STOCKHOLMS UNIVERSITET
Statistiska institutionen
Regressionsanalys och undersökningsmetodik, vårterminen 2017
Jörgen Säve-Söderbergh

TENTAMEN I UNDERSÖKNINGSMETODIK

Datum	2017-05-31
Ansvarig Lärare:	Jörgen Säve-Söderbergh
Antal frågor:	5
Maxpoäng:	50
Hjälpmedel:	1) Språklexikon
	2) Kalkylator utan lagrade formler eller lagrad text
Tentamensgenomgång	Fredag 16 juni kl. 11.00 i sal B 705

Anvisningar

Redovisa dina lösningar i en form som gör det lätt att följa tankegången. Motivera alla väsentliga steg i lösningen. Ange alla antaganden och förutsättningar som du utnyttjar. Skriv endast på en sida av arket. Börja varje ny uppgift på nytt ark.

Lycka Till!

1. En forskare i sociologi studerar villkoren för studier vid landets universitet. Efter att ha undersökt utgifter för boende och datorvanor är sociologen nu intresserad av elevernas skönlitterära intresse. Då Örebro universitet tidigare förekommit och populationen av studenterna då stratifierades i tre stratum efter bostadsort, behölls denna stratifiering ännu en gång av redovisningsskäl. Frågan ställdes till eleverna om de hade läst någon skönlitterär bok under det senaste halvåret eller inte. En stickprovundersökning genomfördes på $n = 100$ individer och proportionell allokering användes. Följande resultat erhölls:

Stratum	N_i	n_i	p_i
Boende i Örebro län, men ej i Örebro kommun	6000	35	0.23
Boende i Örebro kommun	10000	59	0.18
Boende utanför Örebro län	1000	6	0.50

Beräkna ett 95%-igt konfidensintervall för andelen elever med ett skönlitterärt intresse. (10 p)

2. En idrottsforskare önskar studera det totala antalet personer som har spelat bowling under det senaste kalenderåret i Sverige. Från Svenska Bowlingförbundet erhåller hon en databasfil med uppgifter om bokningar av banor och antal personer som enligt bokningen skulle spela. Forskaren är inte så van vid databaser så det går inte att använda ett databasprogram, utan hon har helt enkelt öppnat filen med en enkel texteditor. Då kan hon läsa innehållet och ser att bokningarna är numrerade från ett till tjugotretusen. Forskaren har fått ett erbjudande om att deltaga i Gomorron Sverige för att tala om bowling och gör därför en preliminär analys. Forskaren drar ett OSU på sjuhundra bokningar med hjälp av slumptal från SAS. Antalet spelare vid de sjuhundra bokningarna är fördelade enligt:

Antal spelare per bokad bana	1	2	3	4
Antal bokningar	380	110	90	120

a) Beskriv hur forskaren genomför sitt obundna slumpmässiga urval. Vilken typ av slumptal måste hon använda? (2 p)

b) Beräkna ett 95%-igt konfidensintervall för det totala antalet personer som spelade bowling under det gångna året. (5 p)

c) Vilka problem kan din skattning tänkas ha? (3 p)

3. En population har följande värden:

Individ nr	1	2	3	4	5	6	7	8	9	10	11	12
Värde	20	15	35	70	30	80	25	14	80	35	10	50
Individ nr	13	14	15	16	17	18	19	20	21	22	23	24
Värde	45	92	29	10	65	28	49	29	65	23	43	56
Individ nr	25	26	27	28	29	30	31	32	33	34	35	36
Värde	20	7	24	92	90	80	26	5	19	9	16	48

Drag ett stickprov om $n = 3$ individer genom systematiskt urval. Använd följande utdrag ur en slumptalstabell:

19223 95034 05756 28713 96409 12531

För full poäng måste du redogöra för hur du har använt slumptalstabellen. (10 p)

4. En forskare önskade studera individers givmildhet och frågade ett urval om tio personer hur stora donationer de hade gjort under det senaste året för olika ändamål. Forskaren hade deltagit i en kurs i forskningsmetoder vid Örebro universitet och hade där lärt sig att det är effektivt ur statistisk synvinkel att använda kvotskattningar. I den artikel som blev resultatet av forskningen användes personernas deklarerade bruttoinkomst som hjälpvariabel. Mera exakt utryckt var forskaren intresserad av den genomsnittliga donationsnivån bland alla svenskar och urvalsprincipen som nyttjades var OSU.

Inkomst	Uppgiven donation
440 415	4400
193 822	19840
296 331	130
316 338	3690
529 656	235
112 667	11650
107 199	660
363 954	6970
107 049	980
549 067	5420

a) Beräkna kvotskattingen av den genomsnittliga donationen bland alla svenskar med hjälp av forskarens datamaterial. (Forskaren använde $\mu_Z = 297000$). (3 p)

b) Beräkna den vanliga medelvärdesskattningen av den genomsnittliga donationen bland alla svenskar med hjälp av forskarens datamaterial. (2 p)

c) Beräkna korrelationen mellan X och Z. (Pearson's produktmomentkorrelation). (3 p)

d) Vilken av skattningarna från a) och b) skulle du tro på. Varför? (2 p)

5. En forskare har föreslagit en metod för att göra urval ur ändliga populationer. Antag att populationen är $\{1, 2, 3, 4, 5, 6\}$. Om vi drar stickprov av storleken två, så innebär metoden att följande stickprov är de enda som kan bli utvalda i denna population. Därtill har forskaren angett sannolikheten för respektive stickprov i tabellen:

Stickprov	Sannolikhet för att detta stickprov blir utvalt
$\{1, 3\}$	1/2
$\{4, 6\}$	1/2

a) Beräkna inklusionssannolikheten π_i för samtliga individer i populationen. (3 p)

b) Är detta en bra metod? Vilka problem skulle du få om du använde denna metod? (2 p)

c) Är det möjligt att tolka denna metod utifrån någon av de kända metoder som vi har gått igenom? Är denna metod någon av våra gamla vanliga metoder i förklädnad? (5 p)

