**DATA624 - DATATHON #7**

**GROUP 1 – ADAEZE NWIGWE, ADRIAN MITCHELL, MINKI LEE**

**APRIL 14, 2021**

**STUDIES OF ALBERTA COMMUNITY HEALTH SURVEY DATA**

**Introduction**

2014 in Alberta was the beginning of an economic recession in the province. Between January and December 2014 more than 20,000 Albertans lost their jobs in a declining provincial economy.[[1]](#footnote-2) By 2019, this number would be more than 100,000 (Tombe, 2019)￼￼. Data collected [[2]](#footnote-3)￼ collected over 106 pieces of information from respondents about from demographics, physical health, through mental health, personal perceptions of their abilities to handle stress as well as their recent life experiences. This data could prove useful for health-care providers to proactively manage patient health and pre-emptively mitigate health decline.

With a view to examining the impacts of a declining economy, at the start of the provincial recession, the guiding questions that we wanted to explore, and their rationale, were as follows:

1. *What are the most important contributors to stress and how well do people manage them?*

Any number of things can contribute to stress and people react differently to a stressful situation. Therefore, our team was interested in finding the main causes of stress and how well people manage it.

1. *How do financially related recent life events affect individuals’ ability to handle stress and overall life satisfaction?*

Financial struggles can have a severe impact on individual stress level and can negatively affect life satisfaction. Our team set out to understand how having recently experienced a finance related life crisis can affect individuals’ ability to handle stress and overall life satisfaction.

1. *Is there a statistically significant association between the answers for stress regarding their handling of personal crises and the demands of day-to-day life and any recent financial crises or job losses they may have experienced?*

Given that the respondents answered questions about whether they had recently experienced financial crises, job loss, or cut in pay as well as answered questions about their views on their ability to handle difficult situations and financial crises, our team was interested to see whether these responses followed a predictable pattern that could be modelled. If successful, this could give valuable insight and early-warning to health-care workers and policy makers regarding public health consequences of a poor economy.

**Datasets**

“Alberta Community Health Survey.csv”

This dataset contains 106 columns of data which span 7559 rows. The columns within the dataset included fields we explored including: STS\_Q1-Q3, SATIS\_1, and RLE\_1-4. Since this dataset was so large in nature, it was very important to generate specific research questions that would limit the amount of data used in analysis, allowing for much more specific results. Different sets of variables were used for parts of the analysis, because they gave us the ability to create a story with our data using multiple combinations of variables.

**Analysis**

We investigated Alberta Community Health Survey data. Data cleaning and wrangling were performed using Python and RStudio. Calculations and visualisations were performed using Tableau.

**Likert Chart and Pie Charts**

The Alberta Community Health Survey dataset was first filtered using Python before importing it into RStudio for a Likert chart. Unnecessary columns were dropped, and only the columns of interest, “STS\_Q1”, “STS\_Q2”, and “STS\_Q3”, were extracted. We replaced any NaN values with zeros in Pandas DataFrame and assigned each score value to each corresponding score category – e.g., excellent is 1, poor is 5, and prefer not to say is 0 on a 6-point scale – for “STS\_Q1” and “STS\_Q2”. The “STS\_Q3” column had a 17-point scale – e.g., time pressure is 1, nothing is 16, and prefer not to say is 0. The cleaned dataset was then imported into RStudio to visualize the Likert chart and into Tableau to create pie charts of the top three important things contributing to feelings of stress.

**Square Tile Plot**

The Alberta Community Health Survey dataset was processed using R. The columns that were kept for analysis are: “SATIS\_1”, “RLE\_1”, “RLE\_3”, “RLE\_4”, “STS\_Q1”. Columns that were not necessary to answer the posed question were dropped. “NaN” values in the columns of interest were omitted. Data frames were created in R to observe the relationship between the following variables: overall life satisfaction ratings and recent monetary crisis, overall life satisfaction ratings and recent job change; overall satisfaction ratings and recent pay cut; recent monetary crisis and ability to handle stress ratings; job change and ability to handle stress; and pay cut and ability to handle stress. Square tile plots were then created in R to visualize each of the three relationships.

