**Maths Studies Project**

**The Relationship between Total Income of Families with Children and House Price Index in Canada from 2000 – 2019**

**Student’s Name:**

**Teacher’s Name:**

**Table of Content**

Page

|  |  |
| --- | --- |
| Introduction | 2 |
| Analysis  | 2 |
| * Part 1: Math Calculation
 | 2  |
| * Part 2: Histogram and Line Graph
 | 6 |
| * Part 3: Evaluation
 | 7 |
| Conclusion  | 8 |
| Reference  | 9 |

**Introduction**

The aim of this project is to observe whether there is a correlation between total income of families with children and house price index in Canada from 2000 – 2019 by analyzing measure of central tendency, measure of spread, and linear regression. The house price has been climbing in the past two decades, as well as the family income due to the overall health of the economy. Therefore, the null hypothesis of this project that I’m curious about is: Higher the total income of family with children, higher the house price index. In order to challenge the null hypothesis, the data for total income of families with children were collected from the reliable source Statista (Families are comprised of couples and single parents living with one or more children); and the data for house price index (Index 2010 = 100) were collected from FRED economic data (*Real residential property prices for Canada*. FRED, 2022, March 24; Statista).

**Analysis**

**Part 1: Math Calculations**

1. Measure of Central Tendency of independent variable: Total income of families with children in Canada from 2000 to 2019

Mean: is the sum of the values in a set of data divided by the number of values in the set.

Mean = $\frac{\sum\_{}^{}x}{n}=\frac{53700+56700+58600+…+102550}{20}=\$76,533$

Median: is the middle value when values are ranked from highest to lowest.

Median = $\frac{74690+75980}{2}=\$75,335$

Mode: is the value which occurs the most frequent in a set of data.

Mode = N/A because there is no repeated values

1. Measure of Spread:
2. Independent variable: Total income
3. Standard deviation and Variance: shows how values in a distribution are centered about the mean.

Standard deviation, $σ$ = $\sqrt{\frac{\sum\_{}^{}(x-u)^{2}}{n}}$ and variance = $σ^{2}$

Table 1:

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Income of Families with Children (x) | x-u | (x-u)^2 |
| 2000 | 53700 | -22833 | 521345889 |
| 2001 | 56700 | -19833 | 393347889 |
| 2002 | 58600 | -17933 | 321592489 |
| 2003 | 59300 | -17233 | 296976289 |
| 2004 | 61600 | -14933 | 222994489 |
| 2005 | 64900 | -11633 | 135326689 |
| 2006 | 68400 | -8133 | 66145689 |
| 2007 | 71920 | -4613 | 21279769 |
| 2008 | 74690 | -1843 | 3396649 |
| 2009 | 74190 | -2343 | 5489649 |
| 2010 | 75980 | -553 | 305809 |
| 2011 | 78890 | 2357 | 5555449 |
| 2012 | 81670 | 5137 | 26388769 |
| 2013 | 84270 | 7737 | 59861169 |
| 2014 | 87200 | 10667 | 113784889 |
| 2015 | 89860 | 13327 | 177608929 |
| 2016 | 91650 | 15117 | 228523689 |
| 2017 | 95250 | 18717 | 350326089 |
| 2018 | 99340 | 22807 | 520159249 |
| 2019 | 102550 | 26017 | 676884289 |
| Sum |  |  | 4147293820 |

Standard deviation, $σ$ = $\sqrt{\frac{4147293820}{20}}=\$14,400.16$

variance = $σ^{2}=207,364,691$

1. Interquartile range: is the range for the middle half of the data, or $Q\_{3}-Q\_{1}$

$Q\_{1}=\frac{61600+64900}{2}=\$63,250$ and $Q\_{3}=\frac{87200+89860}{2}=\$88,530$

$Q\_{3}-Q\_{1}=88530-63250=\$25,280$

1. z-score: is a measure of how many standard deviation is a particular data value is from the mean, $z=\frac{x-u}{σ}$.

