Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2012-10-06
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och ringa in vilken sal som avses	TER1
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 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:	2012-10-06, 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)

a) Man planerar för en enkätundersökning av en population som består av 4000 element. Undersökningen ska genomföras som ett OSU, och man accepterar en bredd om 5 procentenheter för ett 90-procentigt konfidensintervall för andelen ja-svarare i en viss viktig ja/nej-fråga i undersökningen. Hur stort stickprov bör man dra? (2p)

b) Antag att man genomfört enkätundersökningen med stickprov dimensionerat enligt ditt svar i a), och att samtliga besvarade enkäten. 80% svarade ja på frågan. Bestäm ett 90% konfidensintervall för andelen ja-svarare i populationen. (2p)

c) Antag att undersökningen genomfördes med stickprov dimensionerat enligt ditt svar i a), men att endast 80% besvarade enkäten, och bland dessa svarade 70% ja på frågan. Bland de som inte svarade gjordes ett nytt urval om 20%, vilka kontaktades på ett annat sätt. Samtliga av dessa besvarade frågan och 90% svarade ja. Bestäm en punktskattning av andelen ja-svar i populationen. (2p)

Uppgift 2 (5p)

Ett fackförbund vill undersöka hur lokalavdelningarna ute i landet ställer sig till ett visst förslag. Man drar ett OSU om 5 av företagets totalt 25 lokalavdelningar och undersöker samtliga anställda vid dessa avdelningar. Följande resultat erhålles.

Avdelning	Antal förtroendevalda	Antal som stödjer förslaget
1	41	23
2	39	25
3	35	25
4	45	27
5	25	17

a) Förklara vilken undersökningsdesign som använts och varför den är lämplig här. (1p)

b) Bestäm en punktskattning av andelen förtroendevalda som stödjer förslaget. (1p)

c) Bestäm ett 95% konfidensintervall för andelen förtroendevalda som stödjer förslaget. (3p)

Uppgift 3 (5p)

Ett företag har verksamhet på tre orter. Man vill från ledningsnivå undersöka den genomsnittliga arbetstiden inom företaget, och man drar därför ett slumpmässigt urval av anställda från respektive ort. Informationen sammanställs i följande tabell.

Ort	Antal anställda	Antal undersökta	Medelarbetstid per månad	Standardavvikelse
A	100	15	158	12
B	200	10	165	21
C	300	20	173	14

a) Förklara vilken urvalsdesign som valts och varför den är lämplig här. (1p)
b) Beräkna ett 95-procentigt konfidensintervall för den genomsnittliga arbetstiden i hela företaget. (3p)
c) Hur många personer skulle ha valts ut ur respektive stratum om man använt proportionell allokering? (1p)

Uppgift 4 (4p)

a) När lämpar sig i allmänhet systematiskt urval bättre än OSU? (1p)
b) När lämpar sig i allmänhet PPS-urval bättre än OSU? (1p)
c) Vad är en *redovisningsgrupp*? (1p)
d) 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

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

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

Konfidensintervall

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

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

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

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

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

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

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

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

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

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

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

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

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

5 Bortfallsstratumansatsen

Punktskattningar och konfidensintervall

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

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

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

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.