Tillägg till formelsamling undersökningsmetodik

Skattning av τ_X. Urval OSU

$$\hat{\tau}_{kvot} = \hat{R} \cdot \tau_Z = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} z_i} \cdot \tau_Z$$

$$\hat{V}\left(\hat{\tau}_{kvot}\right) = N^2 \left(\frac{N-n}{nN}\right) \frac{\sum_{i=1}^{n} \left(x_i - \hat{R} z_i\right)^2}{n-1}$$

$$\sum_{i=1}^{n} \left(x_i - \hat{R} z_i\right)^2 = \sum_{i=1}^{n} x_i^2 + \hat{R}^2 \sum_{i=1}^{n} z_i^2 - 2\hat{R} \sum_{i=1}^{n} x_i z_i$$

Skattning av μ_X. Urval OSU.

$$\hat{\mu}_{reg} = \overline{x} + b\left(\mu_Z - \overline{z}\right)$$

$$b = \frac{\sum_{i=1}^{n} \left(z_i - \overline{z}\right)\left(x_i - \overline{x}\right)}{\sum_{i=1}^{n} \left(z_i - \overline{z}\right)^2}$$

$$\hat{V}\left(\hat{\mu}_{reg}\right) = \left(\frac{N-n}{nN}\right)\left(\frac{1}{n-2}\right)\left[\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 - b^2 \sum_{i=1}^{n} \left(z_i - \overline{z}\right)^2\right]$$

Formelsamling undersökningsmetodik

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \hat{\tau} = N\bar{X}$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1} = \frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n(n-1)}$$

Beräkning av stickprovsstorlek:

$$n \geq \frac{N\sigma^2}{D^2(N-1) + \sigma^2}$$

Stratifierat urval:

$$\bar{X}_{st} = \sum_{l=1}^{L} W_l \bar{X}_l \qquad V(\bar{X}_{st}) = \sum_{l=1}^{L} W_l^2 V(\bar{X}_l) \qquad \text{där } W_l = \frac{N_l}{N}$$

Optimal allokering:

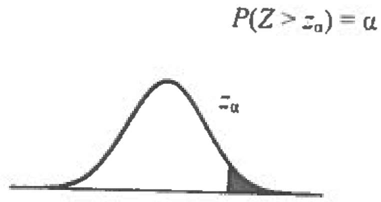
$$n_l = n\frac{N_l \sigma_l}{\sum_{j=1}^{L} N_j \sigma_j}$$

Skattning av medelvärde samt proportion per element:

$$\bar{X}_{kvot} = \frac{\sum_{i=1}^{n} \tau_i}{\sum_{i=1}^{n} m_i} \qquad \bar{X}_{VVR} = N\frac{\bar{\tau}}{M} \qquad p_{kvot} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} \qquad P_{VVR} = N\frac{\bar{a}}{M}$$

Punktskattning OSU	*Varians* m.å.	*Variansskattning* m.å.	*Varians* u.å.	*Variansskattning* u.å.
$\bar{X}$	$\frac{\sigma^2}{n}$	$\frac{s^2}{n}$	$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{s^2}{n}\left(1-\frac{n}{N}\right)$
$\hat{\tau}$	$N^2 \cdot \frac{\sigma^2}{n}$	$N^2 \cdot \frac{s^2}{n}$	$N^2 \cdot \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{s^2}{n}\left(1-\frac{n}{N}\right)$
P	$\frac{P(1-P)}{n}$	$\frac{p(1-p)}{n-1}$	$\frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$
$\hat{A}$	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$

TABELL 1. Normalfördelningens kvantiler, standardiserad $Z \in N(0, 1)$. Vilket värde har z_α om $P(Z > z_\alpha) = \alpha$ där α är en given sannolikhet. Utnyttja även $\Phi(-z) = 1 - \Phi(z)$ för $P(Z \leq -z_\alpha)$.



α	z_α
0,25	0,6745
0,10	1,2816
0,05	1,6449
0,025	1,9600
0,010	2,3263
0,005	2,5758
0,0025	2,8070
0,0010	3,0902
0,0005	3,2905
0,00025	3,4808
0,00010	3,7190
0,00005	3,8906
0,000025	4,0556
0,000010	4,2649
0,000005	4,4172

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jörgen Säve-Söderbergh

Lösningsförslag till skriftlig tentamen i Undersökningsmetodik, den 31 maj 2017

1. Punktskattningen av andelen elever med ett skönlitterärt intresse ges som

$$p_{st} = \frac{N_1}{N}p_1 + \frac{N_2}{N}p_2 + \frac{N_3}{N}p_3 = \frac{6000}{17000}\times 0.23 + \frac{10000}{17000}\times 0.18 + \frac{1000}{17000}\times 0.50 = 0.2165$$

Den skattade variansen för punktskattningen ges av

$$\begin{aligned}
\hat{V}(p_{st}) &= \left(\frac{N_1}{N}\right)^2\left(\frac{N_1 - n_1}{N_1}\right)\frac{p_1(1-p_1)}{n_1 - 1} + \left(\frac{N_2}{N}\right)^2\left(\frac{N_2 - n_2}{N_2}\right)\frac{p_2(1-p_2)}{n_2 - 1} \\
&+ \left(\frac{N_3}{N}\right)^2\left(\frac{N_3 - n_3}{N_3}\right)\frac{p_3(1-p_3)}{n_3 - 1} \\
&= \left(\frac{6000}{17000}\right)^2\frac{6000-35}{6000}\cdot\frac{0.23(1-0.23)}{35-1} + \left(\frac{10000}{17000}\right)^2\frac{10000-59}{10000}\cdot\frac{0.18(1-0.18)}{59-1} \\
&+ \left(\frac{1000}{17000}\right)^2\frac{1000-6}{1000}\cdot\frac{0.50(1-0.50)}{6-1} \\
&= \underline{0.0017}
\end{aligned}$$

Felmarginalen blir då

$$1.96\sqrt{\hat{V}(p_{st})} = 0.0806$$

vilket betyder att vårt konfidensintervall för andelen elever med ett skönlitterärt intresse blir

$$0.2165 \pm 0.0806$$

eller $\underline{(0.1358, 0.2971)}$.

