Evolution of gender differences in the Canadian labor market in the 2010s

# Introduction

The main objective of this project is to explore differences by gender in the Canadian labor market with a focus on wages and participation by sector, industry, and job occupation at the beginning and ending of the 2010s decade. The specific research questions to solve are:

* What variables were relevant to explain hourly wages by gender when the decade started?
  + After 10 years, are these variables still important?
  + Are they the same for women and for men?
* What were the sectors, industries, and job occupations with highest hourly wages by gender at the beginning of the decade?
  + Are these the same by the end of 2019?
  + Which of these groups show more equality in wages by gender?
* What is the gender distribution by sector, industry, and job occupation?
  + What groups are more uneven?
  + Was there a change in the distribution over time?

The data source will be the “Labor Force Survey” records of September 2009 and September 2019 published by the government agency Statistics Canada. The analysis will require the application of statistical techniques such as multiple regression, and tests of hypothesis to compare means and proportions.

# Literature Review

Differences in the labor market by gender in Canada and internationally have been studied extensively over the last decades. One of the more discussed topics within this subject is the gender pay gap, which is defined as the difference between women’s and men’s earnings from paid employment, expressed as a proportion of men’s earnings (Statistics Canada, 2019). Other variables that have been analyzed by researchers to describe differences by gender found in the labor market are educational attainment, marital status and motherhood, hours worked, job tenure, and participation by industry and occupations.

Regarding the gender pay gap, Statistics Canada (2019) states that there are several methods to calculate it. One of them consists of comparing the annual earnings of all employed women and men, which captures gender differences both in pay and in hours and weeks worked. In other words, it takes into consideration not only the net difference in wage, but also the impact in women’s earnings of having less disposable time to work due to home and family duties that have historically been covered by them. Measured this way, employed women aged 16 and older earned an average of $0.69 for every dollar earned by men in 2017.

Another way to estimate the gender pay gap consists of calculating the average hourly wages. It only captures the per unit (hour) price of labor, so it produces a smaller estimate than the method of comparing annual earnings. This approach is more appropriate to explore pay equity issues, that is, equal pay for work of equal or comparable value to the employer. In Canada, employed women aged 15 and older earned an average of $0.87 for every dollar earned by employed men on an hourly basis.

Moyser (2017) analyzed the relationship between wage inequality and occupations, and found that men out‑earned women in every occupational group, except for managers and professionals in art, culture, recreation and sport (i.e., librarians, archivists, conservators and curators; writing, translating and related communications professionals; and creative and performing artists) and middle management occupations in production. Even in traditionally‑female occupations like teaching, nursing, clerical, sales and services, the average hourly wages earned by men were greater than those earned by women. If women earned the same amount as men within occupations, the gender wage ratio would nearly reach equality at 0.97.

Motherhood has a persistent impact in wage disparity, although it has lessened in the last years. According to Statistics Canada (2017), in 1997, mothers with at least one child under the age of 18 earned $0.79 for every dollar earned by fathers, while women without children earned $0.88 for every dollar earned by men without children. By 2015, mothers earned $0.85 for every dollar earned by fathers, and women without children earned $0.90 for every dollar earned by men without children.

Drolet and Mumford (2012), with data from Statistics Canada Workplace and Employee Survey (WES) for private sector employees, discuss the relevance of age, fields of study and type of workplace on gender pay differentials.

Regarding age, they mention that is often assumed to be related to wages for several reasons, not least of which is the ability to acquire skills over time. Nevertheless, the results need to be interpreted with caution if women are more likely than men to have taken time out of the labor market; women's age may not accurately reflect the relationship between their actual work experience and earnings. At the same time, working part‐time may be associated with less accumulated work experience. Of note, it has been identified that gender wage gap is seen to be smallest among young workers (those aged below 40 years) with little significant difference

About fields of study, many continue to be dominated by either men or women. While there have been notable changes in some discipline areas (such as biology, medicine, management and law), men and women tend to choose traditional disciplines, and this choice may also explain part of the gender differences in earnings.

Concerning the type of workplace, women are more commonly employed in low‐paying workplaces than men. These results provide evidence that part of the variation in individual earnings attributable to, for example, differing levels of education may arise because high‐wage workplaces disproportionately employ high‐skilled workers.

Aside from gender pay gap, there are other aspects that show discrepancies and disparity between men and women at the workplace. Moyser (2017), mentions that over the last decades, it has become more likely for women to work. In 1950, only 22% of women in the core working ages of 25 to 54 years participated in the labour market, whereas in 2015 this number reached 82%. However, though women participation in the workforce have largely increased, their experiences of paid work still tend to differ from those of men, being shaped to a greater extent by their caregiving roles and/or their employers’ presumptions of these roles. Based on Statistics Canada Labor Force Survey (LFS) data of 2015, she describes these differences from the optics of several variables:

* *Educational attainment.* The more educated, the more likely is that a woman works, though this likelihood is higher for a man with the same level of education. For example, in 2015, the employment rate of women with a high school diploma was 69% compared 83% for those with a university degree. The employment rate of men with a high school diploma was 82% compared to 90% for those with a university degree.
* *Hours worked.* Women generally perform fewer paid hours than men, as they spend more time on housework and childcare. They are also more likely than men to work part-time.
* *Long and short work absences.* Women’s careers are interrupted more frequently than men’s careers and for longer durations. These absences are frequently related to the role of women at home. In 2015, 22% of women who were absent attributed their absence to family or personal responsibilities, compared to only 9.3% of men who also cited those reasons.
* *Job tenure.* Women have nearly equivalent job tenure with current employer as men. Their average was 93.7 months, close to men’s 94.9 months.
* *Self-employment.* Self-employment is less common among women than men, who represented close to two thirds of the self-employed population.
* *Industry.* There are clear differences in men and women participation across industries. Men are majority in the goods-producing sector, which includes industries related to the exploitation of natural resources, utilities, construction, and manufacturing. In contrast, women outnumber men in the services-producing sector, that consists of industries like wholesale and retail trade, finance, insurance, real estate, business, educational services, healthcare and social assistance, accommodation and food services, public administration, among others.
* *Occupation.* Women and men occupy distinct occupations, with women’s typically being at lower levels than men’s, even in industries dominated by women. For example, in the accommodation and food services industry, 59.7% of chefs and cooks were men, while 71.6% of food counter attendants, kitchen helpers and related support personnel were women.
* *Leadership and high-paying positions.* Women are underrepresented in leadership positions in the private sector, although not in the public sector. In 2015, 54% of legislators and senior government managers and officials were women, whereas in the private sector they represented only 26% of senior managers.

## Conclusions

There are many instances where differences by gender in the labor market have been identified. The relationship or possible impact of some variables to wage disparity has also been studied. When analyzing these attributes and their relevance to explain hourly wages by gender, we will see if they have the same weight on each gender results. For example, we will be able to distinguish whether educational attainment, worked hours, job tenure, age, industry, and occupation have the same level of influence on the wages of men and women. Regarding wages disparity calculation, we will focus on pay equity, meaning that the method that compares average hourly wages will be applied. Furthermore, it will be explored whether some of the discrepancies by gender found in the past, have been persistent as of the end of the last decade.

## References

Statistics Canada. 2019. “Measuring and analyzing the gender pay gap: A conceptual and methodological overview”. Studies on Gender and Intersecting Identities. Ottawa: Statistics Canada.

<https://www150.statcan.gc.ca/n1/daily-quotidien/190830/dq190830d-eng.htm>

Moyser, Melissa. 2017. “Women and Paid Work”. Women in Canada: A Gender-based Statistical Report. Ottawa: Statistics Canada. Catalogue no. 89-503-X. ISSN: 1719-4407.

<https://www150.statcan.gc.ca/n1/pub/89-503-x/2015001/article/14694-eng.htm>

Drolet, Marie and Karen Mumford. 2012. "The Gender Pay Gap for Private‐Sector Employees in Canada and Britain." British Journal of Industrial Relations 50 (3): 529-553.

