13/1
2021
sid 1

Undersökningsmetodik

Anonymkod:
0009-EFL

Upg 1

Andelen positiva

$\frac{\alpha}{2} = 0{,}025$

$z = 1{,}96$: Tabell 2

a) $N = 300$ 95% KI
Felmarginal $= 0{,}05$ } Given info

$ZD = 0{,}05$ $D = \frac{0{,}05}{1{,}96}$

σ^2 okänt → Antar därmed största möjliga. $p = 0{,}5$ → $p(1-p) = \sigma^2$

$\sigma^2 = 0{,}5(1-0{,}5) = 0{,}25$

$$n \geqslant \frac{N\sigma^2}{D^2(N-1)+\sigma^2} \rightarrow \frac{300 \cdot 0{,}25}{\left(\frac{0{,}05}{1{,}96}\right)^2(300-1)+0{,}25}$$

$$n = \frac{75}{0{,}44458} = 168{,}698$$

$\simeq 169$ st i urvalet

b) $n = 80$
64 st positiva } Givet.

90% KI Andelen

Tabell 2: $\alpha/2 = 0{,}05 \rightarrow z = 1{,}6449$

$$\hat{p} = \frac{\Sigma y_K}{n} = \frac{64}{80} = 0{,}80$$

$\frac{80}{300} = 0{,}267 > 0{,}1$
Anv. Ändlighetskorrektion

$$\hat{V}(\hat{p}) = \left(1-\frac{n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}$$

$$= \left(1-\frac{80}{300}\right) \cdot \frac{0{,}8\cdot(1-0{,}8)}{80-1}$$

$$= 0{,}001485232$$

KI: $\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{p})}$

$0{,}80 \pm 1{,}6449 \cdot \sqrt{0{,}001485}$
$0{,}80 \pm 0{,}06339$

$(0{,}73661 \quad 0{,}86339)$

13/1-2021

sid 2 Upg 1. Fort.

Anonymkod:
0009-EFL

c) Det kan anses rimligt att tro att fler än 70% är positiva då KI antyder att i 90 fall av 100 ligger det sanna värdet inom

≃ 73,6% till 86,3%.

Det är även ett "OK" urvalstorlek inte det kanske bäst lämpade men eftersom att

n > 30 "duger" det för att göra generaliseringar utifrån antagandet CGS (centrala gränsvärdesatsen)

Dvs. troligen normalfördelat urval.

13/1
2021

Anonymkod
0009-EFL

sid. 3

Upg (2)

Givet:

* Stratifierat urval
* Ändlighets korrektion.

By	N_K	n_K	$\bar{y}_K$	s_K	W_K
N	890	50	95000	40000	0,20892
C	2200	150	65000	30000	0,51643
S	1170	100	50000	20000	0,27465
					1,00
Σ	4260	300			

a) Uppskatta hushållens medelinkomst i kommunen

95% KI $\alpha/2 = 0,025 \rightarrow z = 1,96$

$$\bar{y}_{STR} = \sum_{k=1}^{L} W_K \bar{y}_K \qquad W_K = \frac{N_K}{N}$$

$95000 \cdot 0,20892 + 65000 \cdot 0,51643 + 50000 \cdot 0,27465$

$= 67147,85$

$$\hat{V}(\bar{y}_{STR}) = \sum_{k=1}^{L} W_K^2 \left(1 - \frac{n_K}{N_K}\right) \frac{s_K^2}{n_K} =$$

$$0,20892^2 \cdot (1 - 50/890) \cdot \frac{40000^2}{50}$$

$$+ \; 0,51643^2 (1 - 150/2200) \cdot \frac{30000^2}{150}$$

$$+ \; 0,27465^2 \left(1 - \frac{100}{1170}\right) \frac{20000^2}{100} =$$

$1318254,59 + 1491095,146 + 275941,5592$

$= 3085291,295$

$$\bar{y}_{STR} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{V}(\bar{y}_{STR})} \rightarrow 67147,85 \pm 1,96 \cdot \sqrt{308529,29}$$

$= 67147,85 \pm 3442,7395$

KI: $(63705,1105 \; ; \; 70590,5895)$

13/1
2021

Anonymkod
0009-EFL

Sid. 4

Upg 2. Fort.

b) Ny undersökning. Utgå från given data.

