data_w4

Celina Jang

2025-07-01

Week 4

```
chsw3 <- read.csv('chs2020_working_w3.csv')

chsw3 <- chsw3 |> mutate(
  # change binary variable values to 1 and 0
  delaypayrent0 = case_when(
    delaypayrent == 1 ~ 0,
    delaypayrent == 2 ~ 1,
    is.na(delaypayrent) ~ NA),
  didntgetcare0 = case_when(
    didntgetcare20 == 1 ~ 0,
    didntgetcare20 == 2 ~ 1,
    is.na(didntgetcare20) ~ NA),
  nspd0 = case_when(
    nspd == 1 ~ 1,
    nspd == 2 ~ 0,
    is.na(nspd) ~ NA
  ),
  rodentsstreet0 = case_when(
    rodentsstreet==1 ~ 1,
    rodentsstreet == 2 ~ 0,
    is.na(rodentsstreet) ~ NA
  ),
  # want reference group to be white, so reorder
  race_ethnicity = fct_relevel(race_ethnicity, 'White'),
  # label employment20
  employment = case_when(
        employment20 == 1 ~ 'Employed for wages or salary',
        employment20 == 2 ~ 'Self-employed',
        employment20 == 3 ~ 'Unemployed for 1 year or more',
        employment20 == 4 ~ 'Unemployed for less than 1 year',
        employment20 == 5 ~ 'A homemaker',
        employment20 == 6 ~ 'A student',
        employment20 == 7 ~ 'Retired',
        employment20 == 8 ~ 'Unable to work',
        employment20 == '.d' ~ 'Dont know',
        employment20 == '.r' ~ 'Refused',
        is.na(employment20) ~ NA
))
```

```r
# linear regression (unweighted)
k6.fit.lm <- lm(k6 ~ social_cohesion_rev, data=chsw3)
summary(k6.fit.lm)
```

```
##
## Call:
## lm(formula = k6 ~ social_cohesion_rev, data = chsw3)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.415 -3.359 -1.124  1.994 20.463
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.00226    0.27653  21.705   <2e-16 ***
## social_cohesion_rev -0.58701    0.09206  -6.376    2e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.309 on 4328 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.009307,   Adjusted R-squared:  0.009078
## F-statistic: 40.66 on 1 and 4328 DF,  p-value: 2.004e-10
```

```r
# logistic regression (unweighted)
nspd.fit.lg <- glm(nspd ~ social_cohesion_rev, data=chsw3)
summary(nspd.fit.lg)
```

```
##
## Call:
## glm(formula = nspd ~ social_cohesion_rev, data = chsw3)
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.878275   0.015276 122.956  < 2e-16 ***
## social_cohesion_rev 0.020976   0.005086   4.125 3.78e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.05665044)
##
##     Null deviance: 246.15  on 4329  degrees of freedom
## Residual deviance: 245.18  on 4328  degrees of freedom
##   (6 observations deleted due to missingness)
## AIC: -138.8
##
## Number of Fisher Scoring iterations: 2
```

```r
exp(coef(summary(nspd.fit.lg))[2, "Estimate"])
```

```
## [1] 1.021198
```

```r
# social cohesion by age
sc_age_lm <- lm(social_cohesion_rev ~ age_band, data = chsw3)
anova(sc_age_lm)
```

```
## Analysis of Variance Table
##
## Response: social_cohesion_rev
##             Df Sum Sq Mean Sq F value    Pr(>F)
## age_band     2   71.5  35.751  73.001 < 2.2e-16 ***
## Residuals 4318 2114.7   0.490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Weighted analysis

```
library(survey)
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
# Setting the weights
chs2020_svy <- svydesign(ids = ~1, strata = ~strata_q1, weights = ~wt21_dual_q1, data = chsw3)

# multiple regression with continuous kessler (mental distress)
svy_lm_k6 <- svyglm(k6 ~ social_cohesion_rev + age_band + gender + race_ethnicity +
                    education + employment + delaypayrent0 + rodentsstreet0,
                    design = chs2020_svy)
summary(svy_lm_k6)
```

