Name: Jonathan St-Onge Course: CDAE 395 Date: 11/10/22 —

Research Question

`Broad question:` What factors influence the perceived benefits and costs of learning to program among individuals and fields that traditionally do not engage in computational methods?

Hypotheses

`Research question:` What factors relates to the perceived benefits and costs of learning to program among individuals and fields that traditionally do not engage in computational methods?

Hypothesis 1:

`H0:` Controlling for years of experience programming, there is no relationship between gender and self-identifying as a programmer (two-tailed)

`HA1:` Controlling for years of experience programming, there is a positive relationship between being a male and self identifying as a programmer (one-tailed)

Hypothesis 2:

`H0:` There is no relationship between feeling that coding skills are valued as much as other types of knowledge and academic background (STEM or Not-STEM) (two-tailed)

`HA1:` There is a relationship between feeling that coding skills are valued as much as other types of knowledge and being in STEM (one-tailed)

Hypothesis 3:

`H0:` Gender is not associated with proportion of social contacts that are likely to participate in a project that require programming in the upcoming academic year.

`HA1:` Gender is associated with proportion of social contacts that are likely to participate in a project that require programming in the upcoming academic year (two-tailed)

Hypothesis 4:

`H0:` There is no relationship between the perceived importance of sharing code associated with an academic paper and early programming experience

`HA1:` There is a relationship between the perceived importance of sharing code associated with an academic paper and early programming experience (one-tailed)

Exact variables:

`Independent variable 1:` Gender (categorical-binary)

`Independent variable 2:` Academic background (categorical-binary)

`Independent variable 3:` Experience programming before 18 (categorical-binary)

`Dependent variable 1:` Proportion of your social contacts are likely to participate in a project that require programming in the upcoming academic year (categorical-3 levels)

`Dependent variable 2:` Importance of sharing code associated with an academic paper (ordinal)

`Dependent variable 3:` Self-identifying as programmer (categorical-3 levels)

`Dependent variable 4:` How coding skills are valued compared to domain expertise? (ordinal)

Levels of measurement of variables

`Independent variable 1:` 2 levels (yes and no) Recoding will occur to map preferred pronouns to binary.

`Independent variable 2:` 2 levels (STEM and not-STEM) Recoding will occur to map department to either being in STEM or not-STEM.

`Independent variable 3:` 2 levels (Experience programming before 18 or not) Recoding will occur to map various years of experience, which was an ordinal variable with 5 levels, to either coding experience before 18 or not. We ignore "Prefer not to say" variable.

`Dependent variable 1:` Proportion (categorical-3 levels; buckets of $1/3 < 2/3 < 1$). Recoding will occur to map the proportion to smaller buckets.

`Dependent variable 2:` 4 levels (important, somewhat important, not important, and don't know). These will not be recoded, we interpret not knowing as not finding it important to share code online.

`Dependent variable 3:` 3 levels (yes, no, maybe). These will not be recoded, we interpret the maybe response as valid response, and "don't know" and "refused" will be excluded.

`Dependent variable 4:` 4 levels (as valued as my research ability, less valued than my research capacity, more valued than my research ability, not valued at all compared to my research ability). These will not be recoded, and "don't know" and "refused" will be excluded.

Hypothesis test:

`Hyp test 1`: Chi-squared test

`Hyp test 2`: Wilcoxon-Mann Whitney test

`Hyp test 3`: Wilcoxon-Mann Whitney test

`Hyp test 4`: Wilcoxon-Mann Whitney test