2. a) Eftersom populationen innehåller 23000 bokningar, så behöver forskaren femsiffriga slumptal för att samtliga bokningar ska kunna bli valda i urvalet. Vid urvalet utnyttjas den numrering som finns av bokningarna. Med hjälp av slumptalen väljer vi de bokningar som utgör stickprovet. T ex om slumptalet 12902 uppträder så ska bokning 12902 väljas ut.

   b) Vi observerar från texten att $N = 23000$. Eftersom vi har data i en frekvenstabell beräknar vi

$$\sum_{i=1}^{4} f_i x_i = 380 \times 1 + 110 \times 2 + 90 \times 3 + 120 \times 4 = 1350.$$

För att beräkna variansen beräknar vi kvadratsumman av observationerna

$$\sum_{i=1}^{4} f_i x_i^2 = 380 \times 1^2 + 110 \times 2^2 + 90 \times 3^2 + 120 \times 4^2 = 3550.$$

Därmed blir kvadratsumman kring det aritmetiska medelvärdet

$$\sum_{i=1}^{4} f_i x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{4} f_i x_i\right)^2 = 3550 - \frac{1}{700}1350^2 = 946.4286$$

och stickprovsvariansen

$$s^2 = \frac{\sum_{i=1}^{5} f_i (x_i - \overline{x})^2}{n-1} = \frac{946.4285714285716}{700-1} = 1.3540.$$

Ett 95%-igt konfidensintervall för τ ges av

$N\bar{x} = 44357$

$$N\overline{x} \pm 1.96\sqrt{N^2\left(\frac{N-n}{N}\right)\frac{s^2}{n}}.$$

Med våra data har vi

$$23000 \cdot \frac{1350}{700} \pm 1.96\sqrt{23000^2 \cdot \left(\frac{23000-700}{23000}\right) \cdot \frac{1.353975066421419}{700}}$$

Vårt konfidensintervall blir alltså

$$44357 \pm 1952.2$$

eller ett intervall som $(42405, 46309)$.

c) Vi vet inte om samma personer kan förekomma på flera bokningar. Det är snarare troligt att många bokningar innehåller samma spelare. Om vi tänker oss att bokning 1 består av 3 personer och bokning 2 av fyra personer, så kan det förstås vara samma tre personer som spelar igen med en kompis som har kommit till. Alltså är det troligt att vår skattning överskattar det totala antalet spelare. Men, totalen kan väl ändå säga något om omfattningen av bowlandet, så undersökningen saknar inte helt värde.

3. Det gäller att

$$r = \frac{N}{n} = \frac{36}{3} = 12.$$

Det finns alltså tolv möjliga stickprov beroende på vilket värde r antar. Samtliga stickprov återges vertikalt i nedanstående tabell:

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
20	15	35	70	30	80	25	14	80	35	10	50
45	92	29	10	65	28	49	29	65	23	43	56
20	7	24	92	90	80	26	5	19	9	16	48

Eftersom r ligger mellan 1 och 12, så måste vi dra tvåsiffriga slumptal. Om vi drar ensiffriga slumptal har ju de tre talen 10, 11 och 12 sannolikheten noll att bli utvalda. De tvåsiffriga slumptalen är 19, 22, 39, 50, 34, **05**, o.s.v. De fem första talen kan vi inte använda, men det sjätte ger oss $r = 5$, så vårt stickprov blir alltså 30, 65, 90.

4. **a)** Vår hjälpvariabel Z är inkomsten och vår undersökningsvariabel X är poängen. Vi vet att medelvärdet för X skattas med hjälp av kvotskattningar enligt

$$\hat{\mu}_{\text{kvot}} = \hat{R} \times \mu_Z,$$

där $\hat{R} = \bar{x}/\bar{z}$.

Från tabellen har vi att

$$\bar{x} = 5397.5, \qquad \bar{z} = 301649.8$$

Alltså blir

$$\hat{\mu}_{\text{kvot}} = \hat{R} \times \mu_Z = \frac{\bar{x}}{\bar{z}} \times \mu_Z = \frac{5397.5}{301649.8} \times 297000 \boxed{= 5314.3.}$$

b) En vanlig medelvärdesskattning är $\bar{x} = 5397.5$.

c) Utan att gå igenom detaljerna i beräkningen blir korrelationen $r_{X,Z} = -0.22721$.

d) Om det finns en positiv korrelation mellan hjälpvariabeln Z och undersökningsvariabeln X, så får man säkrare skattningar av medelvärdet i en population med kvotskattningar än med den vanliga skattningen. Emellertid har vi visat i c) att X och Z korrelerar svagt negativt. Det innebär att vi bör sätta störst tilltro till den vanliga medelvärdesskattningen i detta fall.

5. Urvalsmetoden beskrivs i följande tabell

Stickprov	Sannolikhet
$\{1,3\}$	1/2
$\{4,6\}$	1/2

a) Om vi börjar med det första elementet i populationen, 1, ser vi att det förekommer i ett stickprov, det första, bland två möjliga. Alltså kan vi använda den klassiska sannolikhetsdefinitionen, antalet gynnsamma delat med antalet möjliga, för att finna sannolikheten. Då har vi

$$\pi_1 = \frac{\text{antal gynnsamma}}{\text{antal möjliga}} = \frac{1}{2} = 0.5$$

För det andra elementet, 2, blir motsvarande beräkning

$$\pi_2 = \frac{\text{antal gynnsamma}}{\text{antal möjliga}} = \frac{0}{2} = 0,$$

eftersom det är omöjligt att dra element 2 med denna metod. Att dra 2 är en omöjlig händelse, så sannolikheten för den händelsen ska ju vara noll.

