İST-309-02 ÖRNEKLEME

Tabakalı Rastgele Örnekleme

Ramazan Erduran

21821809
ramazan.erduran@hacettepe.edu.tr

İçindekiler Tablosu:

1. Kısım:

(d)

Veri setimin kaynağı: datacamp.com

Kitlemin ismi: young_people_survey_response

Kitlemin büyüklüğü: $1007 = N$

Kitlemin birimi: Puan

Örneklem birimi: Puan

Tahmini yapılacak değişken: Mathematics

2. Kısım:

(2) → Örneklem tablolarından yararlanarak ($d = 0,04$ ve $z_{0,05} \cong 2$)

için $n = 375$ örneklemde karar kıldım.

→ Tabakalar arası birim başına yapılacak giderler

aynı olduğu için "Neyman değişimi" ni kullandım.

→ Tabakalarım "village" ve "city"

$$n_{village} = \frac{375 \cdot 408,268653}{1364,028724} \cong 112$$

$$n_{city} = \frac{375 \cdot 955,760071}{1364,028724} \cong 263$$

3. Kısım:

(3) Tabaka	N_H	$S_H \cong$	$N_H . S_H \cong$	n_H	W_h
(1) Village	299	1,36545	408,2685	112	0,297
(2) City	707	1,35185	955,7600	263	0,703
Toplam	1006	2,7173	1364,0285	375	

Ortalama $*$ $\;\hookrightarrow\;$

$$\bar{y}_{tb} = \sum W_h . \bar{y}_h$$

$$= 0,297 . 2,3571$$
$$+ 0,703 . 2,2889$$

$$\cong 2,3092$$

$$\bar{y}_{village} = \frac{\sum y_i}{n_i} = \frac{264}{112}$$
$$= 2,3571$$

$$\hat{y}_{city} = \frac{\sum y_i}{n_i} = \frac{602}{263}$$
$$= 2,2889$$

$$\hookrightarrow v(\bar{y}_{tb}) = \sum w_h^2 \frac{(1-f_h)}{n_h} . s_h^2$$

$$= 0,297 . \frac{(1-\frac{112}{295})}{112} . 1,8644 \quad \overset{\frown}{\to} S_h^2$$

$$+ 0,703 . \frac{(1-\frac{263}{707})}{263} . 1,8275 \quad \to S_h^2$$

$$= 0,00310 + 0,003067 \quad = 0,006167$$

Toplam $*$

$$\hat{Y}_{tb} = N . \bar{y}_{tb} = \sum N_h . \bar{y}_h = 299 . 2,3571$$
$$+ 707 . 2,2889$$

$$= 2323,025 \cong 2323$$

$$v(\hat{Y}_{tb}) = N^2 . v(\bar{y}_{tb})$$

$$= \sum N_h^2 \frac{(1-f_h)}{n_h} . s_h^2 \quad = 3,1129 + 3,0852$$

$$\cong 6,1981$$

Oran $*$ 2'den daha fazla puan oranı : $\quad P_1 = \frac{46}{112} = 0,410 \quad P_2 = \frac{161}{263} = 0,6$

$$(P_{tb}) = \sum W_h P_h = 0,297 . 0,410 + 0,703 . 0,612 = 0,552$$

$$v(P_{tb}) = \sum w_h^2 \frac{(1-f_h) P_h . q_h}{n_h} = 0,00012 + 0,00029 = 0,00041$$

Birim Sayısı   *   Şehir puanlarının 4 ve 5 olması istenmektedir,

$$\hat{A} = N \cdot p \quad \rightarrow \quad P_{city[4;5]} = \frac{52}{263} \cong 0,20$$

$$= 707 \cdot 0,20$$

$$\cong 141$$

$$V(\hat{A}) = N^2 \cdot V(p) = 707^2 \cdot 0,00029 \cong 144,96$$

4. Kısım:

(4)   "Village" köy tabakasını kitle kabul ettim;

(a)   BRÖ ile örneklem seçimi: (t≅2 ve d=904 için)

$$n = \frac{\left(\frac{t \cdot s}{d}\right)^2}{1 + \frac{\left(\frac{t \cdot s}{d}\right)^2}{N}} = \frac{\left(\frac{2 \cdot 1,36545}{0,04}\right)^2}{1 + \frac{4661.134}{299}} \cong 281$$

(b)   Ortalama:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{654}{281} = 2,3274$$

$$V(\bar{y}) = (1-f) \frac{s^2}{n} \quad , \quad s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = 1,8352$$

$$= \left(1 - \frac{281}{299}\right) \cdot \frac{1,8352}{281} \cong 0,0004$$

(c)   toplam:

$$\hat{Y} = N \cdot \bar{y} = 299 \cdot 2,3274 \cong 696$$

$$V(\hat{Y}) = N^2 \cdot V(\bar{y}) = 299^2 \cdot 0,0004 \cong 36$$

(d)   Oran:

$$P = \frac{a}{n} = \text{"2'den büyük puan oranı"} = \frac{118}{281}$$

$$\cong 0,42$$

$$V(P) = (1-f) \cdot \frac{P \cdot q}{n-1} = 0,0000523$$

(e) Binim sayısı : "2'den fazla puan veren binim sayısı"

$$\hat{A} = N.P = 299 . 0,42 \cong 123$$

$$V(\hat{A}) = N^2 . v(p) = 299^2 . 0,0000523 \cong 4,6767$$

Kullanılan kodlar (r) :

