ERP

Försättsblad till skriftlig tentamen vid Linköpings Universitet

	(fylls i av ansvarig)
Datum för tentamen	*2009-03-27*
Sal	TER1
Tid	*08-12*
Kurskod	*732G04*
Provkod	*TENC*
Kursnamn/benämning	Surveymetodik Tentamen
Institution	*IDA*
Antal uppgifter som ingår i tentamen	*4*
Antal sidor på tentamen (inkl. försättsbladet)	*3* + 6 s formler
Jour/Kursansvarig	*Lotta Hallberg*
Telefon under skrivtid	
Besöker salen ca kl.	*10.00*
Kursadministratör (namn + tfnnr + mailadress)	*Carita Lilja* *013-2811463 carli@ida.liu.se*
Tillåtna hjälpmedel	Räknedosa, med tentan vidhäftad formelsamling
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	*G=12, VG=16*

Linköpings Universitet
IDA/Statistik
LH

732G02, 732G90 Statistik 1 del C, 7,5 av 15hp
732G04 Surveymetodik, 7,5hp

Tentamen i Surveymetodik, 2009-03-27

Skrivtid: kl: 8-12
Hjälpmedel: Räknedosa, med tentan vidhäftad formelsamling
Jourhavande lärare: Lotta Hallberg
Redovisa och motivera kort alla dina lösningar.

All data in denna tenta är påhittad.

1

I ett visst län vill man skatta andelen mjölkproducerande gårdar som är kravgodkända. Man drar därför ett OSU om 30 gårdar och bland dessa var det 8 gårdar som var godkända. Totalt finns 220 gårdar i länet.

a) Beräkna ett 99% konfidensintervall för andelen kravgodkända gårdar i länet. 2p

b) Om undersökningen ska göras om vid ett senare tillfälle, hur många gårdar ska man då dra om man i ett konfidensintervall för andelen kravgodkända gårdar i länet vill ha en konfidensgrad på 99% och en felmarginal på högst 15%? 2p

2

I det valda länet från uppgiften ovan vill man uppskatta medelantalet mjölkkor per gård. Man drar därför ett OSU utan återläggning om 50 gårdar. Resultatet redovisas nedan efter storlek på gård. x= antal mjölkkor. Totalt finns 220 gårdar i länet.

Storlek på gård	$\bar{x}_i$	n_i	s_i
Liten	10	10	2
Mellan	35	22	4
Stor	110	18	21

Du ska i uppgiften räkna som om du hade stora stickprov. I de fall då det är nödvändigt så får du skatta N_i/N med n_i/n. Du behöver inte ta hänsyn till denna osäkerhet i denna uppgift.

a) Beräkna ett 95% konfidensintervall för medelantalet mjölkkor på stora gårdar. 2p
b) Skatta totala antalet mjölkkor på stora gårdar i länet med ett 95% konfidensintervall. Skatta dessutom (ej konfidensintervall) andelen mjölkkor i länet som finns på stora gårdar. 3p
c) Poststratifiera (dvs räkna som om du hade ett stratifierat urval) med avseende på storlek på gård och skatta medelantalet mjölkkor per gård i hela länet med ett 95% konfidensintervall. 3p
d) Om man vid ett senare tillfälle vill göra om undersökningen och då vill stratifiera från början, hur stora urval ska man då dra från varje stratum om man vill ta hänsyn till de olika standardavvikelserna i varje stratum. Låt även här totala urvalsstorleken vara 50 gårdar. 2p

3

Bland länets mjölkproducerande bönder ville man göra en undersökning av vilken typ av mjölkningsutrustning bönderna använde. Vissa utrustningar kräver mycket manuellt arbete medan andra är nästan helt automatiska. Man beslutar sig för att göra telefonintervjuer. Det visade sig att man fick ett bortfall på hela 60%.
Uppenbarligen har telefonintervjuer inte varit en bra lösning i detta fall. Fundera över varför detta kan ha varit fel metod samt redovisa hur bortfallsstratifieringsmetoden går till och vilken typ av intervjumetod som skulle lämpa sig bra i detta problem. 3p

