0.5, 1)) +
  labs(x = "Age of youngest child", y = "Participation rate", fill = NULL)
```

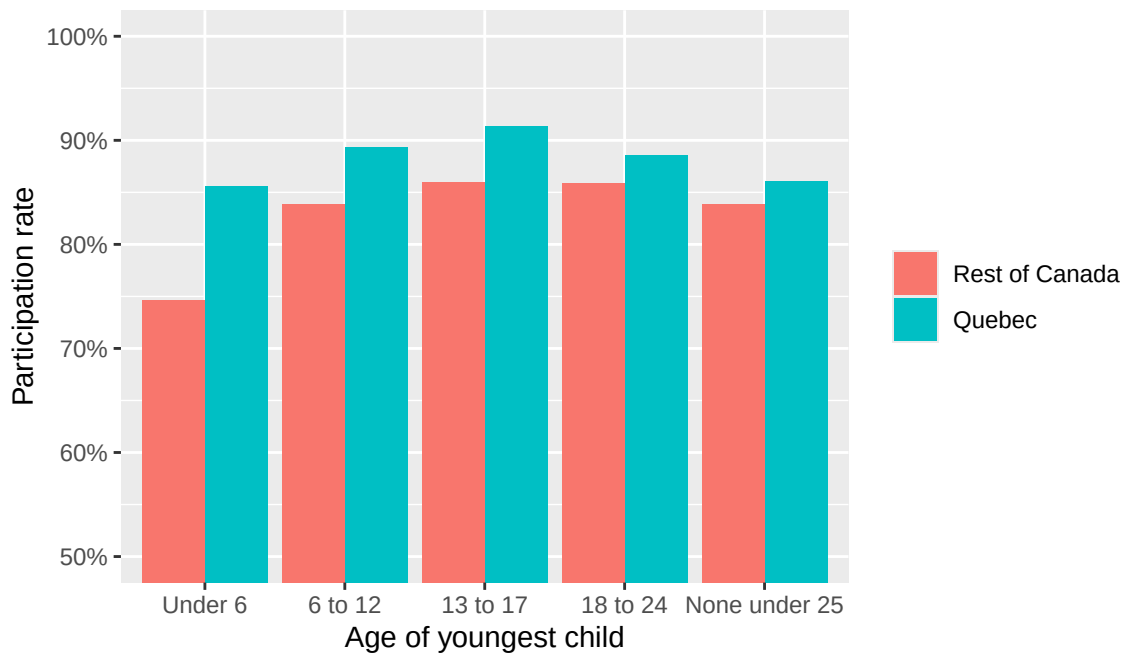


Figure 1: Labour force participation of women aged 25 to 54, by age of youngest child

Comment on the results. What does the graph show? Describe how participation changes as the youngest child gets older, and how the gap between Quebec and the rest of Canada changes.

Answer:

In the rest of Canada, participation is lowest for women whose youngest child is under 6 (74.6%). It rises once the youngest child reaches school age, to 83.9% for a youngest child aged 6 to 12 and 86.0% for 13 to 17. In Quebec the dip for mothers of young children almost disappears: their participation rate is 85.6%, about the same as for Quebec women with no children under 25.

Quebec women participate more than women elsewhere in every group, but the gap is largest, at about 11 percentage points, for women whose youngest child is under 6, which is the group for whom the cost of child care matters most. For women with no children under 25, the gap is only about 2 points. This pattern is what the labour supply model predicts if cheap child care lowers the cost of working. (Note the vertical axis starts at 50% so that the differences are visible.)

Question 3 (5 marks)

Participation is only one way to adjust labour supply. The other is how many hours to work. Keep only women who are employed (at work or absent from work) and whose youngest child is under 6. Make a box plot of usual weekly hours (`hours`) for Quebec and for the rest of Canada.

Hints.

- Start from `women`, and save the filtered data as a new data frame to give to `ggplot()`. Conditions in `filter()` separated by commas must all hold, so you can write the condition on employment and the condition on the child separately. Two categories of `lfsstat` count as employed. See “`filter()` — keep some rows” in the R tutorial.
- Put `factor(quebec)` on the horizontal axis and replace the 0 and 1 on the axis with words. Adding `outliers = FALSE` to `geom_boxplot()` leaves out the individual points beyond the whiskers, which otherwise crowd the graph. See the boxplot of men’s and women’s wages under “Graphs” in the R tutorial.