```
##
## Call:
## svyglm(formula = k6 ~ social_cohesion_rev + age_band + gender +
##     race_ethnicity + education + employment + delaypayrent0 +
##     rodentsstreet0, design = chs2020_svy)
##
## Survey design:
## svydesign(ids = ~1, strata = ~strata_q1, weights = ~wt21_dual_q1,
##     data = chsw3)
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            6.48765    0.70943   9.145  < 2e-16
## social_cohesion_rev                   -0.34938    0.13998  -2.496 0.012600
## age_band45-64                         -0.91928    0.21745  -4.227 2.42e-05
```

```
## age_band65+                                -1.67436    0.35497  -4.717 2.47e-06
## gendermale                                 -0.79051    0.19277  -4.101 4.20e-05
## race_ethnicityAsian/Pacific Islander       -0.71869    0.32110  -2.238 0.025260
## race_ethnicityBlack                        -0.95490    0.27387  -3.487 0.000494
## race_ethnicityHispanic                     -0.42690    0.26652  -1.602 0.109284
## race_ethnicityNorth African/Mid Eastern     1.32593    0.91503   1.449 0.147397
## race_ethnicityOther                         0.96951    0.76695   1.264 0.206262
## education                                  -0.01177    0.09459  -0.124 0.901003
## employmentA student                         1.77358    0.63468   2.794 0.005223
## employmentEmployed for wages or salary      0.98004    0.40407   2.425 0.015334
## employmentRetired                           2.12296    0.51004   4.162 3.22e-05
## employmentSelf-employed                     1.06854    0.49761   2.147 0.031824
## employmentUnable to work                    3.85749    0.61886   6.233 5.03e-10
## employmentUnemployed for 1 year or more     1.49375    0.57259   2.609 0.009120
## employmentUnemployed for less than 1 year   2.05787    0.50108   4.107 4.09e-05
## delaypayrent0                              -1.92756    0.33411  -5.769 8.55e-09
## rodentsstreet0                              0.92688    0.21440   4.323 1.58e-05
##
## (Intercept)                                ***
## social_cohesion_rev                        *
## age_band45-64                              ***
## age_band65+                                ***
## gendermale                                 ***
## race_ethnicityAsian/Pacific Islander       *
## race_ethnicityBlack                        ***
## race_ethnicityHispanic
## race_ethnicityNorth African/Mid Eastern
## race_ethnicityOther
## education
## employmentA student                        **
## employmentEmployed for wages or salary     *
## employmentRetired                          ***
## employmentSelf-employed                    *
## employmentUnable to work                   ***
## employmentUnemployed for 1 year or more    **
## employmentUnemployed for less than 1 year  ***
## delaypayrent0                              ***
## rodentsstreet0                             ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 16.4194)
##
## Number of Fisher Scoring iterations: 2
```

```
# multiple regression with binary psychological distress
svy_glm_nspd <- svyglm(nspd0 ~ social_cohesion_rev + age_band + gender + race_ethnicity +
                       education + employment + delaypayrent0 + rodentsstreet0,
                       design = chs2020_svy, family = quasibinomial())
summary(svy_glm_nspd)
```

```
##
## Call:
## svyglm(formula = nspd0 ~ social_cohesion_rev + age_band + gender +
##     race_ethnicity + education + employment + delaypayrent0 +
```