Urvalsallokering som kan minimera felmarginalen.

$n = 300$.

→ Neyman allokering / optimal allokering är lämpligast för att minimera felmarginalen.

$\sigma_K = S_K$

$$n_K = n \cdot \frac{N_K \sigma_K}{\sum_{j=1}^{L} N_j \sigma_j}$$

$$\sum N_j \sigma_j = 890 \cdot 40000 + 2200 \cdot 30000 + 1170 \cdot 20000 = 125000000$$

$$n_n = \frac{890 \cdot 40000}{125000000} \cdot 300 = 85{,}44 \approx 86$$

$$n_c = \frac{2200 \cdot 30000}{125000000} \cdot 300 = 158{,}40 \approx 158$$

$$n_s = \frac{1170 \cdot 20000}{125000000} \cdot 300 = 56{,}16 \approx 56$$

300

Norrbyn $n = 86$

Centrum $n = 158$

Södbyn $n = 56$

Rimligt att Norrbyn har högre n än Södbyn då dess standardawikelse är dubbelt så stor.

13/1
2021

Anonymkod
0009-EFL

sid. 5

Upg ③

a) Kvalitetskriterier 2st: samt möjlig målkonflikt

1. Punktlighet

2. Nogrannhet.

1. Framställningstid, tillgänglig vid angivet datum.

Målkonflikt: om man ska hålla sig inom given tidsram är det inte säkert att man kan göra allt så noggrant som man skulle behöva för att säkerställa kvaliteten.

2. Hur väl överstämmer skattningarna med det (okända) sanna värdet?

b) Systematiskt urval: Används om man ex har en sorterad ram. Ex personers namn i kursen statistik 1 sorterat i bokstavsordning.
För att dra ett slumpmässigt urval väljs först vilka som ska väljas via $\frac{N}{n} = r$, och därefter väljs en startpunkt (t) t: $1 \leq t \leq r$. Och sedan väljs var r:te person.
Då t väljs slumpmässigt har alla element möjlighet att bli dragna. Fördelar: enkelt att genomföra, om ramen är lämpligt sorterad. Fås hög precision.
Nackdel: Risk för periodicitet!

Ex
$N = 20$
$n = 5$

$r = 4$
$t = 1$

Urval
$i = 1, 2, \ldots, 20$
→
1, 5, 9, 12, 16
=
urvalet

c) Bortfall effekt: Urvalet kan bli snedvridet om ex alla som svarar är "stugsittare" dvs ej representativa för urvalet.
Ex om man undersöker andelen positiva till "A" så finns en risk att bortfallet består av en större majoritet "negativa".

Det leder till en större osäkerhet i skattningarna vilket kan påverka trovärdigheten och kvaliteten på undersökningen

d) Sannolikhetsurval: Definition "varje element i populationen har en känd inklusionssannolikhet, och denna är större än 0! (Behöver dock ej vara lika stora)
Fördelar jämfört med icke-sannolikhetsurval:
* Går att beräkna vvr skattningar såsom medelvärde och totalvärde av de okända populationsparametrarna.
* Vi kan dra slutsatser om populationen utifrån urvalet
* Vi kan beräkna osäkerheten i skattningarna

Ex på sannolikhetsurval är OSU, stratifierat urval & systematiskt urval.

