Homework 2

Xinyi Lin

9/20/2019

Question 1

Problem a

```
# input data
data1 = data.frame("Class" = 1:25, "mean score" = c(51.3, 52.1, 59.6, 46.0,
53.3, 55.5, 59.5, 52.8, 51.6, 45.3, 54.0, 39.4, 54.3, 49.5, 52.4, 50.7, 52.9,
49.1, 49.0, 54.4, 50.0,46.8, 50.7, 50.5, 56.1))

y_bar = mean(data1$mean.score)
y_bar

## [1] 51.472

f_alpha = 25/108
var_y = (1-f_alpha)*var(data1$mean.score)/25
marg = round(qt(0.975, 24)*sqrt(var_y),3)
marg

## [1] 1.576
```

By using R, we can get the estimation of the average score is 51.472. The margin of error is 1.576.

Problem b

```
var_SRS = 0.4
# design effect
deff = round(var_y/var_SRS, 3)
deff

## [1] 1.457

# rate of homogeneity
n0 = 30
roh = round((deff-1)/(n0-1), 3)
roh

## [1] 0.016
```

The design effect is 1.457 and the rate of homogeneity is 0.016.

Problem c

```
deff_new = 1+(40-1)*roh
varSRS_new = var_y/deff_new
s2 = 0.4*750/(1-750/3240)
nSRS = 1*8000/(varSRS_new*8000/s2+1)
nSRS
```

```
## [1] 957.4035
```

```
n_new = nSRS*(1+39*roh)
n_new
```

```
## [1] 1554.823
```

```
k = n_new/40
k
```

```
## [1] 38.87058
```

$$var_{new,SRS}(\overline{y}) = \frac{1 - n_{SRS}/8000}{n_{SRS}} S^2$$

$$var_{SRS}(\overline{y}) = \frac{1 - 750/3240}{750} S^2$$

$$n_{SRS} \approx 958$$

$$n_{new} = (1 + (n_o - 1)roh)n_{SRS} \approx 1555$$

$$k = n_{new}/40 \approx 39$$

So 39 classes need to be selected.

Question 2

```
data2 = data.frame("Job"=c("A", "AR", "NA"), "Mean"=c(7.63,7.74,6.55),
"Standard error"=c(0.15, 0.35, 0.11), "sample size"=c(1347, 163, 1095),
"percentage" = c(0.5, 0.1, 0.4))
mean = sum(data2$Mean*data2$percentage)
mean
```

```
## [1] 7.209
```

```
var_ybar = sum(data2$percentage^2*data2$Standard.error^2)
low_q = round(mean - qt(0.975, sum(data2$sample.size)-3)*sqrt(var_ybar), 3)
low_q
```

```
## [1] 7.025
```

```
high_q = round(mean + qt(0.975, sum(data2$sample.size)-3)*sqrt(var_ybar), 3)
high_q
```

```
## [1] 7.393
```

The mean time for the population is 7.209, and 95% confidence interval is (7.025, 7.393).

Question 3

Problem a

The target population is the noninstitutinalized civilian resident U.S. population. It excludes all person in supervised care or custody in institutional settings, all active-duty military personnel, active-duty family members living overseas, and any other U.S. citizens residing outside of the 50 states and District of Columbia.

Problem b

1. The first stage consisted of selecting PSUs from a frame of all U.S. counties.
2. The second stage of selection for the NHANES 2011-2014 sample included a sample of area segments, comprising census blocks or combinations of blocks.
3. The third stage consisted of dwelling units(DUs), including noninstitutional group quarters such as dormitories.
4. The forth stage consisted of persons within occupied DUs, or households.

Problem c

From the approximately 3100 counties and county equivalents in the United States, 2846 PSUs were formed(most of which consisted of individual counties), a sample of 60 locations was selected and 15 of these locations per year were randomly allocated to each of the years. A total of 8 study locations in the full NHANES 2011-2014 out of the 60-location sample were assigned to certainty PSUs. These locations were in six counties; one county contained multiple study locations.