4

Älgstammen i ett visst landskap har sjunkit på senare år. För att uppskatta hur stor stammen är idag så drogs ett OSU utan återläggning om 4 delområden. Landskapet har delats upp i totalt 70 naturliga delområden som huvudsakligen definieras av skogar. I de 4 områdena räknades antalet älgar. Men eftersom detta är en svår uppgift så togs även ett flygfoto över alla 70 delområdena. I tabellen nedan visas resultatet från de 4 områdena där T är antalet älgar man räknat från marken och M är antalet man sett på flygfotot.

Område nr i	1	2	3	4
T_i	3	1	0	5
M_i	5	2	1	5

Totala antalet älgar på flygfotot uppskattades till 210st.
Skatta med ett 95% konfidensintervall totala antalet älgar i landskapet. Använd kvotskattning. 3p

Formelblad i Surveymetodik (732G04, 732G02-C, 732G90-C)

Normalfördelningskvantiler:

$1-\alpha$	$z_{\alpha/2}$	z_α
0.90	1.645	1.28
0.95	1.96	1.645
0.99	2.576	2.326

1 Obundet slumpmässigt urval

1.1 Urvalsvarians(er)

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

$$s^2 = \frac{n}{n-1}\cdot p\cdot(1-p).$$

$$\sum x^2 = (n-1)\cdot s^2 + n\cdot(\bar{x})^2$$

1.2 Konfidensintervall

$$\mu : \quad \bar{x} \pm z_{\alpha/2}\cdot\sqrt{\frac{s^2}{n}\cdot\left(1-\frac{n}{N}\right)}$$

$$\tau : \quad N\cdot\bar{x} \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{s^2}{n}\cdot\left(1-\frac{n}{N}\right)}$$

$$P : \quad p \pm z_{\alpha/2}\cdot\sqrt{\frac{p(1-p)}{n-1}\cdot\left(1-\frac{n}{N}\right)}$$

Ändlighetskorrektionen $(1-\frac{n}{N})$ kan utelämnas om $\frac{n}{N} \leq 0.05$.

1.3 Urvalsdimensionering

$$n_0 \geq \frac{4\cdot(z_{\alpha/2})^2\cdot\sigma^2}{B^2}$$

$$n_0 \geq \frac{4\cdot(z_{\alpha/2})^2\cdot P(1-P)}{B^2}$$

$$n = \frac{n_0}{1+\frac{n_0-1}{N}}$$

Samband mellan konfidensintervallbredder: $B_\tau = N\cdot B_\mu$

1.4 Redovisningsgrupper

Punktskattningar

$$\begin{aligned}
\bar{x_1} &= \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i} \\
s_1^2 &= \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_{1i}-\bar{x_1})^2 = \frac{1}{n_1-1}\left(\sum_{i=1}^{n_1} x_{1i}^2 - n_1\cdot(\bar{x_1})^2\right)
\end{aligned}$$

Konfidensintervall

$$\begin{aligned}
\mu_1 &: \bar{x_1} \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)} \\
P_1 &: p_1 \pm z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1-1}\left(1-\frac{n_1}{N_1}\right)} \\
\tau_1 &: N\cdot\bar{x}' \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{s'^2}{n}\left(1-\frac{n}{N}\right)} \\
A_1 &: N\cdot p' \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{p'(1-p')}{n-1}\left(1-\frac{n}{N}\right)}
\end{aligned}$$

där

$$x_i' = \begin{cases} x_i & \text{om elementet tillhör redovisningsgruppen} \\ 0 & \text{annars} \end{cases}$$

och

$$\begin{aligned}
\bar{x}' &= \frac{1}{n}\sum_{i=1}^{n} x_i' \\
s'^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i'-\bar{x}')^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i')^2 - n\cdot(\bar{x}')^2\right) \\
p' &= \frac{\text{Antal element från redovisningsgruppen i urvalet med} x=1}{n}
\end{aligned}$$