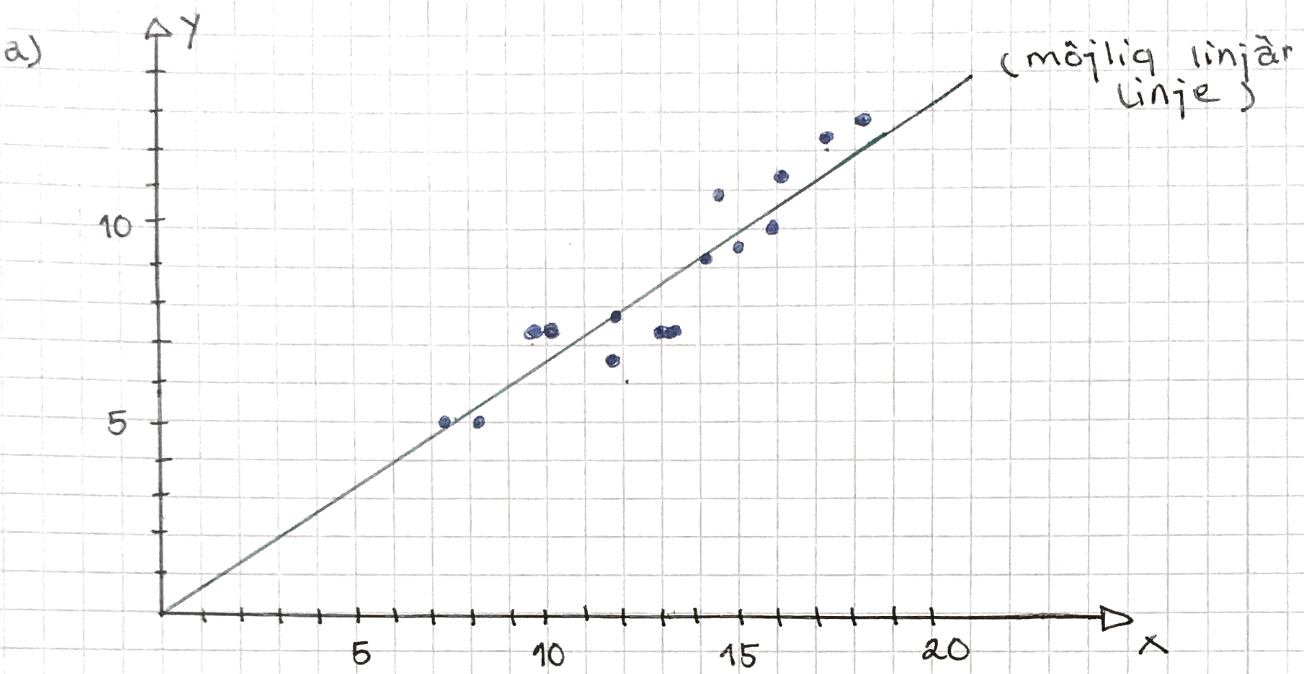
13/1 -2021 Anonymkod 0009-EFL

sid. 6 Upg (4)

Äpplen x_i = vikt

y_i = juicemängd

a)



om man placerar givna värden i en graf kan man utläsa ett positivt samband.

När x ökar, ökar y.

Vi kan även ana att om vi ritar en linjär linje är det möjligt att den går igenom origo. (se linje i grafen)

Det ser även ut att vara ett linjärt samband.

Därav anser jag att kvotestimator är lämplig att använda för att uppskatta mängden juice.

Vi har även informationen τ_x = totala vikten av partiet och då kan vi använda x som en hjälp-information för att skatta y.

13/1 - 2021

sid 7

Anonymkod

0009 - EFL

Upg 4 Fort.

b) Uppskatta tot. mängden juice.

$n = 15$

$$\sum x_i = 195,4 = 10 + 11,8 + \dots + 7,3 + 15,9$$

$$\sum x_i^2 = 2694,32 = 10^2 + 11,8^2 + \dots + 7,3^2 + 15,9^2$$

$\tau_x = 40000\,g = 40\,kg$

$$\bar{x} = \frac{195,4}{15} = 13,02667 = \frac{\sum x_i}{n}$$

$$\sum y_i = 129,4 = 7,3 + 6,8 + \dots + 5 + 10$$

$$\sum y_i^2 = 1198,38 = 7,3^2 + 6,8^2 + \dots + 5^2 + 10^2$$

$$\bar{y} = \frac{129,4}{15} = 8,62667 = \frac{\sum y_i}{n}$$

$$\hat{\tau}_{kvot} = \hat{R} \cdot \tau_x = \frac{\sum y_k}{\sum x_k} \cdot \tau_x$$