```
q3_data <- women |>
  filter(kid_u6 == 1,
         lfsstat %in% c("Employed, at work", "Employed, absent from work"))

ggplot(q3_data, aes(x = factor(quebec), y = hours, fill = factor(quebec))) +
  geom_boxplot(outliers = FALSE, width = 0.5, show.legend = FALSE) +
  scale_x_discrete(labels = c("Rest of Canada", "Quebec")) +
  labs(x = NULL, y = "Usual weekly hours, all jobs")
```

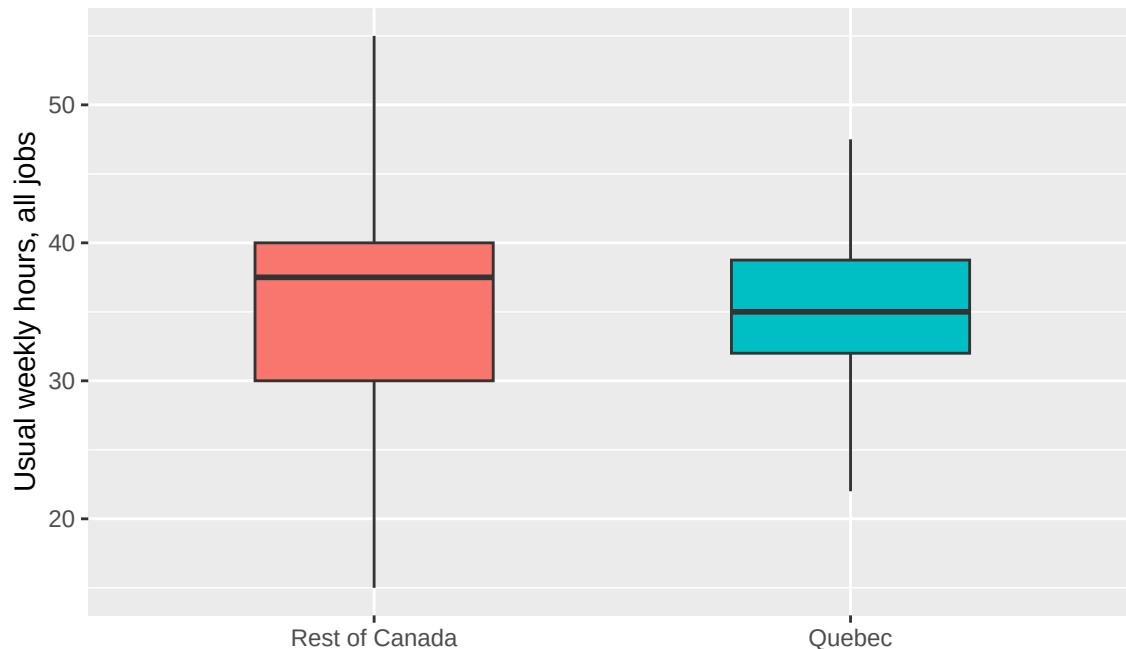


Figure 2: Usual weekly hours of employed women whose youngest child is under 6

Comment on the results. Do employed mothers of young children in Quebec work more hours than those in the rest of Canada? How does the difference compare with the difference in participation you found in Question 2?

Answer:

Hours are similar. The median employed mother of a young child usually works 35 hours a week in Quebec and 37.5 in the rest of Canada, while the means are about an hour apart the other way (34.8 in Quebec and 33.8 elsewhere). The Quebec distribution is more compressed: its 25th percentile is higher (32 hours against 30) and its 75th percentile lower (38.8 against 40). Fewer Quebec mothers work short part-time hours, and fewer work long hours.

Compared with the participation gap of about 11 percentage points, the difference in hours is small. The Quebec difference is mainly at the extensive margin (whether mothers work at all) rather than the intensive margin (how much they work once employed). That fits the idea that child care costs act largely like a fixed cost of working, which mainly affects the decision to participate.

Question 4 (5 marks)

The Quebec gap for mothers of young children in Question 2 may not all be due to child care. Quebec could differ from the rest of Canada in other ways, such as its labour market

or attitudes toward women working, that raise participation for *all* Quebec women. Women whose youngest child is not under 6 rarely need daycare, so the Quebec gap for them estimates that general difference. Subtracting it from the gap for mothers of young children gives a **difference-in-differences** estimate of the effect of child care. For example, with made-up numbers:

Participation rate	Quebec	Rest of Canada	Quebec minus rest of Canada
Youngest child under 6	80%	70%	10 points
All other women	85%	81%	4 points
Difference in differences			6 points

(a) Using women only, calculate the four participation rates by `quebec` and `kid_u6`, the same way as in Question 2, and present them in a table. Then calculate the following, and store each as a single number without rounding:

- `gap_u6`: the Quebec minus rest-of-Canada participation gap for women whose youngest child is under 6
- `gap_other`: the same gap for all other women (`kid_u6 == 0`)
- `did`: `gap_u6 - gap_other`

Hints.

- `datasummary_df()` turns a data frame of results, such as your four rates, into a presentable table. See “Tables” in the R tutorial.
- For `gap_u6`, `gap_other` and `did`, store each of the four rates as its own object first, under a name of your choosing, then subtract.
- Each rate is the mean of `lfp` for one group of women. Square brackets pick out a group: `women$lfp[women$quebec == 1]` is `lfp` for every Quebec woman, and `&` adds a second condition. See “Picking out elements” and “Functions” in the R tutorial.