1.5 Ramproblem

Konfidensintervall vid övertäckning

$$\begin{aligned}
\mu_0 &: \quad \bar{x_0} \pm z_{\alpha/2}\cdot\sqrt{\frac{s_0^2}{n_0}\cdot\left(1-\frac{n}{N}\right)} \\
P_0 &: \quad p_0 \pm z_{\alpha/2}\cdot\sqrt{\frac{p_0(1-p_0)}{n_0-1}\cdot\left(1-\frac{n}{N}\right)}
\end{aligned}$$

För totaler:

$$\widehat{N_0}\cdot\bar{x_0} \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{\frac{1}{n-1}\cdot\left(\sum_{i=1}^{n_0} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n_0} x_i\right)^2\right)}{n}\cdot\left(1-\frac{n}{N}\right)}.$$

$$\widehat{N_0}\cdot p_0 \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{\frac{n_0\cdot p_0}{n}\cdot\left(1-\frac{n_0\cdot p_0}{n}\right)}{n-1}\cdot\left(1-\frac{n}{N}\right)}.$$

Replikatproblem

$$\widehat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{M_i/N}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n}\frac{x_i}{M_i}}{\sum_{i=1}^{n}\frac{1}{M_i}}$$

2 Stratifierat urval

2.1 Punktskattningar:

$$\begin{aligned}
\hat{\mu} &= \bar{x}_{st} = W_1\cdot\bar{x}_1 + W_2\cdot\bar{x}_2 + \ldots + W_L\cdot\bar{x}_L = \sum_{i=1}^{L} W_i\cdot\bar{x}_i \\
\widehat{\tau} &= N\cdot\bar{x}_{st} \\
\widehat{P} &= p_{st} = W_1\cdot p_1 + W_2\cdot p_2 + \ldots + W_L\cdot p_L = \sum_{i=1}^{L} W_i\cdot p_i
\end{aligned}$$

2.2 Konfidensintervall

$$\begin{aligned}
\mu &: \quad \bar{x}_{st} \pm z_{\alpha/2}\cdot\sqrt{\sum_{i=1}^{L} W_i^2\cdot\frac{s_i^2}{n_i}\cdot\left(1-\frac{n_i}{N_i}\right)} \\
\tau &: \quad N\cdot\bar{x}_{st} \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\sum_{i=1}^{L} W_i^2\cdot\frac{s_i^2}{n_i}\cdot\left(1-\frac{n_i}{N_i}\right)} \\
P &: \quad p_{st} \pm z_{\alpha/2}\cdot\sqrt{\sum_{i=1}^{L} W_i^2\cdot\frac{p_i\cdot(1-p_i)}{n_i-1}\cdot\left(1-\frac{n_i}{N_i}\right)} \\
A &: \quad N\cdot p_{st} \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\sum_{i=1}^{L} W_i^2\cdot\frac{p_i\cdot(1-p_i)}{n_i-1}\cdot\left(1-\frac{n_i}{N_i}\right)}
\end{aligned}$$

2.3 Proportionell allokering

$$n_i = n\cdot W_i = n\cdot\frac{N_i}{N}$$

$$n_0 \geq \frac{4\cdot(z_{\alpha/2})^2\cdot\sum_{i=1}^{L} W_i\cdot\sigma_i^2}{B^2}$$

$$n_0 \geq \frac{4\cdot(z_{\alpha/2})^2\cdot\sum_{i=1}^{L} W_i\cdot P_i\cdot(1-P_i)}{B^2}$$

$$n = \frac{n_0}{1+\frac{n_0-1}{N}}$$

2.4 Fullständig optimal allokering

$$n_i = n \cdot \frac{W_i \cdot \sigma_i / \sqrt{c_i}}{\sum_{j=1}^{L} W_j \cdot \sigma_j / \sqrt{c_j}} = n \cdot \frac{N_i \cdot \sigma_i / \sqrt{c_i}}{\sum_{j=1}^{L} N_j \cdot \sigma_j / \sqrt{c_j}}$$