In year 2000, the z-score of family income = $\frac{53700-76533}{14400.16}=-1.59$

The z-score for all data values are shown below:

Table 2:

|  |  |  |
| --- | --- | --- |
| Year | Income of Families with Children (x) | z-score |
| 2000 | 53700 | -1.59 |
| 2001 | 56700 | -1.38 |
| 2002 | 58600 | -1.25 |
| 2003 | 59300 | -1.20 |
| 2004 | 61600 | -1.04 |
| 2005 | 64900 | -0.81 |
| 2006 | 68400 | -0.56 |
| 2007 | 71920 | -0.32 |
| 2008 | 74690 | -0.13 |
| 2009 | 74190 | -0.16 |
| 2010 | 75980 | -0.04 |
| 2011 | 78890 | 0.16 |
| 2012 | 81670 | 0.36 |
| 2013 | 84270 | 0.54 |
| 2014 | 87200 | 0.74 |
| 2015 | 89860 | 0.93 |
| 2016 | 91650 | 1.05 |
| 2017 | 95250 | 1.30 |
| 2018 | 99340 | 1.58 |
| 2019 | 102550 | 1.81 |

1. Dependent variable: Real Residential Property prices Index for Canada
2. Standard deviation: $σ=24.19$
3. Variance: $σ^{2}=585.36$
4. Interquartile range: $IQR=34.1135$
5. z-score in the year of 2019: $z=1.72$
6. Linear regression and correlation:

Scatter Plot:

Figure 1: The relationship between Total income of families with children in Canada and real residential property prices index for Canada from 2000 to 2019

Linear regression: $y=0.0017x-31.4$ and correlation coefficient, $r=0.98615$

The following table shows the steps on how to get the linear regression and correlation coefficient.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Income of Families with Children (x) | Real Residential Property prices Index for Canada (y) | xy | x^2 | y^2 |
| 2000 | 53700 | 58.2334 | 3127133.5800 | 2883690000 | 3391.1289 |
| 2001 | 56700 | 58.0960 | 3294043.2000 | 3214890000 | 3375.1452 |
| 2002 | 58600 | 63.2774 | 3708055.6400 | 3433960000 | 4004.0294 |
| 2003 | 59300 | 65.7826 | 3900908.1800 | 3516490000 | 4327.3505 |
| 2004 | 61600 | 71.9819 | 4434085.0400 | 3794560000 | 5181.3939 |
| 2005 | 64900 | 77.4652 | 5027491.4800 | 4212010000 | 6000.8572 |
| 2006 | 68400 | 80.5446 | 5509250.6400 | 4678560000 | 6487.4326 |
| 2007 | 71920 | 92.9796 | 6687092.8320 | 5172486400 | 8645.2060 |
| 2008 | 74690 | 99.5048 | 7432013.5120 | 5578596100 | 9901.2052 |
| 2009 | 74190 | 92.7926 | 6884282.9940 | 5504156100 | 8610.4666 |
| 2010 | 75980 | 100.2634 | 7618013.1320 | 5772960400 | 10052.7494 |
| 2011 | 78890 | 99.5791 | 7855795.1990 | 6223632100 | 9915.9972 |
| 2012 | 81670 | 101.8205 | 8315680.2350 | 6669988900 | 10367.4142 |
| 2013 | 84270 | 103.6024 | 8730574.2480 | 7101432900 | 10733.4573 |
| 2014 | 87200 | 106.8698 | 9319046.5600 | 7603840000 | 11421.1542 |
| 2015 | 89860 | 110.8042 | 9956865.4120 | 8074819600 | 12277.5707 |
| 2016 | 91650 | 117.2243 | 10743607.0950 | 8399722500 | 13741.5365 |
| 2017 | 95250 | 131.5822 | 12533204.5500 | 9072562500 | 17313.8754 |
| 2018 | 99340 | 138.5269 | 13761262.2460 | 9868435600 | 19189.7020 |
| 2019 | 102550 | 137.1277 | 14062445.6350 | 1.0517E+10 | 18804.0061 |
| Sum | 1530660 | 1908.0586 | 152900851.4100 | 1.2129E+11 | 193741.6784 |

$y=mx+b$, where $m=\frac{n\sum\_{}^{}xy-\sum\_{}^{}x\sum\_{}^{}y}{n\sum\_{}^{}x^{2}-(\sum\_{}^{}x)^{2}}$ and $b=\overbar{y}-m\overbar{x}$

$m=\frac{20\*152900851-(1530660)(1908.0586)}{20\*1.2129\*10^{11}-(1530660)^{2}}=0.0017$