```
q4_table <- women |>
  group_by(quebec, kid_u6) |>
  summarise(lfp_rate = mean(lfp))

datasummary_df(q4_table, fmt = 3)
```

quebec	kid_u6	lfp_rate
0.000	0.000	0.843
0.000	1.000	0.746
1.000	0.000	0.878
1.000	1.000	0.856

```

qc_u6      <- mean(women$lfpr[women$quebec == 1 & women$kid_u6 == 1])
roc_u6      <- mean(women$lfpr[women$quebec == 0 & women$kid_u6 == 1])
qc_other    <- mean(women$lfpr[women$quebec == 1 & women$kid_u6 == 0])
roc_other    <- mean(women$lfpr[women$quebec == 0 & women$kid_u6 == 0])

gap_u6      <- qc_u6 - roc_u6
gap_other    <- qc_other - roc_other
did          <- gap_u6 - gap_other

round(c(gap_u6 = gap_u6, gap_other = gap_other, did = did), 3)

```

gap_u6	gap_other	did
0.110	0.035	0.075

Comment on the results. Report the three numbers in percentage points (0.05 is 5 percentage points) and explain what each one means. What has to be true for `did` to measure the effect of Quebec's child care policy?

Answer:

- `gap_u6` = 11.0 percentage points: among women whose youngest child is under 6, the participation rate is 11.0 points higher in Quebec than in the rest of Canada (85.6% against 74.6%).
- `gap_other` = 3.5 percentage points: among all other women, who rarely need daycare, participation is 3.5 points higher in Quebec. This is the general Quebec difference.
- `did` = 7.5 percentage points: the extra Quebec advantage that is specific to mothers of young children.

So of the 11.0-point Quebec gap for mothers of young children, about 3.5 points reflect the general Quebec difference that shows up among all women, and about 7.5 points are specific to mothers of young children. That 7.5 points is the difference-in-differences estimate of the effect of Quebec's child care policy.

For `did` to measure the effect of the policy, the general Quebec difference has to be the same for both groups: without cheap child care, Quebec mothers of young children would have had the same 3.5-point advantage as other Quebec women, no more and no less. If something else sets Quebec mothers of young children apart, such as more generous parental leave, `did` picks that up too.

(b) Using women only, estimate the regression

$$lfpr_i = \beta_0 + \beta_1 quebec_i + \beta_2 kid_u6_i + \beta_3 (quebec_i \times kid_u6_i) + \epsilon_i$$

and store it as `q4_model`. Present the results in a table with `modelsummary()`.

Hint. Estimate it with `lm()`, using `women` as the data. In an R formula, `quebec * kid_u6` includes both variables and their product. See “Interactions” in the R tutorial.

```
q4_model <- lm(lfp ~ quebec * kid_u6, data = women)

modelsummary(
  list("Women" = q4_model),
  coef_rename = c("(Intercept)" = "Constant", "quebec" = "Quebec",
    "kid_u6" = "Youngest child under 6",
    "quebec:kid_u6" = "Quebec × child under 6"),
  gof_map = c("nobs", "r.squared"),
  stars = TRUE,
  fmt = 3
)
```

	Women
Constant	0.843*** (0.003)
Quebec	0.035*** (0.007)
Youngest child under 6	−0.097*** (0.007)
Quebec × child under 6	0.075*** (0.015)
Num.Obs.	23 849
R2	0.012
+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001	

Comment on the results. Interpret each of the four coefficients. How does β_3 compare with `did` from part (a)?

Answer:

- $\beta_0 = 0.843$: the participation rate of women in the rest of Canada whose youngest child is not under 6.

- $\beta_1 = 0.035$: among women whose youngest child is not under 6, Quebec women's participation is 3.5 points higher. This equals `gap_other`.
- $\beta_2 = -0.097$: in the rest of Canada, women whose youngest child is under 6 participate 9.7 points less than other women.
- $\beta_3 = 0.075$: the Quebec gap is 7.5 points larger for women whose youngest child is under 6 than for other women.

β_3 is exactly equal to `did`. With two dummies and their interaction, the regression has one coefficient for each of the four groups, so it reproduces the four group means exactly, and the interaction is the difference in differences. What the regression adds is a standard error: β_3 is statistically significant at the 1% level.

Question 5 (5 marks)

(a) Estimate the same regression for men, and store it as `q5_model`. Present both regressions side by side in one table, with the columns labelled “Women” and “Men”. Part (a) needs no written answer; you will use the results in part (b).

Hints.

- You need a men-only data frame, made the same way as `women`.
- `modelsummary()` accepts a list of models, and the names you give the elements of the list become the column headings. See “Presenting results” in the R tutorial.