```
##     rodentsstreet0, design = chs2020_svy, family = quasibinomial())
##
## Survey design:
## svydesign(ids = ~1, strata = ~strata_q1, weights = ~wt21_dual_q1,
##     data = chsw3)
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           -2.69328    0.72324  -3.724 0.000199
## social_cohesion_rev                   -0.12984    0.12170  -1.067 0.286088
## age_band45-64                         -0.26768    0.23779  -1.126 0.260357
## age_band65+                           -0.65423    0.38318  -1.707 0.087830
## gendermale                            -0.46312    0.19522  -2.372 0.017726
## race_ethnicityAsian/Pacific Islander   0.09247    0.38305   0.241 0.809251
## race_ethnicityBlack                   -0.28917    0.32759  -0.883 0.377444
## race_ethnicityHispanic                 0.02006    0.28124   0.071 0.943132
## race_ethnicityNorth African/Mid Eastern  1.03498  0.72090   1.436 0.151168
## race_ethnicityOther                    0.65424    0.52003   1.258 0.208434
## education                              0.06809    0.09452   0.720 0.471359
## employmentA student                    1.40703    0.69911   2.013 0.044221
## employmentEmployed for wages or salary 0.59130    0.58296   1.014 0.310494
## employmentRetired                      1.56403    0.66503   2.352 0.018729
## employmentSelf-employed                0.88500    0.67593   1.309 0.190503
## employmentUnable to work               2.49411    0.59675   4.179 2.98e-05
## employmentUnemployed for 1 year or more  0.71237  0.72358   0.984 0.324930
## employmentUnemployed for less than 1 year  1.50127  0.59545  2.521 0.011732
## delaypayrent0                         -1.00706    0.24232  -4.156 3.31e-05
## rodentsstreet0                         0.49222    0.21287   2.312 0.020814
##
## (Intercept)                           ***
## social_cohesion_rev
## age_band45-64
## age_band65+                           .
## gendermale                            *
## race_ethnicityAsian/Pacific Islander
## race_ethnicityBlack
## race_ethnicityHispanic
## race_ethnicityNorth African/Mid Eastern
## race_ethnicityOther
## education
## employmentA student                   *
## employmentEmployed for wages or salary
## employmentRetired                     *
## employmentSelf-employed
## employmentUnable to work              ***
## employmentUnemployed for 1 year or more
## employmentUnemployed for less than 1 year *
## delaypayrent0                         ***
## rodentsstreet0                        *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.9558655)
##
```

```
## Number of Fisher Scoring iterations: 6
```

```
# interaction term (cohesion x age)
svy_lm_k6_int <- svyglm(k6 ~ social_cohesion_rev * age_band + gender + race_ethnicity +
                          education + employment20 + delaypayrent + rodentsstreet,
                        design = chs2020_svy)
summary(svy_lm_k6_int)
```

```
##
## Call:
## svyglm(formula = k6 ~ social_cohesion_rev * age_band + gender +
##     race_ethnicity + education + employment20 + delaypayrent +
##     rodentsstreet, design = chs2020_svy)
##
## Survey design:
## svydesign(ids = ~1, strata = ~strata_q1, weights = ~wt21_dual_q1,
##     data = chsw3)
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             10.735152   0.926029  11.593  < 2e-16
## social_cohesion_rev                     -0.258015   0.198750  -1.298 0.194297
## age_band45-64                           -0.565596   0.980749  -0.577 0.564175
## age_band65+                             -0.074662   1.229824  -0.061 0.951593
## gendermale                              -0.627585   0.187720  -3.343 0.000836
## race_ethnicityAsian/Pacific Islander    -0.816061   0.332641  -2.453 0.014198
## race_ethnicityBlack                     -0.875088   0.276625  -3.163 0.001571
## race_ethnicityHispanic                  -0.439011   0.268199  -1.637 0.101731
## race_ethnicityNorth African/Mid Eastern  1.119137   0.932915   1.200 0.230360
## race_ethnicityOther                      1.071374   0.796098   1.346 0.178448
## education                               -0.002875   0.093827  -0.031 0.975559
## employment20                             0.258765   0.047821   5.411 6.62e-08
## delaypayrent                            -2.031577   0.327011  -6.213 5.74e-10
## rodentsstreet                           -0.934673   0.215670  -4.334 1.50e-05
## social_cohesion_rev:age_band45-64       -0.088322   0.322406  -0.274 0.784140
## social_cohesion_rev:age_band65+         -0.541771   0.375389  -1.443 0.149033
##
## (Intercept)                             ***
## social_cohesion_rev
## age_band45-64
## age_band65+
## gendermale                              ***
## race_ethnicityAsian/Pacific Islander    *
## race_ethnicityBlack                     **
## race_ethnicityHispanic
## race_ethnicityNorth African/Mid Eastern
## race_ethnicityOther
## education
## employment20                            ***
## delaypayrent                            ***
## rodentsstreet                           ***
## social_cohesion_rev:age_band45-64
## social_cohesion_rev:age_band65+
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for gaussian family taken to be 16.67039)
##
## Number of Fisher Scoring iterations: 2
```

Visualizing regression results

```r
# tidy the model
tidy_model <- broom::tidy(svy_lm_k6)