R Kodları:

```
library(readr)
library(dplyr)
library(PracTools)
library(writexl)
library(tibble)

ypsr <- read_csv("C:/Users/lenovo/Desktop/Data_Camp/young-people-survey-responses.csv")
View(ypsr)


###ypsr verisi:
ypsr_math <- ypsr %>%
  filter(`Village - town` %in% c("village","city")) %>%
  summarize(Gender=Gender, `Village - town`=`Village - town`, Mathematics = Mathematics) %>%
  arrange(`Village - town`)


ypsr_math_village <- ypsr %>%
  filter(`Village - town` == "village") %>%
  summarize(Gender=Gender, `Village - town`=`Village - town`, Mathematics = Mathematics , ) %>%
  arrange(`Village - town`)


ypsr_math_city <- ypsr %>%
  filter(`Village - town` == "city") %>%
  summarize(Gender=Gender, `Village - town`=`Village - town`, Mathematics = Mathematics) %>%
  arrange(`Village - town`)

ypsr_math$Mathematics[is.na(ypsr_math$Mathematics)]=0
ypsr_math_city$Mathematics[is.na(ypsr_math_city$Mathematics)] = 0
ypsr_math_village$Mathematics[is.na(ypsr_math_village$Mathematics)] = 0


#write_xlsx(ypsr_math,"D:\\Hacettepe\\Örnekleme\\Hwork\\ypsr_math.xlsx")
#write_xlsx(ypsr_math_village,"D:\\Hacettepe\\Örnekleme\\Hwork\\ypsr_math_village.xlsx")
#write_xlsx(ypsr_math_city,"D:\\Hacettepe\\Örnekleme\\Hwork\\ypsr_math_city.xlsx")

count(ypsr_math_city)
count(ypsr_math_village)
```

```
###Sh lar:
tot<-0
for (i in ypsr_math_city$Mathematics){
  ortdan_farki <- (i- mean(ypsr_math_city$Mathematics))^2
  tot=tot+ortdan_farki
}
sqrt(tot/(706))


tot2<-0
for (i in ypsr_math_village$Mathematics){
  ortdan_farki <- (i- mean(ypsr_math_village$Mathematics))^2
  tot2=tot2+ortdan_farki
}
sqrt(tot2/(298))


### Örneklem çekme:
smpl_village_math <- sample(ypsr_math_village$Mathematics,112,replace=F)
village_math_samp <- as.data.frame(list(village_math=smpl_village_math))

smpl_city_math <- sample(ypsr_math_city$Mathematics,263,replace=F)
city_math_samp <- as.data.frame(list(city_math=smpl_city_math))

#write_xlsx(city_math_samp,"C:\\Users\\lenovo\\Desktop\\Hworks\\ornekleme\\smpl_city_math.xlsx")
#write_xlsx(village_math_samp,"C:\\Users\\lenovo\\Desktop\\Hworks\\ornekleme\\smpl_village_math.xlsx")

ypsr_math <- add_column(ypsr_math, reference=NA)
ypsr_math <- select(ypsr_math, c(-referans,-Referans))
ypsr_math <- select(ypsr_math, -reference)

###Yh lar:
tot_village=0
for (j in village_math_samp$village_math){
  tot_village=tot_village+j
}
village_ort_orn <- tot_village/112

tot_city=0
for (j in city_math_samp$city_math){
  tot_city=tot_city+j
}
city_ort_orn <- tot_city/263
ypsr_ort <- mean(ypsr_math$Mathematics)

###2 den fazla puan oranları:
city_2_oran <- city_math_samp %>%
  count(city_math>2)
city_2_oran

village_2_oran <- village_math_samp %>%
  count(village_math>2)
village_2_oran

city_4 <- city_math_samp %>%
  count(city_math>=4)
```

```
###village değişkenini kitle kabul ettim:
village_samp_2 <- sample(ypsr_math_village$Mathematics,281,replace=F)
village_samp_2 <- as.data.frame(list(village_samp_2))
village_samp_2 <- village_samp_2 %>%
  summarize(datas=village_samp_2$c.1..1..1..1..1..5..5..3..5..3..2..3..1..5..1..4..1..5..3..3..)


tot_village_2=0
tot_suq_village_2=0
for (i in village_samp_2$datas){
  tot_suq_village_2 = (i-mean(village_samp_2$datas))^2 + tot_suq_village_2
  tot_village_2=tot_village_2 + i
}
tot_village_2
tot_suq_village_2/280

count_village_samp_2=0
for (i in village_samp_2$datas){
  if(i >2 ){
    count_village_samp_2=count_village_samp_2+1
  }
  else{
    next
  }
}
count_village_samp_2
```

Kaynakça:

1-) Verilerin (kitle) alındığı site: Datacamp 06/12/2021 tarihinde erişildi.

2-) Verilerin analizinde kullanılan kaynak: R Programming / R Studio