# Dataset

A dataset will be created pulling records from the Public Use Microdata Files (PUMFs) of the Labor Force Survey (LFS) of September of 2009 and September 2019, published by Statistics Canada. The subjects covered in this survey are:

* Employment and unemployment
* Hours of work and work arrangements
* Industries
* Labour
* Occupations
* Unionization and industrial relations
* Wages, salaries and other earnings

The survey is conducted nationwide, and the target population is the non-institutionalized population 15 years of age and over. Excluded groups represent less than 2% of target population: persons living on reserves and other Aboriginal settlements in the provinces, full-time members of the Canadian Armed Forces, Institutionalized population, and households in extremely remote areas with very low population density.

Deployment is performed monthly. The reference period to answer the questions is usually the week containing the 15th day of the month, and the collection period is the week following the reference period.

## Labor Force Survey Sampling

This is a sample survey with a cross-sectional design (snapshot). It uses a probability sample that is based on a stratified multi-stage design. Each province is divided into large geographic stratum. The first stage of sampling consists of selecting smaller geographic areas, called clusters, from within each stratum. The second stage of sampling consists of selecting dwellings from within each selected cluster.

The LFS uses a rotating panel sample design so that selected dwellings remain in the LFS sample for six consecutive months. Each month about one-sixth of the LFS sampled dwellings are in their first month of the survey, one-sixth are in their second month of the survey, and so on. One feature of the LFS sample design is that each of the six rotation groups can be used as a representative sample by itself.

Within selected dwellings, basic demographic information is collected for all household members. Labor force information is collected for all civilian household members who are aged 15 and over.

Recently, the monthly LFS sample size has been approximately 56,000 households, resulting in the collection of labor market information for approximately 100,000 individuals. It should be noted that the LFS sample size is subject to change from time to time to meet data quality or budget requirements.

## Data set for this project

Data sets from September 2009 and September 2019 were selected to represent the beginning and ending of the 2010s decade. Added together in a single table, the data from the two periods sums over 200,000 observations.

### Attributes

#### Differences between LFS 2009 and 2019 variables

The LFS questionnaire is redesigned periodically. Variables out-of-date are removed, new variables can be introduced, and value sets of other variables can also be modified. LFS from September 2009 includes 82 attributes, whereas the one from 2019 consist of only 60 attributes. Here is a list of the attributes that were omitted by 2019, along with the reason of removal or change provided in the LFS documentation:

|  |  |  |
| --- | --- | --- |
| **Attribute 2009** | **Description** | **Reason of Removal** |
| **ED76to89** | Highest educational attainment | Out-of-date, replaced by EDUC90 |
| **NAICS\_18** | Industry of main job, 18 categories | Out-of-date, replaced by NAICS\_21 |
| **NAICS\_43**  **(Not available after 2015)** | Industry of main job, 43 categories | Out-of-date |
| **SOC80\_49**  **(Not available after 2015)** | Occupation at main job, 1980 Standard Occupational Classification, 49 categories | Out-of-date |
| **SOC80\_21**  **(Not available after 2015)** | Occupation at main job, 1980 Standard Occupational Classification, 21 categories | Out-of-date |
| **NOCS\_06\_25**  **(Not available after 2015)** | Occupation at main job, 2001 National Occupational Classification for Statistics, 25 categories | Out-of-date, replaced by NOC\_10 |
| **NOCS\_06\_47**  **(Not available after 2015)** | Occupation at main job, 2001 National Occupational Classification for Statistics, 47 categories | Out-of-date, replaced by NOC\_40 |
| **WHYPTOLD** | Reason for part-time | Out-of-date, replaced by WHYPTNEW |
| **YNOLKOLD** | Reasons for not looking for work in the past four weeks, after looking for work in the last 6 months | Out-of-date, replaced with new variable |
| **RELREFN** | Relationship to reference person | Variables related to family and spouse/partner’s labour force characteristics were removed |
| **EFAMSIZE** | Number of individuals in economic family, 1 to 5+ |
| **EFAMEMPL** | Number of employed persons |
| **EFAMUNEM** | Number of unemployed persons |
| **SP\_AGE** | Age of spouse, if applicable. |
| **SP\_LFSST** | Labour force status of spouse, if applicable. |
| **SPED7689** | Education of spouse, if applicable. |
| **SPED1990** | Education of spouse, if applicable. |
| **SP\_SOC80** | Occupation of spouse, if applicable. |
| **SP\_NOCS06** | Occupation based on NOC-S2006 |
| **SP\_UHRSM** | Spouse's usual hours at main job, employed, |
| **SP\_UHRST** | Spouse's usual hours at all jobs, employed |
| **SP\_COWM** | Spouse's class of worker at main job, employed |
| **SCH1624** | At least one child age 16 to 24 in |

Besides the removals, there were variables which value sets were either expanded or collapsed:

* The **census metropolitan area (CMA)** variable expanded from the three largest CMAs in Canada (Montreal, Toronto and Vancouver) to nine CMAs: Québec, Montreal, Ottawa, Toronto, Hamilton, Winnipeg, Calgary, Edmonton and Vancouver.
* Two **occupation** variables, based on the 2016 National Occupational Classification, were reintroduced: a generic, 10-category variable (NOC\_10) and a detailed 40-category variable (NOC\_40).
* The **labour force status** variable (LFSSTAT) was collapsed from six to four categories: employed, at work; employed, absent from work; unemployed; and not in the labour force.
* Two **industry** variables (NAICS\_18 and NAICS\_43) based on the 2012 North American Industry Classification System were replaced with one industry variable consisting of 21 categories, based on the same classification system.
* The **age of youngest child** variable (AGYOWNK) has been collapsed from six categories to four: under 6 years; 6 to 12 years; 13 to 17 years; and 18 to 24 years.
* The **school attendance** variable (SCHOOLN) has been collapsed from nine categories, which include information on the type of school attended, to three categories indicative of student status only: non-student; full-time student; and part-time student.

Finally, there is one new variable that was introduced in 2017:

* **Immigrant status** (IMMIG) distinguishes between recent immigrants, landed 10 or less years ago; established immigrant, landed more than 10 years ago; and non-immigrant.

#### Adjustments to attributes in project’s data set

As noted, some variables value sets are not the same in 2019 as they were in 2009. The next tables show how those variable values will be matched to be comparable:

**Census Metropolitan Area (CMA)**

|  |  |  |  |
| --- | --- | --- | --- |
| **2009** | | **2019** | |
| **CMA** | | **CMA** | |
| **Code** | **Description** | **Code** | **Description** |
| 1 | Montreal | 2 | Montréal |
| 2 | Toronto | 4 | Toronto |
| 3 | Vancouver | 9 | Vancouver |
| 4 | Other CMA or Non-CMA | 1,3,5,6,7,8,0 | Quebec, Ottawa-Gatineau, Hamilton, Winnipeg, Calgary, Edmonton, Other |

**Labor Force Status (LFSTAT)**

|  |  |  |  |
| --- | --- | --- | --- |
| **2009** | | **2019** | |
| **LFSTAT** | | **LFSTAT** | |
| **Code** | **Description** | **Code** | **Description** |
| 1 | Employed, at work | 1 | Employed, at work |
| 2 | Employed, absent from work | 2 | Employed, absent from work |
| 3, 4, 5 | 3 = Unemployed, temporary layoff 4 = Unemployed, job searcher 5 = Unemployed, future start | 3 | Unemployed |
| 6 | Not in labour force | 4 | Not in labour force |

**Industry (NAICS)**

|  |  |  |  |
| --- | --- | --- | --- |
| **2009** | | **2019** | |
| **NAICS\_18** | | **NAICS\_21** | |
| **Code** | **Description** | **Code** | **Description** |
| 1 | Agriculture | 1 | Agriculture |
| 2 | Forestry, Fishing, Mining, Oil and Gas | 2,3,4 | 2 = Forestry and logging and support activities for forestry 3 = Fishing, hunting and trapping 4 = Mining, quarrying, and oil and gas extraction |
| 3 | Utilities | 5 | Utilities |
| 4 | Construction | 6 | Construction |
| 5 | Manufacturing – durables | 7 | Manufacturing - durable goods |
| 6 | Manufacturing non-durables | 8 | Manufacturing - non-durable goods |
| 7 | Wholesale Trade | 9 | Wholesale trade |
| 8 | Retail Trade | 10 | Retail trade |
| 9 | Transportation and Warehousing | 11 | Transportation and warehousing |
| 10 | Finance, Insurance, Real Estate and Leasing | 12,13 | 12 = Finance and insurance 13 = Real estate and rental and leasing |
| 11 | Professional, Scientific and Technical Services | 14 | Professional, scientific and technical services |
| 12 | Management, Administrative and Other Support | 15 | Business, building and other support services |
| 13 | Educational Services | 16 | Educational services |
| 14 | Health Care and Social Assistance | 17 | Health care and social assistance |
| 15 | Information, Culture and Recreation | 18 | Information, culture and recreation |
| 16 | Accommodation and Food Services | 19 | Accommodation and food services |
| 17 | Other Services | 20 | Other services (except public administration) |
| 18 | Public Administration | 21 | Public administration |