Fix kostnad:

$$n \geq \frac{(C_{max} - C_0) \cdot \sum_{i=1}^{L} (N_i \cdot \sigma_i / \sqrt{c_i})}{\sum_{i=1}^{L} (N_i \cdot \sigma_i \cdot \sqrt{c_i})}.$$

Fix konfidensintervallbredd:

$$n \geq \frac{\left(\sum_{i=1}^{L} (W_i \cdot \sigma_i \cdot \sqrt{c_i}) \right) \cdot \sum_{i=1}^{L} (W_i \cdot \sigma_i / \sqrt{c_i})}{B^2 / (4 \cdot (z_{\alpha/2})^2) + (1/N) \cdot \sum_{i=1}^{L} (W_i \cdot \sigma_i^2)}$$

2.5 Neyman-allokering

$$n_i = n \cdot \frac{W_i \cdot \sigma_i}{\sum_{j=1}^{L} W_j \cdot \sigma_j} = n \cdot \frac{N_i \cdot \sigma_i}{\sum_{j=1}^{L} N_j \cdot \sigma_j}$$

$$n \geq \frac{\left(\sum_{i=1}^{L} (W_i \cdot \sigma_i) \right)^2}{B^2 / (4 \cdot (z_{\alpha/2})^2) + (1/N) \cdot \sum_{i=1}^{L} (W_i \cdot \sigma_i^2)}$$

Vid 0/1–data gäller $\sigma_i = \sqrt{P_i \cdot (1 - P_i)}$.

3 Enstegs klusterurval

3.1 OSU av kluster

Väntevärdesriktig skattning

$$\hat{\mu}_u = \frac{N}{M_0} \cdot \overline{T}$$

$$\text{K.I.:} \quad \hat{\mu}_u \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{N}{M_0}\right)^2 \cdot \frac{s_T^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$$

$$\widehat{\tau} = M_0 \cdot \hat{\mu}_u$$

$$\text{K.I.:} \quad M_0 \cdot \hat{\mu}_u \pm M_0 \cdot z_{\alpha/2} \cdot \sqrt{\left(\frac{N}{M_0}\right)^2 \cdot \frac{s_T^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$$

Kvotskattning, $\hat{\mu}_R = \overline{x}_{cl}$

$$\overline{x}_{cl} = \frac{\sum_{i=1}^{n} T_i}{\sum_{i=1}^{n} M_i}$$

$$\widehat{Var(\overline{x}_{cl})} = \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n \cdot (n-1)} \cdot \frac{1}{(\overline{M})^2} \cdot \left(\sum_{i=1}^{n} T_i^2 + (\overline{x}_{cl})^2 \cdot \sum_{i=1}^{n} M_i^2 - 2 \cdot \overline{x}_{cl} \cdot \sum_{i=1}^{n} T_i \cdot M_i\right)$$

$$\text{K.I.} \quad : \quad \overline{x}_{cl} \pm z_{\alpha/2} \cdot \sqrt{\left(1 - \frac{n}{N}\right) \cdot \frac{1}{n \cdot (n-1)} \cdot \frac{1}{(\overline{M})^2} \cdot \left(\sum_{i=1}^{n} T_i^2 + (\overline{x}_{cl})^2 \cdot \sum_{i=1}^{n} M_i^2 - 2 \cdot \overline{x}_{cl} \cdot \sum_{i=1}^{n} T_i \cdot M_i\right)}$$