Om vi fortsätter på samma vis finner vi att $\pi_3 = \pi_4 = \pi_6 = 0.5$, medan $\pi_4 = 0$.

b) Eftersom $\pi_2 = \pi_4 = 0$ så leder denna metod till icke sannolikhetsurval. Då kan ingen statistik slutledningsteori användas på det stickprov som metoden levererar. Vi vet t ex inget om urvalsfelet.

c) Är det systematiskt urval? Då $r = \frac{N}{n} = \frac{6}{2} = 3$, har vi tre möjliga startpunkter och således tre olika stickprov, men vi har endast två. Detta är inte ett systematiskt urval.

Om vi stratifierar populationen i två strata $\{1, 2, 4\}$ och $\{3, 5, 6\}$, så kan vi tänka på detta som ett stratifierat urval med proportionellt allokering. Det innebär att $n_1 = n\frac{N_1}{N} = 2\frac{3}{6} = 1$ och samma i stratum två. Vi drar alltså ett värde från varje stratum. Då finns det en möjlighet att vi drar 1 i stratum ett och 3 i stratum två, alltså $\{1, 3\}$. På samma sätt $\{4, 6\}$. Men det finns ju ytterligare sju stickprov som denna metod medger. Alltså är det inget stratifierat urval detta heller.

Det är förstås inget OSU heller. Om vi drar stickprov av storleken två ur en population med sex individer är det välkänt att vi kan göra det på $\binom{6}{2} = 15$ olika sätt. Det betyder att vi saknar tretton stickprov.

Metoden är helt sin egen och mycket speciell. En tredjedel av population kan aldrig bli utvald!

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 31/5-2017

Sal: Värtasalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: REU-0071

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					9 3c
Lär.ant. 10	10	10	10	10					

POÄNG	BETYG	Lärarens sign.
50	A	JSS

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: REU-0071 Blad nr: 1

1. Stratifierat urval, proportionell allokering vilket innebär att $\frac{N_i}{N} = \frac{n_i}{n}$. Totala stickprovet har $n = 100$ individer. Populationen, N har $6000 + 10000 + 1000 = 17000$ individer

Tre strata,

① : $N_1 = 6000$, $n_1 = 35$, $p_1 = 0.23$, $W_1 = \frac{N_1}{N} = \frac{6000}{17000} = 0.352941176$

② : $N_2 = 10000$, $n_2 = 59$, $p_2 = 0.18$, $W_2 = \frac{N_2}{N} = \frac{10000}{17000} = 0.588235294$

③ : $N_3 = 1000$, $n_3 = 6$, $p_3 = 0.50$, $W_3 = \frac{N_3}{N} = \frac{1000}{17000} = 0.058823529$

- Först beräknas medelvärdet för andelen elever med skönlitterärt intresse över de tre strata, Detta ges av $\bar{p}_{st} = \sum_{i=1}^{L} W_i p_i =$

$= 0.352941176 \cdot 0.23 + 0.588235294 \cdot 0.18 + 0.058823529 \cdot 0.50 =$

$\bar{p}_{st} = 0.2164705879$ R

- Nu beräknas variansen, $V(\bar{p}_{st}) = \sum W_i^2 V(p_i) =$

$= \sum W_i^2 \cdot \frac{p_i(1-p_i)}{n_i - 1}\left(1 - \frac{n_i}{N_i}\right) = 0.352941176^2 \cdot \frac{0.23(1-0.23)}{34}\left(1 - \frac{35}{6000}\right)$

$+ 0.588235294^2 \cdot \frac{0.18(1-0.18)}{58}\left(1 - \frac{59}{10000}\right) + 0.058823529^2 \cdot \frac{0.50(1-0.5)}{5}$

$\cdot \left(1 - \frac{6}{1000}\right) = 6.4506502982 \cdot 10^{-4} + 8.753678555 \cdot 10^{-4} +$

$+ 1.7197231593 \cdot 10^{-4} = 0.0016924052$ R

$\sqrt{V(\bar{p}_{st})} = 0.0411388526\,97$

Jag antar att normalapproximering kan göras då $n > 30$ och använde då z-fördelningen. $z_{\alpha/2} = z_{0.05/2} = 1.96$ (från tabell)

- 95% konfidensintervall är då $\bar{p}_{st} \pm 1.96 \cdot \sqrt{V(\bar{p}_{st})} =$

$= 0.2164705879 \pm 0.0806321512\,87 \approx [0.1358 ; 0.2971]$

- Svar: 95% konfidensintervall för $\bar{p}_{st}$ är $[0.1358 ; 0.2971]$, dvs för andelen elever med skönlitterärt intresse. R

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: REU-0071 Blad nr: 2

2\. Svar: femsiffrigt tal R !

a) Då bokningarna är numrerade från 1-23000 måste hon använda femsiffriga slumptal så att alla bokningar har chans att bli utvalda i urvalet. (Eftersom det är OSU ska varje element ha samma inklusionssannolikhet $= \frac{n}{N}$) Jag skulle genomföra urvalet genom att helt enkelt använda bokningens numrering, och välja slumpvisa nummer mellan 1-23000 med någon slumpgenerator. Får forskare t.ex slumptalet 20500, då väljer hon ut bokning nr 20500. Får hon nr 1, välja hon bokning nr 1 osv. Man får ha ett system som ser till att varje nummer endast kan väljas en gång.

b) Först måste totala antalet spelare i urvalet beräknas. Eftersom värdena står i en frekvenstabell ges antalet av $\sum f_i x_i$ där f_i = frekvens.