**Statistical Association of Ability to Handle Personal and Financial Crises with Recent Financial Crisis and Job Loss**

Answers to six questions related to an individual’s recent life events (RLE\_1, RLE\_3, and RLE\_4 in the ACHS dataset) and ability to handle stress related to financial crises and day-to-day demands (STS\_Q1, STS\_Q3 and STS\_Q4 in the ACHS dataset) were separated in pandas and imported into R for statistical analysis.

The respondents whose answers were not included for any question were dropped, leaving a sample of 6,789 complete responses.

RLE\_1, RLE\_3, and RLE\_4 were binary yes/no answers.

STS\_Q1 and STS\_Q3 were Likert-type answers with an ordinal scale from 1-Excellent to 5-Poor.

STS\_Q3 asked respondents what “thing” in their lives was the largest contributor to their feelings of stress. The questioners did not prompt the respondents for a predetermined set of answers but had a list of 16 categorical responses to code their answers to. As shown in the pie charts, two of the top 3 responses were “Financial Situation” and “Own Work Situation”. These answers were kept for analysis and all other answers were grouped into one category (“Other”) for the analysis.

The ordinal values for the answers to STS\_Q1 and STS\_Q2 are compared with the answers to RLE\_1, RLE\_3, RLE\_4, and STS\_Q3 in Tables 1-4 (see Appendix). These ordinal values were also compared with binary and categorical answers using an ordinal logistic regression in R. The regression output for the final models for these questions are included in the Appendix.

**Findings**

**Main Stressors Researched**

Out of 16 options about stress in day-to-day life, the top three main stressors people chose were: 1. Own work situation, 2. Other, 3. Financial situation. 20% of survey respondents ranked work-related stress, which may include long hours, heavy workload, and conflict with co-workers, as the number-one stress in life, 18% selected the “Other” answer option, and 12% of people chose financial situation. Moreover, about 8% of people did not select any answer from the list. A 6-scale Likert chart shows the percentage of each score category, including information about the ability to handle day-to-day demands and unexpected and difficult problems in their lives. For the question about the ability to handle the day-to-day demands, the most selected choice was “Very good” with 46%, and the least option chosen was “Poor” with only 2%. Less than 1% of people did not answer the question. For the question about the ability to handle unexpected and difficult problems, the most selected choice was also “Very good” with 43%, and the least selected option was “Poor” with 2%. Less than 1% of people did not answer the question. Both had a very similar trend.

**Recent Finance Related Life Crisis, ability to handle stress and overall life satisfaction**

Experiencing finance related life crisis affects people's rating of their ability to handle stress. Our findings suggest that people who have experienced finance related life crisis tend to give a lower rating to the question about ability to handle unexpected and difficult problems, such as a family or personal crisis, while individuals who have not experienced a finance related life crisis give higher ratings to this question. While 49% of individuals who experience finance related life crisis rate the questions about handling stress as excellent or good, 68% of individuals who have not experienced a finance related crisis rate ability to handle stress as good or excellent. fair instead of good. Experiencing a finance related life crisis also showed to affect people's rating of life satisfaction overall. People who experience finance related life crisis, tend to rate overall life satisfaction as lower compared to individuals who have not experienced a recent finance related life crisis. 81% of individuals who have not recently experienced a finance related crisis rate life satisfaction as 8, 9 or 10 while only 51% of individuals who have experienced a recent finance life crisis rate overall life satisfaction as 8, 9 or 10. These findings suggests that the occurrence of any finance related life crisis can severely affect individual’s stress levels and overall life satisfaction.