```
men <- lfs |> filter(woman == 0)
q5_model <- lm(lfp ~ quebec * kid_u6, data = men)

modelsummary(
  list("Women" = q4_model, "Men" = q5_model),
  coef_rename = c("(Intercept)" = "Constant", "quebec" = "Quebec",
    "kid_u6" = "Youngest child under 6",
    "quebec:kid_u6" = "Quebec × child under 6"),
  gof_map = c("nobs", "r.squared"),
  stars = TRUE,
  fmt = 3
)
```

	Women	Men
Constant	0.843*** (0.003)	0.896*** (0.002)
Quebec	0.035*** (0.007)	−0.002 (0.006)
Youngest child under 6	−0.097*** (0.007)	0.051*** (0.005)
Quebec × child under 6	0.075*** (0.015)	0.013 (0.013)
Num.Obs.	23 849	22 959
R2	0.012	0.005

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

(b) Comment on the results. Is did a causal estimate of the effect of Quebec’s child care policy on mothers’ participation? In your answer, consider:

- what the result for men tells you;
- other ways Quebec differs from the rest of Canada that could affect parents of young children. One example: the LFS counts a parent on parental leave, with a job to return to, as “employed, absent from work”;
- in 2021 the other provinces began cutting their child care fees toward \$10 a day. If you repeated this analysis with data from 2025, what would you expect to happen to did, and why?

Answer:

It is suggestive, but not conclusive. The key assumption is that without the policy, the Quebec gap for mothers of young children would have matched the gap for other women. Several things could make that fail.

Men. For men the interaction is 0.013, small and not statistically significant. Whatever raises participation among Quebec mothers of young children does not show up among Quebec fathers of young children. That supports reading the women’s result as something specific to mothers, which is where child care costs are expected to bite, rather than something that affects all Quebec parents. Men are an imperfect comparison, though: almost all fathers of young children are in the labour force (note the positive coefficient on the child dummy), so there is little room for any policy to raise their participation.

Other differences between Quebec and the rest of Canada. Quebec mothers of young children may differ from mothers elsewhere in age, education, or when they have children, and Quebec

has other family policies. The most relevant is parental leave. Quebec runs its own, more generous parental insurance plan, and because the LFS counts a parent on leave with a job to return to as employed, more generous leave could raise measured participation of mothers of infants without any role for child care. In these data it can explain only a small part of the gap: 19.9% of Quebec mothers of young children are employed but absent from work, against 17.2% elsewhere. That difference of about 3 points is small next to the participation gap; most of the Quebec advantage comes from more mothers being at work.

Fees elsewhere, then and now. In March 2019 the rest of Canada had not yet cut fees broadly, although BC and Alberta had started smaller fee-reduction programs in 2017 and 2018, so some families outside Quebec already had cheaper care. After 2021 the other provinces cut fees toward \$10 a day, so data from 2025 would compare Quebec with provinces that are partly treated themselves. If cheaper care raises mothers' participation, mothers elsewhere should have gained ground, and *did* should shrink. It does: the same calculation with the March 2025 LFS gives about 5.9 points, against 7.5 here. Other things changed between 2019 and 2025, the pandemic above all, so that is suggestive rather than proof. A convincing causal estimate would compare mothers before and after a policy change, like the studies of Quebec's 1997 reform, rather than across provinces at a single point in time.

Use of generative AI

If you used a generative AI tool for any part of this assignment, name the tool and describe what you used it for. If you did not use one, say so.

Not applicable (answer key).

Before you submit

This chunk lists any required objects that are still missing. It should say "All required objects found" before you submit.

All required objects found