$\sum f_i x_i = 1 \cdot 380 + 2 \cdot 110 + 3 \cdot 90 + 4 \cdot 120 = 1350$

Antalet spelare i populationen punktskattas då av $N \cdot \frac{\sum f_i x_i}{n} =$

$= 23000 \cdot \frac{1350}{700} = 44357.1428571$ spelare $\approx$ 44357 spelare R

Nästa steg är att beräkna stickprovsvariansen, s^2 och sedan variansen för skattningen.

s^2 ges av $\frac{\sum f_i \cdot (x_i - \bar{x})^2}{n-1}$ där $\sum f_i \cdot (x_i - \bar{x})^2 =$

$= \sum f_i x_i^2 - \frac{1}{n}\left(\sum f_i x_i\right)^2$ forts. nästa sida (3)

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: REU-0071 Blad nr: 3

2 b) fortsättning. $\sum f_i x_i^2 = 1^2 \cdot 380 + 2^2 \cdot 110 + 3^2 \cdot 90 + 4^2 \cdot 120 =$

$= 380 + 440 + 810 + 1920$

$\sum f_i x_i^2 = 3550$

$\left(\sum f_i x_i\right)^2 = 1350^2 = 1822500$ ser:

$$s^2 = \frac{3550 - \left(\frac{1}{700} \cdot 1822500\right)}{699} = 1{,}3539750664 \quad R$$

Skattningen av Variansen för skattningen av totalantalet bowlingspelare i populationen är då: $N^2 \cdot \frac{s^2}{n}\left(1 - \frac{n}{N}\right) = 23000^2 \cdot \frac{1{,}3539750664}{700} \cdot$

$\cdot \left(1 - \frac{700}{23000}\right) = 992076{,}873672$ Jag normalapproximerar och använder z-fördelningen. $z_{\alpha/2} = z_{0{,}05/2} = 1{,}96$ (från tabell)

95 % konfidensintervall ges då av $44357{,}1428571 \pm 1{,}96 \times \sqrt{\hat{V}}$

$= 44357{,}1428571 \pm 1952{,}2198948 \approx [42405 ; 46309]$ R

b) Svar: Ett 95 % konfidensintervall för totala antalet personer är $[42405 ; 46309]$ R

c) Svar: Jag tror det finns en risk att man överskattar antalet bowlingspelare pga att det är ju inte säkert att antal spelare per bana är samma sak som antal individer. Jag menar alltså att t.ex kan en eller flera spelare bokat en viss bana och samma personer kanske spela vidare vid nästa bana och boka denna osv. Ja

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** BEU-0071 **Blad nr:** 4

3) Systematiskt urval. Innebär att $r = \frac{N}{n}$ stickprov är möjliga och att stegen (intervallet) mellan de urval som görs är r enheter långt. Här är $r = \frac{36}{3} = 12$ då $N = 36, n = 3$. Det innebär att steglängden är 12. Det innebär också att med slumpmässig start av urvalet ska en individ mellan 1 och r dvs mellan 1 och 12 väljas som första individ. Eftersom $r = 12$ måste vi alltså ha tvåsiffriga slumptal för att möjliggöra att alla tänkbara individer kan bli utvalda.

Jag väljer då att läsa slumptalstabellen från vänster till höger tills jag hittar första tvåsiffriga tal mellan 1 och 12: Först ser jag 19, sedan 22, 37, 50, 34 och 05. De fem första talen är inte aktuella eftersom de är $> r$. Jag väljer alltså 05, dvs individ 5 som startpunkt i mitt systematiska urval. Sedan väljer jag ytterligare 2 individer där nästa individ ska ha nr. $05 + r$ och sista (tredje) individen har numret $05 + 2r$. Detta ger stickprovsindividerna: 05, 17, 29 dvs värdena 30, 65, 90 som korresponderar till de valda individerna.

Svar: Se ovan, och mitt stickprov om $n = 3$ individer blir alltså individer nr 5, 17, 29 dvs värdena 30, 65 och 90 ingår i stickprovet.

R

SU, STATISTIK

Skrivsal: Vårtasalen Anonymkod: REU-0071 Blad nr: 5

4.) a) Kvotskattning söks av genomsnittlig donation. OSV gäller.
Vi kallar variabeln donation för X och variabeln inkomst (hjälpvariabeln) för z. X är alltså vår undersökningsvariabel. Känt är: $N_z = 297000$, $n = 10$. Vi behöver räkna ut $\sum_{i=1}^{n} x_i$ och $\sum_{i=1}^{n} z_i$ för de tio observationerna:

$\sum x_i = 4400 + 19840 + 130 + 3690 + 235 + 11650 + 660 + 6970 + 980 + 5420$
$= 53975$ (summa av donationerna)

$\sum z_i = 440415 + 193822 + \ldots = 3016498$ (summa av inkomster)

En kvotskattning av den genomsnittliga donationen ges då av

$$N_{kvot_X} = \frac{\sum x_i}{\sum z_i} \cdot N_z = \frac{53975}{3016498} \cdot 297000 = 5314.29985345$$

≈ 5314 kr.

a Svar: Genomsnittlig donation skattad mha kvotskattning är 5314 kr R

4.) b) Vanlig medelvärdesskattning av donationen ges av

$$\frac{\sum_{i=1}^{n} x_i}{n} = \frac{53975}{10} = 5397.5 \text{ kr} \approx 5398 \text{ kr}$$ om man vill svara i heltal

b) Svar: 5398 kr. R

4 c) Korrelationen, $r = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum(x_i-\bar{x})^2 \cdot \sum(z_i-\bar{z})^2}}$

Vi behöver alltså först räkna ut ett antal kvadratsummor.

Forts. nästa sida.

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** REU-0091 **Blad nr:** 6

4 c) fortsättning.

$$\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z}) = \sum x_i z_i - \frac{1}{n}\left(\sum x_i\right)\times\left(\sum z_i\right)$$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2$$

$$\sum_{i=1}^{n}(z_i-\bar{z})^2 = \sum z_i^2 - \frac{1}{n}\left(\sum z_i\right)^2$$

Vi behöver alltså räkna ut $\sum x_i z_i$, $\sum x_i^2$ och $\sum z_i^2$. Övriga har redan räknats ut i 4 a).

$x_i z_i$	x_i^2	z_i^2
1937826000	4400^2	440415^2
3845428480	19840^2	193822^2
...	...	...
Σ 14114466320	641749625	$1{,}16953166155\times10^{12}$

Enskilda värden redovisas ej här pga det stora antalet värdesiffror, utan slås direkt på miniräknaren. Annars risk att göra fel.