**Statistical Association of Ability to Handle Personal and Financial Crises with Recent Financial Crisis and Job Loss**

Recent experience with a financial crisis was found to be very influential on an individual’s perception of their ability to handle difficulties and stresses of day-to-day life.

The response to STS\_Q1 (“How would you rate your ability to handle unexpected and difficult problem?”) was found to have a significant association with whether the respondent had recently experienced a financial crisis, a negative change in work, and with whether their primary source of stress was financial or work related. The magnitudes of the effects were surprisingly different however. For those who had not experienced a crisis, a negative change in work or for whom their primary source of stress was not financial nor their work situation the cumulative odds of a rating of at least Very Good was 3.1, rating of at least Good was 0.45, rating of at least Fair was 0.1, rating of at least poor was 0.02. If they had recently experienced a financial crisis, their odds of a worse rating increased by a factor of 18.5 (1850%) [CI: 11.48, 30.2], while if they had experienced a negative change in job, the odds of a worse rating was actually reduced by a factor of 0.09 (-91%) [CI: 0.06, 0.14]. For those whose primary source of stress was their own work situation the odds of a worse rating were increased by a factor of 1.01 (1%) [CI: 0.9,1.13]. Lastly, for those whose primary source of stress was financial the odds of a worse rating increased by a factor of 1.43 (43%) [CI: 1.25, 1.65]. A reduction in pay, or a demotion, was not found to have a significant association with the responses.

The response to STS\_Q2 (“How would you rate your ability to handle the day-to-day demands in your life?”) was found to have a significant association with whether the respondent had recently experienced a financial crisis and a negative change in work. It was not found to be associated with a reduction in pay (or a demotion), nor with their primary source of stress. For those who had not experienced a crisis, nor a negative change in job, the cumulative odds of a rating of at least Very Good was 3.1, rating of at least Good was 2.5, rating of at least Fair was 0.07, rating of at least poor was 0.01. If they had recently experienced a financial crisis, their odds of a worse rating increased by a factor of 7.03 (703%) [CI: 4.72, 10.49] and if they had experienced a negative change in job, the odds of a worse rating decreased by a factor of 0.25 (-75%) [CI: 0.17,0.37].

**Limitations**

The data was a telephone conducted survey of Albertans. Phone survey is constrained by time. Telephone survey may interrupt the personal time of the respondents and thus there is a restriction on the number, length, and complexity of questions. Moreover, sometimes the pressure to provide a response may lead to errors and affect the result.

A deeper exploration of related responses may have enriched he statistical findings and could be pursued in future works.

**Conclusion**

We explored the data collected from the ACHS and visualized the provided datasets in a creative and insightful way. We discovered that the main stressor for Albertans was work-related and their ability to handle stress was perceived to be very good overall. We also found and individual’s recent experience with a finance crisis can very negatively affects their perception of their overall ability to handle stress, much more than a negative change in job situation.

Health care professionals can use this information to engage with the public to identify the main stressors that Albertans face daily, educate health behaviours, and to create stress-reducing intervention or programs to improve Albertan’s health. They can also be more informed about the impact that different life events have on individuals and use that information to proactively address and prevent negative health consequences that these events can cause. In a time of recession, these impacts could, in the worse cases, mirror those of an epidemic.

# **Reference**

Geddes, L., 2015. *Timeline: Tracking the layoffs in Alberta’s oilpatch.* [Online]   
Available at: https://globalnews.ca/news/1889598/timeline-tracking-the-layoffs-in-albertas-oilpatch/  
[Accessed April 2021].

Government of Canada, 2018. *Alberta Community Health Survey (ACHS) - Telephone conducted.* [Online]   
Available at: https://open.canada.ca/data/en/dataset/c8671772-bfb2-483f-8d31-81ac8b9b603f  
[Accessed 31 March 2021].

Shyllon, O. & Scruggs, E., 2015. *City of Calgary - Labour Market Review.* [Online]   
Available at: https://www.calgary.ca/home.html  
[Accessed 13 April 2021].