# filter for significant variables (p < 0.05)
signif_vars <- tidy_model |>
  filter(p.value < 0.05) |>
  mutate(across(where(is.numeric), ~ round(., 3))) |>
  rename(
    Variable = term,
    Estimate = estimate,
    "Std. Error" = std.error,
    "p-value" = p.value,
    "t value" = statistic
  ) |>
  select(Variable, Estimate, 'Std. Error', 't value', "p-value")

# create the table
signif_vars |>
  kable(caption = "Significant Predictors of K6 (Survey-Weighted Linear Regression)", escape=T)
```

Table 1: Significant Predictors of K6 (Survey-Weighted Linear Regression)

Variable	Estimate	Std. Error	t value	p-value
(Intercept)	6.488	0.709	9.145	0.000
social_cohesion_rev	-0.349	0.140	-2.496	0.013
age_band45-64	-0.919	0.217	-4.227	0.000
age_band65+	-1.674	0.355	-4.717	0.000
gendermale	-0.791	0.193	-4.101	0.000
race_ethnicityAsian/Pacific Islander	-0.719	0.321	-2.238	0.025
race_ethnicityBlack	-0.955	0.274	-3.487	0.000
employmentA student	1.774	0.635	2.794	0.005
employmentEmployed for wages or salary	0.980	0.404	2.425	0.015
employmentRetired	2.123	0.510	4.162	0.000
employmentSelf-employed	1.069	0.498	2.147	0.032
employmentUnable to work	3.857	0.619	6.233	0.000
employmentUnemployed for 1 year or more	1.494	0.573	2.609	0.009
employmentUnemployed for less than 1 year	2.058	0.501	4.107	0.000
delaypayrent0	-1.928	0.334	-5.769	0.000
rodentsstreet0	0.927	0.214	4.323	0.000

```r
# Step 1: Tidy the model
log_table <- tidy(svy_glm_nspd)

# Step 2: Filter significant variables (before modifying p-value format)
```

```r
significant_terms <- log_table %>%
  filter(p.value < 0.05 & term != "(Intercept)")

# Step 3: Add formatting (after filtering)
significant_terms <- significant_terms %>%
  mutate(
    "p-value" = ifelse(p.value < 1e-4, "<0.0001", round(p.value, 4)),
    "Odds Ratio" = round(exp(estimate), 3),
    Estimate = round(estimate, 3),
    "Std. Error" = round(std.error, 3)
  ) %>%
  select(Term = term, Estimate, "Std. Error", "Odds Ratio", "p-value")


# Display full table or significant terms only
significant_terms %>%
  kable(caption = "Significant Predictors of Psychological Distress (nspd)",
        escape = T, align = "lcccc")
```

Table 2: Significant Predictors of Psychological Distress (nspd)

Term	Estimate	Std. Error	Odds Ratio	p-value
gendermale	-0.463	0.195	0.629	0.0177
employmentA student	1.407	0.699	4.084	0.0442
employmentRetired	1.564	0.665	4.778	0.0187
employmentUnable to work	2.494	0.597	12.111	<0.0001
employmentUnemployed for less than 1 year	1.501	0.595	4.487	0.0117
delaypayrent0	-1.007	0.242	0.365	<0.0001
rodentsstreet0	0.492	0.213	1.636	0.0208

```r
# tidy table
reg_table <- tidy(svy_lm_k6_int)

# select significant variables
reg_table_signif <- reg_table |>
  filter(p.value < 0.05) |>
  select(
    Term = term,
    Estimate = estimate,
    "Std. Error" = std.error,
    "t value" = statistic,
    "p-value" = p.value
  ) |>
  mutate(across(where(is.numeric), ~ round(., 3)))

# put into table
reg_table_signif |>
  kable(caption = "Significant Predictors in Cohesion X Age", align = "lcccc")
```

Table 3: Significant Predictors in Cohesion X Age

Term	Estimate	Std. Error	t value	p-value
(Intercept)	10.735	0.926	11.593	0.000
gendermale	-0.628	0.188	-3.343	0.001
race_ethnicityAsian/Pacific Islander	-0.816	0.333	-2.453	0.014
race_ethnicityBlack	-0.875	0.277	-3.163	0.002
employment20	0.259	0.048	5.411	0.000
delaypayrent	-2.032	0.327	-6.213	0.000
rodentsstreet	-0.935	0.216	-4.334	0.000