Vi har från 4 a) att: $\sum x_i = 53975 \Rightarrow \left(\sum x_i\right)^2 = 2913300625$

$\sum z_i = 3016498 \Rightarrow \left(\sum z_i\right)^2 = 9{,}099260184\cdot10^{12}$

$\therefore \sum x_i z_i - \frac{1}{n}\left(\sum x_i\right)\times\left(\sum z_i\right) = 14114466320 - \left(\frac{1}{10}\times 53975 \times \times 3016498\right) =$

$= -2167081635$

$$\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 641749625 - \left(\frac{1}{10}\times 2913300625\right) = = 350419562{,}5$$

$$\sum z_i^2 - \frac{1}{n}\left(\sum z_i\right)^2 = 1{,}16953166155\times10^{12} - \left(\frac{1}{10}\times 9{,}099260184\times10^{12}\right) = = 259605643150$$

forts nästa sida

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: REU-0071 Blad nr: 7

4 c) fortsättning:

Detta ger att $r = \dfrac{-2167081635}{\sqrt{350419562.5 \times 25960564315\text{0}}} =$

$= -0.227208218168 \approx -0.227$

Svar: Korrelationen mellan X och Z, r, är −0.227 R

d) I det här fallet är korrelationen mellan X och Z negativ. Det innebär att jag skulle tro mest på den vanliga medelvärdesskattningen (OSU-skattningen) i detta fall. Anledningen är att för att kvotskattningen ska kunna ge en säkrare skattning än den vanliga OSU-skattningen krävs att det finns en positiv korrelation mellan undersökningsvariabeln X och hjälpvariabeln Z. Här gäller ju inte detta, utan istället det omvända. R

Bra!

d) Svar: Se ovan.

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** SEU-0071 **Blad nr:** 8

5 a) N = 6, individer väljs ut direkt genom två stickprov.

- Man kan direkt säga att inklusionssannolikheten, π_i, för individerna 2 och 5 är = 0 i båda fallen, eftersom de inte ingår i något av de möjliga två stickproven, vardera med slh 1/2, att bli utvalda. R
- För individer 1 och 3 kan man se att de alltid ingår i det ena stickprovet av två möjliga. Eftersom detta stickprov av {1,3} har sannolikheten $\frac{1}{2}$ att bli utvalt måste också sannolikheten för individerna 1 och 3 att bli utvalda vara densamma dvs $\frac{1}{2}$. Dvs 1 av 2 möjliga händelse gör att individerna 1 och 3 blir utvalda $\Rightarrow$ slh = $\frac{1}{2}$
- För individer 4 och 6 kan man på motsvarande sätt se att de alltid ingår i det andra stickprovet {4,6} som har slh $\frac{1}{2}$ att bli utvalt. Det innebär att även individerna 4 och 6 har respektive sannolikhet = $\frac{1}{2}$ att väljas ut.
- Detta ger alltså att $\pi_1 = \frac{1}{2}$, $\pi_2 = 0$, $\pi_3 = \frac{1}{2}$, $\pi_4 = \frac{1}{2}$, $\pi_5 = 0$ och $\pi_6 = \frac{1}{2}$ R
- Man kan alltså tänka enligt "klassisk" sannolikhetslära, dvs sannolikheten för ett element att bli utvalt är = $\frac{\text{"Antal gynnsamma"}}{\text{"Antal möjliga händelser"}}$ vilket blir = $\frac{1}{2}$ för vardera element 1,3,4,6. och = 0 för elementen 2 resp. 5

i) Svar: Inklusionssannolikheterna är för individ 1: $\frac{1}{2}$, för individ 2: 0, för individ 3: $\frac{1}{2}$, för individ 4: $\frac{1}{2}$, för individ 5: 0, för individ 6: $\frac{1}{2}$, se även diskussion och beräkningssätt ovan. R

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** REU-0071 **Blad nr:** 9

5b) Svar: Nej, i statistiska sammanhang vill jag påstå att detta inte är någon bra metod. Det är inget sannolikhetsurval eftersom 2 individer dvs $\frac{2}{6} = \frac{1}{3}$ av populationen, har slh = 0 att bli utvalda.
På grund av detta skulle jag få mycket stora problem att göra några valida statistiska slutledningar. Beräkning av urvalsfelet går t.ex inte att göra. Det skulle ju också kunna vara möjligt att de individer som aldrig blir utvalda representerar unika, viktiga egenskaper som är mkt viktiga för den statistiska undersökningen, och då missas dessa helt. R

5c) Svar: Vad jag kan bedöma är detta ingen variant av någon av de kända metoder vi gått igenom på kursen. Det verkar vara en egen, speciell metod. Det är inte OSU eftersom i OSU har alla element lika sannolikhet $= \frac{n}{N}$ att komma med i urvalet. I denna metod har ju två element slh = 0 och resterande slh $= \frac{1}{2}$. Det är inte heller systematiskt urval som också är ett sannolikhetsurval där med inklusionsslh $= \frac{n}{N}$ per element och med $r = \frac{N}{n} = \frac{6}{2} = 3$ stickprov. Här finns ju bara två stickprov.
Gruppurval kan också uteslutas direkt. Här görs ju direkt-urval av element (individer) och inte grupper som i gruppurval. Stratifierat urval kan också uteslutas, där man delar av populationen i ett antal strata (områden) innan urval dras. Här har man inte delat av hela populationen, utan istället uteslutit en del av den innan urval dras. Urvalen är ju också sannolikhetsurval. Nej, jag kan inte tolka metoden utifrån kända metoder

BPA/

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 31/5-2017

Sal: Värtasalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

REV-

ANONYMKOD: 0005

[X] Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					10 3L
Lär.ant. 10	10	10	10	10					

POÄNG	BETYG	Lärarens sign.
50	A	JSS

SU, STATISTIK

Skrivsal: Värta Anonymkod: REV-0005 Blad nr: 1

Fråga 1

Stratum	N_i	n_i	p_i	$w_i = \frac{N_i}{N}$
1	6000	35	0,23	6000/17000
2	10 000	59	0,18	10000/17000
3	1000	6	0,5	1000/17000
Σ	17000	100	1	