$b=95.4029-0.0017\*76533=-31.4$

$r=\frac{n\sum\_{}^{}xy-\sum\_{}^{}x\sum\_{}^{}y}{\sqrt{(n\sum\_{}^{}x^{2}-(\sum\_{}^{}x)^{2})(n\sum\_{}^{}y^{2}-(\sum\_{}^{}y)^{2})}}$

$ =\frac{20\*152900851.41-(1530660)(1908.0586)}{\sqrt{(20\*1.2129\*10^{11}-(1530660)^{2})(20\*193741.6784-(1908.0586)^{2})}}$

$r=0.98615$

**Part 2: Graphs (Histogram and line graph)**

Figure 2: Canada – total income of families with children 2000 – 2019

Figure 3: Canada – Real residential property prices index 2000 – 2019 (Index 2010 = 100)

**Part 3: Evaluation**

Total income of families with children in Canada is independent variable, and real residential property price index for Canada is dependent variable. Both independent variable and dependent variable are continuous because they can be any values, either integers or decimals. In the past 20 years from 2000 to 2019, both total income of families with children and house price has been increasing constantly (Figure 2, Figure 3). Year to year increase is 3-5% except the year of 2008 because of the great rescission negatively impacted on the housing market and employment market during that time.

The average and median of total income is $76,533 and $75,335 respectively. The two values are similar because the independent variable doesn’t show any outliers. To perform the outlier check, the interval of ($Q\_{1}+1.5IQR,Q\_{3}+1.5IQR)$ needs to be calculated, which turns out ($25,330, $126,450). None of the data points is outside of this interval, hence no outlier detected. Moreover, the interquartile range $25,280 indicates the middle half of the data has a range of $25,280 around the median. The z-score shows a measure of how many standard deviation is a particular data value is from the mean. Both variables show an increasing pattern, so the z-score from 2000 to 2010 are negative because they are under the average, whereas the data points from 2011 to 2019 are positive as they are above the average. More particularly, the z-score of income in year of 2000 equals -1.59 indicating $53,700 is 1.59 standard deviation below $76,533.

The standard deviation of total income and house price index is $14,400.16 and 24.19. A higher standard deviation shows that the data is widely spread out and less reliable because they are inconsistent distributed, and a low standard deviation shows that the data are clustered closely around the mean and it’s more reliable.

The correlation coefficient is 0.98615 which implies total income of families with children is strongly and positively correlated to house price index in Canada from 2000 to 2019. The trend can be observed from the scatter plot (Figure 1), house price index increases as total income increases. The data points are very close to the trendline. The equation of linear regression is y = 0.0017x – 31.4, which can be used to interpolate and extrapolate by plugging in any x values. For example, if total income were $90,000 on June of 2016, the house price index would be 0.0017 \* 90,000 - 31.4 = 121.6; if total income were increased to $150,000 in 10 years in the future, the house price index would go up to 0.0017 \* 150,000 – 31.4 = 223.6. Even the scatter plot shows a strong positive correlation between income and house market, there are so many other common factors could drive both income level and real estate market. Canada is an immigration country, nearly 200,000 new immigrants become permanent resident per year. Such population growth increases the house demand so that to drive the house price going up. At the meanwhile, the growth of economy in the past several decades enhanced manufacturing activity, the prices of good, employment rate, family income, as well as house market. (Investopedia, Feb, 2022)

**Conclusion:**

The equation of linear regression$ y=0.0017x-31.4$ with correlation coefficient of $r=0.98615$ provides a strong evidence to support the null hypothesis, so higher the total income of family with children, higher the house price index in Canada from 2000 – 2019. The common factors that cause both income and house market to increase are large number of immigration and sustainable growth of economy in Canada. However, one of the possible bias could be household bias due to the sample size is too small. In this project, 20 years data were collected but an overall look from 1970 to 2021 would be more adequate and sufficient.

**Reference:**

*Real residential property prices for Canada*. FRED. (2022, March 24). Retrieved June 17, 2022, from https://fred.stlouisfed.org/series/QCAR628BIS

Jeudy, L. (2021, November 24). *Canada: Median income of families with children 2019*. Statista. Retrieved June 17, 2022, from https://www.statista.com/statistics/478958/median-total-income-of-families-with-children-in-canada/

Nguyen, J. (2022, February 16). *4 key factors that drive the real estate market*. Investopedia. Retrieved June 17, 2022, from https://www.investopedia.com/articles/mortages-real-estate/11/factors-affecting-real-estate-market.asp