Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2011-08-18
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och ringa in vilken sal som avses	TER2
Tid	8-12
Kurskod	732G19
Provkod	TENT
Kursnamn/benämning **Provnamn/benämning**	Utredningskunskap I 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 1463
Tillåtna hjälpmedel	Valfri räknedosa samt bifogad formelsamling
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt
Antal exemplar i påsen	

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

Kurskod och namn:	732G19 Utredningskunskap I, samt 732G04 Surveymetodik
Datum och tid:	2011-08-18, 8-12
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 (6p)

Ett värdetransportföretag har verksamhet fördelat på tre områden: normaltransporter, säkerhetstransporter och iltransporter. Under ett år har antalet transportuppdrag fördelat sig enligt följande.

	Normal	Säkerhet	Il
Antal transporter	2340	480	950

Till varje uppdrag finns en så kallad transportlogg, där all information samlas. Man vill undersöka andelen transporter det aktuella året, där föreskrivna rutiner ej följts (vilket kan utläsas ur loggen). Man tror sig veta att detta sker i cirka 10% av alla normaltransporter, i runt 2% av alla säkerhetstransporter och i runt 15% av alla iltransporter.

a) Hur stort urval av loggar ska göras för att ett 95% konfidensintervall för andelen transporter där föreskrivna rutiner ej följts ska få bredden högst 0.10? Använd mesta möjliga information i dina beräkningar. (3p)

b) Antag att man gjort ett proportionellt urval omfattande 200 transporter och därvid fått följande resultat:

Transportgrupp	Antal transporter där rutiner ej följts
Normal	14
Säkerhet	0
Il	6

Beräkna ett 95% konfidensintervall för andelen transporter där rutiner ej följts. (3p)

Uppgift 2 (6p)

På en skola vill man undersöka elevernas matvanor på fritiden och bland annat uppskatta hur många pizzor, μ, en slumpmässigt vald elev vid skolan äter per vecka. Man valde slumpmässigt ut 5 klasser bland de totalt 24 klasserna på skolan och i varje klass fick varje elev svara på frågan om hur många pizzor man ätit den senaste veckan. Resultatet var

Klass (i)	Antal elever (M_i)	Totalt antal ätna pizzor (T_i)
1	22	12
2	24	16
3	27	19
4	26	10
5	21	14

a) Beräkna ett 95% konfidensintervall för μ. (3p)

b) I efterhand visar det sig att den som gjorde urvalet av klasser gjorde detta med återläggning och med dragningssannolikheter proportionella mot klasstorlek. Beräkna med utgångspunkt från detta ett nytt 95% konfidensintervall för μ. (3p)

Uppgift 3 (4p)

Genomsnittslönen för förstajobbet bland de utexaminerade från en viss utbildning ett år har baserat på ett urval om 37 personer beräknats till 29 tkr, med en standardavvikelse om 4.3 tkr.

a) Beräkna ett 95% konfidensintervall för genomsnittslönen för förstajobbet bland de utexaminerade från utbildningen om antalet utexaminerade från utbildningen det studerade året var 400 personer. (2p)

b) Beräkna ett 95% konfidensintervall för genomsnittslönen för förstajobbet bland de utexaminerade från utbildningen om antalet utexaminerade från utbildningen det studerade året var 80 personer. (2p)

Uppgift 4 (4p)

Bedöm för vart och ett av följande påståenden om det är sant eller falskt.

a) Optimal allokering bör användas om spridningen varierar mycket mellan olika strata.
b) Vid användning av bortfallsstratumansatsen kan man inte få ned det slutliga bortfallet till 0%.
c) Vid klusterurval finns inga väntevärdesriktiga skattningar av populationsandelar.
d) Imputering innebär att en person som ej vill vara med i en undersökning ersätts med en ny person som accepterar att vara det.

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

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

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

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

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

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}$$

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)^2 = \frac{1}{n_1-1}\left(\sum_{i=1}^{n_1} x_{1i}^2 - n_1\cdot(\bar{x}_1)^2\right)$$

Konfidensintervall

$$\mu_1 \quad : \quad \bar{x}_1 \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)}$$

$$P_1 \quad : \quad p_1 \pm z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1-1}\left(1-\frac{n_1}{N_1}\right)}$$

$$\tau_1 \quad : \quad N\cdot\bar{x}' \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{s'^2}{n}\left(1-\frac{n}{N}\right)}$$

$$A_1 \quad : \quad N\cdot p' \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{p'(1-p')}{n-1}\left(1-\frac{n}{N}\right)}$$

där

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

och

$$\bar{x}' = \frac{1}{n}\sum_{i=1}^{n} x_i'$$

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

$$p' = \frac{\text{Antal element från redovisningsgruppen i urvalet med } x=1}{n}$$

1.5 Ramproblem

Konfidensintervall vid övertäckning

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

$$P_0 \quad : \quad p_0 \pm z_{\alpha/2}\cdot\sqrt{\frac{p_0(1-p_0)}{n_0-1}\cdot\left(1-\frac{n}{N}\right)}$$

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

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

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

2 Stratifierat urval

2.1 Punktskattningar:

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

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

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

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

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

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

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

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

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

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

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

5 Bortfallsstratumansatsen

Punktskattningar och konfidensintervall

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

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

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

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

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.