- söker först P_{st}:

$$P_{st} = \Sigma W p_i = \frac{6000}{17000} \cdot 0{,}23 + \frac{10000}{17000} \cdot 0{,}18 + \frac{1000}{17000} \cdot 0{,}5 = 0{,}216470588$$ R

- söker sedan $V(P_{st})$:

$$V(P_{st}) = \Sigma W^2 V(P_i) \quad \text{där} \quad V(P_i) = \frac{p(1-p)}{n_i - 1}\left(1 - \frac{n_i}{N_i}\right)$$

$$V(P_{st}) = \left(\frac{6000}{17000}\right)^2 \cdot \frac{0{,}23(1-0{,}23)}{35-1} \cdot \left(1 - \frac{35}{6000}\right)$$

$$+ \left(\frac{10\,000}{17\,000}\right)^2 \frac{0{,}18(1-0{,}18)}{59-1} \cdot \left(1 - \frac{59}{10\,000}\right)$$

$$+ \left(\frac{1000}{17000}\right)^2 \cdot \frac{0{,}5 \cdot (1-0{,}5)}{6-1} \left(1 - \frac{6}{1000}\right)$$

$$= 0{,}0006450650315 + 0{,}000875367355\,9$$

$$+ \; 0{,}000171972 3183$$

$$= \underline{0{,}0016924043 18} = V(P_{st})$$ R

SU, STATISTIK

Skrivsal: Värtan Anonymkod: REV-0005 Blad nr: 2

Forts. fråga 1

95%-igt K.I:

$$P_{st} \pm 1{,}96 \cdot \sqrt{V(P_{st})} =$$

$$= 0{,}216470588 \pm 1{,}96 \cdot \sqrt{0{,}001692404318}$$

$$= 0{,}216470588 \pm \underbrace{0{,}08063213}_{\text{felmarginal}}$$

$$[0{,}135838458 \,;\, 0{,}297102718]$$

$$\approx [0{,}136 \,;\, 0{,}297]$$

R

SU, STATISTIK

Skrivsal: Värta **Anonymkod:** REU-0005 **Blad nr:** 3

Fråga 2

$N = 23\,000$ x = antal bowlingspelare

$n = 700$ $\hat{\tau}$ = totalt — u —

a/ Hon behöver femsiffriga slumptal, mellan 00001 - 23000
Sedan låter hon en dator (SAS) slumpa fram ett tal mellan 00001 och 23000, väljer datorn t.ex. nr 04300 väljs bowling nr 4300 till urvalet.
Detta upprepas 700 ggr, tills $n = 700$.

Bra! R

b/ Söker $\hat{\tau}_{OSU} = N \cdot \bar{x}$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{380 \cdot 1 + 110 \cdot 2 + 90 \cdot 3 + 120 \cdot 4}{700} =$$

$$= \frac{1350}{700} = 1{,}928571429$$

$$\hat{\tau}_{osu} = N \cdot \bar{x} = 23000 \cdot \frac{1350}{700} = \underline{44\,357{,}14286}$$ R

$$V(\hat{\tau}_{osu}) = N^2 \cdot \frac{s^2}{n}\left(1 - \frac{n}{N}\right) \quad \text{där} \quad s^2 = \frac{\sum(x - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum x^2 - \frac{1}{n}(\sum x)^2}{n-1} = \frac{1^2 \cdot 380 + 2^2 \cdot 110 + 3^2 \cdot 90 + 4^2 \cdot 120 - \left(\frac{1}{700} \cdot 1350^2\right)}{700-1}$$

$$= \frac{3550 - \left(\frac{1350^2}{700}\right)}{699} = 1{,}353975066$$

$$V(\hat{\tau}_{osu}) = 23000^2 \cdot \frac{1{,}353975066}{700}\left(1 - \frac{700}{23000}\right)$$

$$= 992\,076{,}8734$$

SU, STATISTIK

Skrivsal: Värta Anonymkod: REV-0005 Blad nr: 4

Forts. fråga 2

95%-igt KI

$$\hat{\tau}_{OSU} \pm 1{,}96\sqrt{V(\hat{\tau}_{OSU})}$$

$$44\,357{,}14286 \pm 1{,}96\sqrt{992\,076{,}8734}$$

$$= 44\,357{,}14286 \pm \underbrace{1952{,}219895}_{\text{felmarg.}}$$ R

$$[42\,404{,}923 \; ; \; 46\,309{,}36275]$$

$$\approx [42\,405 \; ; \; 46\,309]$$ R → detta är människor, så måste avrunda till heltal.

c/ Vid OSU finns alltid risken att få ett "dåligt" urval, t.ex om det råkats slumpats fram 500 pers som endast kommer från Stockholm, och stockholmarna spelar ovanligt mycket bowling. Detta är dock ett relativt stort urval, så risken för detta är ganska liten. Troligen Denna risk kan dock hanteras genom att tex stratifiera urvalet; detta kan ge högre precision än OSU om det verkligen existerar olika nivåer på antalet bowlingspelare i olika delar av landet, t.ex, och då göra strata efter exempelvis län. I detta fall, när en väldefinierad ram existerar, är det även möjligt med ett systematiskt urval! Detta kan också ge bättre precision än OSU, om ramen är välsorterad (i detta exempel om ordningen av bokningarna ligger på ett sätt så att ev. geografiska skillnader kan fångas upp)

SU, STATISTIK

Skrivsal: Värtan   Anonymkod: REV-0005   Blad nr: 5

Forts. fråga 2

Problem med denna skattning är således att den riskerar en bristande precision, i form av högre varians och bredare konfidensintervall än som eventuellt hade kunnat uppnås med t.ex. stratifierat urval.

Mycket bra.