**Occupation (NOC)**

|  |  |  |  |
| --- | --- | --- | --- |
| **2019** | | **2009** | |
| **NOC\_10** | | **NOC\_25** | |
| **Code** | **Description** | **Code** | **Description** |
| 1 | Management occupations | 1, 2 | 1 = Senior Management Occupations 2 = Other Management Occupations |
| 2 | Business, finance and administration occupations | 3, 4, 5 | 3 = Professional Occupations in Business and Finance 4 = Financial, Secretarial and Administrative Occupations 5 = Clerical Occupations, Including Supervisors |
| 3 | Natural and applied sciences and related occupations | 6 | Natural and Applied Sciences and Related Occupations |
| 4 | Health occupations | 7, 8 | 7 = Professional Occupations in Health, Nurse Supervisors and Registered Nurses 8 = Technical, Assisting and Related Occupations in Health |
| 5 | Occupations in education, law and social, community and government services | 9, 10 | 9 = Occupations in Social Science, Government Service and Religion 10 = Teachers and Professors |
| 6 | Occupations in art, culture, recreation and sport | 11 | Occupations in Art, Culture, Recreation and Sport |
| 7 | Sales and service occupations | 12, 13, 14, 15, 16, 17 | 12 = Wholesale, Technical, Insurance, Real Estate Sales Specialists, and Retail, Wholesale and Grain Buyers  13 = Retail Salespersons, Sales Clerks, Cashiers, Including Retail Trade, Supervisors 14 = Chefs and Cooks, and Occupations in Food and Beverage Service, Including Supervisors 15 = Occupation in Protective Services 16 = Childcare and Home Support Workers 17 = Sales and Service Occupations n.e.c. (not elsewhere classified), Including Occupations in Travel and Accommodation, Attendants in Recreation and Sport as well as Supervisors |
| 8 | Trades, transport and equipment operators and related occupations | 18, 19, 20, 21, 22 | 18 = Contractors and Supervisors in Trades and Transportation 19 = Construction Trades 20 = Other Trades Occupations 21 = Transport and Equipment Operators 22 = Trades Helpers, Construction, and Transportation Labourers and Related Occupations |
| 9 | Natural resources, agriculture and related production occupations | 23 | Occupations Unique to Primary Industry |
| 10 | Occupations in manufacturing and utilities | 24, 25 | 24 = Machine Operators and Assemblers in Manufacturing, Including Supervisors 25 = Labourer in Processing, Manufacturing and Utilities |

**Age of youngest child (AGYOWNKN)**

|  |  |  |  |
| --- | --- | --- | --- |
| **2009** | | **2019** | |
| **AGYOWNKN** | | **AGYOWNKN** | |
| **Code** | **Description** | **Code** | **Description** |
| 1, 2 | 1 = Youngest child under 3 2 = Youngest child 3 to 5 | 1 | Youngest child less than 6 years |
| 3 | Youngest child 6 to 12 | 2 | Youngest child 6 to 12 years |
| 4, 5 | 4 = Youngest child 13 to 15 5 = Youngest child 16 to 17 | 3 | Youngest child 13 to 17 years |
| 6 | Not in labour force | 4 | Youngest child 18 to 24 years |

**Current Student Status (SCHOOLN)**

|  |  |  |  |
| --- | --- | --- | --- |
| **2009** | | **2019** | |
| **SCHOOLN** | | **SCHOOLN** | |
| **Code** | **Description** | **Code** | **Description** |
| 1 | Non-student | 1 | Non-student |
| 2, 4, 6, 8 | 2 = Primary or secondary school, full-time 4 = University full-time 6 = Community college or CEGEP full-time 8 = Other full-time | 2 | Full-time student |
| 3, 5, 7, 9 | 3 = Primary or secondary school, part-time 5 = University part-time 7 = Community college or CEGEP part-time 9 = Other part-time | 3 | Part-time student |

#### Variables to be excluded from project’s data set

LFS includes several questions about many aspects of unemployment, nevertheless the focus of this project is on the employed population. Thus, 18 variables related to this topic will be omitted in the data set created for this project.

Also, there are two variables that contain group age data. One of them includes information for all respondents (age\_12), while the other only covers people between 15 and 29 years (age\_6). The latter will be excluded.

Besides, there are 4 variables related to hours worked per week:

* utothrs. Usual hours worked per week at all jobs
* atothrs. Actual hours worked per week at all jobs
* uhrsmain. Usual hours worked per week at main job
* ahrsmain. Actual hours worked per week at main job

Only the first one (utothrs) will be kept in the project’s data set, since is the one that provide the usual hours considering all jobs, and not just actual hours worked on reference week or only at main job.

On the other hand, there are 2 attributes that give information about part-time employees. One of them identifies full time and part time workers (ftptmain), and the other provides reasons for part-time work (whypt). Only the former will be kept, to explore if there is an impact in hourly wages.

There are 3 variables related to overtime worked in the reference week that will be excluded. There is no evidence to confirm whether the number of extra hours worked during this timeframe is recurring or just a one-time event. And wages data focuses on “usual hourly wages”.

Lastly, 5 variables that provide details regarding reasons and time duration of current work absence will be excluded. This information is only available for people who declared being absent from work in the reference week. Furthermore, the impact of being absent from work will be captured by the labor force status variable (lfsstat), which distinguishes employees at work and employees absent from work.

#### Complete list of attributes to be included in project’s data set

The new data set will consist of 27 attributes selected from the original microdata files:

|  |  |  |  |
| --- | --- | --- | --- |
| **VARIABLE** | **DESCRIPTION** | **TYPE** | **COMMENTS** |
| **rec\_num** | Order of record in file | Nominal |  |
| **survyear** | Survey year | Nominal |  |
| **survmnth** | Survey month | Nominal |  |
| **Lfsstat** | Labour force status | Nominal | 2009 values adjusted to match 2019 |
| **Prov** | Province | Nominal |  |
| **Cma** | Three largest CMAs | Nominal | 2019 values adjusted to match 2009 |
| **age\_12** | Five-year age group of respondent | Ordinal |  |
| **Sex** | Sex of respondent | Nominal |  |
| **Marstat** | Marital status of respondent | Nominal |  |
| **Educ** | Highest educational attainment | Ordinal |  |
| **Mjh** | Single or multiple jobholder | Nominal |  |
| **cowmain** | Class of worker, main job | Nominal |  |
| **Immig** | Immigrant status | Nominal | Only 2019 data |
| **naics\_18** | Industry of main job | Nominal | 2019 values adjusted to match 2009 |
| **noc\_10** | Occupation at main job | Nominal | 2009 values adjusted to match 2019 |
| **noc\_40** | Occupation at main job | Nominal | Only 2019 data |
| **Ftptmain** | Full- or part-time status at main or only job | Nominal |  |
| **Utothrs** | Usual hours worked per week at all jobs | Quantitative |  |
| **Tenure** | Job tenure with current employer (months) | Quantitative |  |
| **Hrlyearn** | Usual hourly wages, employees only | Quantitative |  |
| **Union** | Union status, employees only | Nominal |  |
| **permtemp** | Job permanency, employees only | Nominal |  |
| **Estsize** | Establishment size | Ordinal |  |
| **Firmsize** | Firm size | Ordinal |  |
| **School** | Current student status | Nominal | 2009 values adjusted to match 2019 |
| **efamtype** | Type of economic family | Nominal |  |
| **agyownk** | Age of youngest child | Ordinal | 2009 values adjusted to match 2019 |

### Dataset descriptive statistics

The total number of records in LFS September 2009 and 2019 are 107,593 and 100,011, respectively. However, the focus of the research questions posed for this project is on the employed population. Looking at the “Labor force status” variable, it is possible to identify the respondents that declared to be employed:

Hourly wages data for self-employed workers are not shared on the public records of LFS. Also there is a small portion of employees that self-identified as unpaid family worker. Because of the lack of wages data, both groups of workers will be excluded.