Tombe, T., 2019. *Why earnings in Alberta have been stagnant for years.* [Online]   
Available at: https://www.cbc.ca/news/canada/calgary/alberta-wages-recession-tombe-1.5229086  
[Accessed April 2021].

**Appendix**

Table 1 – RLE\_1

| Characteristic | Overall, N = 6,7891 | Financial Crisis, N = 1,4431 | No Crisis, N = 5,3461 | p-value2 |
| --- | --- | --- | --- | --- |
| STS\_Q1 |  |  |  | <0.001 |
| 1 | 1,516 (22%) | 272 (19%) | 1,244 (23%) |  |
| 2 | 2,920 (43%) | 533 (37%) | 2,387 (45%) |  |
| 3 | 1,664 (25%) | 388 (27%) | 1,276 (24%) |  |
| 4 | 531 (7.8%) | 173 (12%) | 358 (6.7%) |  |
| 5 | 158 (2.3%) | 77 (5.3%) | 81 (1.5%) |  |
| STS\_Q2 |  |  |  | <0.001 |
| 1 | 1,538 (23%) | 261 (18%) | 1,277 (24%) |  |
| 2 | 3,138 (46%) | 568 (39%) | 2,570 (48%) |  |
| 3 | 1,633 (24%) | 415 (29%) | 1,218 (23%) |  |
| 4 | 366 (5.4%) | 138 (9.6%) | 228 (4.3%) |  |
| 5 | 114 (1.7%) | 61 (4.2%) | 53 (1.0%) |  |
| 1n (%) |
| 2Pearson's Chi-squared test |

Table 2 – RLE\_3

| Characteristic | Overall, N = 6,7891 | Negative Job Change, N = 1,4221 | No Job Change, N = 5,3671 | p-value2 |
| --- | --- | --- | --- | --- |
| STS\_Q1 |  |  |  | <0.001 |
| 1 | 1,516 (22%) | 275 (19%) | 1,241 (23%) |  |
| 2 | 2,920 (43%) | 553 (39%) | 2,367 (44%) |  |
| 3 | 1,664 (25%) | 380 (27%) | 1,284 (24%) |  |
| 4 | 531 (7.8%) | 160 (11%) | 371 (6.9%) |  |
| 5 | 158 (2.3%) | 54 (3.8%) | 104 (1.9%) |  |
| STS\_Q2 |  |  |  | <0.001 |
| 1 | 1,538 (23%) | 257 (18%) | 1,281 (24%) |  |
| 2 | 3,138 (46%) | 600 (42%) | 2,538 (47%) |  |
| 3 | 1,633 (24%) | 393 (28%) | 1,240 (23%) |  |
| 4 | 366 (5.4%) | 124 (8.7%) | 242 (4.5%) |  |
| 5 | 114 (1.7%) | 48 (3.4%) | 66 (1.2%) |  |
| 1n (%) |
| 2Pearson's Chi-squared test |

Table 3 – RLE\_4

| Characteristic | Overall, N = 6,7891 | No Pay Cut, N = 5,6311 | Pay Cut, N = 1,1581 | p-value2 |
| --- | --- | --- | --- | --- |
| STS\_Q1 |  |  |  | <0.001 |
| 1 | 1,516 (22%) | 1,290 (23%) | 226 (20%) |  |
| 2 | 2,920 (43%) | 2,486 (44%) | 434 (37%) |  |
| 3 | 1,664 (25%) | 1,340 (24%) | 324 (28%) |  |
| 4 | 531 (7.8%) | 399 (7.1%) | 132 (11%) |  |
| 5 | 158 (2.3%) | 116 (2.1%) | 42 (3.6%) |  |
| STS\_Q2 |  |  |  | <0.001 |
| 1 | 1,538 (23%) | 1,323 (23%) | 215 (19%) |  |
| 2 | 3,138 (46%) | 2,653 (47%) | 485 (42%) |  |
| 3 | 1,633 (24%) | 1,311 (23%) | 322 (28%) |  |
| 4 | 366 (5.4%) | 261 (4.6%) | 105 (9.1%) |  |
| 5 | 114 (1.7%) | 83 (1.5%) | 31 (2.7%) |  |
| 1n (%) | | | | |
| 2Pearson's Chi-squared test | | | | |