$$= \underbrace{\frac{129,4}{195,4}}_{\hat{R} = 0,66223132} \cdot 40000 = 26489,25281$$

$$\simeq 26489,25\,g$$

$$\simeq 26\,kg \text{ och } 489,25\,g$$

c) $\hat{V}(\hat{\tau}_{kvot})$ utan Ändlighetskorr. $\left(1 - \frac{n}{N}\right)$

$$= N^2 \cdot \frac{1}{n} \cdot \left(\frac{\sum (y_k - \hat{R} x_k)^2}{n-1} \right)$$

$$\sum (y_k - \hat{R} x_k)^2 = \sum y_k^2 - 2\hat{R} \sum x_k y_k + \hat{R}^2 \sum x_k^2$$

$$\sum x_k y_k = (10 \cdot 7,3) + (11,8 \cdot 6,8) + \dots (7,3 \cdot 5) + (15,9 \cdot 10) = 1789,52$$

$$\Rightarrow 1198,38 - (2 \cdot 0,66223132 \cdot 1789,52) + (0,66223132^2 \cdot 2694,32)$$

$$= 9,822517853$$

$\Rightarrow$ Vi vet inte N $\Rightarrow$ skattar via att ta totala vikten delat på medel och får då en möjlig skattning av antalet.

$$\frac{\tau_x}{\bar{x}} = \frac{40000}{13,02667}$$

$$= 3070,62436$$

$$\simeq 3071 = \hat{N}$$

$\Rightarrow$

 Anonymkod 0009-EFL

13/1-2021

$$3071^2 \cdot \frac{1}{15} \cdot \left(\frac{9{,}8225178853}{15-1} \right)$$

$$= 441126{,}5171 = \hat{V}(\hat{\tau}_{KVOT})$$

Vi kan däremot skatta $\hat{V}(\hat{\mu}_{kvot})$ då det inte krävs att vi vet N.

$$\Rightarrow \quad \frac{1}{n} \cdot \left(\frac{\Sigma (y_k - \hat{R} x_k)^2}{n-1} \right)$$

vilket är "samma" formel som för τ_{kvot} men utan N^2

$$\Rightarrow \quad \frac{1}{15} \cdot \left(\frac{9{,}8225178853}{15-1} \right) =$$

$$0{,}046773894$$

och då vet vi att $\hat{V}(\hat{\tau}_{kvot})$ är $N^2 \cdot 0{,}046773894$ för det sanna värdet på N

13/1
2021

Sid 9

Anonymkod

0009 - EFL

Upg 5.

$N = 5 \qquad U = \{1, 4, 5, 6, 9\}$

a) $\mu_y = \frac{1}{N} \cdot \sum y_k = \frac{1+4+5+6+9}{5} = 5$

$\sigma_y^2 = \frac{\sum y_k^2 - N\mu_k^2}{N} = \frac{159 - 5 \cdot 5^2}{5}$

$\sum y_k^2 = 1^2+4^2+5^2+6^2+9^2 = 159$ $\quad = 6,8$

b) Möjliga urval $n=2$ OSU med återläggning

Med hänsyn till ordning: $N^n = 25$

Möjliga urval	$\bar{y}$
1, 1	1
1, 4	2,5
1, 5	3
1, 6	3,5
1, 9	5
4, 1	2,5
4, 4	4
4, 5	4,5
4, 6	5
4, 9	6,5
5, 1	3
5, 4	4,5
5, 5	5
5, 6	5,5
5, 9	7
6, 1	3,5
6, 4	5
6, 5	5,5
6, 6	6
6, 9	7,5
9, 1	5
9, 4	6,5
9, 5	7
9, 6	7,5
9, 9	9

... utan hänsyn till ordning:

$C_n^{N+n-1} = C_2^{5+2-1} = 15$

Möjliga urval

		$\bar{y}$	s_y^2
1	1, 1	1	0
2	1, 4	2,5	4,5
3	1, 5	3	8
4	1, 6	3,5	12,5
5	1, 9	5	32
6	4, 4	4	0
7	4, 5	4,5	0,5
8	4, 6	5	2
9	4, 9	6,5	12,5
10	5, 5	5	0
11	5, 6	5,5	0,5
12	5, 9	7	8
13	6, 6	6	0
14	6, 9	7,5	4,5
15	9, 9	9	0
	Σ	75	85

då $n=2 \Rightarrow$ $s^2 = \frac{(y_i - y_j)^2}{2}$

$\leftarrow \frac{(4-5)^2}{2}$

$\leftarrow \frac{(5-9)^2}{2}$

$\leftarrow \frac{(9-9)^2}{2}$

$\bar{\bar{y}} = \frac{75}{15} = 5 = \frac{\sum \bar{y}_i}{n} \qquad n = 15$

$\Rightarrow$

13/1-2021
sid 10

Anonymkod
0009-EFL

Upg 5. Fort.

b) $\sigma^2_{\bar{y}} = \frac{\sum(\bar{y}_i - \bar{\bar{y}}_i)^2}{N} = \frac{59,5}{15}$

$N = 15$ = Antal olika stickprov här.

	$\bar{y}_i - \bar{\bar{y}}_i$		$(\bar{y}_i - \bar{\bar{y}})^2$
1	1 - 5	= -4	16
2	2,5 - 5	-2,5	6,25
3	3 - 5	-2	4
4	3,5 - 5	-1,5	2,25
5	5 - 5	0	0
6	4 - 5	-1	1
7	4,5 - 5	-0,5	0,25
8	5 - 5	0	0
9	6,5 - 5	1,5	2,25
10	5 - 5	0	0
11	5,5 - 5	0,5	0,25
12	7 - 5	2	4
13	6 - 5	1	1
14	7,5 - 5	2,5	6,25
15	9 - 5	4	16
		Σ	59,5

$\frac{59,5}{15} = 3,96667 = \sigma^2_{\bar{y}}$

c) Nej enbart $\bar{\bar{y}}$ är vvr (medelvärdet av medelvärden)

men $\sigma^2_{\bar{y}}$ är inte vvr då

$\sigma^2_{\bar{y}} \neq \sigma^2$ $\quad 3,96667 \neq 6,8$

Detta beror nog på att vi har dragit ett urval med återläggning och det bidrar till att "samma" kan komma med i urvalet och då blir variansen för det urvalet 0.

Men i verkligheten finns det en skillnad mellan alla element då och därmed bör det inte uppkomma 0 i varians i ett representativt urval!

$1 \neq 4 \neq 5 \neq 6 \neq 9$

d) Upg 5 Forts.

$$\hat{V}(\bar{y}) = \left(1-\frac{n}{N}\right)\frac{s^2}{n} \Rightarrow \left(1-\frac{2}{5}\right)\frac{s_i^2}{2}$$

	s^2	$\hat{V}(\bar{y})$
1	0	0
2	4,5	1,35
3	8	2,4
4	12,5	3,75
5	32	9,6
6	0	0
7	0,5	0,15
8	2	0,6
9	12,5	3,75
10	0	0
11	0,5	0,15
12	8	2,4
13	0	0
14	4,5	1,35
15	0	0
		25,5

$$\frac{25,5}{15} = 1,7 = \frac{\sum \hat{V}(\bar{y}_i)}{m}$$

$n=15$

$$V(\bar{y}) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n} = \left(\frac{5-2}{5-1}\right)\frac{6,8}{2} = 2,55$$

Inte heller här blir

$$V(\bar{y}) = \hat{V}(\hat{y}) \quad \text{då}$$

$$2,55 \neq 1,7.$$

Hade vi tagit bort alla "lika" och då alla med variansen 0 att ta hänsyn till i n hade vi haft $n=10$ och då hade

$V(\bar{y}) = \hat{V}(\bar{y})$ stämt eftersom

$$\frac{25,5}{10} = 2,55 \,!$$

Så för att få vvr skattning krävs det ett urval UTAN återläggning.