The total number of observations of only employed population from the private and public sector are 54,557 for 2009 and 50,931 for 2019, that sum up a global of 105,488 cases. For subsequent analysis, only these records will be included.

#### Missing Values

Here is the list of variables that contain missing values and how they will be handled:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Description | Number of cases with missing data | % of total cases (Out of 105,488 cases) | Solution |
| School | Current student status | 3,002 | 3% | Replace with 4, label “Unknown” |
| Agyownk | Age of youngest child | 65,144 | 62% | Replace with 5, label “Youngest > 24 years or no children” |

Besides, only 2019 records have data for immigrant status (immig) and detailed occupation (noc\_40) attributes. They should not be used in analysis which results will be compared to 2009, but they can provide insights when exploring the 2019 data set independently.

#### Categorical Attributes

The next tables show frequencies and proportions for the main categorical attributes in the dataset, by period:

**Province**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **Province** | **2009** | **2019** | **2009** | **2019** |
| AB | 5,885 | 5,292 | 11% | 10% |
| BC | 5,996 | 6,096 | 11% | 12% |
| MN | 5,166 | 4,166 | 9% | 8% |
| NB | 2,827 | 2,484 | 5% | 5% |
| NL | 1,830 | 1,685 | 3% | 3% |
| NS | 2,632 | 2,565 | 5% | 5% |
| ON | 15,563 | 14,253 | 29% | 28% |
| PEI | 1,447 | 1,424 | 3% | 3% |
| QC | 9,370 | 9,294 | 17% | 18% |
| SK | 3,841 | 3,672 | 7% | 7% |

**Age**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **Age Groups** | **2009** | **2019** | **2009** | **2019** |
| 15 to 19 years | 3,572 | 2,576 | 7% | 5% |
| 20 to 24 years | 5,055 | 4,215 | 9% | 8% |
| 25 to 29 years | 5,649 | 5,288 | 10% | 10% |
| 30 to 34 years | 5,495 | 5,550 | 10% | 11% |
| 35 to 39 years | 5,870 | 5,605 | 11% | 11% |
| 40 to 44 years | 6,305 | 5,477 | 12% | 11% |
| 45 to 49 years | 7,491 | 5,461 | 14% | 11% |
| 50 to 54 years | 6,835 | 5,545 | 13% | 11% |
| 55 to 59 years | 4,784 | 5,450 | 9% | 11% |
| 60 to 64 years | 2,509 | 3,754 | 5% | 7% |
| 65 to 69 years | 736 | 1,384 | 1% | 3% |
| 70 and over | 256 | 626 | 0% | 1% |

**Sex**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **Sex** | **2009** | **2019** | **2009** | **2019** |
| Female | 27,615 | 25,432 | 51% | 50% |
| Male | 26,942 | 25,499 | 49% | 50% |

**Marital Status**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **Marital Status** | **2009** | **2019** | **2009** | **2019** |
| Married | 26,641 | 23,434 | 49% | 46% |
| Living in common-law | 7,732 | 8,450 | 14% | 17% |
| Widowed | 635 | 540 | 1% | 1% |
| Separated | 1,576 | 1,410 | 3% | 3% |
| Divorced | 2,663 | 2,172 | 5% | 4% |
| Single, never married | 15,310 | 14,925 | 28% | 29% |

**Education**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **Education Attainment** | **2009** | **2019** | **2009** | **2019** |
| 0 to 8 years | 1,211 | 792 | 2% | 2% |
| Some high school | 5,892 | 3,891 | 11% | 8% |
| High school graduate | 11,614 | 9,980 | 21% | 20% |
| Some postsecondary | 4,454 | 2,946 | 8% | 6% |
| Postsecondary certificate or diploma | 20,090 | 19,724 | 37% | 39% |
| Bachelor's degree | 8,023 | 9,355 | 15% | 18% |
| Above bachelor's degree | 3,273 | 4,243 | 6% | 8% |

**Sector**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **Class of Worker** | **2009** | **2019** | **2009** | **2019** |
| Public sector employees | 14,434 | 13,700 | 22% | 23% |
| Private sector employees | 40,123 | 37,231 | 62% | 62% |

**Industry of main job**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **NAICS\_18** | **2009** | **2019** | **2009** | **2019** |
| Agriculture | 718 | 709 | 1% | 1% |
| Forestry, Fishing, Mining, Oil and Gas | 1,566 | 1,481 | 3% | 3% |
| Utilities | 633 | 495 | 1% | 1% |
| Construction | 3,590 | 3,746 | 7% | 7% |
| Manufacturing - durables | 3,245 | 2,736 | 6% | 5% |
| Manufacturing non-durables | 2,783 | 2,308 | 5% | 5% |
| Wholesale Trade | 1,769 | 1,590 | 3% | 3% |
| Retail Trade | 7,149 | 6,125 | 13% | 12% |
| Transportation and Warehousing | 2,537 | 2,559 | 5% | 5% |
| Finance, Insurance, Real Estate and Leasing | 2,794 | 2,506 | 5% | 5% |
| Professional, Scientific and Technical Services | 2,160 | 2,494 | 4% | 5% |
| Management, Administrative and Other Support | 1,813 | 1,621 | 3% | 3% |
| Educational Services | 4,574 | 4,465 | 8% | 9% |
| Health Care and Social Assistance | 7,280 | 7,432 | 13% | 15% |
| Information, Culture and Recreation | 2,185 | 1,858 | 4% | 4% |
| Accommodation and Food Services | 3,947 | 3,540 | 7% | 7% |
| Other Services | 2,056 | 1,716 | 4% | 3% |
| Public Administration | 3,758 | 3,550 | 7% | 7% |

**Occupation of main job**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Frequency (#)** |  | **Percentage (%)** |  |
| **NOC\_10** | **2009** | **2019** | **2009** | **2019** |
| Management occupations | 3,587 | 2,956 | 7% | 6% |
| Business, finance and administration occupations | 9,822 | 8,085 | 18% | 16% |
| Natural and applied sciences and related occupations | 3,380 | 3,583 | 6% | 7% |
| Health occupations | 3,923 | 4,077 | 7% | 8% |
| Occupations in education, law and social, community and government services | 5,143 | 6,371 | 9% | 13% |
| Occupations in art, culture, recreation and sport | 952 | 861 | 2% | 2% |
| Sales and service occupations | 14,578 | 12,782 | 27% | 25% |
| Trades, transport and equipment operators and related occupations | 8,766 | 8,055 | 16% | 16% |
| Natural resources, agriculture and related production occupations | 1,581 | 1,572 | 3% | 3% |
| Occupations in manufacturing and utilities | 2,825 | 2,589 | 5% | 5% |

#### Quantitative Attributes

##### Real Values

Quantitative variables codes do not include decimal points and must be divided by 10 or 100 to get the actual number. Only “tenure” variable codes are equal to the real value.

|  |  |  |  |
| --- | --- | --- | --- |
| VARIABLE | DESCRIPTION | CODE | REAL VALUE |
| utothrs | Usual hours worked per week at all jobs | 1-990 | 0.1-99.0 (one decimal implied) |
| Tenure | Job tenure with current employer (months) | 1-240 | 1-240 |
| hrlyearn | Usual hourly wages, employees only | 1-999999 | $0.01-$9,999.99 (two decimals implied) |

##### Overall Descriptive Statistics

Here are the descriptive statistics for the main quantitative attributes:

|  |  |  |  |
| --- | --- | --- | --- |
| VARIABLE | STATISTICS | 2009 | 2019 |
| Usual hours worked per week | MIN | 0.4 | 1.0 |
| MAX | 99.0 | 99.0 |
| AVERAGE | 36.1 | 36.3 |
| STD DEV | 11.2 | 11.2 |
| Job tenure with current employer (months) | MIN | 1 | 1 |
| MAX | 240 | 240 |
| AVERAGE | 85.3 | 86.8 |
| STD DEV | 83.4 | 82.6 |
| Usual hourly wages, employees only | MIN | 2.0 | 3.0 |
| MAX | 115.4 | 108.0 |
| AVERAGE | 21.5 | 27.5 |
| STD DEV | 11.3 | 13.7 |

# Approach

## Step 1: Build data set

Add records from data sets of Septembers 2009 and 2019 into a single data set. Remove variables that are not needed and matched the values of variables that have different value sets for each year. Assign the right type to each attribute.

