Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2012-08-14
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER2
Tid	14-18
Kurskod	732G04
Provkod	TENC
Kursnamn/benämning **Provnamn/benämning**	Surveymetodik Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Karl Wahlin
Telefon under skrivtiden	0709-719096
Besöker salen ca kl.	
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	carita.lilja@liu.se tel 1463
Tillåtna hjälpmedel	Valfri räknedosa, till tentamen vidhäftad formelsamling.
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt
Antal exemplar i påsen	

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

Kurskod:	732G19 och 732G04
Datum och tid:	2012-08-14, 14-18
Jourhavande lärare:	Kalle Wahlin
Tillåtna hjälpmedel:	Valfri räknedosa, till tentamen vidhäftad formelsamling.
Betygsgränser:	Tentamen omfattar totalt 20p. Godkänt från 12p, väl godkänt från 16p.
	Siffrorna i uppgifterna är påhittade.

Redovisa och motivera tydligt alla dina lösningar!

Uppgift 1 (9p)

En trafikskola bedriver verksamhet på fem orter. Man vill undersöka hur stor andel av trafikskolans elever som tagit körkort under sin första uppkörning efter genomgången utbildning, och drar ett stickprov bland de körskoleelever som avslutat utbildningen. Stickprovsstorlek, antalet elever som klarat uppkörningen på första försöket samt totalantalet körskoleelever under de senaste fem åren framgår ur följande tabell.

Ort	Totalantal elever	Stickprovsstorlek	Antal som klarat uppkörningen på första försöket
A	2500	500	140
B	3000	500	98
C	14000	500	249
D	6000	500	76
E	1500	500	195

a) Förklara hur ett stratifierat urval går till och varför det kan vara lämpligt att stratifiera på ort i detta fall. (1p)

b) Skatta andelen som klarat uppkörningen på första försöket bland de som genomgått utbildningen med 99 procents konfidensnivå. (3p)

c) Om man vill göra om undersökningen nästa säsong hur stora urval bör man då dra ur varje stratum om man använder Neymanallokering. Låt den totala urvalsstorleken vara 2500 även denna gång. Anta vidare att det totala antalet elever på varje ort är desamma som detta år och att andelen som klarat uppkörningen på första försöket är ungefär densamma. (3p)

d) Skatta med 95 procents konfidensnivå det totala antalet elever som klarat uppkörningen på första försöket vid ort C. (2p)

Uppgift 2 (5p)

Trafikskolan vill kartlägga vad uppkörningsledarna har för inställning till företaget. Var och en av de fem orterna har ett uppkörningskontor, och man väljer genom PPS-urval ut tre av dessa orter. För vart och ett av kontoren undersöker man samtliga de anställda uppkörningsledarna med avseende på deras inställning till trafikskolan.

Kontor på ort	Antalet undersökta uppkörningsledare	Antal uppkörningsledare som är positivt inställda till trafikskolan
A	4	3
C	6	4
D	2	1

Skatta med konfidensnivån 95 procent andelen uppkörningsledare på de fem orterna som är positivt inställda till trafikskolan.

Uppgift 3 (3p)

Trafikskolan vill även undersöka hur det går för privatister som kör upp. Bland det stora antalet privatister som gjort uppkörning på de fem orter som trafikskolan verkar visar ett slumpmässigt urval om 500 privatister att 64 av dessa klarade uppkörningen på första försöket. Bilda ett 95-procentigt konfidensintervall för andelen privatister som klarar uppkörningen på första försöket.

Uppgift 4 (3p)

a) Vilken typ av information bör ingå i ett introduktionsbrev vid en enkätundersökning? (1p)
b) I ett urval om 200 personer erhölls svar från 139. Bland de återstående gjordes ett urval om 50 personer, varav svar erhölls från 21. Vad blev det slutliga bortfallet? (1p)
c) Ge ett exempel på när det kan vara av nytta att skatta *totalmängd*. (1p)

Formelblad i Surveymetodik ht 2007 (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

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

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

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

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

$$\begin{aligned}\mu_0 &: \ \bar{x}_0 \pm z_{\alpha/2}\cdot\sqrt{\frac{s_0^2}{n_0}\cdot\left(1-\frac{n}{N}\right)} \\ P_0 &: \ p_0 \pm z_{\alpha/2}\cdot\sqrt{\frac{p_0(1-p_0)}{n_0-1}\cdot\left(1-\frac{n}{N}\right)}\end{aligned}$$

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

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

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

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

3.2 *PPS*–urval av kluster

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

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

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

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

3.3 Urvalsdimensionering

pps–**urval**

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

4 Tvåstegs klusterurval

4.1 OSU av kluster

Väntevärdesriktiga skattningar

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

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

Kvotskattningar

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

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

4.2 *PPS*–urval

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

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

5 Bortfallsstratumansatsen

Punktskattningar och konfidensintervall

$$\hat{\mu}_{st} = \frac{n'_S}{n} \cdot \bar{x}_S + \frac{n'_B}{n} \cdot \bar{x}_B$$

$$\hat{P}_{st} = \frac{n'_S}{n} \cdot p_S + \frac{n'_B}{n} \cdot p_B$$

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

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

Slutligt bortfall

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

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