

## Differences in Output

Overall, there were very few differences in output between the original code and the reproduced code despite taking two different approaches to recoding/one-hot encoding.

However, when defining partnership, the original and reproduced code differ in approach and output. Partnership, named 'partner' in reposited R code, consists of two values: 1 = have a partner, and 0 = does not have a partner. These values are derived from the original variable MARSTAT, or marital status, in the 2017 Canadian GSS codebook.

Following the removal of missing data, MARSTAT consists of several values listed below:

- 1 = Married
- 2 = Living in common-law
- 3 = Widowed
- 4 = Separated
- 5 = Divorced
- 6 = Single, never married

The original code defines 'have a partner' (1) by the condition  $\text{MARSTAT} < 2$  and defines "does not have a partner" (0) by the two conditions  $\text{MARSTAT} \geq 3 \mid \text{MARSTAT} \leq 6$ . This means that 'have a partner' consists of individuals ONLY married respondents. On the other hand, 'does not have a partner' consists of common-law, widowed, separated, divorced, or single respondents. Because common-law relationships are still partnerships, the original code defines 'partner' inaccurately. According to the assigned logic of 'partner', those living in a common-law relationship are grouped into respondents who do not have a partner. This is an easily overlook error and can be adjusted by changing the 'have a partner' condition to  $\text{MARSTAT} \leq 2$ .

The reproduced code defines 'partner' according to the corrected assigned logic above, where 'have a partner' (1) is defined by the condition  $\text{MARSTAT} \leq 2$  and 'does not have a partner' (0) by the two conditions  $\text{MARSTAT} \geq 3 \mid \text{MARSTAT} \leq 6$ .

Because of these differences, the descriptive statistics of 'partner' found in the RRWM\_DescStats.R file and corresponding .pdf files differs slightly. Therefore, the logit regression outputs of the original and reproduced code differ, where the significance of the predictor 'minority' shifts from a very significant to slightly significant predictor.

## Classification

The differences in code were identified and adjusted (see RRWM\_Reprod.R file), therefore the reproducibility of this activity is classified as **full reproduction with minor issues**.