ETC 2420/5242 Lab 7 2017

Di Cook

SOLUTION

```
#
# Call:
# glm(formula = science_std ~ gender + JOYSCIE + nbooks + ntvs,
#     data = aus_nomiss, weights = SENWT)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -2.4275  -0.2905  -0.0084   0.2845   1.9705
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) -0.32446    0.04307   -7.53  5.3e-14 ***
# genderm      0.05295    0.01530    3.46  0.00054 ***
# JOYSCIE      0.27737    0.00664   41.74  < 2e-16 ***
# nbooks       0.21646    0.00547   39.58  < 2e-16 ***
# ntvs        -0.11374    0.01072  -10.61  < 2e-16 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for gaussian family taken to be 0.254)
#
#     Null deviance: 4231.3  on 12355  degrees of freedom
# Residual deviance: 3139.9  on 12351  degrees of freedom
# AIC: 35006
#
# Number of Fisher Scoring iterations: 2
```

Question 1 (3 pts)

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

The 95% confidence interval is (0.206 , 0.227)

b. Use this to test the hypothesis that `nbooks` is not important for the model. Write down the null and alternative hypotheses you are using.

$H_o : \beta_4 = 0$ vs $H_o : \beta_4 \neq 0$

`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 `nbooks` 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.

```r
library(boot)
calc_stat <- function(d, i) {
  x <- d[i,]
  mod <- glm(science_std~gender+JOYSCIE+nbooks+ntvs, data=x, weights=SENWT)
  stat <- as.numeric(coefficients(mod)[4])
  return(stat)
}
stat <- boot(aus_nomiss, statistic=calc_stat, R=1000,
     weights=aus_nomiss$SENWT)
stat
#
# WEIGHTED BOOTSTRAP
#
#
# Call:
# boot(data = aus_nomiss, statistic = calc_stat, R = 1000, weights = aus_nomiss$SENWT)
#
#
# Bootstrap Statistics :
#     original   bias    std. error  mean(t*)
# t1*    0.216 -0.00101     0.00603     0.215
```

The 95% confidence interval is (0.205 , 0.227).

b. How does the bootstrap interval compare with the t-interval? Compare and constrast them. `The two intervals are virtually identical.`

Question 3 (3 pts)

Now make a 95% bootstrap confidence interval for predicted value for a girl, enjoys science (JOYSCIE=1.0) two TVs and 26-100 books in the home. The weight for a student like this is 0.2. Be sure to convert the values back into the actual science score range.

```r
calc_pred <- function(d, i, newd) {
  x <- d[i,]
  mod <- glm(science_std~gender+JOYSCIE+nbooks+ntvs, data=x,
             weights=SENWT)
  pred <- predict(mod, newd)
  return(pred)
}
new_data <- data.frame(gender=1, JOYSCIE=1.0, nbooks=3, ntvs=3, SENWT=0.2)
new_data$gender <- factor(new_data$gender, levels=c(1,2), labels=c("f","m"))
pred <- boot(aus_nomiss, statistic=calc_pred, R=1000,
     weights=aus_nomiss$SENWT, newd=new_data)
pred
#
# WEIGHTED BOOTSTRAP
#
#
# Call:
# boot(data = aus_nomiss, statistic = calc_pred, R = 1000, weights = aus_nomiss$SENWT,
#     newd = new_data)
#
#
```

```
# Bootstrap Statistics :
#     original  bias    std. error  mean(t*)
# t1*    0.261  0.0103      0.0145     0.271
sort(pred$t)[25]
# [1] 0.243
sort(pred$t)[975]
# [1] 0.299
```

The 95% confidence interval for the predicted value is (529.862 , 535.455).

Question 4 (3 pts)

Do the same 95% bootstrap confidence interval for predicted value for a boy, all other values remain the same. Compare and contrast the intervals for the girl and boy, on the raw science score scale.

```
#
# WEIGHTED BOOTSTRAP
#
#
# Call:
# boot(data = aus_nomiss, statistic = calc_pred, R = 1000, weights = aus_nomiss$SENWT,
#     newd = new_data)
#
#
# Bootstrap Statistics :
#     original  bias    std. error  mean(t*)
# t1*    0.314   0.011      0.0155     0.325
# [1] 0.294
# [1] 0.353
```

The 95% confidence interval for the predicted value is (534.95 , 540.842).

The confidence interval for the predicted score for a girl was shifted down by five points from the boys. That is, an average boy with the same demographics would be expected to score five points more on science than the average girl.

Question 5 (3 pts)

Compute a bootstrap 95% prediction interval for the girl as in question 3. Be sure to convert the values back into the actual science score range.

```
calc_res <- function(d, i) {
  x <- d[i,]
  mod <- glm(science_std~gender+JOYSCIE+nbooks+ntvs, data=x,
             weights=SENWT)
  res <- residuals(mod)
  return(c(l=min(res), u=max(res)))
}
res <- boot(aus_nomiss, statistic=calc_res, R=1000,
     weights=aus_nomiss$SENWT)
l <- sort(res$t[,1])[25]
u <- sort(res$t[,2])[975]
new_data <- data.frame(gender=1, JOYSCIE=1.0, nbooks=3, ntvs=3, SENWT=0.2)
```

```r
new_data$gender <- factor(new_data$gender, levels=c(1,2), labels=c("f","m"))
pred <- predict(aus_glm, new_data)
```

The 95% prediction interval is (285.162 , 729.347)