3.2 *PPS*–urval av kluster

$$\hat{\mu}_{pps} = \frac{1}{n} \sum_{i=1}^{n} \overline{x}_i$$

$$\text{K.I.} \quad \overline{x}_{pps} \pm z_{\alpha/2} \cdot \sqrt{\frac{1}{n \cdot (n-1)} \cdot \left(\sum_{i=1}^{n} (\overline{x}_i)^2 - n \cdot (\overline{x}_{pps})^2\right)}$$

$$\hat{p}_{pps} = p_{pps} \frac{1}{n} \sum_{i=1}^{n} p_i$$

$$\text{K.I.} \quad p_{pps} \pm z_{\alpha/2} \cdot \sqrt{\frac{1}{n \cdot (n-1)} \cdot \left(\sum_{i=1}^{n} (p_i)^2 - n \cdot (p_{pps})^2\right)}$$

3.3 Urvalsdimensionering

pps–**urval**

$$n \geq \frac{4 \cdot (z_{\alpha/2})^2 \cdot \sigma_{\overline{x}_i}^2}{B^2}$$

4 Tvåstegs klusterurval

4.1 OSU av kluster

Väntevärdesriktiga skattningar

$$\hat{\mu}_{u2} = \left(\frac{N}{M_0}\right) \cdot \hat{\overline{T}} = \left(\frac{N}{M_0}\right) \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} M_i \cdot \bar{x}_i$$

$$\hat{P}_{u2} = \left(\frac{N}{M_0}\right) \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} M_i \cdot p_i$$

Kvotskattningar

$$\hat{\mu}_{R2} = \frac{\sum_{i=1}^{n} \hat{T}_i}{\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n} M_i \cdot \bar{x}_i}{\sum_{i=1}^{n} M_i}$$

$$\hat{P}_{R2} = \frac{\sum_{i=1}^{n} M_i \cdot p_i}{\sum_{i=1}^{n} M_i}$$

4.2 *PPS*–urval

$$\hat{\mu}_{pps2} = \frac{1}{n} \cdot \sum_{i=1}^{n} \bar{x}_i$$

$$\widehat{Var(\hat{\mu}_{pps2})} = \frac{s^2_{\bar{x}_i}}{n} = \frac{1}{n \cdot (n-1)} \cdot \left(\sum_{i=1}^{n} (\bar{x}_i)^2 - n \cdot (\bar{x}_{pps})^2\right)$$

5 Bortfallsstratumansatsen

Punktskattningar och konfidensintervall

$$\begin{aligned} \hat{\mu}_{st} &= \frac{n'_S}{n} \cdot \bar{x}_S + \frac{n'_B}{n} \cdot \bar{x}_B \\ \hat{P}_{st} &= \frac{n'_S}{n} \cdot p_S + \frac{n'_B}{n} \cdot p_B \end{aligned}$$

$$\hat{\mu}_{st} \pm z_{\alpha/2} \cdot \sqrt{\frac{n'_S}{n} \cdot \frac{s_S^2}{n} + \left(\frac{n'_B}{n}\right)^2 \cdot \frac{s_B^2}{n_B} + \frac{1}{n} \cdot \left(\frac{n'_S}{n} \cdot (\bar{x}_S - \hat{\mu}_{st})^2 + \frac{n'_B}{n} \cdot (\bar{x}_B - \hat{\mu}_{st})^2\right)}$$

$$\hat{P}_{st} \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{n'_S}{n}\right)^2 \cdot \frac{p_S \cdot (1-p_S)}{n'_S - 1} + \left(\frac{n'_B}{n}\right)^2 \cdot \frac{p_B \cdot (1-p_B)}{n_B - 1} + \frac{1}{n} \cdot \left(\frac{n'_S}{n} \cdot (p_S - \hat{P}_{st})^2 + \frac{n'_B}{n} \cdot (p_B - \hat{P}_{st})^2\right)}$$

Slutligt bortfall

$$\begin{aligned} b &= (\text{Bortfallsandel i fas 1}) \times (\text{Bortfallandel i fas 2}) \\ &= \frac{n'_B}{n} \cdot \frac{n''_B}{n_B} \end{aligned}$$

där n''_B =antal bortfall i fas 2–urvalet.