ETC 2420/5242 Lab 7 2016

Di Cook

SOLUTION

Model building will be done using:

- Response: `math` (standardised)
- Explanatory variables: `ST04Q01, ST06Q01, ST15Q01, ST19Q01, ST26Q01, ST26Q04, ST26Q06, ST27Q02, ST28Q01`.

Question 1 (3 pts)

a. Compute and report the 95% confidence interval for the parameter for the number of books in the household (ST28Q01), using classical t-interval methods.

The 95% confidence interval is (0.201 , 0.225)

b. Use this to test the hypothesis that ST28Q01 is not important for the model.

`0 is outside the interval, so it is not a plausible value for the parameter.` H_o `would be rejected, and the conclusion would be that the parameter is not 0.`

Question 2 (3 pts)

a. The boot package can generate bootstrap samples for weighted data. To use the `boot` function for drawing samples, you need a function to compute the statistic of interest. Write the function to return the slope for ST28Q01 after fitting a `glm` to a bootstrap sample. The skeleton of the function `calc_stat` is below, where `d` is the data, and `i` is the vector of indices of the bootstrap sample.

```
library(boot)
calc_stat <- function(d, i) {
  x <- d[i,]
  mod <- glm(math_std~ST04Q01+ST06Q01+ST15Q01+ST19Q01+ST26Q01+
               ST26Q04+ST26Q06+ST27Q02+ST28Q01, data=x,
             weights=SENWGT_STU)
  stat <- as.numeric(coefficients(mod)[14])
  return(stat)
}
stat <- boot(aus_nomiss, statistic=calc_stat, R=1000,
     weights=aus_nomiss$SENWGT_STU)
stat
#
# WEIGHTED BOOTSTRAP
#
#
# Call:
# boot(data = aus_nomiss, statistic = calc_stat, R = 1000, weights = aus_nomiss$SENWGT_STU)
#
```

```
#
# Bootstrap Statistics :
#     original   bias    std. error  mean(t*)
# t1*    0.213 -0.00507     0.00656     0.208
```

The 95% confidence interval is (0.195 , 0.221).

b. How does the bootstrap interval compare with the t-interval? `The bootstrap interval is considerably wider than the classical t-interval.`

Question 3 (2 pts)

Now make a 95% bootstrap confidence interval for predicted value for a new student who is FEMALE, started school at 4, mother and father both work full-time, has a desk, computer and internet, two TVs and 26-100 books in the home. The weight for a student like this is 0.1041. Be sure to convert the values back into the actual math score range.

```
calc_pred <- function(d, i, newd) {
  x <- d[i,]
  mod <- glm(math_std~ST04Q01+ST06Q01+ST15Q01+ST19Q01+ST26Q01+
               ST26Q04+ST26Q06+ST27Q02+ST28Q01, data=x,
             weights=SENWGT_STU)
  pred <- predict(mod, newd)
  return(pred)
}
new_data <- data.frame(ST04Q01=1, ST06Q01=0, ST15Q01=1, ST19Q01=1,
        ST26Q01=1, ST26Q04=1, ST26Q06=1, ST27Q02=3, ST28Q01=3,
        math_std=0, SENWGT_STU=0.1041)
new_data$ST04Q01 <- factor(new_data$ST04Q01, levels=c(1,2))
new_data$ST15Q01 <- factor(new_data$ST15Q01, levels=c(1,2,3,4))
new_data$ST19Q01 <- factor(new_data$ST19Q01, levels=c(1,2,3,4))
new_data$ST26Q01 <- factor(new_data$ST26Q01, levels=c(1,2))
new_data$ST26Q04 <- factor(new_data$ST26Q04, levels=c(1,2))
new_data$ST26Q06 <- factor(new_data$ST26Q06, levels=c(1,2))
pred <- boot(aus_nomiss, statistic=calc_pred, R=1000,
     weights=aus_nomiss$SENWGT_STU, newd=new_data)
pred
#
# WEIGHTED BOOTSTRAP
#
#
# Call:
# boot(data = aus_nomiss, statistic = calc_pred, R = 1000, weights = aus_nomiss$SENWGT_STU,
#     newd = new_data)
#
#
# Bootstrap Statistics :
#     original  bias    std. error  mean(t*)
# t1*    0.154  0.0314      0.0241     0.186
sort(pred$t)[25]
# [1] 0.138
sort(pred$t)[975]
# [1] 0.236
```

The 95% confidence interval for the predicted value is (518.368 , 527.391).

Question 4 (2 pts)

Compute a bootstrap 95% prediction interval for the same student as in the previous question. Be sure to convert the values back into the actual math score range.

```
calc_res <- function(d, i) {
  x <- d[i,]
  mod <- glm(math_std~ST04Q01+ST06Q01+ST15Q01+ST19Q01+ST26Q01+
               ST26Q04+ST26Q06+ST27Q02+ST28Q01, data=x,
             weights=SENWGT_STU)
  res <- residuals(mod)
  return(c(l=min(res), u=max(res)))
}
res <- boot(aus_nomiss, statistic=calc_res, R=1000,
     weights=aus_nomiss$SENWGT_STU)
l <- sort(res$t[,1])[25]
u <- sort(res$t[,2])[975]
new_data <- data.frame(ST04Q01=1, ST06Q01=0, ST15Q01=1, ST19Q01=1,
        ST26Q01=1, ST26Q04=1, ST26Q06=1, ST27Q02=3, ST28Q01=3,
        math_std=0, SENWGT_STU=0.1041)
new_data$ST04Q01 <- factor(new_data$ST04Q01, levels=c(1,2))
new_data$ST15Q01 <- factor(new_data$ST15Q01, levels=c(1,2,3,4))
new_data$ST19Q01 <- factor(new_data$ST19Q01, levels=c(1,2,3,4))
new_data$ST26Q01 <- factor(new_data$ST26Q01, levels=c(1,2))
new_data$ST26Q04 <- factor(new_data$ST26Q04, levels=c(1,2))
new_data$ST26Q06 <- factor(new_data$ST26Q06, levels=c(1,2))
pred <- predict(aus_glm, new_data)
```

The 95% prediction interval is (412.921 , 606.29)