## Step 2: Numeric variable analysis

Performed analysis of outliers using boxplots and verify numeric variables distribution. Analyze correlation and decide whether there are variables that should be excluded in the next steps.

## Step 3: Multiple Regression

To answer the first research question regarding the relevant variables that explain hourly wages by gender, at the start and end of the decade, multiple regression analysis will be applied. The resulting coefficients weights will be compared to determine whether there are significant differences.

## Step 4: Analysis of Variance

To answer the second research question concerning comparison of average hourly wages by sector, industries and occupations, analysis of variance will be applied to determine significant differences among the two periods studied and gender.

## Step 5: Hypothesis Testing for Proportions

Finally, to answer the last research question about gender distribution by sector, industry and occupation, test of hypothesis to compare proportions and identify significant differences by gender and through time.

# Results

## Numeric Variable Analysis

### Descriptive statistics by gender

Next, we have the basic statistics of each group by gender:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Male | | Female | |
| VARIABLE | **STATISTICS** | **2009** | **2019** | **2009** | **2019** |
| Usual hours worked per week | MIN | 0.4 | 1.0 | 0.4 | 1.0 |
| MAX | 99.0 | 99.0 | 99.0 | 99.0 |
| AVERAGE | 39.0 | 38.8 | 33.3 | 33.9 |
| STD DEV | 10.9 | 11.0 | 10.8 | 10.7 |
| Job tenure with current employer (months) | MIN | 1.0 | 1.0 | 1.0 | 1.0 |
| MAX | 240.0 | 240.0 | 240.0 | 240.0 |
| AVERAGE | 86.3 | 85.7 | 84.4 | 87.9 |
| STD DEV | 84.7 | 82.5 | 82.2 | 82.7 |
| Usual hourly wages, employees only | MIN | 2.1 | 3.0 | 2.0 | 3.1 |
| MAX | 115.4 | 108.0 | 89.7 | 106.7 |
| AVERAGE | 23.3 | 29.3 | 19.7 | 25.7 |
| STD DEV | 11.9 | 14.5 | 10.3 | 12.6 |

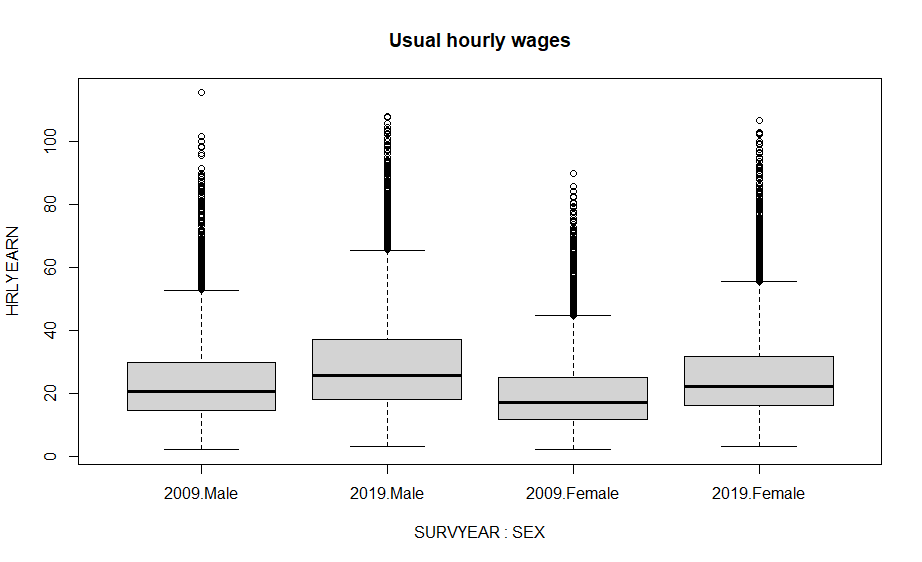
### Outliers

To identify outliers in the numeric variables, boxplot analysis was applied to groups by year and gender. Next are the summary of boxplot statistics and the graphs:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Male | | Female | |
| VARIABLE | **STATISTICS** | **2009** | **2019** | **2009** | **2019** |
|  | **TOTAL CASES** | **26,942** | **25,499** | **27,615** | **25,432** |
| Usual hours worked per week (utothrs) | LOWER FENCE | 33.8 | 33.8 | 15 | 15 |
| Q1 | 37.5 | 37.5 | 30 | 30 |
| MEDIAN | 40 | 40 | 37 | 37.5 |
| Q3 | 40 | 40 | 40 | 40 |
| UPPER FENCE | 43.5 | 43.7 | 55 | 55 |
| TOTAL OUTLIERS | 8,453 | 8,080 | 2,324 | 2,096 |
| % OUTLIERS | 31% | 32% | 8% | 8% |
| Job tenure with current employer in months (tenure) | LOWER FENCE | 1 | 1 | 1 | 1 |
| Q1 | 15 | 15 | 16 | 16 |
| MEDIAN | 51 | 53 | 50 | 57 |
| Q3 | 144 | 144 | 135 | 148 |
| UPPER FENCE | 240 | 240 | 240 | 240 |
| TOTAL OUTLIERS | 0 | 0 | 0 | 0 |
| % OUTLIERS | 0% | 0% | 0% | 0% |
| Usual hourly wages, employees only (hrlyearn) | LOWER FENCE | 2.14 | 3 | 2 | 3.07 |
| Q1 | 14.5 | 18 | 11.75 | 16 |
| MEDIAN | 20.51 | 25.65 | 17.14 | 22 |
| Q3 | 29.8 | 37 | 24.91 | 31.79 |
| UPPER FENCE | 52.69 | 65.38 | 44.62 | 55.38 |
| TOTAL OUTLIERS | 675 | 618 | 708 | 681 |
| % OUTLIERS | 3% | 2% | 3% | 3% |







### Distribution of numeric variables

To verify normality of the quantitative variables, histograms and q-q plots were built, and Kolmogorov-Smirnov tests were run. Shapiro-Wilk’s method was not applied because is restricted to samples of size smaller than 5,000.

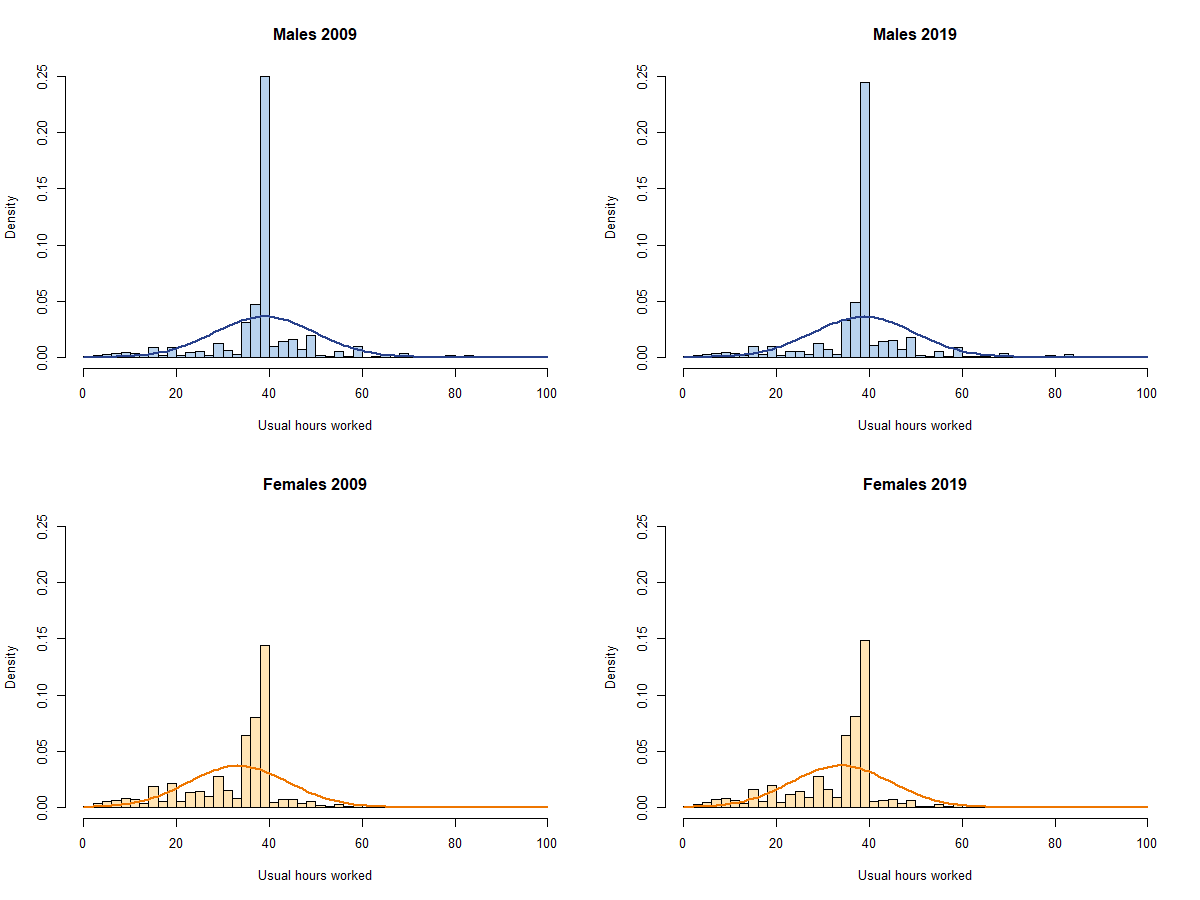
The histograms show the shape of the distribution and compare it to a curve that would follow a normal distribution with the same mean and standard deviation as the ones from the variable tested.