SU, STATISTIK

Skrivsal: Värtan Anonymkod: REU-0005 Blad nr: 6

Fråga 3

$N = 36$, $n = 3$

$r = \frac{N}{n} = \frac{36}{3} = 12$ dvs finns totalt 12 olika möjliga stickprov att dra. R

$01 \leq a \leq 12 \Rightarrow$ vi använder tvåsiffriga slumptal för att få reda på a, dvs var i listan vi ska börja.

Ur slumptalstabellen får vi därför talen:

19 22 39 50 34 | 05 | 75 62 ...

Vi väljer det första talet inom intervallet 01-12 : i detta fall 05.

Ur vår population drar vi då urvalet med värdena: 30, 65 och 90.

Dvs

30
65
90

ur listan, eftersom vi då börjat på individ 5, och sedan tagit var 12:e (r:te) individ.

(= individ nr 5, 17 & 29)

R

SU, STATISTIK

Skrivsal: Värtа Anonymkod: REU-0005 Blad nr: 7

Fråga 4

a/ z_i = inkomst, x_i = donation

$$\hat{\mu}_x = \hat{\mu}_{kvot} = \hat{R} \cdot \mu_z = \frac{\sum x_i}{\sum z_i} \cdot \mu_z$$

Inkomst = z_i Donation = x_i

$\sum z_i = 3'016\,498$ $\sum x_i = 53\,975$

$$\hat{\mu}_{kvot} = \frac{53\,975}{3\,016\,498} \cdot 297\,000 = 5314{,}299893$$

Svar $\hat{\mu}_{kvot} \approx 5314{,}2999$ R

b/ $$\bar{x} = \frac{\sum x_i}{n} = \frac{53\,975}{10} = 5397{,}5$$

Svar: $\bar{x} = 5397{,}5$ R

c/ $$corr = \frac{\sum(z-\bar{z})(x-\bar{x})}{\sqrt{\sum(z-\bar{z})^2 \sum(x-\bar{x})^2}} = \frac{\sum zx - \frac{1}{n}(\sum z)(\sum x)}{\sqrt{[\sum z^2 - \frac{1}{n}(\sum z^2)][\sum x^2 - \frac{1}{n}(\sum x)^2]}}$$

- $\sum z_i x_i = 1{,}411465082 \cdot 10^{10}$
- $\sum z^2 = 1{,}169531662 \cdot 10^{12}$
- $\sum x^2 = 628\,133\,525$
- $\sum zx - \frac{1}{n}(\sum z)(\sum x) =$

$= 1{,}411465082 \cdot 10^{10} - \frac{1}{10} \cdot 3\,016\,498 \cdot 53\,975$

$= -2166897135$

- $\sum(z-\bar{z})^2 = 1{,}169531662 \cdot 10^{12} - \frac{1}{10} \cdot 3016498^2 =$

$= 2{,}596056436 \cdot 10^{11}$

SU, STATISTIK

Skrivsal: Värtan   Anonymkod: REV-0005   Blad nr: 8

Forts. fråga 4

$$\sum(x-\bar{x})^2 = 628\,133\,525 - \frac{1}{10}\,53\,975^2$$

$$= 336\,803\,462{,}5$$

$$\sqrt{\sum(z-\bar{z})^2 \cdot \sum(x-\bar{x})^2} = \sqrt{2{,}596056436 \cdot 10^{4} \cdot 336\,803\,462}$$

$$= 93\,507\,26156$$

$$\text{Corr} = \frac{-216\,689\,7135}{93\,507\,26156} = -0{,}231735706$$

$$\text{corr} \approx -0{,}23$$ R

d/ Enligt c/ får vi fram en negativ korrelation mellan x och z.
Kvotskattning kan ge bättre precision om en positiv korrelation existerar mellan hjälp- och undersökningsvariabel.
Detta gör att den vanliga medelvärdesskattningen i b) är att föredra, framför kvotskattningen i a).

R

SU, STATISTIK

Skrivsal: Värta Anonymkod: REV-0005 Blad nr: 9

Fråga 5

a)

individer	π_i
1	1/2
2	0
3	1/2
4	1/2
5	0
6	1/2

Inklusionsslh:ten är således

$\pi_1 = 0{,}5$, $\pi_2 = 0$, $\pi_3 = 0{,}5$, $\pi_4 = 0{,}5$, $\pi_5 = 0$, $\pi_6 = 0{,}5$. R

b) Detta är inte en sannolikhetsmetod, så vi kommer inte att kunna dra några slutsatser kring t.ex medelfel R

c) Om OSU: skulle antal möjliga stickprov vara $\binom{N}{n} = \binom{6}{2} = \frac{6!}{2!4!} = 15$

I denna metod finns bara två möjliga stickprov. Detta är alltså inte OSU.

- Dessutom har inte alla individer lika stor chans att bli valda R

Om systematiskt urval skulle antal möjliga stickprov vara $\frac{N}{n} = \frac{6}{2} = 3$.

Dvs är inte systematiskt urval. R

1	2	3
4	5	6

→ (de skulle dessa ha varit de möjliga urvalen).

SU, STATISTIK

Skrivsal: Värta   Anonymkod: REU-0005   Blad nr: 10

Forts fråga 5

Om stratifierat urval: vi kan tänka oss urval ur två stratum, t.ex.

|1, 2, 5| och |3, 4, 6|
män — kvinnor, t.ex.

Enligt proportionell allokering vill vi då ha $n \cdot \frac{N_i}{N} = 2 \cdot \frac{3}{6} = 1$ individ ur varje stratum till vårt stickprov. Två möjliga urval är då (1, 3) och (4, 6), men finns ytterligare 7 möjliga urval (finns tot. 9 st). Är därför inte ett stratifierat urval. R

Om gruppurval:

Liknande resonemang som vid stratifierat urval: om det istället hade varit lämpligt med gruppindelning, och |1, 2, 5| och |3, 4, 6| var två grupper med 3 element var, hade två möjliga urval varit (1, 3) och (4, 6), men saknas fortf. 7 stickprov som också kan bli valda.

Svar: Nej, detta är ingen av våra "gamla vanliga" metoder, utan en för denna kurs okänd metod.

Mycket genomgripande diskussion! BRA