Table 4 – STS\_Q4

| Characteristic | Overall, N = 6,7891 | Other, N = 4,3901 | Own Work Situation, N = 1,5261 | Financial Situation, N = 8731 |
| --- | --- | --- | --- | --- |
| STS\_Q1 |  |  |  |  |
| 1 | 1,516 (22%) | 1,003 (23%) | 344 (23%) | 169 (19%) |
| 2 | 2,920 (43%) | 1,838 (42%) | 720 (47%) | 362 (41%) |
| 3 | 1,664 (25%) | 1,076 (25%) | 360 (24%) | 228 (26%) |
| 4 | 531 (7.8%) | 361 (8.2%) | 88 (5.8%) | 82 (9.4%) |
| 5 | 158 (2.3%) | 112 (2.6%) | 14 (0.9%) | 32 (3.7%) |
| STS\_Q2 |  |  |  |  |
| 1 | 1,538 (23%) | 1,046 (24%) | 331 (22%) | 161 (18%) |
| 2 | 3,138 (46%) | 1,937 (44%) | 779 (51%) | 422 (48%) |
| 3 | 1,633 (24%) | 1,062 (24%) | 366 (24%) | 205 (23%) |
| 4 | 366 (5.4%) | 255 (5.8%) | 43 (2.8%) | 68 (7.8%) |
| 5 | 114 (1.7%) | 90 (2.1%) | 7 (0.5%) | 17 (1.9%) |
| 1n (%) |

STS\_Q1 – R Regression Output

## --------------------------------------------   
## Test for X2 df probability   
## --------------------------------------------   
## Omnibus -30.28 12 1  
## CrisisYes 1.2 3 0.75  
## JobYes 0 3 1  
## Thing1 5.3 3 0.15  
## Thing2 11.85 3 0.01  
## --------------------------------------------   
##   
## H0: Parallel Regression Assumption holds

## Warning in brant(ch.polr): 35 combinations in table(dv,ivs) do not occur.  
## Because of that, the test results might be invalid.

##   
## Re-fitting to get Hessian

## Call:  
## polr(formula = Handle ~ Crisis + Job + Thing, data = stsq1)  
##   
## Coefficients:  
## Value Std. Error t value  
## CrisisYes 2.92007 0.2465 11.848  
## JobYes -2.41736 0.2471 -9.783  
## Thing1 0.00793 0.0574 0.138  
## Thing2 0.36005 0.0712 5.060  
##   
## Intercepts:  
## Value Std. Error t value  
## 1|2 -1.116 0.038 -29.280  
## 2|3 0.792 0.037 21.631  
## 3|4 2.375 0.049 48.478  
## 4|5 3.981 0.087 45.930  
##   
## Residual Deviance: 17836.83   
## AIC: 17852.83

##   
## Re-fitting to get Hessian

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## CrisisYes 2.92007 0.24645 11.85 < 2e-16 \*\*\*  
## JobYes -2.41736 0.24710 -9.78 < 2e-16 \*\*\*  
## Thing1 0.00793 0.05736 0.14 0.89   
## Thing2 0.36005 0.07115 5.06 4.3e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Re-fitting to get Hessian

## Value Std. Error t value  
## CrisisYes 18.543 1.3 1.4e+05  
## JobYes 0.089 1.3 5.6e-05  
## Thing1 1.008 1.1 1.1e+00  
## Thing2 1.433 1.1 1.6e+02  
## 1|2 0.327 1.0 1.9e-13  
## 2|3 2.207 1.0 2.5e+09  
## 3|4 10.753 1.1 1.1e+21  
## 4|5 53.559 1.1 8.9e+19