The q-q plot (or quantile-quantile plot) draws the correlation between a given sample and the normal distribution. A 45-degree reference line is also plotted. If both sets of quantiles come from the same distribution, we should see the points forming a line that is roughly straight. In contrast, if points form a curve instead of a straight line it usually means the sample data are skewed. Moreover, if the points fall along a line in the middle of the graph, but curve off in the extremities it is a sign that the data have more extreme values than would be expected if they truly came from a normal distribution.

The Kolmogorov-Smirnov test compares the observed distribution with a theoretically specified distribution, in this case, the normal distribution. It is important that this distribution has identical mean and standard deviation as the sample evaluated. The null hypothesis of the K-S test is that the distribution is normal, therefore, if p-value of the test is >0.05, we do not reject the null hypothesis and conclude that the distribution in question is not statistically different from a normal distribution.

#### Usual hours worked per week (utothrs)

**Histograms**



**Q-Q Plots**

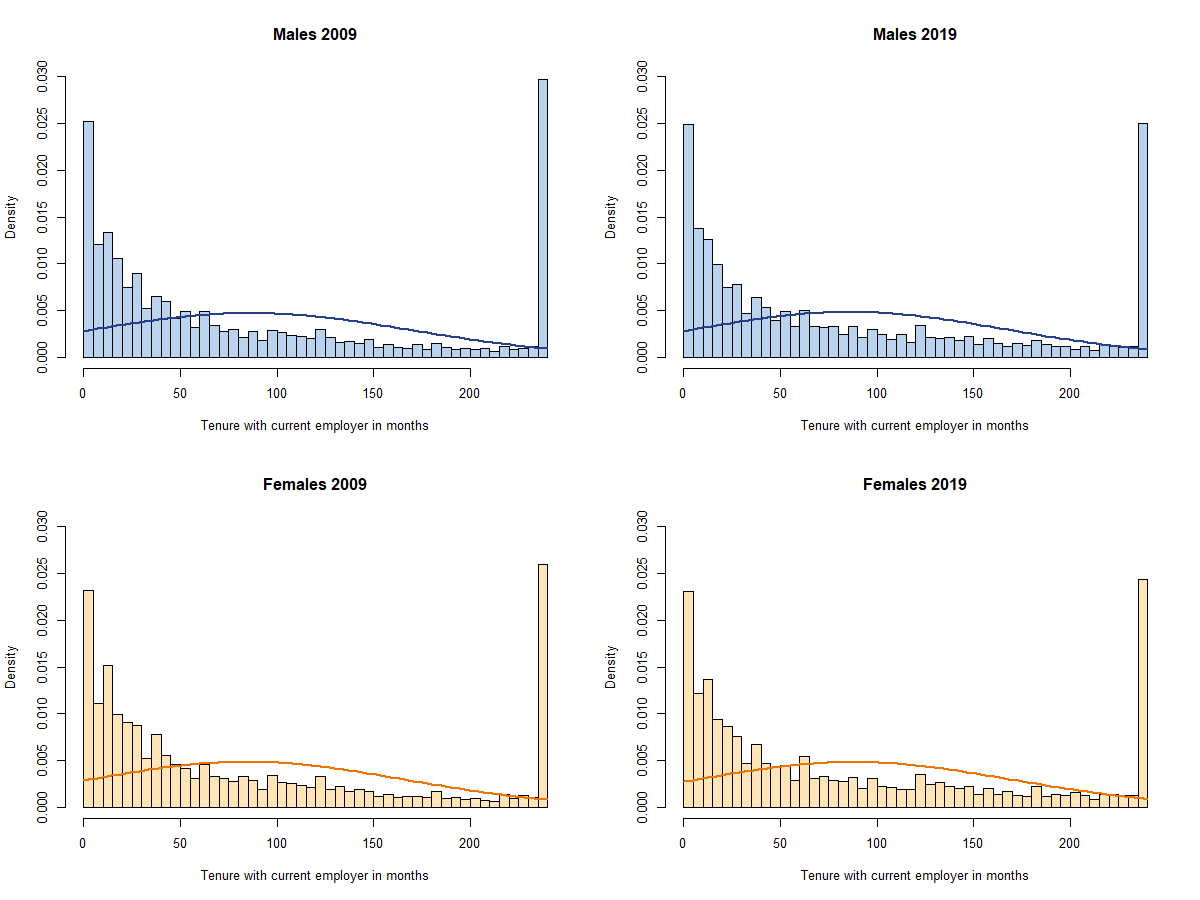
|  |  |
| --- | --- |
| **Males 2009** | **Males 2019** |
| **Females 2009** | **Females 2019** |

**Kolmogorov-Smirnov**

|  |  |
| --- | --- |
| **Males 2009**  data: data.all.09.male$UTOTHRS  D = 0.266, p-value < 2.2e-16  alternative hypothesis: two-sided | **Males 2019**  data: data.all.19.male$UTOTHRS  D = 0.26344, p-value < 2.2e-16  alternative hypothesis: two-sided |
| **Females 2009**  data: data.all.09.fem$UTOTHRS  D = 0.21438, p-value < 2.2e-16  alternative hypothesis: two-sided | **Females 2019**  data: data.all.19.fem$UTOTHRS  D = 0.21143, p-value < 2.2e-16  alternative hypothesis: two-sided |

