Assignment 4

Huong Tran

11/24/2021

```
data <- read.table("cap3.txt", header = T)
attach(data)
n <- nrow(data)
library(survey)
```

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

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

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

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

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

1. Compute an estimate of population mean for variable "growth" along with the standard error.

```
fpc <- 1:n
fpc[stype == 'E'] <- 4421
fpc[stype == 'M'] <- 1018
fpc[stype == 'H'] <- 755

sv <- svydesign(id=~ dnum, strata = stype, data = data, fpc =~ fpc, nest = TRUE)
svymean(~growth, sv)
```

```
##          mean     SE
## growth 35.875 2.7662
```

b. Regression analysis to assess if there is a linear relationship between the "growth" and "awards,"meal" and "colgrad".

```
summary(svyglm(growth ~ awards + meals + colgrad, design = sv))
```

```
##
## Call:
## svyglm(formula = growth ~ awards + meals + colgrad, design = sv)
##
## Survey design:
## svydesign(id = ~dnum, strata = stype, data = data, fpc = ~fpc,
##     nest = TRUE)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.83294    5.66469  -0.677 0.503997
```

```
## awardsYes   34.27696    4.21659   8.129 5.79e-09 ***
## meals        0.30672    0.08024   3.822 0.000647 ***
## colgrad      0.05256    0.14454   0.364 0.718753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 533.8214)
##
## Number of Fisher Scoring iterations: 2
```

Based on P-value, we conclude that eligibility for awards and the percentage of students eligible for subsidized meals can help predict the change in academic performacne index from 1999 to 2000. c.

```
summary(svyglm(I(grsign == 1) ~awards + meals + colgrad , design = sv, family=binomial(link = "logit")))
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
##
## Call:
## svyglm(formula = I(grsign == 1) ~ awards + meals + colgrad, design = sv,
##     family = binomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~dnum, strata = stype, data = data, fpc = ~fpc,
##     nest = TRUE)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.031537   0.500851   2.060   0.0485 *
## awardsYes   19.485105   0.391997  49.707   <2e-16 ***
## meals        0.003852   0.008405   0.458   0.6502
## colgrad     -0.007620   0.027345  -0.279   0.7825
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 0.3150747)
##
## Number of Fisher Scoring iterations: 19
```

Using p-value, we conclude that there is relationship between the eligibility for award and "grsign".

d.

```
svytable(~grsign + awards, sv)
```

```
##       awards
## grsign       No      Yes
##      0  813.775    0.000
##      1 2365.625 6989.633
```

```
svychisq(~grsign + awards, sv,statistic = "Wald")
```

```
##
##  Design-based Wald test of association
##
## data:  svychisq(~grsign + awards, sv, statistic = "Wald")
## F = 13.224, ndf = 1, ddf = 32, p-value = 0.0009608
```

Using Wald test, we obtain small p-value, which indicate the relationship between “grsign” and “awards”, and this result agrees with the conclusion in part c.