## Waiting for profiling to be done...  
##   
## Re-fitting to get Hessian

## 2.5 % 97.5 %  
## CrisisYes 11.480 30.20  
## JobYes 0.055 0.14  
## Thing1 0.901 1.13  
## Thing2 1.247 1.65

ci <- confint(ch.polr)

## Waiting for profiling to be done...

##   
## Re-fitting to get Hessian

exp(cbind(OR=coef(ch.polr),ci))

## OR 2.5 % 97.5 %  
## CrisisYes 18.543 11.480 30.20  
## JobYes 0.089 0.055 0.14  
## Thing1 1.008 0.901 1.13  
## Thing2 1.433 1.247 1.65

Cumulative odds of rating Ability to handle unexpected and difficult problems, levels are 1=excellent to 5=poor:

exp(1.116)

## [1] 3.1

exp(-0.792)

## [1] 0.45

exp(-2.357)

## [1] 0.095

exp(-3.981)

## [1] 0.019

STS\_Q2 – R Regression Output

## --------------------------------------------   
## Test for X2 df probability   
## --------------------------------------------   
## Omnibus -123.03 6 1  
## CrisisYes -156.34 3 1  
## JobYes -143.33 3 1  
## --------------------------------------------   
##   
## H0: Parallel Regression Assumption holds

## Warning in brant(ch.polr2): 5 combinations in table(dv,ivs) do not occur.  
## Because of that, the test results might be invalid.

##   
## Re-fitting to get Hessian

## Call:  
## polr(formula = Handle ~ Crisis + Job, data = stsq2)  
##   
## Coefficients:  
## Value Std. Error t value  
## CrisisYes 1.95 0.203 9.59  
## JobYes -1.39 0.204 -6.85  
##   
## Intercepts:  
## Value Std. Error t value  
## 1|2 -1.135 0.031 -36.941  
## 2|3 0.916 0.029 31.282  
## 3|4 2.733 0.050 54.550  
## 4|5 4.246 0.096 44.116  
##   
## Residual Deviance: 16970.47   
## AIC: 16982.47

##   
## Re-fitting to get Hessian

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## CrisisYes 1.950 0.203 9.59 <2e-16 \*\*\*  
## JobYes -1.394 0.204 -6.85 8e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Re-fitting to get Hessian

## Value Std. Error t value  
## CrisisYes 7.03 1.2 1.5e+04  
## JobYes 0.25 1.2 1.1e-03  
## 1|2 0.32 1.0 9.1e-17  
## 2|3 2.50 1.0 3.8e+13  
## 3|4 15.37 1.1 4.9e+23  
## 4|5 69.81 1.1 1.4e+19

## Waiting for profiling to be done...  
##   
## Re-fitting to get Hessian

## 2.5 % 97.5 %  
## CrisisYes 4.72 10.49  
## JobYes 0.17 0.37

ci2 <- confint(ch.polr2)

## Waiting for profiling to be done...

##   
## Re-fitting to get Hessian

exp(cbind(OR=coef(ch.polr2),ci2))

## OR 2.5 % 97.5 %  
## CrisisYes 7.03 4.72 10.49  
## JobYes 0.25 0.17 0.37

Cumulative odds of rating Ability to handle the day-to-day demands of life, levels are 1=excellent to 5=poor:

exp(1.135)

## [1] 3.1

exp(0.916)

## [1] 2.5

exp(-2.733)

## [1] 0.065

exp(-4.246)

## [1] 0.014

1. 18,000 January to November (Geddes, 2015) and 2,600 in December (Shyllon & Scruggs, 2015) [↑](#footnote-ref-2)
2. Alberta Community Health Survey (ACHS) (Government of Canada, 2018) [↑](#footnote-ref-3)