#### Job Tenure

**Histogram**



**Q-Q Plots**

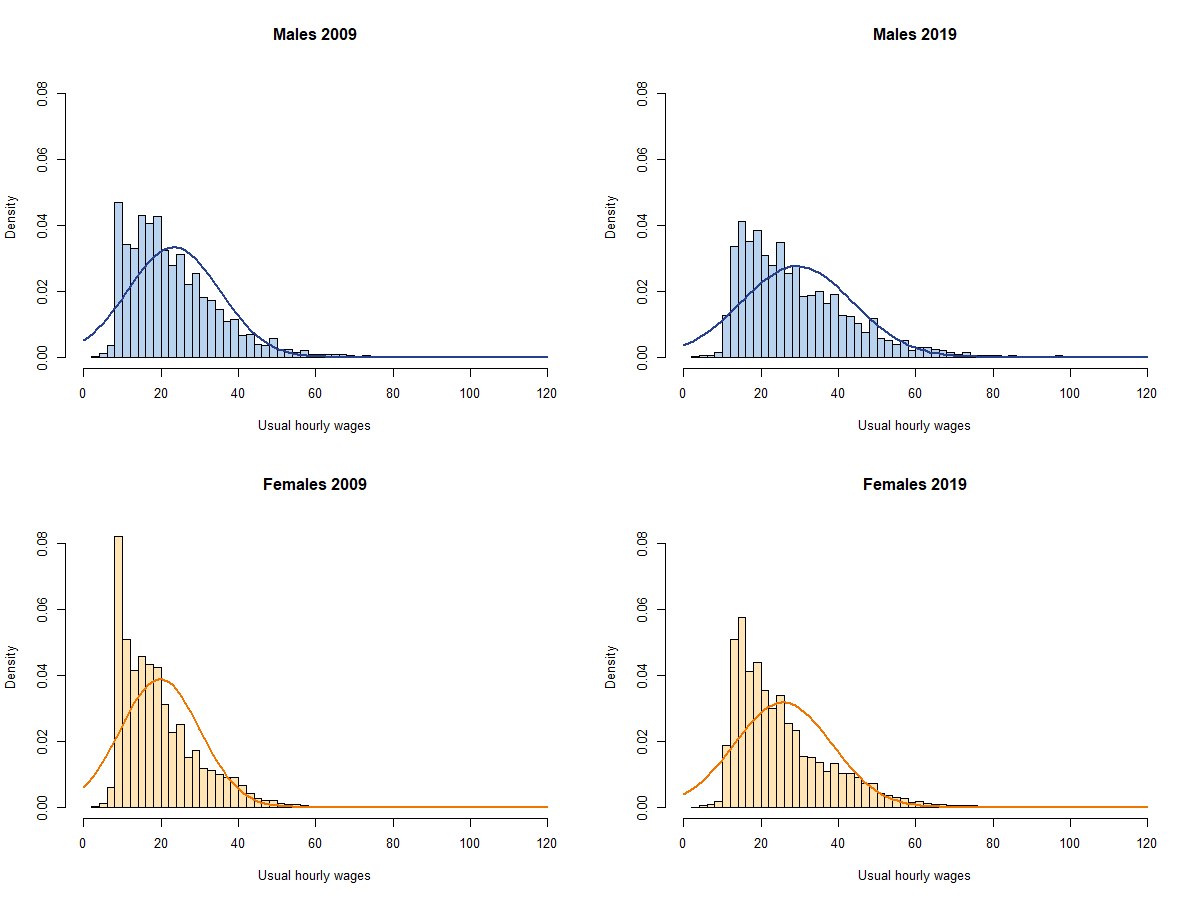
|  |  |
| --- | --- |
| **Males 2009** | **Males 2019** |
| **Females 2009** | **Females 2019** |

**Kolmogorov-Smirnov**

|  |  |
| --- | --- |
| **Males 2009**  data: data.all.09.male$TENURE  D = 0.16707, p-value < 2.2e-16  alternative hypothesis: two-sided | **Males 2019**  data: data.all.19.male$TENURE  D = 0.15484, p-value < 2.2e-16  alternative hypothesis: two-sided |
| **Females 2009**  data: data.all.09.fem$TENURE  D = 0.16497, p-value < 2.2e-16  alternative hypothesis: two-sided | **Females 2019**  data: data.all.19.fem$TENURE  D = 0.1523, p-value < 2.2e-16  alternative hypothesis: two-sided |

#### Usual Hourly Wages

**Histogram**



**Q-Q Plots**

|  |  |
| --- | --- |
| **Males 2009** | **Males 2019** |
| **Females 2009** | **Females 2019** |

**Kolmogorov-Smirnov**

|  |  |
| --- | --- |
| **Males 2009**  data: data.all.09.male$HRLYEARN  D = 0.098075, p-value < 2.2e-16  alternative hypothesis: two-sided | **Males 2019**  data: data.all.19.male$HRLYEARN  D = 0.10788, p-value < 2.2e-16  alternative hypothesis: two-sided |
| **Females 2009**  data: data.all.09.fem$HRLYEARN  D = 0.11895, p-value < 2.2e-16  alternative hypothesis: two-sided | **Females 2019**  data: data.all.19.fem$HRLYEARN  D = 0.12557, p-value < 2.2e-16  alternative hypothesis: two-sided |

Looking at the graphs and the results from Kolmogorov-Smirnov, we can conclude that none of the numeric variables are normally distributed.

Parametric tests usually to be normally distributed, however, there are cases where they still can be applied when the sample size is large enough.

On the other hand, linear regression remains a statistically sound technique in studies of large sample sizes even when a dependent variable is not distributed normally. By the law of large numbers and the central limit theorem, the ordinary least squares (OLS) estimators in linear regression technique still will be approximately normally distributed around the true parameter values, which implies the estimated parameters and their confidence interval estimates remain robust. Hence, in a large sample, the use of a linear regression technique, even if the dependent variable violates the “normality assumption” rule, remains valid.

### Correlation

Pearson correlation measures linear correlation between two numerical variables:

UTOTHRS TENURE HRLYEARN

UTOTHRS 1.0000000 0.1332259 0.1988056

TENURE 0.1332259 1.0000000 0.3074330

HRLYEARN 0.1988056 0.3074330 1.0000000

Here, we can see that there is no evidence of relationship between the variables, except for a weak correlation between tenure and hourly wages.

Besides, Spearman test allows to include ordinal variable to analyse the correlation:

UTOTHRS TENURE HRLYEARN AGE\_12 EDUC

UTOTHRS 1.00000000 0.1165325 0.2166884 0.10925217 0.03582495

TENURE 0.11653250 1.0000000 0.3715966 0.48665429 0.10148136

HRLYEARN 0.21668838 0.3715966 1.0000000 0.25836345 0.41618973

AGE\_12 0.10925217 0.4866543 0.2583635 1.00000000 0.06745240

EDUC 0.03582495 0.1014814 0.4161897 0.06745240 1.00000000

ESTSIZE 0.04903734 0.1926394 0.3001594 0.07583280 0.17116224

FIRMSIZE -0.03107291 0.1770767 0.2376289 0.03684549 0.15247882

AGYOWNK -0.07936159 -0.1051187 -0.1973903 0.01913263 -0.16673495

ESTSIZE FIRMSIZE AGYOWNK

UTOTHRS 0.04903734 -0.03107291 -0.07936159

TENURE 0.19263945 0.17707666 -0.10511872

HRLYEARN 0.30015941 0.23762892 -0.19739033

AGE\_12 0.07583280 0.03684549 0.01913263

EDUC 0.17116224 0.15247882 -0.16673495

ESTSIZE 1.00000000 0.58094125 -0.06851089

FIRMSIZE 0.58094125 1.00000000 -0.05406980

AGYOWNK -0.06851089 -0.05406980 1.00000000

There are only moderate correlation between Age and Tenure, as well as Hourly Wages and Education. There is also a strong correlation between establishment and firm size, which is to be expected since large firms usually will have offices with many employees.

## Feature Selection

### Training and Test Set

The dataset was split into 70% of training and 30% of test sets, selecting observations randomly.

### Forward Selection

Here are the results of the Forward Selection analysis applied to training data:

Call:

lm(formula = HRLYEARN ~ NOC\_10 + NAICS\_18 + EDUC + TENURE + PROV +

ESTSIZE + AGE\_12 + SEX + FIRMSIZE + PERMTEMP + EFAMTYPE +

FTPTMAIN + COWMAIN + UTOTHRS + AGYOWNK + UNION + CMA + MARSTAT +

LFSSTAT + MJH, data = train)

Residuals:

Min 1Q Median 3Q Max

-42.150 -5.465 -0.979 4.084 83.307

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 34.394399 0.603287 57.012 < 2e-16 \*\*\*

NOC\_10Business, finance and administration -11.496253 0.165983 -69.261 < 2e-16 \*\*\*

NOC\_10Natural and applied sciences -5.761783 0.199661 -28.858 < 2e-16 \*\*\*

NOC\_10Health -5.098106 0.220114 -23.161 < 2e-16 \*\*\*

NOC\_10Education, law and social, community and government services -6.456751 0.192576 -33.528 < 2e-16 \*\*\*

NOC\_10Art, culture, recreation and sport -10.107108 0.306934 -32.929 < 2e-16 \*\*\*

NOC\_10Sales and service -12.930467 0.166751 -77.543 < 2e-16 \*\*\*

NOC\_10Trades, transport and equipment operators -10.804708 0.183130 -59.000 < 2e-16 \*\*\*

NOC\_10Natural resources and agriculture -11.880225 0.319340 -37.202 < 2e-16 \*\*\*

NOC\_10Manufacturing and utilities -13.382497 0.242719 -55.136 < 2e-16 \*\*\*

NAICS\_18Forestry, Fishing, Mining, Oil and Gas 9.220595 0.380762 24.216 < 2e-16 \*\*\*

NAICS\_18Utilities 8.236889 0.514284 16.016 < 2e-16 \*\*\*

NAICS\_18Construction 5.813064 0.407672 14.259 < 2e-16 \*\*\*

NAICS\_18Manufacturing durables 2.430302 0.415166 5.854 4.82e-09 \*\*\*

NAICS\_18Manufacturing non-durables 1.489599 0.420267 3.544 0.000394 \*\*\*

NAICS\_18Wholesale Trade 2.793519 0.427513 6.534 6.43e-11 \*\*\*

NAICS\_18Retail Trade -1.225378 0.397647 -3.082 0.002060 \*\*

NAICS\_18Transportation and Warehousing 1.317339 0.414219 3.180 0.001472 \*\*

NAICS\_18Finance, Insurance, Real Estate and Leasing 3.302142 0.412331 8.008 1.18e-15 \*\*\*

NAICS\_18Professional, Scientific and Technical Services 4.420618 0.416593 10.611 < 2e-16 \*\*\*

NAICS\_18Management, Administrative and Other Support -0.460703 0.413205 -1.115 0.264875

NAICS\_18Educational Services 1.312515 0.427536 3.070 0.002142 \*\*

NAICS\_18Health Care and Social Assistance -0.964540 0.408595 -2.361 0.018247 \*

NAICS\_18Information, Culture and Recreation 0.470978 0.418771 1.125 0.260734

NAICS\_18Accommodation and Food Services -1.143406 0.409307 -2.794 0.005215 \*\*

NAICS\_18Other Services 0.834436 0.420744 1.983 0.047345 \*

NAICS\_18Public Administration 4.187362 0.422196 9.918 < 2e-16 \*\*\*

EDUC.L 9.395202 0.186979 50.247 < 2e-16 \*\*\*

EDUC.Q 3.354446 0.170323 19.695 < 2e-16 \*\*\*

EDUC.C 1.135280 0.135562 8.375 < 2e-16 \*\*\*

EDUC^4 -0.515425 0.122868 -4.195 2.73e-05 \*\*\*

EDUC^5 -0.554527 0.087690 -6.324 2.57e-10 \*\*\*

EDUC^6 -0.024085 0.103369 -0.233 0.815759

TENURE 0.021418 0.000512 41.836 < 2e-16 \*\*\*

PROVPEI -1.549582 0.278442 -5.565 2.63e-08 \*\*\*

PROVNS -1.428673 0.243097 -5.877 4.20e-09 \*\*\*

PROVNB -1.834636 0.243355 -7.539 4.79e-14 \*\*\*

PROVQc 0.392609 0.213197 1.842 0.065547 .

PROVON 1.884180 0.203093 9.277 < 2e-16 \*\*\*

PROVMB -0.097977 0.222217 -0.441 0.659283

PROVSK 2.300858 0.228873 10.053 < 2e-16 \*\*\*

PROVAB 4.690275 0.217067 21.608 < 2e-16 \*\*\*

PROVBC 2.986442 0.226745 13.171 < 2e-16 \*\*\*

ESTSIZE.L 2.962233 0.101435 29.203 < 2e-16 \*\*\*

ESTSIZE.Q 0.674583 0.085070 7.930 2.23e-15 \*\*\*

ESTSIZE.C 0.126906 0.077248 1.643 0.100420

AGE\_12.L 2.334002 0.280904 8.309 < 2e-16 \*\*\*

AGE\_12.Q -3.235781 0.260918 -12.402 < 2e-16 \*\*\*

AGE\_12.C -1.465221 0.218419 -6.708 1.98e-11 \*\*\*

AGE\_12^4 0.254909 0.196182 1.299 0.193828

AGE\_12^5 -1.173816 0.179385 -6.544 6.05e-11 \*\*\*

AGE\_12^6 0.383617 0.162512 2.361 0.018251 \*

AGE\_12^7 -0.322107 0.143726 -2.241 0.025021 \*

AGE\_12^8 0.154170 0.126597 1.218 0.223303

AGE\_12^9 0.022915 0.113877 0.201 0.840523

AGE\_12^10 0.085832 0.105822 0.811 0.417315

AGE\_12^11 -0.007170 0.101236 -0.071 0.943538

SEXFemale -2.793067 0.085710 -32.587 < 2e-16 \*\*\*

FIRMSIZE.L 1.409881 0.093332 15.106 < 2e-16 \*\*\*

FIRMSIZE.Q -0.041170 0.089565 -0.460 0.645754

FIRMSIZE.C 0.008198 0.093211 0.088 0.929919

PERMTEMPTemporary, seasonal -1.988367 0.184070 -10.802 < 2e-16 \*\*\*

PERMTEMPTemporary, term or contract -1.441414 0.148590 -9.701 < 2e-16 \*\*\*

PERMTEMPTemporary, casual or other -1.432111 0.191765 -7.468 8.23e-14 \*\*\*

EFAMTYPEHWDENC 0.311675 0.186471 1.671 0.094639 .

EFAMTYPEHWDE17 0.310602 0.239247 1.298 0.194207

EFAMTYPEHWDE24 -0.307748 0.234600 -1.312 0.189593

EFAMTYPEHWSHNC 0.138287 0.242827 0.569 0.569026

EFAMTYPEHWSH17 0.140773 0.288197 0.488 0.625225

EFAMTYPEHWSH24 -0.132596 0.392703 -0.338 0.735628

EFAMTYPEHWSWNC -0.655901 0.253341 -2.589 0.009627 \*\*

EFAMTYPEHWSW17 -1.025395 0.372463 -2.753 0.005906 \*\*

EFAMTYPEHWSW24 -1.630267 0.452259 -3.605 0.000313 \*\*\*

EFAMTYPEHWNENC -1.045643 0.413061 -2.531 0.011361 \*

EFAMTYPEHWNE17 -1.111293 1.069662 -1.039 0.298846

EFAMTYPEHWNE24 -0.851020 0.808786 -1.052 0.292702

EFAMTYPESPE17 -0.130313 0.280367 -0.465 0.642080

EFAMTYPESPE24 -0.283631 0.308267 -0.920 0.357532

EFAMTYPESPN17 -0.380748 0.700241 -0.544 0.586624

EFAMTYPESPN24 -0.109977 0.783216 -0.140 0.888331

EFAMTYPEOther -0.759658 0.163530 -4.645 3.40e-06 \*\*\*

FTPTMAINPart-time -1.923942 0.149238 -12.892 < 2e-16 \*\*\*

COWMAINPrivate sector -1.229040 0.144257 -8.520 < 2e-16 \*\*\*

UTOTHRS -0.042972 0.005302 -8.105 5.34e-16 \*\*\*

AGYOWNK.L -0.685752 0.228197 -3.005 0.002656 \*\*

AGYOWNK.Q -0.064969 0.155101 -0.419 0.675302

AGYOWNK.C -0.180315 0.172613 -1.045 0.296201

AGYOWNK^4 -0.535790 0.158112 -3.389 0.000703 \*\*\*

UNIONNot a member but covered by a union contract 0.249071 0.248336 1.003 0.315885

UNIONNon-unionized -0.389969 0.097487 -4.000 6.33e-05 \*\*\*

CMAToronto -0.086860 0.250809 -0.346 0.729105

CMAVancouver -1.015192 0.284722 -3.566 0.000363 \*\*\*

CMAOther -0.070826 0.192105 -0.369 0.712364

MARSTATLiving in common-law 0.207016 0.108946 1.900 0.057415 .

MARSTATWidowed -1.280200 0.362518 -3.531 0.000414 \*\*\*

MARSTATSeparated 0.343679 0.255216 1.347 0.178107

MARSTATDivorced -0.132868 0.221003 -0.601 0.547707

MARSTATSingle, never married -0.149997 0.175384 -0.855 0.392416

LFSSTATEmployed, absent from work -0.509135 0.133573 -3.812 0.000138 \*\*\*

MJHMultiple jobholder -0.240789 0.166549 -1.446 0.148251

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.257 on 73741 degrees of freedom

Multiple R-squared: 0.4806, Adjusted R-squared: 0.4799

F-statistic: 689.3 on 99 and 73741 DF, p-value: < 2.2e-16